



Universidad
Carlos III de Madrid
www.uc3m.es

Data Warehouse Marco de Calidad

Autor: Pablo Martín Gutiérrez
Tutor: Antonio García Carmona

Defensa del Proyecto de Final de Carrera

Índice

- Introducción.
- Objetivos del PFC.
- Fundamentos
 - ▣ Business Intelligence
 - ▣ Data Warehouse.
 - ▣ ETL.
- Calidad.
 - ▣ Calidad en los procesos
 - ▣ Calidad del dato
 - ▣ Marco de Calidad
- Herramientas.
 - ▣ Aplicación práctica
 - ▣ Demo.



Introducción



➤ Business Intelligence (BI)

- Técnicas que permiten extraer conocimiento del negocio.
 - Sobre el qué y cómo pasa.
 - El porqué pasa.
- Para facilitar la **toma de decisiones**.

➤ Situación Actual

- Desde el estallido de la **crisis económica mundial**:
 - **lidiar con la incertidumbre**
 - **miedos de no poder controlar los procesos de nuestra organización.**

➤ Estado del arte

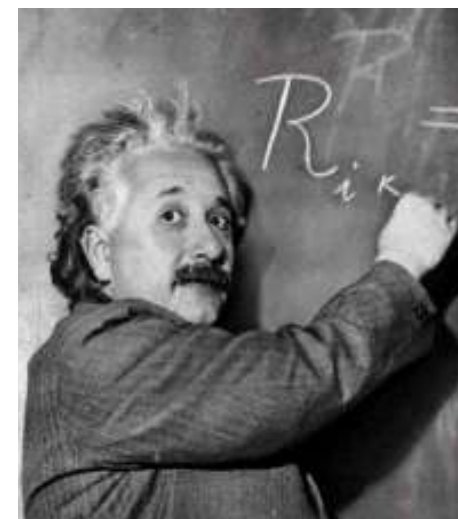
- **La documentación** asociada a sus componentes de desarrollo :
 - **No es muy completa**
 - Idioma (Castellano muy escasa).
 - **No se integra con facilidad** con los estándares, normas y guías de buenas prácticas



Objetivos del PFC

➤ En lo teórico

- ❑ Crear un **Marco de Calidad** para *el desarrollo de Data Warehouse*..
- ❑ **Integración** en el mismo de la **Calidad en el Dato y en el Proceso**.
- ❑ **idea clara y precisa** sobre qué es el **Business Intelligence**.
- ❑ **Principales componentes**
- ❑ **Principales metodologías de desarrollo** de un Data Warehouse.
- ❑ **Principales normas de calidad, estándares y guías de buenas prácticas** que se pueden aplicar.



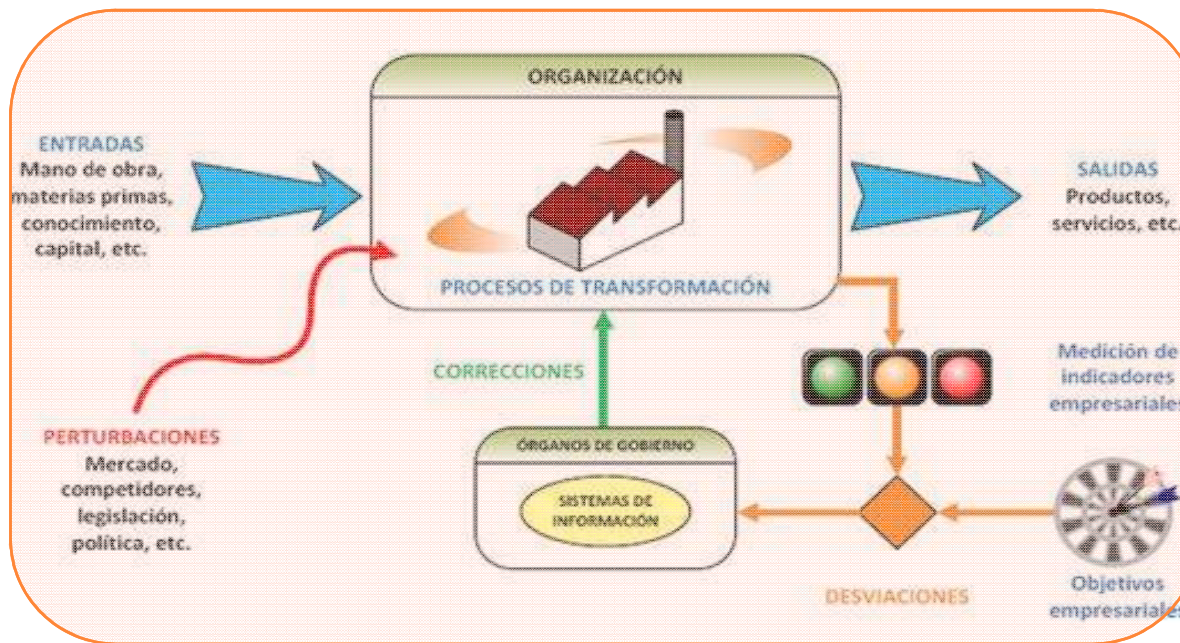
➤ En lo práctico

- ❑ **Acercar el modelo teórico a la realidad** a través del estudio de dos **herramientas** muy útiles para la implementación del principal componente del desarrollo de un DW, el proces de Extracción, Transformación y Carga (ETL).

Fundamentos (1 de 4)

Business Intelligence

- *Una organización se debe entender como un sistema ...*



- *Si la información es comprensible, y llega en el tiempo y formato adecuado, ésta permite a la empresa reducir su incertidumbre y tomar mejores decisiones con el objetivo de aportarle una ventaja competitiva.*

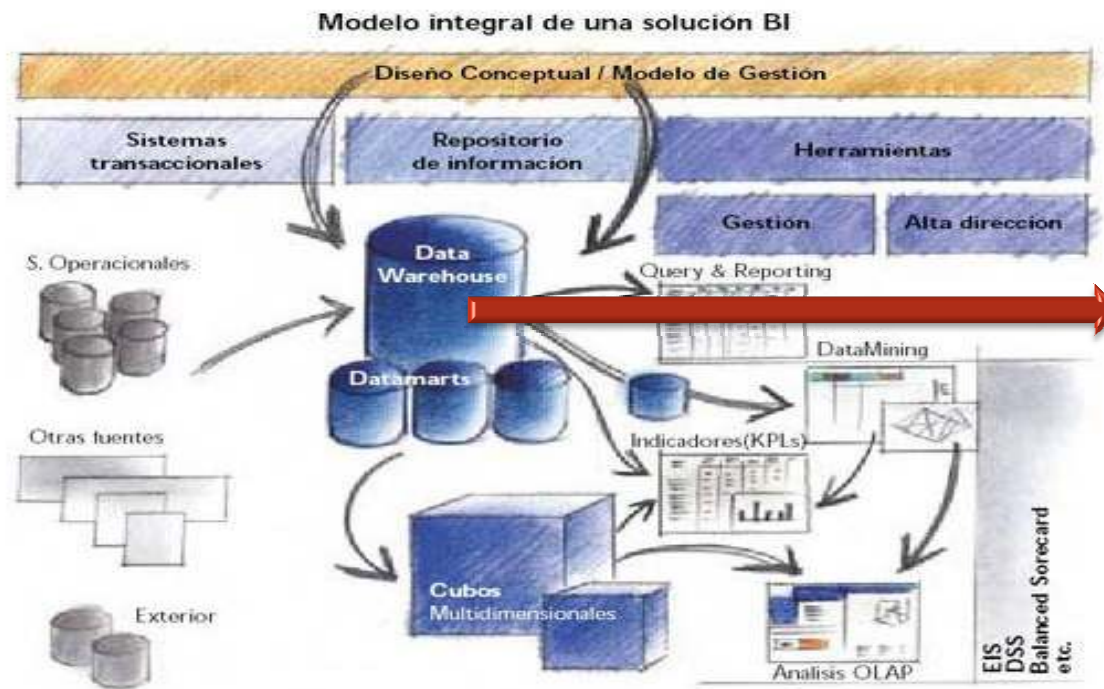
- **Beneficios:**

- Ahorro de costes.
- Acceso a la información.
- Apoyo en la toma de decisiones.
- Orientación al usuario final.
- Mayor rentabilidad.
- Mayor agilidad empresarial.
- Menores costes operativos.
- Mayor fidelidad de los clientes.
- Optimización de la adquisición de clientes.
- Interacción directa entre los usuarios.

Fundamentos (2 de 4)

Business Intelligence y Data Warehouse

- Pero ... ¿Realmente, **como te ayudan los sistemas de BI a tomar decisiones?**
 1. **Recopilando datos** sobre los que trabajar.
 2. **Dotar a los analistas de negocio de herramientas** para explotarlos de forma sencilla, útil y rápida.



□ **Data Warehouse**

- Componente **más importante** solución de **BI**
- **Depósito de datos.**
- **Integra todos** los datos que se generan en **los SI de la empresa.**
- **Datos de carácter histórico.**
- **La información** que contiene se enfoca **para generar informes en tiempos óptimos.**

Fundamentos (3 de 4)

Arquitectura DW

- “Un Data Warehouse no se compra, se construye”
- Arquitectura de un DW:

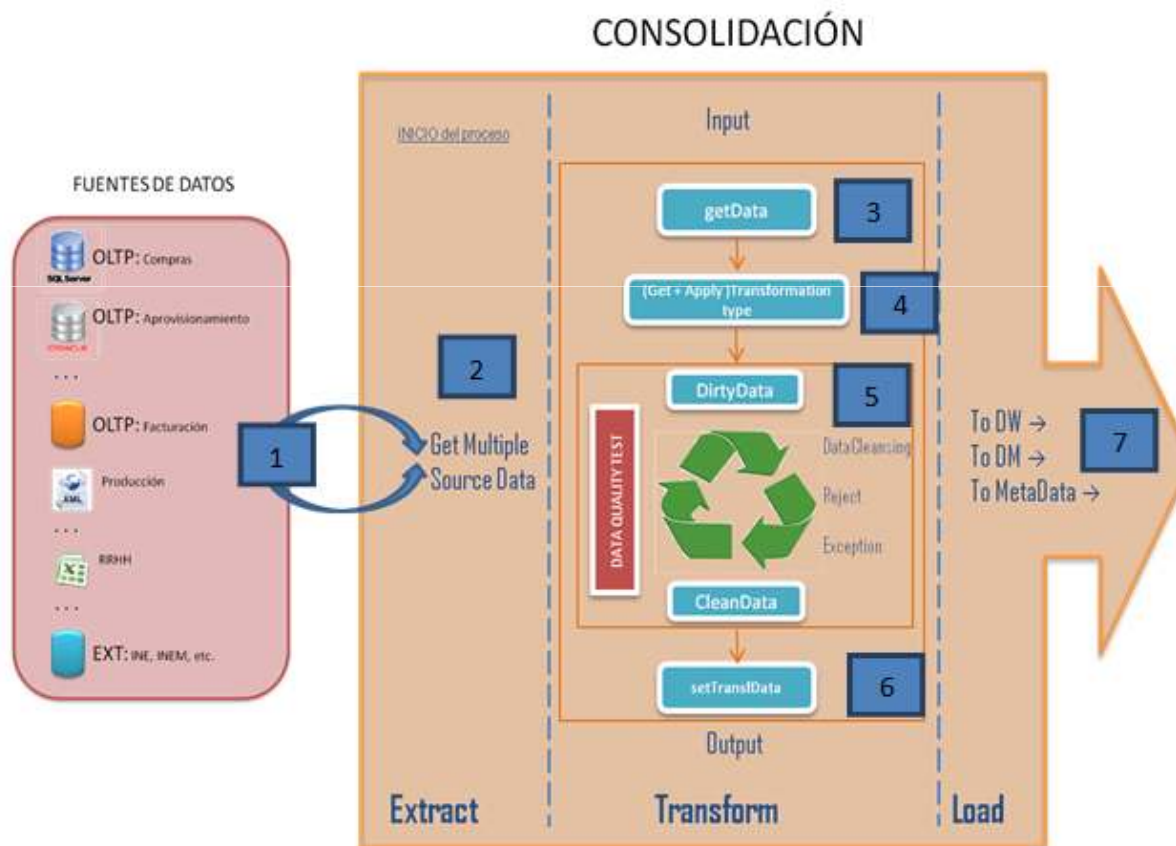


1. **Fuentes de datos:** Es la información que se recoge de los SI que participan en la empresa.
2. **Consolidación:** Es el proceso que integra y transforma los datos para insertarse en el DW.
 1. Es el proceso más importante a la hora de formar el DW → Incluye el proceso ETL.
3. **Almacenamiento:** Recoge la información consolidada físicamente en la BBDD que forma el DW.
 1. Es la persistencia del DW.
4. **Acceso:** Es SW de conectividad (Middleware) para las herramientas de explotación.
5. **Explotación:** Son las herramientas que explotan el DW para obtener información útil para la toma de decisiones.

Fundamentos (4 de 4)

ETL: Extract, Transform and Load.

- **ETL: Proceso de consolidación de las fuentes de datos en el Data Warehouse.**

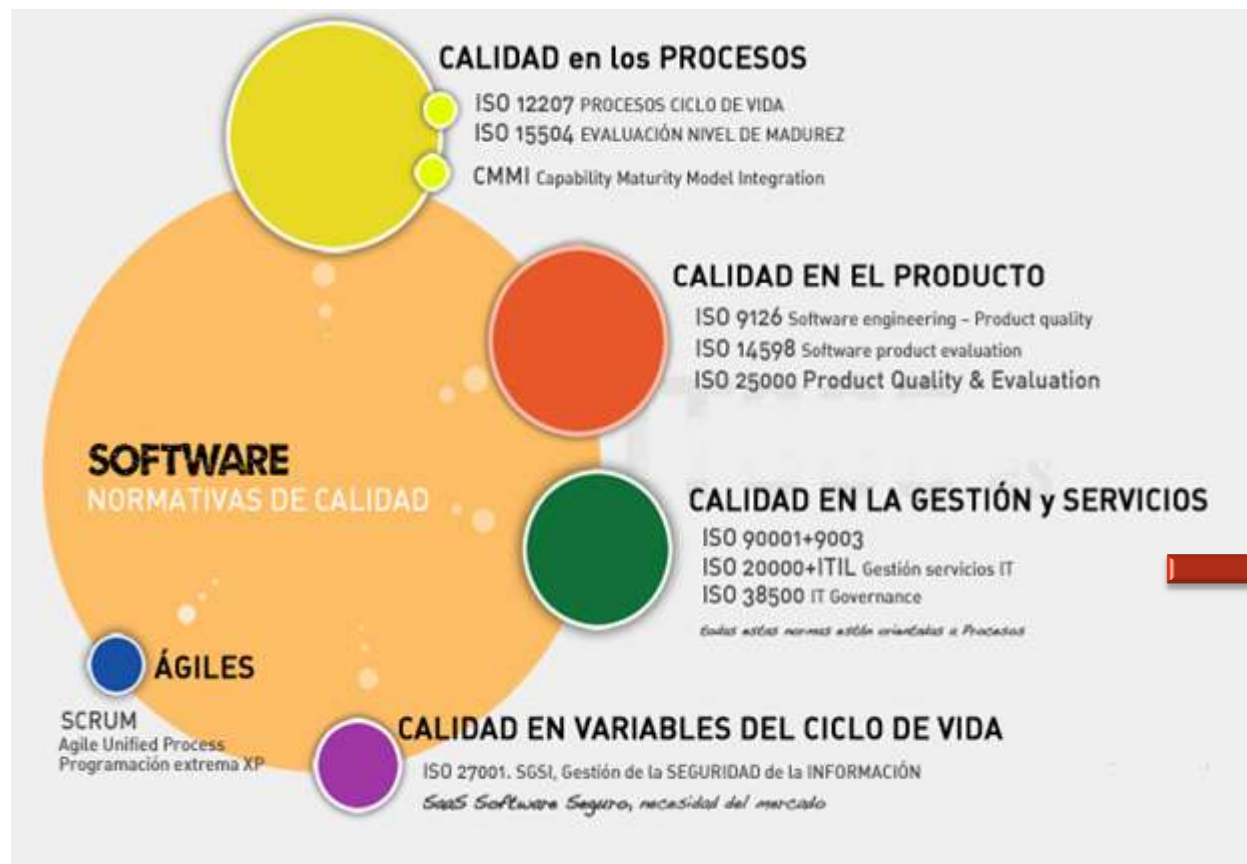


- **(1) Interoperabilidad**
 - Comunicación con las fuentes de datos.
 - Seguridad
- **(2) Extracción → Extract.**
 - De las fuentes de datos a la base de datos intermedia.
- **(3 - 6) Transformación:**
 - Recuperar el dato.
 - Aplicar transformaciones
 - Ciclo de calidad y limpieza del dato
 - Escritura del dato transformado.
- **(7) Carga → Load**
 - A la base de datos destino.

Calidad (1 de 3)

En los Procesos

- **Estándares, Normas de Calidad y Guías de Buenas Prácticas.**



Calidad (2 de 3)

De los Datos

- ❑ ¿Qué propiedades debe tener un dato de calidad?
- ❑ Estudio del estado del arte.
 - ❑ ISO 25012
 - ❑ Wang y Strong.
 - ❑ Leo L. Pipino et. al.
 - ❑ DWQ Project
 - ❑ Rudra y Yeo
 - ❑ Leithisier R.
- ❑ Estudio depurado de las características comunes
- ❑ Unificación de criterios y definiciones
- ❑ Selección



❑ *Características de la calidad del dato*

- ❑ ACCESIBILIDAD.
- ❑ ACCESO SEGURO.
- ❑ CANTIDAD APROPIADA DE DATOS.
- ❑ COMPLETITUD.
- ❑ CONFIDENCIALIDAD.
- ❑ CONFORMIDAD.
- ❑ CONSISTENCIA.
- ❑ CREDIBILIDAD.
- ❑ DISPONIBILIDAD.
- ❑ EFICIENCIA.
- ❑ ENTENDIBILIDAD.
- ❑ EXACTITUD.
- ❑ FACILIDAD DE MANIPULACIÓN.
- ❑ GRUPO FORMADO POR LA PROFUNDIDAD, ACTUALIDAD Y VOLATILIDAD.
- ❑ INTERPRETABILIDAD.
- ❑ LEGALIDAD.
- ❑ OBJETIVIDAD.
- ❑ PORTABILIDAD.
- ❑ PRECISIÓN.
- ❑ RECUPERABILIDAD.
- ❑ RELEVANCIA.
- ❑ REPRESENTACIÓN CONCISA.
- ❑ REPRESENTACIÓN CONSISTENTE.
- ❑ REPUTACIÓN.
- ❑ TRAZABILIDAD.
- ❑ VALOR AÑADIDO.

Calidad (3 de 3)

Marco de Calidad

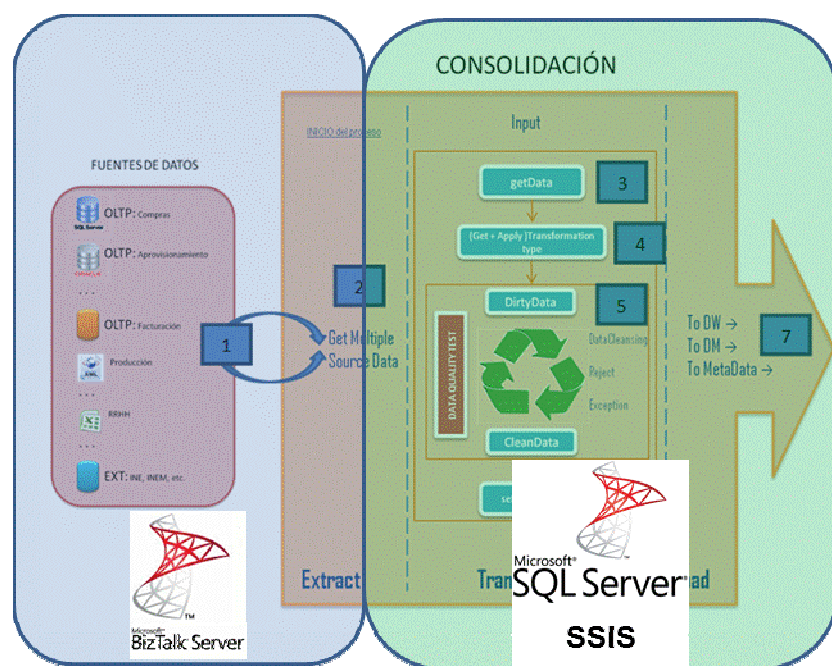


Características de calidad del dato.

- Se han establecido **26 características** relacionadas con la calidad en los datos.
- Se debe elegir una **métrica** para la evaluación de la calidad del dato. (ISO 2504n)
- Se debe establecer un conjunto de **propiedades** que medir, para cada dato.
- **Evaluación:** Se deben especificar unos niveles de calidad.
- **Mejora** continua.

Herramientas (1 de 2)

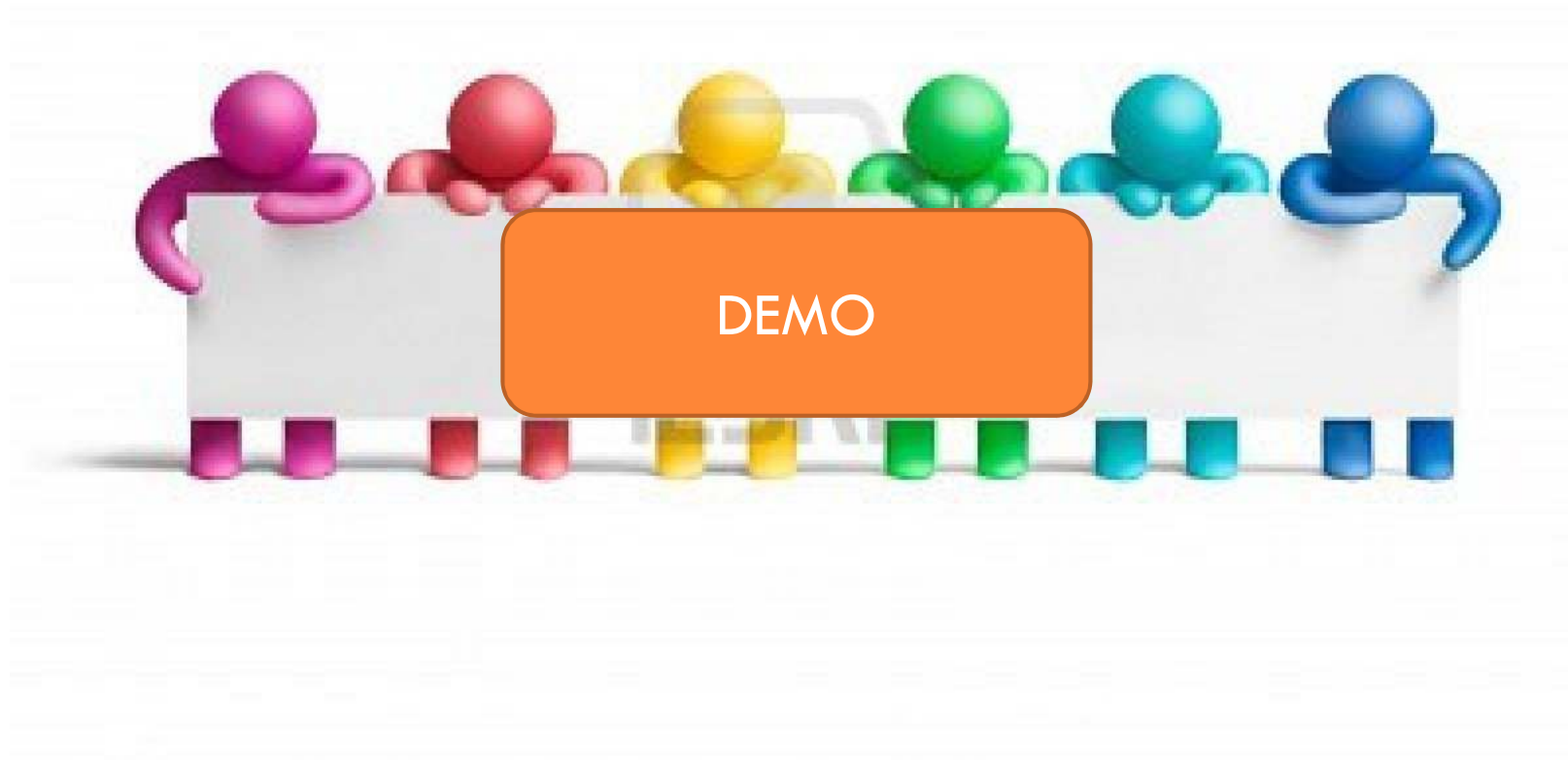
- Para aproximar el proyecto al mundo real:
 1. Para la interoperabilidad → *Microsoft Biztalk Server 2010*.
 2. Para el ETL → *SQL server Integration Services*.



- **Microsoft BizTalk Server** es un servidor Gestión de procesos de negocio (BPM).
 - Usa adaptadores diseñados para comunicarse con diferentes tipos de software
 - Los más usados en el entorno empresarial.
- **SSIS** son las siglas en inglés de **SQL Server Integration Services**.
 - Herramienta diseñada para simplificar los proyectos de ETL.
 - Reducir al mínimo las necesidades de intervención manual.

Herramientas (2 de 2)

- A continuación mostraremos un pequeño ejemplo de como podemos recuperar datos de varias fuentes e integrarlos en nuestro DW gracias a SSIS.





Universidad
Carlos III de Madrid
www.uc3m.es

Data Warehouse Marco de Calidad

Autor: Pablo Martín Gutiérrez

Tutor: Antonio García Carmona

Defensa del Proyecto de Final de Carrera