



Sistema Gestor de Documentos Electrónicos para la Automatización de la Gestión de Reservas

PROYECTO DE FIN DE CARRERA
CARLOS MARTIN MARTINEZ

Tabla de contenido

1.	Introducción.....	6
2.	Estado del Arte	8
2.1.	Estructura de un Sistema Gestor de Documentos Electrónicos	8
2.1.1.	Tecnologías o sistemas de Digitalización	9
2.1.2.	Tecnologías de Servicios de Documento o Librería.....	9
2.1.3.	Tecnologías de Workflow	10
2.1.4.	Procesamiento de Formularios	11
2.1.5.	Aplicaciones de Gestión de Registros	11
2.1.6.	Captura de Documentos y Contenido	11
2.2.	SGDE para la Automatización de Gestión de Reservas	13
2.2.1.	Situación de los mercados de Gestión Documental y Gestión de Procesos 14	
2.2.1.1.	Alfresco Software Inc.	15
2.2.1.1.1.	Descripción.....	15
2.2.1.1.2.	Productos y Soluciones.....	15
2.2.1.2.	IBM	16
2.2.1.2.1.	Descripción.....	16
2.2.1.2.2.	Productos y Soluciones.....	16
2.2.1.3.	Lexmark.....	16
2.2.1.3.1.	Descripción.....	16
2.2.1.3.2.	Productos y soluciones	17
2.2.1.4.	OpenText.....	17
2.2.1.4.1.	Descripción.....	17
2.2.1.4.2.	Productos y Servicios.....	18
2.2.1.5.	Oracle	18
2.2.1.5.1.	Descripción.....	18
2.2.1.5.2.	Productos y servicios	18
3.	Justificación y descripción detallada de la solución elegida.....	19
3.1.	Solución de múltiples proveedores frente a solución de proveedor único 20	
3.2.	Riesgos más frecuentes en la implantación de un proyecto Software	20
3.2.1.	Análisis de Riesgos	21
3.2.2.	Conclusión del análisis: elección de Oracle como proveedor.....	23
3.2.3.	Descripción a alto nivel de la solución de Oracle	24
3.3.	Oracle Fusion Middleware.....	25
3.4.	Oracle WebLogic Server.....	25
3.4.1.	Dominios de WebLogic.....	26
3.5.	Oracle WebCenter Content.....	28
3.5.1.	WebCenter Enterprise Capture.....	29
3.5.1.1.	Descripción General.....	29
3.5.1.2.	Extensibilidad	32
3.5.2.	WebCenter Imaging.....	32
3.5.2.1.	Consola de Imaging	33
3.5.2.2.	WebCenter Imaging Input Agent.....	36
3.5.2.3.	Extensibilidad de WebCenter Imaging	37
3.5.2.4.	URL Tools	38
3.5.2.5.	Anexos Gestionados.....	38
3.6.	Unified Business Process Management Suite	39
3.6.1.	Business Process Composer.....	41
3.6.2.	BPM Studio	43
3.6.3.	Business Activity Monitoring	44
3.7.	WebCenter Forms Recognition.....	44
3.7.1.	WebCenter Forms Recognition Designer.....	45
3.7.2.	WebCenter Forms Recognition Verifier	48

3.7.3. WebCenter Forms Recognition Runtime Service	50
3.8. Solución de Gestión Documental para la Automatización de la Gestión de Reservas	53
4. Requisitos	57
5. Arquitectura	60
5.1 Dimensionamiento predictivo	60
5.1.1 Enterprise Capture	60
5.1.2. WebCenter Content	61
5.1.3. WebCenter Forms Recognition	64
5.1.4. Unified Business Process Management Suite	65
5.2. Alta Disponibilidad y Balanceo de Carga	65
5.2. Arquitectura física	66
5.3. Arquitectura lógica	69
5.3.1 Dominio WebLogic y Esquemas para WebCenter Content	69
5.3.2. Dominio WebLogic y Esquemas para BPM Suite	70
5.3.3. Arquitectura lógica WebCenter Forms Recognition	71
6. Instalación y Configuración de los Productos de Oracle	72
6.1 Instalación de Oracle Fusion Middleware	72
6.1.1. Comprobar prerrequisitos	72
6.1.2. Descargar Software Necesario	73
6.1.3. Instalar Oracle WebLogic Server	74
6.1.4. Instalación de Oracle HTTP Server	75
6.1.5. Instalación de productos FMW	76
6.1.5.1 Instalación de WebCenter Content	76
6.1.5.2. Instalación de SOA y BPM Suite	77
6.1.6. Creación de Dominios WebLogic	78
6.1.6.1 Creación del Dominio WebCenter Content	78
6.1.6.2 Creación del Dominio BPM Suite	80
6.1.7. Tareas de post-configuración comunes	80
6.1.7.1 Configuración de Oracle HTTP Server	80
6.1.7.2 Configuración de Node Manager	81
6.1.7.3 Configuración de Microsoft Active Directory como proveedor de identidad	83
6.1.8. Tareas de post-configuración específicas	83
6.2. Instalación de WebCenter Forms Recognition	84
6.3 Configuración de Enterprise Capture	85
6.3.1. Metadatos a capturar	85
6.3.2. Perfil de Documento	86
6.3.3 Procesamiento automático	86
6.3.3.1 Trabajo de Importación de Email	87
6.3.3.2 Trabajo de Conversión a TIFF	89
6.3.3.3 Trabajo de Confirmación y paso a WebCenter Forms Recognition	90
6.4 Configuración de Forms Recognition	92
6.4.3. Clasificación	92
6.4.3.1. Pre-configuración de OCR	92
6.4.3.2. Estructura y motores de clasificación	93
6.4.3. Extracción de campos	98
6.4.4. Configuración de Runtime Service	104
6.5 Configuración de WebCenter Imaging	106
6.5.1 Configuración de Conexiones	106
6.5.2 Aplicación de Reservas	107
6.5.3 Entrada automatizada	109
7. Ejecución de la Confirmación de Reservas	111
7.1. Proceso Manual	111
7.2. Proceso Automático	111

8. Costes y Gestión del Proyecto	117
8.1. Licenciamiento	117
8.2. Plan de proyecto	118
9. Referencias	119
Tabla 1 Componentes de un SGDE genérico	8
Tabla 2 Explicación del cuadrante Mágico de Gartner	14
Tabla 3 Resumen Cuadrante Mágico 2014 para ECM e iBPMS	14
Tabla 4 Comparativa de proveedores	19
Tabla 5 Ventajas e inconvenientes de las soluciones de proveedor único	20
Tabla 6 Dimensiones del riesgo y sus frecuencias	22
Tabla 7 Riesgos y sus frecuencias.....	22
Tabla 8 Integración vertical de IBM, OpenText y Oracle	23
Tabla 9 Portfolio de Oracle Fusion Middleware	25
Tabla 10 Arquitectura esquematizada de Oracle Fusion Middleware	25
Tabla 11 Comparativa resultados SPECjEnterprise2010 de Oracle vs IBM	26
Tabla 12 Resumen de componentes de la solución de Oracle	54
Tabla 13 Formato de captura de requisitos	57
Tabla 14 Tamaños Máquinas Virtuales.....	59
Tabla 15 Dimensionamiento predictivo de Enterprise Capture.....	61
Tabla 16 Dimensionamiento predictivo de WebCenter Content	62
Tabla 17 Dimensionamiento esquema de base de datos de WebCenter Content.....	63
Tabla 18 Comparativa tamaño de Base de Datos	63
Tabla 19 Tamaño de Base de Datos con compresión activada	63
Tabla 20 Dimensionamiento predictivo Forms Recognition.....	64
Tabla 21 Selección máquina virtual para Forms Recognition.....	64
Tabla 22 Software Necesario	73
Tabla 23 Detalles configuración dominio wcc_domain	80
Tabla 24 Detalles configuración dominio bpm_domain	80
Tabla 25 Parámetros del Trabajo de Importación de Enterprise Capture	89
Tabla 26 Parámetros del Trabajo de Conversión de Enterprise Capture	89
Tabla 27 Parámetros del Trabajo de Confirmación de Enterprise Capture	91
Tabla 28 Motores OCR Disponibles en Forms Recognition	93
Tabla 29 Motores de Clasificación disponibles en Forms Recognition.....	95
Tabla 30 Técnicas de Extracción de Campos disponibles en Forms Recognition	99
Tabla 31 Configuración Fichero Exportado por Forms Recognition	103
Tabla 32 Parámetros de Runtime Service de Forms Recognition.....	105
Tabla 33 Parámetros de conexiones en WebCenter Imaging	107
Tabla 34 Parámetros de la Aplicación de Reservas en WebCenter Imaging.....	108
Tabla 35 Parámetros de la Entrada Automatizada en WebCenter Imaging.....	110
Tabla 36 Información Correos Electrónicos Entrantes.....	111
Tabla 37 Adjuntos de los Correos Electrónicos	112
Tabla 38 Costes de las licencias requeridas para Proyecto	117
Tabla 39 Costes de servicios de implantación del Proyecto.....	118
Ilustración 1 Servidores, Clústeres y Dominios WebLogic	27
Ilustración 2 Dominio WebCenter Content Genérico	28
Ilustración 3 Cliente de Captura Enterprise Capture.....	30
Ilustración 4 Flujo de ejecución típico de Enterprise Capture	31
Ilustración 5 Consola de WebCenter Imaging.....	33
Ilustración 6 Búsquedas en WebCenter Imaging.....	33
Ilustración 7 Subida manual en WebCenter Imaging.....	34
Ilustración 8 Gestión de búsquedas en WebCenter Imaging.....	34
Ilustración 9 Gestión de aplicaciones en WebCenter Imaging	35
Ilustración 10 Gestión de entradas en WebCenter Imaging	35
Ilustración 11 Controles del visor de WebCenter Imaging	36
Ilustración 12 Flujo de ejecución típico de WebCenter Imaging	37
Ilustración 13 Diagrama de los Anexos Gestionados de WebCenter Imaging	38
Ilustración 14 Interfaz de los Anexos Gestionados de WebCenter Imaging.....	39

Ilustración 15 Dominio de BPM Suite genérico	39
Ilustración 16 BPM Composer.....	41
Ilustración 17 Procesos en BPM Composer	42
Ilustración 18 Diseño de proceso conforme al estándar BPMN en BPM Composer	43
Ilustración 19 Business Activity Monitoring	44
Ilustración 20 Forms Recognition Designer	45
Ilustración 21 Resaltado de texto en Forms Recognition Designer	46
Ilustración 22 Resaltado de bloques en Forms Recognition Designer	47
Ilustración 23 Vista de lotes en Forms Recognition Verifier	48
Ilustración 24 Formulario de Forms Recognition Verifier.....	49
Ilustración 25 Forms Recognition Runtime Service	50
Ilustración 26 Flujo de ejecución típico de Forms Recognition	52
Ilustración 27 Diagrama de alto nivel del proyecto a implantar	55
Ilustración 28 Arquitectura Física	66
Ilustración 29 Arquitectura lógica de WebCenter Content	69
Ilustración 30 Arquitectura Lógica de BPM Suite.....	70
Ilustración 31 Arquitectura Lógica Forms Recognition.....	71
Ilustración 32 Pasos de instalación de Fusion Middleware	72
Ilustración 33 Repository Creation Utility	74
Ilustración 34 Instalación JRockit.....	75
Ilustración 35 Instalación WebLogic.....	75
Ilustración 36 Instalación Oracle HTTP Server	76
Ilustración 37 Instalación WebCenter Content.....	76
Ilustración 38 Instalación BPM y SOA	77
Ilustración 39 Creación del Dominio de WebCenter Content	78
Ilustración 40 Conexión a los esquemas creados por RCU	78
Ilustración 41 Configuración de Servidores Gestionados	79
Ilustración 42 Configuración de Clústeres	79
Ilustración 43 Instalación Forms Recognition	84
Ilustración 44 Configuración seguridad en Forms Recognition	84
Ilustración 45 Activación de Forms Recognition Service como Servicio de Windows.....	85
Ilustración 46 Workspace de Enterprise Capture.....	85
Ilustración 47 Campos de Enterprise Capture	86
Ilustración 48 Perfil de documento	86
Ilustración 49 Configuración OCR en Forms Recognition	92
Ilustración 50 Jerarquía de Clasificación	93
Ilustración 51 Nueva Reserva (New Booking)	96
Ilustración 52 Cancelación (Cancellation).....	96
Ilustración 53 Corrección (Amendment).....	97
Ilustración 54 No es reserva (Void).....	97
Ilustración 55 Realce de la palabra identificada por el motor de Clasificación por Frases	98
Ilustración 56 Configuración de los campos de extracción en Forms Recognition, con los campos reconocidos por formato resaltados	100
Ilustración 57 Configuración de búsqueda asociativa.....	101
Ilustración 58 Extensión de Forms Recognition mediante Visual Basic Script.....	102
Ilustración 59 Configuración Runtime Service de Forms Recognition	104
Ilustración 60 Configuración de conexiones en WebCenter Imaging	106
Ilustración 61 Creación de la Aplicación de Reservas en WebCenter Imaging.....	107
Ilustración 62 Mapeo de campos de entrada en WebCenter Imaging.....	109
Ilustración 63 Correos electrónicos en el servidor de correo	111
Ilustración 64 Resultados de captura de Enterprise Captura	112
Ilustración 65 Documentos preparados para ser procesados por Forms Recognition	113
Ilustración 66 Runtime Service en Ejecución, procesando los 3 documentos	113
Ilustración 67 Resultados de exportación de Forms Recognition.....	114
Ilustración 68 Documentos listos para ser cargados por el Input Agent de Imaging.....	115
Ilustración 69 Los 3 documentos correctamente importados en WebCenter Imaging	115
Ilustración 70 Una de las reservas vista con el Visor de Imaging.....	115
Ilustración 71 Plan de Proyecto.....	118

1. Introducción

El presente documento supone la memoria del Proyecto de Fin de Carrera de Carlos Martín Martínez.

El sector turístico se ha revolucionado desde el auge del comercio online. Mientras que antes dominaban los mayoristas de estancias, que intermedian entre hoteles y agencias, ahora, gracias a internet, las agencias pueden vender directamente las estancias, pudiendo ofrecer por tanto precios más competitivos que si estuviera el mayorista de estancias de por medio.

Al mismo tiempo, surge un nuevo negocio conocido como Bancos de Camas, empresas con una extensa base de datos de hoteles, con sus correspondientes estancias, que ofrecen el acceso a dicha base de datos a mayoristas, agencias, buscadores y hoteles. Para mayoristas, agencias y buscadores supone acceso a un mayor número de estancias que las que podían acceder anteriormente mientras que para los hoteles ingresar en las bases de datos de los Bancos de Camas supone ser localizables por más personas a través de distintos canales, lo que finalmente se refleja en un aumento de la ocupación. [1][2][3]

La operativa habitual de estos Bancos de Camas es la siguiente:

1. El Banco de Camas recibe una solicitud de reserva a través de una agencia, buscador o mayorista.
2. Envía un correo electrónico con un documento con los detalles de dicha reserva al hotel.
3. El hotel recibe los detalles, los revisa y responde con otro correo electrónico, confirmando dicha reserva
4. El Banco de Camas actualiza su base de datos correspondientemente, quedando la reserva completamente confirmada.

Más adelante en este documento se detalla este proceso con todas sus posibles consideraciones. En este punto basta con comprender que este proceso es enteramente manual.

El proyecto presentado en esta memoria trata de automatizar los pasos 3 y 4, es decir, la actualización de la base de datos del Banco de Camas con la confirmación definitiva de una reserva, consistente en la recepción de un correo electrónico con un documento adjunto que contiene información acerca de la reserva, que deberá ser extraída automáticamente. El correo electrónico lleva un documento adjunto con información sobre la reserva que deberá ser almacenado junto con los datos extraídos y presentados a un responsable de verificar los datos en una interfaz web en la que podrá corregir los errores de extracción automática y finalmente actualizar la base de datos con la reserva confirmada.

La necesidad de implantar este proyecto viene por la probabilidad de error intrínseca a la introducción manual de datos. En ocasiones la interfaz de introducción de datos es complicada o no está refinada para evitar la introducción errores. Otras veces, incluso con todo a favor de la interfaz, errar es humano y es inevitable cometer algún fallo tras horas de realizar una tarea tan repetitiva como la lectura de documentos e introducción de datos en un formulario. [4]

En ciencia y estadística, la introducción errónea de información puede invalidar completamente los resultados de la investigación [5] mientras que, para una empresa, errores en la introducción de datos financieros o económicos puede suponer la pérdida de grandes sumas de dinero o, incluso, cometer delitos de fraude. [6]

Para un Banco de Camas, cuyo negocio gira en torno al mantenimiento actualizado de una base de datos de acomodaciones y la comercialización del acceso a la misma, errores en la introducción manual de la información de una confirmación llevan a que el Banco de Camas ofrezca información incongruente a sus clientes, indicando en un buscador, por ejemplo, que un hotel tiene más plazas disponibles de las que realmente quedan, llevando a que usuarios queden frustrados al tratar de realizar reservas sin éxito o, por el contrario, mostrar algún hotel con más ocupación que la real, privándolo de potenciales clientes. Al final, esto provocaría una pérdida de confianza en la calidad de la información que ofrece el Banco de Camas, que al ser su principal negocio tiene consecuencias directas en sus ingresos.

Por lo tanto, un sistema que ayude a asegurar la calidad de la información, reduciendo la entrada manual al automatizar la introducción de los datos es crítico para obtener ventajas competitivas frente a otros Bancos de Camas con procesos menos automatizados. Además de la automatización de la introducción de datos, contar con un repositorio centralizado en el que almacenar todos los documentos referentes a reservas permite poder recuperarlos rápidamente para resolver incidencias o corroborar información y el servidor de procesos de negocio les permite obtener indicadores de sus procesos con los que antes no contaban y así

poder optimizar dicho proceso de confirmación de reservas con el fin de obtener aún más calidad en su base de datos.

El proyecto en su totalidad trata de la implantación de un sistema robusto y escalable de digitalización de documentos con extracción automática de información de los mismos, así como un flujo de aprobación para comprobar los datos extraídos antes de introducción en una base de datos. En este documento se entra en mayor detalle en la parte de digitalización y extracción de información y se harán menciones a la parte del flujo de aprobación donde corresponda.

El sistema se compone de soluciones propietarias de Oracle para la captura y digitalización de documentos, la clasificación y extracción automática de datos aplicando técnicas de Optical Character Recognition (OCR), gestión documental y gestión de procesos. En este documento se hace un estudio del mercado en el que se engloban dichas soluciones, se comparan las soluciones líder del mercado y finalmente se justifica por qué la elección de las soluciones de Oracle. Se opta por una solución basada en componentes de software empresarial por la robustez y escalabilidad que ofrecen probada, teniendo que dedicar esfuerzo únicamente a la configuración de los mismos y la parametrización para cubrir los requisitos específicos. Se decide además por optar por obtener una solución compuesta íntegramente por componentes de un único proveedor para minimizar riesgos y complicaciones en la integración. Finalmente, se decide optar por las soluciones de Oracle por presentar una solución pre-integrada tanto horizontal, los distintos componentes entre sí, como vertical, cada componente se basa en tecnología 100% Oracle. Cada solución de Oracle se describe en detalle.

A continuación, se analizan los requisitos funcionales y técnicos para un Banco de Camas concreto y se diseña una solución basada en los componentes de Oracle descritos con anterioridad. Se explica cómo desplegar los productos que componen la solución ateniéndose a los requisitos de escalabilidad, rendimiento y disponibilidad pertinentes. Se describen las configuraciones y parametrizaciones a aplicar sobre los productos y finalmente se detalla mediante explicaciones textuales, gráficos y ejemplos de código las modificaciones realizadas para cubrir todos los requisitos.

2. Estado del Arte

La Association for Information and Image Management (AIIM) es la comunidad global de profesionales de la información que cuenta con un consejo de liderazgo ejecutivo formado por las compañías ABBY, Adobe, Box, Canon, Capgemini, Ccube Solutions, DataSolutions, DICOM DocuWare, EASY SOFTWARE, EMC², Fabasoft, FileBound, Fiserv, FUJITSU, HP,, IBM, IBML, InfoMentum, KnowledgeLake, Kodak Alaris, Kofax, Konica Minolta, Metalogix, Microsoft, Onbase by Hyland, OpenText, Oracle, Software Innovation, SPS, Systemware, Unisys, y Workshare [7]. El propósito de la AIIM es aconsejar y recomendar a las personas que dentro de sus organizaciones se responsabilizan de la gestión de la información y contenidos acerca de mejores prácticas y cómo invertir mejor su presupuesto en soluciones. Estos consejos y recomendaciones vienen principalmente en 4 formas [8]:

1. Estudios de mercado
 - Industry Watch Reports: cubren los impulsores clave y las tendencias del mercado.
 - White Papers: centrados en temas de negocio particulares y problemas del mundo real
 - ELC Trendscapes: previsiones a 18 y 24 meses para identificar qué es lo que realmente vendrá en el futuro
2. Consejos de Expertos
 - Webinars Online: educar acerca de los temas y tendencias principales que afectan a los profesionales de la información
 - Seminarios presenciales: seminarios educativos gratuitos de un día realizado en determinadas ciudades de Norte América y Gran Bretaña
 - Conferencia AIIM: la asamblea anual de los líderes en gestión de la información
3. Comunidad
 - Capítulos AIIM: 40 grupos locales alrededor del mundo
 - Red interactiva: bloggers expertos, grupos de discusión y foros sobre temas específicos
 - Estándares y *Best Practices*: aprobadas por la industria y el ANSI (American National Standards Institute)
4. Desarrollo de habilidades
 - Programas de entrenamiento: aprendizaje online o en persona acerca de Gestión de Contenidos Empresarial (ECM), Gestión de Registros Empresariales (ERM) y Gestión de Procesos de Negocio (BPM)
 - Certified Information Professional: certificación reconocida mundialmente que demuestra el amplio entendimiento de todos los activos de información y la valía de la persona certificada como administrador de la información en una organización.
 - Cursos cortos y tutoriales: ofrecidos mediante suscripción para el aprendizaje bajo demanda.

2.1. Estructura de un Sistema Gestor de Documentos Electrónicos

En el informe Analysys, Selection and Implementation of Electronic Document Management Systems (EDMS, **SGDE** por sus siglas en castellano) acerca de prácticas recomendadas de la AIIM aprobado el 5 de Junio de 2009 se describe un sistema de gestión de documentos electrónicos mediante el siguiente esquema [9]:

Digitalización	Servicios de Documento y Librería	Enrutamiento / Workflow	Procesamiento de formularios	Aplicaciones de gestión de registros
Servicios de aplicación básicos				
Servicios de Base de Datos		Controladores de dispositivos de almacenamiento		
Sistema Operativo				

Tabla 1 Componentes de un SGDE genérico

Cómo la mayoría del software, su base es un Sistema Operativo. A continuación se encuentran los servicios de Base de Datos y los controladores de los dispositivos de almacenamiento. Debido al uso de herramientas Open DataBase Connectivity (ODBC), la Base de Datos se ha convertido prácticamente en lo que se conoce como una *commodity*, es decir, que la elección de una u otra opción no aporta un gran valor diferencial para el resto del SGDE. Los controladores de los dispositivos de almacenamiento permiten conectar con la tecnología de almacenamiento como discos duros, cintas o discos de estado sólido, que se necesite para ofrecer dicho el espacio de almacenamiento disponible al SGDE (principalmente a los documentos). La tercera capa son los servicios de aplicación básicos que incluyen, entre otros, herramientas de configuración, interfaces de programador (APIs) y componentes de integración con la base de datos.

En la cuarta y última capa definida por la AIIM para los SGDE se ubican las tecnologías principales y módulos específicos de aplicaciones. Cualquier SGDE suele incorporar al menos 3 de estas tecnologías. Algunas soluciones integran algunas de estas tecnologías como la Digitalización y los Servicios de Librería en una única aplicación mientras que otras soluciones integran tecnologías en función de las necesidades de una organización.

2.1.1. Tecnologías o sistemas de Digitalización

Comúnmente por su nombre anglosajón *Imaging*, esta tecnología permite el escaneado de documentos en papel, introduciéndolos en el sistema y almacenándolos en formato digital. Una solución de digitalización suele constar de los siguientes 4 componentes.

1. Entrada o Captura: Típicamente uno o varios escáneres de una (simplex) o doble (dúplex) cara.
2. Identificación o indexación: Da la capacidad al usuario final de asignar metadatos a los documentos para poder ser localizados posteriormente.
3. Almacenamiento: Consiste en varios componentes del sistema de gestión documental o servidor de workflow que permiten almacenar, recuperar y gestionar información digital.
4. Recuperación: Consiste en la entrega de la información almacenada de acuerdo a la petición lanzada.

Para la captura de imágenes además de los escáneres, suele utilizarse alguna solución de captura que los gestione desde una consola centralizada. Estas soluciones se caracterizan por ofrecer las siguientes capacidades:

- Gestión de documentos por lotes
- Gestión y control de escáneres, incluyendo capacidades para determinar la resolución, compresión, etc.
- Determinar formatos de salida
- Comprobación de calidad que incluye monitorización de escáneres, gestión de lotes y comprobación de calidad de imágenes.

2.1.2. Tecnologías de Servicios de Documento o Librería

Permiten la gestión de documentos originalmente electrónicos. Estos servicios controlan la autoría, el ingreso y retirada de documentos (conocido como check-in/check-out) y control de versiones de documentos en elaboración o almacenados. Existen muchas similitudes entre las tecnologías de Imaging y las de Servicios de Documento. La principal diferencia radica en que los sistemas de Imaging incorporan tecnologías para el escaneado de documentos físicos lo que a su vez haga que sea más complicado la edición y el versionado de los documentos digitalizados. Las capacidades básicas de estas tecnologías permiten a usuarios autorizados:

1. Capturar, cargar o importar documentos originariamente electrónicos directamente al sistema
2. Introducir metadatos relevantes asociados al documento
3. Crear carpetas virtuales que enlacen documentos entre sí
4. Ingresar y retirar información del repositorio
5. Realizar cambios y reingresar la información modificada al repositorio
6. Establecer niveles de seguridad para los documentos

2.1.3. Tecnologías de Workflow

Proporcionan diferentes niveles de enrutamiento, auditoría y administración y se pueden dividir en 3 categorías: Administrativos, ad-hoc y producción. Las tecnologías de workflow administrativo están pensadas para organizaciones cuyos procesos cambian poco o nada; Tecnologías ad-hoc permiten redefinir ciertos procesos para determinadas tareas; Los workflows de producción combinan las características de ambas junto con amplias capacidades de auditoría y registro. Las tecnologías de workflow permiten la automatización de procesos de negocio, a través de interfaces gráficas y otras herramientas.

A día de hoy, las tecnologías de Workflow se han hecho muy populares gracias a la definición de estándares por parte de distintas empresas y organizaciones no gubernamentales. De entre todos, destacan 2 estándares, cada uno definido y mantenido por una gran organización no gubernamental.

La primera gran organización fue SGML Open, fundada en 1993. Las siglas SGML provienen de **Standard Generalized Markup Language** o “Estándar de Lenguaje de Marcado Generalizado” y es un sistema para la organización y etiquetado de documentos. En 1998 y con el fin de ampliar el foco de la organización, SGML Open pasa a denominarse OASIS, **Organization for the Advancement of Structured Information Standards**) [10]. Entre sus 4 miembros fundadores se encuentran IBM y Microsoft mientras que entre sus patrocinadores se encuentran compañías como Adobe, Alfresco, Dell, EMC, Fujitsu, HP, Liferay, NEC, NetApp, Nokia, Oracle, Red Hat, SAP, Symantec, Tibco, US Department of Defense (DoD) o VMWare. Además de fundadores y patrocinadores, OASIS cuenta con numerosos contribuidores entre los que se encuentran compañías de servicios de consultoría como Accenture, proveedoras de sistemas de digitalización como Canon, compañías de software como Google, Informatica o WSO2, fabricantes como ARM, Alcatel o Huawei. También contribuyen numerosas universidades de prestigio como el Massachusetts Institute of Technology, la Universidad Politécnica de Madrid y la Universidad Carlos III de Madrid. Así mismo, contribuyen otras organizaciones de estandarización como la anteriormente descrita AIIIM y la que se describirá a continuación, Object Management Group (OMG) [11]. El estándar definido por OASIS se conoce como **Web Services Business Process Execution Language** y se trata de un lenguaje para la orquestación de procesos de negocio basada en servicios web. WSBPEL, o BPEL, extiende la infraestructura de servicios web, utilizada para la comunicación entre aplicaciones y que aprovecha protocolos y estándares comunes como Hypertext Transport Protocol (HTTP) y eXtensible Markup Language (XML) para definir sus propios estándares como **Web Service Description Language (WSDL)** para la descripción de interfaces de servicios web y **Simple Object Access Protocol (SOAP)** para permitir a aplicaciones llamar al servicio web de otra. BPEL extiende los servicios web con la inclusión de cabeceras y etiquetas que orquestan las llamadas a servicios web entre aplicaciones e interacciones con los usuarios (generalmente para la introducción de datos) para construir los procesos de negocio.[12]

La segunda gran organización es el Object Management Group. Desde su fundación en 1989, produce y mantiene especificaciones para aplicaciones empresariales interoperables, portales y reutilizables en entornos distribuidos y heterogéneos.[13] Las categorías de membresía y algunas compañías y organizaciones que forman parte de la OMG son: [14][15]

1. Miembros Contribuyentes: miembros con derecho a voto en los comités técnicos de Dominio y de Plataforma
 1. Dell, HP, IBM, The Open Group (responsables de la famosa certificación TOGAF), W3 Consortium
2. Miembros de Dominio: Generalmente son organizaciones en determinadas industrias verticales que implementan y despliegan tecnologías internamente
 1. EMC, HL7, Microsoft, OASIS, SAP, TIBCO
3. Miembros de Plataforma: Generalmente son organizaciones que crean productos de infraestructura que son ampliamente utilizadas por aplicaciones de dominio. Típicamente son los que envían e implementan las especificaciones de la OMG
 1. CA, Fujitsu, Red Hat, Toyota
4. Miembros Influyentes: son participantes críticos y tienen voz y voto en grupos de interés especial y Task Forces
 1. Airbus, Appian, AT&T, Boeing, General Electric, Huawei, ISIS Papyrus Software, Micro Focus, Oracle, PwC, SoftwareAG
5. Miembros gubernamentales: representan agencias de gobierno y tienen los mismos privilegios que los Miembros Influyentes
 1. NASA, NIST, Secretaría de Defensa estadounidense

6. Miembros en pruebas: sólo puede asistir una persona a las reuniones como máximo y no tiene derecho a voto
 1. Mitsubishi, NEC, Toshiba
7. Miembros Universitarios: pueden ir a todas las reuniones y tienen derecho a voto n grupos de interés especial y Task Forces
 1. Tecnalía, MIT, Universidad de Cantabria, Universidad Politécnica de Madrid

Entre los numerosos estándares de modelado definidos por la OMG, cabe destacar el Unified Modeling Language (UML), la especificación más utilizada a nivel mundial para modelar la estructura, comportamiento y arquitectura de aplicaciones. Precisamente de modelado se trata el estándar que define y mantiene la OMG que ha popularizado las herramientas de Workflow. Se trata del estándar Business Process Model & Notation (BPMN) y define aquello que BPEL precisamente no define: cómo se representan gráficamente los procesos de negocio. [15]

Las soluciones de BPM más potentes hoy en día integran los estándares BPEL y BPMN, ofreciendo una interfaz gráfica basada en el estándar de notación BPMN para permitir a usuarios sin conocimientos técnicos diseñar y probar procesos de negocio. Estos procesos se ejecutan sobre un motor de ejecución de procesos basado en el estándar BPEL que ejecuta la llamada a los servicios web en el orden y la manera que se haya definido gráficamente. Los procesos en sí son uno de 4 elementos básicos que según la AIIIM conforman una solución de Workflow o BPM. Los otros 3 elementos son:

1. Colas de trabajo: Cada instancia de un proceso que alguien debe completar se conoce como trabajo. Un sistema BPM incluye una funcionalidad de colas de trabajo a las que asignar usuarios o grupos de usuarios en función de ciertas reglas automáticas o manualmente. Estos usuarios acceden a sus colas de trabajo para desempeñar su siguiente tarea.
2. Herramientas: Son las distintas formas que tienen los usuarios de avanzar el proceso. Ya sea mediante formularios, procesadores de texto, terminales, aplicaciones, etc. Se accede a estas herramientas para completar una tarea y avanzar el proceso.
3. Objetos: Con la sofisticación de las tecnologías que ahora permiten el tratamiento de información en forma de video, audio u otros formatos más allá de los meros documentos, se acuñó este término genérico para referirse a aquello sobre lo que se está trabajando en el proceso.

2.1.4. Procesamiento de Formularios

La creación y utilización de formularios electrónicos permite la recolección de datos de una manera estandarizada y automáticamente cargar estos datos a la solución SGDE. Estos formularios son típicamente creados con algún módulo de creación de los mismos o mediante editores HTML estándar. El uso de estos formularios está totalmente extendido en el mercado de las soluciones SGDE por lo que la mayoría de estas soluciones incorporan algún módulo para la creación y gestión de los mismos.

2.1.5. Aplicaciones de Gestión de Registros

Estas aplicaciones permiten la implementación de reglas asociadas al tiempo que determinados tipos de información, independientemente de si son documentos, imagen, videos etc. deben mantenerse en el sistema y el tipo de disposición que deberá ocurrir al finalizar este tiempo. Estas aplicaciones dependen de capacidades de alguna de las tecnologías mencionadas anteriormente (especialmente de las de Imaging y Documento/Librería) para el almacenamiento, indexación y recuperación de información. Sin ellas, las aplicaciones de Gestión de Registros no ofrecen más que un programa de retención automatizado.

2.1.6. Captura de Documentos y Contenido

Las tecnologías de Digitalización o Imaging y Servicios de Documento/Librería tienen numerosos elementos en común y lo que distingue una de otras es principalmente el origen de los documentos, físico o papel en el caso de Imaging y electrónico en el caso de Documentos. Por esta razón, las tecnologías de Servicios de Documentos tienen mayores capacidades de revisión debido a que es mucho menos frecuente que se actualicen los documentos digitalizados que los electrónicos en origen, principalmente por la imposibilidad de manipular el documento digitalizado. Por esta última razón, las tecnologías de Imaging incluyen

herramientas de anotación y marcado para añadir notas y marcas a los documentos digitalizados con el fin de poder someterlos a una cierta edición.

En cuanto a la captura de los documentos, los documentos en papel o faxes deben pasar a través de escáneres para poder ser digitalizados. Una vez digitalizados, pueden ser introducidos al SGDE como son introducidos los documentos originalmente electrónicos: directamente o enviados a una carpeta, bandeja de correo electrónico o servidor FTP para ser importados.

En ambos casos, el proceso de captura de documentos se puede resumir en 3 pasos, que van dando más valor de negocio a los mismos:

1. **Capturar** contenido entrante, tanto en papel como digital. Mientras que los documentos electrónicos suelen llevar ya una importante cantidad de información asociada que nos puede ayudar a su indexación como es su formato de archivo, su título u otros metadatos, así como su contenido ya electrónico, los documentos en papel requieren de un procesamiento a través de técnicas de reconocimiento óptico de caracteres.
2. **Procesar** el contenido con el fin de darle un formato adecuado y prepararlo para su indexación. En este paso se utilizan los metadatos disponibles para clasificar e indexar el documento. Nuevamente, para los documentos en papel digitalizados, podemos apoyarnos en tecnologías de reconocimiento óptico de caracteres para agilizar este proceso.
3. **Salida** de los documentos al SGDE o a una aplicación. Por defecto, si se tiene una estrategia de gestión documental que cubra todos los departamentos de una organización, el destino de cualquier documento, independientemente de su origen debería ser el SGDE, vaya a ser utilizado desde alguna otra aplicación o no. Es común que los documentos estén asociados a alguna aplicación, como facturas a una aplicación de Enterprise Resource Planning (ERP) o currículums a una aplicación de Human Capital Management (HCM). En estos casos es interesante hacer el documento accesible desde dichas aplicaciones e incluso actualizar los datos de las aplicaciones con los datos indexados del documento, para lo cual suele dispararse un proceso para validar y completar la información necesaria para la actualización de la aplicación. La implantación de un sistema BPM ayuda a gestionar correctamente estos procesos.

Estas son las tecnologías principales que permiten la captura de documentos y contenidos: [16]

1. Captura de imágenes: tecnologías que convierten documentos en papel a formato electrónico. Definidos en las tecnologías de Imaging pero recordemos sus características principales:
 1. Gestión de documentos por lotes
 2. Gestión y control de escáneres
 3. Determinar el formato de salida
 4. Verificación de calidad
2. Reconocimiento: tecnologías utilizadas para extraer datos de documentos para su indexación o como entrada para procesos de negocio. Existen distintos tipos de reconocimiento.
 1. Tecnologías que no reconocen caracteres, son 2 principalmente:
 1. OMR: de sus siglas en inglés, **O**ptical **M**ark **R**ecognition, utilizada para capturar marcas o “checks” de documentos en papel de los documentos
 2. Patch Code Recognition, utilizado generalmente para separar documentos de múltiples páginas entre sí. También pueden utilizarse y reconocerse códigos de barras para este fin. Los códigos de barras pueden proporcionar información adicional.
 2. Tecnologías que reconocen caracteres
 1. OCR: de sus siglas en inglés **O**ptical **C**haracter **R**ecognition, utilizada para capturar información generada por ordenador de los documentos.
 2. ICR: de sus siglas en inglés, **I**ntelligent **C**haracter **R**ecognition, utilizada para la captura de información manuscrita confinada en un área específica.
 3. MICR: de sus siglas en inglés, **M**agnetic **I**nk **C**haracter **R**ecognition, principalmente utilizado en cheques bancarios.

3. Tecnologías que reconocen documentos
 1. ODR: de sus siglas en inglés, **O**ptical **D**ocument **R**ecognition. También conocido como reconocimiento basado en plantillas. Las plantillas deben ser definidas a priori y le indican al sistema como identificar un formulario en concreto, qué campos reconocer y qué tipo de información contienen. Esta tecnología es la más apropiada para el procesamiento de formularios estructurados como encuestas o solicitudes.
 2. IDR: de sus siglas en inglés, **I**ntelligent **D**ocument **R**ecognition. Al contrario que en la tecnología de ODR, ésta no requiere la definición de plantillas. En su lugar, IDR tiene la capacidad de definir conjuntos de aprendizaje para cada clase de documento y sus campos. Los documentos entrantes se contrastan automáticamente con los conjuntos de aprendizaje y en base a una serie de umbrales y pesos, los documentos son clasificados y la información de sus campos extraída automáticamente. IDR se utiliza mucho en el procesamiento de cuentas a pagar, mailroom e historiales clínicos entre otros.

2.2. SGDE para la Automatización de Gestión de Reservas

Como se ha visto en el apartado anterior, una solución SGDE está compuesta de diversos módulos y éstos incorporan distintas tecnologías. Esto permite una gran flexibilidad a la hora de adaptar la solución a nuestras necesidades. En concreto, para el caso de uso de procesamiento de Reservas, se requieren los siguientes componentes:

1. **Captura:** dado que la información a procesar está contenida en documentos de diverso formato y origen, es necesario contar con un módulo que homogeneice la entrada de los mismos al sistema. Los documentos llegarán principalmente como adjuntos de correos electrónicos en formato PDF o TIFF aunque también puede haber información incluida en el propio cuerpo del email.
 - (1) PDF, Portable Document Format, en castellano, formato de documento portátil. Inventado por Adobe Systems y perfeccionado durante más de 20 años, el formato de documento portátil (PDF) es un estándar abierto para el intercambio de documentos electrónicos que mantiene la Organización Internacional de Normalización (ISO). Al convertir documentos, formularios, gráficos y páginas web a PDF, su aspecto será igual que si estuvieran impresos. Sin embargo, a diferencia de los documentos impresos, los archivos PDF pueden contener vínculos y botones en los que se puede hacer clic, campos de formulario, vídeo y audio, así como lógica para contribuir a automatizar los procesos empresariales cotidianos. Cuando compartes un archivo PDF, prácticamente cualquiera puede leerlo utilizando el software gratuito Adobe Reader® o la app para dispositivos móviles Adobe Reader. (<http://www.adobe.com/es/products/acrobat/adobepdf.html>)
 - (2) TIFF, Tagged Image File Format. Es un formato de basado en etiquetas (tags) para el almacenamiento e intercambio de imágenes. El formato TIFF es comúnmente utilizado por escáneres, capturadores de fotogramas y programas de edición de fotos. Inventado por Aldus Corporation en 1986, actualmente está en su versión 6 y pertenece a Adobe Systems. (<http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>)
2. **Reconocimiento Inteligente de Documentos (IDR):** dado que los documentos entrantes pueden hacer referencia a nuevas reservas, cancelaciones o correcciones a reservas existentes y es necesario clasificarlos automáticamente en base a esto. A continuación es necesario extraer cierta información automáticamente para actualizar la base de datos de camas disponibles. También es necesario detectar si un documento lleva información sobre una o varias reservas así como detectar documentos que no correspondan a una reserva como promociones que pueden venir adjuntas a correos electrónicos.
3. **Gestión Documental:** para almacenar todos los documentos y poder ser localizados a posteriori realizando búsquedas en el sistema SGDE y también para poder acceder a los documentos directamente desde las aplicaciones de gestión de las reservas.
4. **Gestión de Procesos:** para asegurar que todos los documentos pasan por un flujo de aprobación para comprobar la validez de su clasificación y extracción automática además de contrastar la información extraída con posibles actualizaciones externas a cualquier sistema como pudiera ser una llamada telefónica.

2.2.1. Situación de los mercados de Gestión Documental y Gestión de Procesos

Muchas organizaciones acuden a analistas como Gartner para conocer qué soluciones hay en el mercado para resolver su problemática sin tener que plantear un desarrollo completamente nuevo. Gartner se fundó en 1979, con sus oficinas centrales en Stamford Connecticut, se ha convertido en una de las organizaciones dedicadas al análisis e investigación de mercados y tendencias de tecnología de la información más relevante del mundo, si no la que más. Uno de los recursos más famosos que ofrece Gartner es el denominado Cuadrante Mágico o Magic Quadrant (MQ).

Se trata de un informe que realiza Gartner anualmente en el que comienza por definir el mercado que se está analizando y las características que se evalúan en las distintas soluciones. A continuación evalúa a los proveedores en 2 medidas:

- **Habilidad de ejecución:** es una combinación de la estabilidad financiera del proveedor y su historial de éxitos en implantaciones de las soluciones que ofrece
- **Complejidad de visión:** esta métrica corresponde principalmente a características técnicas de la solución tanto actuales como futuras

En base a estas 2 métricas Gartner construye el siguiente diagrama situando a cada proveedor en uno de los 4 cuadrantes:

	Complejidad de visión (-)		(+)
Habilidad de ejecución (+)	Competidores: su oferta no es especialmente innovadora pero tienen un importante número de referencias y/o una solidez que les permitirá innovar y convertirse en líderes pronto.	Líderes: son un referente en lo que a este mercado se refiere.	
	Jugadores de Nicho: tienen una solución dentro de este mercado pero muy enfocada a un determinado caso de uso o industria.	Visionarios: tienen grandes ideas y soluciones innovadoras pero carecen de referencias o solidez para confirmar que lo que ofrecen es realmente factible a corto y medio plazo.	
(-)			

Tabla 2 Explicación del cuadrante Mágico de Gartner

Para encontrar los proveedores que tienen soluciones para los 4 componentes del SGDE para la Automatización de la Gestión de Reservas se debe contrastar el último informe de Gartner para el mercado de Gestión de Contenidos Empresarial (ECM) [17] que incluye soluciones de Captura, Reconocimiento Inteligente de Documentos y Gestión Documental; y el de Suites de Gestión de Procesos Inteligentes (iBPMS) [18] que incluye soluciones para la Gestión de Procesos.

ECM		iBPMS	
Competidores: HP	Líderes: EMC, Hyland (SAP), IBM , Microsoft, OpenText , Oracle , Perceptive Software (Lexmark)	Competidores:	Líderes: Appian, IBM , Pegasystems
Jugadores de Nicho: Ever Team, Fabasoft, Laserfiche, M-Files, Objective, Software Innovation, SunGard, Systemware, Upland Software	Visionarios: Alfresco, Newgen Software , Xerox	Jugadores de Nicho: DST Systems, Kofax (Lexmark) , Newgen Software Technologies , OpenText , PNMsoft	Visionarios: Bosch Software Innovations, Oracle , Software AG, Tibco Software, Vitria, Whitestein Technologies

Tabla 3 Resumen Cuadrante Mágico 2014 para ECM e iBPMS

Como se puede observar, IBM, OpenText, Oracle y Perceptive Software son los únicos líderes en el mercado de ECM que además cuentan con una solución valorada por Gartner en el mercado de iBPMS. Esto significa que las capacidades de gestión de procesos de fabricantes como EMC o Microsoft se limitan a flujos de aprobación de documentos y no están pensados

para extenderse más allá del gestor documental e integrarse con otros sistemas. Además, no ofrecen capacidades de monitorización e informes sobre la ejecución de los procesos. Con el fin de estudiar las soluciones más completas posibles para la implantación proyecto, se considerarán únicamente los 4 fabricantes mencionados anteriormente por tener una propuesta líder en ECM, que cubre 3 de los 4 componentes a considerar para el proyecto, y además una propuesta en el mercado de los iBPMS. No se evaluará Newgen Software puesto que su propuesta en ECM se clasifica de visionario y en iBPMS de jugador de nicho por lo que se prevé que presentará importantes limitaciones en más de uno de los 4 componentes que se requieren. Por el contrario, si se evaluará Alfresco, visionario en EMC, como alternativa OpenSource al predominante ecosistema de Software propietario.

A continuación se van a estudiar las soluciones que ofrecen los mencionados proveedores y se hará un breve análisis tanto de la compañía como de las soluciones ofertadas. Finalmente, se condensarán los módulos que ofrecen los distintos proveedores para construir un EDMS como el que requiere la casuística de la gestión de reservas en una tabla a modo de resumen comparativo.

El siguiente estudio está total y exclusivamente basado en información disponible en la página web pública de cada proveedor. La información que se busca de cada compañía es:

- Descripción
- Soluciones y productos dentro de las 4 áreas identificadas como claves para un Gestor Documental para la Automatización de la Gestión de Reservas
 - o Captura
 - o Reconocimiento Inteligente de Documentos (IDR)
 - o Gestión Documental
 - o Gestión de Procesos

2.2.1.1. *Alfresco Software Inc.*

2.2.1.1.1. Descripción

Fundada en 2005 por John Newton, cofundador de Documentum (adquirida por EMC) y John Powell, jefe de operaciones de Business Objects (adquirida por SAP), ofrecen al mercado una solución de gestión contenidos empresarial de código abierto basado en Java y, por lo tanto, multiplataforma. [19]

2.2.1.1.2. Productos y Soluciones

Ofrece su producto principal, el Gestor Documental homónimo, Alfresco, en 2 versiones: [20]

- **Alfresco One:** anteriormente conocido como Enterprise Edition. Esta versión requiere de un pago recurrente a modo de suscripción y da acceso a todas las capacidades de gestión de contenidos tradicional junto con ciertas características empresariales como el despliegue de la solución sobre servidores de aplicaciones Oracle WebLogic o IBM WebSphere y la creación de clústers para escalar Alfresco y asegurar una alta disponibilidad del servicio
- **Community Edition:** es la versión gratuita de Alfresco con ciertas restricciones en cuanto a las posibilidades de despliegue, pudiendo desplegarse únicamente sobre Apache Tomcat, así como el uso de ciertos módulos como el servidor de transformación, pieza clave para obtener un conversiones fidedignas de documentos para su visualización en la web. Además, no cuenta con soporte de ningún tipo por parte de Alfresco Software, quedando la resolución de incidencias completamente en manos del administrador y el apoyo de la comunidad. Esta versión es completamente gratuita.

En cuanto a la Captura y Reconocimiento Inteligente de Documentos, Alfresco Software certifica la integración con el componente o módulo para Alfresco desarrollado por la compañía Ephesoft, conocido como Ephesoft Enterprise Edition. [21]

Ephesoft es una compañía fundada en 2010 y que lanzó su producto de Captura y Reconocimiento Inteligente de Documentos en 2011. Es por tanto de los proveedores con menor tiempo en el mercado con un producto bastante innovador con respecto a los demás. La principal innovación viene dada por su modelo de despliegue que es 100% web, careciendo de clientes pesados.

Aunque, al igual que otras soluciones OpenSource, cuenta con una versión completamente gratuita para la comunidad, Alfresco certifica la integración con la versión Enterprise Edition, la cual requiere una suscripción.

Finalmente, para la gestión de procesos, Alfresco Software ofrece **Alfresco Activiti**, un gestor de procesos empresarial de código abierto, basado en Java, con un motor de ejecución de procesos BPM y con editor gráfico de acuerdo al estándar BPMN 2.0.

En cuanto a soluciones verticales por industria ofrecen soluciones específicas para educación o sanidad y en cuanto a horizontales ofrecen soluciones como la automatización del proceso de cuentas a pagar. No ofrecen nada en concreto para la industria hotelera ni para la gestión de reservas para los Bancos de Camas.

2.2.1.2. IBM

2.2.1.2.1. Descripción

IBM es una empresa histórica en el sector. Fue fundada el 16 de Junio de 1911 en Endicott, Nueva York (EEUU). En la actualidad cuenta con más de 435.000 empleados en todo el mundo y ofrece multitud de soluciones y productos tanto hardware como software, además de servicios de consultoría y de hosting. Dos de los productos por los que mejor se conoce a IBM son sus computadores Mainframe y sus procesadores POWER. A lo largo del tiempo, IBM ha ido prescindiendo de líneas de negocio “commodizadas”, esto es, que aportan escaso valor añadido por si solas y su éxito depende en gran medida de su precio, como por ejemplo su línea de servidores x86 y ordenadores personales, vendidos a la empresa china Lenovo. Al mismo tiempo y a base de adquisiciones, ha ido aumentando su oferta de software empresarial, entrando en mercados como el Big Data, Análisis o la nube.

2.2.1.2.2. Productos y Soluciones

Encontrar las soluciones de Captura, Reconocimiento Inteligente de Documentos, Gestión Documental y de Procesos no es sencillo dado el inmenso portfolio de productos de Gestión de Contenidos Empresarial con el que cuenta IBM. Leyendo las descripciones disponibles en su página web, se puede concluir que los productos correspondientes de IBM son:

- **Datacap Taskmaster Capture**. Producto proveniente de la adquisición por parte de IBM en 2010, tiene funcionalidades tanto para capturar documentos en papel o electrónicos así como para ejecutar reconocimiento inteligente sobre los mismos.
- Para la Gestión de Contenidos, IBM cuenta con, al menos, 2 soluciones: **IBM Content Manager Enterprise Edition** y **IBM FileNet Content Manager**. Haciendo un estudio somero de ambas ofertas no queda claro por qué existen ambos productos, si están completamente solapados o tienen fines distintos. Lo único que si se sabe es que el producto IBM FileNet viene de la adquisición de FileNet en el año 2006.
- El 5 de Abril de 2011, IBM lanzó **IBM Business Process Manager**, solución BPM única nacida a partir de los anteriores **WebSphere Process Server** y **WebSphere Lombardi Edition** [22], quedando ambas obsoletas.

IBM ofrece una serie de soluciones verticales para el sector de viajes y transportes como programas de fidelización para hoteles y compañías de transporte pero ninguna en concreto para Hoteles ni Bancos de Camas.

2.2.1.3. Lexmark

2.2.1.3.1. Descripción

La tradicional compañía dedicada a soluciones y productos de impresión e imagen conocida por sus impresoras que opera desde 1991 a raíz de que IBM se deshiera del negocio de fabricación de impresoras, adquirió el 21 de Mayo de 2010 una empresa llamada Perceptive Software [23], cuyas soluciones se englobaban en el marco de la gestión de contenidos (ECM). Desde entonces, Perceptive Software cambia su marca a Perceptive Software from Lexmark pero continua operando como un negocio independiente y Lexmark continua extendiendo dicho negocio a base de nuevas adquisiciones. Concretamente, el 5 de Marzo de 2012 adquiere Brainware, Inc. [24], una empresa líder en software de captura inteligente de datos. Su plataforma se denomina Brainware Distiller y se trata del mismo software que Oracle incorpora en su solución de digitalización mediante una relación de fabricante de equipos original (OEM por sus siglas en ingles Original Equipment Manufacturer), bajo el nombre Oracle WebCenter Forms Recognition. A raíz de la adquisición de Brainware Inc. y su inclusión en la línea de negocio Perceptive Software de Lexmark, el día 27 de mayo de 2012, 22 días tras el anuncio oficial de la adquisición, Perceptive Software y Oracle anuncian que extienden su relación OEM [25]. De Brainware Distiller/WebCenter Forms Recognition se hablará en detalle más adelante

en esta memoria. Otra adquisición importante por parte de Lexmark en su objetivo por extender su negocio a la oferta de soluciones software y servicios de digitalización y gestión de documentos y de procesos es la de ReadSoft, compañía sueca especializada en la automatización de procesos departamentales como el proceso de cuentas a pagar. A día 8 de Septiembre de 2014, Lexmark anuncia que poseyendo el 90% de las acciones de ReadSoft y con el 98% de los inversores habiendo aceptado la oferta de adquisición revisada, se disponen a reclamar el resto de acciones pendientes hasta completar la adquisición [26]. Con estas y otras adquisiciones, Lexmark consigue entrar con Perceptive Software en el cuadrante de Gartner para soluciones ECM de 2014.

El 24 de Marzo de 2015, Lexmark anuncia que ha iniciado el proceso de adquirir para su marca Perceptive Software a otro gran fabricante en el área de la Captura y los procesos orientados a la digitalización: Kofax [27].

Podría decirse que Lexmark está pasando por la misma fase de crecimiento por adquisiciones por el que pasaron IBM, OpenText y Oracle hace aproximadamente una década. Esta racha de adquisiciones ha aumentado en gran manera el portfolio de Lexmark, complicando sobremedida su estudio y añadiendo cierta inseguridad al no saber con certeza qué productos se considerarán estratégicos y cuales se dejarán de soportar.

Lexmark cuenta actualmente con 12.700 empleados a nivel mundial a los que se deben sumar los más de 1.400 empleados de Kofax una vez finalice el proceso de adquisición.

2.2.1.3.2. Productos y soluciones

A continuación se listan los productos de Lexmark que cubren las 4 áreas anteriormente estudiadas:

- **Captura:** Perceptive Capture. La adquisición de Kofax añade otros productos a la lista como Kofax Capture, Kofax Web Capture o Kofax Front Office Server
- **Reconocimiento Inteligente de Documentos:** Perceptive Intelligent Capture, parte de Perceptive Capture
- **Gestión Documental:** Perceptive Content
- **Gestión de Procesos:** Perceptive Process. La adquisición de Kofax añade Kofax TotalAgility y Kofax Markview.

Lexmark es el único fabricante que además ofrece sus propios escáneres y dispositivos multifunción.

En cuanto a soluciones verticales, ni Lexmark ni la recientemente adquirida Kofax ofrecen nada específico para la industria hotelera ni para el proceso de confirmación de reservas de los Bancos de Camas.

2.2.1.4. OpenText

2.2.1.4.1. Descripción

Empresa canadiense fundada en 1991 y que actualmente cuenta con 8.500 empleados. Se definen como la empresa líder en Enterprise Information Management (EIM), término acuñado por primera vez en 2006 por Ian Kiernan de Data Agility como “los procesos, tecnologías y herramientas necesarias para transformar datos en información, información en conocimiento y conocimiento en planes que lleven a acciones de negocio rentables”. Desde entonces, compañías como Gartner o la propia OpenText han dado sus propias definiciones, recogidas en el estudio del Dr. Ulrich Kampffmeyer de PROJECT CONSULT donde él mismo define EIM como “La gestión y utilización exhaustiva de toda la información de una empresa, sin importar la ubicación, el usuario, el autor, la fuente, la aplicación, la plataforma, el caso de uso, el formato el medio o el tiempo. Para cubrir este enfoque, EIM soporta, usa y unifica componentes funcionales como *Social Business*, *Business Intelligence*, *Enterprise Search*, Clasificación Automática, *Big Data*, *Knowledge Management*, *Cloud*, Movilidad y Ciclo de vida de la información. Según el Dr. Ulrich Kampffmeyer, el término EIM es una profecía autocumplida haciendo referencia a la evolución de cómo se comenzaron a utilizar y definir los documentos o contenidos no estructurados en la empresa, abstrayéndolos cada vez más de su formato o tipo y teniendo cada vez más en cuenta su valor en el negocio. La profecía se cumple cuando hemos desestructurado por completo los contenidos, abstrayéndolos de formato, origen o tipo pero los gestionamos con procesos y herramientas completamente estructurados para sacar el máximo rendimiento de la información.[28]

2.2.1.4.2. Productos y Servicios

A continuación se listan los productos de OpenText que cubren las 4 áreas anteriormente estudiadas:

- **Captura: OpenText Enterprise Scan** (Papel) y **OpenText Capture Center** (Electrónico)
- **Reconocimiento Inteligente de Documentos: OpenText Capture Center**
- **Gestión Documental: OpenText Content Server**
- **Gestión de Procesos: OpenText Cordys**, de la adquisición de la compañía Cordys en 2013.

En cuanto a soluciones verticales, OpenText no ofrece nada en concreto para la industria del turismo.

2.2.1.5. Oracle

2.2.1.5.1. Descripción

La segunda compañía del presente listado que comenzó su actividad antes de los 90, es decir, antes de la explosión de internet y cuyo portfolio va mucho más allá de la Gestión de Contenidos. Fundada en 1977 por el ahora famoso multimillonario Larry Ellison como una empresa de software de Base de Datos, tecnología a la que Oracle debe la mayor parte de su fama hoy en día, hoy en día cuenta con 122.458 empleados en todo el mundo.

Oracle comenzó al igual que IBM una estrategia de expansión vía adquisiciones a partir de mediados de los 2000. Las adquisiciones más relevantes fueron PeopleSoft (2004) y Siebel (2006) que junto con el desarrollo propio de e-Business Suite han sido la base de la línea de negocio de aplicaciones hasta la fecha; Stellent (2006) para la Gestión de Contenidos y BEA (2008) para toda la gama de tecnologías de Middleware. Finalmente, desde la adquisición de Sun Microsystems en 2010, Oracle no sólo obtiene una nueva línea de productos de servidores y almacenamiento si no que se convierte en el custodio principal de Java, el lenguaje de programación más extendido en todo tipo de aplicaciones a nivel mundial.

Si hay algo en lo que muchos analistas y clientes están de acuerdo acerca de la estrategia de adquisiciones de Oracle es la rapidez con la que técnicamente el producto adquirido se integra con los productos existentes además de que financieramente, los resultados obtenidos tras la adquisición son mayores que la suma de los resultados de Oracle y la compañía adquirida por separado.

2.2.1.5.2. Productos y servicios

A continuación se listan los productos de OpenText que cubren las 4 áreas anteriormente estudiadas:

- **Captura:** Originalmente llamado **Document Capture**, obtenido a través de la adquisición en 2008 de la compañía Captovation, ha sido rehecho completamente, reconstruyéndolo como aplicación web Java desplegada sobre Oracle WebLogic Server y se conoce como **WebCenter Enterprise Capture**.
- **Reconocimiento Inteligente de Documentos:** Para este componente, Oracle mantiene un contrato de Original Equipment Manufacturer (OEM) con Lexmark Perceptive Intelligent Capture, previamente conocido como Brainware Distiller, y lo rebautiza como **WebCenter Forms Recognition**.
- **Gestión Documental: WebCenter Content.**
- **Gestión de Procesos: Unified Business Process Management Suite.**

Oracle tiene una línea de productos específicos para hoteles y cadenas hoteleras tras la adquisición de la compañía Micros. No obstante, estas son aplicaciones que ayudan a hoteles y cadenas hoteleras de manera individual y no ofrecen soluciones a las problemáticas de los Bancos de Camas. En realidad, dichas aplicaciones ayudarían a los hoteles o cadenas hoteleras a conectar con los bancos de camas.

3. Justificación y descripción detallada de la solución elegida

Se han visto los componentes que deben formar una solución de SGDE, insiendiendo especialmente en la digitalización de documentos en papel con las tecnologías de captura inteligente y procesos de aprobación que ello conlleva, y se ha estudiado la oferta actual de los fabricantes OpenText, Lexmark, IBM y Oracle así como proyectos OpenSource como Alfresco. En la siguiente tabla se compara la cobertura de cada fabricante en los 4 aspectos principales que conforman un SGDE para la gestión de reservas, con el fin de recordar, **Captura, Reconocimiento Inteligente de Documentos (IDR), Gestión Documental y Gestión de Procesos.**

	Captura	Reconocimiento Inteligente	Gestión Documental	Gestión de Procesos
Alfresco	No tiene. Capacidad ofrecida por diversos add-ons como Ephesoft (Certificado), ChronoScan o Capturesco (No certificados)	No tiene. Capacidad ofrecida por diversos add-ons como Ephesoft (Certificado), Orion ScanIT o FILLIN Desktop (No certificados)	Si	Si, mediante Alfresco Activiti
IBM	Si, mediante Datacap Taskmaster Capture	Si, mediante Datacap Taskmaster Capture	Si, mediante FileNet Content Manager o Content Manager Enterprise Edition	Si, mediante Business Process Manager
Lexmark	Si, con Perceptive Capture o Kofax Capture, entre otros. Ofrecen además sus propios escáneres.	Si, con Perceptive Intelligent Capture	Si, con Perceptive Content	Si, con Perceptive Process, Kofax TotalAgility o Kofax Markview.
OpenText	Si, con Capture Center	Si, con Capture Center	Si, mediante Content Server	Si, mediante Cordys
Oracle	Si, mediante WebCenter Enterprise Capture	Si, mediante WebCenter Forms Recognition	Si, mediante WebCenter Content	Si, mediante Unified BPM Suite

Tabla 4 Comparativa de proveedores

3.1. Solución de múltiples proveedores frente a solución de proveedor único

Lo primero que se debe considerar es si confiar en un único proveedor para obtener una solución única e integrada o por el contrario integrar soluciones de distintos proveedores con el fin de, a priori, conseguir lo mejor posible para cada componente.

Existen numerosos artículos y documentos que exponen las ventajas y los inconvenientes de ambos tipos de soluciones y todos coinciden en gran manera con lo expuesto por la compañía SAP en su documento *Evaluating a Single-Vendor Versus Multivendor Approach to Modern Data Management*. [29]¹

Múltiples Proveedores	Proveedor único
Ventajas <ul style="list-style-type: none"> • Permite utilizar la mejor tecnología disponible para cada tarea • Costes de licenciamiento iniciales más reducidos (o nulos) al competir todos los proveedores en cada tecnología • Evita “encerrarse” o el “lock-in” con un único proveedor 	Ventajas <ul style="list-style-type: none"> • Tecnologías pre-integradas reducen complejidad • Menores costes de formación y mantenimiento • Una plataforma única reduce los tiempos de evaluación, testing, pruebas de concepto y despliegues • Reducción de costes de licencias al obtener descuentos por volumen del mismo proveedor
Inconvenientes <ul style="list-style-type: none"> • La complejidad en la arquitectura provoca altos costes de mantenimiento e integración a la larga • Probar y ejecutar pruebas de concepto a lo largo de diversas plataformas y arquitecturas incrementa el tiempo para el despliegue • Costes de formación más altos al tener que aprender tecnologías muy diversas • Mayor riesgo ocasionado por incompatibilidades de versiones de productos de distintos proveedores y por decisiones de ciertos proveedores de cesar de soportar productos de otro 	Inconvenientes <ul style="list-style-type: none"> • Riesgo de encerrarse con un único proveedor • Riesgo de limitar el ritmo de innovación al dictado por el proveedor elegido

Tabla 5 Ventajas e inconvenientes de las soluciones de proveedor único

3.2. Riesgos más frecuentes en la implantación de un proyecto Software

Ya sea un desarrollo completamente nuevo, la instalación y configuración de un producto software ya existente o, como en el caso del SGDE para la Automatización de la Gestión de Reservas, la instalación, configuración, integración y parametrización de diversos productos de mercado, todos los proyectos tienen los siguientes factores en común:

- **Usuarios:** la implantación del proyecto afectará en mayor o menor medida, directa o indirectamente a uno o más usuarios
- **Requisitos:** La captura de requisitos no es exclusiva de los nuevos desarrollos. La simple instalación de un único producto ya está sujeta a numerosos requisitos como el

1. Se opta por presentar las conclusiones de la compañía SAP en torno a este debate puesto que coincide con el consenso general y es un proveedor que no entra en competición con los que se están evaluando para la implantación de un Gestor Documental para la Automatización de la Gestión de Reservas

entorno en el que se debe integrar o los componentes específicos que se deben habilitar.

- **Complejidad:** Integrar dos productos de mercado puede llegar a ser tanto o más complejo que desarrollar una solución desde cero.
- **Planificación y Control:** La instalación de una solución basada en componentes requiere una planificación y control clara para asegurar que cada componente está preparado para ser integrado con los demás en el momento que se haya establecido, con el fin de no incurrir en retrasos de la implantación del proyecto.
- **Equipo:** contar con un equipo experimentado en la implantación de una solución determinada o que tengan un conocimiento en profundidad de la misma puede suponer una importante reducción en el número de jornadas de la implantación del sistema.
- **Entorno Organizacional:** se trate de lo que se trate el proyecto, finalmente se implantará dentro de una organización con mayor o menor complejidad.

3.2.1. Análisis de Riesgos

En el análisis “*Top Ten Lists of Software Project Risks: Evidence from the Literature Survey*” de Tharwon Arnuphaptrairong [30], investigador del departamento de Estadística de la Chulalongkorn Business School, se listan los principales riesgos al implantar cualquier solución software de acuerdo a 12 estudios realizados entre 1981 y 2003 por distintos autores como McFarlan, Boehm o Summer. Cada estudio se estructuró de una forma, dividiendo los riesgos en entre ninguna y 14 categorías o dimensiones y dentro de éstas, cada estudio identificó entre 10 riesgos el que menos y 55 los que más. De entre los 12 estudios, 8 confeccionaron una lista de los 10 riesgos más frecuentes.

Con el fin de simplificar el análisis, Arnuphaptrairong resume las dimensiones de los riesgos en los 6 que se presentan en “*Software Project Risk and their Effect on Outcomes*” de L. Wallace y M. Keil:

1. Usuario (U)
 - a. Resistencia al cambio por parte de los usuarios
 - b. Conflictos entre usuarios
 - c. Usuarios con actitud negativa ante el proyecto
 - d. Usuarios no comprometidos con el proyecto
 - e. Falta de cooperación por parte de los usuarios
2. Requisitos (R)
 - a. Cambios constantes en los requisitos
 - b. Inadecuada identificación de requisitos
 - c. Requisitos poco claros
 - d. Requisitos incorrectos
3. Complejidad (C)
 - a. Uso de tecnología nueva
 - b. Alto nivel de complejidad técnica
 - c. Tecnología inmadura
 - d. Uso de tecnología no utilizada previamente
4. Planificación y Control (P)
 - a. Falta de tecnología de gestión de proyectos efectiva
 - b. Monitorización insuficientemente detallista del progreso del proyecto
 - c. Inadecuada estimación de los recursos requeridos
 - d. Pobre planificación de proyecto
 - e. Hitos del proyecto no definidos de forma clara
 - f. Inexperiencia de los Jefes de proyecto
 - g. Comunicaciones inefectivas
5. Equipo (E)
 - a. Inexperiencia de los miembros del equipo
 - b. Equipo de desarrollo indebidamente entrenado
 - c. Falta de habilidades especializadas requeridas por el proyecto
6. Entorno Organizacional (O)
 - a. Cambio en la dirección durante el proyecto
 - b. Políticas corporativas con efectos negativos sobre el proyecto
 - c. Entorno organizacional inestable
 - d. Reestructuración organizacional durante el proyecto

A continuación, Arnuphaptrairong sintetiza bajo esta categorización unificada los 10 requisitos más frecuentes de cada uno de los 8 estudios que hicieron dicho listado con el fin de encontrar la dimensión bajo la que se encuentran los riesgos más frecuentemente.

Dimensión	Frecuencia
Usuario (U)	14
Requisitos (R)	17
Complejidad (C)	4
Planificación y Control (P)	27
Equipo (E)	9
Entorno Organizacional (O)	9
Total	80

Tabla 6 Dimensiones del riesgo y sus frecuencias

Se puede ver que, con gran diferencia, la mayoría de riesgos identificados en los 8 estudios son de Planificación y Control. En el recuento de requisitos individuales realizado por Arnuphaptrairong, se obtienen las siguientes frecuencias:

Riesgo	Frecuencia
Malentendido en los requisitos (R)	5
Falta de compromiso y apoyo por parte de la alta dirección (O)	5
Falta de involucración adecuada de los usuarios (U)	4
Incapacidad de conseguir compromiso de los usuarios (U)	3
Incapacidad de gestionar las expectativas de los usuarios (U)	3
Cambios en los requisitos (R)	3
Falta de metodología de gestión de proyectos efectiva (P)	3

Tabla 7 Riesgos y sus frecuencias

Se puede observar que, mientras la mayoría de los riesgos son de Planificación y Control, los 2 riesgos que 5 de los 8 estudios identifican como más frecuentes son el malentendido de los requisitos y la falta de compromiso por parte de la alta dirección.

Mientras que el compromiso y apoyo de la alta dirección con el proyecto no estará directamente relacionado con las tecnologías o proveedores que se seleccionen para implementar un EDMS para la Gestión de Reservas, los siguientes factores indicados como inconvenientes a la hora de elegir una solución multi-proveedor sí que pueden incrementar los riesgos de malentendido en los requisitos:

- La complejidad en la arquitectura provoca altos costes de mantenimiento e integración a la larga
- Probar y ejecutar pruebas de concepto a lo largo de diversas plataformas y arquitecturas incrementa el tiempo para el despliegue
- Mayor riesgo ocasionado por incompatibilidades de versiones de productos de distintos proveedores y por decisiones de ciertos proveedores de cesar de soportar productos de otro

Se puede concluir, por tanto, que optar por una solución integrada de un único proveedor reducirá en mayor medida uno de los riesgos que con mayor frecuencia han identificado 8 estudios de Riesgos en la Implantación de Proyectos Software: el malentendido en los requisitos. El principal motivo es que una plataforma pre-integrada necesita un menor número de requisitos, sobretodo de compatibilidad e infraestructura, y a menor número de requisitos, menor será la probabilidad de malinterpretarlos.

Quedan, por tanto, descartados Alfresco, ya que la Captura y el Reconocimiento lo hacen productos de 3^{er}s y Lexmark ya que, como se ha analizado, ha completado los 4 componentes requeridos a base de adquisiciones muy recientes lo que complicará la integración de los mismos, al no estar aún pre-integrados por Lexmark.

La elección ahora está entre IBM, OpenText y Oracle pues ofrecen una solución pre-integrada desde hace más tiempo.

Mientras que los 3 proveedores ofrecen soluciones que cubren los 4 aspectos identificados como críticos para la implementación de un EDMS para la Gestión de Reservas, únicamente IBM y Oracle las ofrecen basadas completamente en su propia pila tecnológica.

	IBM	OpenText	Oracle
Aplicación (EDMS)	FileNet	CS	WebCenter
Middleware	WebSphere		WebLogic
Base de Datos	DB2		Oracle DBMS
Virtualización	PowerVM/AIX		Oracle VM/Solaris Zones
Sistema Operativo	AIX, Linux z/OS		Solaris, Enterprise Linux
Servidor	Blades, x86, Mainframe		SPARC, x86
Almacenamiento	Disco, cinta, SAN NAS		Disco, cinta, SAN NAS

Tabla 8 Integración vertical de IBM, OpenText y Oracle

OpenText requiere de bases de datos y middleware de 3^{os} o tecnologías propietarias y no estándar. Esto acentúa el riesgo ocasionado por incompatibilidades de versiones de productos de distintos proveedores y por decisiones de ciertos proveedores de dejar de soportar productos de otro lo que directamente aumenta el número de requisitos a capturar, con su correspondiente aumento de probabilidad de fallo en el entendimiento de los mismos. [31]

Entre IBM y Oracle la decisión es más complicada y, generalmente, tendrá un impacto directo en el otro riesgo identificado como más frecuente: la falta de compromiso y apoyo por parte de la alta dirección. Esto se debe a que la mayoría de empresas medianas y grandes tienen una fuerte inversión en servidores, bases de datos y/o middleware de uno de los 2 proveedores, en ocasiones en ambos, y por tanto existe una fuerte relación comercial entre la compañía y el proveedor, lo que garantiza el apoyo de la dirección en el caso de optar por una solución del proveedor con el que mejor relación se tiene.

En el caso concreto de la implantación del proyecto de SGDE para la Gestión de Reservas presentado en esta memoria el cliente tiene una fuerte relación comercial con Oracle debido a que gran parte de su infraestructura tecnológica está basada en soluciones Oracle. Por ejemplo, su intranet está construida con tecnología de portales de Oracle conocida como WebCenter Portal que, como su nombre indica, está fuertemente relacionado y pre-integrado con la solución de Gestión Documental de Oracle, WebCenter Content. De igual manera, la base de datos de reservas, en torno a la cual se construye este proyecto, es una base de datos Oracle.

No obstante, con el fin de justificar la decisión de Oracle frente a IBM como proveedor único para la solución de SGDE para la Gestión de Reservas, sirva la siguiente cita del informe de Cuadrante Mágico para Gestión de Contenidos Empresarial (ECM) de Gartner 2014 acerca de la oferta de productos de IBM:

Una de las mayores fuerzas de IBM es además su mayor problema: la amplitud de su portfolio de productos y soluciones puede complicar el entender por dónde comenzar [...] Los equipos comerciales de IBM en ocasiones se presentan con diferentes propuestas, lo que añade confusión a su hoja de ruta. IBM soporta 3 repositorios de contenido y ofrece numerosas soluciones BPM [...]

Esta complicación a la hora de entender cuál de todas las soluciones que ofrece IBM es la que se necesita no hace más que aumentar el riesgo de un malentendido en los requisitos. Como prueba de esto está la existencia de 2 Gestores Documentales identificados en el capítulo anterior.

3.2.2. Conclusión del análisis: elección de Oracle como proveedor

Por lo tanto y a modo de resumen del desarrollo de la justificación expuesta, se opta por implementar el SGDE para la Gestión de Reservas únicamente con tecnología de Oracle por ser la solución que, a priori, presenta menor riesgo de malentender los requisitos y falta de compromiso y apoyo por parte de la alta dirección.

En los siguientes apartados de esta sección se hará una descripción detallada de los productos de Oracle con los que se implementará la solución de SGDE para la Gestión de Reservas.

3.2.3. Descripción a alto nivel de la solución de Oracle

Como se ha comentado en el análisis, las 4 soluciones con las que se construirá el SGDE son:

- **WebCenter Enterprise Capture o Enterprise Capture:** solución que permite de manera automática o manual importar grandes volúmenes de documentos tanto electrónicos como en papel desde diversas fuentes como escáneres, carpetas o bandejas de entrada de correo electrónico.
- **WebCenter Forms Recognition o Forms Recognition:** solución que permite crear y configurar proyectos de reconocimiento inteligente de documentos que aplican reconocimiento óptico sobre los documentos, los clasifican y extraen los campos que se necesiten automáticamente.
- **WebCenter Content: Imaging, WebCenter Imaging o Imaging:** módulo dentro de la solución de gestión documental completa de Oracle, WebCenter Content, orientada a la gestión específica de contenido transaccional proveniente de documentos en papel.
- **BPM Suite o BPM:** Solución de diseño, ejecución y gestión de procesos de negocio que permite a usuarios técnicos y no técnicos colaborar en la construcción y prueba de los mismos para finalmente desplegarlos a toda la organización.

De Enterprise Capture se aprovechará la capacidad de configurar trabajos de importación para recuperar automáticamente los correos electrónicos que contengan adjuntos. Forms Recognition juega el papel más importante del proyecto ya que se encargará de reconocer ópticamente dichos documentos adjuntos, clasificarlos para identificar si se tratan de documentos referentes a confirmaciones de reservas y extraer información relevante de ellos de manera automática. La confirmación de reserva es a continuación ingresada en WebCenter Imaging asignando los valores extraídos automáticamente como metadatos. Finalmente, el ingreso en WebCenter Imaging provoca el disparo automático de un proceso en BPM Suite para la comprobación de los datos extraídos y confirmar la actualización en la base de datos del Banco de Camas.

Las 3 primeras soluciones, forman parte de un subconjunto de soluciones conocida como WebCenter, que engloba todo el portfolio de soluciones de Oracle orientadas al diseño de webs y portales y la gestión de contenidos, siendo estas 3 las soluciones orientadas a la digitalización de documentos en papel.

BPM Suite y las soluciones WebCenter a su vez se engloban dentro de un área de productos conocida como Fusion Middleware (FMW). Prácticamente todas las soluciones englobadas dentro de Fusion Middleware son una o varias aplicaciones J2EE desplegadas sobre el servidor de aplicaciones Java WebLogic, también parte del portfolio de Fusion Middleware.

3.3. Oracle Fusion Middleware

Bajo esta familia de soluciones se encuentran, entre otros, los siguientes productos: [32]

- Servidor de aplicaciones Java WebLogic Server
- Servidor web Oracle HTTP Server
- Bus de servicios Oracle Service Bus
- Plataforma de integración servicios y procesos SOA Suite
- Gestor de procesos de negocio BPM Suite
- Gestor documental WebCenter Content
- Portales horizontales, extranets e intranets WebCenter Portal
- Gestión de contenidos y experiencia web WebCenter Sites
- Gestor de identidades Oracle Identity Governance Suite
- Gestor de accesos Oracle Access Management Suite
- Directorios LDAP
- Frameworks de desarrollo J2EE y móvil: Application Development Framework y Mobile Application Framework

Mobile Application Framework	Application Development Framework	WebCenter Portal	WebCenter Sites	Identity Governance	Access Management
		WebCenter Content			
		BPM Suite			
		SOA Suite			
		Service Bus	BPEL PM		
Java SE		WebLogic Server		Directory Services	

Tabla 9 Portfolio de Oracle Fusion Middleware

Todos tienen una arquitectura común lo cual permite a Oracle homogeneizar el proceso de instalación y parchado de cada producto lo que, en definitiva, permite un gran ahorro de tiempo y costes en el mantenimiento de cualquiera de las soluciones Oracle Fusion Middleware. Además, esto mismo reduce el tiempo de aprendizaje y adopción de nuevas tecnologías Oracle Fusion Middleware, dedicando ese tiempo únicamente a aspectos funcionales.

La arquitectura de cualquier solución Oracle Fusion Middleware puede resumirse en una o varias aplicaciones J2EE desplegadas sobre Oracle WebLogic con uno o varios esquemas de base de datos Oracle asociados como repositorio. [33]

FMW J2EE App1	FMW J2EE App2	FMW J2EE App3
WebLogic Server		
Database Schema		

Tabla 10 Arquitectura esquematizada de Oracle Fusion Middleware

Para entender mejor la arquitectura de cualquier solución Oracle Fusion Middleware es necesario comprender el concepto de dominios de WebLogic Server que se explicará a continuación.

3.4. Oracle WebLogic Server

Oracle WebLogic Server es el servidor de aplicaciones Java más robusto y potente del mercado. Estudios de mercado demuestran que, gracias a su robustez, Oracle WebLogic es hasta un 38% más económico que la alternativa OpenSource más popular, Red Hat JBoss en el plazo de 5 años. Los estudios destacan que pese a que el coste de inversión inicial en Oracle WebLogic puede ser hasta un 22% mayor que Red Hat JBoss, principalmente por el precio de la licencia y el soporte anual. Sin embargo, los costes iniciales de licencias, soporte, servidores e instalación se estiman que únicamente son un 20% del coste total de lo que cuesta tener una plataforma de servidor de aplicaciones Java empresarial. El otro 80%, los costes de mantenimiento de la plataforma, despliegue, recuperación y resolución de errores, en definitiva, los costes operativos derivados de utilizar y mantener la plataforma de servidor de aplicaciones Java son hasta un 56% más económicos en Oracle WebLogic que en Red Hat JBoss. Los principales motivos que alega el estudio son la robustez y facilidad de administración y uso de Oracle WebLogic frente a Red Hat JBoss. [34] [35]

Por otro lado, Oracle WebLogic bate constantemente records de rendimiento frente al principal competidor, IBM WebSphere, en la prueba de referencia o benchmarking SPECjEnterprise2010

de SPEC². SPECjEnterprise2010 es una prueba de rendimiento para tanto servidores de aplicaciones Java como la infraestructura que lo soporta como la máquina virtual de Java, Base de Datos, almacenamiento y servidor.

La prueba consiste en medir la cantidad de transacciones por segundo que pueden realizar una aplicaciones web que utiliza tecnologías Java EE como Remote Method Invocation (RMI), Web Services, Clases, Servlets, Java Server Pages (JSP), Enterprise Java Beans (EJB). Estas transacciones se conocen como EjOPS.

Los últimos resultados publicados por Oracle e IBM son: [36] [37]

Servidor Java EE	Hardware			Base de Datos	Hardware			Resultado
	Modelo	Chips	Núcleos		Modelo	Chips	Núcleos	
Oracle WebLogic Standard Edition 12.1.3	Oracle Server X5-2	2	36	DB2 10.5	Oracle Server X5-2	2	36	21.504,30
IBM WebSphere Application Server V8.5.5.4	IBM System x3650	2	28	Oracle 12.1.0	IBM System x3850	4	60	19,282.14

Tabla 11 Comparativa resultados SPECjEnterprise2010 de Oracle vs IBM

Se puede observar que en condiciones similares en las que Oracle decide utilizar servidores homogéneos para el servidor de aplicaciones y el de base de datos, de 2 CPUs de 18 núcleos e IBM decide optar por una máquina de 2 CPUs de 14 núcleos para el servidor de aplicaciones y un servidor más potente, de 60 núcleos, para la base de datos, la configuración de Oracle aventaja a la de IBM en 2222,16 EjOPS o, lo que es lo mismo, la misma aplicación desplegada sobre servidor de aplicaciones, base de datos y hardware Oracle realizó 2222,16 transacciones más que sobre la configuración equivalente de IBM.

Estos dos hechos, el menor coste a lo largo del tiempo frente a la solución OpenSource Red Hat JBoss y el mayor rendimiento que IBM WebSphere, no hacen sino reforzar la justificación presentada en el apartado anterior de la decisión de Oracle como proveedor único para construir la solución de EDMS para la Gestión de Reservas ya que, como ya se ha intuido en el apartado anterior y se explicará con mayor detalle a continuación, los principales componentes de dicha solución se despliegan sobre Oracle WebLogic.

3.4.1. Dominios de WebLogic

Un dominio se define como la unidad administrativa básica para instancias WebLogic Server. Se compone de uno a varias instancias del servidor WebLogic, conocidos como Servidores Gestionados o Managed Servers, administradas por un único Servidor de Administración, conocido como Administration Server o Admin Server. La función de los Servidores Gestionados es separar las aplicaciones o componentes de una misma solución en instancias distintas, de manera que es posible tener un control más granular sobre qué partes de la solución se están usando o desactivar componentes opcionales para ahorrar recursos. Por su lado, entre otras tareas, desde el Servidor de Administración se puede:

- Arrancar y parar los Servidores Gestionados
- Configurar distintas fuentes de datos, o Datasources, que puede aprovechar cualquier aplicación desplegada sobre cualquier servidor del dominio
- Configurar la seguridad y el directorio de usuarios para las aplicaciones desplegadas sobre cualquier servidor del dominio

Sobre un mismo servidor, físico o virtualizado, pueden coexistir numerosos dominios, cada uno con una función y usuarios distintos. Por ejemplo, se puede tener un dominio dedicado a las aplicaciones de gestión documental, con sus propios usuarios y recursos, y otro dominio dedicado a desarrollo y pruebas de aplicaciones Java. Un mismo dominio puede contener uno o más clústers, esto es, 2 o más Servidores Gestionados ejecutándose en servidores físicos o

² The Standard Performance Evaluation Corporation (SPEC) es una organización sin ánimo de lucro, formada para establecer, mantener y avalar una serie estandarizada de pruebas de referencia o benchmarks que pueden ser aplicados a la última generación de computadores de alto rendimiento.

virtualizados diferentes de manera que se puede distribuir la carga y ante la caída de uno de los servidores, las aplicaciones continúan dando servicio en el servidor superviviente.

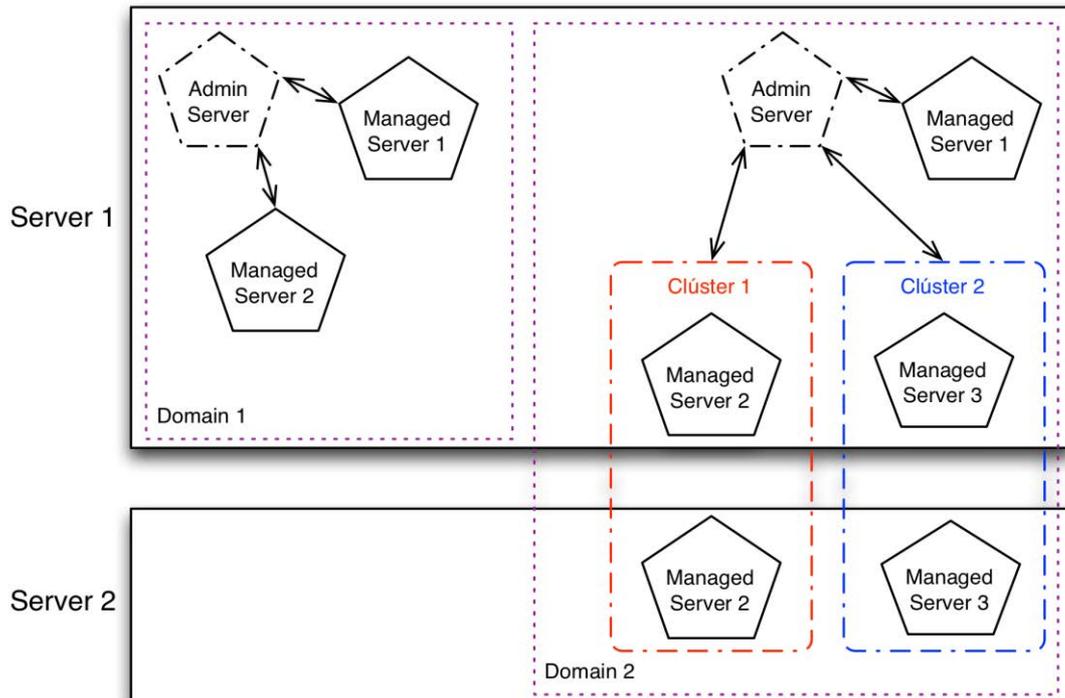


Ilustración 1 Servidores, Clústeres y Dominios WebLogic

Esto facilita sobremanera la gestión de entornos WebLogic y permite tener las aplicaciones desplegadas sobre la plataforma WebLogic constantemente disponibles.

La instalación de la gran mayoría de soluciones Oracle Fusion Middleware consta de 2 pasos principales:

1. Creación del dominio WebLogic con los Servidores Gestionados pertinentes y despliegue automático de las aplicaciones Java sobre los mismos
2. Creación del/los esquemas de base de datos relativos a las soluciones Fusion Middleware que se están instalando.

Para el primer paso, Oracle ofrece un instalador por cada solución de Fusion Middleware y cuyo único prerrequisito es instalar Oracle WebLogic Server. Al ser un instalador Java, es independiente de la plataforma donde vayamos a instalarlo, siempre y cuando esté instalada la Máquina Virtual de Java.

Para el segundo paso, se ofrece una herramienta conocida como Repository Creation Utility (RCU) o Utilidad de Creación de Repositorio. Ésta herramienta es común para todas las soluciones Fusion Middleware y requiere conectar con una base de datos, Oracle u otras soportadas, con permisos de administrador para crear los esquemas, usuarios y tablas pertinentes para cada solución Middleware.

3.5. Oracle WebCenter Content

WebCenter Content es la solución de Oracle Fusion Middleware para la gestión de contenidos empresarial. Un dominio WebLogic de WebCenter Content tiene por defecto los siguientes Servidores Gestionados: [38]

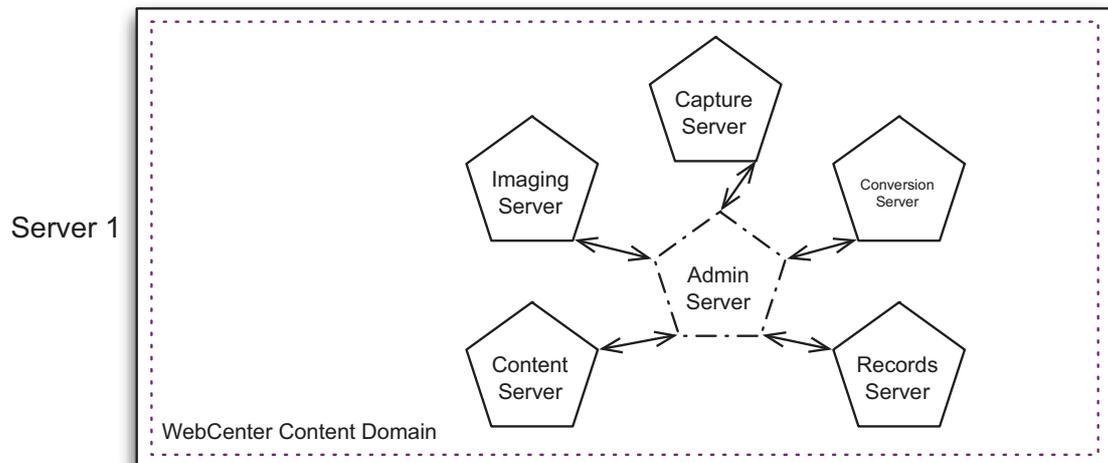


Ilustración 2 Dominio WebCenter Content Genérico

- **Content Server**, Servidor de Contenidos. También conocido como Universal Content Manager (UCM). Es el componente principal de WebCenter Content ya que ejerce de repositorio documental para los demás componentes. Este componente es el encargado de indexar los documentos para su posterior recuperación.
- **Conversion Server**, Servidor de Conversiones, conocido como Inbound Refinery (IBR). Se encarga de la conversión de ficheros a distintos formatos según se necesite. Por defecto, convierte formatos ofimáticos (Word, Excel y PowerPoint) a PDF para poder ser visualizados desde el navegador web sin necesidad de tener una suite ofimática instalada.
- **Imaging Server**, conocido como **WebCenter Imaging**, anteriormente conocido como Imaging & Process Manager (IPM), es una aplicación especialmente diseñada para la gestión de contenido transaccional (TCM), típicamente ligado a la digitalización de documentos y la integración en aplicaciones y procesos.
- **Records Server**, anteriormente conocido como Universal Records Manager (URM). Puede configurarse para utilizar el repositorio de Content Server o utilizar uno propio. En cualquier caso, es capaz de aplicar reglas de disposición a los documentos así como extender la gestión documental más allá de documentación electrónica pudiendo gestionar la ubicación de contenido físico como archivadores, CDs, etc. Esto se conoce como Gestión de Contenidos Físicos o Physical Content Management (PCM).
- **Capture Server**, conocido como **WebCenter Enterprise Capture** se encarga de la captura e importación de documentos desde distintas fuentes tanto físicas como electrónicas

Entre sus aplicaciones se encuentran 2 de los 4 componentes necesarios para la creación del EDMS para la Gestión de Reservas:

- **WebCenter Enterprise Capture**, para la captura e importación de reservas hoteleras en papel desde faxes, escáneres y servidores de correo electrónico.
- **WebCenter Imaging**, para visualizar las reservas almacenadas en formato electrónico en el Servidor de Contenidos e integrar las reservas en el resto de aplicaciones y procesos.

Se analizarán estas dos aplicaciones en mayor detalle, teniendo en cuenta aquellos aspectos relevantes para el diseño del EDMS para la Gestión de Reservas.

3.5.1. WebCenter Enterprise Capture

3.5.1.1. Descripción General

Esta aplicación consta a su vez de dos aplicaciones web, una orientada a usuarios finales, conocida como Cliente de Captura y otra para administradores y súper-usuarios, conocida como Consola de Espacios de Trabajo. Los administradores utilizan la Consola para crear Espacios de Trabajo a los que pueden asignar súper-usuarios. En los Espacios de Trabajo se pueden definir:

- **Perfiles de Cliente.** están directamente relacionados con el Cliente de Captura y tiene que existir al menos 1 para poder utilizar el Cliente de Captura. Los súper-usuarios dan permiso a los usuarios sobre los Perfiles de Cliente y configuran cómo se hace la separación de páginas y qué metadatos hay que introducir al indexar. Cuando los usuarios acceden al Cliente de Captura, eligen el Perfil de Cliente que deseen utilizar de entre los que tienen permiso para ver
- **Metadatos.** Los súper-usuarios definen metadatos que deben ser rellenados o bien manualmente por los usuarios a través de la interfaz del Cliente de Captura o automáticamente mediante Scripts. Es importante definir metadatos ya que hay ciertos datos que únicamente se conocen en el momento de captura de un documento, como por ejemplo, la cuenta de correo electrónico de la que proviene una reserva, y es a través del uso de metadatos como podemos capturarlo y propagarlo hacia los siguientes pasos de la indexación de la reserva.
- **Perfiles de Documentos.** Sirven para clasificar los documentos de forma manual, mostrando distintos metadatos a introducir en función del perfil que se seleccione.
- **Trabajos de procesamiento automático.** Con el fin de incorporar numerosos documentos de manera automática se pueden definir trabajos de procesamiento automático. Existen 4 tipos de trabajo y pueden configurarse para formar una cadena de procesamiento:
 - **Trabajos de Importación.** Son procesos que se ejecutan en segundo plano de manera automática. Los súper-usuarios deben configurar desde dónde se van a importar documentos: Carpeta del servidor, Servidor de Correo electrónico o FAX, la frecuencia con la que se comprobará la existencia de nuevos documentos y qué hacer con los documentos procesados tanto en caso de importación satisfactoria así como en caso de error. Todo procesado debe empezar por un Trabajo de Importación y no puede haber más de uno en el mismo procesado.
 - **Trabajos de Conversión.** Permite configurar procesos en segundo plano que convierten documentos electrónicos como PDF, Word o Excel a imágenes JPEG o TIFF. Puede haber varios Trabajos de Conversión seguidos pero siempre debe haber un Trabajo de Importación o una importación manual antes.
 - **Trabajos de Reconocimiento.** Permite configurar reconocimiento de códigos de barras, códigos de separador conocidos como Patch Codes (ver Figura) o códigos QR o Bidireccionales, entre otros. Especialmente útiles para separar páginas automáticamente en base al reconocimiento de alguno de estos códigos. Puede haber varios Trabajos de Reconocimiento seguidos pero siempre debe haber un Trabajo de Importación o una importación manual antes.
 - **Trabajos de Confirmación.** Como su nombre indica, permite configurar el destino final de los documentos y los metadatos capturados. Por defecto, hay 3 posibles destinos: una carpeta del sistema, WebCenter Content o WebCenter Imaging (aunque técnicamente WebCenter Content e Imaging usan el mismo repositorio, ingresar un documento para poder ser explotado por Imaging requiere unas consideraciones especiales que se verán más adelante; esto permitirá, entre otras cosas, ver los documentos utilizando el visor de Imaging). También se configura el formato de salida de los documentos, a elegir entre formato nativo (el mismo que el de entrada), PDF Imagen o PDF Texto. Los Trabajos de Confirmación indican el final de un procesado automático y por tanto no puede definirse ningún otro trabajo tras ellos.

El Cliente de Captura es una sencilla aplicación Web en la que los usuarios finales pueden:

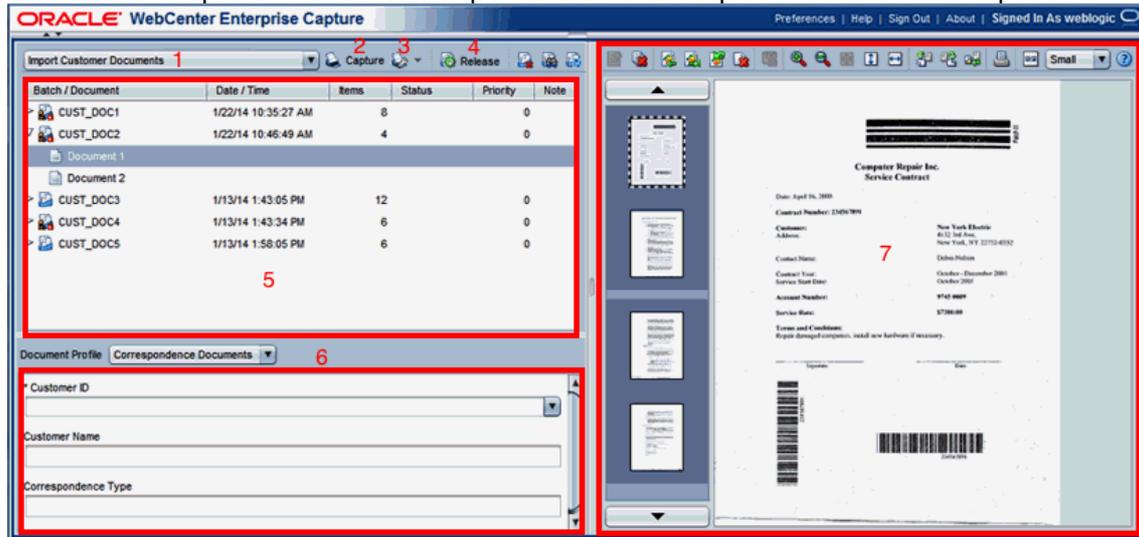


Ilustración 3 Cliente de Captura Enterprise Capture

1. Seleccionar el Perfil de Captura que deseen utilizar. Únicamente podrán elegir de entre aquellos Perfiles de Captura a los que el súper-usuario les ha dado permiso.
2. Capturar nuevos documentos manualmente ya sea desde una carpeta ubicada en el sistema operativo del usuario o bien desde un escáner conectado al equipo.
3. Seleccionar la carpeta desde la cual capturar los documentos manualmente o, por el contrario, seleccionar el escáner como fuente.
4. Liberar lotes para ser procesados por un trabajo de Reconocimiento, Conversión o Confirmación.
5. Seleccionar lotes de documentos capturados
6. Introducir metadatos del documento que se está visualizando
7. Previsualizar documentos, navegar por sus distintas páginas y rotarlas o borrarlas

En el siguiente diagrama de flujo puede observarse la operativa típica de WebCenter Enterprise Capture [39]

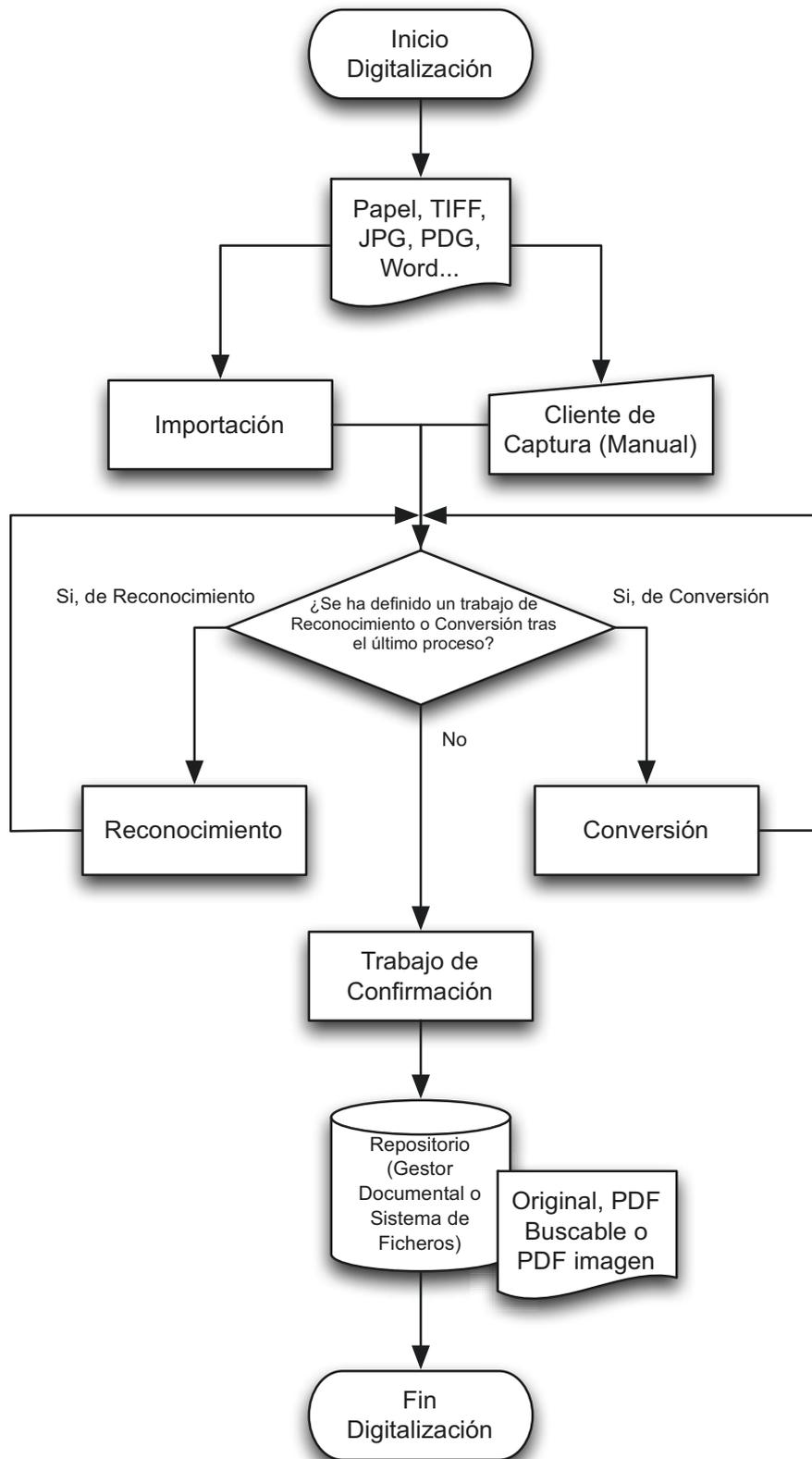


Ilustración 4 Flujo de ejecución típico de Enterprise Capture

La posibilidad de encadenar diversos trabajos de Reconocimiento y Conversión permite configurar flujos alternativos en caso de error, por ejemplo.

3.5.1.2. Extensibilidad

WebCenter Enterprise Capture permite extender su funcionalidad más allá de la ofrecida de serie mediante la creación de scripts JavaScript que utilizan la extensa API de Enterprise Capture.

Existen 3 tipos de Scripts:

1. Scripts de Cliente: modifican el comportamiento del cliente de Captura como, por ejemplo, una acción que se llama cuando el usuario selecciona un campo de metadatos a rellenar
2. Scripts de Reconocimiento: son invocados por un Trabajo de Reconocimiento y modifican el comportamiento por defecto del mismo para, por ejemplo, utilizar el valor reconocido de un código de barras para actualizar una base de datos.
3. Scripts de Importación: son invocados por un Trabajo de Importación y modifican el comportamiento por defecto del mismo para, por ejemplo, capturar de manera automática datos como la cuenta de correo desde la que se recibe un correo electrónico

3.5.2. WebCenter Imaging

Esta aplicación aprovecha el servidor de contenidos de WebCenter Content como repositorio de documentos por un lado y, por el otro, el motor de procesos de negocio de Unified Business Process Management Suite (BPM Suite) que se describirá a continuación, para ofrecer una solución de Gestión de Contenido Transaccional. El contenido transaccional es, típicamente, imágenes y documentos vinculados a procesos y aplicaciones y tienen, por tanto, mucha carga informativa. El objetivo del contenido transaccional es actualizar las bases de datos y aplicaciones de una organización. El ejemplo más típico de contenido transaccional son las facturas que, como ya se ha visto, la mayoría de proveedores ofrecen una solución para la digitalización y gestión de las mismas y suelen estar vinculadas a aplicaciones de gestión de los recursos de una empresa (ERP). Las reservas de un hotel también son contenido transaccional y, por lo tanto, se prestan a ser gestionadas por una solución como WebCenter Imaging.

WebCenter Imaging ofrece 2 interfaces principales, la Consola y el Visor, conocido como Imaging Viewer.

3.5.2.1. Consola de Imaging

En esta interfaz hay distintos menús cuyo permiso de acceso es configurable a nivel de usuario o de grupo de usuarios.

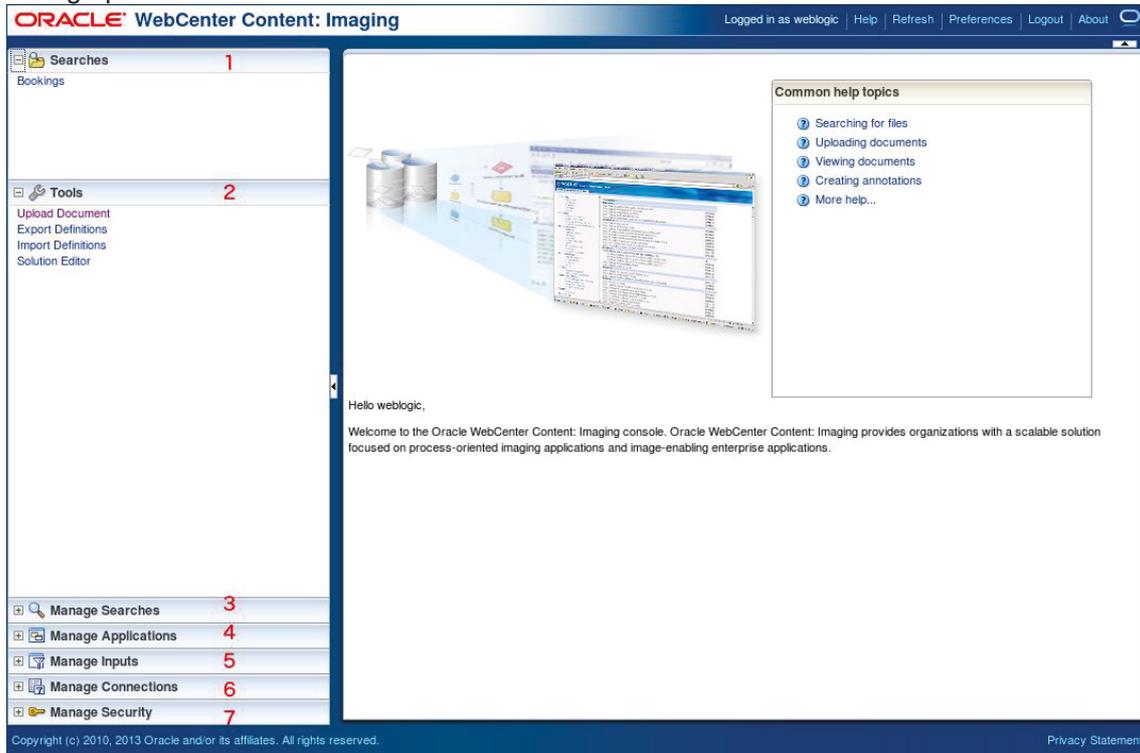


Ilustración 5 Consola de WebCenter Imaging

1. **Búsquedas:** Permite realizar búsquedas de documentos en base a los campos configurados desde la **Gestión de Búsquedas**, buscando por metadatos propios como, por ejemplo, el proveedor en el caso de buscar facturas u hotel en el caso de reservas hoteleras, además de por metadatos genéricos como la fecha de subida del documento o el creador.

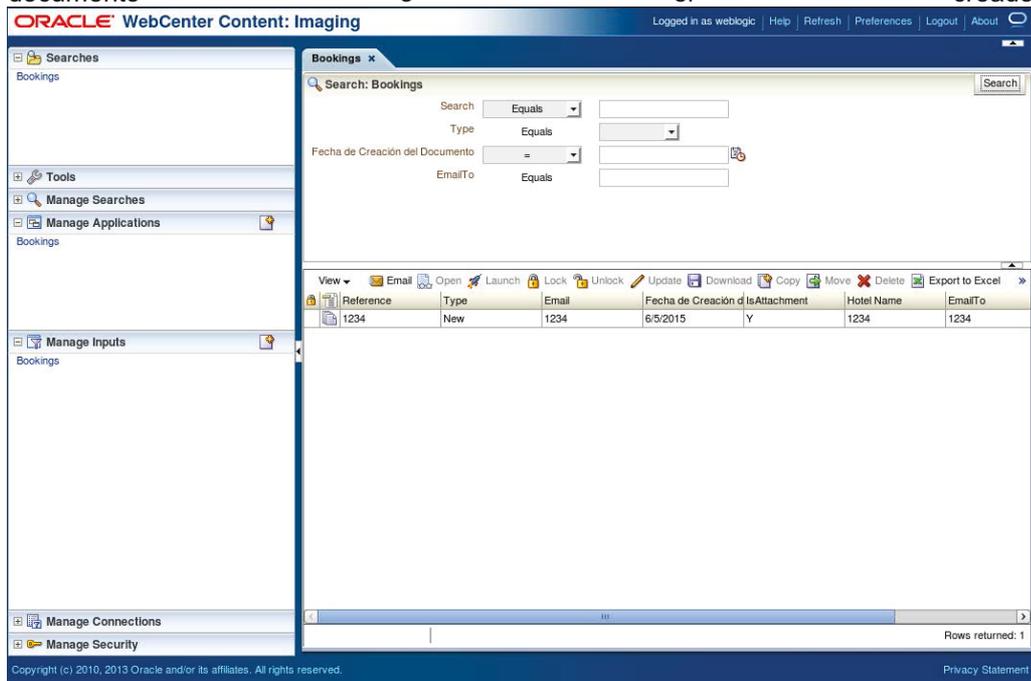


Ilustración 6 Búsquedas en WebCenter Imaging

- Herramientas:** Permite subir documentos manualmente, seleccionando la aplicación a la que se desea subir y rellenando los metadatos relevantes. En este menú también se encuentran funciones para importar y exportar la configuración de WebCenter Imaging de manera que podemos migrar las aplicaciones, metadatos y otras configuraciones de un servidor de WebCenter Imaging a otro.

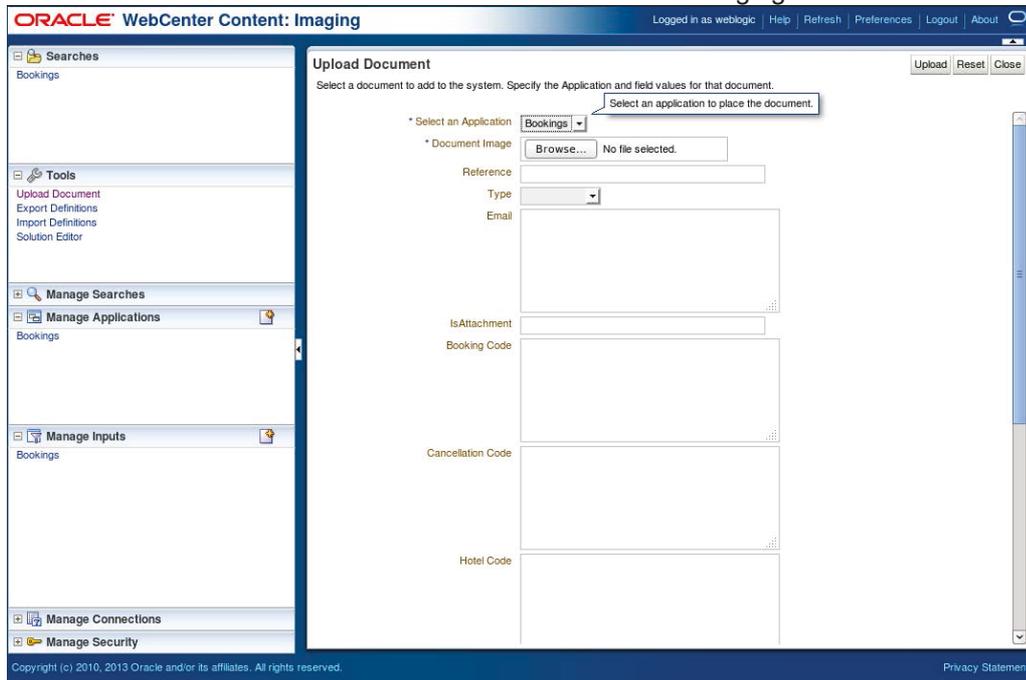


Ilustración 7 Subida manual en WebCenter Imaging

- Gestión de Búsquedas:** Esto es un menú de configuración que permite crear nuevas búsquedas, configurando los parámetros que los usuarios pueden seleccionar para realizarlas. Se pueden preparar búsquedas a través de más de una aplicación.

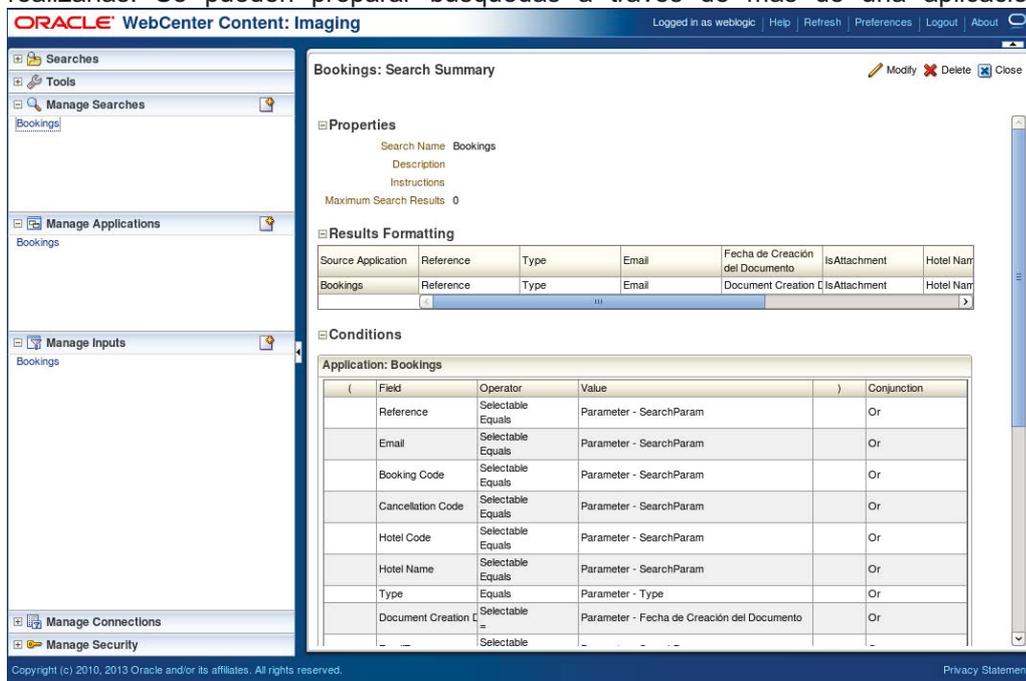


Ilustración 8 Gestión de búsquedas en WebCenter Imaging

- Gestión de Aplicaciones:** Esto es un menú de configuración que permite crear nuevas aplicaciones, asignándoles metadatos, indicando qué conexión con WebCenter Content utilizar como repositorio y qué proceso de BPM lanzar cuando se suban

nuevos documentos a dicha aplicación. El concepto de aplicación en WebCenter Imaging es el del uso o utilidad que tienen los documentos que se suben a dicha aplicación.

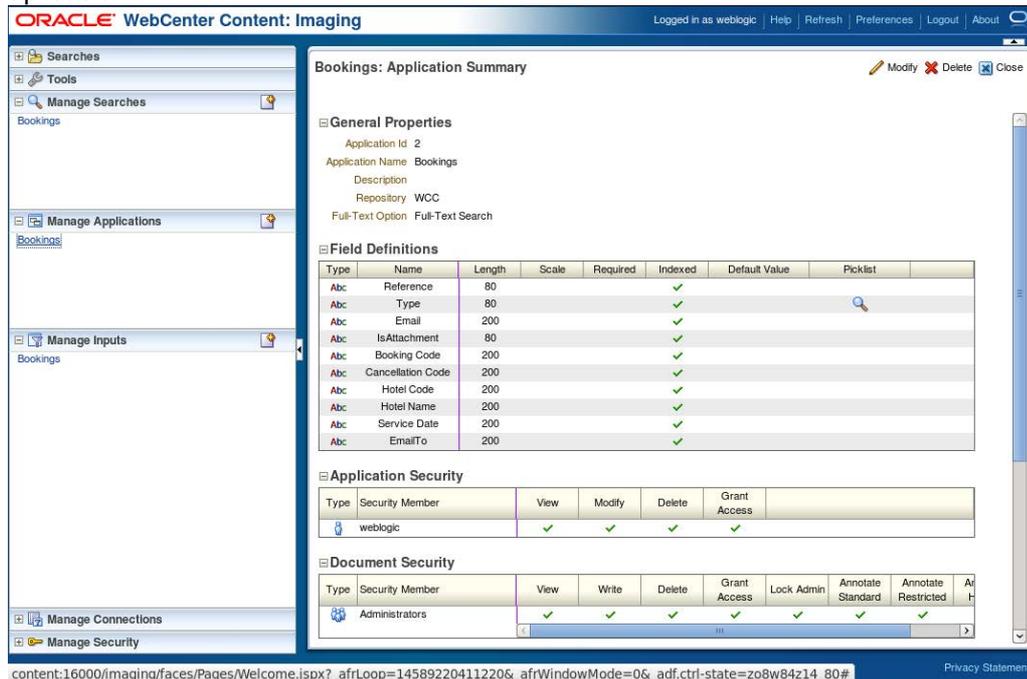


Ilustración 9 Gestión de aplicaciones en WebCenter Imaging

- Gestión de Entradas:** Esto es un menú de configuración que permite mapear el formato de un fichero separado por comas con los metadatos de una aplicación concreta, de manera que se puede configurar una ingesta automatizada de documentos a WebCenter Imaging, utilizando un fichero separado por comas para asignar metadatos. La ingesta automatizada de documentos se realiza a través de un proceso conocido como Input Agent y se explicará su operativa con más detalle en el siguiente apartado

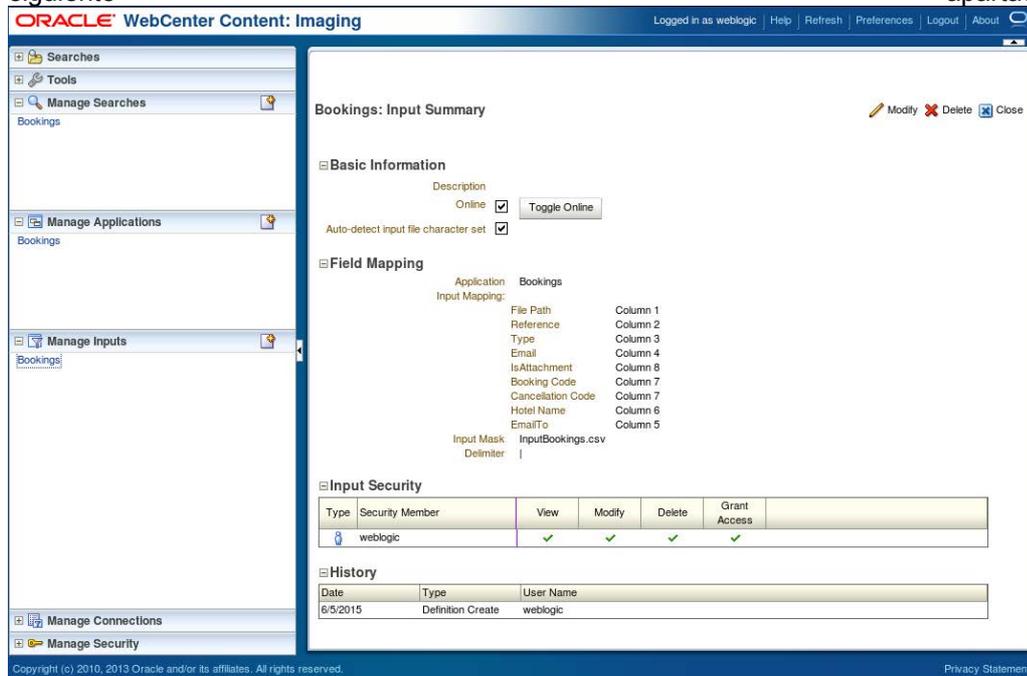


Ilustración 10 Gestión de entradas en WebCenter Imaging

- Gestión de Conexiones:** Permite crear conexiones con distintos servidores de contenidos (WebCenter Content) y servidores de procesos (BPM Suite). Es obligatorio

establecer al menos una conexión con un servidor de contenidos. Suele ser con el servidor de contenidos local, presente en el mismo dominio que WebCenter Imaging como se explicó en el apartado anterior.

7. **Gestión de Seguridad:** Permite a los administradores otorgar a los demás usuarios permisos sobre los distintos menús de configuración.

La otra interfaz que ofrece WebCenter Imaging es tanto o más importante que la propia consola: el Visor, conocido como Imaging Viewer.

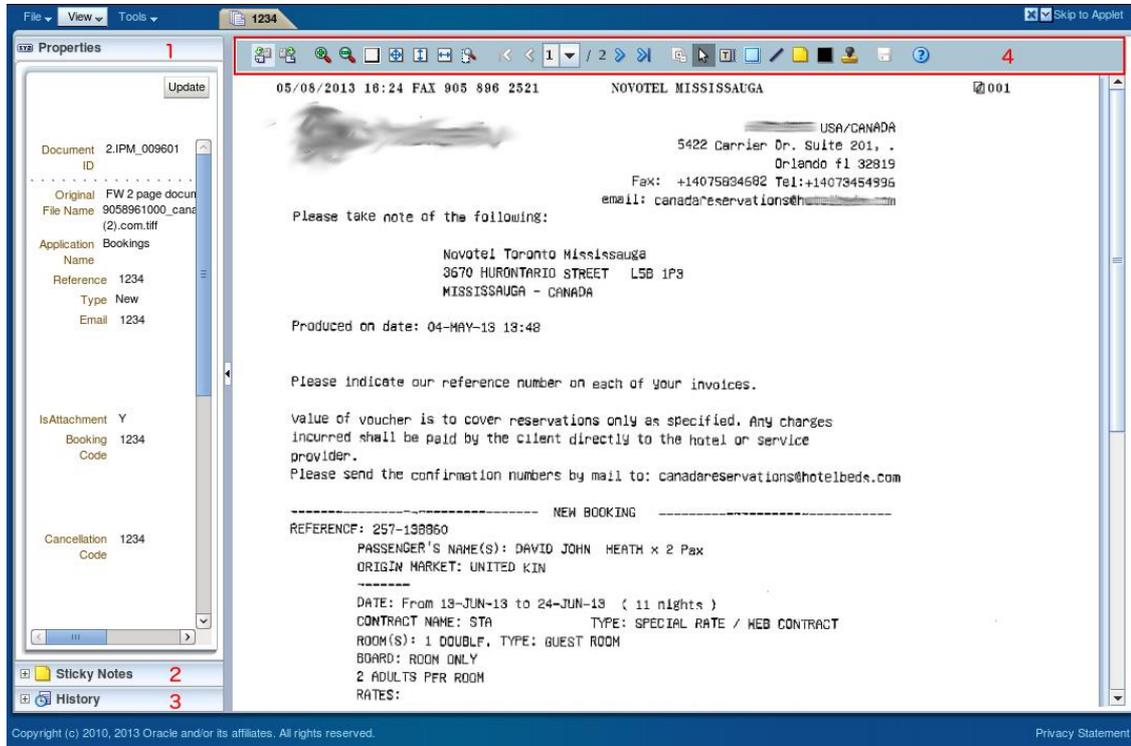


Ilustración 11 Controles del visor de WebCenter Imaging

1. **Propiedades:** Muestra los metadatos del documento que se está visualizando
2. **Notas adhesivas:** lista todas las notas adhesivas que distintos usuarios han añadido a este documento
3. **Historial:** Muestra el historial de cambios y acciones realizados sobre el documento, indicando el usuario y la fecha de la acción
4. **Controles del Visor:** estos controles permiten a los usuarios realizar distintas acciones sobre el documento como añadir notas adhesivas o marcas para resaltar o censurar ciertas partes. Cabe destacar que todos los elementos se superponen en una capa sobre el documento original, de manera que las modificaciones no afectan al documento original. Además, las distintas marcas y notas que se añadan están sujetas a configuraciones de seguridad por lo que se pueden añadir notas que únicamente vean ciertos usuarios o se añadan censuras para otros.

3.5.2.2. WebCenter Imaging Input Agent

El Input Agent es un proceso que se ejecuta en 2º plano en el servidor de WebCenter Imaging y establece una estructura de carpetas automáticamente para gestionar las ubicaciones satisfactorias y las erróneas así como un fichero que explica el motivo del error. Dentro de esa estructura también se almacenan los ejemplos de ficheros de entrada subidos a través de la consola de Imaging. El siguiente diagrama de flujo explica su operativa en detalle.

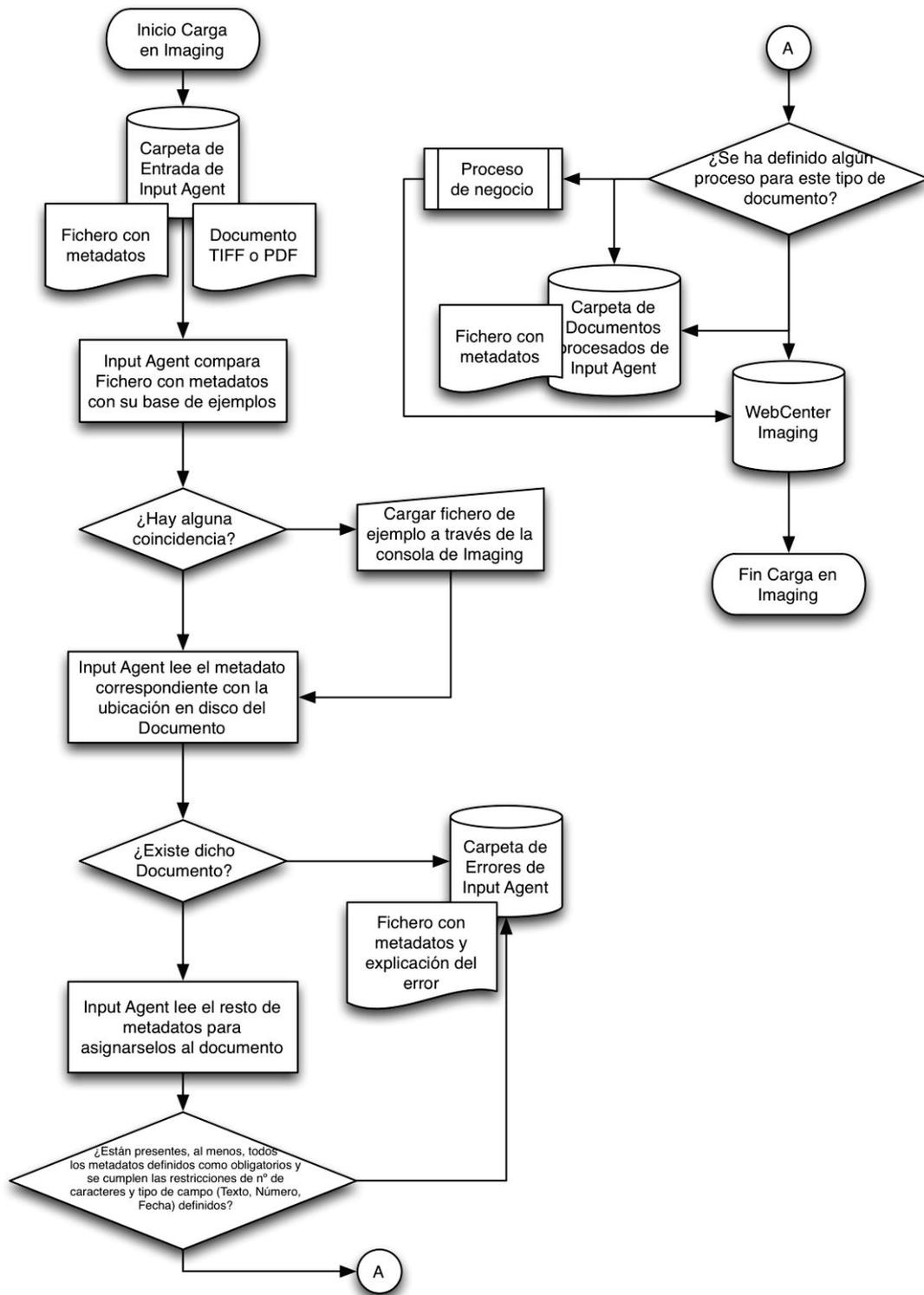


Ilustración 12 Flujo de ejecución típico de WebCenter Imaging

3.5.2.3. Extensibilidad de WebCenter Imaging

Como se ha explicado, el propósito de WebCenter Imaging es integrarse con las aplicaciones y procesos de negocio para complementar las transacciones con documentos. Además de una API nativa de Java y una API basada en servicios web SOAP, WebCenter Imaging ofrece principalmente dos maneras de integrarse con aplicaciones y procesos: URL Tools y Managed Attachments o Anexos Gestionados.

3.5.2.4. URL Tools

Son unas herramientas que permiten acceder a ciertas funcionalidades y ofrecer ciertas interfaces de WebCenter Imaging con tan solo incluir ciertos parámetros en una URL concreta. La URL es <http://<servidor>:<puerto>/imaging/faces/Pages/UrlTools.jspx> y, además de ciertos parámetros cosméticos que harán que la ventana se muestre de una manera u otro, el parámetro principal es *ToolName* y, dependiendo del valor que se le dé, se deben pasar otros parámetros. Los valores que pueden darse a *ToolName*, junto con la funcionalidad que da dicha elección y los parámetros obligatorios son:

- **ExecuteSearch:** muestra la interfaz de búsqueda anteriormente descrita directamente sin tener que acceder a la consola. Se debe pasar obligatoriamente el nombre de la búsqueda como parámetro. Por ejemplo, para buscar los documentos que tengan asociado el empleado Jon Doe dentro de la búsqueda Find HR DOCS, se llamaría a la siguiente URL:

```
http://<server>:<port>/imaging/faces/Pages/UrlTools.jspx?ToolName=ExecuteSearch&SearchName=Find+HR+DOCS&EmployeeName=Jon+Doe
```

- **ViewDocument:** Abre el documento especificado por parámetro en el Imaging Viewer directamente. Por ejemplo, para abrir el documento con identificador 123.RPO_456, se llamaría a la siguiente URL:

```
http://<server>:<port>/imaging/faces/Pages/UrlTools.jspx?ToolName=ViewDocument&DocumentId=123.RPO_456
```

De esta sencilla manera podemos mediante enlaces o componentes HTML iFrame mostrar documentos almacenados en Imaging dentro de cualquier aplicación o proceso de negocio.

3.5.2.5. Anexos Gestionados

Esta solución permite crear integraciones ligeras entre aplicaciones y WebCenter Imaging mediante una sencilla llamada a un servicio web SOAP. En la aplicación únicamente es necesario crear una ligera modificación para añadir un enlace o botón que llame al servicio web, pasando como parámetros los valores de los campos que se están visualizando en ese momento, en otras palabras, los valores de la transacción que se está gestionando en la aplicación. La respuesta es una ventana que muestra los documentos anexos a la transacción y controles para subir nueva documentación anexa desde diversas fuentes. El siguiente diagrama explica cómo funciona exactamente:

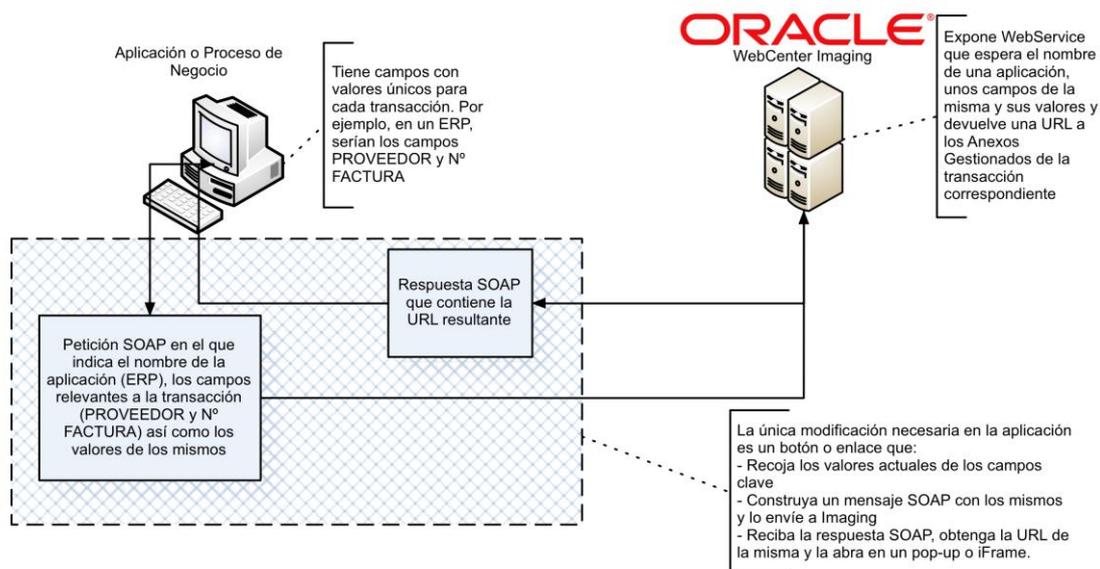


Ilustración 13 Diagrama de los Anexos Gestionados de WebCenter Imaging

Esta es la ventana por defecto de los anexos gestionados, puede modificarse para adaptarse a los estilos de la aplicación:

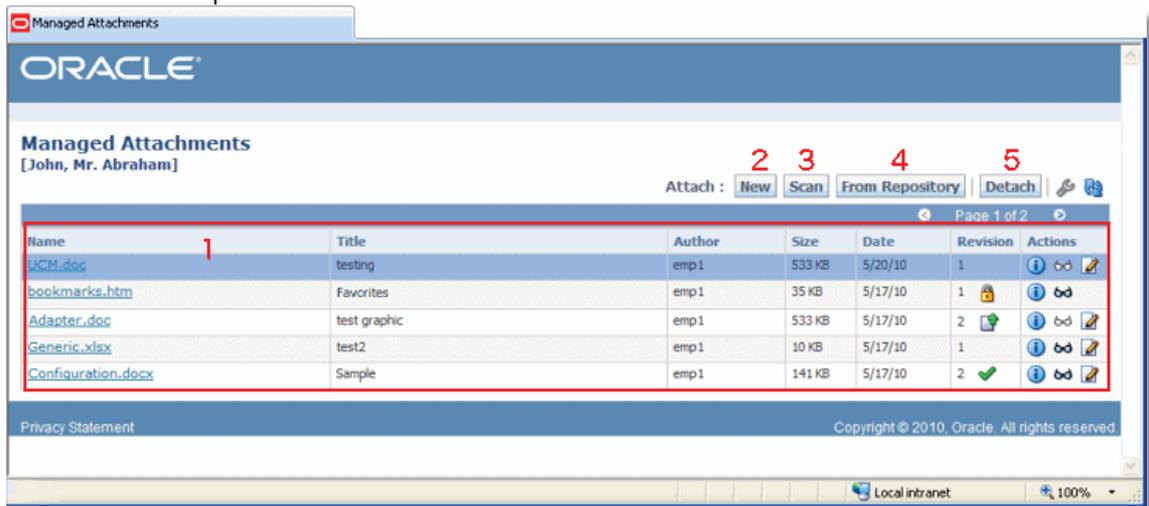


Ilustración 14 Interfaz de los Anexos Gestionados de WebCenter Imaging

1. Lista de documentos anexos a la transacción
2. Subida de nuevos anexos desde el sistema de ficheros del usuario
3. Escanear un documento en papel y anexarlo a la transacción utilizando WebCenter Enterprise Capture
4. Anexar un documento ya subido previamente al repositorio
5. Desanexar el documento seleccionado

Tanto las Herramientas de URL como los Anexos Gestionados son dos soluciones que permiten integrar WebCenter Imaging dentro de otras aplicaciones y procesos de una manera sencilla y lo más desacoplada posible.

3.6. Unified Business Process Management Suite

BPM Suite es la solución de Oracle Fusion Middleware pensada para el diseño, implementación y ejecución de procesos conforme al estándar BPMN definido en el capítulo anterior. BPM Suite es una extensión a SOA Suite, la solución de integración de datos, servicios y procesos, y por tanto, en lo que a configuración del dominio WebLogic, tiene diversas variantes. [41]

- 2 Dominios: soa_domain y bpm_domain
- 1 Único Dominio donde las aplicaciones de BPM se ejecutan sobre el mismo servidor gestionado que las aplicaciones SOA: soa_server
- 1 único Dominio con un servidor gestionado diferente para BPM y otro para SOA. Se representará gráficamente esta opción al considerarse la más práctica y gestionable.

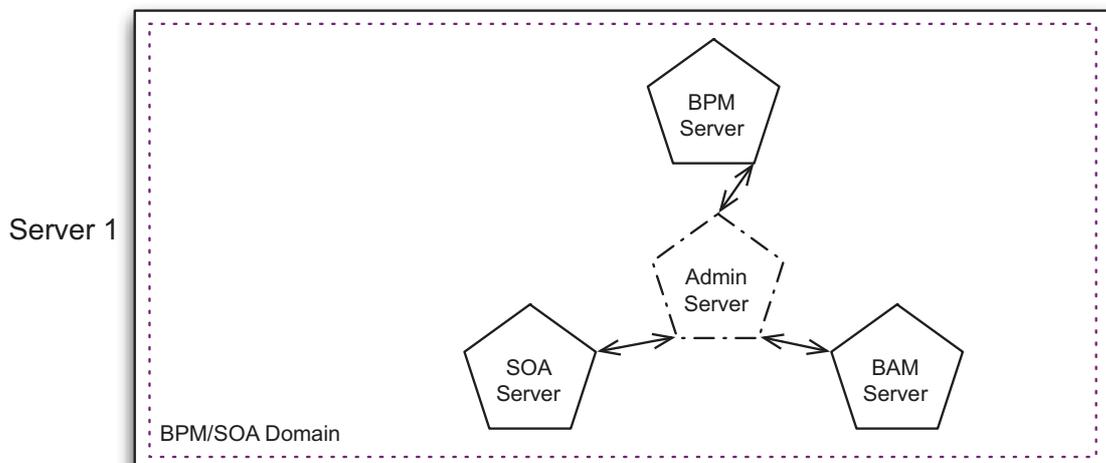


Ilustración 15 Dominio de BPM Suite genérico

Un dominio BPM consta, por defecto, de un servidor de Administración y 3 servidores gestionados.

- **SOA Server:** Servidor donde se ejecutan los componentes de infraestructura de la solución de arquitectura orientada a servicio como el motor de procesos BPEL, fundamento para la ejecución de procesos BPM.
- **BAM Server:** este servidor ejecuta la aplicación Business Activity Monitoring, una aplicación que recoge métricas de los procesos de negocio y permite diseñar cuadros de mando en base a las mismas.
- **BPM Server:** el servidor gestionado encargado de ejecutar las 2 aplicaciones que forman BPM Suite.
 - **Business Process Composer:** aplicación web para el modelado de procesos de negocio
 - **BPM Workspace:** interfaz web donde los participantes de los procesos ejecutan sus tareas encomendadas como adjuntar información, revisar datos, incluirlos, etc.

BPM Suite es el tercer de los 4 componentes necesarios para construir un EDMS para la Gestión de Reservas. Se pasará a describir en detalle las aplicaciones que lo conforman así como la aplicación de Business Activity Monitoring como herramienta de monitorización. Del mismo modo que WebCenter Content es una solución completa que incluye numerosos módulos y funciones de modo que la descripción se limitará a aquellas partes relevantes para la implementación del proyecto de Gestión Automatizada de Reservas.

3.6.1. Business Process Composer

Esta aplicación permite a los diseñadores de procesos modelar procesos conforme al estándar BPMN. También permite la creación de las denominadas Arquitecturas de Negocio o Business Architecture (BA) que permiten organizar y modelar todos los objetivos de una compañía, desde los planes estratégicos hasta los procesos individuales. No obstante, al no ser relevante para el proyecto de Automatización de la Gestión de Reservas, la siguiente descripción se limitará a lo relevante a los procesos BPM. Las distintas interfaces permiten gestionar numerosos proyectos de BPM así como todos los elementos de cada proceso, desde el modelo hasta los formularios, pasando por las reglas de negocio. Se presentan a continuación las 3 interfaces más relevantes.

En primer lugar, la interfaz que condensa todos los proyectos BA y BPM. [42]

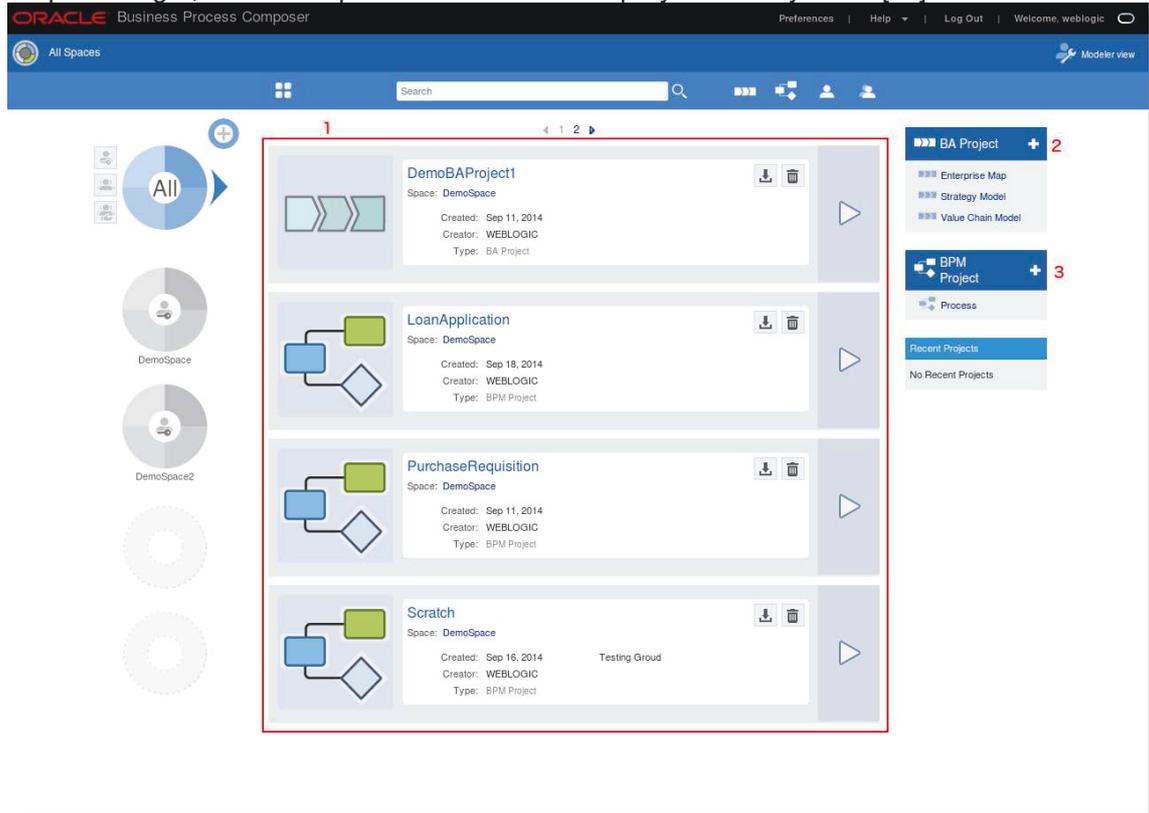


Ilustración 16 BPM Composer

1. Lista de proyectos de Business Architecture (BA) y proyectos BPM.
2. Creación de nuevo proyecto BA
3. Creación de nuevo proyecto BPM

El siguiente nivel de granularidad, es la vista de un proyecto BPM en concreto.

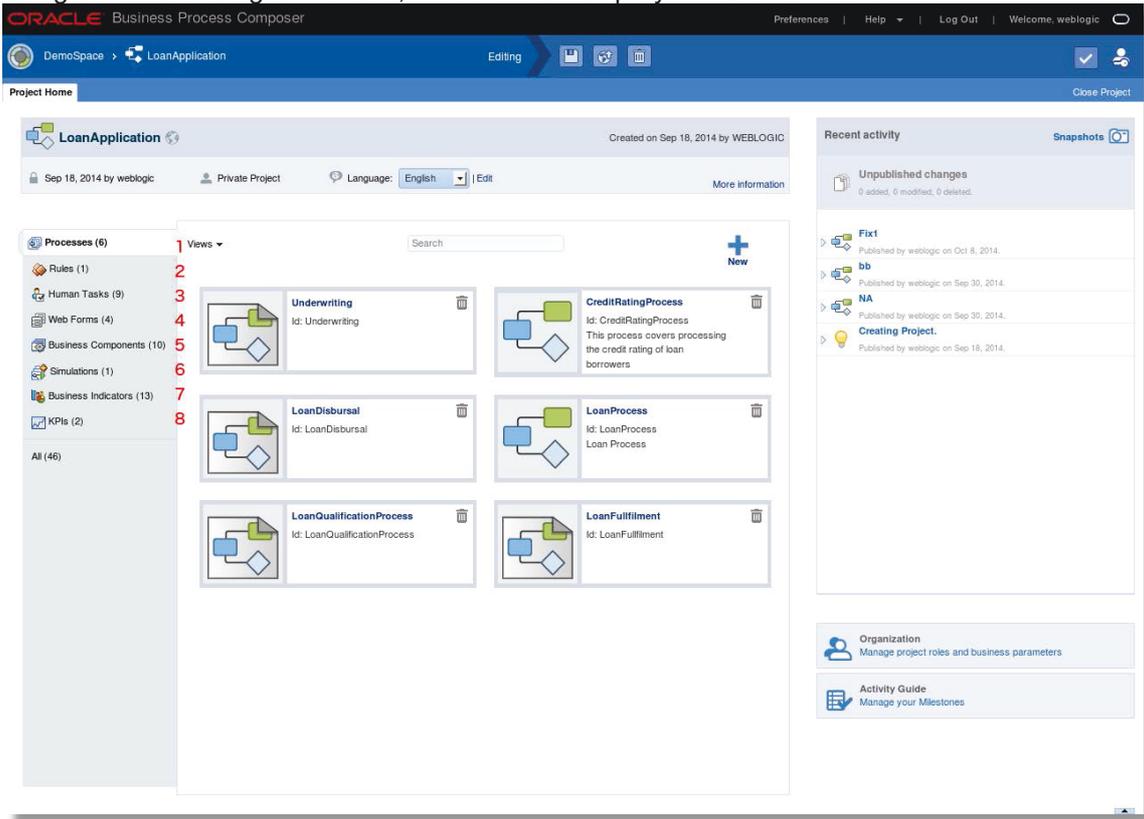


Ilustración 17 Procesos en BPM Composer

1. **Procesos:** Procesos que se han definido para este proyecto de BPM. Se verá con algo más de detalle en la siguiente interfaz.
2. **Reglas:** Reglas de negocio reutilizables a través de los distintos procesos.
3. **Tareas Humanas:** Son partes de un proceso que deben ser completadas por una persona, ya sea introduciendo datos en un formulario, subiendo un documento o interactuando manualmente con algún sistema.
4. **Formularios Web:** Son los formularios que se han creado para la introducción manual de datos por parte de los usuarios en determinados puntos de los procesos. Estos formularios son reutilizables de manera que cuando se esté definiendo un nuevo proceso no sea necesario crear todos los formularios desde el principio.
5. **Componentes de Negocio:** permite exponer datos de los sistemas para que los procesos interactúen con ellos.
6. **Simulaciones:** permite definir datos de prueba para simular el comportamiento de los interventores de un proceso para analizar el flujo del mismo.
7. **Business Indicators:** En esta sección se pueden definir una serie de indicadores sobre los que recolectar información cuando se llega a determinados puntos del proceso.
8. **KPIs:** Son agregados de los Business Indicators. Se explotan a continuación en herramientas como Business Activity Monitoring.

Por último, la interfaz de diseño de un proceso específico agrega todos los componentes anteriormente definidos en determinado orden para servir un proceso de negocio.

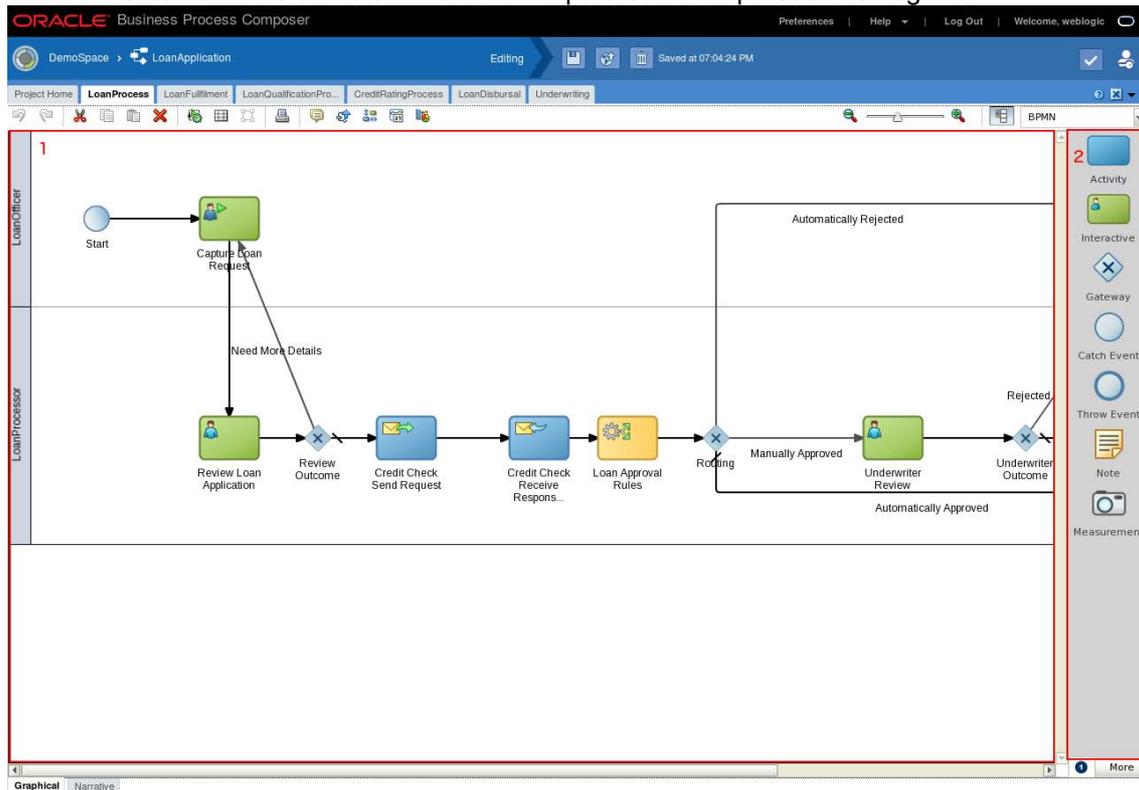


Ilustración 18 Diseño de proceso conforme al estándar BPMN en BPM Composer

1. Lienzo en el que diseñar el proceso. De acuerdo al estándar BPMN 2.0, cada “pista” horizontal corresponde a un responsable o grupo de responsables del proceso.
2. Paleta de componentes del estándar BPMN 2.0 para construir e implementar el proceso.³

3.6.2. BPM Studio

Para la implementación de ciertas actividades complejas que requieran la integración de diversos sistemas a veces no es suficiente con la interfaz web. Para ello, BPM ofrece un entorno de desarrollo conocido como BPM Studio que se sincroniza con la aplicación web mediante control de versiones con el fin de no crear conflictos cuando 2 o más personas trabajan sobre un mismo proyecto. La idea es que usuarios no técnicos utilicen las herramientas web para diseñar y definir los procesos y los desarrolladores recojan esos diseños e implementen las actividades más complejas.

³ BPNN 2.0 Manual de Referencia y Guía Práctica Ed. Hispana, Marzo 2014, Bernhard Hitpass

3.6.3. Business Activity Monitoring

Esta aplicación permite crear cuadros de mando a partir de los datos recolectados mediante Business Indicators o KPIs en los procesos o accediendo a Componentes de Negocio (Vistas y tablas) directamente. Una vez terminado de diseñar y crear un Cuadro de Mando, puede compartirse con otras personas a través de una URL para visualizarlo en un navegador.

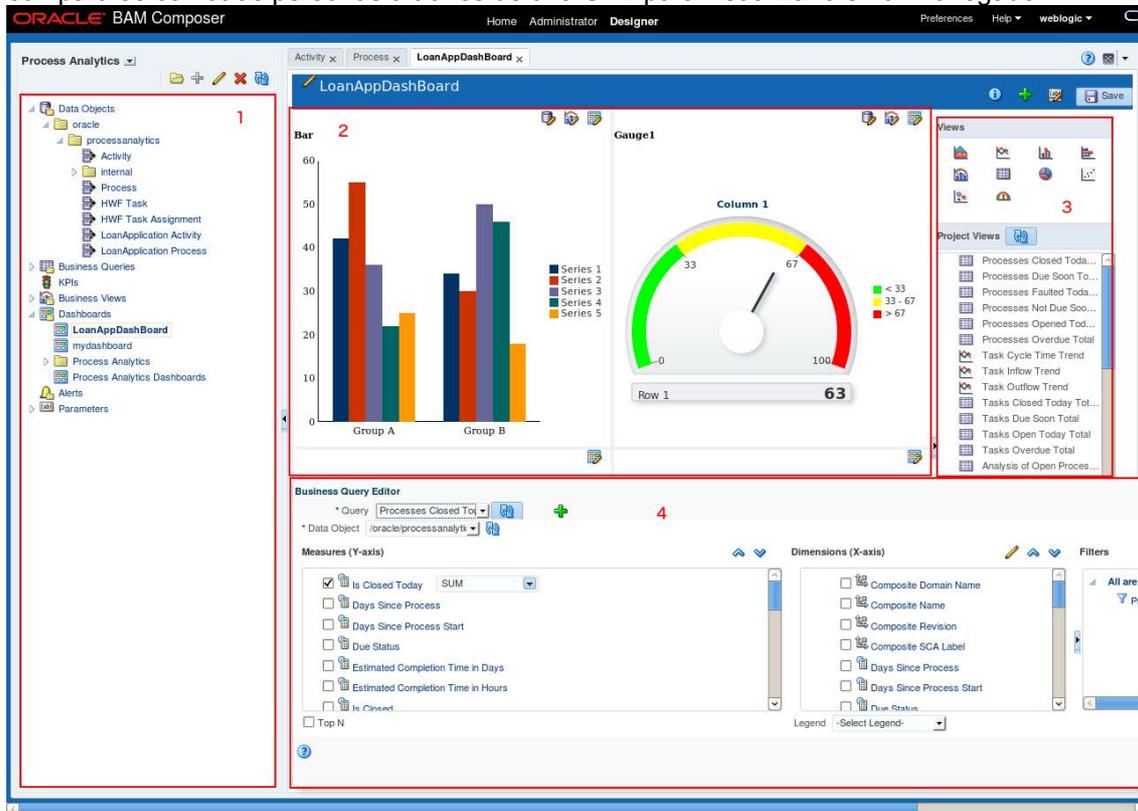


Ilustración 19 Business Activity Monitoring

1. Panel de navegación por los distintos datos que se pueden presentar en los cuadros de mando así como los distintos cuadros de mando que se han diseñado. Desde este menú se consiguen las URL para compartir los cuadros de mando con el resto de la organización
2. Lienzo para diseñar el cuadro de mando. Puede personalizarse el número de visualizaciones que se tienen.
3. Lista de visualizaciones como diagramas de barras, de tarta, “velocímetros”, etc así como las distintas vistas y tablas que pueden arrastrar y soltarse para componer el Cuadro de Mando.
4. Editor para seleccionar los datos a representar en cada visualización o diagrama.

Con esta última descripción se han analizado en mayor detalle 3 de los 4 componentes necesarios para construir el proyecto de Automatización de la Gestión de Reservas, identificando las aplicaciones que forman cada solución y las distintas interfaces que ofrecen para distintos perfiles de usuario. Falta por describir la pieza quizás más relevante del proyecto ya que es la encargada de clasificar y extraer datos automáticamente: WebCenter Forms Recognition.

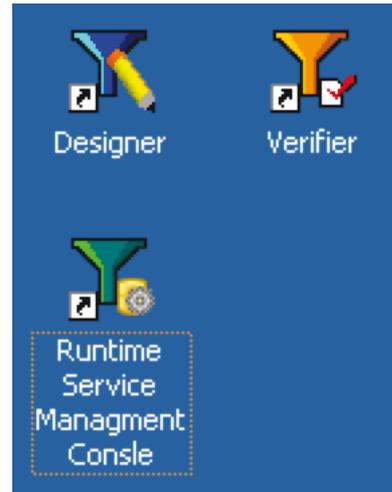
3.7. WebCenter Forms Recognition

La pieza de reconocimiento inteligente de documentos de la solución de Oracle es la única aplicación que no es una aplicación J2EE desplegada sobre WebLogic. Es un contrato Original Equipment Manufacturer que Oracle mantiene con la compañía Brainware, recientemente adquirida por Lexmark, para su producto Distiller y que Oracle ha rebautizado como WebCenter Forms Recognition.

Esto garantiza que el reconocimiento sea óptimo, al contar con una pieza de un fabricante que se dedica única y exclusivamente a tecnologías de reconocimiento, pero además está pre-integrado con el resto de la solución de Oracle.

WebCenter Forms Recognition consta de 3 aplicaciones principales: [43]

1. **Designer:** es la aplicación con la que los diseñadores crean los proyectos de reconocimiento. En un proyecto de reconocimiento se configuran la clasificación de los documentos, los campos a extraer y los conjuntos de aprendizaje, entre otros detalles.
2. **Verifier:** es la aplicación con la que los usuarios pueden comprobar la calidad de la clasificación y extracción automática, reclasificar e introducir datos manualmente en caso de error y, si se desea, añadir la corrección al conjunto de aprendizaje. También está disponible en modo Web, denominándose **Web Verifier**.
3. **Runtime Service (RTS) Management Console:** es una consola que permite crear, arrancar y parar servicios que automatizan el proceso de clasificación de documentos, extracción de campos y exportación del trabajo realizado.



Como se puede observar, son aplicaciones de tipo cliente pesado o Thick Client que requieren instalarse sobre un servidor Windows.

3.7.1. WebCenter Forms Recognition Designer

Es la aplicación principal para desarrollar el proyecto de reconocimiento. A continuación se describen los componentes principales de su interfaz.

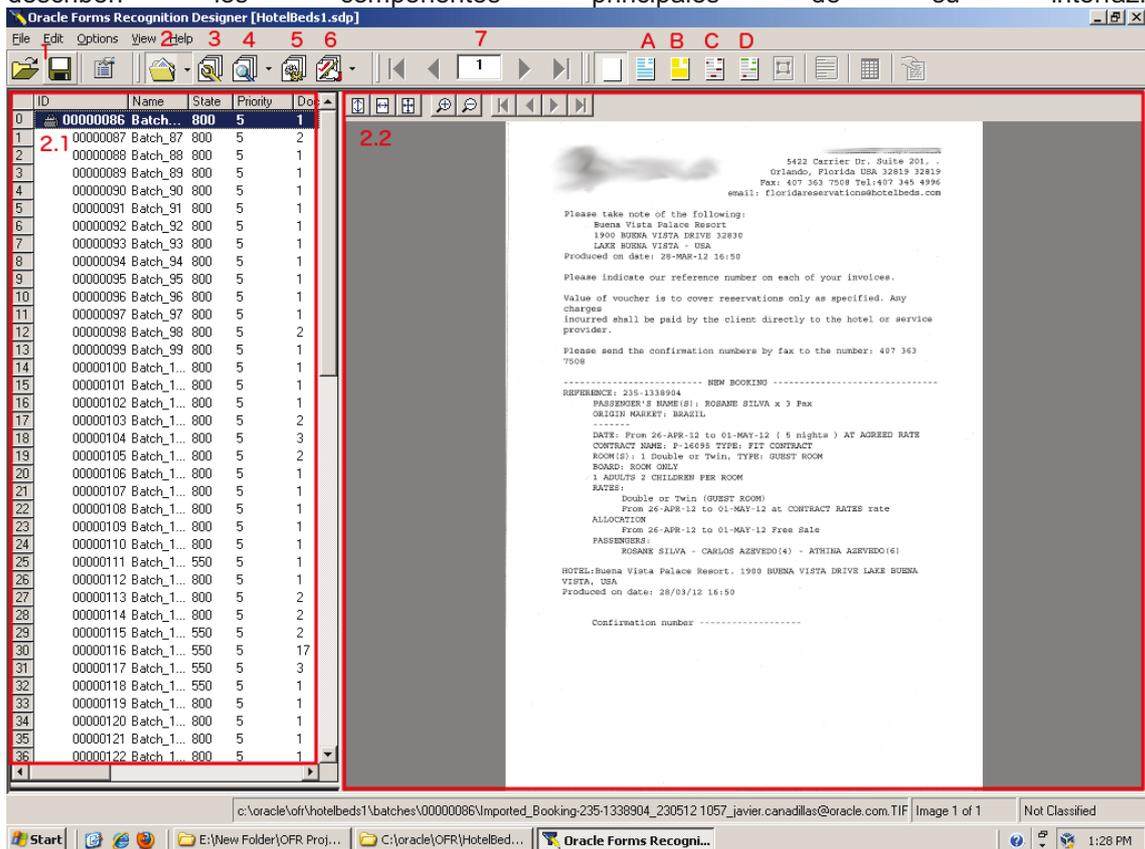


Ilustración 20 Forms Recognition Designer

1. Controles para abrir y guardar proyectos de Forms Recognition (ficheros .sdp).
2. Menú de selección de lotes o archivos sobre los que trabajar. Se pueden cargar documentos directamente desde el sistema de ficheros, de lotes importados mediante el Runtime Service (RTS), o del conjunto de aprendizaje.

- 2.1. Dependiendo de si se desea explorar los lotes, el conjunto de aprendizaje o el sistema de ficheros, en este cuadro se muestra la lista de lotes, el conjunto de aprendizaje o el explorador del sistema de ficheros.
 - 2.2. Visualizador del documento seleccionado con controles para ampliar la previsualización o navegar entre las distintas páginas de un archivo multipágina
 3. Menú de configuración de la clasificación y extracción de campos. Es donde se crean las distintas clases de documentos y sus campos asociados y donde se configuran las reglas para clasificar los documentos y las técnicas de extracción de datos para los campos.
 4. Menú de entrenamiento donde se construye el conjunto de aprendizaje a base de añadir documentos de ejemplo a las clases configuradas así como para la extracción de documentos.
 5. Menú de ejecución donde se pueden pasar una serie de documentos por la clasificación y extracción configurada y obtener métricas de éxito o fracaso en la clasificación y extracción así como estadísticas.
 6. Menú de configuración de formularios de verificación donde se crean y configuran los formularios que los usuarios del **Verifier** usarán para comprobar la clasificación y extracción automatizada de los documentos.
 7. Navegación entre los documentos de una misma carpeta o lote, en función de si en el Menú de selección de archivos (2) hemos seleccionado lotes o sistema de ficheros.
- Los controles resaltados con letras son controles disponibles en los menús resaltados con 2, 3 y 4, los menús en los que está disponible el visor de documentos, y permiten resaltar distintos elementos reconocidos por el motor OCR.
- A. Resalta todas las palabras reconocidas y si se pasa el ratón por cualquiera de ellas se puede ver exactamente qué ha reconocido el motor OCR. Es útil para comprobar que la resolución del documento es suficiente y WebCenter Forms Recognition es capaz de reconocer los caracteres en él de manera que será posible trabajar sobre él.

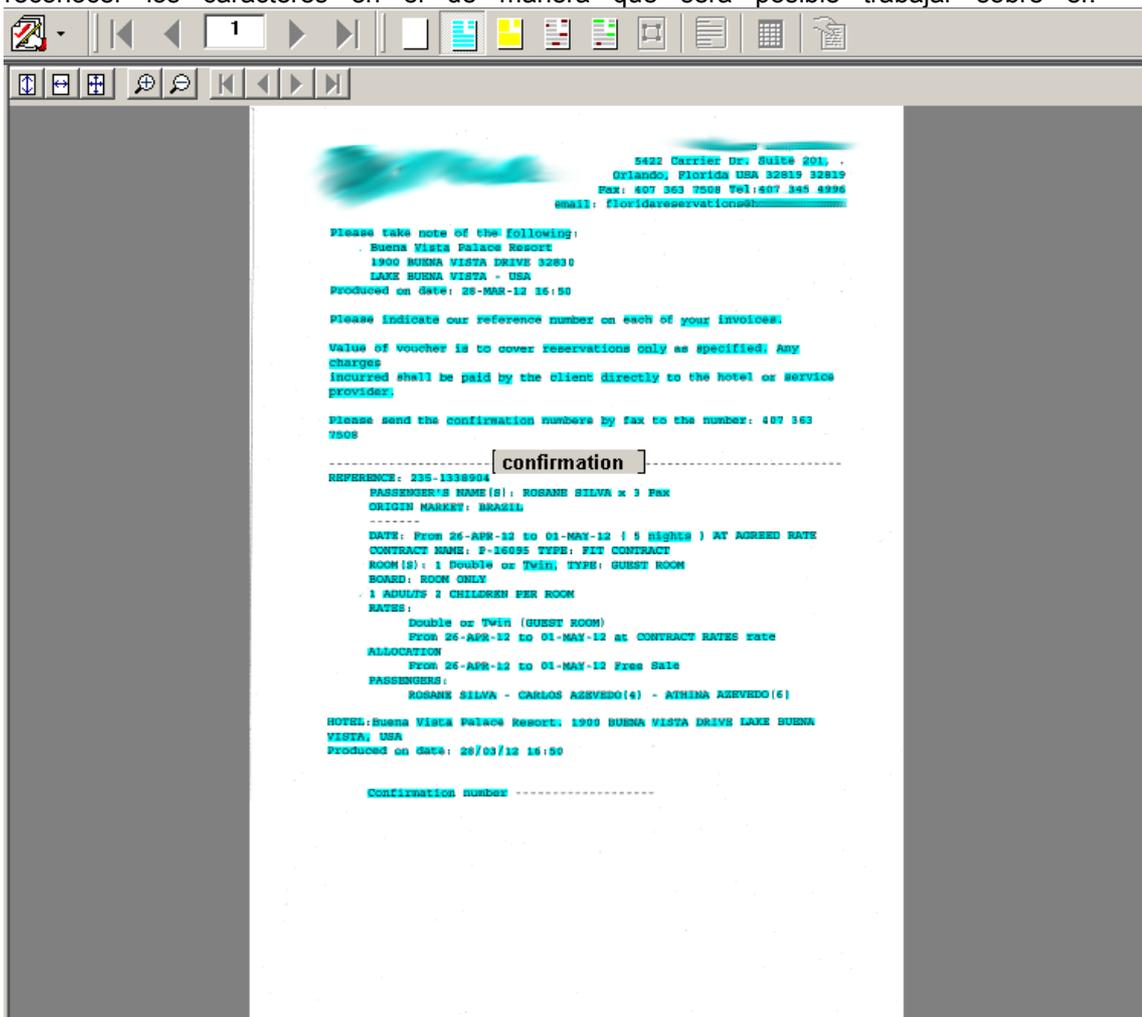


Ilustración 21 Resaltado de texto en Forms Recognition Designer

- B. Resalta bloques de contenido o párrafos. Esto sirve para comprender cómo WebCenter Forms Recognition distingue entre un formulario u otro.

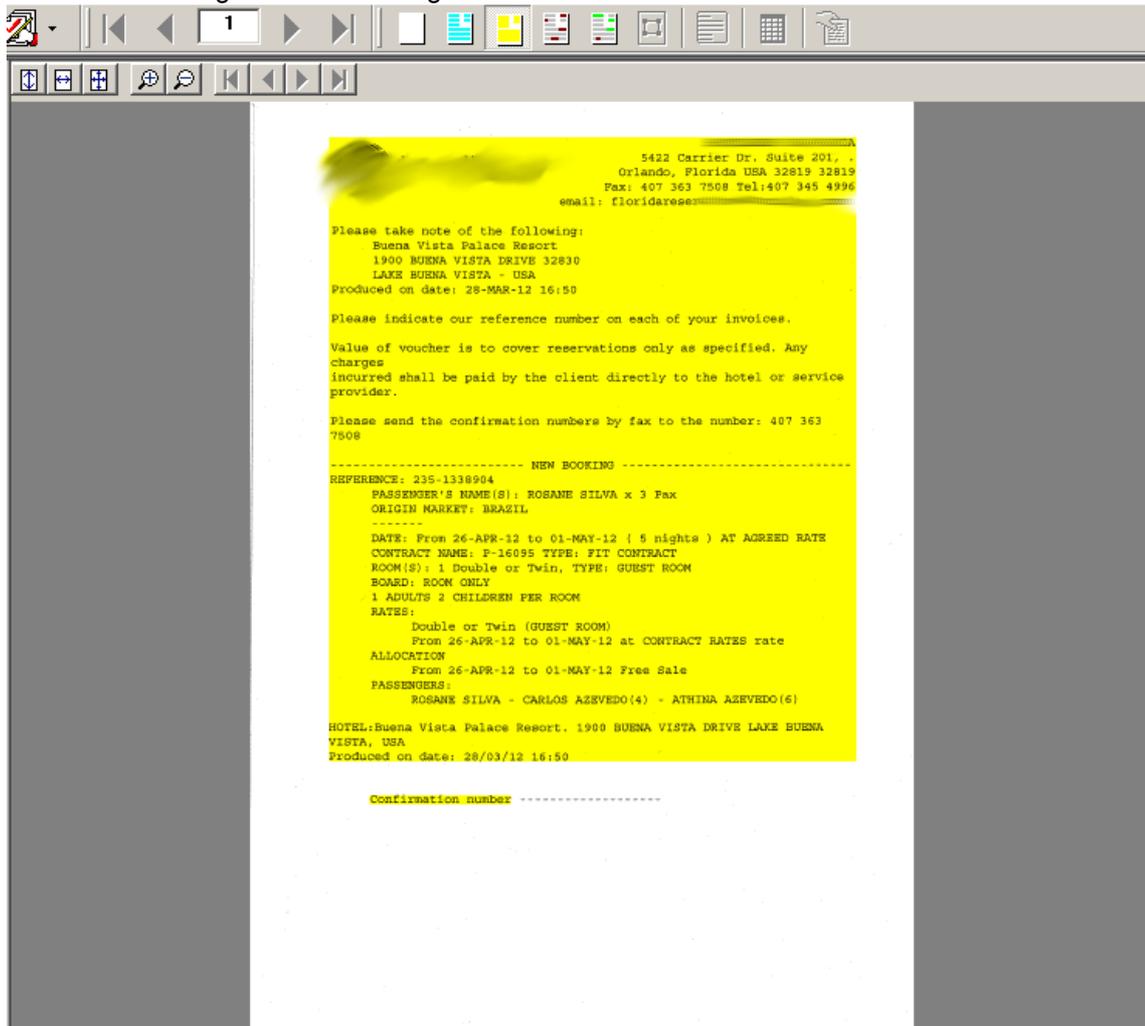


Ilustración 22 Resaltado de bloques en Forms Recognition Designer

- C. Resalta los candidatos que encajan con las reglas establecidas para la extracción automática con un determinado campo. Se verá más en detalle cuando se describa el diseño del proyecto de reconocimiento específico para la Gestión de Reservas.
- D. Resalta los candidatos como el anterior y además destaca en verde aquellos que ya están asignado a un determinado campo como dato más relevante. Se verá más en detalle cuando se describa el diseño del proyecto de reconocimiento específico para la Gestión de Reservas.

3.7.2. WebCenter Forms Recognition Verifier

De esta aplicación cabe destacar 2 interfaces principalmente. Por un lado, la selección del lote a revisar. Está interfaz muestra todos los detalles importantes a tener en cuenta de cada lote como su estado o la última estación de trabajo y fecha desde la que se accedió al lote. Los usuarios administradores pueden configurar esta aplicación para que o bien muestre todos los lotes asociados a un determinado proyecto o bien que cargue los lotes que se encuentran en una carpeta y que cargue la configuración de cada proyecto asociado una vez se seleccione un lote.

Batch ID	State	Priority	Name	Folders	Documents	Client	Last Workstation	Last Module	Last Access
00000086	800	5	Batch_86	1	1	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 4:53:16 PM
00000087	800	5	Batch_87	2	2	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 1:27:56 PM
00000088	800	5	Batch_88	1	1	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 1:27:59 PM
00000089	800	5	Batch_89	1	1	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 1:28:00 PM
00000090	800	5	Batch_90	1	1	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 1:28:02 PM
00000091	800	5	Batch_91	1	1	N/A	OSC-VIRTUALBOX	Designer	6/4/2015 1:28:01 PM
00000092	800	5	Batch_92	1	1	N/A	TUI-91BD9C9CA34	Designer	5/23/2012 5:07:18 PM
00000093	800	5	Batch_93	1	1	N/A	SA5DGT01	Designer	11/20/2012 7:29:22 PM
00000094	800	5	Batch_94	1	1	N/A	SA5DGT01	Designer	11/20/2012 7:29:24 PM
00000095	800	5	Batch_95	1	1	N/A	SA5DGT01	Designer	11/21/2012 9:52:34 AM
00000096	800	5	Batch_96	1	1	N/A	SA5DGT01	Designer	11/20/2012 7:29:26 PM
00000097	800	5	Batch_97	1	1	N/A	SA5DGT01	Designer	11/21/2012 9:52:36 AM
00000098	800	5	Batch_98	2	2	N/A	TUI-91BD9C9CA34	Server	5/24/2012 8:52:00 PM
00000099	800	5	Batch_99	1	1	N/A	SA5DGT01	Designer	11/21/2012 9:52:36 AM
00000100	800	5	Batch_100	1	1	N/A	SA5DGT01	Designer	11/20/2012 7:31:58 PM
00000101	800	5	Batch_101	1	1	N/A	SA5DGT01	Designer	11/21/2012 9:52:38 AM
00000102	800	5	Batch_102	1	1	N/A	SA5DGT01	Designer	11/21/2012 10:09:32 AM
00000103	800	5	Batch_103	2	2	N/A	TUI-91BD9C9CA34	Server	5/25/2012 11:13:02 AM
00000104	800	5	Batch_104	3	3	N/A	SA5DGT01	Designer	11/21/2012 10:09:34 AM
00000105	800	5	Batch_105	2	2	N/A	SA5DGT01	Designer	11/21/2012 8:01:40 PM
00000106	800	5	Batch_106	1	1	N/A	SA5DGT01	Designer	11/21/2012 8:01:42 PM
00000107	800	5	Batch_107	1	1	N/A	TUI-91BD9C9CA34	Server	5/25/2012 12:42:26 PM
00000108	800	5	Batch_108	1	1	N/A	TUI-91BD9C9CA34	Server	5/25/2012 12:45:14 PM
00000109	800	5	Batch_109	1	1	N/A	SA5DGT01	Designer	11/21/2012 8:01:38 PM
00000110	800	5	Batch_110	1	1	N/A	SA5DGT01	Designer	11/21/2012 10:09:42 AM
00000111	550	5	Batch_111	1	1	N/A	SA5DGT01	Designer	11/21/2012 10:09:46 AM
00000112	800	5	Batch_112	1	1	N/A	SA5DGT01	Designer	11/21/2012 10:10:42 AM
00000113	800	5	Batch_113	2	2	N/A	SA5DGT01	Designer	11/21/2012 10:11:36 AM
00000114	800	5	Batch_114	2	2	N/A	SA5DGT01	Designer	11/21/2012 10:43:48 AM
00000115	550	5	Batch_115	2	2	N/A	SA5DGT01	Designer	11/21/2012 8:01:30 PM
00000116	550	5	Batch_116	17	17	N/A	SA5DGT01	Designer	11/21/2012 11:17:42 AM
00000117	550	5	Batch_117	3	3	N/A	SA5DGT01	Designer	11/21/2012 10:58:06 AM
00000118	550	5	Batch_118	1	1	N/A	SA5DGT01	Designer	11/21/2012 8:01:50 PM
00000119	800	5	Batch_119	1	1	N/A	SA5DGT01	Designer	11/21/2012 8:02:24 PM
00000120	800	5	Batch_120	1	1	N/A	SA5DGT01	Server	11/21/2012 7:53:54 PM
00000121	800	5	Batch_121	1	1	N/A	SA5DGT01	Server	11/21/2012 7:54:08 PM
00000122	800	5	Batch_122	1	1	N/A	SA5DGT01	Server	11/21/2012 7:54:12 PM
00000123	550	5	Batch_123	1	1	N/A	SA5DGT01	Designer	11/21/2012 8:02:34 PM

Ilustración 23 Vista de lotes en Forms Recognition Verifier

Los datos que se muestran para cada lote son:

- **Batch ID:** identificador único y autoincrementado para cada lote
- **State:** un valor numérico que indica cómo se encuentra el lote. Se pueden configurar distintos valores numéricos en función de las necesidades del proyecto. Por defecto los estados más relevantes son:
 - **100** - Lote importado ya sea a través del Runtime Server o cargado mediante el Designer
 - **150** - Lote importado pero error al ejecutar el motor OCR sobre alguno de sus documentos
 - **200** – Lote importado y con OCR aplicado
 - **250** – Error al tratar de clasificar algún documento del lote. Necesita verificación desde el Verifier.
 - **300** – Lote con todos sus documentos correctamente clasificados
 - **550** – Lote con todos sus documentos clasificados pero con errores en la extracción de campos. Necesita verificación desde el Verifier.
 - **700** – Lote correctamente clasificado y extraído, listo para ser exportado
 - **750** – Lote con errores en la exportación
 - **800** – Lote correctamente exportado
 - Por defecto únicamente es posible abrir lotes en los estados **250** (Error de clasificación) y **550** (Error de extracción) en el Verifier ya que presentan errores

corregibles por un usuario desde esta herramienta. Los administradores pueden modificar este comportamiento.

- **Name:** similar al **Batch ID** pero, a diferencia de este, los administradores pueden asignar un prefijo o sufijo para identificar mejor el origen o propósito de los documentos incluidos en el lote
- **Folders y Documents:** de cuantas carpetas y documentos está formado el lote
- **Last Workstation:** nombre del último servidor desde el que se abrió el lote. Corresponde con el nombre de la máquina Windows desde la que se abrió por última vez el lote.
- **Last Module:** indica el último módulo o aplicación de Forms Recognition que accedió al lote. Puede ser cualquiera de los 3 anteriormente descritos: Designer, Verifier o Runtime Service (indicado como Server).
- **Last Access:** fecha del último acceso al lote.

Cuando los usuarios verificadores seleccionan un lote se les presenta el formulario que el diseñador del proyecto preparó en el Designer.

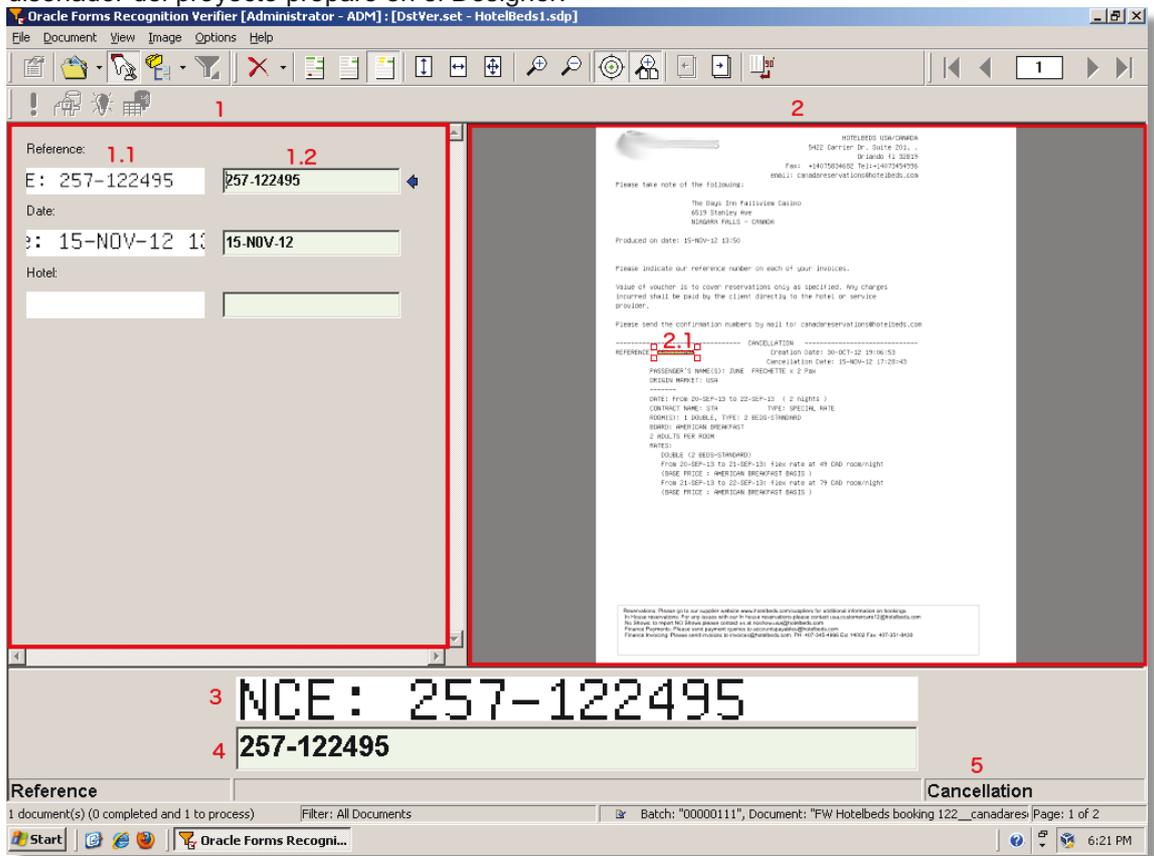


Ilustración 24 Formulario de Forms Recognition Verifier

Además de los controles en la parte superior para eliminar el documento del lote, orientar la imagen o navegar a través de los distintos documentos, estos son los elementos más destacables de esta interfaz:

1. Formulario de Verificación, previamente configurado en el Designer
 - 1.1. Parte de la imagen del documento de la que se ha extraído el campo 1.2
 - 1.2. El dato extraído mediante OCR. Este campo se resalta en rojo si debe ser corregido
2. Visualización de la imagen del documento completa
 - 2.1. Recuadro que resalta el campo que se selecciona del formulario en la imagen completa
3. Visualización aumentada de la parte de la imagen que corresponde al dato extraído
4. Visualización aumentada del campo seleccionado
5. Clasificación a la que pertenece el documento seleccionado

3.7.3. WebCenter Forms Recognition Runtime Service

Con este componente se pueden configurar servicios que ejecuten de manera automática el OCR, la clasificación y extracción sobre documentos ubicados en una carpeta del sistema o compartida en red.

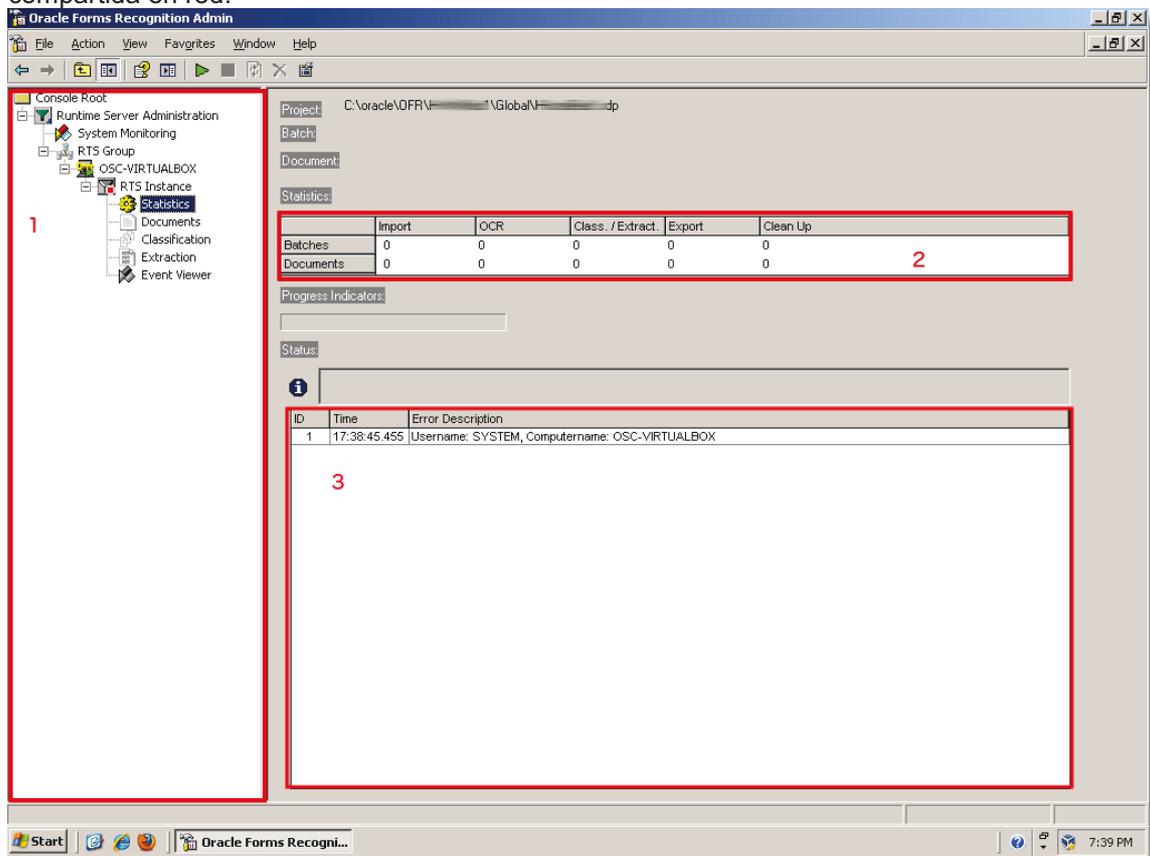


Ilustración 25 Forms Recognition Runtime Service

- 1) En esta navegación se crean agrupaciones de servidores y en cada servidor se crean instancias de Runtime Service. Para cada instancia se debe configurar un proyecto que utilizará para obtener las reglas y el conjunto de aprendizaje, así como las carpetas de importación y/o exportación y qué pasos ejecutará dicha instancia. Los pasos se definen en el siguiente punto.
- 2) Estado del proceso de reconocimiento dividido en 5 pasos. Los estados de 3 cifras mencionados en el apartado de descripción del Verifier sirven para identificar en cuál de estos pasos se encuentra un lote. Como se indicó, existe una numeración de estados por defecto pero puede ampliarse y modificarse.
 - a) **Import:** En este paso se importan los documentos desde una carpeta del sistema o compartida en red y se crean lotes. Puede crearse un lote por cada documento, crear un lote nuevo al llegar a un determinado número o crear un lote por cada subcarpeta de la carpeta de importación.
 - i) **Éxito:** 100
 - ii) **Fracaso:** 50
 - b) **OCR:** este es el paso en el que se aplican los distintos motores de OCR configurados en el proyecto al documento. Supone bastante carga para el servidor por lo que es buena práctica configurar una instancia de RTS dedicada exclusivamente a realizar OCR y en cualquier caso no se recomienda tener más de una instancia de RTS realizando OCR al mismo tiempo.
 - i) **Éxito:** 200
 - ii) **Fracaso:** 150
 - c) **Classify / Extraction:** En realidad son dos pasos en uno, primero se clasifica el documento y, en base a dicha clasificación, se extraen los campos pertinentes.
 - i) **Éxito en la clasificación:** 300

- ii) **Fracaso en la clasificación:** 250
 - iii) **Éxito en la extracción:** 700
 - iv) **Fracaso en la extracción:** 550
 - d) **Export:** En este paso se copian o mueven las imágenes de los documentos junto con un fichero con los campos extraídos separados por comas. Puede configurarse que se exporte una conversión a PDF del documento. Además, en este paso reside la mayor capacidad de extensibilidad de Forms Recognition ya que pueden invocarse Scripts Visual Basic que ejecuten la lógica que se desee. Entre otras cosas, suele utilizarse para personalizar el formato de los ficheros de salida para adaptarlo al formato de entrada de otras aplicaciones.
 - i) **Éxito:** 800
 - ii) **Fracaso:** 750
 - e) **Clean Up:** Este paso únicamente elimina los lotes del sistema una vez han sido exportados con éxito. Es habitual tener una instancia de RTS que únicamente da este paso a intervalos regulares con el fin de mantener los lotes en el sistema un tiempo prudencial en caso de ser necesario recuperarlos pero sin llegar a colapsar el almacenamiento del servidor.
- 3) Salida de la ejecución del proceso de reconocimiento. Todo lo que se muestra también se escribe a ficheros log para realizar auditorías. Tiene 4 niveles de configuración
- a) **No logging:** no se traza ninguna actividad
 - b) **Info:** se trazan todos los mensajes
 - c) **Warning:** se trazan errores y advertencias
 - d) **Error:** únicamente se trazan los errores

En el siguiente diagrama de flujo se resume la operación habitual de WebCenter Forms Recognition, incluyendo los estados que por defecto tienen los lotes en los distintos pasos.

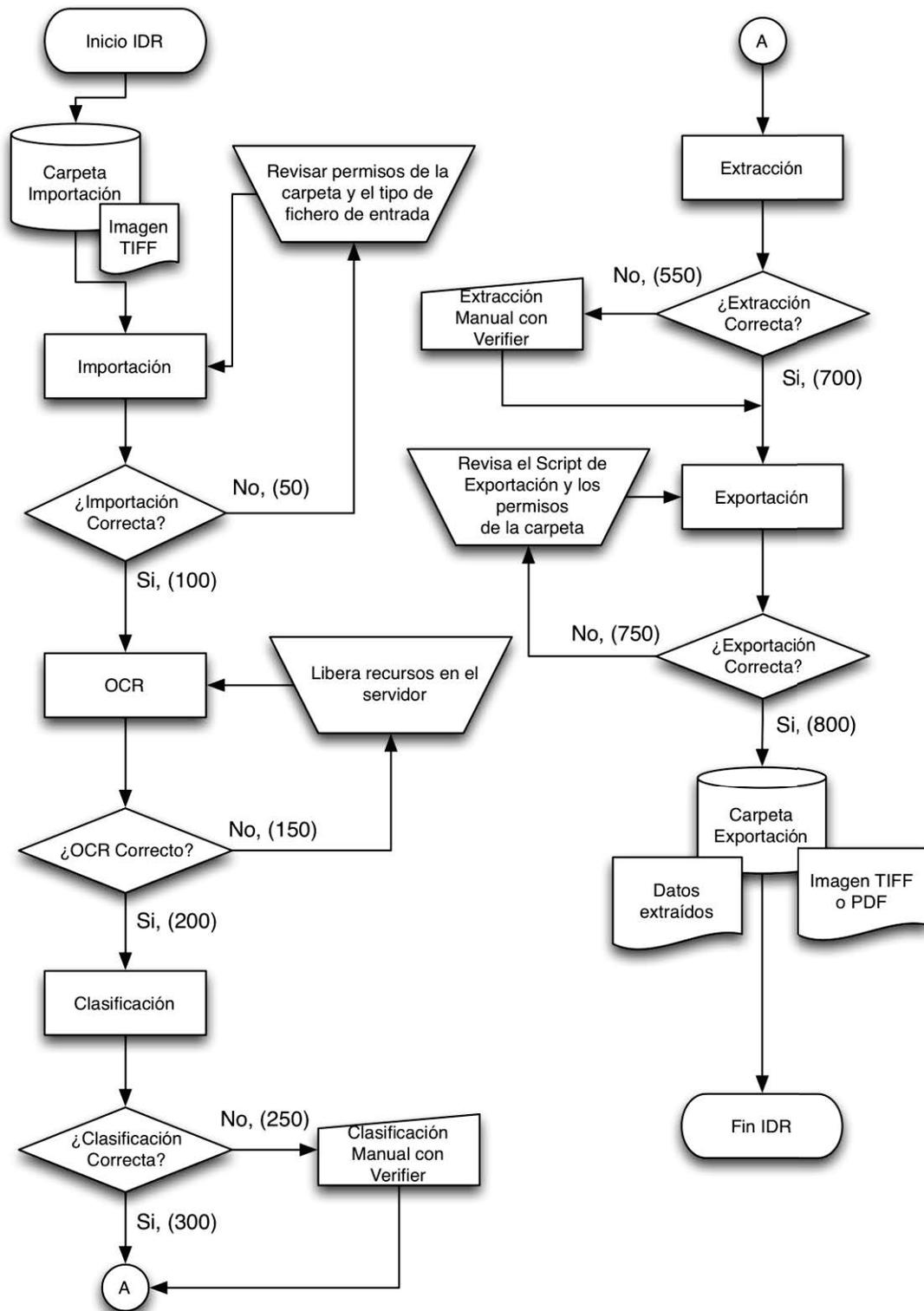


Ilustración 26 Flujo de ejecución típico de Forms Recognition

3.8. Solución de Gestión Documental para la Automatización de la Gestión de Reservas

Hasta este punto se han estudiado los componentes que pueden tener los Gestores Documentales según la AIIM, máxima autoridad en lo que a la Gestión de la Información se refiere. De ese estudio y en base a las necesidades identificadas para la Gestión de Reservas, se identificaron 4 componentes básicos para la construcción de un Gestor Documental para la Automatización de la Gestión de Reservas.

Acto seguido, se estudió la oferta actual de estos componentes en el mercado así como soluciones pre-integradas para otras áreas como es la Gestión de Cuentas a Pagar. De este estudio se concluye que no existe una solución específica para la Gestión de Reservas y, por lo tanto, se decide construir dicha solución a partir de los 4 componentes identificados.

Al inicio de esta sección, se justifica la decisión de optar por Oracle como proveedor único para los 4 componentes y se describen detalladamente cada uno de los 4, insistiendo en aquellas aplicaciones e interfaces más relevantes para la construcción del Gestor Documental para la Automatización de la Gestión de Reservas.

En este apartado se esquematizan las distintas aplicaciones descritas anteriormente, indicando el perfil de usuario al que están orientadas y resumiendo su utilidad con el fin de explicar brevemente el trabajo realizado para la construcción del presente proyecto.

Solución	Aplicación	Tipo	Perfil	Utilidad
WebCenter Enterprise Capture	Capture Client	Interfaz de Usuario	Personal administrativo.	Subir o escanear documentos manualmente
	Capture Workspace	Interfaz de Usuario	Administrador / desarrollador.	Permite configurar Trabajos Automatizados de Captura y Metadatos a Capturar
	Import Processor Job	Proceso en 2º plano	N/A	Captura imágenes desde carpetas, servidores de correo, faxes...
	Recognition Processor Job	Proceso en 2º plano	N/A	Reconoce códigos de barras
	Conversion Processor Job	Proceso en 2º plano	N/A	Convierte documentos electrónicos a imágenes
	Commit Processor Job	Proceso en 2º plano	N/A	Envía las imágenes capturadas al Gestor Documental o Sistema de Ficheros
WebCenter Forms Recognition	Designer	Interfaz de Usuario	Administrador / Desarrollador	Crea y configura proyectos de reconocimiento
	Verifier	Interfaz de Usuario	Personal administrativo	Supervisa la clasificación y extracción automatizada
	Runtime Service Management	Interfaz de Usuario	Administrador	Crea y monitoriza servicios de

	Console			reconocimiento automáticos
	Runtime Service	Proceso en 2º plano	N/A	Ejecuta las reglas definidas en un proyecto de reconocimiento automáticamente sobre imágenes
WebCenter Imaging	Imaging Console	Interfaz de Usuario	Administrador	Configura Entradas automatizadas al gestor documental, metadatos a indexar, búsquedas, etc
	Imaging Viewer	Interfaz de Usuario	Usuario de Negocio	Permite ver, anotar, censurar, remarcar los documentos digitalizados y sus metadatos asociados
	Input Agent	Proceso en 2º plano	N/A	Recoge imágenes junto con sus metadatos asociados en documentos separados por comas y las ingresa en WebCenter Imaging
BPM Suite	Business Process Composer	Interfaz de usuario	Administrador	Permite diseñar y ejecutar procesos de negocio
	BPM Studio	Interfaz de usuario	Desarrollador	Permite a los desarrolladores de BPM colaborar con los usuarios de Process Composer para desarrollar las actividades o interfaces más complejas.
	BPM Workspace	Interfaz de Usuario	Usuario de Negocio	Espacio donde los usuarios pueden ejecutar sus tareas dentro de un proceso en ejecución

Tabla 12 Resumen de componentes de la solución de Oracle

Habiendo dejado claro el listado de aplicaciones, en el siguiente diagrama se muestran cuáles y como se han utilizado en el contexto de la problemática de los Bancos de Cama explicada en la introducción.

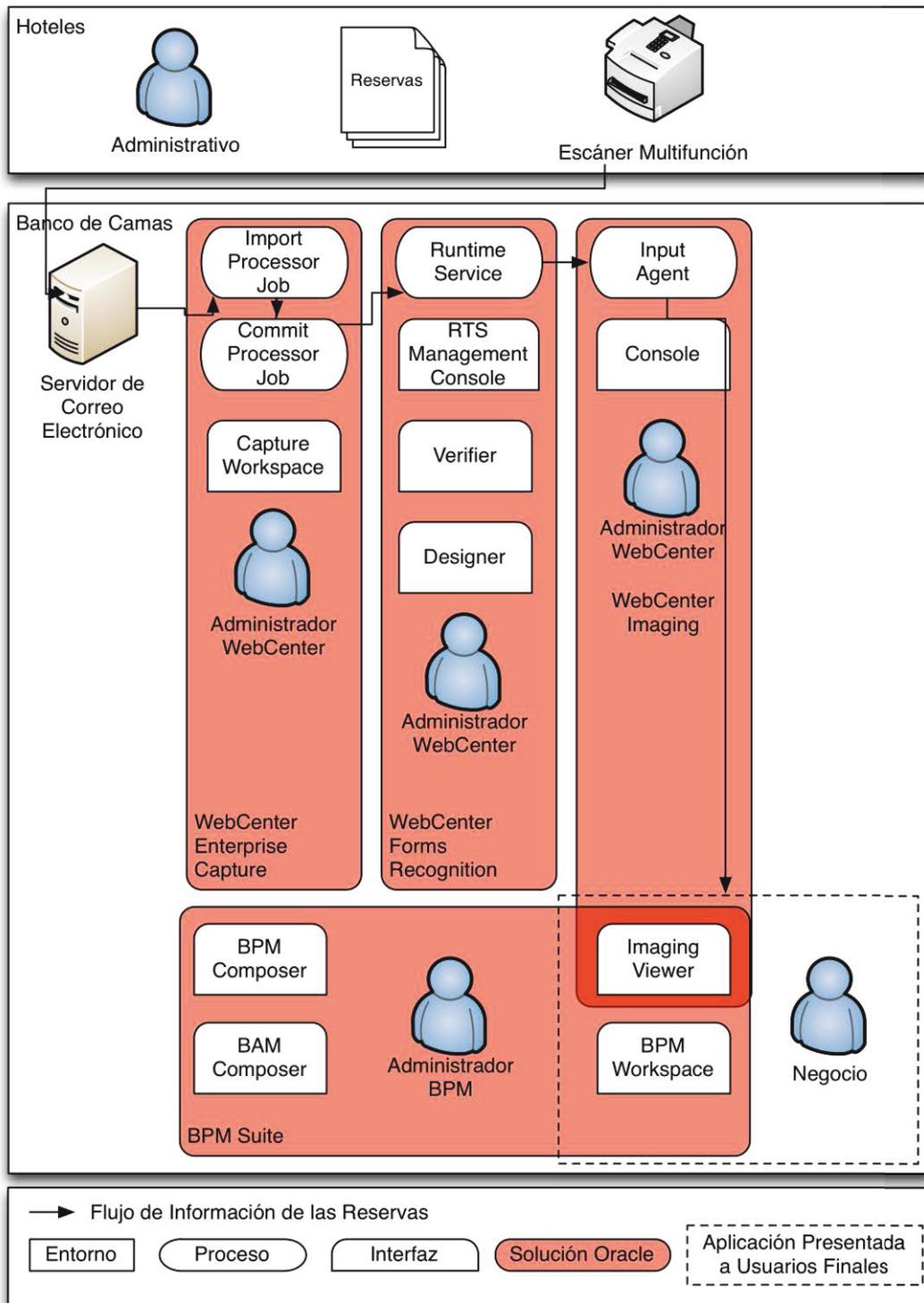


Ilustración 27 Diagrama de alto nivel del proyecto a implantar

El anterior diagrama trata de dejar claro los siguientes conceptos, vitales para comprender la solución completa en conjunto, tanto lo que van a ver los usuarios finales como lo que no, así como el trabajo realizado por el autor de la presente memoria dentro de la misma:

- Los hoteles escanean y envían documentos de Reservas al Banco de Camas como adjuntos a correos electrónicos

- La solución captura dichos adjuntos mediante un trabajo de Importación de WebCenter Enterprise Capture y los envía a una carpeta mediante un trabajo de Confirmación.
 - El Administrador/Desarrollador de WebCenter utiliza la interfaz de Capture Workspace para configurar debidamente dichos trabajos
- La solución procesa los adjuntos, clasificándolos y extrayendo información automática con el Runtime Service de WebCenter Forms Recognition
 - El Administrador/Desarrollador de WebCenter utiliza las 3 interfaces de Forms Recognition para crear el proyecto de reconocimiento, entrenarlo y probar los resultados
- La solución recoge los adjuntos junto con su información extraída y lo introduce en WebCenter Imaging mediante el Input Agent
 - El Administrador/Desarrollador de WebCenter utiliza la consola de Imaging para configurar correctamente el Input Agent y los metadatos a indexar
- Los usuarios de negocio acceden a BPM Workspace o reciben notificaciones por correo cuando tienen que revisar la entrada de una nueva reserva.
 - El Administrador/Desarrollador de BPM debe construir el proceso y los formularios utilizando BPM Composer. Integrará el Viewer de Imaging en los formularios con el fin de que los usuarios vean el documento original del que provienen los datos que deben aprobar
- Los usuarios de negocio podrán ver cuadros de mando de Business Activity Monitoring para monitorizar la actividad que está generando el proceso de recepción de reservas
 - El Administrador/Desarrollador de BPM debe construir dichos cuadros de mando con BAM Composer.

Los perfiles identificados como Administrador/Desarrollador de WebCenter y BPM son, durante la implantación del proyecto, además los encargados de:

- Analizar los requisitos funcionales y técnicos
- Asegurar que el entorno cumple con las certificaciones para poder instalar el Software de Oracle
- Instalar y configurar el Software de Oracle aplicando las mejores prácticas para obtener entornos productivos de alta disponibilidad
- Realizar los desarrollos y extensiones a las aplicaciones que sean necesarias para cubrir los requisitos
- Identificar posibles riesgos y sugerir alternativas
- Documentar todas las tareas de instalación y configuración para poder realizar un traspaso de conocimiento adecuado
- Traspaso de conocimiento a las personas que tomarán la responsabilidad de Administrador/Desarrollador de BPM y WebCenter una vez concluida la implantación del proyecto

El autor de la presente memoria de proyecto desempeña la labor de Administrador y Desarrollador de WebCenter por lo que en las próximas secciones de esta memoria se describirán los aspectos que atañen a dichas soluciones, haciendo puntuales menciones a BPM Suite allá donde sea necesario para no perder la visión global de la solución, tal y como se ha hecho en esta sección.

4. Requisitos

En esta sección se describirán los requisitos tanto técnicos como funcionales para la implantación del EDMS para la automatización del proceso de Gestión de Reservas. La captura de requisitos ayudará a entender que componentes de un EDMS deben implementarse y cómo. Los requisitos se clasificarán en los siguientes tipos:

- Requisito de Captura, identificados por el prefijo CAP
- Requisito Reconocimiento, identificados por prefijo OCR
- Requisito de Proceso de Negocio, identificados por prefijo BPM
- Requisito de Gestión Documental, identificados por prefijo ECM
- Requisitos de Infraestructura, identificados por el prefijo INFRA

Identificador: prefijo-XXX	Nombre	Dependencias de otros requisitos
Descripción		

Tabla 13 Formato de captura de requisitos

CAP-001	Entrada de documentos por correo electrónico	
El sistema debe capturar los archivos adjuntos de 6 cuentas de correo electrónico que reciben un promedio de 1100 correos diarios entre todas.		

CAP-002	Formato electrónico	CAP-001
Independientemente del origen del documento, deben convertirse a formato TIFF con, al menos, 300 dpi de resolución. Esto facilitará la labor de la aplicación del OCR.		

CAP-003	Separación de reservas	CAP-001
Deben separarse automáticamente los correos que contengan información de reservas de aquellos con información sobre promociones, comúnmente denominados como "spam".		

OCR-001	Clasificación automática	CAP-003
Deben reconocerse las reservas y clasificarse en función de si son: <ul style="list-style-type: none"> • Nuevas (New Bookings) • Correcciones (Amendments) • Cancelaciones (Cancellations) 		

OCR-002	Extracción automática	OCR-001
Deben extraerse de manera automatizada, al menos, los siguientes campos: <ul style="list-style-type: none"> • Código de reserva • Fecha de la reserva • Hotel de la reserva 		

ECM-001	Ingreso en el gestor documental	OCR-002
Deben ingresarse de forma automática las imágenes correspondientes a las reservas junto con sus metadatos en el gestor documental para su posterior recuperación.		

ECM-002	Metadatos en el gestor documental	OCR-002
---------	-----------------------------------	---------

El gestor documental debe utilizar los siguientes metadatos para indexar las imágenes:

- Por realizar estadísticas:
 - Remitente
 - Destinatario
- Para poder actualizar la base de datos del banco de camas
 - Tipo
 - Nueva (reserva)
 - Corrección (a una reserva existente)
 - Cancelación (de una reserva existente)
 - Código de la reserva
 - Fecha de la reserva
 - Hotel de la reserva

BPM-001	Arranque de proceso	ECM-001
Debe arrancar un proceso de verificación manual/visual cada vez que se ingresa un documento en el sistema para comprobar la extracción automática de metadatos y corregirlos.		

BPM-002	Interfaz de verificación	
Se debe ofrecer una interfaz gráfica que permita a los usuarios visualizar la imagen a la que corresponden los metadatos extraídos para su verificación.		

BPM-003	Interfaz de verificación	
Se debe ofrecer una interfaz gráfica que permita a los usuarios visualizar la imagen a la que corresponden los metadatos extraídos para su verificación.		

INFRA-001	Sistema Operativo	
Se debe instalar el software en cualquiera de los siguientes Sistemas Operativos disponibles:		
<ul style="list-style-type: none"> • Windows <ul style="list-style-type: none"> ○ Windows Server 2003 SP2+ ○ Windows Server 2008 SP1+ • Linux <ul style="list-style-type: none"> ○ Red Hat Enterprise Linux 4.7+ ○ Red Hat Enterprise Linux 5.3+ ○ Red Hat Enterprise Linux 6.1+ ○ Oracle Enterprise Linux 4.7+ ○ Oracle Enterprise Linux 5.3+ ○ Oracle Enterprise Linux 6.1+ 		

INFRA-003	Base de Datos de Preproducción	
El sistema debe conectar con la Base de Datos de Preproducción durante la fase de pilotaje		

INFRA-004	Directorio Activo	
Los usuarios deben ser tomados y creados del Directorio Activo corporativo		

INFRA-005	Virtualización	
-----------	----------------	--

La organización cuenta con una estrategia de virtualización de servidores por lo que los servidores deben ser virtualizables con Oracle Virtual Machine (OVM). Por motivos de agilidad, rapidez en el aprovisionamiento, mejorar la gobernabilidad de las tecnologías de la información y poder realizar auditorías e inventarios de servidores virtuales de manera más sencilla, la organización ha establecido 4 tamaños o configuraciones estándar. Los servidores virtuales pueden tener las siguientes formas o tamaños:

	S	M	L	XL
vCPUs ⁴	1	2	4	8
RAM	4 GB	8 GB	16 GB	32 GB
Almacenamiento	50 GB	100 GB	200 GB	500 GB

Tabla 14 Tamaños Máquinas Virtuales

Esto agiliza el proceso de aprovisionamiento. No obstante, presentando las justificaciones debidas, pueden obtenerse excepciones.

⁴ Virtual CPUs o procesadores virtuales.

5. Arquitectura

Teniendo en cuenta las características de los servidores virtuales que pueden seleccionarse para instalar el Software, es necesario hacer un dimensionamiento de los recursos que necesitarán las soluciones de Oracle para elegir el tamaño de servidor virtual que mejor convenga de entre las configuraciones estándar ofrecidas por la organización, indicadas en el requisito **INFRA-005** de la sección anterior.

5.1 Dimensionamiento predictivo

Oracle ofrece unas tablas de dimensionamiento las cuales tienen en cuenta la arquitectura de sus aplicaciones y las cargas de trabajo que son capaces de soportar, de manera que a la hora de decidir la infraestructura que se va a necesitar, únicamente es necesario aportar datos relativos al uso que se hará de la aplicación.

5.1.1 Enterprise Capture

Para dimensionar los recursos necesarios para ejecutar Enterprise Capture es necesario considerar la siguiente información:

Consideración	Impacto en el rendimiento	Constante	Respuesta
¿Se van a realizar conversiones de Documentos? (Trabajos de Conversión)	Si se realiza → Alto	CONVERSION_SI	Si, para la conversión de correos electrónicos y adjuntos a imagen.
	No se realiza → Bajo	CONVERSIO_NO	
¿Se van a reconocer códigos de barra o códigos de separación?	Si se realiza → Alto	RECONOCIMIEN O_SI	No
	No se realiza → Bajo	RECONOCIMIEN O_NO	
¿En qué formato dejará los documentos Enterprise Capture al ejecutar la confirmación: Imagen TIFF, PDF Imagen o PDF Texto?	PDF Texto → Alto	FORMATO_PDFB	Imagen TIFF
	PDF Imagen → Medio	FORMATO_PDFI	
	Imagen TIFF → Bajo	FORMATO_TIFF	
¿Dónde dejará Enterprise Capture los documentos procesados: en el sistema de ficheros, en WebCenter Content o en WebCenter Imaging?	WebCenter Content → Alto	SALIDA_WC	Sistema de Ficheros
	WebCenter Imaging → Alto	SALIDA_WC	
	Sistema de Ficheros → Bajo	SALIDA_SF	
Nº de Documentos a ser procesados en HORAS_PROCESADO	Bajo	NUM_DOCUMENTOS	1100
Nº medio de páginas por documento	Alto	NUM_PAGINAS	3
¿Cuántas horas se estarán procesando documentos al día?	Inversamente proporcional	HORAS_PROCESADO	24
% de Importación Automática (a través de un Trabajo de Importación)	Bajo	PCT_IMPORTACION	99%
% de Documentos no-imagen (PDF, Word, Excel, PPT)	Alto	PCT_NOIMAGEN	50%

Nº de Usuarios Dedicados concurrentes	Alto	NUM_U_DEDICADOS	
Nº de Usuarios de Ocasiones Concurrentes	Medio	NUM_U_OCASIONALES	6

Tabla 15 Dimensionamiento predictivo de Enterprise Capture

A través de la siguiente fórmula:

$$\begin{aligned}
 M\text{Values} = & \text{NUM}_{U\text{DEDICADOS}} + \text{NUM}_{U\text{OCASIONALES}} \\
 & + \frac{\text{NUM}_{\text{DOCUMENTOS}} \times \text{NUM}_{\text{PAGINAS}} \times \text{PCT}_{\text{IMPORTACION}}}{\text{HORAS}_{\text{PROCESADO}}} \\
 & \times [(PCT_{\text{NOIMAGEN}} \times \text{CONVERSION}_{\text{SI}} | \text{CONVERSION}_{\text{NO}}) \\
 & + (\text{RECONOCIMIENTO}_{\text{SI}} | \text{RECONOCIMIENTO}_{\text{NO}}) \\
 & + (\text{FORMATO}_{\text{PDFB}} | \text{FORMATO}_{\text{PDFI}} | \text{FORMATO}_{\text{TIFF}}) + (\text{SALIDA}_{\text{WC}} | \text{SALIDA}_{\text{SF}})]
 \end{aligned}$$

Se obtienen los M-Values correspondientes a la carga de trabajo que supondría para el servidor el software de Enterprise Capture. Los M-Values son una métrica interna y confidencial que mantiene Oracle para comparar cargas de trabajo. Prácticamente todos los servidores disponibles en el mercado tienen un M-Value asociado.

Para el ejercicio de dimensionamiento basta con saber que la anterior fórmula con los datos específicos aportados, resulta en unos 8.000 M-Values que corresponde aproximadamente al 10% de un núcleo de un procesador x86 Intel XEON de 8 núcleos a 2.5 GHz. Este 10% de carga de trabajo no es sobre el 100% de la capacidad sino sobre el 70% para dejar margen para el procesamiento de tareas del sistema operativo, red y demás por lo que la carga real sería apenas un 7% de la capacidad total del procesador mencionado.

Por lo tanto, con los requisitos de procesamiento actuales, para Enterprise Capture sería suficiente con una máquina virtual de tipo S. No obstante, al formar parte de la misma instalación que WebCenter Content, debemos dimensionar dicha solución para decidir sobre qué tamaño de máquina desplegar este software.

5.1.2. WebCenter Content

De manera similar que con WebCenter Enterprise Capture, Oracle también tiene una serie de métricas y constantes que esta vez se traducen en Gigahercios directamente para predecir de manera aproximada la carga de trabajo que supondrá una instalación de WebCenter Content dependiendo de cómo se utilice. Para WebCenter Content se debe dimensionar por un lado la propia aplicación y por otro el almacenamiento ya que se trata del repositorio donde van a parar todos los documentos.

Con respecto a la aplicación se debe considerar:

Consideración	Impacto en el rendimiento	Constante	Respuesta
Interacción			
Usuarios concurrentes	Medio	USUARIOS	50
Páginas por minuto por usuario	Alto	PAGINAS_USUARIO	10
Subidas por segundo (tanto automáticas como manuales)	Alto	SUBIDAS_SEGUNDO	10
Conversiones (afecta principalmente al servidor de conversiones Inbound Refinery)			
Conversión de documentos a PDF por cada 10 minutos	Bajo	CONVERSIONES_PDF	0 (los documentos se recogen en TIFF o PDF y se ingresan sin aplicarles conversión)

			adicional)
Tamaño medio de los documentos	Bajo para el servidor de contenidos. Bajo para el servidor de conversiones	MEDIA_DOC	60 KB
Conversión de imágenes por cada 10 minutos	Medio	CONVERSIONES_IMG	0
Tamaño medio de las imágenes	Bajo para el servidor de contenidos. Medio para el servidor de conversiones	MEDIA_IMG	N/A
Conversión de videos por cada 10 minutos	Alto	CONVERSIONES_VID	0
Duración media de los videos	Bajo para el servidor de contenidos. Alto para el servidor de conversiones	MEDIA_VID	N/A
Configuración			
Contenido Indexado o Transaccional	Indexado → Alto	CONTENIDO_INDEXADO	Transaccional
	Transaccional → Bajo	CONTENIDO_TRANSACCIONAL	
Almacenamiento	Base de datos IBM DB2 → Alto	ALMACENAMIENTO_DB2	Sistema de Ficheros
	Base de datos Microsoft SQLServer → Alto	ALMACENAMIENTO_SQL	
	Sistema de Ficheros → Medio	ALMACANAMIENTO_FS	
	Base de datos Oracle → Medio	ALMACENAMIENTO_ORA	
	Sistema de ingeniería conjunta Exadata → Bajo	ALMACENAMIENTO_EXA	
Indexación del contenido del documento (Full-Text)	Si → Medio	FULLTEXT_SI	Si
	No → Bajo	FULLTEXT_NO	
Metadatos adicionales	0 → Ninguno	META_0	Se deben crear, al menos, 5 metadatos adicionales tal y como se ha analizado en los requisitos
	Hasta 10 → Bajo	META_10	
	Entre 10 y 50 → Medio	META_50	
	Entre 50 y 100 → Alto	META_100	

Tabla 16 Dimensionamiento predictivo de WebCenter Content

La fórmula a aplicar sería:

$$GHZ_{Totales} = GHZ_{CONTENIDO} + GHZ_{CONVERSIONES}$$

Donde

$$GHZ_{CONTENIDO} = GHZ_{ENTRADAS} + GHZ_{INTERFAZ}$$

Donde

$$GHZ_{ENTRADAS} = SUBIDAS_{SEGUNDO} \times [(CONTENIDO_{INDEXADO} | CONTENIDO_{TRANSACCIONAL}) + (ALMACENAMIENTO_{DB2} | ALMACENAMIENTO_{SQL} | ALMACENAMIENTO_{FS} | ALMACENAMIENTO_{ORA} | ALMACENAMIENTO_{EXA})]$$

Y

$$GHZ_{INTERFAZ} = USUARIOS \times PAGINAS_{USUARIO} \times [(ALMACENAMIENTO_{DB2} | ALMACENAMIENTO_{SQL} | ALMACENAMIENTO_{FS} | ALMACENAMIENTO_{ORA} | ALMACENAMIENTO_{EXA}) + ((FULLTEXT_{SI} | FULLTEXT_{NO}) + (META_0 | META_{10} | META_{50} | META_{100}))]$$

Como no se van a realizar conversiones con WebCenter Content, los GHz Totales corresponden a los GHz necesarios para el servidor de Contenidos. Aplicando las fórmulas anteriores combinando las constantes internas y confidenciales de Oracle con la información referida al proyecto de Sistema Gestor de Documentos Electrónicos para la Gestión de Reservas, se obtiene que serían necesarios 1.6 GHz, que equivaldría al 20% de capacidad de un procesador Intel XEON de 8 núcleos a 2.5 GHz. En este caso ya se está asumiendo cierta carga implícita en el procesador.

El otro aspecto a dimensionar para WebCenter Content es el almacenamiento de los documentos. WebCenter Content por defecto utiliza una base de datos para las tablas internas y los metadatos de los documentos mientras que los propios documentos son almacenados en sistema de ficheros. Aunque es posible utilizar la base de datos como repositorio para los documentos también, esto hará que la base de datos aumente drásticamente de tamaño. Esto puede verse en la siguiente tabla comparativa.

Teniendo en cuenta los siguientes datos:

Documentos y Configuración	
Nº inicial de documentos	0 (No se va a digitalizar el histórico)
Documentos nuevos mensuales	33.000 (1.100 documentos diarios)
Revisiones mensuales	0 (Al tratarse de documentos puramente transaccionales, rara vez serán modificados)
Borrados mensuales	0 (Por el momento, no es necesario establecer políticas de borrado de los documentos)
Tamaño medio de los documentos	60 KB
Indexación del contenido (Full-Text)	Sí
% de Documentos con su contenido Indexado	100%
Nº de metadatos personalizados definidos	5
% de utilización de metadatos personalizados	70%

Tabla 17 Dimensionamiento esquema de base de datos de WebCenter Content

Comparamos el tamaño de la base de datos si se utiliza como repositorio o si, por el contrario, se utiliza el sistema de ficheros como repositorio:

	Tamaño de la Base de Datos utilizándola como repositorio para los documentos					Tamaño de la base de datos utilizándola únicamente para indexación y utilizando el sistema de ficheros como repositorio				
Tamaño Inicial	20 MB					20 MB				
Crecimiento Mensual	3609 MB					1676 MB				
Tamaño al final de año (GB)	1	2	3	4	5	1	2	3	4	5
	42	85	127	169	212	20	39	59	79	98

Tabla 18 Comparativa tamaño de Base de Datos

Se puede observar que hay una diferencia de aproximadamente un 210% en el tamaño de la base de datos al optar por un almacenamiento u otro. La ventaja de utilizar la base de datos como repositorio es que se pueden aplicar ciertas tecnologías como la replicación de datos o, más relevante para el dimensionamiento, la compresión. Aplicando compresión de datos en la base de datos, los tamaños a 5 años quedarían de la siguiente manera:

Año	1	2	3	4	5
Tamaño con compresión	34 GB	69 GB	103 GB	137 GB	172 GB

Tabla 19 Tamaño de Base de Datos con compresión activada

La diferencia del tamaño de la base de datos frente a utilizar el sistema de ficheros como repositorio se reduce hasta ser aproximadamente un 170% superior. No hay que engañarse, los documentos seguirán ocupando espacio en el Sistema de Ficheros en y la ocupación total será menor si se decide utilizar la base de datos como repositorio y aplicar compresión. No obstante, esta tecnología tiene un coste adicional en el licenciamiento de la base de datos de Oracle. Por ello, además de porque es más económico y sencillo aumentar el tamaño de un sistema de ficheros que el de una base de datos y porque los documentos tienen una criticidad dentro de los parámetros habituales, se decide utilizar el Sistema de Ficheros como repositorio para WebCenter Content, dejando la opción de utilizar una Base de Datos con encriptación como punto de mejora al proyecto. Debemos asumir una media de 1100 documentos de 200 KB diarios lo que suponen aproximadamente 80 GB al año.

En conclusión, la parametrización expuesta para el proyecto del Sistema Gestor de Documentos Electrónicos para la Automatización de la Gestión de Reservas para las soluciones WebCenter Enterprise Capture y WebCenter Content supone, a priori, una carga de trabajo equivalente al 27% de la capacidad de un procesador Intel Xeon de 8 núcleos a 2,50 GHz y un almacenamiento de 400 GB a lo largo de 5 años

La máquina virtual de tamaño L, cuyas características son:

- 4 vCPUs
- 16 GB RAM
- 200 GB de almacenamiento

Sería más que suficiente puesto que las vCPU asignadas quedarían aproximadamente a un 54% de carga, lo que daría cabida a crecimiento, y el almacenamiento es suficiente para no tener que considerar un aumento del mismo en casi 3 años sin aplicar si quiera reglas de borrado de documentos.

5.1.3. WebCenter Forms Recognition

Como se describió anteriormente, es recomendable que el proceso de OCR se ejecute en exclusiva en un procesador. Por esto y porque WebCenter Forms Recognition requiere Windows como Sistema Operativo, se instalará en una máquina virtual distinta a donde se instale WebCenter Content. Para dimensionar WebCenter Forms Recognition hay que considerar los siguientes parámetros:

Parámetro	Valor
Nº medio de documentos a procesar al día	1100
Nº medio de páginas por documento	3
Resolución típica de los documentos	300 DPI
Tamaño medio de los documentos	60 KB
Nº de Proyectos	1
Nº medio de Clases por Proyecto	4
Nº medio de Campos por Clase	3

Tabla 20 Dimensionamiento predictivo Forms Recognition

Estas características son similares a las presentadas en el manual de instalación oficial de WebCenter Forms Recognition en el que se presentan los requerimientos de sistema recomendados para un proyecto de reconocimiento de facturas que reconoce 4000 páginas diarias. Estos requerimientos son:

1. CPU de 4 núcleos Intel Xeon a 2.8 GHz
2. 8 GB de RAM
3. 100 GB de Disco duro

Teniendo en cuenta los tamaños de máquina virtual disponibles, la elección quedaría entre:

Máquina M	Máquina L
2 vCPU	4 vCPU
8 GB RAM	16 GB RAM
100 GB HDD	200 GB HDD

Tabla 21 Selección máquina virtual para Forms Recognition

Una máquina M se podría quedar corta, a priori, de procesador mientras que una L supondría un malgasto de recursos al duplicar la cantidad de RAM y Almacenamiento recomendados.

Como el proyecto de reconocimiento requerido para el Sistema Gestor de Documentos para la Automatización de la Gestión de Reservas es de características inferiores que el escenario documentado, es suficiente con la máquina M.

5.1.4. Unified Business Process Management Suite

De igual forma que se ha realizado para los productos WebCenter, en paralelo se realiza un dimensionamiento de BPM Suite tras el cual se deduce que se va a necesitar una máquina XL.

5.2. Alta Disponibilidad y Balanceo de Carga

Con el fin de desplegar el proyecto de manera que sea capaz de soportar grandes picos de carga de trabajo, más allá de los previstos mediante el dimensionamiento predictivo, es necesario comprender y configurar los siguientes componentes

- **Balanceador de carga corporativo de la organización:** configurado para, mediante algoritmo de Round Robin distribuir las peticiones equitativamente a los servidores que forman un mismo clúster
- **Oracle HTTP Server (OHS):** Servidor Web basado en Apache. Puede configurarse como balanceador de carga en caso de que la organización no contase con uno. En el caso de esta implantación, el balanceador de carga es externo y se comunicará con OHS para el envío de la petición al servidor WebLogic adecuado, no siendo necesario especificar el puerto al que se manda la petición, indicando únicamente la aplicación que se quiere utilizar.
- **Node Manager:** Proceso o demonio en constante ejecución en todos los servidores que conforman el clúster WebLogic. Requiere los parámetros de dirección IP del servidor físico donde se ejecuta cada servidor WebLogic y permisos suficientes para ejecutar las rutinas de arranque y parada del servidor. Permite el arranque y parada de los servidores gestionados desde la consola de administración del dominio y, por tanto permite configurar el arranque automatizado de los servidores que forman un clúster en caso de caída de alguno de ellos.
- **Clúster WebLogic:** un clúster WebLogic permite replicar todos objetos clave de una aplicación Java EE como es la sesión, los Enterprise Java Beans (EJB) o las colas de mensajería Java (Java Message Queue, JMS) de manera que de cara al usuario final la caída de un nodo del clúster es totalmente transparente, al estar ejecutándose su transacción de manera distribuida a lo largo de todo el clúster.
- **Alta disponibilidad y balanceo de carga en WebCenter Forms Recognition:** al no estar basado en la infraestructura Oracle Fusion Middleware y, por tanto, no desplegarse sobre WebLogic, las consideraciones de Alta Disponibilidad y Balanceo de Carga son diferentes para WebCenter Forms Recognition.
 - **Balanceo de carga:** El componente Runtime Server balancea la carga automáticamente cuando detecta que otro proceso igual está ejecutando los pasos de OCR, Clasificación, Extracción y Exportación con el mismo proyecto de reconocimiento. La carga se reparte automáticamente mediante el algoritmo de Round Robin entre todos los procesos Runtime Server aplicando las reglas de un mismo proyecto de reconocimiento. Es buena práctica tener un único servidor ejecutando el paso de importación para evitar interbloqueos o corrupción de ficheros.
 - **Alta Disponibilidad:** De igual manera que para balancear la carga, en caso de caída de uno de los servidores que está ejecutando el Runtime Server, el Runtime Server superviviente continuará ejecutando los pasos pendientes desde donde el proceso caído lo dejó. Tanto para Runtime Server como las aplicaciones de Designer y Verifier, es recomendable ubicar el fichero de proyecto, los documentos y carpetas de trabajo en una unidad de almacenamiento compartida en red de manera que siempre esté disponible independientemente desde qué servidor se accede. Esta unidad deberá estar configurada en modo Raid 0 mínimo para garantizar la consistencia de los datos.

5.2. Arquitectura física

El siguiente diagrama presenta la arquitectura física de acuerdo a los requerimientos analizados anteriormente.

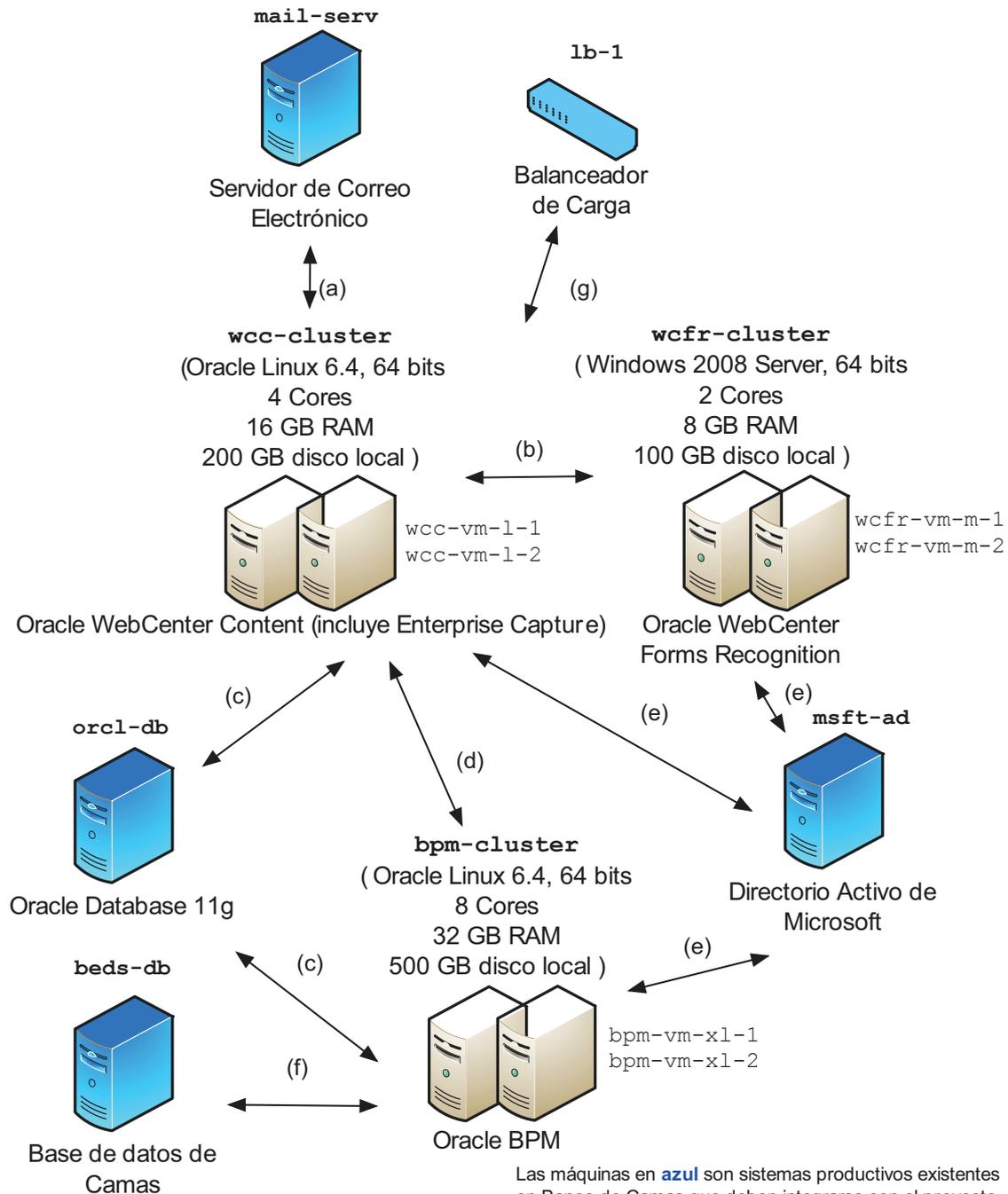


Ilustración 28 Arquitectura Física

En el diagrama se pueden ver los siguientes servidores:

- Servidores existentes en la organización
 - **mail-serv**: Servidor de correo de la organización al que llegan los correos electrónicos con la información de las reservas
 - **orcl-db**: Servidor Base de datos Oracle en producción para su uso en proyectos como el presente
 - **msft-ad**: Servidor del Directorio Activo de Microsoft corporativo, repositorio de todos los usuarios de la organización.
 - **beds-db**: Base de datos de las camas. Esta es la base de datos con información de las reservas y camas disponibles. Se trata de la base de datos

que monetiza la organización y por lo tanto, este proyecto trabajará con el entorno de pruebas.

- Servidores a crear:
 - **wcc-vm-l**: Servidor virtual de tamaño L con Oracle Enterprise Linux 6.4. En este servidor se instalará el software de WebCenter Enterprise Capture y WebCenter Content. Requiere la apertura de los siguientes puertos:
 - **80**: Acceso a internet para la descarga de paquetes Linux e instaladores de Oracle.
 - **7001**: acceso a la consola de administración del dominio WebLogic
 - **16000**: para poder acceder a la aplicación WebCenter Imaging a través de un navegador web.
 - **16200**: para poder acceder al servidor de contenidos de WebCenter Content, también conocido como UCM a través de un navegador web.
 - **16400**: para poder acceder al espacio de trabajo y al cliente de captura de WebCenter Enterprise Capture a través de un navegador web.
 - **wcfr-vm-m**: Servidor virtual de tamaño M con Windows Server 2008 R2. En este servidor se instalará el software de WebCenter Forms Recognition. Únicamente requiere la apertura del puerto 80 para poder acceder a internet para descargar el software de Oracle
 - **bpm-vm-l-xl**: Servidor virtual de tamaño XL con Oracle Enterprise Linux 6.4. En este servidor se instalará el software de BPM Suite. Requiere la apertura de los siguientes puertos:
 - **80**: Acceso a internet para la descarga de paquetes Linux e instaladores de Oracle.
 - **7001**: acceso a la consola de administración del dominio WebLogic
 - **8001**: acceso a las aplicaciones BPM Workspace y BPM Composer
 - **9001**: acceso a la aplicación Business Activity Monitoring (BAM) Composer

Además de presentarse las máquinas a crear para el proyecto y los sistemas existentes con los que la solución deberá interactuar, se detallan las interconexiones entre los distintos sistemas. A continuación se describe cada interconexión:

- a) **Comunicación entre mail-serv y wcc-vm-l**: Conexión de WebCenter Enterprise Capture al servidor de correo para recuperar los correos electrónicos referentes a reservas. Requisitos:
 - i. Cuenta de correo y contraseña para acceder a recuperar los correos de la bandeja de entrada
 - ii. **Protocolos y/o puertos a habilitar**: IMAP vía TLS (993), SMTP vía TLS (465)
- b) **Comunicación entre wcc-vm-l y wcfr-vm-m**: Integración vía carpeta compartida entre Enterprise Capture y Forms Recognition para iniciar el reconocimiento, clasificación y extracción de datos de los correos capturados. Integración inversa en la que WebCenter Content, a través del Input Agent de WebCenter Imaging, recoge los resultados depositados por WebCenter Forms Recognition en una carpeta. Requisitos:
 - i. Permisos de creación y compartición de carpetas en red en Windows
 - ii. Paquete `cifs-utils` o permiso de instalación de paquetes en Linux para el montaje de la carpeta compartida como carpeta en Linux
- c) **Comunicación con orcl-db**: Conexión que establecen los servidores wcc-vm-l y bpm-vm-l con la base de datos en producción para utilizarla como repositorio de configuración (requisito de la arquitectura de Oracle Fusion Middleware). Requisitos:
 - i. **Protocolos y/o puertos a habilitar**: Puerto estándar de comunicación con base de datos Oracle (1521).
- d) **Comunicación entre wcc-vm-l y bpm-vm-xl**: Comunicación entre la solución de WebCenter Content y BPM para provocar el disparo de procesos BPM y recuperar documentos durante el propio proceso BPM de validación. Requisitos:
 - i. **Protocolos y/o puertos a habilitar**: Puerto por defecto de Remote Intradoc Client (RIDC) (4444), puerto por defecto para el acceso a las aplicaciones de WebCenter Content (16000 para Imaging y 16200 para Content Server), puerto por defecto para las aplicaciones de BPM (8001 para Workspace y BPM Composer, 9001 para Business Activity Monitoring)
- e) **Comunicación entre los servidores y el servidor del Directorio Activo de Microsoft para autenticación**: Conexión para que las soluciones Oracle utilicen el

Directorio Activo como repositorio de usuarios en lugar del propio de la infraestructura WebLogic para evitar la creación de duplicidades. Requisitos:

- i. **Protocolos y/o puertos a habilitar:** Puerto por defecto de comunicación con Directorio Activo de Microsoft (389).
- f) **Comunicación entre bpm-vm-xl y beds-db:** el fin último del proyecto que se está implantando es la actualización de dicha base de datos. Requisitos:
 - i. Parámetros de conexión con dicha base de datos.
- g) **Comunicación entre el Balanceador de Carga y cada uno de los servidores:** El balanceador de carga debe conocer las direcciones de los servidores del proyecto así como poder hacer peticiones a ciertos puertos de manera que pueda enrutar debidamente las peticiones entrantes. Requisitos:
 - i. **Protocolos y/o puertos a habilitar:**
 - i. **7001:** acceso a la consola de administración del dominio WebLogic
 - ii. **7001:** acceso a la consola de administración del dominio WebLogic
 - iii. **8001:** acceso a las aplicaciones BPM Workspace y BPM Composer
 - iv. **9001:** acceso a la aplicación Business Activity Monitoring (BAM) Composer
 - v. **16000:** para poder acceder a la aplicación WebCenter Imaging a través de un navegador web.
 - vi. **16200:** para poder acceder al servidor de contenidos de WebCenter Content, también conocido como UCM a través de un navegador web.
 - vii. **16400:** para poder acceder al espacio de trabajo y al cliente de captura de WebCenter Enterprise Capture a través de un navegador web.

5.3. Arquitectura lógica

Como ya se explicó anteriormente, las soluciones Oracle Fusion Middleware tienen una arquitectura común que consta de un dominio WebLogic y uno o varios esquemas en una base de datos. A continuación se muestran los dominios WebLogic junto con sus esquemas relacionados.

5.3.1 Dominio WebLogic y Esquemas para WebCenter Content

En el dominio WebCenter Content se prescindirá del servidor de registros Universal Records Management (URM). Se incluye el servidor de conversiones Inbound Refinery (IBR) puesto que es necesario para convertir correctamente documentos ofimáticos a PDF en el caso de que se presenten documentos con dicho formato.

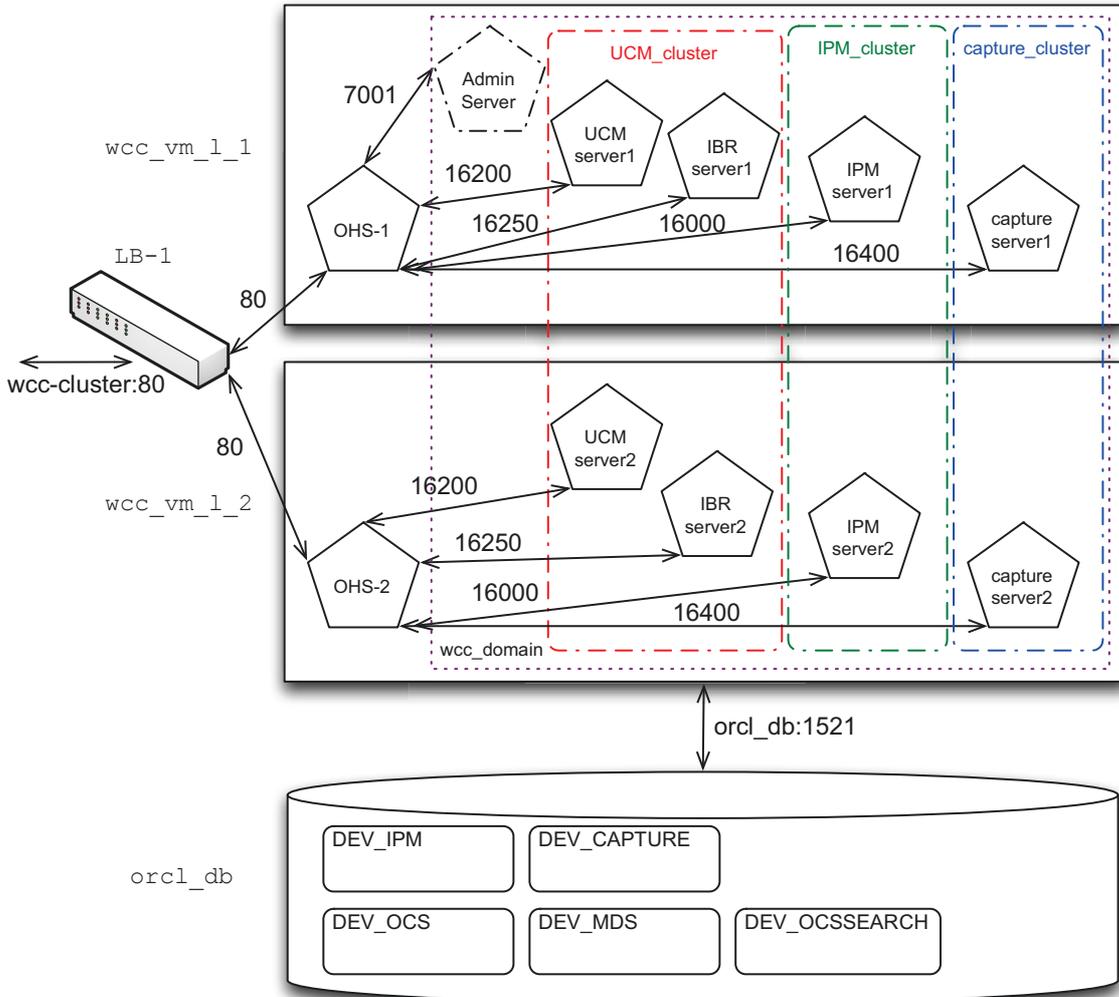


Ilustración 29 Arquitectura lógica de WebCenter Content

- **wcc_domain:** dominio WebLogic para WebCenter Content. Consta del servidor de administración y de 4 servidores gestionados repartidos en 3 clúster.
 - **Admin Server:** servidor de administración del dominio WebLogic llamado wcc_domain. Escucha por el puerto **7001**. El esquema **DEV_MDS** está asociado con el servidor de administración. Se trata de un esquema siempre presente en las instalaciones de Oracle Fusion Middleware que contiene información de configuración general y de personalización de interfaces.
 - **UCM_cluster:** clúster WebLogic formado por los servidores gestionados **UCM_server1**, **UCM_server2**, **IBR_server1** e **IBR_server2**
 - **UCM_serverX:** servidor de contenidos de WebCenter Content. Escucha en el puerto **16200**. Tiene los siguientes esquemas de base de datos asociados:

- **DEV_OCS:** esquema en el que se encuentran todas las tablas de configuración, información de roles, metadatos, etc.
- **DEV_OCSSEARCH:** esquema que incluye las tablas de metadatos y contenidos indexados.
- **IBR_serverX:** servidor de conversiones de WebCenter Content. Escucha en el puerto **16250**. No tiene ningún esquema asociado.
- **IPM_cluster:** clúster WebLogic formado por los servidores gestionados **IPM_server1** e **IPM_server2**
 - **IPM_serverX:** servidor gestionado sobre el que se ejecuta la aplicación WebCenter Imaging (Visor y Consola descritos anteriormente) para el contenido transaccional. Escucha peticiones en el puerto 16000 y tiene asociado el esquema de base de datos **DEV_IPM** para almacenar información de configuración de seguridad, aplicaciones, búsquedas etc.
- **capture_cluster:** clúster WebLogic formado por los servidores gestionados **capture_server1** y **capture_server2**.
 - **capture_serverX:** servidor gestionado sobre el que se ejecutan las 2 aplicaciones de Enterprise Capture: el espacio de trabajo y el cliente de captura. Escucha peticiones en el puerto 16400. Tiene asociado el esquema **DEV_CAPTURE** para almacenar información de configuración y lotes de documentos con los que se esté trabajando antes de ser liberados al sistema de ficheros.

5.3.2. Dominio WebLogic y Esquemas para BPM Suite

La configuración del dominio WebLogic para BPM Suite se hará utilizando únicamente 2 servidores gestionados, desplegando las aplicaciones de BPM Suite sobre el servidor de infraestructura orientada a servicio (SOA).

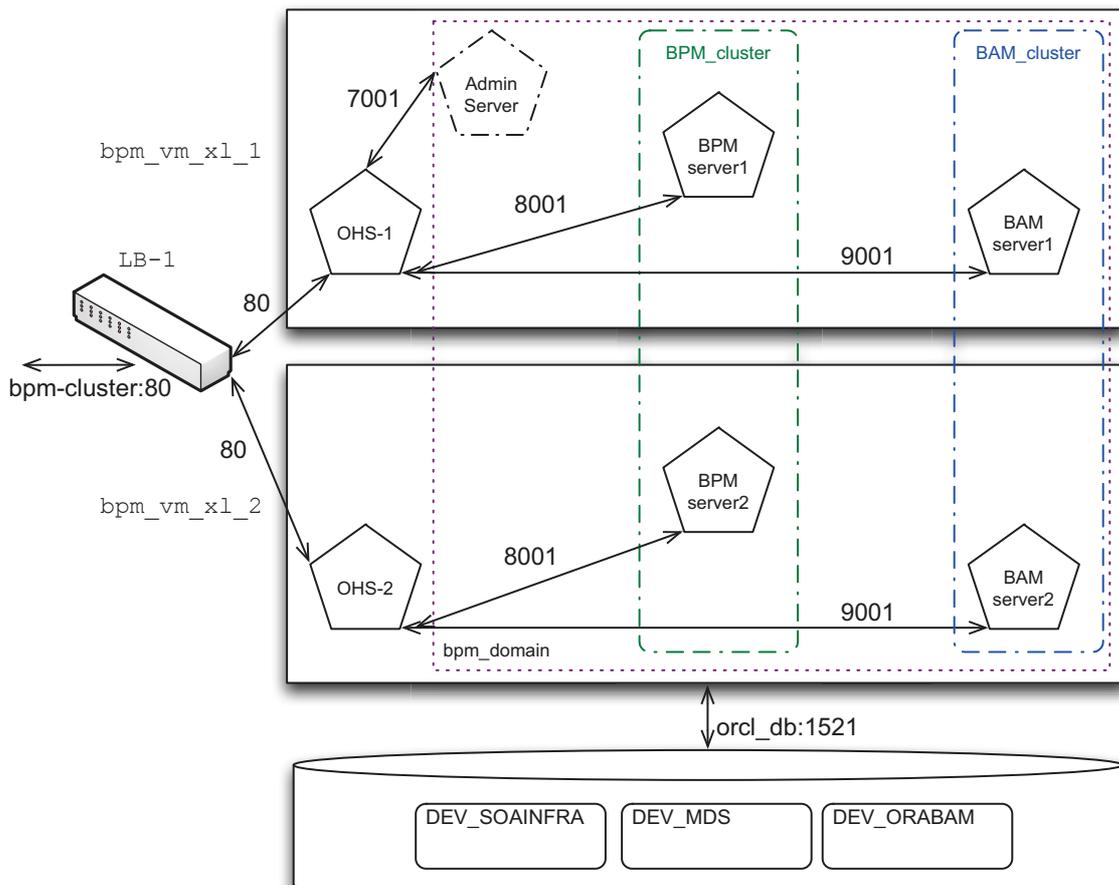


Ilustración 30 Arquitectura Lógica de BPM Suite

- **bpm_domain:** dominio WebLogic para BPM Suite. Consta del servidor de administración y 2 servidores gestionados repartidos en 2 clúster y
- **Admin Server:** servidor de administración del dominio WebLogic llamado wcc_domain. Escucha por el puerto **7001**. El esquema **DEV_MDS** está asociado con el servidor de administración. Se trata de un esquema siempre presente en las instalaciones de Oracle Fusion Middleware que contiene información de configuración general y de personalización de interfaces.
- **BPM_cluster:** clúster WebLogic formado por los servidores gestionados **BPM_server1** y **BPM_server2**
 - **BPM_serverX:** servidor gestionado sobre el que se ejecuta tanto la infraestructura de la arquitectura orientada a servicio como es el motor de ejecución de procesos BPEL. Escucha peticiones a través del puerto **8001** y tiene asociado el esquema **DEV_SOAIFNRA** para el almacenamiento de información de configuración de procesos, formularios y seguridad.
- **BAM_cluster:** clúster WebLogic formado por los servidores gestionados **BAM_server1** y **BAM_server2**
 - **BAM_serverX:** servidor gestionado sobre el que se ejecuta la aplicación Business Activity Monitoring Composer así como los cuadros de mando diseñados por la misma. Escucha peticiones a través del puerto **9001** y tiene asociado el esquema **DEV_ORABAM** para almacenar y explotar las métricas que se definan para ser monitorizadas.

5.3.3. Arquitectura lógica WebCenter Forms Recognition

WebCenter Forms Recognition no es una aplicación desplegada sobre WebLogic por lo que su arquitectura lógica difiere de las anteriores dos.

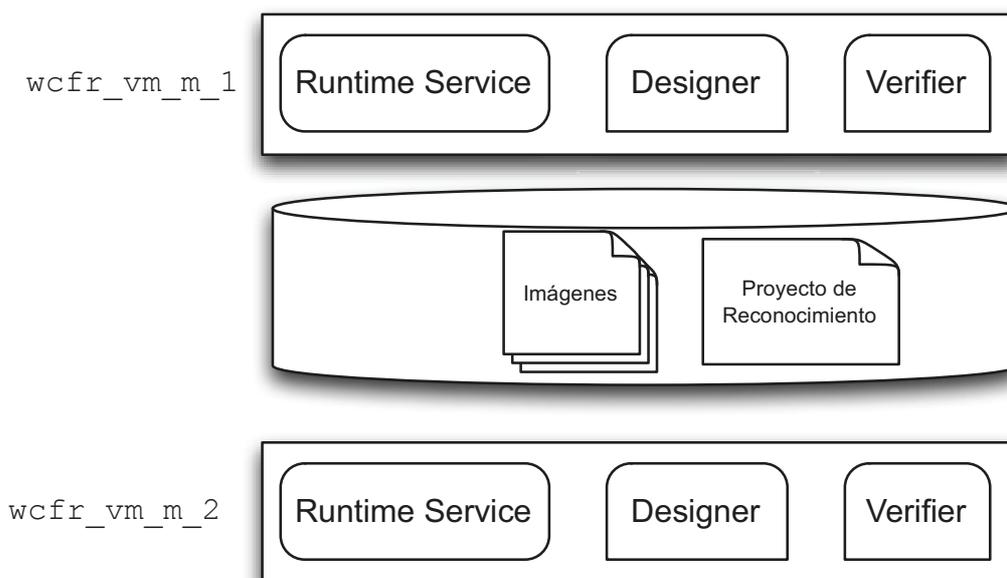


Ilustración 31 Arquitectura Lógica Forms Recognition

Como se ha indicado, el Runtime Service balancea la carga de manera automática. Basta con ubicar las imágenes y el fichero relativo al proyecto de reconocimiento en una ubicación compartida por ambos servidores para que, en caso de caída de uno, se pueda continuar el trabajo desde el otro.

6. Instalación y Configuración de los Productos de Oracle

A continuación se describe el proceso de instalación de los productos de manera esquematizada. Se explicarán en detalle todas las configuraciones realizadas para dar respuesta a la casuística de la Gestión de Reservas de un Banco de Camas, siendo la configuración de WebCenter Forms Recognition el componente con más relevancia para ello.

6.1 Instalación de Oracle Fusion Middleware

El proceso para instalar WebCenter Enterprise Capture, WebCenter Content y BPM Suite es idéntico, siendo la única diferencia la elección de los componentes específicos de cada solución. Esto facilita enormemente el proceso de instalación y reduce sobremanera la necesidad de documentación en esta memoria.

El proceso típico de instalación de productos de FMW se presenta en el siguiente diagrama

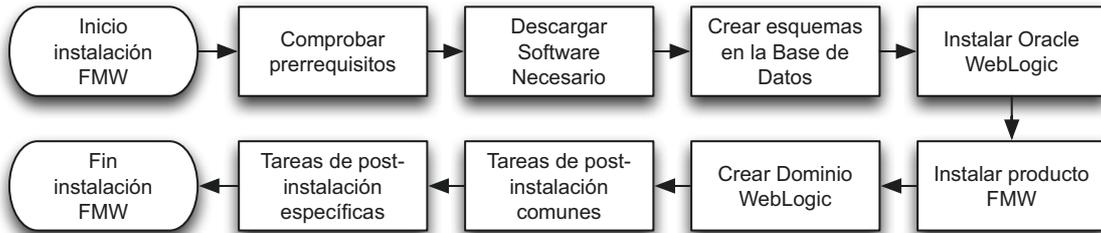


Ilustración 32 Pasos de instalación de Fusion Middleware

Este proceso debe ejecutarse tanto en las máquinas de WebCenter Content como las de BPM Suite. A continuación se explicará brevemente los principales puntos en los que consiste cada paso con el fin de definir lo más importante y relevante sin resultar tedioso.

6.1.1. Comprobar prerequisites

Antes de empezar a instalar cualquier producto de FMW sobre Oracle Linux 6.4 deben comprobarse los siguientes prerequisites.

- El `hostname` y Fully Qualified Domain Name (FQDN) de los servidores deben ser resueltos por el DNS de la organización a la IP de *frontend* de éstos.
- Las máquinas deberán tener creado un usuario de trabajo temporal. Dicho usuario deberá poder tener acceso a comandos de `root` (p.e. `sudo su -`). Este usuario temporal será eliminado una vez finalice la instalación.
- El FQDN del balanceador debe ser resuelto por el DNS de la organización.
- Disponibilidad del servidor de Base de Datos `orcl-db` mencionado en la arquitectura
- Disponibilidad del servidor de Active Directory, `msft-ad` mencionado en la arquitectura
- El sistema operativo deberá tener instalados los siguientes paquetes⁵, mencionados en la captura de requisitos:

- | | |
|--|--|
| • <code>binutils-2.20.51.0.2-5.11.e16-x86_64</code> | • <code>sysstat-9.0.4-11.e16-x86_64</code> |
| • <code>compat-libcap1-1.10-1-x86_64</code> | • <code>gcc-4.4.4-13.e16-x86_64</code> |
| • <code>compat-libstdc++-33-3.2.3-69.e16-x86_64</code> | • <code>gcc-c++-4.4.4-13.e16-x86_64</code> |
| • <code>compat-libstdc++-33-3.2.3-69.e16-i686</code> | • <code>glibc-2.12-1.7.e16-i686</code> |
| • <code>libgcc-4.4.4-13.e16-i686</code> | • <code>glibc-2.12-1.7.e16-x86_64</code> |
| • <code>libgcc-4.4.4-13.e16-x86_64</code> | • <code>glibc-devel-2.12-1.7.e16-x86_64</code> |
| • <code>libstdc++-4.4.4-13.e16-x86_64</code> | • <code>glibc-devel-2.12-1.7.e16</code> |
| • <code>libstdc++-4.4.4-13.e16-i686</code> | • <code>libaio-0.3.107-10.e16-x86_64</code> |
| • <code>libstdc++-devel-4.4.4-13.e16-x86_64</code> | • <code>libaio-devel-0.3.107-10.e16</code> |

- El sistema operativo deberá tener los siguientes parámetros configurados en `/etc/security`:

⁵ La versión de los paquetes se corresponde con RHEL u OL 6. Versiones anteriores de SO hará que las versiones de las librerías sean distintas.

- o hardnofiles = 4096
- o softnofiles = 4096
- **Configurar las siguientes variables de entorno:**
- ORACLE_BASE = /u01/app/ofm/
- FMW_HOME = \$ORACLE_BASE/fmwhome
- JAVA_HOME = \$ORACLE_BASE/javahome
- ECM_HOME = \$FMW_HOME/Oracle_ECM1
 - o (Sólo en wcc-vm-1-1 y wcc-vm-1-2)
- SOA_HOME = \$FMW_HOME/Oracle_SOA1
 - o (Sólo en bpm-vm-xl-1 y bpm-vm-xl-2)
- OHS_HOME = \$FMW_HOME/Oracle_WT1
- DOMAIN_HOME = \$ORACLE_BASE/user_projects
- INSTANCE_HOME = \$DOMAIN_HOME/instances
 - o (Sólo en bpm-vm-xl-1 y bpm-vm-xl-2)
- WLS_HOME = \$FMW_HOME/wlserver_10.3

6.1.2. Descargar Software Necesario

En la siguiente tabla se muestra el software a descargar, junto con su versión y dónde deberá descargarse. El software se descargará desde la nube de descarga de Oracle, <http://edelivery.oracle.com>.

Software	Versión	Máquina(s) destino
Oracle Fusion Middleware Repository Creation Utility (RCU) 11g	11.1.1.9.0	Todas
Oracle JRockit JDK	R28.2.7	Todas
Oracle WebLogic Server 11g	10.3.6	Todas
Oracle WebCenter Content 11g	11.1.1.9.0	wcc-vm-1-1 y wcc-vm-1-2
Oracle SOA Suite 11g	11.1.1.7.0	bpm-vm-xl-1 y bpm-vm-xl-2
Oracle Fusion Middleware Web Tier Utilities 11g	11.1.1.7.0	Todas

Tabla 22 Software Necesario

Pese a existir la versión 12c (12.1.3) de Oracle SOA Suite, que aporta grandes novedades a BPM como la posibilidad de crear Arquitectura de Negocio o visualizar los informes de BAM en más navegadores, se opta por instalar la versión 11g porque dichas novedades no son críticas para la implantación del proyecto y, sin embargo, introducirían una mayor complejidad al requerir una versión de Oracle WebLogic Server distinta (la versión 12c, 12.1.1). Se recomienda la migración a la versión 12c una vez esté disponible también para WebCenter Content.

La diferencia entre la versión menor del Repository Creation Utility y WebCenter Content (11.1.1.9.0) y la de SOA Suite (11.1.1.7.0) no es motivo de preocupación ya que están certificadas para funcionar unas con otras ya que, mientras que la siguiente versión a SOA Suite es la versión 12c, WebCenter Content ha tenido 2 versiones más en 11g (11.1.1.8.0 y 11.1.1.9.0) por lo que lo relativo a la integración con BPM no se ha modificado desde la versión 11.1.1.7.0.

La principal diferencia entre las versiones 11.1.1.7.0 y 11.1.1.9.0 de WebCenter Content es que desde la versión 11.1.1.8.0, WebCenter Enterprise Capture se incluye como parte de la instalación del producto y forma parte del mismo dominio WebLogic, como se ha visto anteriormente. Hasta dicha versión, WebCenter Capture era una solución que debía instalarse por separado en un servidor Windows, similar a WebCenter Forms Recognition.

La herramienta recién descargada conocida como Repository Creation Utility (RCU) permite crear de una manera gráfica los esquemas en la base de datos necesarios para los productos de Oracle Fusion Middleware. Basta con ejecutarlo, indicar la base de datos en la que se crearán los esquemas (se deben tener permisos de administrador sobre la misma) y

seleccionar los productos de Oracle Fusion Middleware como se muestra en la siguiente imagen para crear los esquemas y sus respectivas tablas necesarias.

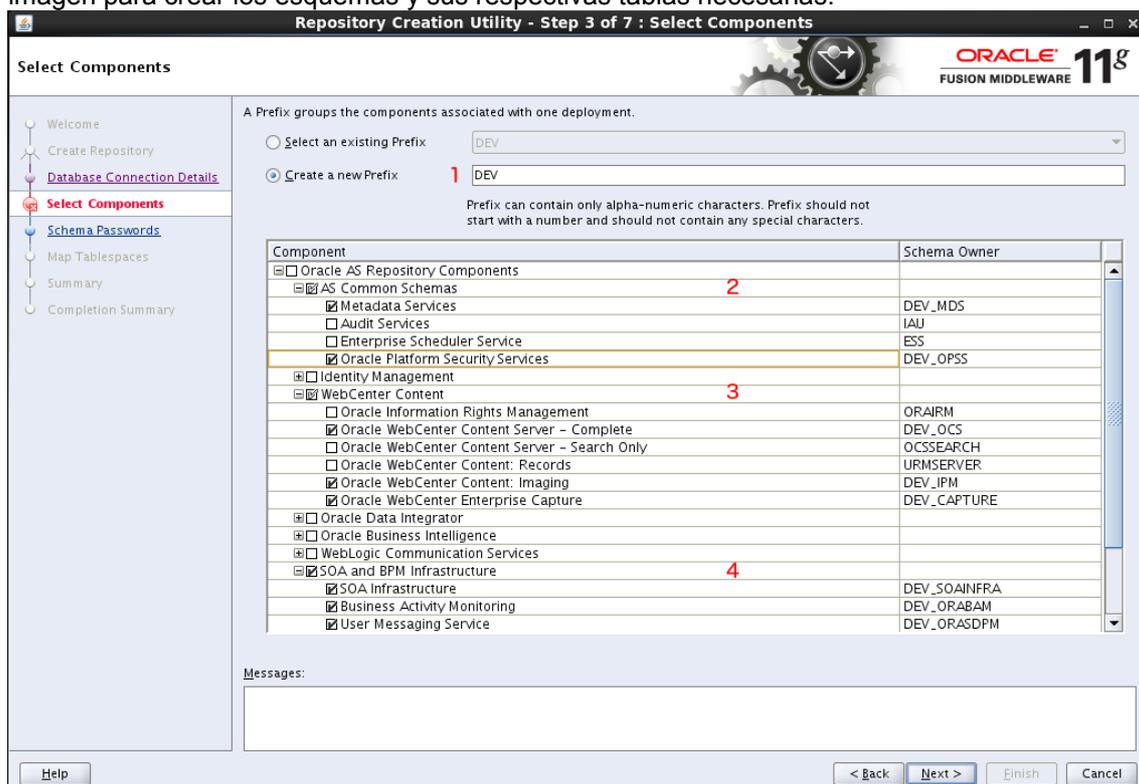


Ilustración 33 Repository Creation Utility

1. Prefijo que se añadirá a los esquemas. Sirve para distinguir los esquemas creados para distintos entornos (Desarrollo, Producción, Pruebas...).
2. Todos los esquemas comunes a todas las soluciones FMW.
3. Todos los esquemas disponibles para la instalación de WebCenter Content. Se añadirá el prefijo seleccionado anteriormente al nombre del esquema.
4. Todos los esquemas disponibles para la instalación de SOA y BPM Suite, Se añadirá el prefijo seleccionado anteriormente al nombre del esquema.

Como la base de datos es común a ambas instalaciones, únicamente se tiene que ejecutar este paso una vez. Los parámetros a introducir son:

- **Prefijo:** DEV, al tratarse de un proyecto en desarrollo (DEvelopment)
- **Esquemas**
 - **Comunes**
 - Metadata Services (**MDS**)
 - Oracle Platform Security Services (**DEV_OPSS**)
 - **WebCenter Content**
 - Content Server (**OCS**)
 - Imaging (**IPM**)
 - Enterprise Capture (**CAPTURE**)
 - **SOA and BPM**
 - SOA Infrastructure (**SOAINFRA**)
 - Business Activity Monitoring (**ORABAM**)
 - User Messaging Service (**ORASDPM**)

6.1.3. Instalar Oracle WebLogic Server

Previo a la instalación de Oracle WebLogic debe instalarse una máquina virtual de Java (JVM). Existen principalmente 2 opciones: HotSpot JDK y JRockit JDK. Mientras que HotSpot JDK es la JVM por defecto para aplicaciones cliente, JRockit es una JVM pensada para servidores que requieren ejecutar grandes aplicaciones y largas tareas que requieren mucha memoria y ejecutarse en un entorno escalable.

Por lo tanto, el primer paso es instalar JRockit en la ubicación \$JAVA_HOME. Gracias al instalador previamente descargado esta tarea se puede hacer de manera gráfica y sencilla.

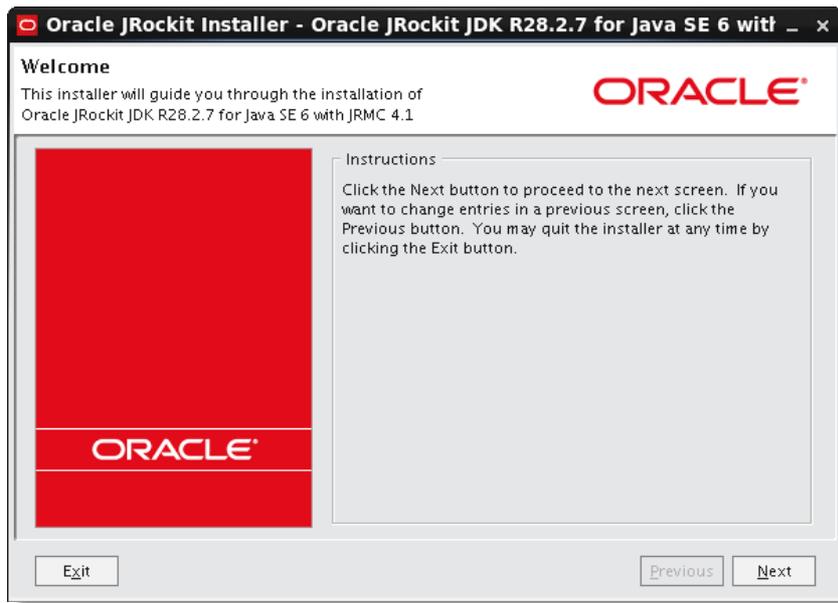


Ilustración 34 Instalación JRockit

Una vez finalizada la instalación de JRockit, se puede continuar con la instalación de Oracle WebLogic, indicando que se utilice el Java Development Kit (JDK) de JRockit anteriormente instalado en \$JAVA_HOME y finalmente instalando WebLogic en \$FMW_HOME. Una vez más, el instalador gráfico simplifica esta tarea.



Ilustración 35 Instalación WebLogic

6.1.4. Instalación de Oracle HTTP Server

Este componente es un servidor web que se utilizará principalmente para reenviar todas las peticiones recibidas por el puerto 80 a cualquier aplicación desplegada sobre el dominio WebLogic. Oracle HTTP Server reenvía la petición al servidor gestionado y puerto apropiado. Con el instalador gráfico únicamente es necesario indicar dónde se ubica \$FMW_HOME y el instalador creará los archivos necesarios para OHS en \$OHS_HOME.

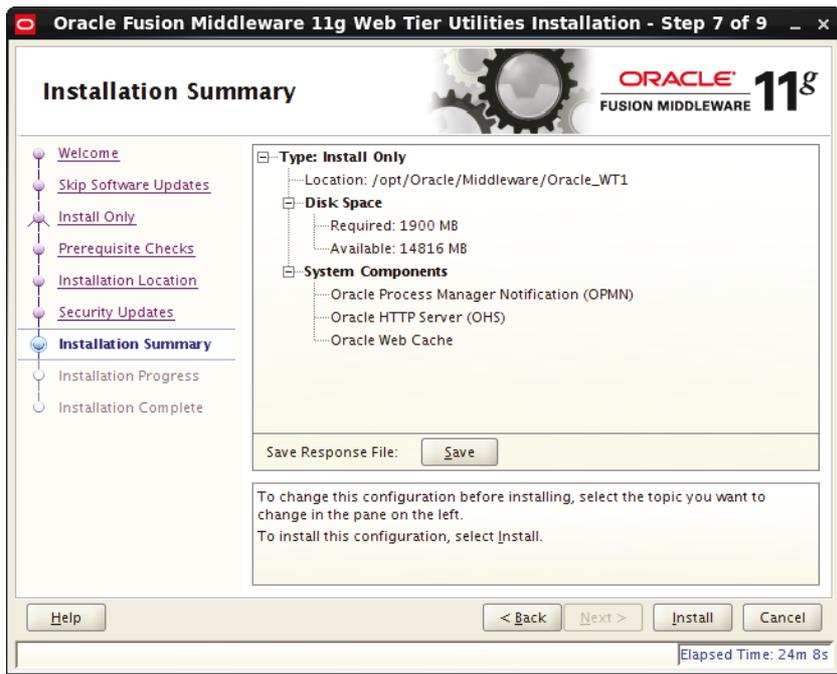


Ilustración 36 Instalación Oracle HTTP Server

Hasta este punto se ha instalado los elementos de infraestructura comunes, Oracle WebLogic Server y Oracle HTTP Server, en todos los servidores del clúster de WebCenter Content y el de BPM Suite. A partir de este punto y hasta las Tareas de Post-Configuración específicas, los pasos a dar son homólogos para WebCenter Content y BPM Suite.

6.1.5. Instalación de productos FMW

A continuación se detalla la instalación de cada producto de FMW.

6.1.5.1 Instalación de WebCenter Content

La instalación de WebCenter Content también se apoya en un instalador gráfico en el que se debe indicar la ruta \$FMW_HOME, donde anteriormente se ha instalado Oracle WebLogic, y el instalador creará los archivos necesarios en la ubicación \$ECM_HOME.

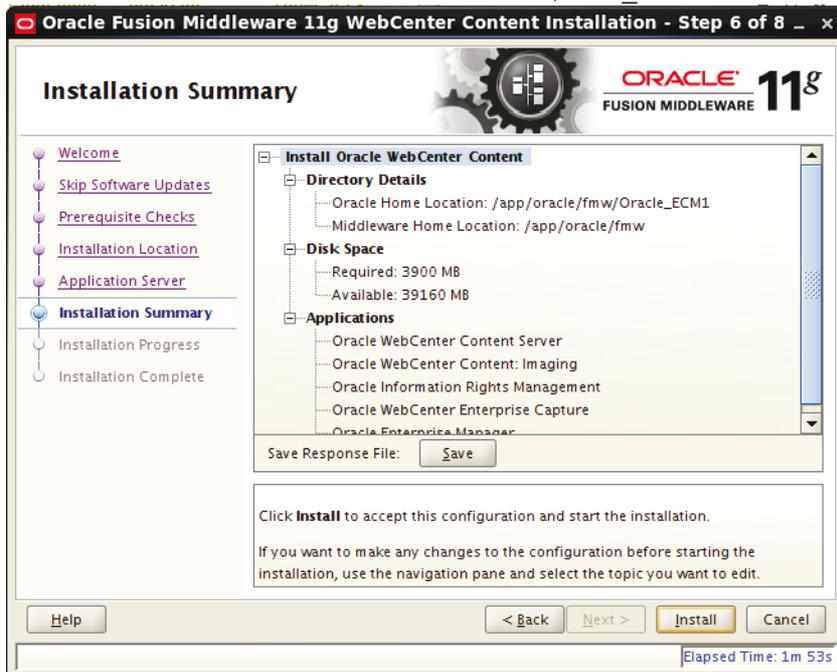


Ilustración 37 Instalación WebCenter Content

Este proceso en realidad ha dejado en nuestro sistema los archivos necesarios para crear un dominio WebLogic con las aplicaciones de WebCenter Content por lo que la instalación de WebCenter Content aún no ha concluido.

6.1.5.2. Instalación de SOA y BPM Suite

De igual manera que con WebCenter Content, SOA y BPM Suite se instala de manera gráfica también, indicando la ubicación \$FMW_HOME, crea los archivos necesarios para la creación del dominio de BPM en la ubicación \$SOA_HOME. Los archivos necesarios para BPM se encuentran dentro del instalador de SOA Suite.

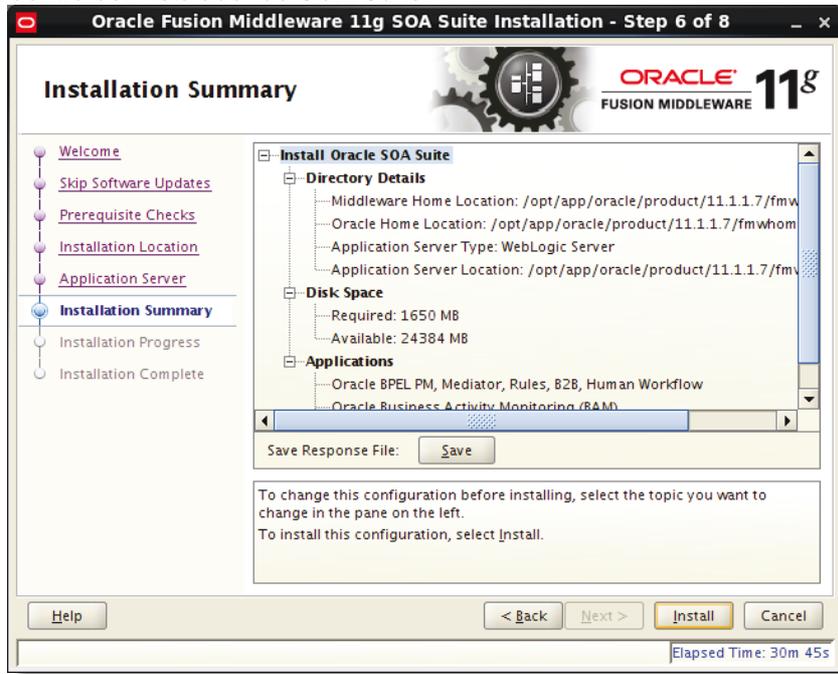


Ilustración 38 Instalación BPM y SOA

6.1.6. Creación de Dominios WebLogic

6.1.6.1 Creación del Dominio WebCenter Content

Una vez concluida la instalación de todos los archivos necesarios, se procede a la creación y configuración del dominio WebLogic. Todo se realiza mediante un configurador gráfico. En primer lugar, se seleccionan las aplicaciones de WebCenter Content a instalar.

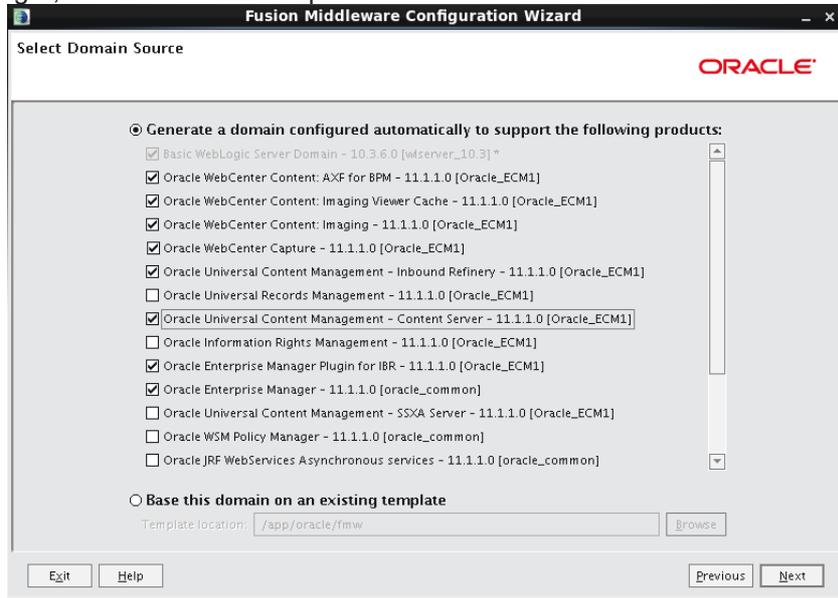


Ilustración 39 Creación del Dominio de WebCenter Content

Los componentes a instalar son:

- Oracle WebCenter Universal Content Management – Content Server
- Oracle WebCenter Universal Content Management – Inbound Refinery
- Oracle WebCenter Content: Imaging
- Oracle WebCenter Capture

El configurador selecciona automáticamente los componentes requeridos por los que se han seleccionado. Tras indicar el nombre del dominio, **wcc_domain**, e indicar donde será creado, **\$DOMAIN_HOME**, el configurador solicita los datos de conexión a los esquemas previamente creados con la herramienta RCU.

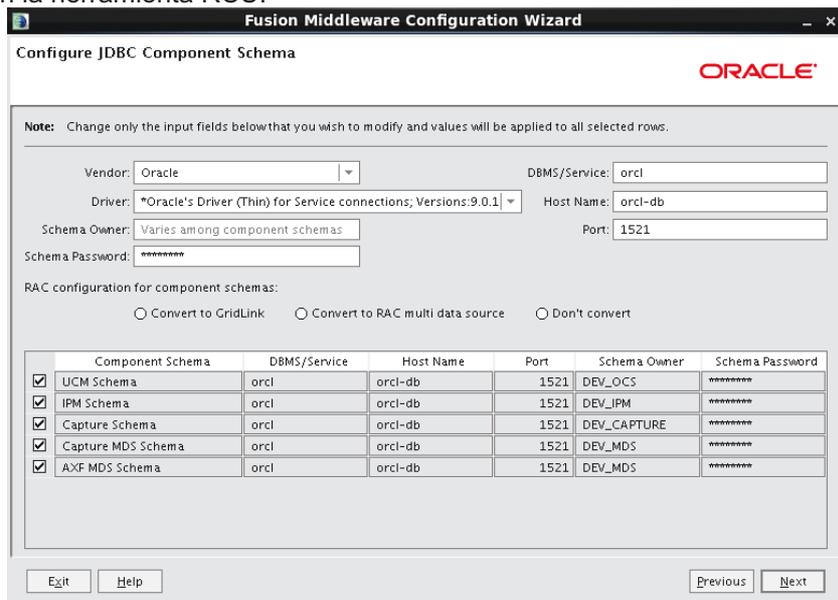


Ilustración 40 Conexión a los esquemas creados por RCU

Para cada esquema se solicita:

- Usuario del esquema, anteriormente creado con RCU, incluyendo el prefijo.
- Contraseña, establecida con RCU
- Servidor, puerto y nombre del servicio. Se utiliza la base de datos proporcionada por la organización en el servidor orcl-db, el nombre del servicio es orcl y el puerto es el puerto por defecto, 1521.

Cuando no se desea crear clústers, los servidores gestionados se crean por defecto. Como se van a crear clústers, es necesario crear todos los servidores gestionados que van a formar parte del dominio para repartirlos en clústers junto con los puertos por los que recibirán peticiones.

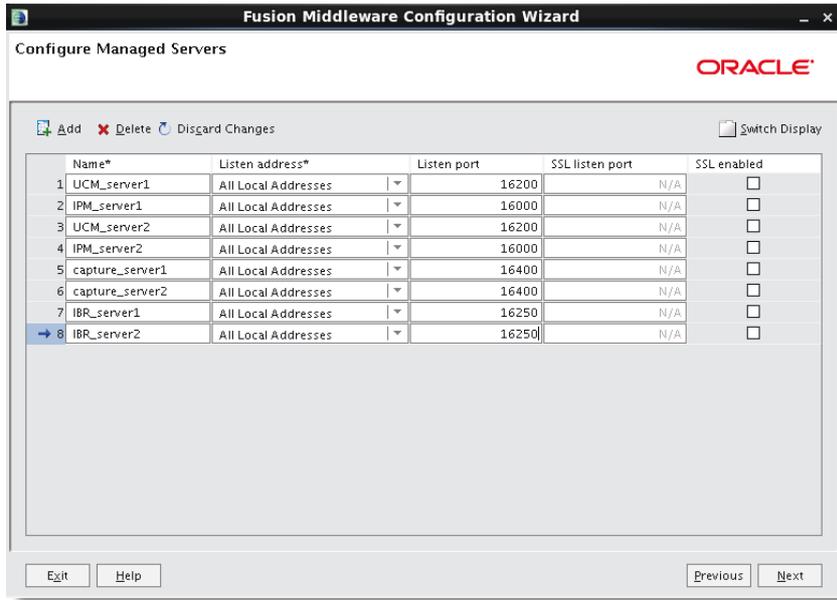


Ilustración 41 Configuración de Servidores Gestionados

A continuación, se crean los clústers, indicando qué servidores forman cada clúster:

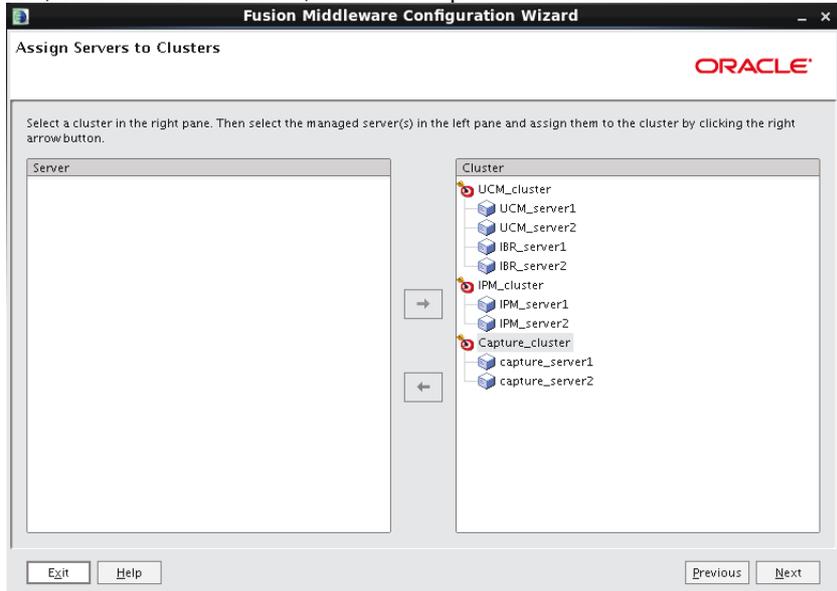


Ilustración 42 Configuración de Clústeres

Por último, se crean las máquinas sobre las que se ejecutan los nodos del clúster y se asocian los servidores gestionados con ellas, de manera muy similar a como se han añadido los servidores a los clústers. En realidad lo que se configura es la dirección y puerto en la que está escuchando el Node Manager en cada máquina. Este componente, como ya se ha descrito, es

un demonio o proceso en segundo plano que gestiona el arranque y parada de los servidores WebLogic presentes en una máquina.

Una vez concluida la configuración, se procede a la creación del dominio con los servidores gestionados, clústers y máquinas especificados, quedando el dominio configurado de la siguiente manera:

Dominio: wcc_domain	
Máquina: wcc-vm-l-1	Máquina: wcc-vm-l-2
Clúster: UCM_cluster	
UCM_server1	UCM_server2
IBR_server1	IBR_server2
Clúster: IPM_cluster	
IPM_server1	IPM_server2
Capture_cluster	
capture_server1	capture_server2

Tabla 23 Detalles configuración dominio wcc_domain

Quedan únicamente los pasos de post-configuración para que WebCenter Imaging y Capture estén listos para ser adaptados para cubrir las necesidades del Sistema Gestor de Documentos Electrónicos para la Automatización de la Gestión de Reservas.

6.1.6.2 Creación del Dominio BPM Suite

El proceso de crear el dominio de BPM Suite es homólogo a la creación del dominio de WebCenter Content, únicamente especificando en las respectivas pantallas los detalles analizados en la sección de la arquitectura, quedando el dominio configurado de la siguiente manera:

Dominio: bpm_domain	
Máquina: bpm-vm-xl-1	Máquina: bpm-vm-xl-2
Clúster: SOA_cluster	
SOA_server1	SOA_server2
Clúster: BAM_cluster	
BAM_server1	BAM_server2

Tabla 24 Detalles configuración dominio bpm_domain

6.1.7. Tareas de post-configuración comunes

Las tareas de post-configuración comunes se resumen en 3:

1. Configuración de Oracle HTTP Server
2. Configuración de Node Manager
3. Configuración de Microsoft Active Directory como proveedor de identidad

6.1.7.1 Configuración de Oracle HTTP Server

Con el fin de que OHS sea capaz de enrutar peticiones vía el puerto 80 hacia las aplicaciones en clúster, es necesario editar el fichero ubicado en `$OHS_HOME/instances/ohs_instance1/config_OHS/ohs1/mod_wl_ohs.conf` y añadir una entrada como la siguiente

```
<Location /<<Aplicación>>>
WebLogicCluster <<Servidor X>>:<<Puerto>>,<<Servidor Y>>:<<Puerto>>
SetHandler weblogic-handler
WLCookieName JSESSIONID
</Location>
```

por cada aplicación desplegada sobre el dominio. En el caso del dominio **wcc-domain** se tendrían que añadir las siguientes entradas:

```
# UCM
<Location /cs>
WebLogicCluster wcc-vm-l-1:16200,wcc-vm-l-2:16200
SetHandler weblogic-handler
WLCookieName JSESSIONID
</Location>
<Location /adfAuthentication>
WebLogicCluster wcc-vm-l-1:16200, wcc-vm-l-2:16200
```

```

    SetHandler weblogic-handler
    WLCookieName JSESSIONID
</Location>
# IBR
<Location /ibr>
WebLogicCluster wcc-vm-1-1:16250, wcc-vm-1-2:16250
SetHandler weblogic-handler
WLCookieName JSESSIONID
</Location>
<Location /ibr/adfAuthentication>
WebLogicCluster wcc-vm-1-1:16250, wcc-vm-1-2:16250
SetHandler weblogic-handler
WLCookieName JSESSIONID
</Location>
# Imaging Application
<Location /imaging>
    SetHandler weblogic-handler
    WebLogicCluster wcc-vm-1-1:16000, wcc-vm-1-2:16000
</Location>
# AXF WS Invocation
<Location /axf-ws>
    SetHandler weblogic-handler
    WebLogicCluster wcc-vm-1-1:16000, wcc-vm-1-2:16000
</Location>
# Enterprise Capture Client
<Location /dc-client>
    SetHandler weblogic-handler
    WebLogicCluster wcc-vm-1-2:16400, wcc-vm-1-2:16400
</Location>
# Enterprise Capture Workspace Console
<Location /dc-console>
    SetHandler weblogic-handler
    WebLogicCluster wcc-vm-1-2:16400, wcc-vm-1-2:16400
</Location>

```

6.1.7.2 Configuración de Node Manager

Para configurar Node Manager, se debe modificar manualmente 2 ficheros de configuración. El primero está ubicado en `$WLS_HOME/common/nodemanager`. El fichero es `nodemanager.properties` y se deben modificar las siguientes entradas:

```

ListenAddress=<<IP de la máquina>>
SecureListener=false // Para indicar que no utilice el puerto SSL si no el normal
StopScriptEnabled=true // Para permitir la parada de Node Manager mediante un Script
StartScriptEnabled=true // Para permitir el arranque de Node Manager

```

El segundo se encuentra en `$DOMAIN_HOME/domains/<<dominio WebLogic>>/config/nodemanager` donde el dominio WebLogic puede ser **wcc-domain** o **bpm-domain** y se llama `nm_password.properties`. En él se deben indicar el usuario administrador de WebLogic y su contraseña. El fichero será encriptado la primera vez que arranque Node Manager.

A continuación, se debe crear el siguiente script de arranque en `/etc/init.d/nodemgr`

```

#!/bin/sh
#
# nodemgr Oracle Weblogic NodeManager service
#
# chkconfig: 345 85 15
# description: Oracle Weblogic NodeManager service
# The script needs to be saved as /etc/init.d/nodemgr and then issue chkconfig --add
# nodemgr as root

### BEGIN INIT INFO
# Provides: nodemgr
# Required-Start: $network $local_fs
# Required-Stop:
# Should-Start:
# Should-Stop:
# Default-Start: 3 4 5
# Default-Stop: 0 1 2 6
# Short-Description: Oracle Weblogic NodeManager service.
# Description: Starts and stops Oracle Weblogic NodeManager.
### END INIT INFO

. /etc/rc.d/init.d/functions

```

```

export MW_HOME="/app/oracle/fmwhome"
export JAVA_HOME="/ app/oracle/javahome"
DAEMON USER="oracle"
PROCESS_STRING="^.* /app/oracle/fmwhome/. *weblogic.NodeManager.*"
source $MW_HOME/wlserver 10.3/server/bin/setWLSEnv.sh > /dev/null
export NodeManagerHome="$WL_HOME/common/nodemanager"
NodeManagerLockFile="$NodeManagerHome/nodemanager.log.lck"
PROGRAM="$MW_HOME/wlserver_10.3/server/bin/startNodeManager.sh >
$NodeManagerHome/nodemanager.out 2>&1"
SERVICE_NAME=`/bin/basename $0`
LOCKFILE="/var/lock/subsys/$SERVICE_NAME"
RETVAL=0

start() {
    OLDPID=`/usr/bin/pgrep -f $PROCESS_STRING`
    if [ ! -z "$OLDPID" ]; then
        echo "$SERVICE_NAME is already running (pid $OLDPID) !"
        exit
    fi
    echo -n $"Starting $SERVICE_NAME: "
    /bin/su $DAEMON USER -c "$PROGRAM &"
    RETVAL=$?
    [ $RETVAL -eq 0 ] && touch $LOCKFILE
    sleep 1
    OLDPID=`/usr/bin/pgrep -f $PROCESS_STRING`
    if [ "$OLDPID" != "" ]; then
        /bin/echo $"OK (pid: $OLDPID)"
    else
        /bin/echo $"ERROR"
    fi
}

stop() {
    echo -n $"Stopping $SERVICE_NAME: "
    OLDPID=`/usr/bin/pgrep -f $PROCESS_STRING`
    if [ "$OLDPID" != "" ]; then
        /bin/kill -TERM $OLDPID
        /bin/echo $"OK"
    else
        /bin/echo "$SERVICE_NAME is stopped"
    fi
    /bin/rm -f $NodeManagerLockFile
    [ $RETVAL -eq 0 ] && rm -f $LOCKFILE
}

restart() {
    stop
    sleep 10
    start
}

case "$1" in
    start)
        start
        ;;
    stop)
        stop
        ;;
    restart|force-reload|reload)
        restart
        ;;
    condrestart|try-restart)
        [ -f $LOCKFILE ] && restart
        ;;
    status)
        OLDPID=`/usr/bin/pgrep -f $PROCESS_STRING`
        if [ "$OLDPID" != "" ]; then
            /bin/echo "$SERVICE_NAME is running (pid: $OLDPID)"
        else
            /bin/echo "$SERVICE_NAME is stopped"
        fi
        RETVAL=$?
        ;;
    *)

```

```
        echo "Usage: $0 {start|stop|status|restart|reload|force-
reload|condrestart}"

        exit 1

esac

exit $RETVAL
```

Se le deben dar los permisos de ejecución adecuados de manera que se ejecute junto al inicio del sistema operativo.

Finalmente, desde la consola de administración de WebLogic se debe configurar que Node Manager no escuche peticiones por el puerto seguro (SSL).

Con esta configuración de Node Manager, ahora los servidores WebLogic podrán arrancar automáticamente tras una caída no planeada de los sistemas.

6.1.7.3 Configuración de Microsoft Active Directory como proveedor de identidad

Como se ha indicado en el diagrama de arquitectura, las organizaciones cuentan típicamente con un directorio corporativo. Éste suele ser Microsoft Active Directory debido a que es una pieza inevitable si se tienen entornos Windows. La proliferación de múltiples directorios no es algo deseable por lo que, con el fin de evitar que el proyecto del SGDE para la Automatización de la Gestión de Reservas utilice el proveedor de identidad por defecto de WebLogic, creando en definitiva un directorio distinto del corporativo, se deben configurar los dominios WebLogic para que utilicen Microsoft Active Directory como proveedor de identidad, utilizando para ello la consola de administración Web de WebLogic de cada uno de los dominios.

6.1.8. Tareas de post-configuración específicas

Estas tareas difieren del producto específico. En el caso de WebCenter, estas tareas son:

- Configuración inicial del Servidor de Contenidos UCM (Universal Content Manager) mediante un formulario web que se presenta al administrador la primera vez que arranca.
- Comunicar el Servidor de Contenidos con el Servidor de Conversiones IBR (Inbound Refinery)
- Habilitar los siguientes componentes
 - Conversiones a PDF
 - Framework de integración con aplicaciones Application Extension Framework que incluye las capacidades de integración de Enterprise Application Adapter Framework y Managed Attachments

Tras estos pasos, queda la plataforma Oracle Fusion Middleware formada por WebCenter Content y BPM Suite lista para ser configurados para cubrir los requisitos del SGDE para la Automatización de la Gestión de Reservas.

6.2. Instalación de WebCenter Forms Recognition

Como se ha indicado en varias ocasiones, WebCenter Forms Recognition es una aplicación de tipo cliente pesado Win32 que se instala sobre Windows por lo que para instalarlo únicamente es necesario ejecutar el instalador asociado y esto instalará automáticamente el Designer, el Verifier y el Runtime Service.

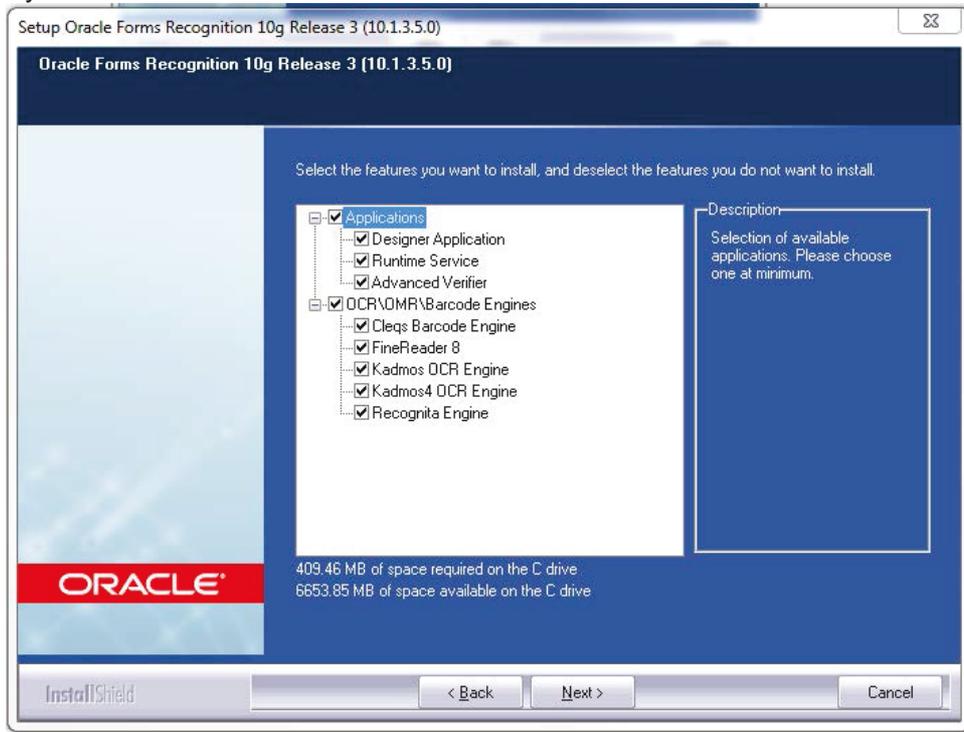


Ilustración 43 Instalación Forms Recognition

Cómo se puede ver en la imagen, además de las 3 aplicaciones principales, también se instalan distintos motores OCR para ser utilizados con WebCenter Forms Recognition. Se explicarán con más detalle en el apartado de configuración del proyecto de reconocimiento.

Una vez se ha terminado de instalar WebCenter Forms Recognition es necesario realizar 2 tareas de post-configuración similares a las realizadas en el entorno Fusion Middleware para utilizar usuarios existentes en Active Directory para trabajar con WebCenter Forms Recognition así como configurar el Runtime Server como servicio de Windows para que éste arranque automáticamente tras la caída de un servidor, similar a Node Manager.

Al tratarse de una aplicación sobre Windows ya está integrada con el Directorio Activo, de modo que únicamente es necesario abrir la herramienta Designer y desde ella localizar en el Directorio Activo a los usuarios pertinentes. La primera vez que se abre Designer, utiliza una cuenta genérica denominada Administrator y que carece de contraseña.

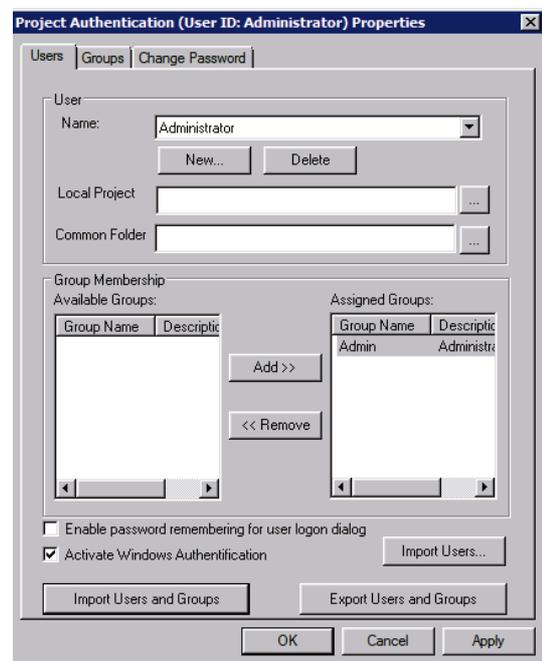


Ilustración 44 Configuración seguridad en Forms Recognition

De igual manera, establecer Runtime Service como servicio que arranque automáticamente al inicio de Windows también es una tarea de administración convencional de Windows. En el menú de Servicios de Windows debe modificarse el servicio de Forms Recognition para que arranque de forma automática y además indicar un usuario administrador con el que se arrancará dicho servicio para poder acceder a carpetas protegidas por permisos.

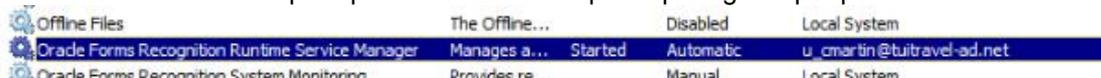


Ilustración 45 Activación de Forms Recognition Service como Servicio de Windows

6.3 Configuración de Enterprise Capture

A continuación se detallarán todas las parametrizaciones realizadas a Enterprise Capture con el fin de cubrir los requisitos de captura identificados.

Lo primero que debe crearse es un espacio de trabajo o Workspace específico para este proyecto.

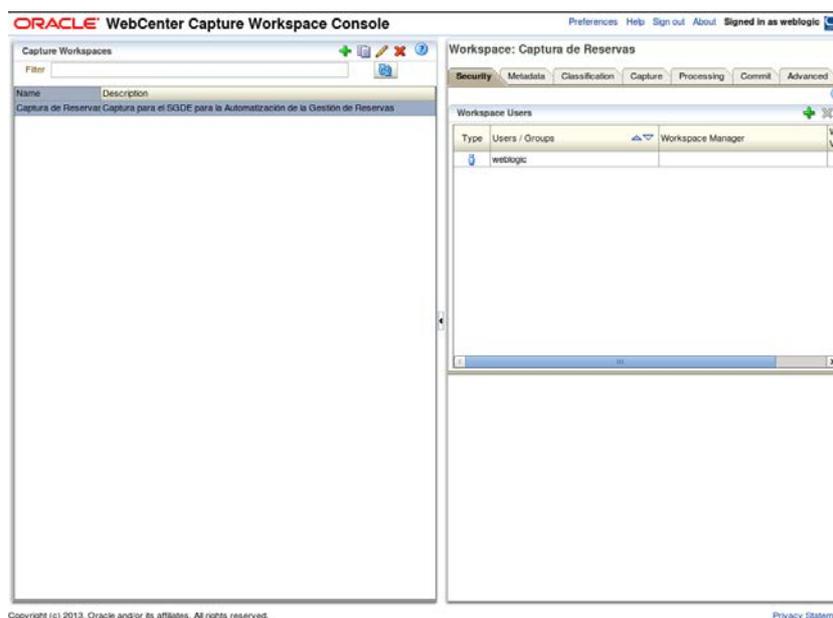


Ilustración 46 Workspace de Enterprise Capture

En el espacio de trabajo se deben crear y configurar:

- Metadatos a capturar:
- Perfil de Documento
- Procesamiento automático
 - Trabajo de Importación
 - Trabajo de Conversión
 - Trabajo Confirmación

6.3.1. Metadatos a capturar

En la pestaña de metadatos, se deben crear 3 metadatos con las siguientes características:

- **Tipo:** Alfanumérico ya que pueden contener cualquier carácter
- **Longitud Máxima:** 255, con el fin de controlar desde la captura la longitud máxima del nombre del fichero para evitar problemas ocasionados por este tanto en el servidor Windows como en WebCenter Imaging. Si fuera necesario, es más sencillo desarrollar un pequeño Script que trunque la longitud del metadato que corregir errores más adelante en el proceso.
- **Ni requerido, ni bloqueado:** De esta manera se evitan problemas al procesar correos que tengan el asunto en blanco o todos sus recipientes ocultos, por ejemplo.
- **Sin formato de muestra ni máscara de entrada:** se tratan de campos con una amplia libertad de formato por lo que no tiene sentido restringirlos mediante máscaras o formatos
- **Sin auto-rellenar:** los metadatos serán capturados de campos del correo electrónico y ni se utilizarán datos del sistema como la fecha de creación del lote o su identificador.

Los 3 metadatos que se deben crear

- **Origen:** Cuenta de correo electrónico desde el que se envía la reserva, generalmente perteneciente a un hotel
- **Destino:** Una de las 6 cuentas de correo electrónico de la organización cuyas bandejas de entrada se monitorizan.
- **Asunto:** Asunto del correo, suele delatar si se trata de una reserva en cuyo caso suele indicar el nº de reservas que van incluidas en el adjunto

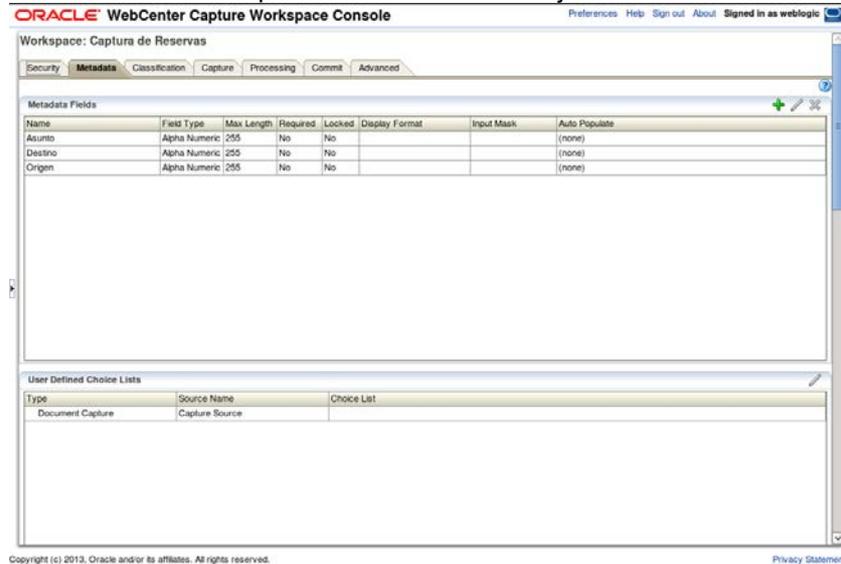


Ilustración 47 Campos de Enterprise Capture

Estos metadatos serán capturados de elementos del correo electrónico entrante como su origen o el asunto y se propagarán añadiéndolos al nombre de la imagen que se genere.

6.3.2. Perfil de Documento

Los perfiles de documentos sirven para agrupar metadatos. Se va a crear un perfil de documento para capturar todas las reservas.

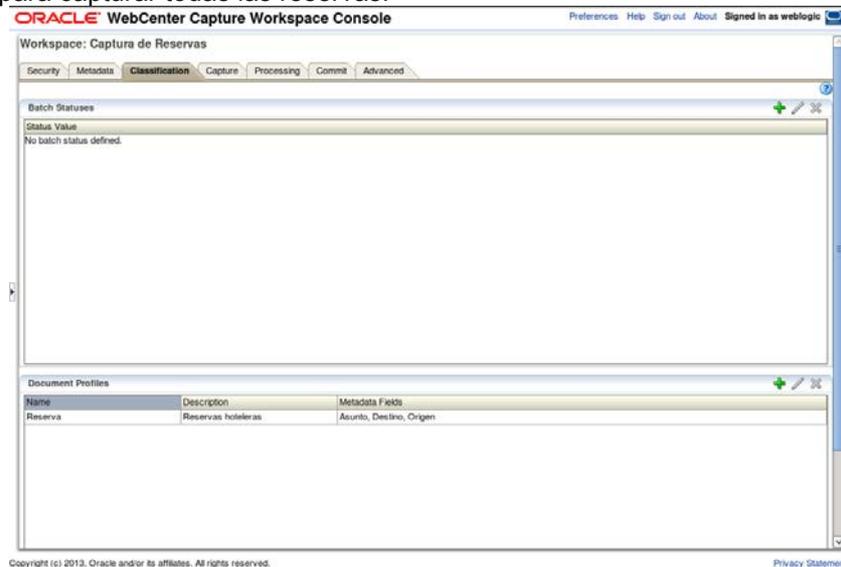


Ilustración 48 Perfil de documento

6.3.3 Procesamiento automático

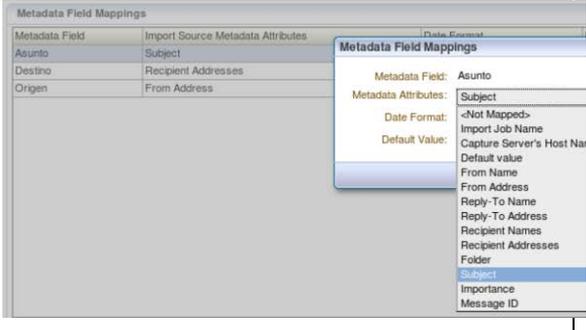
Se deben definir 3 trabajos de procesamiento automático

1. Trabajo de importación: para capturar los correos entrantes de las bandejas de entrada
2. Trabajo de Conversión: para convertir aquellos adjuntos que se reciban en formato PDF a TIFF para su posterior procesado

- Trabajo de confirmación para dejar los documentos capturados en una ubicación compartida con WebCenter Forms Recognition.

6.3.3.1 Trabajo de Importación de Email

Se debe crear el trabajo de importación con los siguientes parámetros relevantes

Campo	Descripción	Valor
1.- General		
Prefijo del lote	Prefijo para identificar el origen del lote de entre los demás lotes.	EMAIL
Fuente	Origen de los documentos a Capturar: Carpeta del sistema, Carpeta con fichero de listado o Email	Email
Frecuencia de Importación	La frecuencia con la que el trabajo va a buscar correos nuevos en la bandeja de entrada: Cada 30 Segundos, 1 Minuto, 5 Minutos, 15 Minutos, 30 Minutos, 1 Hora o 1 Día	1 Minuto
2.- Perfil de Documento y Metadatos		
Perfil de Documento	El perfil de documento que se asignará al documento capturado	Reserva, previamente definido
Metadatos	Se deben mapear los metadatos del perfil de documento con alguno de los valores genéricos disponibles en la importación (Nombre del trabajo de importación, Nombre del servidor donde se ejecuta Enterprise Capture), valores únicamente disponibles en la importación de correos electrónicos (Nombre del Remitente, Dirección del Remitente, Nombre del Reply-to, Dirección de Reply-To, Nombres de los Destinatarios, Direcciones de los Destinatarios, Carpeta, Asunto, ID de Mensaje e Importancia) o el valor por defecto que se desee.	Asunto → Asunto Origen → Dirección del Remitente Destino → Direcciones de los Destinatarios
		
3.- Configuración de la fuente de importación		
3.1.- Cuentas de Correo		
Servidor de correo	Servidor de correo IMAP donde se encuentran las cuentas a monitorizar.	mail-serv (ver diagrama de arquitectura)
Puerto	Puerto por el que establecer la comunicación con el servidor de correo. Por defecto 143 (Sin Secure Socket Layer) y 993 (Con SSL)	993
Direcciones de correo	Todas las cuentas de correo electrónico que se desee monitorizar.	reservascanada@banco camas.com, reservaseste@banco decamas.com, reservasoeeste@banco decamas.com, reservasflorida@banco

		ecamas.com, reservaspuertorico@ban codecamas.com, reservascentro@bancode camas.com
3.2.- Filtros de mensajes		
Carpetas a procesar	Carpetas de la cuenta de correo que se deben procesar, pueden especificarse varias, separándolas por comas. Por defecto se procesa únicamente la bandeja de entrada o inbox .	Inbox
Filtros	Búsqueda de palabras o cadenas de texto en cualquier combinación posible de los siguientes campos: <ul style="list-style-type: none"> • Dirección de remitente • Asunto del mensaje • Cuerpo del mensaje 	Filtro por dirección del remitente: que contenga el término fax2mail ⁶
Operador de búsqueda	<ul style="list-style-type: none"> • And: se deben cumplir <u>todos</u> los filtros anteriores. • Or: se debe cumplir al menos <u>uno</u> de los filtros anteriores 	And
3.3.- Procesado		
Opciones de importación	Permite configurar si se desea importar el cuerpo del mensaje además del adjunto.	No se va a importar el cuerpo del adjunto en este proyecto.
Máscara para incluir adjuntos	Patrón para filtrar adjuntos. Puede usarse *.* para siempre importar los adjuntos o pueden combinarse varias máscaras para importar distintos formatos.	*.pdf, *.PDF, *.tif, *.TIF, *.tiff, *.TIFF
Orden del documento	Establece el orden en el que se incorporarán los adjuntos y el cuerpo de mensaje	1. Adjuntos Únicamente se van a procesar adjuntos.
3.4.- Post-Procesado		
Importación satisfactoria	Permite indicar qué hacer en caso de importación satisfactoria del email. <ul style="list-style-type: none"> • Borrar el mensaje • Mover el mensaje a una carpeta determinada 	Mover el mensaje a la carpeta Importados . Debe crearse dicha carpeta en la cuenta de correo.
Importación fallida	Permite indicar qué hacer en caso de importación fallida del email <ul style="list-style-type: none"> • Borrar el mensaje • Mover el mensaje a una carpeta determinada • No hacer nada. 	Mover el mensaje a la carpeta Errores . Debe crearse dicha carpeta en la cuenta de correo. Esto es preferible en lugar de no hacer nada para evitar que Enterprise Capture reintente importar un email que ha fallado una vez y probablemente vuelva a fallar debido a que lo más probable es que el adjunto contenga algún tipo de error. Moviendo los errores a una carpeta diferente permite establecer un proceso de recuperación de errores de importación al margen del procesado normal.

⁶ Este proyecto únicamente pretende automatizar la confirmación de las reservas que son enviadas como adjuntos de correos fax2mail, dejando el procesado de cuerpos de correos electrónicos para una posible segunda fase del mismo. Las confirmaciones que llegan como adjuntos de correos fax2mail suponen más del 80% de las confirmaciones.

Trabajo de Post-Procesado	Permite indicar qué trabajo de Conversión, de Reconocimiento o de Confirmación ejecutar una vez concluida la importación.	Trabajo de Conversión a TIFF.
---------------------------	---	-------------------------------

Tabla 25 Parámetros del Trabajo de Importación de Enterprise Capture

6.3.3.2 Trabajo de Conversión a TIFF

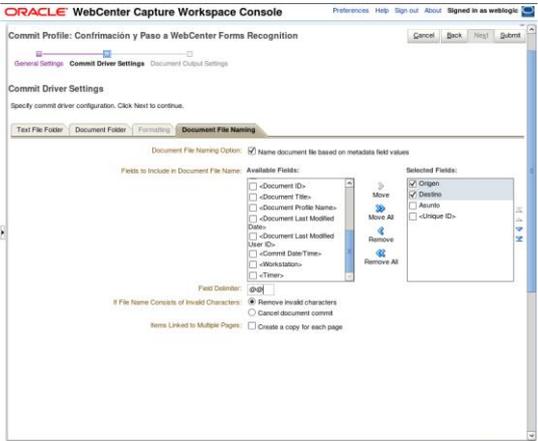
Se debe crear el trabajo de conversión con los siguientes parámetros relevantes

Campo	Descripción	Valor
1.- Selección de Documentos		
Documentos a Convertir	Indica qué documentos deben procesarse <ul style="list-style-type: none"> • Todos los documentos que no sean imágenes • Documentos que no sean imágenes que cumplan determinado filtro de nombre 	Todos
2.- Formato de Salida		
Formato de Imagen	Indica el formato al que se convertirán los documentos: <ul style="list-style-type: none"> • TIFF en blanco y negro • JPEG en color 	TIFF en blanco y negro para asegurar la máxima calidad posible del OCR al no haber mayor contraste posible que negro sobre blanco. Los colores únicamente pueden generar confusión a los motores OCR.
Puntos por pulgada	Define la resolución con la que se debe crear la imagen: 100, 150, 200, 240, 300, 400, 600	300 . Es la resolución mínima que WebCenter Forms Recognition para garantizar un reconocimiento automatizado de, al menos, el 80% de los documentos.
3.- Post-Procesado		
Conversión satisfactoria	Permite indicar qué trabajo de Conversión, de Reconocimiento o de Confirmación ejecutar una vez concluida la conversión. Permite además enviar una notificación a una cuenta de correo electrónico determinada	Trabajo de Confirmación y paso a Forms Recognition. No es necesario enviar notificación por correo electrónico.
Importación fallida	Permite indicar qué trabajo de Conversión, de Reconocimiento o de Confirmación ejecutar en caso de conversión fallida	No invocar ningún trabajo de procesado pero enviar una notificación al Administrador/Desarrollador de WebCenter para que intervenga manualmente.

Tabla 26 Parámetros del Trabajo de Conversión de Enterprise Capture

6.3.3.3 Trabajo de Confirmación y paso a WebCenter Forms Recognition

Se debe crear el trabajo de confirmación con los siguientes parámetros relevantes

Campo	Descripción	Valor
1.- General		
Controlador de Confirmación (Commit Driver)	<p>Elección de qué controlador se utilizará para ejecutar la confirmación:</p> <ul style="list-style-type: none"> • Sistema de Ficheros: Deposita los documentos capturados en una carpeta del sistema • WebCenter Content: Sube los documentos directamente al Servidor de Contenidos • WebCenter Content: Imaging: Sube los documentos al Servidor de Contenidos de manera que pueda ser visualizado y gestionado por la aplicación de contenido Transaccional WebCenter Imaging. 	Sistema de Ficheros. Se deben dejar los documentos en una carpeta compartida entre el clúster de WebCenter Content, dónde se está ejecutando Enterprise Capture y el clúster de WebCenter Forms Recognition de manera que éste pueda recogerlos y aplicarles reconocimiento.
Formato de Salida	<p>Formato en el que se dejarán los documentos en la ubicación definida por el Controlador de Confirmación:</p> <ul style="list-style-type: none"> • Imagen TIFF multi-página • PDF Imagen: PDF sin texto seleccionable • PDF Texto: PDF con texto seleccionable 	Imagen TIFF. Para que WebCenter Forms Recognition pueda aplicarles reconocimiento.
2.- Configuración del Controlador de Confirmación		
Carpeta	Carpeta donde Enterprise Capture dejará los documentos para que Forms Recognition los procese.	/u01/proyectos/SGDEAGR/cap2fr.
Nombre de los ficheros	<p>Permite definir la estructura del nombre de las imágenes generados por Enterprise Capture durante la confirmación. Para ello se deben seleccionar que campos de captura o metadatos propios del sistema, como fecha de captura o identificador interno del documento, utilizar para nombrar el fichero y definir un delimitador de los mismos.</p> 	<ul style="list-style-type: none"> • Origen (Campo capturado) • Destino (Campo capturado) • Asunto (Campo capturado) • <Identificador Único> (Metadato del sistema) <p>Como delimitador, se utilizará la cadena @@ para poder extraer los campos de Origen, Destino y Asunto directamente del nombre del fichero en Forms Recognition sin tener que realizar OCR, lo que supone un ahorro en procesamiento y reducción de la probabilidad de error. La cadena elegida nunca figurará en una dirección de correo ya que está prohibida la aparición de más de un símbolo de @ en las direcciones y no es una cadena que habitualmente se escriba en el asunto de un email. Se podrían incluir más campos o hacer un separador</p>

		<p>más extenso pero se debe tener en cuenta las restricciones de longitud de nombres de ficheros de los Sistemas Operativos.</p> <p>Con dicha configuración, Enterprise Capture generará una imagen TIFF a partir de un email enviado desde direccion1@correo.com a direccion2@correo.com con el asunto Asunto con el siguiente nombre: direccion1@correo.com@@direccion2@correo.com@@asunto@@000000001.TIFF</p>
--	--	--

Tabla 27 Parámetros del Trabajo de Confirmación de Enterprise Capture

Con esto queda configurado Enterprise Capture, el componente de captura del SGDE para la Automatización de la Gestión de Reservas.

6.4 Configuración de Forms Recognition

La pieza de reconocimiento inteligente de documentos, Forms Recognition, es la más crítica para permitir la automatización de la gestión de reservas. En esta sección se describirán todos los parámetros más relevantes para cubrir los requisitos de la solución.

En primer lugar, debe crearse el proyecto de reconocimiento con la herramienta Designer, esto generará un fichero .sdp correspondiente al proyecto de reconocimiento. El fichero es **IDR_SGDE_AGR.sdp** de Intelligent Document Recognition para el **Sistema Gestor de Documentos Electrónicos** para la **Automatización de la Gestión de Reservas**. Para dicho proyecto debe configurarse lo siguiente:

1. Clasificación
2. Extracción
3. Servicio de Ejecución en Tiempo Real (Runtime Service)

6.4.3. Clasificación

La clasificación en Forms Recognition requiere que previamente se configure el OCR.

6.4.3.1. Pre-configuración de OCR

Previo a crear las clases y configurar la clasificación, es necesario establecer cuál será el motor de OCR que el proyecto utilizará por defecto para reconocer los textos y con ellos clasificar los documentos y, a continuación, extraer los campos.



Ilustración 49 Configuración OCR en Forms Recognition

En la siguiente tabla se describen brevemente los motores OCR disponibles en WebCenter Forms Recognition, junto con su relevancia para el proyecto.

Motor	Tipo	Descripción	Relevancia
FineReader	OCR	Ideado para el reconocimiento de caracteres electrónicos, especialmente de documentos con mala calidad de impresión al poder ejecutar ciertas técnicas de pre-procesado. Disponible en multitud de idiomas, es decir, capaz de reconocer caracteres especiales como Ñ o Ç.	Alta. Los documentos a reconocer adjuntos generados electrónicamente.
Kadmos	OCR	Reconocimiento de caracteres manuscritos.	Media. De manera

			excepcional, en ocasiones algunos campos como la referencia de la reserva están escritos a mano en el adjunto del correo electrónico. El número de casos no es suficiente como para complicar el proyecto aplicando reconocimiento manuscrito
Recognita	OCR	Capaz de reconocer caracteres electrónicos y manuscritos. Soporta menos idiomas y tiene menos potencia de pre-procesado que FineReader lo que lo hace menos flexible y tolerante a fallos o ruido en los documentos.	Nula. Mientras que los correos electrónicos tienen alta calidad al no haber sido impresos, la mayoría de documentos adjuntos han sido impresos y escaneados por lo que pierden cierta definición. FineReader es simplemente mejor opción que Recognita.
Cairo	OMR	Motor muy básico para reconocer marcas de una manera rápida y eficiente, utilizado generalmente para reconocer respuestas en formularios.	Nula. No existen campos rellenables con marcas.
Recognita Barcode	Códigos de barras	La versión de Recognita utilizada para el reconocimiento de códigos de barras unidimensionales	Nula. Las reservas no utilizan códigos de barras.

Tabla 28 Motores OCR Disponibles en Forms Recognition

Se utilizará FineReader como motor para reconocer caracteres y, en base a dicho reconocimiento, realizar la clasificación automática de documentos y extracción de campos.

6.4.3.2. Estructura y motores de clasificación.

Como se ha indicado anteriormente, se desea distinguir entre aquellos adjuntos que contienen información de reservas de aquellos que no. De las reservas, se desea saber si se trata de nuevas reservas (New), correcciones (Amendment) o cancelaciones (Cancellation) de reservas ya existentes. De este modo queda la siguiente jerarquía de Clasificación en Forms Recognition:



Ilustración 50 Jerarquía de Clasificación

- **Booking:** Súper clase de Reservas. Agrupar los 3 tipos de reservas bajo esta súper clase evita tener que definir los campos de extracción comunes de manera repetida.
 - **Amendment:** Corrección a una reserva existente
 - **Cancelation:** Cancelación de una reserva existente

- **New:** Nueva reserva
- **Void:** todo lo que no sea clasificado como una reserva. Idealmente quedarán aquí clasificados todos aquellos adjuntos que no hagan referencia a una reserva como publicidad o promociones.

Una vez establecida la estructura de clasificación, debe elegirse que motor o motores de clasificación utilizar para identificar cada una de las clases. Forms Recognition ofrece 5 motores de clasificación. A continuación se describen los 5 motores junto con su relevancia para el proyecto:

Clasificación	Descripción	¿Requiere aprendizaje?	Relevancia para el proyecto de SGDE
Por Contenido con Brainware	Basado en tecnología de redes neuronales Brainware, clasifica los documentos en función del contenido sin tener en cuenta la disposición del mismo. Requiere formar un conjunto de aprendizaje suficientemente grande y con suficiente diferencia entre los ejemplos para cada clase para funcionar correctamente.	Si	Alta ya que en todos los correos referentes a reservas se utilizan términos como "Booking", "Reference" y otra serie de contenidos que hacen que sea fácil reconocer para el motor que se trata de una reserva.
Por Frases	Permite complementar al motor de Clasificación por Contenido, forzando la elección de una clase si se encuentra una o más frases en concreto. Aunque no está recomendado en la mayoría de casos, también puede utilizarse de manera individual, siempre y cuando se pueda asegurar que con pocas frases se puede clasificar un documento. De lo contrario se estaría haciendo un esfuerzo innecesario al poder aprovechar la potencia del motor de Clasificación por Contenido. No requiere usar conjuntos de aprendizaje ya que clasifica el documento en función de lo que el reconocimiento óptico lea en el momento.	No	Alta ya que permite distinguir rápidamente de entre los documentos referentes a las reservas, cuales hacen referencia a una reserva nueva, cuáles a una cancelación y cuáles a una corrección.
Por tamaño de imagen	Clasifica los documentos en función del tamaño del fichero. No es muy útil para distinguir unos documentos de otros ya que no tiene en cuenta el contenido del documento pero sirve para descartar documentos cuyo tamaño se salga de ciertos umbrales que se definan.	No	Nula porque la información referente a reservas puede encontrarse en emails muy cortos o en adjuntos de gran tamaño.
De formularios	Utilizado para distinguir documentos estructurados como formularios que tengan un campo identificador que pueda ser reconocido para ser rápidamente clasificados.	No	Media, podría complementar a la clasificación por Contenido, sustituyendo la clasificación por Frases.
Por disposición	También hace uso de la tecnología de redes neuronales Brainware, por lo que requiere conjunto de aprendizaje. Mientras que la clasificación por contenido tiene en	Si	Media. Puede complementar la clasificación por contenido dado que los adjuntos tienen un formato reconocible. No obstante, los correos

	<p>cuenta la información contenida en el documento, esta clasificación tiene en cuenta la ubicación de esa información, dividiendo el documento en bloques o párrafos de manera que pueden distinguirse unas plantillas de otras.</p>		<p>electrónicos no tienen ningún formato específico y es más importante el contenido del correo.</p>
--	---	--	--

Tabla 29 Motores de Clasificación disponibles en Forms Recognition

A continuación se presentan una serie de ejemplos de cada clase:



Please take note of the following:
Buena Vista Palace Resort
1900 BUENA VISTA DRIVE 32830
LAKE BUENA VISTA - USA
Produced on date: 28-MAR-12 16:50

Please indicate our reference number on each of your invoices.

Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider.

Please send the confirmation numbers by fax to the number: 407 363 7508

----- NEW BOOKING -----
REFERENCE: 235-1338904
PASSENGER'S NAME(S): ROSANE SILVA x 3 Fax
ORIGIN MARKET: BRAZIL

DATE: From 26-APR-12 to 01-MAY-12 (5 nights) AT AGREED RATE
CONTRACT NAME: P-16095 TYPE: FIT CONTRACT
ROOM(S): 1 Double or Twin, TYPE: GUEST ROOM
BOARD: ROOM ONLY
1 ADULTS 2 CHILDREN PER ROOM
RATES:
Double or Twin (GUEST ROOM)
From 26-APR-12 to 01-MAY-12 at CONTRACT RATES rate
ALLOCATION
From 26-APR-12 to 01-MAY-12 Free Sale
PASSENGERS:
ROSANE SILVA - CARLOS AZEVEDO(4) - ATHINA AZEVEDO(6)
HOTEL:Buena Vista Palace Resort. 1900 BUENA VISTA DRIVE LAKE BUENA VISTA, USA
Produced on date: 28/03/12 16:50

Confirmation number -----

Ilustración 51 Nueva Reserva (New Booking)



Please take note of the following:

The Dunes Hotel overlooking the Falls
6546 Fallsview Boulevard L2S 3M2
NIAGARA FALLS - CANADA

Produced on date: 21-NOV-12 06:02

Please indicate our reference number on each of your invoices.

Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider.

Please send the confirmation numbers by mail to: canadareservations@hotelbeds.com

----- CANCELLATION -----
REFERENCE: 257-123427
Creation Date: 19-NOV-12 19:21:33
Cancellation Date: 21-NOV-12 12:53:16
PASSENGER'S NAME(S): IOHNNIS KYRIATSOUIS x 2
ORIGIN MARKET: GREECE

DATE: From 26-DEC-12 to 29-DEC-12 (3 nights)
CONTRACT NAME: STA TYPE: SPECIAL RATE
ROOM(S): 1 DOUBLE, TYPE: 2 QUEEN BEDS-FALLSIDE-STANDARD
BOARD: AMERICAN BREAKFAST
2 ADULTS PER ROOM
RATES:
DOUBLE (2 QUEEN BEDS-FALLSIDE-STANDARD)
From 26-DEC-12 to 29-DEC-12: flex rate at 55 CAD room/night
(BASE PRICE : AMERICAN BREAKFAST BASIS)
ALLOCATION
From 26-DEC-12 to 29-DEC-12 Free Sale

Reservations: Please go to our supplier website www.hotelbeds.com/suppliers for additional information on bookings
In House reservations: For any issues with our In house reservations please contact usa.customercare1@hotelbeds.com
No Shows: to report NO Shows please contact us at nohow.usa@hotelbeds.com
Finance Payments: Please send payment queries to accountspayables@hotelbeds.com
Finance Invoicing: Please send invoices to invoices@hotelbeds.com PH: 407-345-4996 Ext 14002 Fax: 407-351-8430

Ilustración 52 Cancelación (Cancellation)



HOTELBEDS USA/CANADA
 5422 Carrisa Dr., Suite 201, .
 Orlando FL 32819
 Fax: +14075834682 Tel: +14073454556
 email: canadareservations@hotelbeds.com

Please take note of the following:

The Fairmont Royal York
 100 FRONT STREET WEST MSJ 1E3
 TORONTO - CANADA

Produced on date: 19-NOV-12 19:47

Please indicate our reference number on each of your invoices.

Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider.

Please send the confirmation numbers by mail to: canadareservations@hotelbeds.com

----- AMENDMENT -----
 REFERENCE: 257-122363
 ----- CURRENT BOOKING -----
 PASSENGER'S NAME(S): JAMES THOMPSON x 2 Pax
 INITIAL MAKEUP: UNKNDH

 DATE: From 28-NOV-12 to 04-DEC-12 (6 nights)
 CONTRACT NAME: S242-NETPACKG TYPE: SPECIAL RATE / AEB CONTRACT
 ROOM(S): 1 DOUBLE, TYPE: FALKMUNT
 BOARD: ROOM ONLY
 2 ADULTS PER ROOM
 RATES:
 FROM 28-NOV-12 TO 04-DEC-12 (6 nights) flex rate at 109.84 CAD
 ROOM/NIGHT
 (BASE PRICE : ROOM ONLY BASIS)
 ALLOCATION:
 From 28-NOV-12 to 04-DEC-12 Included Allotment

Reservations: Please go to our supplier website www.hotelbeds.com/suppliers for additional information on bookings
 In House Reservations: For any issues with our in house reservations please contact usa.customercare12@hotelbeds.com
 For Shows to report NCI Shows please contact nci.newshows@hotelbeds.com
 Finance Payments: Please send payment queries to accounts.payables@hotelbeds.com
 Finance Invoicing: Please send invoices to invoices@hotelbeds.com PH: 407-345-4686 Ext 14002 Fax: 407-361-8130

Ilustración 53 Corrección (Amendment)

ANEXO CONTRATO M2M PARA TRÁFICO GPRS DE LOS TPV'S DE EXCURSIONES.

Toda la información contenida en este documento tiene carácter confidencial y el acceso a la misma debe restringirse al personal del cliente que deba tener relación con el mismo para su evaluación e ejecución.
No se revelará el contenido de este documento, en todo o en parte, por ningún medio, a terceros, sin que medie previamente la autorización escrita de Telefónica Móviles España.



Ilustración 54 No es reserva (Void)

Como se puede apreciar, es muy sencillo distinguir los adjuntos pues tienen un formato común en el que únicamente cambia una palabra que indica si se trata de una nueva reserva, una

corrección o una cancelación. Además, se cumple que dicha palabra o frase no aparece en los otros 2 subtipos. Por lo tanto, para clasificar los 3 subtipos de Reserva sería suficiente con utilizar el sencillo motor de clasificación por frases, quedando la configuración de la clasificación de la siguiente manera:

- **Booking:** Esta clase únicamente se define para organizar los 3 subtipos y poder crear campos de extracción comunes a todos los subtipos de una manera más conveniente. Se fuerza a ejecutar la sub-clasificación para detectar si es una corrección (Amendment), cancelación (Cancellation) o nueva (New).
 - **Amendment:** Clasificación por frases. Únicamente se necesita una palabra para para sub-clasificar la reserva como una corrección → *Ammendment*.
 - **Cancellation:** Clasificación por frases. Únicamente se necesita una palabra para para sub-clasificar la reserva como una corrección → *Cancellation*.
 - **New Booking:** Clasificación por frases. Únicamente se necesita una frase para para sub-clasificar la reserva como una corrección → *New Booking*.
- **Void:** Como se ha indicado anteriormente, bajo esta clase quedarán todos los documentos que no sean reservas. Por lo tanto, no se aplicará ningún motor de clasificación específico para esta clase si no que se configura el proyecto de manera que cuando el motor de clasificación por frases no sea capaz de clasificar un documento por no tratarse de una reserva, se clasificará automáticamente como esta clase.

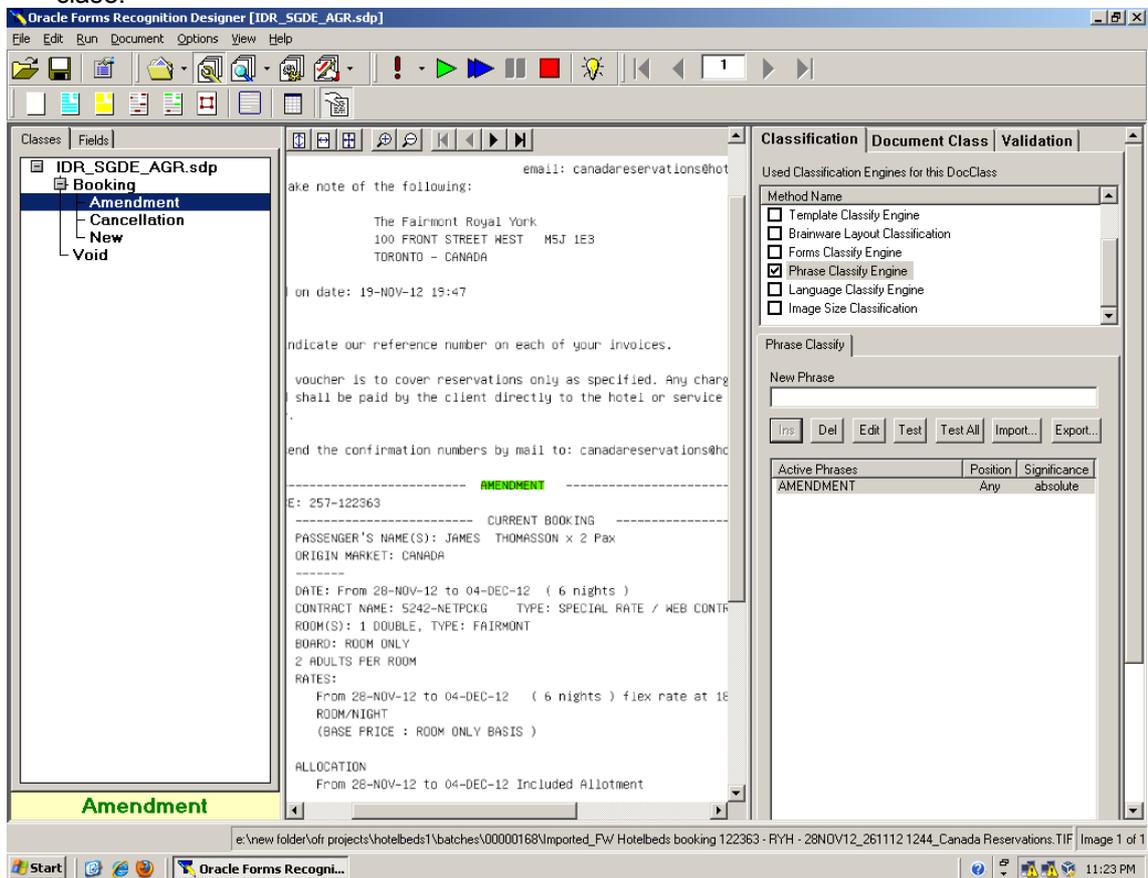


Ilustración 55 Realce de la palabra identificada por el motor de Clasificación por Frases

6.4.3. Extracción de campos

Gracias a la organización de clases definida anteriormente, únicamente se deben crear los campos de extracción en la súper clase Booking y los mismos serán replicados a cada uno de los 3 subtipos. Los campos que se desea extraer de manera automática son:

- **Nº de Referencia:** Un identificador que identifica unívocamente cada reserva. Está formado por un grupo inicial de 3 dígitos y otro de 6, separados por un guion. Por ejemplo, 257-123110.

- Fecha: La fecha de la reserva. El formato con que se encuentra en los documentos es DD-MMM-AA donde DD es el día del mes indicado con 2 dígitos, MMM son las 3 primeras letras del mes en inglés y AA el año representado con 2 dígitos nuevamente. Por ejemplo, el 20 de Noviembre de 2012 sería 20-NOV-12.
- Hotel de la reserva

Forms Recognition ofrece las siguientes técnicas de extracción de campos:

Técnica	Descripción	Relevancia para el proyecto
Análisis de Tabla	Utilizado para identificar elementos agrupados u organizados como si en una tabla estuvieran. Por ejemplo, los ítems de línea de una factura.	Nula, ninguno de los datos a extraer se encuentra organizado como si de una tabla se tratase.
Análisis de Formato	Se vale de expresiones regulares para identificar campos que cumplan un cierto patrón como pueden ser cantidades monetarias o fechas.	Alta, dado que las referencias y las fechas tienen un patrón identificable.
Análisis de Zona	Permite definir zonas concretas de las que aplicar OCR local y asignar el valor extraído al campo. Permite definir anclajes en los documentos como las esquinas de la hoja o ciertas palabras clave para que la zona siempre quede relativa a esos anclajes, de manera que la extracción sea óptima aunque el documento esté movido. Suele utilizarse para el reconocimiento de formularios.	Baja. Podría utilizarse para extraer el hotel que siempre se encuentra en la misma posición dentro del documento.
Motor de búsqueda asociativo	Este motor requiere de una fuente de datos como una base de datos o un fichero separado por comas de la que comparar los valores con todas las palabras reconocidas del documento. Trata de asociar una de las líneas o filas de la fuente de datos con los datos que aparecen en el documento que se está reconociendo y si una de las filas tiene un número suficiente de ocurrencias dentro del documento, se asociarán los valores de dicha fila al campo de extracción. Suele utilizarse para la extracción del proveedor en una factura.	Alta. Al tratarse de confirmaciones de reservas, los datos de los hoteles ya figuran en la base de datos del banco de camas por lo que puede utilizarse dicha fuente de datos para extraer el hotel al que pertenece la reserva que se está analizando.

Tabla 30 Técnicas de Extracción de Campos disponibles en Forms Recognition

Dadas las características de los campos a extraer y las capacidades de cada técnica de extracción, es sencillo deducir que la mejor estrategia para automatizar la extracción de los campos deseados consiste en aplicar las siguientes técnicas en cada campo:

- **Referencia:** Análisis de Formato. Las referencias pueden ser reconocidas por la expresión regular `[0-9]{3}[-]{1}[0-9]{6}`
- **Fecha:** Análisis de Formato. Las fechas pueden ser reconocidas por la expresión regular `[0-9]{1,2}[-]{1}[A-Z]{3}[-]{1}[0-9]{2}`
- **Hotel:** Motor de Búsqueda Asociativo. Al tratarse de confirmaciones, la información del hotel ya figura en las bases de datos del banco de cama por lo que para asociar el documento con su hotel correspondiente con la mayor confianza posible, se opta por el Motor de Búsqueda Asociativo.

La extracción de campos quedaría configurada de la siguiente manera en Forms Recognition:

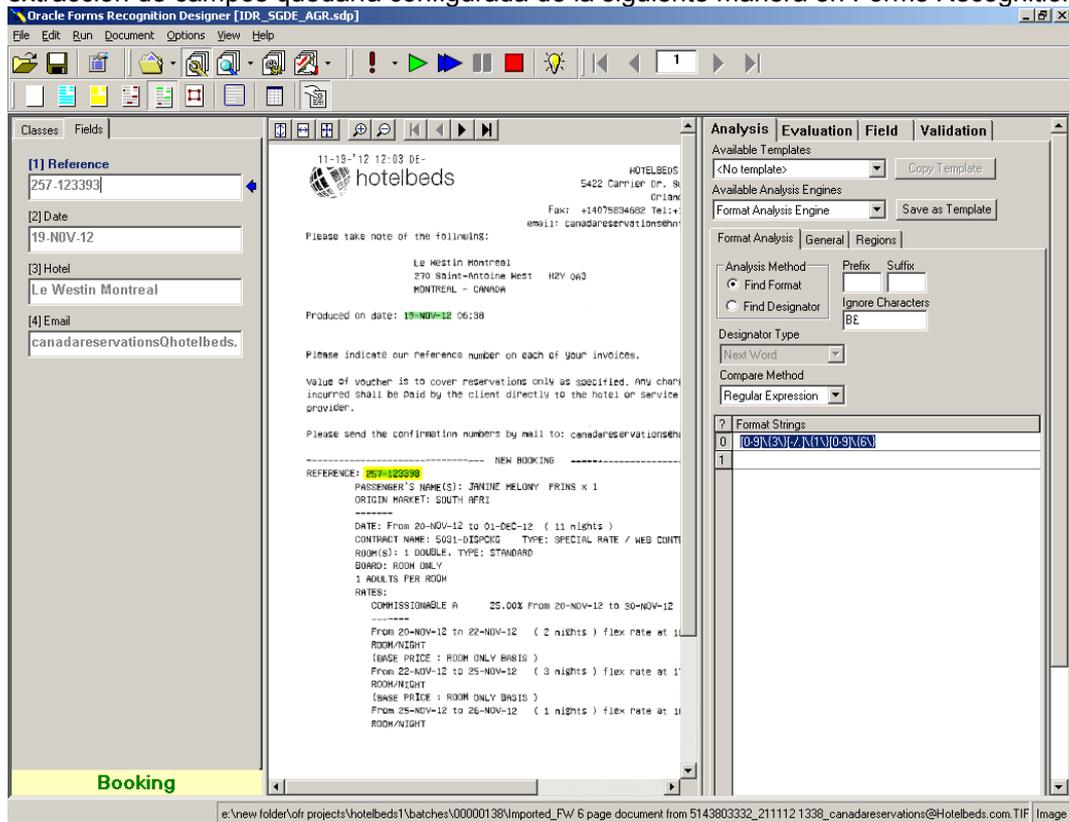


Ilustración 56 Configuración de los campos de extracción en Forms Recognition, con los campos reconocidos por formato resaltados

En la ilustración puede observarse como Forms Recognition resalta en verde los campos definidos con extracción por análisis de formato: la referencia y la fecha. En la siguiente ilustración se verá la configuración de la búsqueda asociativa.

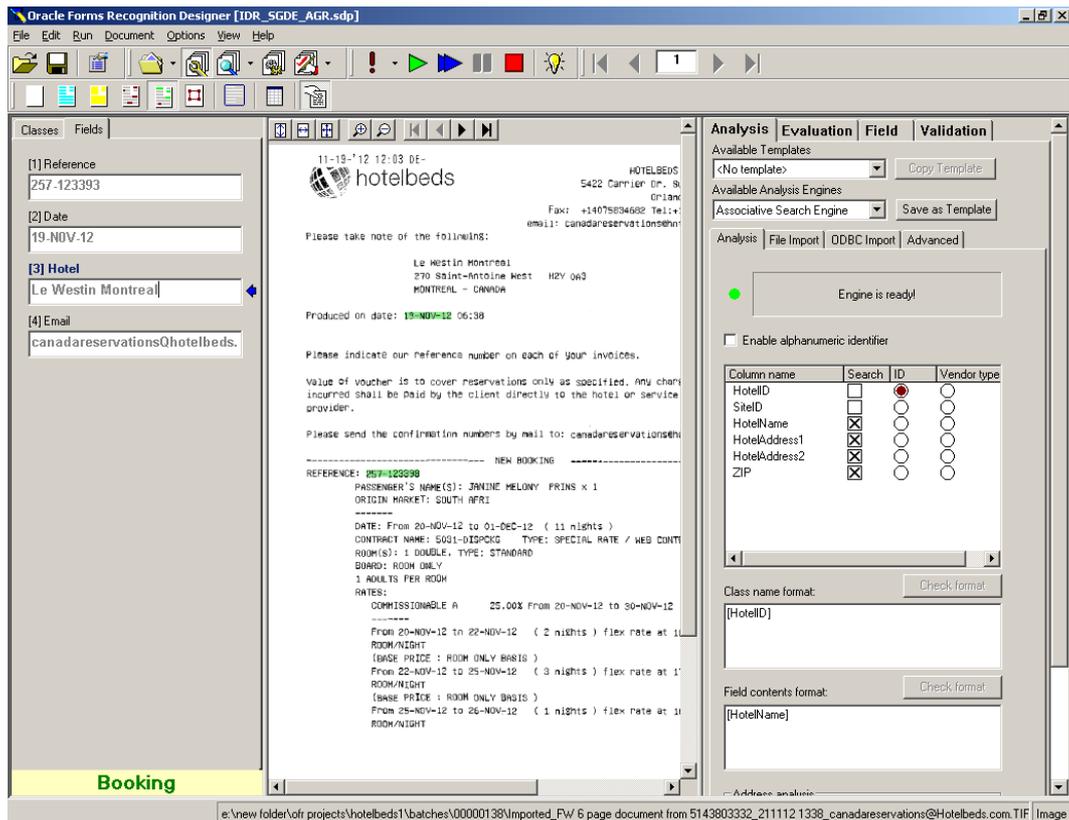


Ilustración 57 Configuración de búsqueda asociativa

En la ilustración puede verse como, una vez importado el fichero separado por comas y conectado a una base de datos, se debe configurar que campos se desea buscar en el documento. A continuación se presenta un extracto del fichero separado por comas de donde se identifica el hotel al que pertenece esta reserva:

```

HotelID;SiteID;HotelName;HotelAddress1;HotelAddress2;ZIP
[...]
00023;00023;Doubletree by Hilton Toronto Airport;655 Dixon
road;TORONTO - CANADA;M9W 1J0
00024;00024;Holiday Inn Vancouver Airport-Richmond;10720 CAMBIE
ROAD;RICHMOND - CANADA; V6X 1K9
00025;00025;Le Westin Montreal;270 Saint-Antoine West;MONTREAL -
CANADA;H2Y 0A3
00026;00026;Novotel Montreal Center;1180 RUE DE LA
MONTAGNE;MONTREAL - CANADA;H3G 1Z1
00027;00027;Crowne Plaza Toronto Airport;33 Carlson Court;Toronto -
[...]

```

Adicionalmente, con fines estadísticos que se mostrarán en Business Activity Monitoring, se desea recoger tanto el correo original del remitente del adjunto así como la dirección de correo electrónico del banco de camas al que se envía dicho adjunto. Como se describió anteriormente, Enterprise Capture capturaba el correo electrónico del remitente y lo escribía en el nombre del fichero TIFF que exportaba a Forms Recognition con el siguiente formato asunto@@direccion1@correo.com@@direccion2@correo.com@@

0000000001.TIFF. Forms Recognition debe recuperar dichos email del nombre del fichero con el fin de incluirlo en los resultados de extracción de manera que WebCenter Imaging pueda importarlos y presentarlos a los usuarios finales a través del proceso de negocio, cómo es requisito del proyecto. Esto se hará mediante una sencilla rutina aprovechando la capacidad de extensión mediante Visual Basic Script que ofrece Forms Recognition.

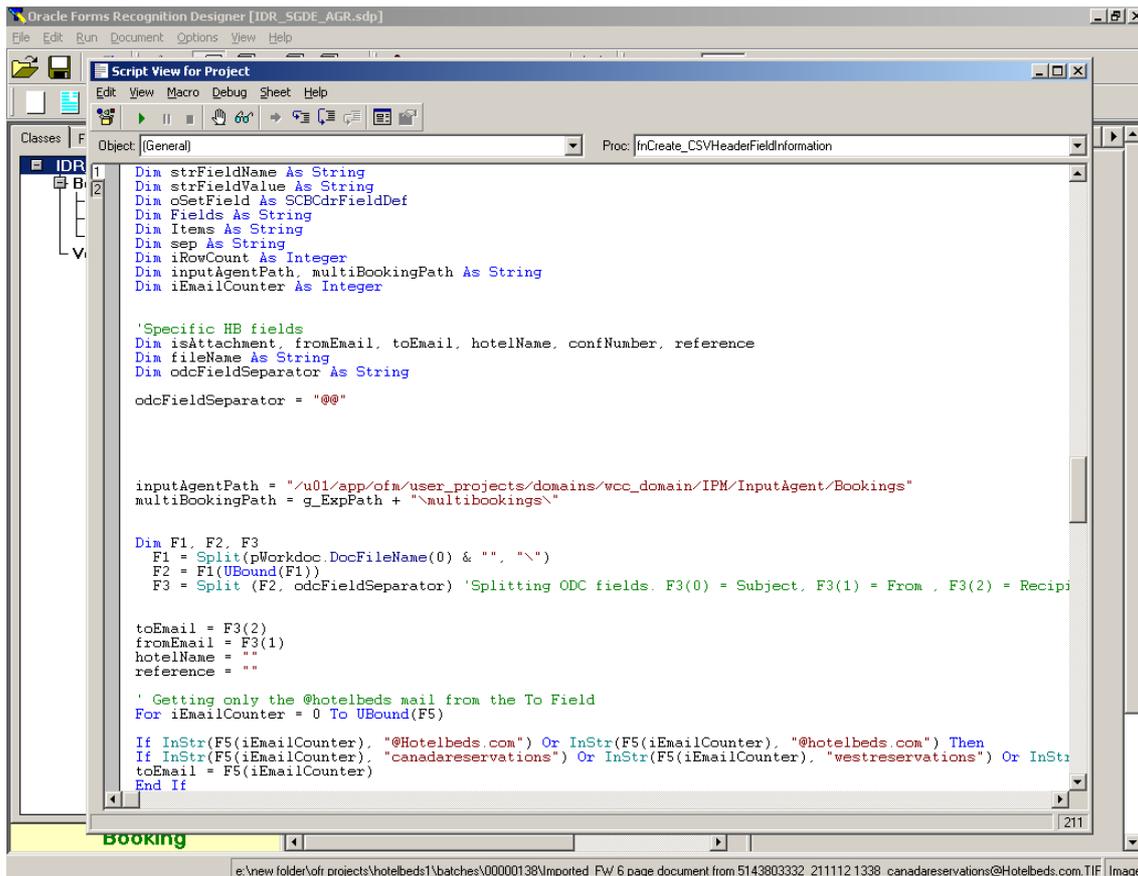


Ilustración 58 Extensión de Forms Recognition mediante Visual Basic Script

La siguiente rutina captura los correos electrónicos del nombre del fichero y los campos extraídos automáticamente mediante reconocimiento y los escribe en una variable que en otra rutina son escritas a un fichero que utilizará WebCenter Imaging para importar las reservas junto con sus campos extraídos.

```

Public Function fnExport_to_Imaging(pWorkdoc As
SCBCdrPROJLib.SCBCdrWorkdoc)
'pWorkdoc es el objeto que contiene toda la información extraída por
Forms Recognition
Dim Fields As String
Dim sep As String
Dim inputAgentPath As Integer

' Campos que se deben pasar a WebCenter Imaging
Dim fromEmail, toEmail, hotelName, date, reference, fileName,
odcFieldSeparator As String

odcFieldSeparator = "@@"

inputAgentPath =
"/u01/app/ofm/user_projects/domains/wcc_domain/IPM/InputAgent/Bookings"

Dim F1, F2, F3
F1 = Split(pWorkdoc.DocFileName(0) & "", "\") 'Separando el nombre
del fichero del resto de la ruta obtenido por el atributo DocFileName
F2 = F1(UBound(F1)) 'Obteniendo el nombre del fichero, el último
ítem del array anterior.
F3 = Split(F2, odcFieldSeparator) 'Separando los campos del nombre
del fichero F3(0) = Asunto, F3(1) = Remitente , F3(2) = Destinatario

```

```

toEmail = F3(2)
fromEmail = F3(1)
hotelName = ""
reference = ""
date = ""

If pWorkdoc.DocClassName <> "Void" Then 'Si el documento no ha sido
clasificado como Void, es decir, es una reserva, se capturan los
valores extraídos para Hotel, Referencia y Fecha

hotelName = pWorkdoc.Fields("Hotel")
reference = pWorkdoc.Fields("Reference")
date = fnConvertDate(pWorkdoc.Fields("Date"))'Función para convertir
el formato de la fecha de DD-MMM-AA a DD/MM/AAAA

End If

fileName = F2

sep = "|"

Fields = inputAgentPath + "/" + fileName + sep + reference + sep +
pWorkdoc.DocClassName & sep & fromEmail & sep & toEmail & sep &
hotelName & sep & date & sep

fnAttach_CSVLine(Fields)

End If

End Function

```

El resultado final es un fichero csv llamado InputBookings.csv con una línea como la siguiente por cada reserva procesada:

```

/u01/app/ofm/user_projects/domains/wcc_domain/IPM/InputAgent/Bookin
gs/Booking Confirmation for 257-123393@@fax-
123456@reply.fax2mail.com@@canadareservations@hotelbeds.com@@000000
1.pdf|257-123393|New|fax-123456@reply.fax2mail.com|
canadareservations@hotelbeds.com.|Le Westin Montreal|19/11/2012|

```

Se han resaltado los separadores en rojo. Dicho fichero separado tendría el siguiente formado de columnas.

Columna 1	Columna 2	Columna 3	Columna 4	Columna 5	Columna 6	Columna 7
Ubicación completa del fichero	Referencia	Tipo	Remitente	Destinatario	Hotel	Fecha
/u01/app/[...]0000001.pdf	257-123393	New	fax-123456@reply.fax2mail.com	canadareservations@hotelbeds.com	Le Westin Montreal	19/11/2012

Tabla 31 Configuración Fichero Exportado por Forms Recognition

Este fichero será el que utilice WebCenter Imaging para importar el fichero que figura en el primer campo y asignarle los metadatos de los campos siguientes.

Con esto queda configurado el proyecto de reconocimiento. Se debe configurar el proceso de Runtime Service para que ejecute dicha configuración de manera automática.

6.4.4. Configuración de Runtime Service

Utilizando la Consola de Administración de Runtime Service, debe crearse un proceso con la siguiente configuración (se indican los valores de configuración más relevantes a lo largo de las distintas pestañas):

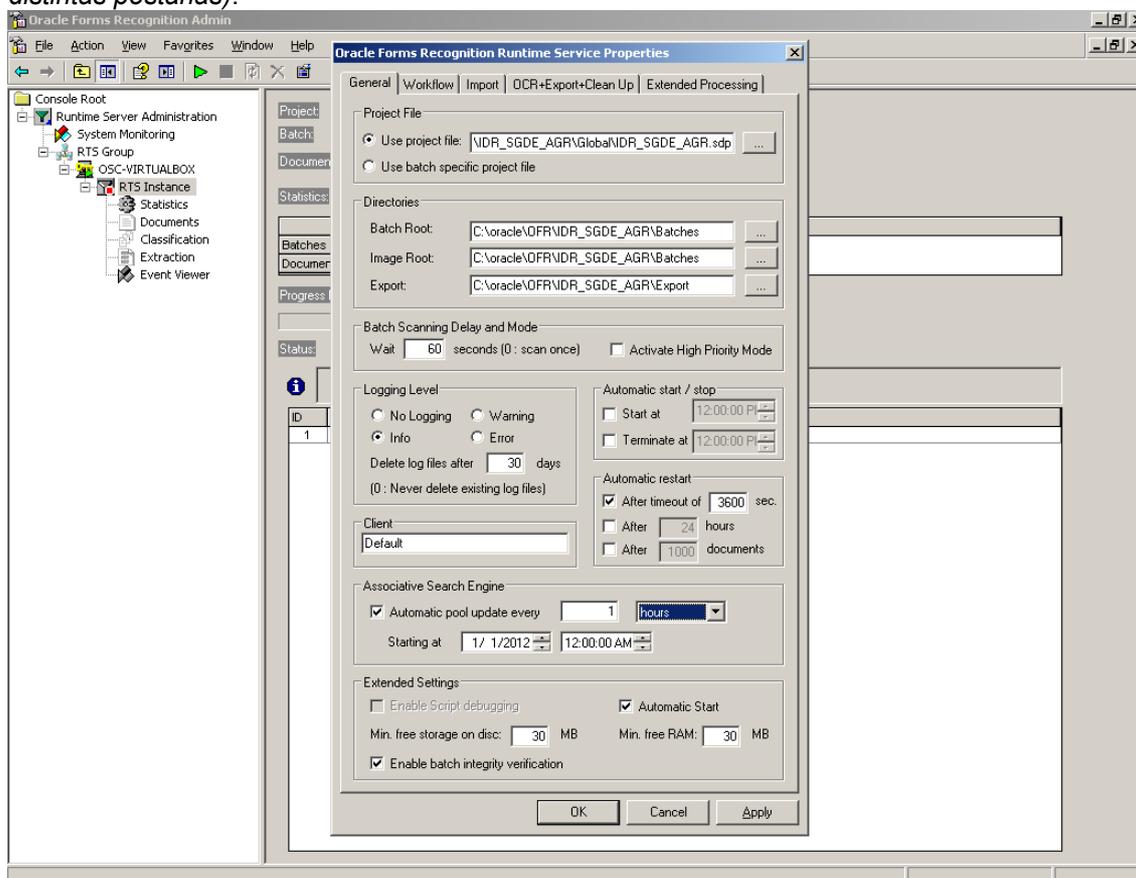


Ilustración 59 Configuración Runtime Service de Forms Recognition

Campo	Descripción	Valor
1.- General		
Proyecto de Reconocimiento	Ruta completa hacia el fichero .SDP que contiene la información de clasificación y extracción de campos que se desea utilizar.	%FR_HOME% ⁷ / IDR_SGDE_AGR.sdp
Carpeta de lotes de imágenes	Ruta completa donde Forms Recognition almacenará los lotes de imágenes	%FR_HOME%/Batches
Carpeta de exportación	Carpeta en la que se situarán los resultados de exportación como son las imágenes procesadas (opcionalmente convertidas a PDF) y/o los ficheros que contienen los campos extraídos a dichas imágenes.	%FR_HOME%/Export ⁸

⁷ %FR_HOME% es la ubicación en el servidor Windows donde se encuentran todos los ficheros relacionados con el proyecto de reconocimiento inteligente del SGDE para la Automatización de la Gestión de Reservas (por ejemplo, C:\oracle\OFR\IDR_SGDE_AGR)

⁸ Esta ubicación está compartida con el clúster wcc de manera que WebCenter Imaging pueda recoger los resultados de exportación mediante el Input Agent

Intervalo de importación	Intervalo de tiempo que el proceso recupera nuevos documentos de la ubicación definido en la carpeta de importación, creando un nuevo lote con ellos y procesándolos.	60 Segundos
Actualización del pool de datos para el motor asociativo	Si se desea actualizar automáticamente el fichero del que el motor asociativo obtiene sus datos de manera automática y, en caso positivo, con qué frecuencia.	Si, con una frecuencia de 1 hora
Arranque automático	Indica si esta instancia concreta de Runtime Service arranca automáticamente cuando arranque Runtime Service. El arranque automático de Runtime Service fue configurado anteriormente como parte de la instalación y configuración inicial de Forms Recognition.	Activado.
2.- Workflow		
Pasos activados	Indica qué pasos del proceso de reconocimiento automatizado realizará esta instancia: <ul style="list-style-type: none"> • Importación • OCR • Clasificación • Extracción • Exportación • Limpieza 	<ul style="list-style-type: none"> • Importación • OCR • Clasificación • Extracción • Exportación • Limpieza
3.- Import		
Carpeta de importación	Ubicación en el servidor de Windows de donde esta instancia de Runtime Service recuperará los documentos a ser procesados	%FR_HOME%/Import ⁹
4.- OCR+Export+Clean Up		
Restricción de OCR	Indica si se desea limitar a qué páginas aplicar el OCR	Únicamente a la primera página ya que en ella se encuentra toda la información necesaria. Esto ayudará en gran medida a reducir la carga de trabajo del servidor de Forms Recognition.
Generación de PDF	Indica si se desea generar un PDF a partir del documento analizado	Activado
Ejecutar script de exportación	Indica si se debe ejecutar el script de exportación específico diseñado en Designer	Activado

Tabla 32 Parámetros de Runtime Service de Forms Recognition

Con esto queda configurado Forms Recognition, el componente de reconocimiento inteligente de documentos del SGDE para la Automatización de la Gestión de Reservas.

⁹ Esta ubicación está compartida con el clúster wcc de manera que Enterprise Capture pueda dejar en esta ubicación los documentos que se capturen de los correos electrónicos.

6.5 Configuración de WebCenter Imaging

WebCenter Imaging es el componente de Gestión Documental del SGDE para la Automatización de la Gestión de Reservas y es el componente que almacenará de manera permanente las reservas, disparando un proceso de negocio BPM por cada reserva que importe y ofreciendo un visor rico en funcionalidad para visualizar las reservas dentro del proceso. Mediante su consola de administración se debe configurar:

- Conexión al servidor de contenidos local, donde se almacenarán las reservas
- Aplicación de Reservas y sus respectivos metadatos a asociar
- El formato del fichero de entrada para la importación automatizada y mapeo con los metadatos
- La conexión con el componente de procesos, BPM Suite, y la selección del proceso apropiado.

6.5.1 Configuración de Conexiones

Mediante el menú de gestión de conexiones se configuran la conexión al Servidor de Contenidos local y al Servidor de Procesos, BPM Suite, instalado en el clúster bpm-cluster.

The screenshot shows the Oracle WebCenter Content: Imaging administration console. The main area displays the configuration for a connection named 'WCC'. The configuration is organized into several sections:

- Basic Information:** Name: WCC, Description: Content Server Repository, Connection Type: Content Server Repository.
- Connection Settings:** SSL: false, Use Local Content Server: true, Server Port: 4444.
- Content Server Pool:** A table with columns 'Machine' and 'Server Port'. The row shows 'content' and '4444'.
- Security:** A table with columns 'Type', 'Security Member', 'View', 'Modify', 'Delete', and 'Grant Access'. The row shows 'weblogic' with green checkmarks in the 'View', 'Modify', 'Delete', and 'Grant Access' columns.
- Audit History:** A table with columns 'Date', 'Type', and 'User Name'. The row shows '6/5/2015', 'Definition Create', and 'weblogic'.

The left sidebar contains navigation options: Searches, Tools, Manage Searches, Manage Applications, Manage Inputs, and Manage Connections. The bottom of the console shows the copyright notice: Copyright (c) 2010, 2013 Oracle and/or its affiliates. All rights reserved.

Ilustración 60 Configuración de conexiones en WebCenter Imaging

Es necesario facilitar los siguientes parámetros:

Campo	Descripción	Valor
1.- Servidor de Contenidos		
Nombre	Nombre para distinguir esta conexión	WCC
Tipo de Conexión	Si se trata de una conexión a un Servidor de Contenidos o a un Servidor de Procesos	Servidor de Contenidos
Servidor de Contenidos Local	Indica si se va a utilizar el servidor de Contenidos local, instalado en el mismo dominio que Imaging. Esto establece el nombre del servidor y el puerto automáticamente.	Si

2.- Servidor de Procesos		
Nombre	Nombre para distinguir esta conexión	BPM
Tipo de Conexión	Si se trata de una conexión a un Servidor de Contenidos o a un Servidor de Procesos	Servidor de Procesos
Dirección HTTP y puerto	Dirección HTTP y puerto donde se encuentra BPM Suite	http://<<IP de bpm-vm-x1>>:8001
Credenciales	Usuario y contraseña de administrador para establecer la conexión	Usuario y contraseña de administración configurados por el consultor/responsable de BPM Suite.

Tabla 33 Parámetros de conexiones en WebCenter Imaging

Con estos parámetros quedarían configuradas ambas conexiones necesarias para los siguientes pasos de configuración de WebCenter Imaging.

6.5.2 Aplicación de Reservas

Una aplicación en Imaging es un tipo de documento junto a sus metadatos y su proceso de BPM asociado. Desde el menú de configuración de aplicaciones debe crearse la aplicación de Bookings (Reservas) con los siguientes parámetros.

The screenshot shows the Oracle WebCenter Content: Imaging configuration page for an application named 'Reservas'. The page is titled 'Reservas: Application Summary' and includes several sections:

- General Properties:** Application Id: 2, Application Name: Reservas, Description: Reservas, Repository: WCC, Full-Text Option: Full-Text Search.
- Field Definitions:** A table with columns: Type, Name, Length, Scale, Required, Indexed, Default Value, Picklist.

Type	Name	Length	Scale	Required	Indexed	Default Value	Picklist
Abc	Referencia	80			✓		
Abc	Tipo	80			✓		
Abc	Remitente	200			✓		
Abc	Hotel	200			✓		
Abc	Fecha	200			✓		
Abc	Destinatario	200			✓		
- Application Security:** A table with columns: Type, Security Member, View, Modify, Delete, Grant Access.

Type	Security Member	View	Modify	Delete	Grant Access
weblogic		✓	✓	✓	✓
- Document Security:** A table with columns: Type, Security Member, View, Write, Delete, Grant Access, Lock Admin, Annotate Standard, Annotate Restricted, Annotate Hidden.

Type	Security Member	View	Write	Delete	Grant Access	Lock Admin	Annotate Standard	Annotate Restricted	Annotate Hidden
Administrators		✓	✓	✓	✓	✓	✓	✓	✓
- Storage Policy:** Document Storage: Volume: File default; Supporting Content Storage: Volume: File default.
- Workflow Configuration:** (Empty section)

Ilustración 61 Creación de la Aplicación de Reservas en WebCenter Imaging

Campo	Descripción	Valor
1.- General		
Nombre	Nombre de la aplicación	Reservas
Repositorio	Nombre de la conexión de Servidor de Contenidos	WCC
Indexación del texto del documento	Indica si se desea indexar todo el texto del documento de manera que se puedan realizar búsquedas más	Activado

	adelante por el contenido del mismo, además de los metadatos	
2.- Definición de metadatos		
Referencia	Referencia de la reserva	Importado automáticamente de la extracción de Forms Recognition
Tipo	Tipo de Reserva: <ul style="list-style-type: none"> • Nueva • Corrección • Cancelación • Void (no es reserva) 	Importado automáticamente de la extracción de Forms Recognition
Remitente	Correo electrónico del remitente de la reserva	Importado automáticamente de la extracción de Forms Recognition
Hotel	Hotel al que hace referencia la reserva	Importado automáticamente de la extracción de Forms Recognition
Fecha	Fecha de la reserva	Importado automáticamente de la extracción de Forms Recognition
Destinatario	Una de las 6 cuentas de correo del banco de camas que reciben reservas	Importado automáticamente de la extracción de Forms Recognition
3.- Seguridad		
Seguridad de la aplicación	Define quien puede acceder a la definición de esta aplicación desde la consola de administración de WebCenter Imaging. Útil para delegar tareas de administración.	Responsable de WebCenter y responsable del proceso de Confirmación de Reservas.
Seguridad de documentos	Define quien puede acceder a los documentos que se suben a esta aplicación	Todo el personal encargado de gestionar la Confirmación de Reservas
4.- Proceso de Negocio		
Conexión con Servidor de Procesos	Nombre de la conexión con el Servidor de Procesos que se desea utilizar	BPM
Proceso y versión	Proceso configurado y desplegado sobre el Servidor de Procesos que se desea lanzar ante la subida de un documento a esta aplicación.	Confirmación de Reservas 1.0 ¹⁰

Tabla 34 Parámetros de la Aplicación de Reservas en WebCenter Imaging

Con esta configuración es posible subir reservas de manera manual a Imaging que disparen procesos de Confirmación. En el siguiente apartado se configurará el fichero de entrada de manera que se puedan importar reservas de manera automática.

¹⁰ Creado y configurado por el consultor/responsable de BPM Suite en el clúster `bpm-cluster`

6.5.3 Entrada automatizada

Desde el menú de configuración de entradas, configurar la entrada automatizada consiste en cargar un fichero de importación de ejemplo como los que genera de manera automática el proyecto de reconocimiento de Forms Recognition previamente configurado, indicar cuál es el separador y mapear las columnas del fichero con los metadatos de la aplicación de Imaging.

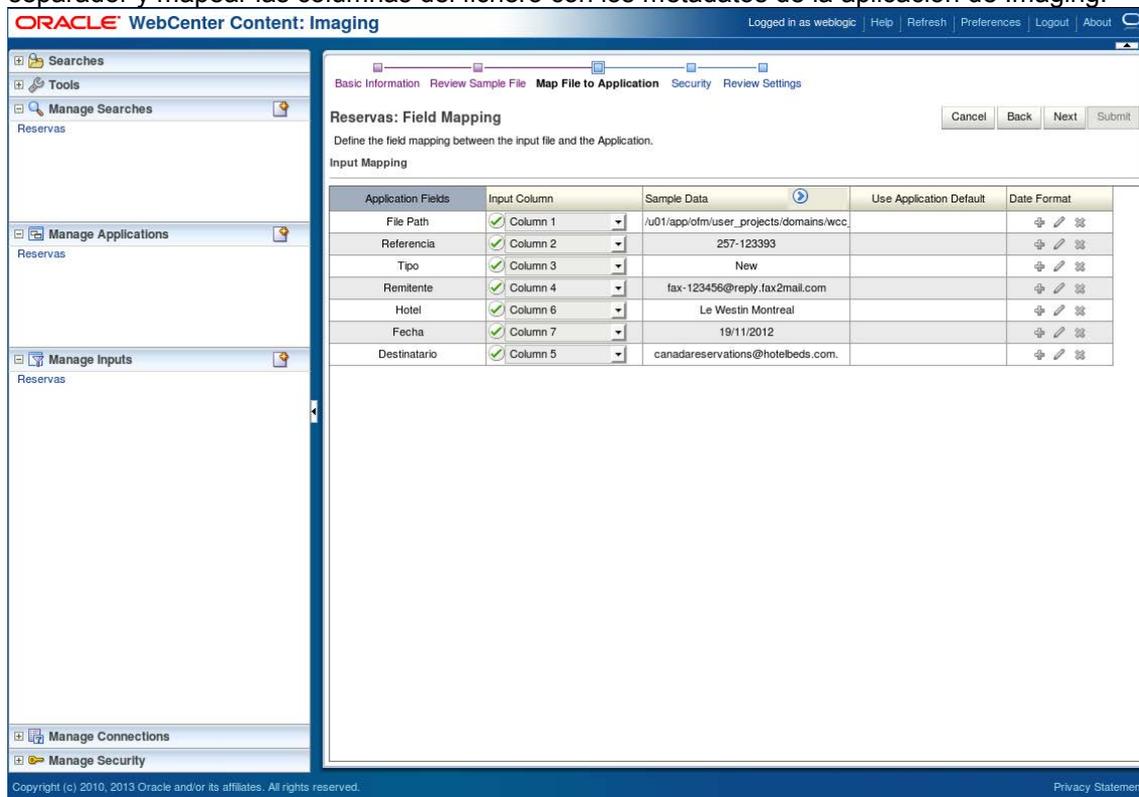


Ilustración 62 Mapeo de campos de entrada en WebCenter Imaging

Campo	Descripción	Valor
1.- General		
Nombre	Nombre de la entrada	Reservas
Carpeta de Input Agent	Carpeta del sistema que el componente de importación de Imaging, Input Agent, monitoriza buscando el fichero de importación	\$DOMAIN_HOME ¹¹ /wcc_domain/IPM/InputAgent/Bookings
Máscara del fichero de importación	Nombre o máscara de nombre que Input Agent reconocerá como el fichero de importación	InputBookings.csv
Frecuencia de Importación	Periodicidad con la que Input Agent se activa para localizar ficheros de importación y realiza las respectivas importaciones	1 Minuto
Delimitador	Carácter que utiliza el fichero para delimitar los campos	" "

¹¹ \$DOMAIN_HOME es una variable de entorno definida previamente que hace referencia a la ubicación donde se encuentra el dominio WebLogic que incluye a Imaging.

2.- Mapeos		
Metadato	Columna del fichero	Ejemplo
Ubicación	Columna 1	/u01/app/ofm/user_projects/ domains/wcc_domain/IPM/ InputAgent/Bookings/Booking Confirmation for 257-123393@@fax- 123456@reply.fax2mail.com@@ canadareservations@hotelbeds.com@@ 0000001.pdf
Referencia	Columna 2	257-123393
Tipo	Columna 3	New
Remitente	Columna 4	fax-123456@reply.fax2mail.com
Hotel	Columna 6	Le Westin Montreal
Fecha	Columna 7	19/11/2012
Destinatario	Columna 5	canadareservations@hotelbeds.com

Tabla 35 Parámetros de la Entrada Automatizada en WebCenter Imaging

Con esto queda configurado WebCenter Imaging, el componente de gestión documental del SGDE para la Automatización de la Gestión de Reservas y, con él, quedan configurados 3 de los 4 componentes del proyecto, siendo el 4º, el componente de Gestión de Procesos BPM Suite, instalado, configurado y parametrizado por el responsable/consultor de BPM Suite.

En el siguiente apartado se expondrá un ejemplo de ejecución desde la recepción de un correo electrónico hasta la actualización de los sistemas con la confirmación de la reserva, la cancelación o la corrección.

7. Ejecución de la Confirmación de Reservas

En este apartado se ilustrará la ejecución del proceso de Confirmación de Reservas de un Banco de Camas. Con el fin de comparar el proceso antes y después de implantar el Sistema Gestor de Documentos Electrónicos para la Automatización de la Gestión de Reservas, se volverá a explicar cómo es el proceso manual actual antes de explicar cómo es con el sistema.

7.1. Proceso Manual

El proceso previo a la implantación del sistema descrito en esta memoria se ejecuta de la siguiente manera:

Varios operadores, empleado del banco de camas, están encargados de monitorizar y gestionar los correos electrónicos que llegan a 6 buzones de entrada. Estos 6 buzones de entrada reciben un total de 1100 correos electrónicos diarios. Cada operador debe realizar los siguientes pasos manuales desde su aplicación de gestión de correo electrónico

- Abrir los correos electrónicos
- Descubrir de qué se trata
- En caso de ser una confirmación de una reserva, cancelación o corrección, localizar el adjunto que contiene toda la información
- Abrir el adjunto y localizar el nombre del hotel, la fecha y el código de referencia
- Abrir la aplicación de gestión de reservas e introducir manualmente dichos datos, con la consiguiente probabilidad de equivocaciones al teclear, para confirmar la reserva, cancelación o corrección, con el fin de actualizar la base de datos del Banco de Camas

7.2. Proceso Automático

A continuación se mostrará de manera gráfica el procesado de una reserva de cada tipo.¹²

Se reciben los siguientes 3 correos electrónicos en el servidor de correo de la organización:

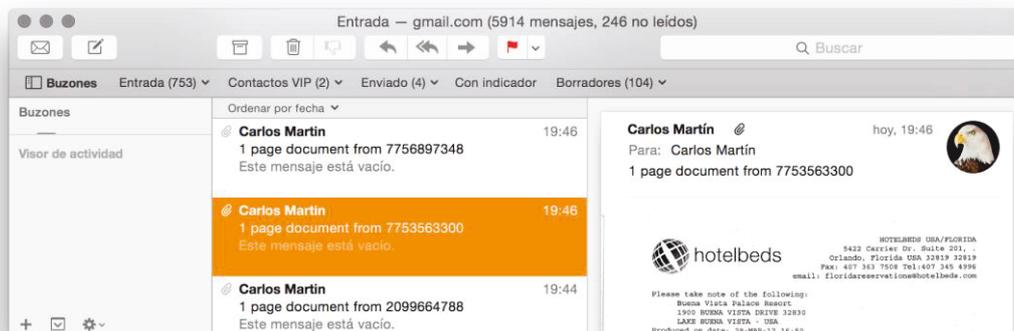


Ilustración 63 Correos electrónicos en el servidor de correo

ID	Remitente	Destinatario	Asunto
001	fax-123456@reply.fax2mail.com	canadareservations@hotelbeds.com	1 page document from 7756897348
002	fax-112233@reply.fax2mail.com	floridareservations@hotelbeds.com	1 page document from 7753563300
003	fax-789123@reply.fax2mail.com	canadareservations@hotelbeds.com	1 page document from 2099664788

Tabla 36 Información Correos Electrónicos Entrantes

¹² Se simulará el proceso utilizando una cuenta de correo electrónico particular puesto que no se tiene acceso a las direcciones de correos del Banco de Camas concreto donde se implementa el sistema

Cada correo electrónico contiene un PDF adjunto:

001	002	003
<p>hotelbeds HOTELBEDS USA/ONTARIO 5402 Centre Dr., Suite 201, Orlando, FL 32819 Fax: +14075564682 Tel: +14075564676 Email: canadareservations@hotelbeds.com</p> <p>Please take note of the following: The Dates Hotel Overlapping the Falls 5546 FALLSIDE RESORTS L25-246 MONTANA FALLS - CANADA Produced on date: 21-MAY-12 08:02</p> <p>Please indicate our reference number on each of your invoices. Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider. Please send the confirmation numbers by email to: canadareservations@hotelbeds.com</p> <p>REFERENCE: 251-125427 CANCELLATION Cancellation Date: 23-MAY-12 19:21:53 PASSENGER'S NAME(S): DOWNES KRYSTALLOU x 2 ORIGIN MARKET: GRECE DATE: From 28-DEC-12 to 29-DEC-12 (2 NIGHTS) CONTRACT NAME: 024 TWEEL SPECIAL RATE ROOM(S): 2 DOUBLE, TYPE: 2 GREEN BEDS-FULLSIZE-STANDARD BONUS: AMERICAN BREAKFAST 2 Adults Per Room NOTES: DOUBLE (2 GREEN BEDS-FULLSIZE-STANDARD) From 28-DEC-12 to 29-DEC-12: Free Rate at 105 CAD ROOM/NIGHT (ENR PRICE 1 AMERICAN BREAKFAST 0615) ALLOCATION From 28-DEC-12 to 29-DEC-12 Free Rate</p> <p>Reservations: Please go to our website www.hotelbeds.com or contact us for additional information on bookings. To make reservations, go to our website www.hotelbeds.com or contact us at canadareservations@hotelbeds.com To make bookings, go to our website www.hotelbeds.com or contact us at canadareservations@hotelbeds.com Financial Payments: Please send payment orders to canadareservations@hotelbeds.com Financial Invoices: Please send payment orders to canadareservations@hotelbeds.com. Tel: +14075564682 Fax: +14075564676</p>	<p>hotelbeds HOTELBEDS USA/FLORIDA 5422 Center Dr., Suite 201, Orlando, Florida USA 32819 32819 Fax: +14075564682 Tel: +14075564676 Email: floridareservations@hotelbeds.com</p> <p>Please take note of the following: Boona Vista Palace Resort 1500 BOONA VISTA DRIVE 32830 LAKE BUONA VISTA USA Produced on date: 28-MAY-12 16:50</p> <p>Please indicate our reference number on each of your invoices. Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider. Please send the confirmation numbers by fax to the number: 407 563 7508</p> <p>REFERENCE: 235-133824 REEM BOOKING PASSENGER'S NAME(S): ROSABE SILVA x 3 Fax ORIGIN MARKET: BRAZIL DATE: From 24-APR-12 to 01-MAY-12 (8 NIGHTS) AT AGREED RATE CONTRACT NAME: P-18795 TYPE: 212 CONTRACT ROOM(S): 1 Double or Twin, TYPE: GUEST ROOM BONUS: ROOM ONLY 1 ADULTS 2 CHILDREN PER ROOM NOTES: Double or Twin (GUEST ROOM) From 24-APR-12 to 01-MAY-12 at CONTRACT RATES rate ALLOCATION From 24-APR-12 to 01-MAY-12 Free Rate PASSENGER(S): ROSABE SILVA - CARLOS AZEVEDO(4) - ATHINA AZEVEDO(16) HOTEL: Boona Vista Palace Resort, 1500 BOONA VISTA DRIVE LAKE BUONA VISTA, USA Produced on date: 28/05/12 16:50</p> <p>Confirmation number:</p> <p>Reservations: Please go to our website www.hotelbeds.com or contact us for additional information on bookings. To make reservations, go to our website www.hotelbeds.com or contact us at floridareservations@hotelbeds.com To make bookings, go to our website www.hotelbeds.com or contact us at floridareservations@hotelbeds.com Financial Payments: Please send payment orders to canadareservations@hotelbeds.com Financial Invoices: Please send payment orders to canadareservations@hotelbeds.com. Tel: +14075564682 Fax: +14075564676</p>	<p>hotelbeds HOTELBEDS USA/ONTARIO 5402 Centre Dr., Suite 201, Orlando, FL 32819 Fax: +14075564682 Tel: +14075564676 Email: canadareservations@hotelbeds.com</p> <p>Please take note of the following: The Fairmont Royal York 100 FINE STREET WEST HQ1 J23 TORONTO - CANADA Produced on date: 18-MAY-12 18:17</p> <p>Please indicate our reference number on each of your invoices. Value of voucher is to cover reservations only as specified. Any charges incurred shall be paid by the client directly to the hotel or service provider. Please send the confirmation numbers by email to: canadareservations@hotelbeds.com</p> <p>REFERENCE: 051-122563 REEM BOOKING PASSENGER'S NAME(S): SIMES TRANSSION x 2 Fax ORIGIN MARKET: LAOS DATE: From 28-DEC-12 to 04-DEC-12 (6 NIGHTS) CONTRACT NAME: 024-HEMLOCK TYPE: SPECIAL RATE / 400 CONTRACT ROOM(S): 1 DOUBLE, TYPE: 1400PMT BONUS: ROOM ONLY 2 Adults Per Room NOTES: From 28-DEC-12 to 04-DEC-12: 1 6 nights) Free Rate at 109.04 USD ROOM/NIGHT (ENR PRICE 1 ROOM ONLY 0420) ALLOCATION From 28-DEC-12 to 04-DEC-12 Included Allotment</p> <p>Reservations: Please go to our website www.hotelbeds.com or contact us for additional information on bookings. To make reservations, go to our website www.hotelbeds.com or contact us at canadareservations@hotelbeds.com To make bookings, go to our website www.hotelbeds.com or contact us at canadareservations@hotelbeds.com Financial Payments: Please send payment orders to canadareservations@hotelbeds.com Financial Invoices: Please send payment orders to canadareservations@hotelbeds.com. Tel: +14075564682 Fax: +14075564676</p>

Tabla 37 Adjuntos de los Correos Electrónicos

El trabajo de importación configurado en **Enterprise Capture** localiza estos 3 nuevos correos electrónicos porque el remitente contiene la palabra clave configurada **fax2mail**. Comienza el procesamiento de Enterprise Capture:

1. El trabajo de Importación crea un lote con cada uno de los adjuntos
2. Captura los valores del remitente, el destinatario y el asunto para asignarlos como metadatos al lote.
3. Finaliza el trabajo de Importación y comienza el de Conversión
4. El trabajo de conversión convierte los PDF a TIFF
5. Finaliza el trabajo de Conversión y comienza el de Confirmación
6. El trabajo de Confirmación exporta los TIFFs con un nombre que sigue la norma asunto@@remitente@@destinatario@id a una ubicación compartida por el servidor Linux `wcc-vm-1-1` y el servidor Windows `wcfr-vm-m-1`

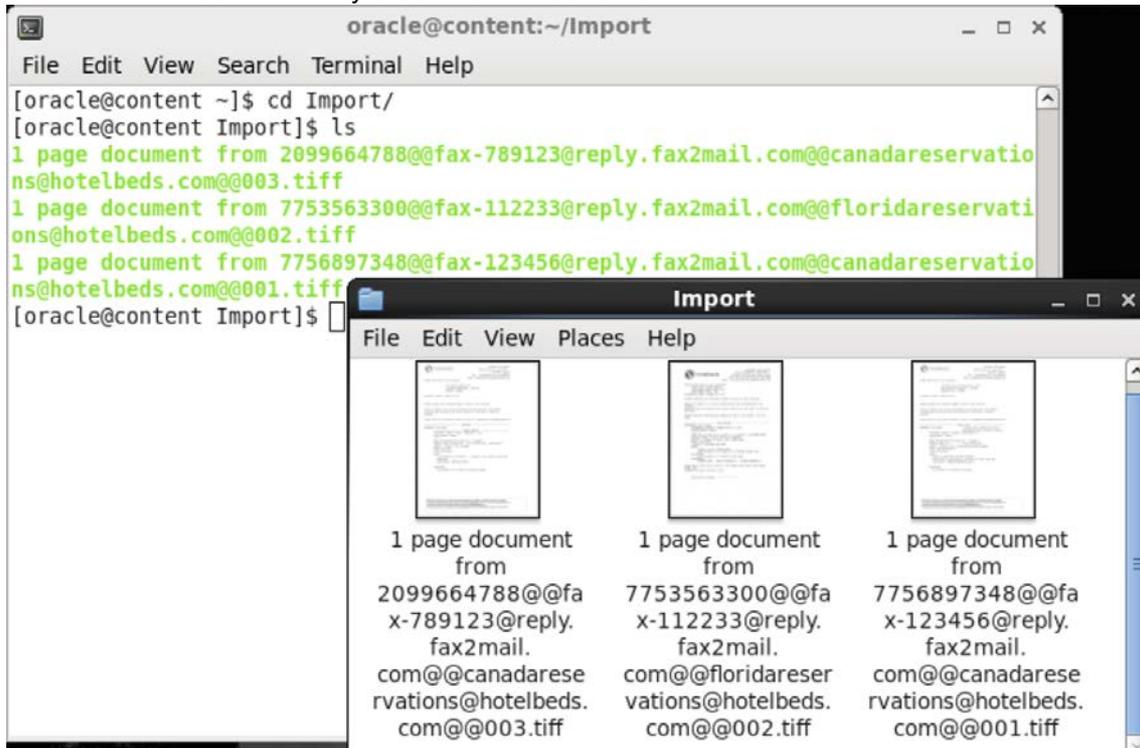


Ilustración 64 Resultados de captura de Enterprise Capture

Aquí finaliza el trabajo de Enterprise Capture y comienza el de Forms Recognition. La instancia de Runtime Service configurada para el proyecto de reconocimiento de reservas monitoriza la carpeta %FR_HOME%/Import, carpeta compartida con el sistema Linux de wcc-vm-1-1 dónde Enterprise Capture deja los 3 adjuntos capturados del correo electrónico.

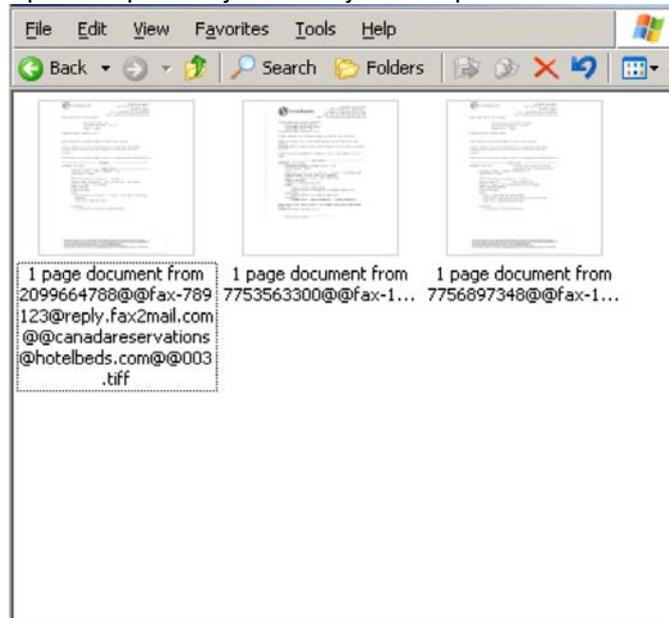


Ilustración 65 Documentos preparados para ser procesados por Forms Recognition

La instancia de Runtime Service

1. Importa los 3 documentos
2. Les aplica OCR
3. Los clasifica con la configuración establecida en el proyecto de reconocimiento
4. Extrae los campos de Fecha, Hotel y Referencia con las técnicas configuradas
5. Exporta los documentos convertidos nuevamente a PDF junto con el fichero InputBookings.csv, producto de la ejecución del script de exportación programado en el Designer que además capturaba los correos electrónicos desde el nombre del fichero, a la carpeta %FR_HOME%/Export, carpeta compartida nuevamente con el servidor Linux wcc-vm-1-1, ésta vez para que WebCenter Imaging pueda utilizar dicho fichero CSV para importar los documentos con los resultados de clasificación y extracción automatizada.

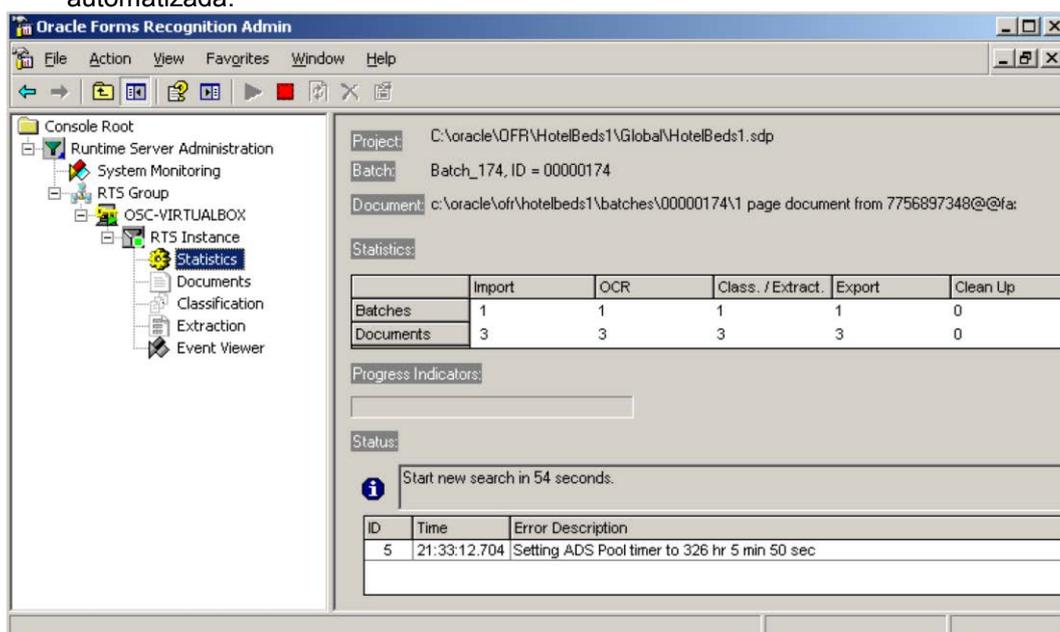


Ilustración 66 Runtime Service en Ejecución, procesando los 3 documentos

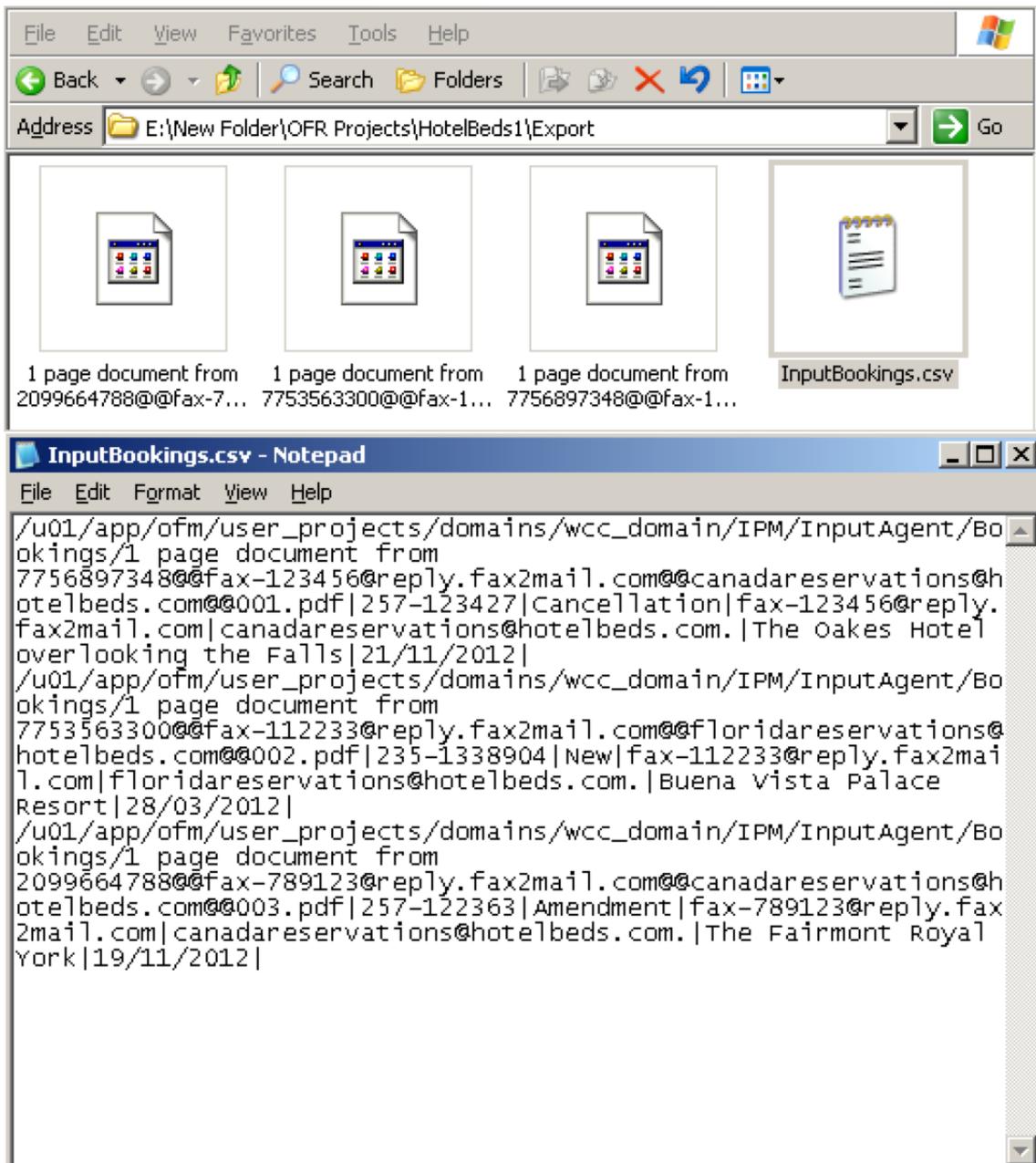


Ilustración 67 Resultados de exportación de Forms Recognition

Finalmente, el Input Agent de WebCenter Imaging descubre que existe un fichero con la máscara InputBooking.csv en la ubicación que está monitorizando \$DOMAIN_HOME/wcc_domain/IPM/InputAgent/Bookings, por lo que lo lee e importa los ficheros ubicados en el primer campo del fichero csv y les asocia los siguientes campos como metadatos, de acuerdo al mapeo configurado previamente.

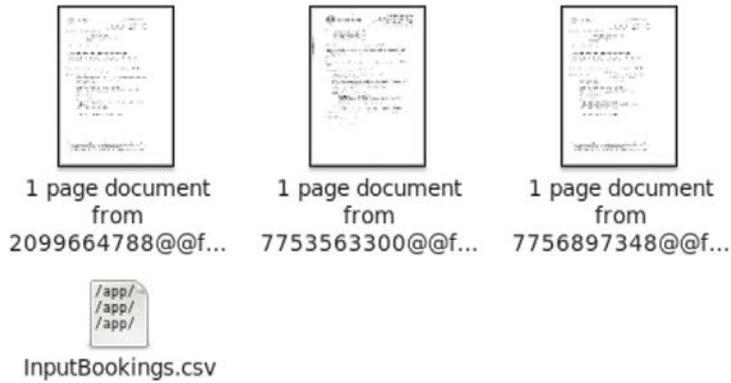


Ilustración 68 Documentos listos para ser cargados por el Input Agent de Imaging

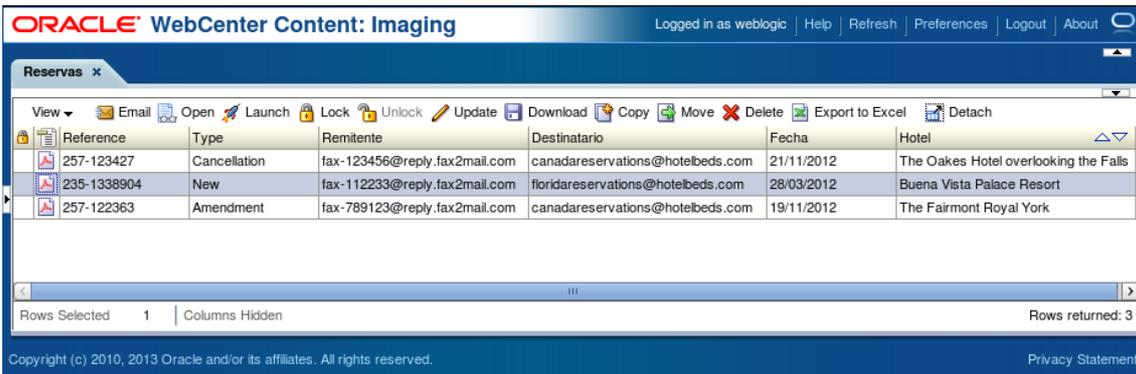


Ilustración 69 Los 3 documentos correctamente importados en WebCenter Imaging

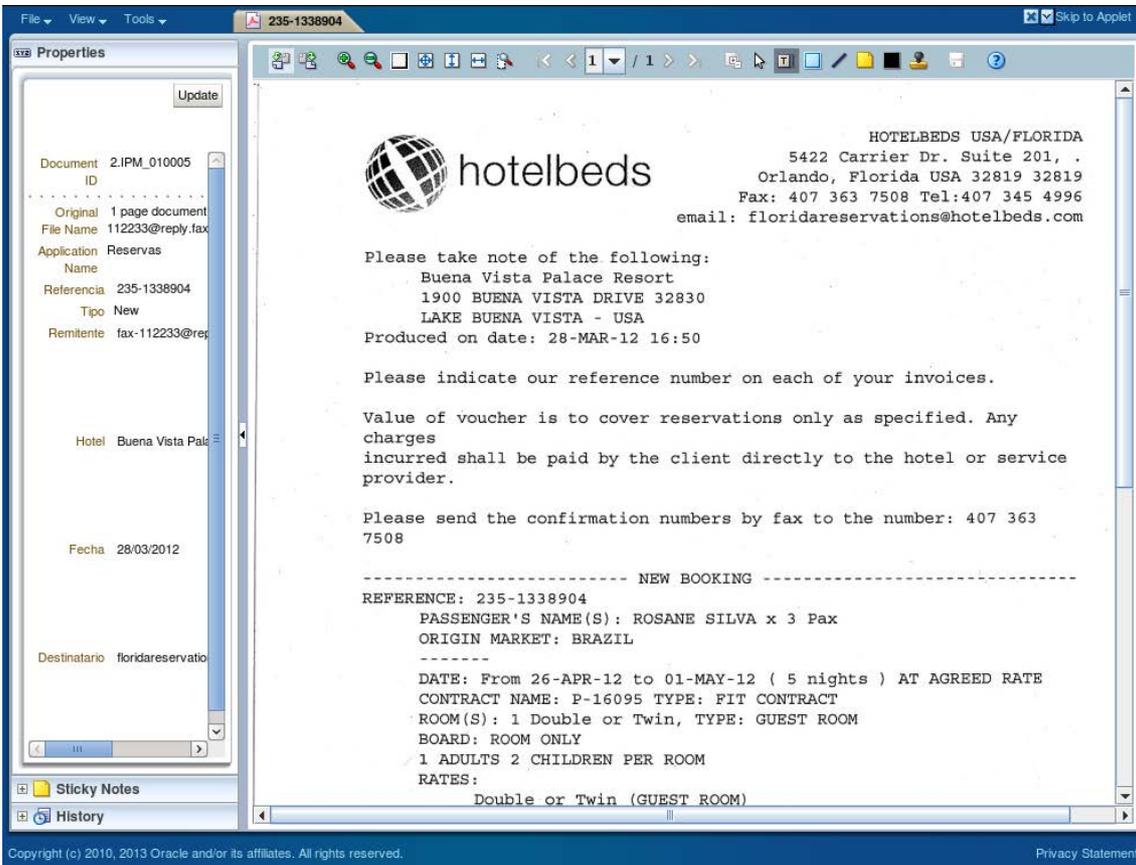


Ilustración 70 Una de las reservas vista con el Visor de Imaging

La entrada de adjuntos a Imaging provoca el disparo de instancias del proceso BPM de Confirmación de Reservas. Este proceso ofrece a los operarios y empleados del Banco de Camas una aplicación con la que revisar la extracción automatizada, corregirla si fuera necesario y, finalmente, confirmar los cambios en el sistema.

En resumen, ahora los operarios únicamente deben comprobar en una interfaz si los datos extraídos son correctos y confirmar o corregir, pudiendo aumentar su ancho de banda en la gestión de reservas e invertir más tiempo en otras tareas de mayor valor.

8. Costes y Gestión del Proyecto

Para la implantación del Sistema Gestor de Documentos Electrónicos para la Automatización de la Gestión de Reservas hay que contabilizar los costes de servicios de instalación, configuración y parametrización además de las licencias del software de Oracle. Los costes de los servidores físicos así como la tecnología de virtualización no serán tenidos en cuenta puesto que son marginales y se asume que la organización que vaya a implantar este proyecto ya cuenta con una infraestructura equivalente a la presentada en esta memoria, ya sea virtualizada o completamente física.

8.1. Licenciamiento

El licenciamiento del software de Oracle se calcula de la siguiente manera[44]:

$$Total_{Licencias} = Precio_{Procesador} * Procesadores * Núcleos * Factor_{Correccion}$$

De donde el Precio por Procesador se obtiene de la lista de precios pública [45] y el Factor de Corrección depende de la arquitectura del procesador. En el caso de Intel x86 es 0.5 [46]. La lista de precios está en dólares por lo que hay que hacer una conversión a euros. El cambio actual está aproximadamente a 1\$ = 0.92€. Además, todo el software de Oracle está sujeto a un contrato de soporte que supone el 22% del precio al que se adquieran las licencias.

Producto	Precio por procesador	Nº procesadores y núcleos	Total Licencias	Soporte anual
Productos WebCenter (Captura, Reconocimiento y Gestión Documental)				
Enterprise Capture	44.160,00 €	2 Procesadores de 4 Núcleos	176.640,00 €	38.860,80 €
Forms Recognition	73.600,00 €	2 Procesadores de 2 Núcleos	147.200,00 €	32.386,00 €
WebCenter Imaging	67.712,00 €	2 Procesadores de 4 Núcleos	270.848,00 €	59.586,56 €
WebLogic Enterprise Edition ¹³	18.400,00 €	2 Procesadores de 4 Núcleos	73600,00 €	16.192,00 €
Subtotal WebCenter			668.288,00 €	147.025,36 €
Productos BPM Suite (Gestión de Procesos)				
WebLogic Suite ¹⁴	33.120,00 €	2 Procesadores de 8 Núcleos	264.960,00 €	58.291,20 €
SOA Suite ¹⁵	42.320,00 €	2 Procesadores de 8 Núcleos	338.560,00 €	74.483,20 €
BPM Suite	42.320,00 €	2 Procesadores de 8 Núcleos	338.560,00 €	74.483,20 €
Subtotal BPM Suite			942.080,00 €	207.257,60 €
Subtotal			1.610.368,00 €	354.282,96 €
Total (21% IVA)			1.948.545,28 €	428.682,38 €

Tabla 38 Costes de las licencias requeridas para Proyecto

¹³ WebLogic Enterprise Edition es un prerrequisito para poder ejecutar Enterprise Capture en clúster. WebCenter Imaging incluye un uso restringido de WebLogic Enterprise Edition al despliegue de WebCenter Imaging

¹⁴ WebLogic Suite es un prerrequisito para SOA Suite

¹⁵ SOA Suite es un prerrequisito para BPM Suite

8.2. Plan de proyecto

El proyecto se plantea en 4 fases principales:

1. Planteamiento
2. Instalación
3. Parametrización
4. Pruebas.

En el siguiente diagrama de Gantt se observa la división de tareas en cada una de estas fases, indicando en rojo las tareas a realizar por el consultor de WebCenter, en amarillo las tareas a realizar por el consultor de BPM y en naranja las tareas que ambos realizarán en conjunto.



Ilustración 71 Plan de Proyecto

Cada línea equivale a una jornada de trabajo. En base a este plan, se estimarían los siguientes costes de implantación del proyecto.

Consultor	Precio por jornada	Planteamiento	Instalación	Parametrización	Pruebas	Jornadas	Total
WebCenter	920 €	10	4	6	12	32	29.440,00€
BPM	920 €	10	2	3	12	27	24.840,00€
Subtotal							54.280,00€
Total (21% IVA)							65.678,80€

Tabla 39 Costes de servicios de implantación del Proyecto

Por tanto, el coste total del proyecto es de **2.442.906,46 €**, incluyendo el precio del primer año de soporte.

9. Referencias

- [1] Vilà J., Director General de TUI Travel A&D y Presidente de HotelBeds (2012) *Hotelbeds / TUI Travel: The global market for accommodation wholesalers will grow 7.3% by 2015, with Asia-Pacific and Americas being the fastest growing segments* <<http://www.traveldailynews.asia/news/article/49464/the-global-market-for-accommodation>> [Consulta: 1 de Enero de 2015]
- [2] De la Rosa, J.M. (2013) *Mayoristas de estancias, un modelo en cuestión* <http://www.hosteltur.com/127170_mayoristas-estancias-modelo-cuestion.html> [Consulta: 1 de Enero de 2015]
- [3] De la Rosa, J.M. (2013) *Opentours: "Las mayoristas son imprescindibles para hoteles y agencias"* <http://www.hosteltur.com/127430_opentours-mayoristas-son-imprescindibles-hoteles-agencias.html> [Consulta: 1 de Enero de 2015]
- [4] Maydanchik A. (2007) *Manual Data Entry en Data Quality Assesment* p. 11
- [5] Scott, J.R., Thompson, A.R., Wright-Thomas, D., Xu, X. & Barchard, K.A. (April, 2008) *Data entry methods: Is double entry the way to go?* Poster presented at the Western Psychological Association Annual Convention, Irvine, CA.
- [6] Sonderegger, D. *Top 10 Tax and Accounting Mistakes Cost Companies Billions* <<http://www2.cfo.com/accounting-tax/2015/03/top-10-tax-accounting-mistakes-cost-companies-billions/>> [Consulta: 22 de Julio de 2015]
- [7] AIIM *Executive Leadership Council* <<http://www.aiim.org/Membership/Executive-Leadership-Council>> [Consulta: 1 de Febrero de 2015]
- [8] About the AIIM <<http://www.aiim.org/About>> [Consulta: 1 de Febrero de 2015]
- [9] AIIM (2009) *Recommended Practice, Analysis, Selection and Implementation of Electronic Document Management Systems (EDMS) p.12-18*
- [10] OASIS *About Us* <<https://www.oasis-open.org/org>> [Consulta: 1 de Febrero de 2015]
- [11] OASIS *Members* <<https://www.oasis-open.org/member-roster>> [Consulta: 2 de Febrero de 2015]
- [12] OASIS (2007) *Web Services Business Process Execution Language Version 2.0*
- [13] The Object Management Group *About OMG* <<http://www.omg.org/gettingstarted/gettingsstartedindex.htm>> [Consulta: 2 de Febrero de 2015]
- [14] The Object Management Group *Search for Member Companies* <<http://www.omg.org/cgi-bin/apps/membersearch.pl>> [Consulta: 2 de Febrero de 2015]
- [15] The Object Management Group *OMG Specifications* <<http://www.omg.org/spec/>> [Consulta: 2 de Febrero de 2015]
- [16] Strategy Partners International Ltd (2003) *Document and Content Capture* p. 3-4
- [17] Gartner Inc. (2014) *Magic Quadrant for Enterprise Content Management*
- [18] Gartner Inc. (2014) *Magic Quadrant for Intelligent Business Process Management Suites*
- [19] Alfresco Software *Acerca de Alfresco Simple + Smart* <<https://www.alfresco.com/es/acerca-de-alfresco>> [Consulta: 16 de Febrero de 2015]
- [20] Alfresco Software *¿Cuál es la diferencia entre Community y Enterprise?* <<https://www.alfresco.com/es/alfresco-community-edition>> [Consulta: 16 de Febrero de 2015]
- [21] Ephesoft *Ephesoft Enterprise Edition* <<https://www.alfresco.com/es/node/1978>> [Consulta: 16 de Febrero de 2015]
- [22] IBM (2011) *Migration Existing Entitlements to IBM Business Manager* <<http://www-01.ibm.com/software/integration/business-process-manager/library/entitlement-migration/>> [Consulta: 16 de Febrero de 2015]
- [23] Lexmark (2010) *Lexmark announces definitive agreement to acquire Perceptive Software* <<http://newsroom.lexmark.com/index.php?s=13630&item=23833>> [Consulta: 16 de Febrero de 2015]
- [24] Lexmark (2012) *Lexmark acquires Brainware* <<http://newsroom.lexmark.com/index.php?s=13630&item=122999>> [Consulta: 16 de Febrero de 2015]
- [25] Lexmark (2012) *Perceptive Software Extends Support for Embedded Brainware Intelligent Data Capture Capabilities through Oracle WebCenter* <<http://www.perceptivesoftware.com/company/news/news-announcements/2012/perceptive-software-extends-support-for-embedded-brainware->

- intelligent-data-capture-capabilities-through-oracle-webcenter.html*> [Consulta: 16 de Febrero de 2015]
- [26] Lexmark (2014) *Lexmark announces results of ReadSoft tender offer* <<http://newsroom.lexmark.com/2014-09-08-Lexmark-announces-results-of-ReadSoft-tender-offer>> [Consulta: 16 de Febrero de 2015]
- [27] Lexmark (2015) *Lexmark to acquire Kofax* <<http://newsroom.lexmark.com/2015-03-24-Lexmark-to-acquire-Kofax>> [Consulta: 16 de Febrero de 2015]
- [28] Dr. Kampffmeyer U. (2013) *Enterprise Information Management*
- [29] SAP *Evaluating a Single-Vendor Versus Multivendor Approach to Modern Data Management* p. 3
- [30] Arnuphaptrairong T. *Top Ten Lists of Software Project Risks*
- [31] Cuartero J., Roux N., Smith M., Tahiri A. *Vertically Integrated Stacks*
- [32] Oracle Corporation *Fusion Middleware Products* <<https://www.oracle.com/middleware/products/index.html>> [Consulta: 20 de Febrero de 2015]
- [33] Oracle Corporation (2014) 1.3. *Understanding Middleware Architecture Design* en *Understanding Oracle Fusion Middleware Concepts*
- [34] Crimson Consulting (2011) *Oracle WebLogic Server® Costs versus JBoss® Application Server*
- [35] Pique Solutions (2013) *Application Server Cost of Ownership*
- [36] SPEC (2015) *Second Quarter 2015 SPECjEnterprise2010 Results* <<https://www.spec.org/jEnterprise2010/results/res2015q2/>> [Consulta: 4 de Junio de 2015]
- [37] SPEC (2015) *First Quarter 2015 SPECjEnterprise2010 Results* <<https://www.spec.org/jEnterprise2010/results/res2015q1/>> [Consulta: 4 de Junio de 2015]
- [38] Oracle Corporation (2015) 1.1 *Overview of the Enterprise Deployment Reference Topology* en *Enterprise Deployment Guide for Oracle WebCenter Content*
- [39] Oracle Corporation (2015) *Managing Oracle WebCenter Enterprise Capture*
- [40] Oracle Corporation (2015) 3 *Managing Oracle WebCenter Content: Imaging* en *Understanding Oracle WebCenter Content*
- [41] Oracle Corporation (2015) *Installing and Configuring Oracle SOA Suite and Business Process Management*
- [42] Oracle Corporation (2015) *Developing Business Process with Oracle Business Process Composer*
- [43] Oracle Corporation (2013) *Oracle WebCenter Forms Recognition Installation Guide*
- [44] Oracle Corporation (2014) *Software Investment Guide*
- [45] Oracle Corporation (2015) *Oracle Technology Global Price List*
- [46] Oracle Corporation (2009) *Oracle Processor Core Factor Table*