

UNIDAD 1

Introducción a Datawarehouse

Introducción a DataWarehouse

Contenido de la Unidad

- **Conceptos, motivación y marco histórico.**
- Utilidad y funciones.
- Generación y construcción de DW: ETL
- Tipos de DataWarehouse.
- Comparativa con Bases de Datos Relacionales.
- Ejemplos prácticos / Casos de Éxito.
- Resumen.

Introducción a DataWarehouse

Expectativas

Luego de esta unidad podremos:

- Entender de qué manera las organizaciones satisfacen sus necesidades de información.
- Reconocer los problemas típicos que las organizaciones modernas enfrentan al trabajar con datos.
- Entender definiciones y conceptos básicos de Data Warehousing como solución.
- Diferenciar entre una arquitectura frecuente de una mejorada.
- Identificarán distintos pasos metodológicos para la construcción de un DW

Para hablar de lo mismo...

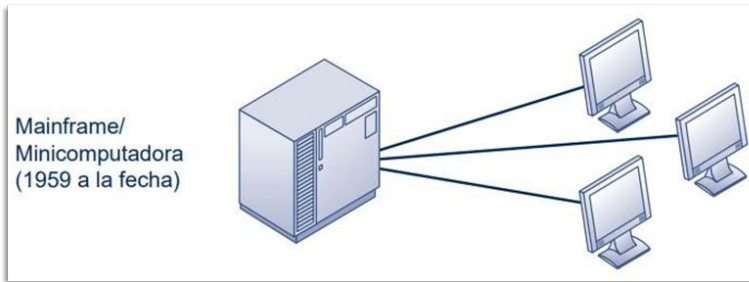
- **Datos:** Son las **partículas de conocimiento** que le llegan al **observador** desde la **realidad** que está sucediendo. Los datos son el reflejo de una realidad que ya sucedió (instancias). Jaime Cabrera, 1994.
- **Información:** Es el conjunto de datos que interesan a un **observador en particular** por un **motivo específico** que él conoce, para un **fin determinado**, en una **circunstancia dada**, en un **momento especial** y que **tiene sentido** para él. JC, 1994
- **Conocimiento:** es una mezcla de **experiencia, valores, información y “know-how”** que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Davenport y Prusak, 1999



El Conocimiento se deriva de la Información como ésta de los datos. Su transformación se produce gracias a la Comparación, Predicción, Conexiones e Interacción.

Introducción a DataWarehouse

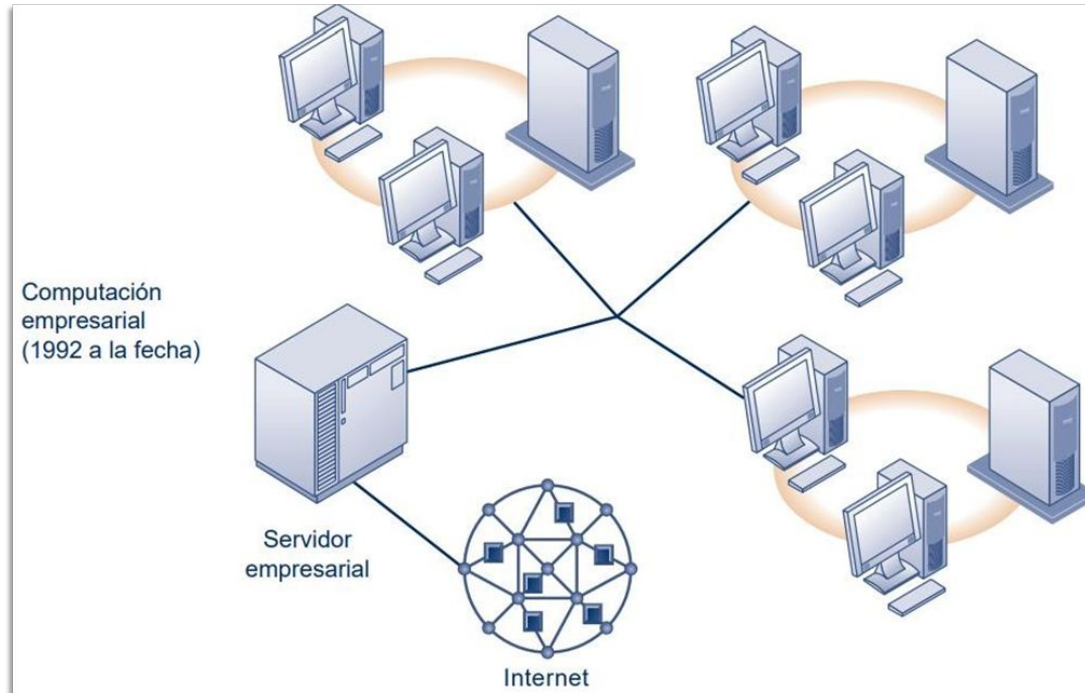
Evolución de Plataformas de TI



Fuente: Laudon y Laudon, 14th ed.

Introducción a DataWarehouse

Evolución de Plataformas de TI



Introducción a DataWarehouse

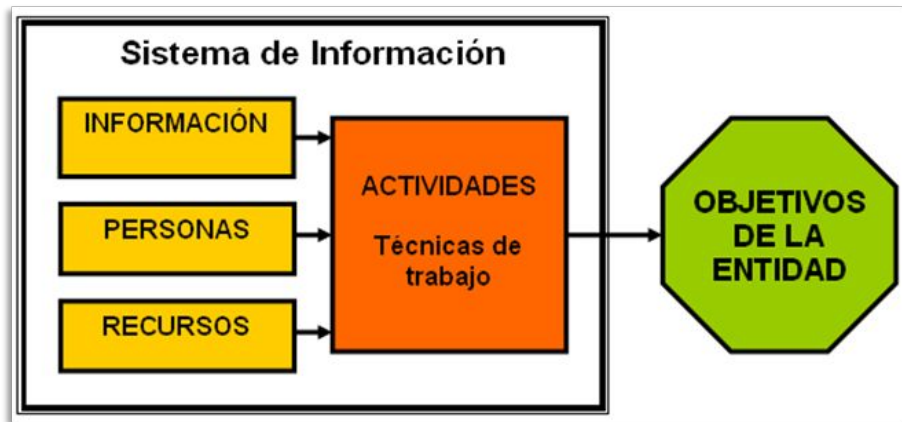
Evolución de Plataformas de TI



Introducción a DataWarehouse

Sistemas de Información

Sistemas de Información es un conjunto de elementos orientados al tratamiento y administración de Datos e Información, con el fin de colaborar con los Objetivos de una Organización o Empresa.

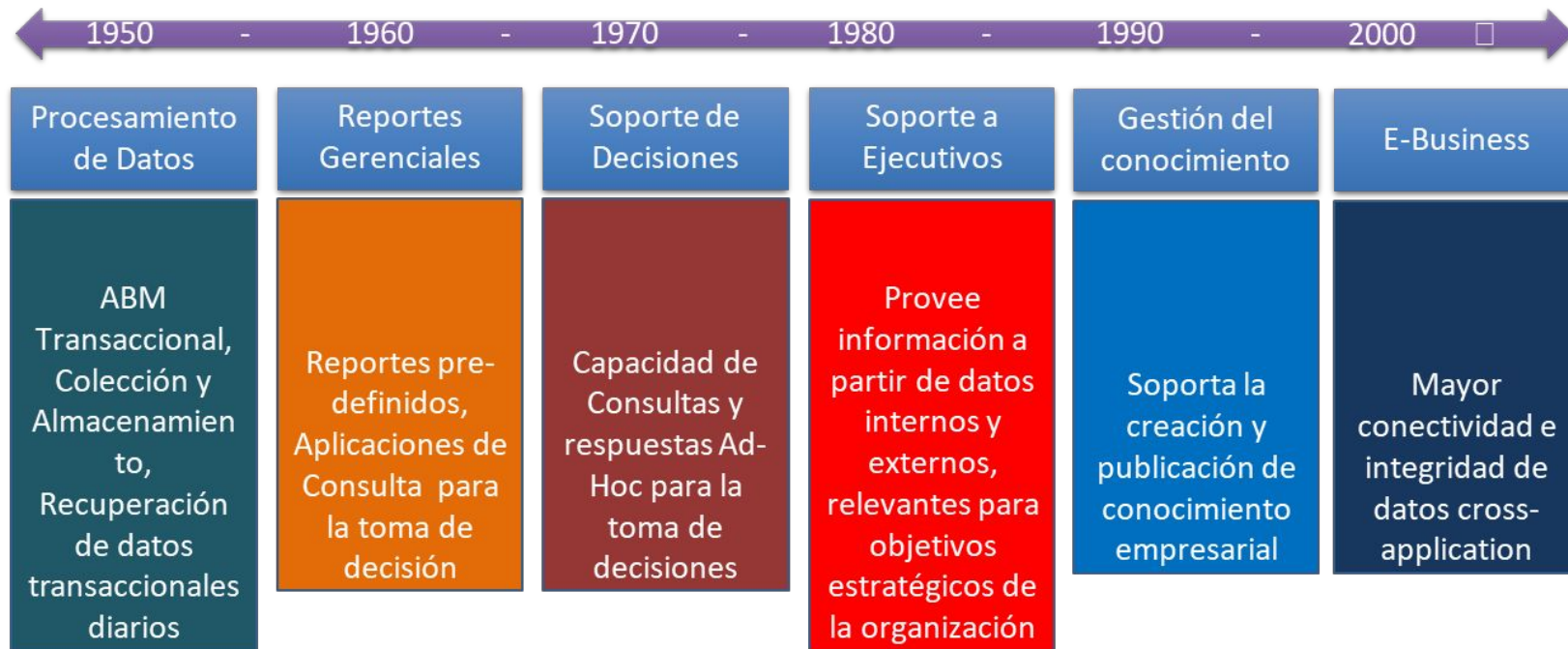


En Informática, estas Actividades pueden ser: Obtener, Transformar, Limpiar, Manipular, Almacenar, Administrar, Proteger, Distribuir, Publicar, Intercambiar, Transmitir, Datos e Información

Fuente: P. Vaish - 2017

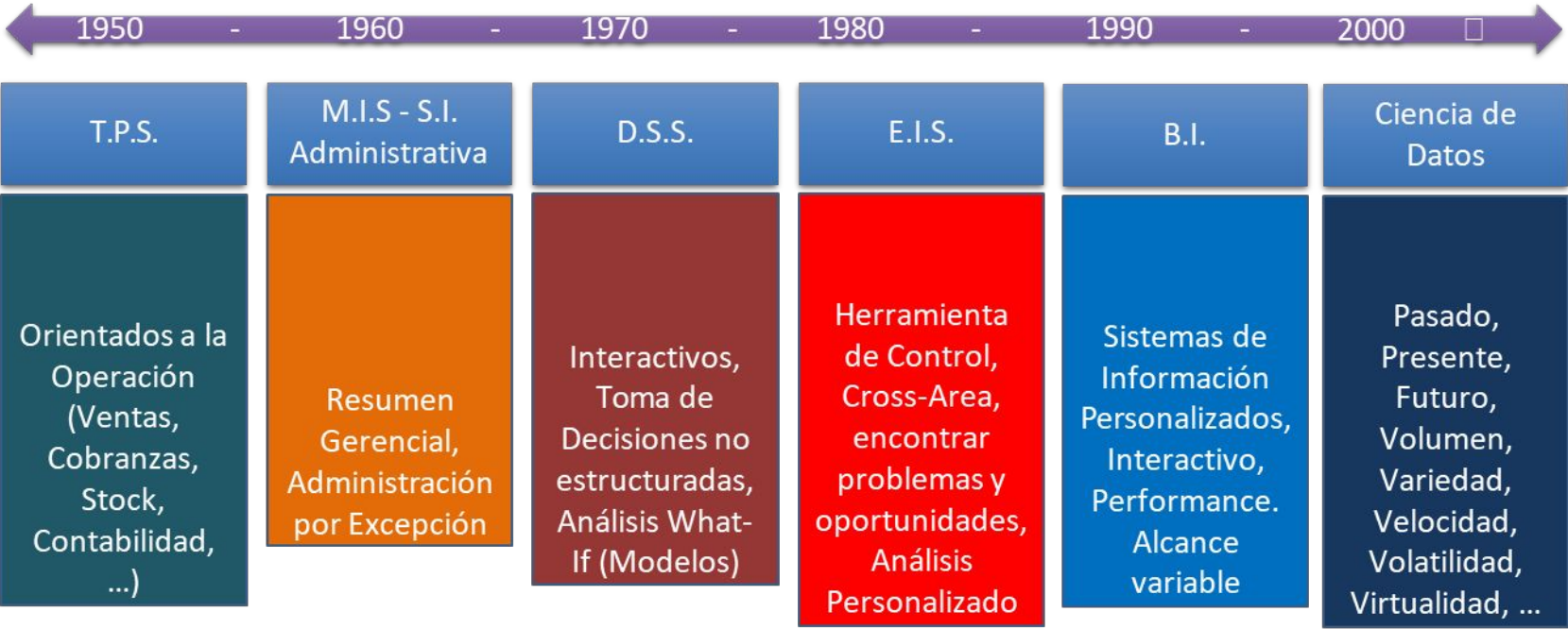
Introducción a DataWarehouse

Evolución de los Sistemas de Información



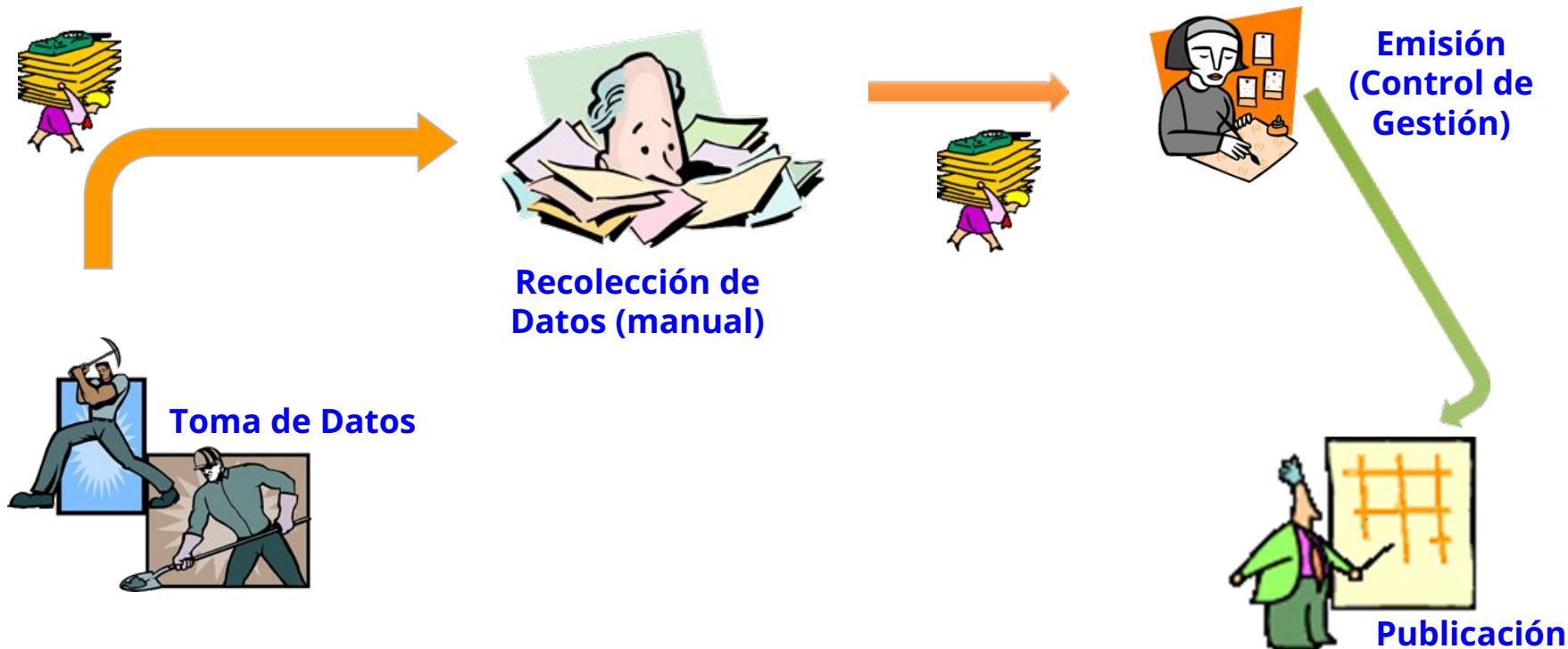
Introducción a DataWarehouse

Evolución de los Sistemas de Información



Introducción a DataWarehouse

Evolución de los Sistemas de Información - Previo a IT



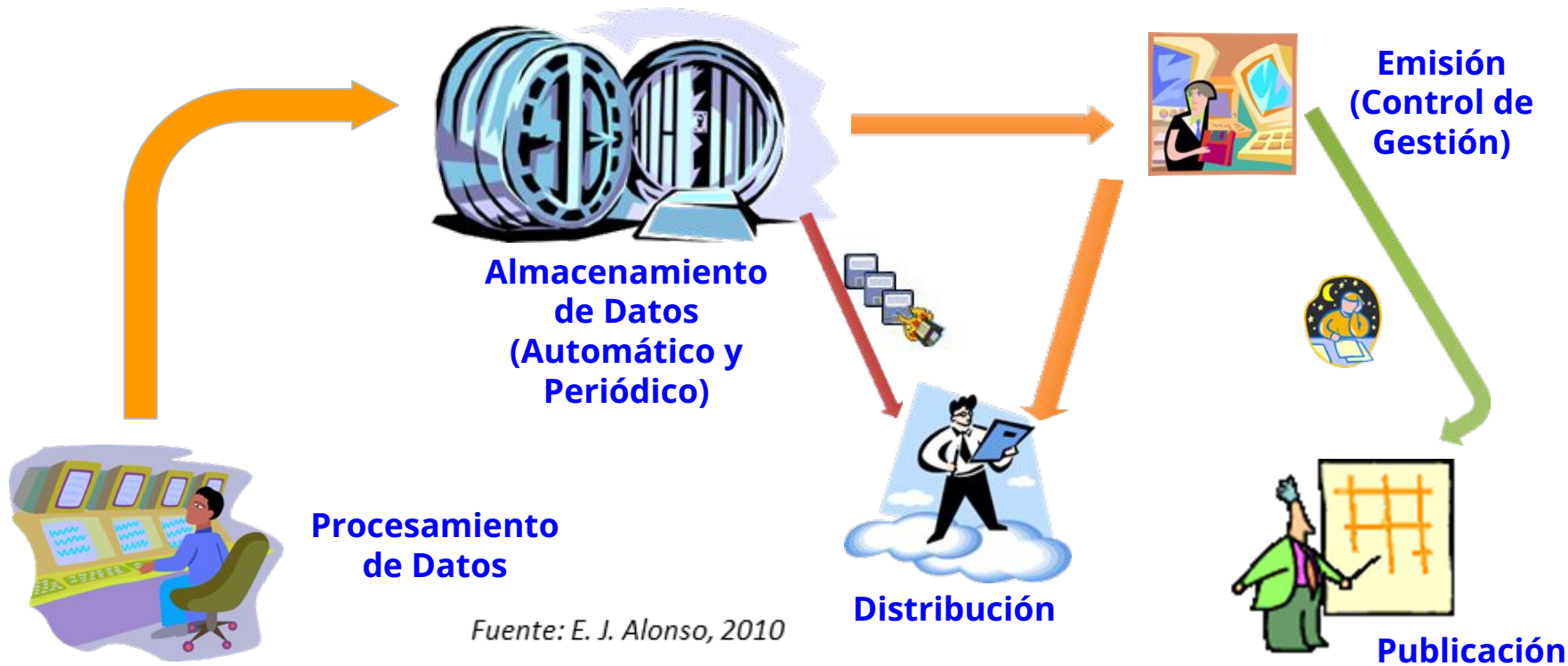
Introducción a DataWarehouse

Evolución de los Sistemas de Información - Con IT



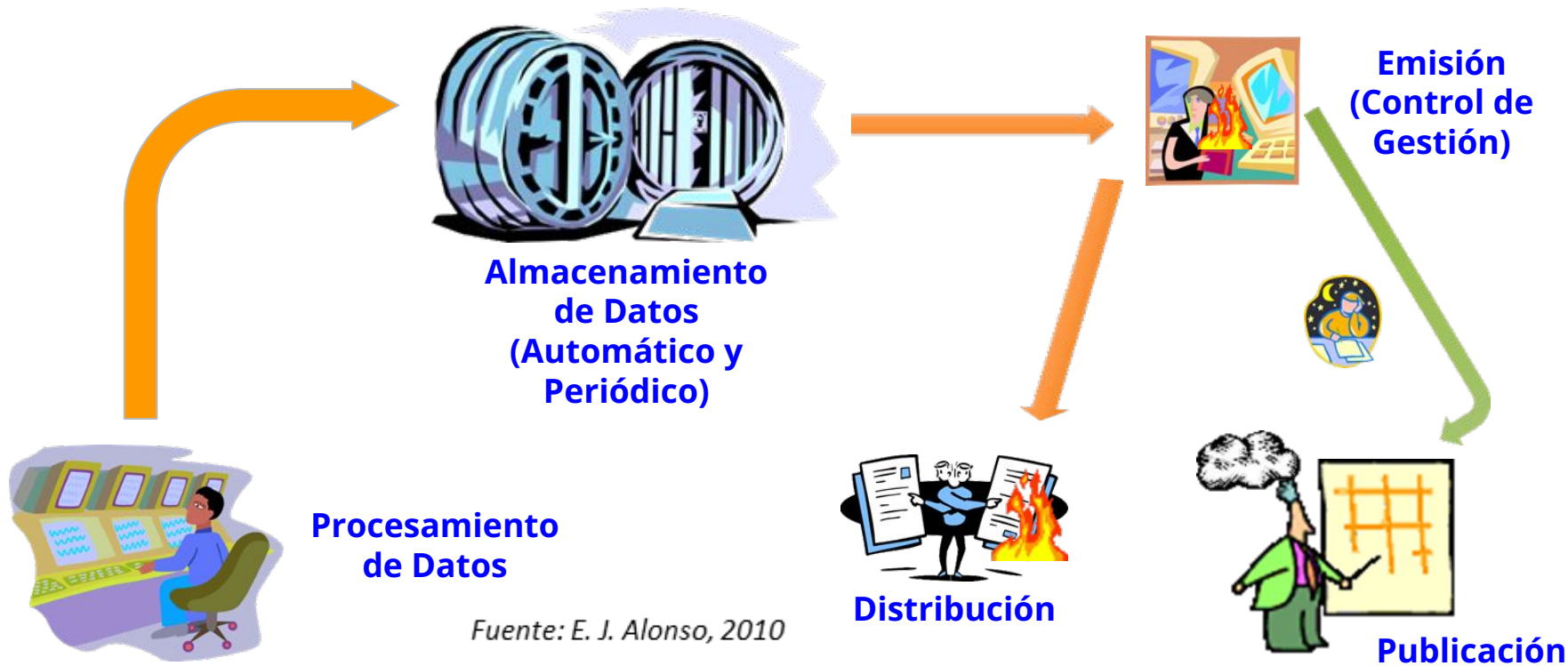
Introducción a DataWarehouse

Evolución de los Sistemas de Información - Con DW



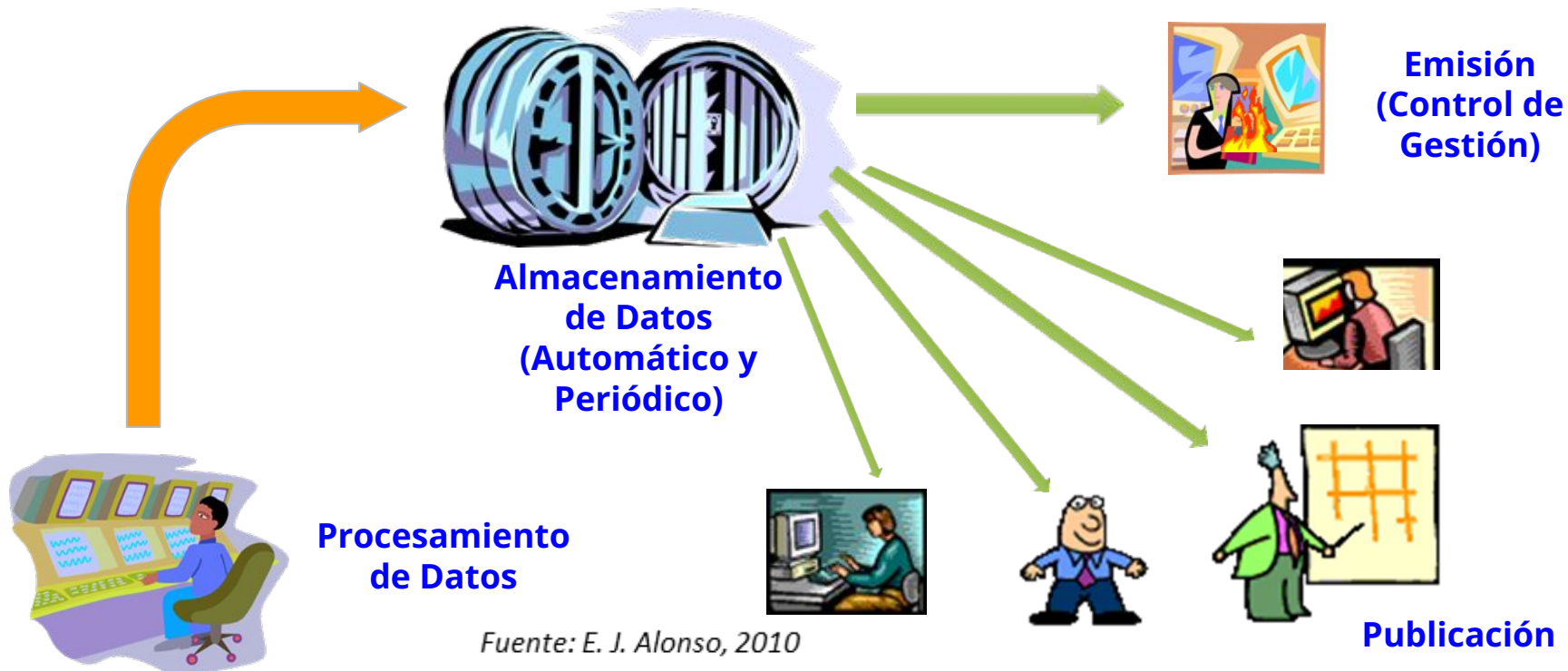
Introducción a DataWarehouse

Evolución de los Sistemas de Información - Con DW + Cultura



Introducción a DataWarehouse

Evolución de los Sistemas de Información - Modelo Superador



Introducción a DataWarehouse

Éxito y Fracaso de los Sistemas de Información

- **Falta de alineación entre los sistemas de información y la estrategia empresarial:** muchas organizaciones siguen considerando los sistemas de información como un mero instrumento que simplifica la burocracia sin valorar las ventajas estratégicas que estos presentan
- **Escaso apoyo de la administración:** la alta dirección de la compañía ha de percibir realmente que los sistemas de información constituyen un arma estratégica. Además, ha de existir una predisposición a cambiar la organización empresarial si lo requieren los nuevos sistemas de información
- **Mala identificación de las necesidades de información:** las empresas implantan las tecnologías de información sin previamente haber realizado un proceso de determinación de las necesidades de información y cómo estas pueden ser satisfechas utilizando adecuadamente los sistemas de información

Fuente: H. Trasobares, 2019

Introducción a DataWarehouse

Éxito y Fracaso de los Sistemas de Información

- **Escaso involucramiento o influencia del usuario final:** a la hora de diseñar el sistema de información resulta fundamental contar con la opinión del usuario final, el cual va a ser quien utilice el sistema de información. Por ello este usuario ha de estar motivado e incentivado a colaborar en el diseño del sistema
- **Escasa o nula formación del personal:** se requiere siempre la realización de actividades formativas para el aprendizaje de las nuevas herramientas informáticas a utilizar en la empresa.

Fuente: H. Trasobares, 2019

Introducción a DataWarehouse

Éxito y Fracaso de los Sistemas de Información



**Capacidad de Recolectar y
Almacenar Datos**

Capacidad de Generar Información

- Las grandes cantidades de datos, por sí solas, no aportan mayor información a la Organización (particularmente a la alta dirección)
- Paradójicamente, cuantos más datos tienen las organizaciones, menor facilidad para producir información
- Las organizaciones actuales cuentan con múltiples Sistemas Operacionales, algunos especializados en un aspecto en particular del Negocio o Foco y administran gran cantidad de Datos y cierta Información
- Los Organismos actuales han visto aumentada su capacidad de generar y recolectar datos (archivos compartidos entre empresas, tecnologías de entrada de datos, sistemas no integrados, ...)

“ Una organización que no pueda identificar, categorizar y resumir los datos, puede ser rica en datos, pero será pobre en información”.
(Stuart Madnick, 1993).

Introducción a DataWarehouse

Datos e Información - Problemas Frecuentes

- Datos dispersos en distintas plataformas y sistemas
- Discrepancias de definiciones entre personas de la misma organización
- Datos Duplicados e inconsistentes entre los distintos sistemas
- Sin perspectivas temporales o historia que permitan deducir tendencias
- Los plazos de entrega no compatibles con las necesidades
- Falta de disponibilidad – Ubicación, tecnología, Compliance, ...
- Altos volúmenes de datos y baja capacidad para el procesamiento y análisis
- Problemas de Distribución / Publicación

Introducción a DataWarehouse

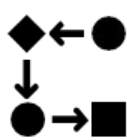
Datos e Información - Problemas Frecuentes

- “En esta organización tenemos montañas de Datos, pero no podemos accederlos”
- “Necesitamos manejar libremente los datos”
- “Tienen que hacer esto mucho más fácil para la gente del negocio”
- “Sólo muéstrenme lo que es importante”
- “No es posible que me reúna con dos personas que tengan el mismo indicador con iguales valores”
- “Queremos que la gente use la información como soporte para la toma de decisiones”
- “El área M tiene la información que necesito, pero no la comparte”

Introducción a DataWarehouse

Conceptos y Definiciones

“**Data Warehousing (DWH)** describe los procesos necesarios para la extracción, limpieza, transformación, control y carga de datos para mantenerlos en un **Data Warehouse (DW)**”



Procesos



Herramientas



Técnicas



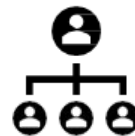
Metodologías



Contruir



Utilizar



Administrar



Mantener

El foco de DWH está en brindar un entorno de datos históricos e integrados sobre datos de la operación del negocio, aplicando reglas de negocio y relaciones entre distintos conceptos que pudieran no estar presentes en el mundo transaccional.

Importante: ¡No es un Producto! (ni varios)

Introducción a DataWarehouse

Conceptos y Definiciones

- “DW es una vista consolidada de datos corporativos, optimizados para reportes y análisis”
- “DW es una copia sumariada de transacciones y datos no transaccionales específicamente estructurados para consultas y análisis.
- “DW es un almacén o repositorio de datos categorizados, que concentra un gran volumen de información de interés para toda una organización, la cual se distribuye por medio de diversas herramientas de consulta y de creación de informes orientadas a la toma de decisiones”

Introducción a DataWarehouse

Conceptos y Definiciones

- “DW es un repositorio de datos específicamente construido y mantenido para responder sosteniblemente a las necesidades de información estratégica, táctica y operativa, de los principales responsables del negocio. DW representa en cada organización el ideal de la única fuente de datos necesaria para construir todo tipo de sistema que brinde soporte a la toma de decisiones”. (E.J. Alonso).

Introducción a DataWarehouse

Conceptos y Definiciones

“Data Warehouse (DW) es la combinación de dos componentes principales: una base de datos integrada para soporte a la decisión y software relacionado para recopilar, limpiar, transformar y almacenar datos desde una variedad de fuentes operativas y externas” (Autores varios).



Orientado a temas



Integrado



Variante en el tiempo



No Volátil



Histórico



Decisión

Un **Enterprise Data Warehouse (EDW)** es un DW centralizado que es diseñado para responder a las necesidades de BI de la organización. Considera un modelo de negocio global y consistente que soporte el proceso de toma de decisiones across-the-Enterprise (Bill Inmon, 1992).

Importante: No se compra; se construye / adapta

Introducción a DataWarehouse

Conceptos y Definiciones

“Data Warehouse (DW) es una copia de datos transaccionales específicamente estructurados para consultas y reportes” (Fuente: Ralph Kimball).



Producción



Compras



Contabilidad



Ventas



Finanzas



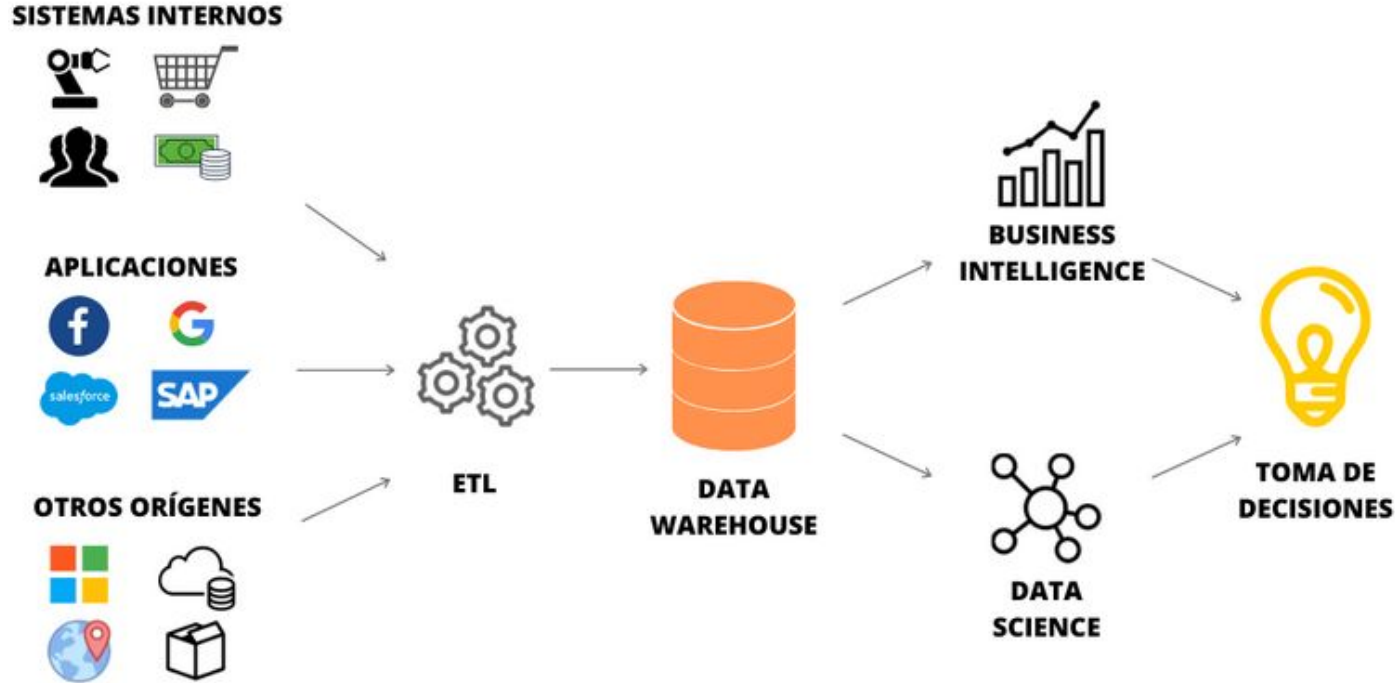
Modelo
Dimensional

DW es un almacén de datos al que se extrae, limpia, conforma y entrega una fuente de datos dimensional para la consulta y el análisis.

Importante: No se compra; se construye / adapta

Introducción a DataWarehouse

El DataWarehouse centro de la Toma de Decisiones



Introducción a DataWarehouse

Orígenes e Historia

- En la década de 1970 surge el concepto de Data Mart Dimensional para el comercio minorista.
- En 1983 se presenta el primer sistema de gestión de bases de datos diseñado específicamente para ayudar a la toma de decisiones (Teradata).
- A fines de los años 80, Devlin y Murphy (ambos investigadores de IBM) presentan el concepto de Datawarehouse empresarial.

Introducción a DataWarehouse

Conceptos, motivación y marco histórico

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- **Conceptos, motivación y marco histórico.**
- **Utilidad y funciones.**
- Generación y construcción de DW: ETL
- Tipos de DataWarehouse.
- Comparativa con Bases de Datos Relacionales.
- Ejemplos prácticos / Casos de Éxito.
- Resumen.

Introducción a DataWarehouse

¿Qué “promete” DWH?

- ~~Datos dispersos en distintas plataformas y sistemas.~~
 - **Única fuente (modelo) de información de la empresa.**
- ~~Discrepancias de definiciones entre personas de la misma organización.~~
 - **Única visión de negocio compartida across-the-company**
- ~~Datos duplicados e inconsistentes entre distintos sistemas.~~
 - **Datos integrados, únicos, consistentes.**
- ~~Sin perspectivas temporales o historia que permitan deducir tendencias.~~
 - **Información histórica e invariante.**

Introducción a DataWarehouse

¿Qué “promete” DWH?

- ~~Plazos de entrega no compatibles con necesidades.~~
 - **Tiempo mínimo para satisfacer al negocio. Respuestas a consultas no previamente formuladas.**
- ~~Falta de disponibilidad Ubicación, tecnología, Compliance, ...~~
 - **Disponibles de acuerdo al perfil de usuarios; restricciones de acuerdo al perfil**
- ~~Altos volúmenes de datos y baja capacidad para el procesamiento y análisis.~~
 - **Alta Performance**
- ~~Problemas de Distribución / Publicación.~~
 - **Gobierno.**

Introducción a DataWarehouse

¿Qué “promete” DWH?

- Separar el mundo de la operación del mundo de la gestión de negocio, pero con compatibilidad entre ambos.
- Crear conceptos del negocio ausentes en el mundo transaccional.
- Una base de comparación y medición entre entidades dentro y fuera del negocio.
- Posibilidad de definir y crear métricas e indicadores clave de performance (KPI).
- Facilitar la visualización de información de acuerdo con las necesidades (¿exigencias?) del negocio.

Introducción a DataWarehouse

¿Qué “promete” DWH?

- Dar una respuesta efectiva, escalable y sostenible a la alta dirección.
- Acompañar a la empresa en su constante transformación y generación de negocios sin rehacer aplicaciones, BD, procesos,... pero sí adaptarlos.
- Generar un ecosistema ordenado y gobernado por políticas, procesos, metodología.
- Unificar y difundir conceptos y definiciones de negocio mal empleados.

Introducción a DataWarehouse

Principios a Considerar al Diseñar un DataWarehouse

- **Foco en los Objetivos del Negocio:** asegurarse de que el DW está alineado con las prioridades de la empresa y soluciona sus problemas de información.
- **Comenzar con “el fin” en mente:** es fundamental consensuar un alcance final con el negocio, independientemente de la metodología.
- **Pensar y diseñar un todo; construir de a poco:** tanto los enfoques top-down como bottom-up son compatibles con esta recomendación: construir y entregar incrementalmente.
- **Colabore:** con otros involucrados e iniciativas (gobierno, Calidad, Metadata, ...)

Introducción a DataWarehouse

Principios a Considerar al Diseñar un DataWarehouse

- **Sumarizar y optimizar al final:** Construir sobre la base de datos atómicos; no reemplazar detalle por agregaciones y sumalizaciones.
- **Promover transparencia y self-service:** cuanto más claro se muestre el contexto, mayor será el consumo.
- Divulgue cómo los datos son procesados e integrados.
- **Construya Metadata mientras define, analiza, construye:** el éxito de los DWs tiene relación directa con la explicación que acompaña a los datos e información. Desde el comienzo es más fácil.

Introducción a DataWarehouse

Ventajas de Implementar un DW

- Acceso a los datos de forma rápida y fácil.
- Acceso a información coherente y actualizada.
- Reducción de tiempo en el análisis de datos y generación de informes.
- Análisis de tendencias(histórico).

Introducción a DataWarehouse

Desventajas de Implementar un DW

- Complejidad en el desarrollo.
- No es una buena solución para datos “no estructurados”.
- Tanto la creación como la implementación de un DataWarehouse lleva mucho tiempo y requiere mucho trabajo.
- Es complejo realizar cambios en los tipos de datos, los esquemas fuentes de datos y las búsquedas.
- Requiere una alta inversión inicial.

Introducción a DataWarehouse

Utilidad y funciones

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- Conceptos, motivación y marco histórico.
- Utilidad y funciones.
- **Generación y construcción de DW: ETL**
- Tipos de DataWarehouse.
- Comparativa con Bases de Datos Relacionales.
- Ejemplos prácticos / Casos de Éxito.
- Resumen.

Introducción a DataWarehouse

Construcción de un DataWarehouse - Diseño

La primera etapa consiste en el diseño de la solución que vamos a implementar.

- Se deben considerar los Requerimientos de negocio y definir el tipo de DataWarehouse (EDW, DM, ODS) a implementar.
- Definir la Arquitectura / Metodología a utilizar
 - Top Down (Bill Inmon) □ Modelo Relacional Normalizado
 - Bottom Up (Ralph Kimball) □ Modelo Dimensional
- Identificar los orígenes de datos disponibles y requeridos.
 - Transformaciones
 - Limpieza / Completado
 - Integración

Introducción a DataWarehouse

Construcción de un DataWarehouse - Construcción

La segunda etapa implica construir el DataWarehouse y los mecanismos de obtención de los datos que incluirá:

- Definir Tecnología (tanto backend del DW como integración y conectividad).
- Definir herramientas de BI que lo utilizarán.
- Identificar forma de detección de cambios
 - CDC / Log / Timestamp
- Construir los procesos de ETL (Extract/Transform/Load) o ELT (Extract/Load/Transform).
 - Carga Inicial / Actualización Periódica

Introducción a DataWarehouse

Construcción de un DataWarehouse - Explotación

La tercer etapa consiste en utilizar y mantener el DW:

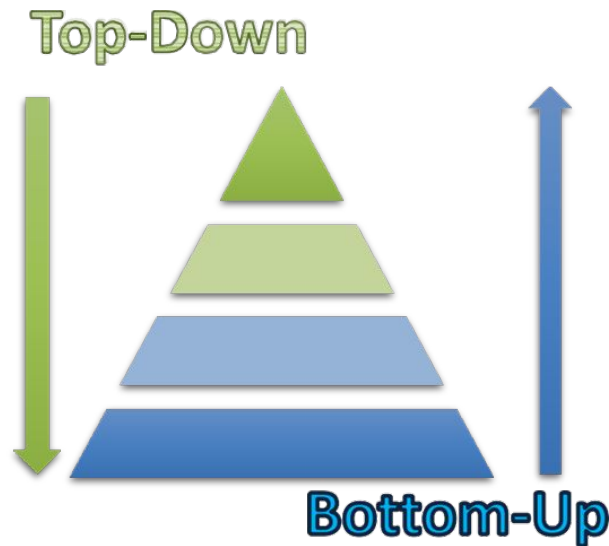
- Construir Reportes
 - Reportes
 - Tableros / KPI
- Monitorear y Mejorar
 - Identificar y Agregar Información Nueva.
 - Monitorear y Mejorar Performance / Recursos.
 - Validar Calidad de los Datos (datos faltantes / errores).

Introducción a DataWarehouse

Arquitecturas de DW

Existen **dos metodologías generalmente aceptadas** al momento de diseñar un DW:

- Bill Inmon “Top-Down”
 - Modelo Relacional Normalizado - Atómico
- Ralph Kimball “Bottom-up”
 - Modelo Dimensional – Data Marts



Introducción a DataWarehouse

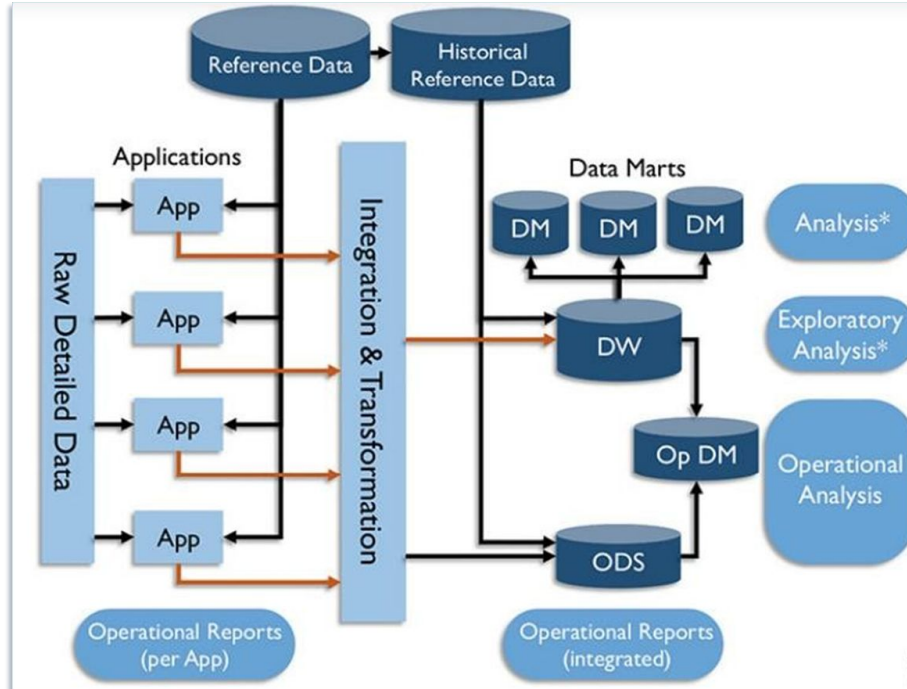
Arquitecturas de DW

Sin embargo, para ambos modelos:

- El DataWarehouse almacena datos de otros sistemas.
- Los datos están organizados de modo tal que garantice valor agregado.
- DW permite acceder y analizar datos. Sirven para / desde el soporte a la operación hasta el análisis predictivo.
- Las organizaciones construyen DW porque necesitan brindar datos confiables e integrados a la organización y personas autorizadas.

Introducción a DataWarehouse

Arquitecturas según B. Inmon - Corporate Information Factory



Introducción a DataWarehouse

Arquitecturas según B. Inmon - Corporate Information Factory

La Fábrica de Información Corporativa contiene:

- **Aplicaciones:** que ejecutan procesos operacionales. Datos detallados desde estas aplicaciones son incluidas en los DW y ODS
- **Integración y Transformación:** los datos son transformados de manera que puedan integrarse al **DW**.
- **ODS:** es una BD integrada de datos operacionales. Se cargan desde BD operacionales y desde otras fuentes. Poca historia (1 a 3 meses). Puede ser fuente para cargar el DW.

Introducción a DataWarehouse

Arquitecturas según B. Inmon - Corporate Information Factory

La Fábrica de Información Corporativa contiene:

- **DataWarehouse:** el DW provee una única fuente integrada de datos corporativos para la toma de decisiones gerenciales, análisis estratégico y planeamiento. Los datos fluyen en una única dirección: desde las fuentes y ODS a los DM. Datos erróneos son rechazados y reingresados luego de ajustes.
- **Data Marts:** datos preparados para analizar. A veces son parte del DW para necesidades puntuales (datos agregados; modelos dimensionales desnormalizados).

Introducción a DataWarehouse

Arquitecturas según B. Inmon - Corporate Information Factory

La Fábrica de Información Corporativa contiene:

- **Data Marts Operacionales:** son DM enfocados en la toma de decisiones tácticas. Se alimentan desde el ODS en lugar del DW (a veces). Historia y volatilidad similar al ODS.
- **Reportes Operativos:** Salidas de todos los repositorios (DW, DM, ODS).
- **Datos Externos, Maestros y de Referencia:** Datos adicionales y necesarios para controlar y contextualizar otros datos

Introducción a DataWarehouse

Arquitecturas según B. Inmon - Corporate Information Factory

La recolección de datos y su transformación en información y análisis, provoca cambios:

- **Funciones:** de la ejecución de funciones operativas al análisis del negocio.
- **Usuarios:** de personal de la línea a decisores
- **Uso:** desde operaciones y procesos estándar a usos personalizados / ad-hoc
- **Volumen:** pocos datos necesarios para la operación a muchos para una consulta o proceso.

Introducción a DataWarehouse

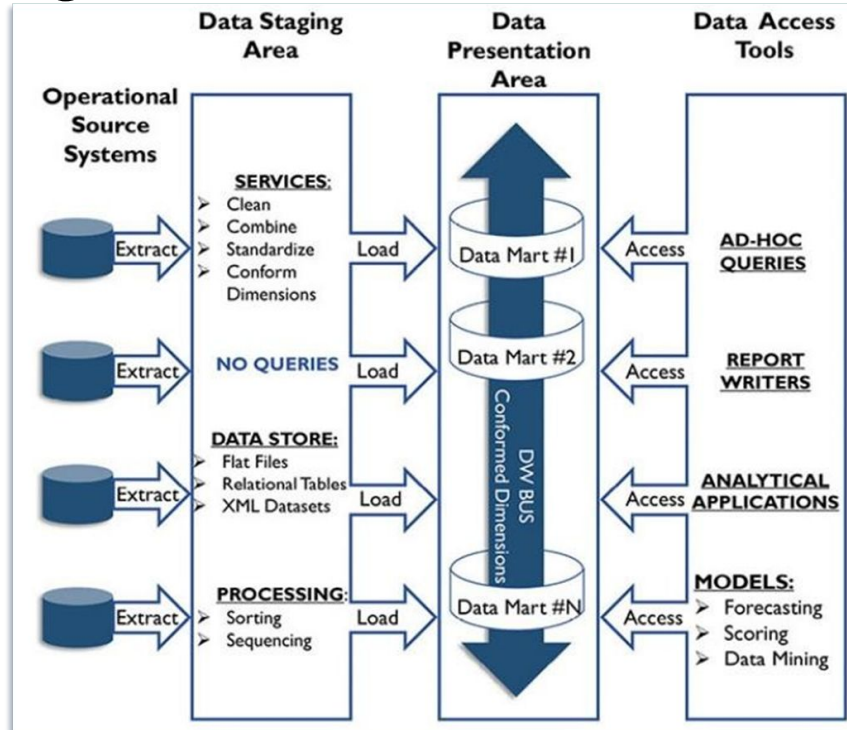
Arquitecturas según B. Inmon - Corporate Information Factory

La recolección de datos y su transformación en información y análisis, provoca cambios:

- **Organización de datos:** de acuerdo a visión de gestión en lugar de tareas operativas.
- **Integración:** de aislados por generación y uso a integrados en un único modelo.
- **Tiempo:** datos actuales y corrientes a variantes en el tiempo.
- **Historia:** mucha más historia en DW que en aplicaciones de origen.

Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional



Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional

El DataWarehouse Dimensional contiene:

- **Sistemas Fuente Operativos:** TPS de la empresa. Estas aplicaciones crean los datos que serán integrados en el ODS y DW.
- **Staging Area:** El área de staging de Kimball incluye un conjunto de procesos necesarios para integrar y transformar datos para su presentación. Puede ser comparado con una combinación de las capas de Integración y Transformación y más componentes del CIF de Inmon.

Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional

El DataWarehouse Dimensional contiene:

- **Área de Presentación:** similar a los Data Marts en el CIF. La diferencia arquitectónica clave es un paradigma integrador de un "DW Bus", como dimensiones compartidas o conformadas que unifican los múltiples mercados de datos
- **Herramientas de Acceso a Datos:** el approach de Kimball se enfoca en los requerimientos de usuarios finales para seleccionar las herramientas adecuadas.

Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional

El DataWarehouse Dimensional tiene diversas áreas de almacenamiento:

- **Staging Area:** Es un área intermedia de almacenamiento transitorio de datos. Aquí pueden ser transformados, integrados y preparados para su carga en el DW.
- **Datos Maestros y de Referencia / Dimensiones Conformadas:** Datos Maestros y de Referencia pueden almacenarse en distintos repositorios. El DW alimenta nuevos datos maestros y de referencia, manteniendo las Dimensiones Conformadas para distintos repositorios

Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional

El DataWarehouse Dimensional tiene diversas áreas de almacenamiento:

- **Central Warehouse:** una vez transformados y preparados, el DW mantiene datos en una capa “atómica”. La estructura de esta área es desarrollada teniendo en cuenta aspectos de Performance y patrones de uso. Algunos elementos de diseño pueden ser:
 - La relación entre la clave de negocio y claves subrogadas (performance).
 - Índices y claves foráneas para soportar dimensiones.
 - Técnicas de CDC para detectar, mantener y almacenar historia.

Introducción a DataWarehouse

Arquitecturas según R. Kimball - DataWarehouse Dimensional

El DataWarehouse Dimensional tiene diversas áreas de almacenamiento:

- **ODS:** una versión de almacenamiento de datos con baja latencia para uso operativo. Se actualiza más rápido que el DW (poca historia). Útil para intervalos de refresco frecuentes, con el tiempo podrían integrarse con DW o DM para datos real-time.
- **Data Marts:** Es un tipo de almacenamiento frecuentemente usado para la capa de presentación. Orientado para un área temática, departamento o proceso. Pueden formar la base de un DW virtualizado ("suma" de los DM).
- **Cubos:** Modelos OLAP. Pueden implementarse como Relacionales, Multidimensionales o Híbridos.

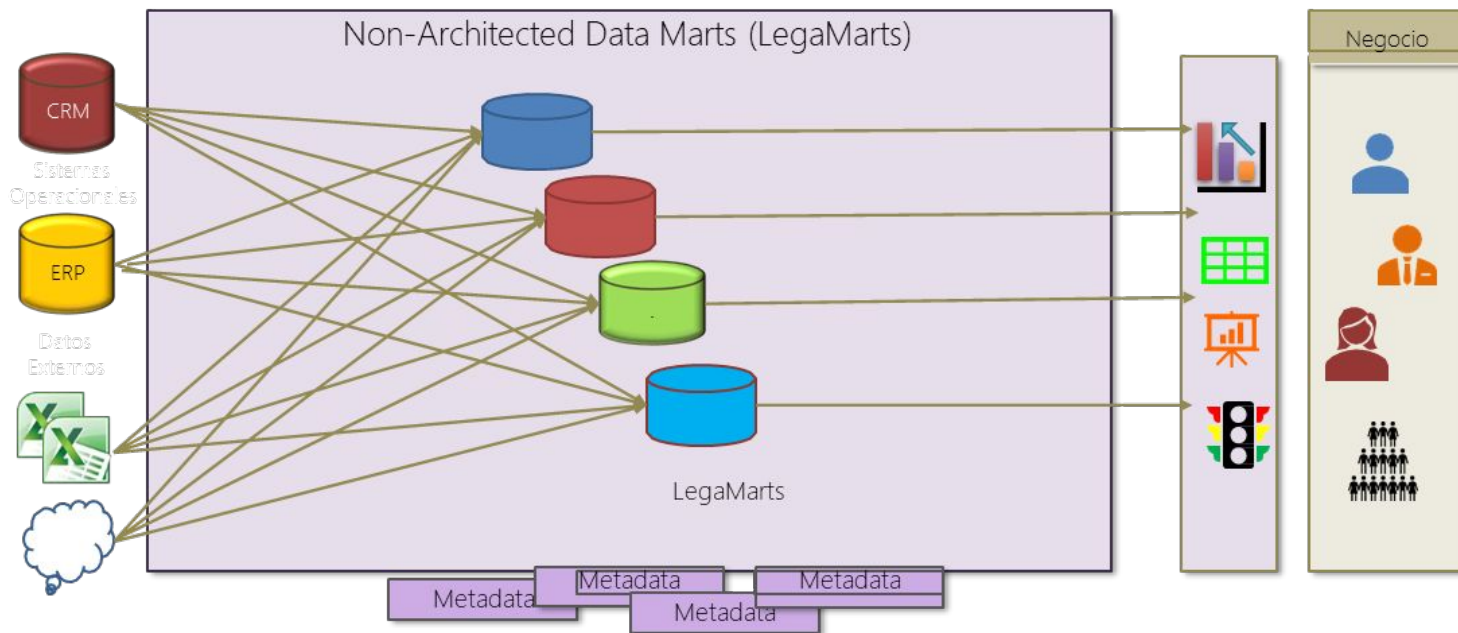
Introducción a DataWarehouse

Interpretación de Arquitecturas

A continuación analizaremos algunas arquitecturas según los dos modelos analizados y variaciones a los mismos.

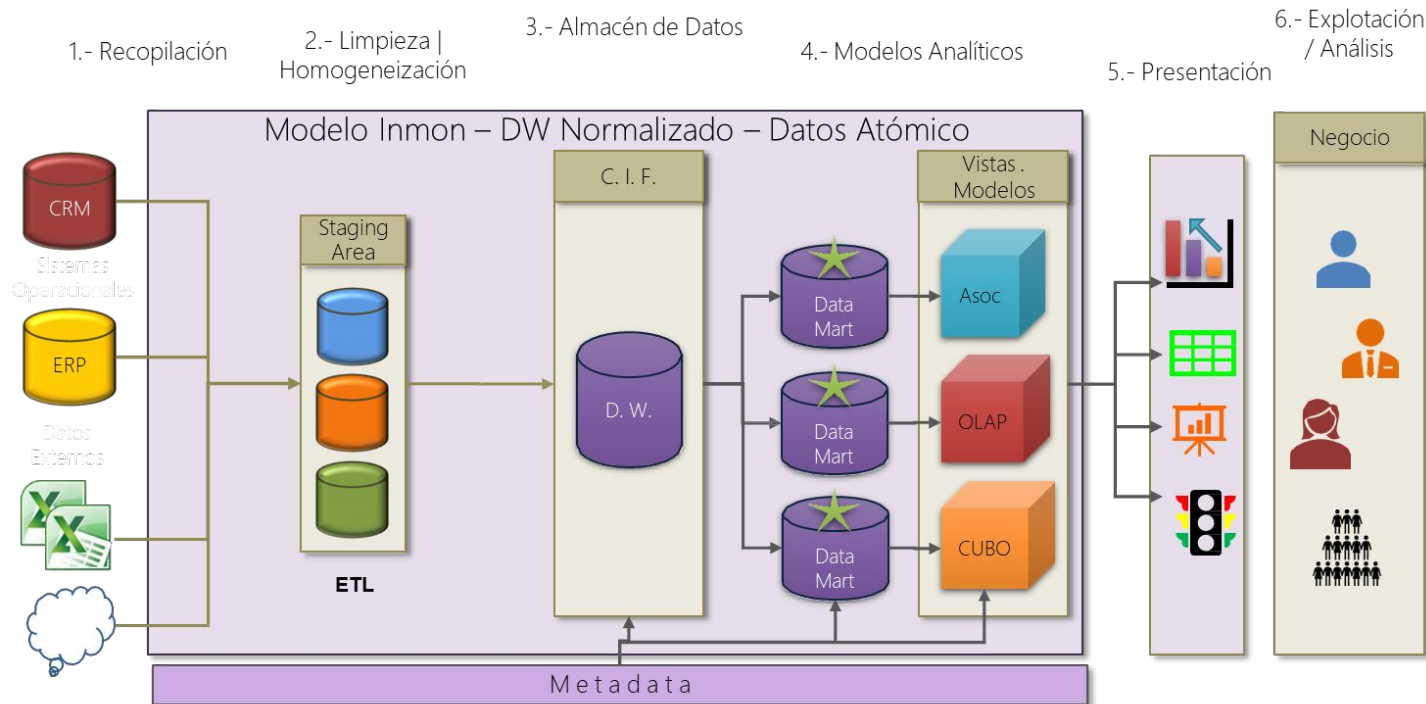
Introducción a DataWarehouse

Interpretación de Arquitecturas - Sin Arquitectura



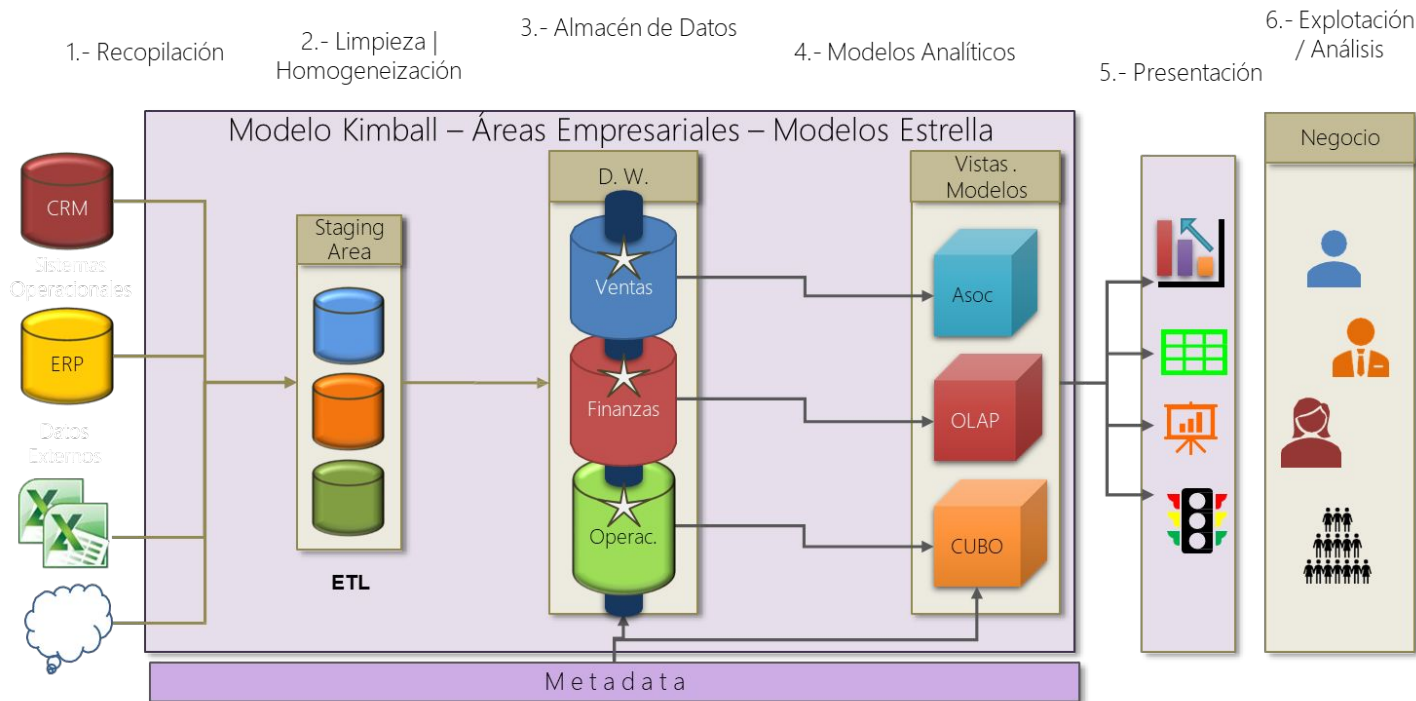
Introducción a DataWarehouse

Interpretación de Arquitecturas - Según B. Inmon



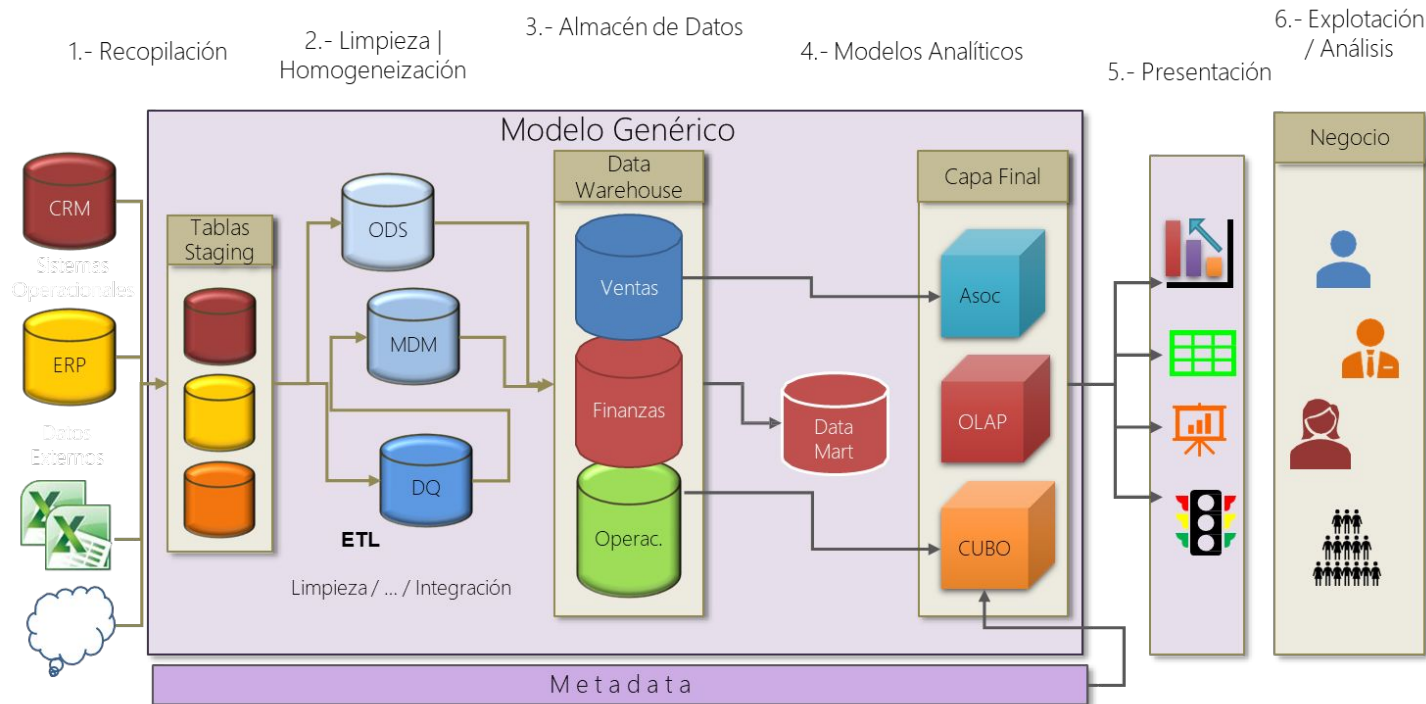
Introducción a DataWarehouse

Interpretación de Arquitecturas - Según R. Kimball



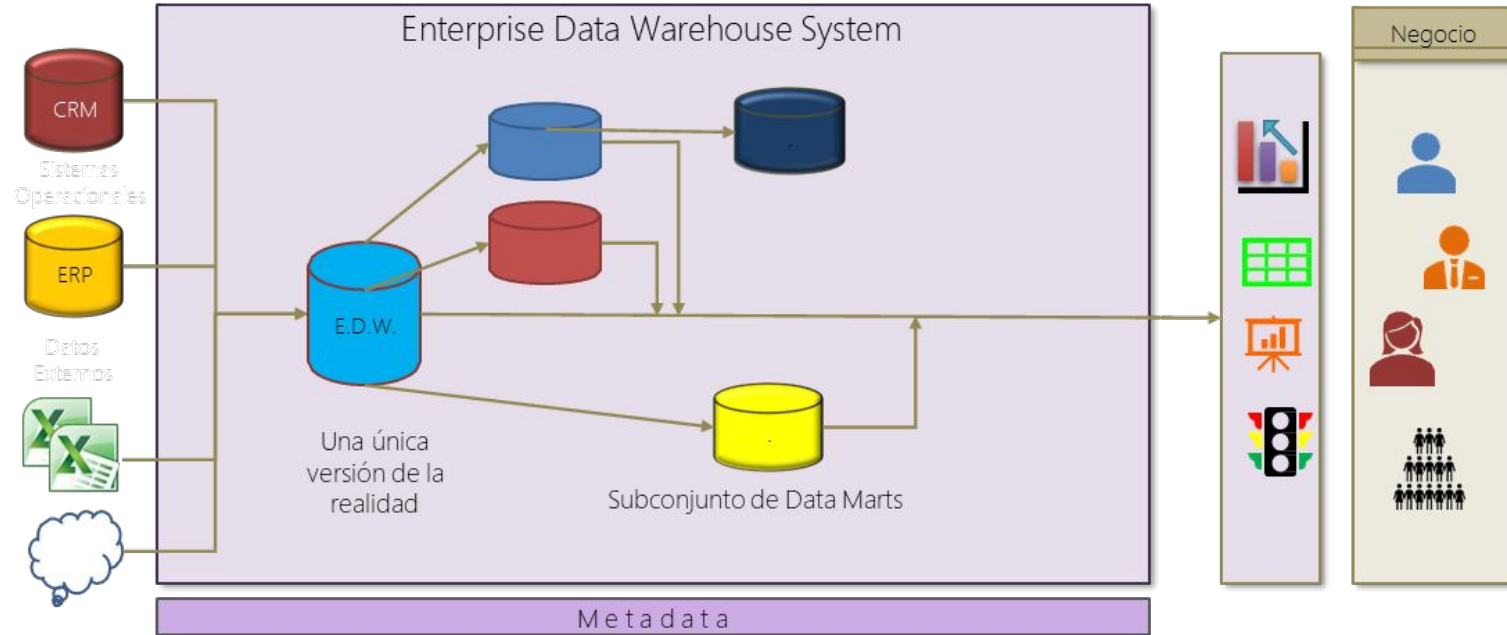
Introducción a DataWarehouse

Interpretación de Arquitecturas - Modelo Mixto



Introducción a DataWarehouse

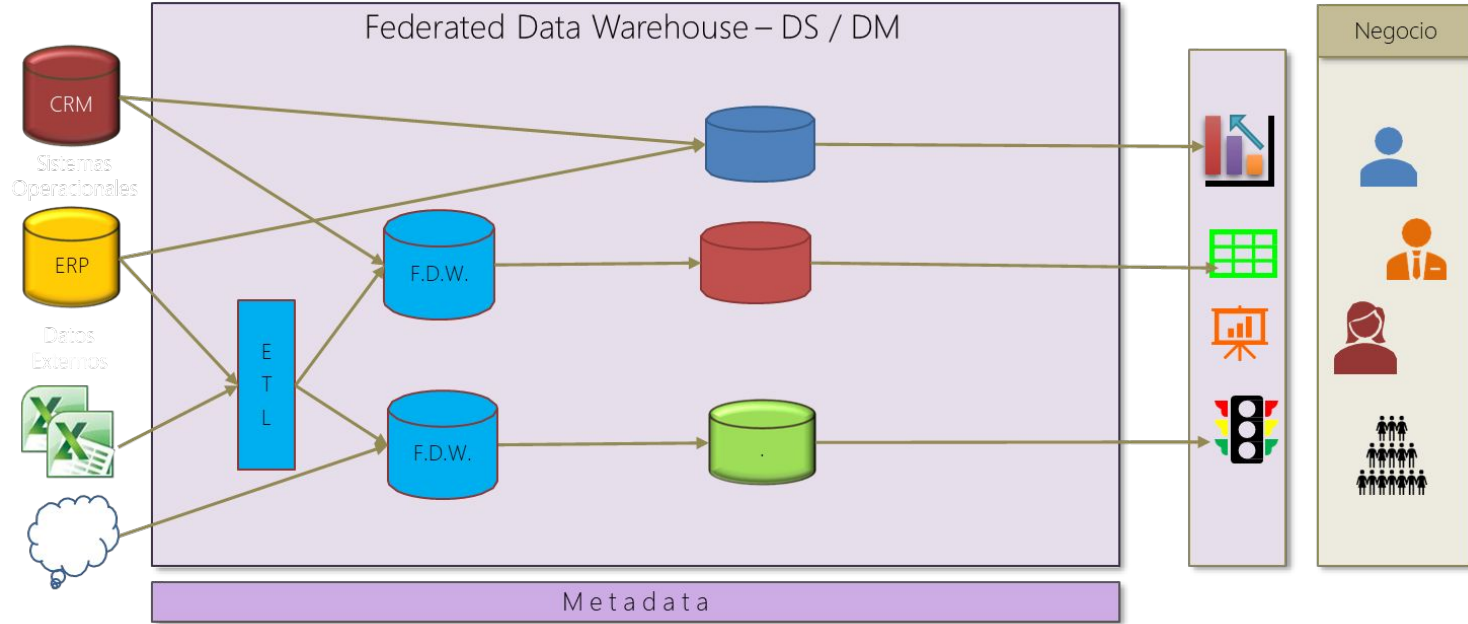
Interpretación de Arquitecturas - Enterprise DataWarehouse



La diferencia está en la metodología, el approach, las herramientas, las técnicas, el enfoque, las prioridades, los destinatarios, ... Todo

Introducción a DataWarehouse

Interpretación de Arquitecturas - DataWarehouse Federado



Introducción a DataWarehouse

Arquitecturas - Componentes en Común

Independientemente de la Arquitectura seleccionada, todos los DW tienen elementos en común:

- **ETL:** El proceso de Extracción – Transformación – Carga
- **Arquitectura:** alguna, clásica o adaptación / interpretación
 - Interacción entre los distintos componentes
 - Repositorios de Información
 - Procesos
- **Explotación:** capa de visualización / consultas / análisis en línea / Balanced Scorecard / KPIs, etc. etc.

Introducción a DataWarehouse

Generación y construcción de DW: ETL

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- Conceptos, motivación y marco histórico.
- Utilidad y funciones.
- Generación y construcción de DW: ETL
- **Tipos de DataWarehouse.**
- Comparativa con Bases de Datos Relacionales.
- Ejemplos prácticos / Casos de Éxito.
- Resumen.

Introducción a DataWarehouse

Tipos de DW - **DataWarehouse Empresarial (EDW)**

- Orientado a la toma de decisiones estratégicas **a nivel organización**. Contiene información histórica.
- Incorpora e integra datos de **todas las áreas de la organización** y también puede incluir datos externos.
- Fácil y rápido de consultar (auge de herramientas de explotación de datos).
- Información unificada y homogeneizada. Organización multidimensional y jerárquica de los datos.

Introducción a DataWarehouse

Tipos de DW - **DataMart (DM)**

- Orientado a la toma de decisiones estratégicas **departamentales** (no a nivel organización). Contiene información histórica.
- Incorpora e integra datos de una o varias áreas de la organización y datos externos, todos ellos **relacionados a un tema específico** (Ventas, Compras, RRHH, etc.)
- Podríamos considerarlo una “partición” de un Datawarehouse empresarial.
- Implica **redundancia** de datos entre distintos datamarts.

Introducción a DataWarehouse

Tipos de DW - **DataMart (DM)**

Algunos **MITOS** sobre los **DataMarts**:

- Son pequeños.
- Son menos complejos de construir / más rápido de obtener que un Data Warehouse.
- Son escalables a un Data Warehouse.
- Pueden ser dinámicamente integrados para formar un Data Warehouse.
- Contienen una única área temática.
- Son más flexibles que los Data Warehouse.

Introducción a DataWarehouse

Tipos de DW - Comparación **DataWarehouse** vs **DataMart**

	Data Warehouse	Data Mart
Alcance	<ul style="list-style-type: none"> • Neutral a las aplicaciones • Centralizado, compartido • Cruza toda la Empresa 	<ul style="list-style-type: none"> • Requerido por una aplicación específica • Departamental • Orientado a procesos de negocio • Múltiples BD con datos redundantes
Perspectiva de los Datos	<ul style="list-style-type: none"> • Detallados e Históricos • Poca sumarización • Suavemente desnormalizado 	<ul style="list-style-type: none"> • Detallado (algo de historia) • Sumarizado • Fuertemente desnormalizado
Temas	<ul style="list-style-type: none"> • Múltiples áreas temáticas 	<ul style="list-style-type: none"> • Única y múltiples áreas temáticas parciales • Snapshot de fuentes operacionales

Introducción a DataWarehouse

Tipos de DW - Comparación **DataWarehouse** vs **DataMart**

	Data Warehouse	Data Mart
Fuentes de Datos	<ul style="list-style-type: none"> • Muchas • Operacionales, datos externos 	<ul style="list-style-type: none"> • Pocas • Operacionales, datos externos • Snapshot de BD OLTP
Tiempo de Implementación	<ul style="list-style-type: none"> • 9-18 meses para la 1era etapa (2 o 3 áreas temáticas) • Implementación en múltiples etapas 	<ul style="list-style-type: none"> • 4 a 12 meses
Características	<ul style="list-style-type: none"> • Flexible, extensible • Durable/estratégico • Orientación a los Datos 	<ul style="list-style-type: none"> • Restrictivo, no-extensible • Corta vida/táctico • Orientación al Proyecto

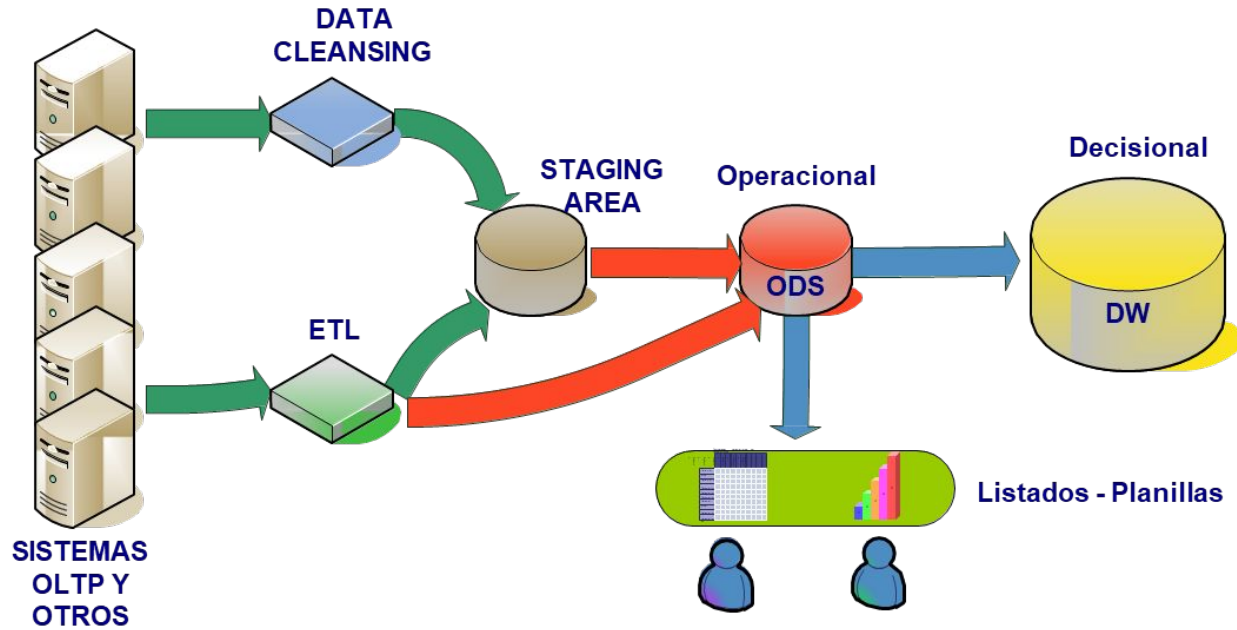
Introducción a DataWarehouse

Tipos de DW - **Operational Data Store (ODS)**

- Orientado a la toma de **decisiones tácticas (corto plazo) u operativas**. Posee información de un período reciente (**near real time**) y acotado de tiempo la cual es actualizada frecuentemente.
- Incorpora e integra datos de una o varias áreas de la organización y datos externos, todos ellos **relacionados a un tema específico** (Ventas, Compras, RRHH, etc.).
- Puede usarse como paso previo para preparar datos y luego importarlos en un Datawarehouse o Datamart.

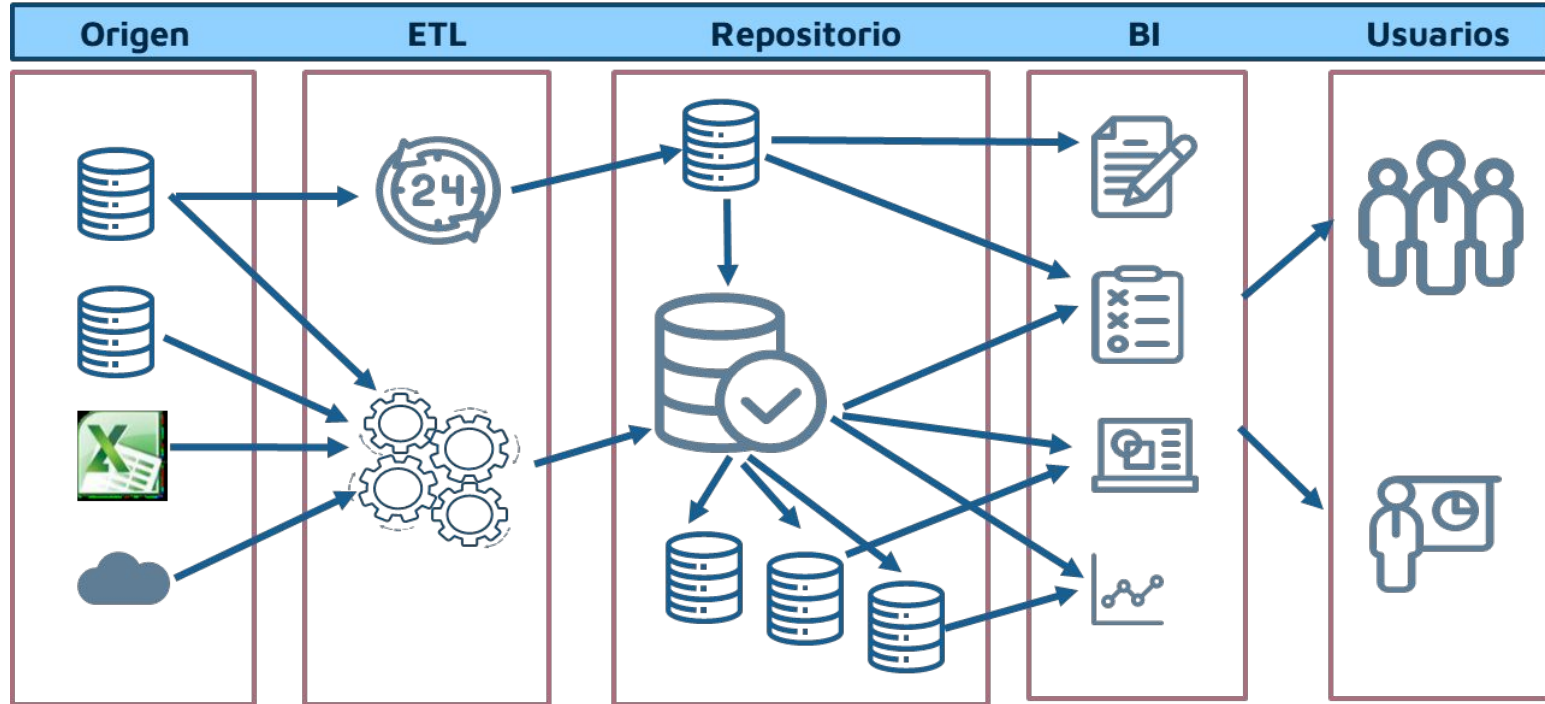
Introducción a DataWarehouse

Tipos de DW - **Operational Data Store (ODS)**



Introducción a DataWarehouse

Tipos de DW - Un modelo completo de EDW + DM + ODS



Introducción a DataWarehouse

Tipos de DW - Base de Datos Híbrida o DW Virtual

- Algunos motores de base de datos permiten la ejecución de consultas / reportes masivos directamente sobre las bases de datos operacionales (sistemas transaccionales u OLTP).
- El uso de tecnologías como “Columnar Store” o “Database In Memory” permite ejecutar consultas para la toma de decisiones en los sistemas OLTP sin afectar la performance ni la operatoria normal de los sistemas transaccionales.
- No requieren desarrollo de ETLs ni servidores / bases de datos adicionales al OLTP pero requieren recursos (memoria / almacenamiento) y diseños adicionales (Columnar Store / In Memory / Materialized Views).

Introducción a DataWarehouse

Tipos de DW - Base de Datos Híbrida o DW Virtual

Características:

- **Acceso:** Los usuarios finales pueden acceder a bases de datos operacionales directamente, usando cualquier herramienta que posibilite “la red de acceso de datos”.
- **Consultas:** Se pueden colocar las cargas de consulta no planificadas más grandes sobre sistemas operacionales.
- **Solución:** Surge de la necesidad de obtener datos operacionales, a partir de una clase relativamente grande de usuarios finales donde la frecuencia probable de pedidos es baja.
- **Conclusión:** Si no hay ETL ni repositorio destino → **no hay DW.**

Introducción a DataWarehouse

Tipos de DataWarehouse

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- Conceptos, motivación y marco histórico.
- Utilidad y funciones.
- Generación y construcción de DW: ETL
- Tipos de DataWarehouse.
- **Comparativa con Bases de Datos Relacionales.**
- Ejemplos prácticos / Casos de Éxito.
- Resumen.

Introducción a DataWarehouse

DataWarehouse versus Relacional

Tanto un DW como una base de datos relacional son sistemas de almacenamiento de datos, que permiten almacenar grandes cantidades de datos estructurados. Ambos pueden ser consultados y actualizados con transacciones.

La principal diferencia entre los dos es que un **DW está diseñado específicamente para el análisis**, mientras que las bases de datos están **diseñadas principalmente para uso "transaccional"**. Además, los DataWarehouse almacenan **datos históricos y agregados** (a menudo de fuentes dispares), mientras que las bases de datos a menudo sólo almacenan estados de información recientes o actuales asociados a una aplicación específica.

Introducción a DataWarehouse

DataWarehouse versus Relacional - General

	DataWarehouse	Base de Datos Relacional
Propósito	Análisis de Datos (Data Out)	Registro de Datos (Data In)
Datos	Históricos / Sumarizados	Actuales / Detallados / Completo
Alcance	Global (EDW) / Temático (DM / ODS)	Específico a la Aplicación
Consulta	Analíticas / Reportes Complejos (OLAP)	Individuales / Limitados (OLTP)
Usuarios	Directores / Gerentes (Reducido)	Todos (Masivo)

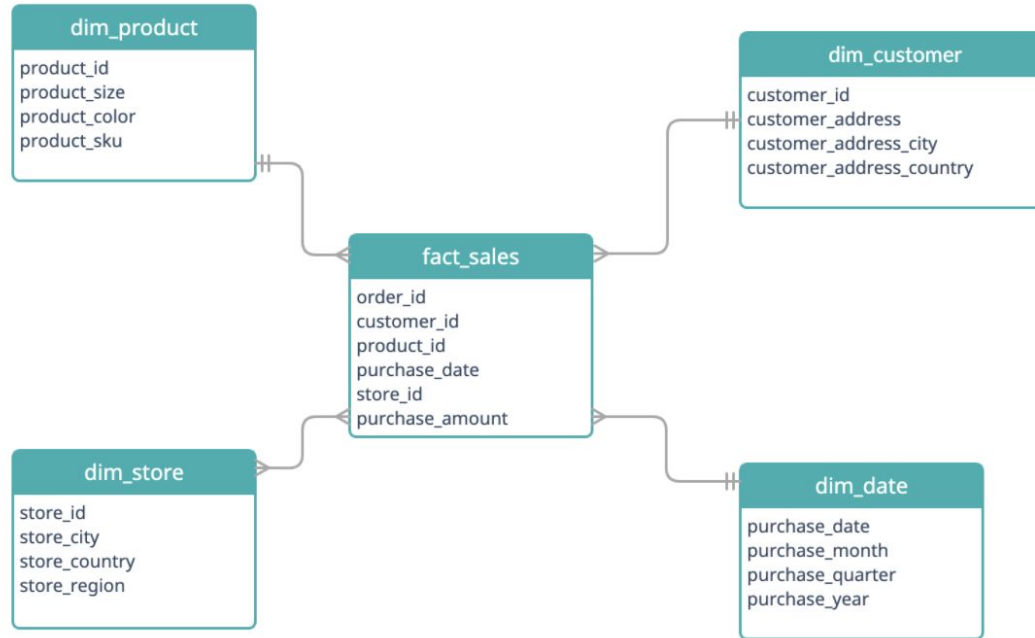
Introducción a DataWarehouse

DataWarehouse versus Relacional - General

	DataWarehouse	Base de Datos Relacional
Transacciones	Pocas y Masivas (ETL diario / semanal / mensual)	Mucha y Atómicas (emisión de comprobantes, etc.)
Disponibilidad	Según Necesidad	24 * 7 * 365
Normalización	Variable (usualmente Mínima)	Alta
Integridad	Garantizada por el ETL	Implementada en la DB
Duración	Histórica (versionado)	Volátil (datos se sobrescriben)
Modelo	Star o Snowflake	Relacional

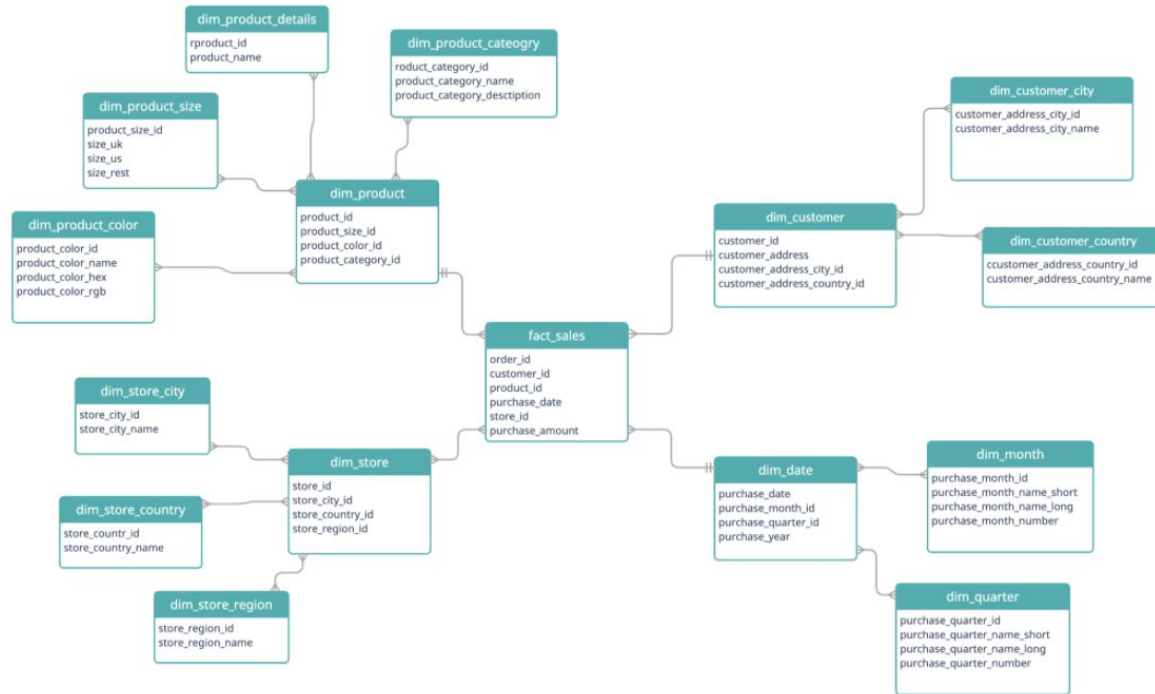
Introducción a DataWarehouse

Modelo Star (Estrella) - DataWarehouse “Denormalizado”



Introducción a DataWarehouse

Modelo Snowflake (Copo de Nieve) - DataWarehouse “Normalizado”



Introducción a DataWarehouse

Star Versus Snowflake

	Star	Snowflake
Normalización	No Normalizado	Normalizado
Redundancia	Alta	Baja
Complejidad de Queries	Baja	Media a Alta
Performance	Alta	Depende del nivel de análisis
Almacenamiento	Requiere más espacio	Requiere menos espacio
Integridad	Menor	Mayor
Diseño	Fácil Creación / Difícil Cambio	Difícil Creación / Fácil Cambio

Introducción a DataWarehouse

Comparativa con Bases de Datos Relacionales

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- Conceptos, motivación y marco histórico.
- Utilidad y funciones.
- Generación y construcción de DW: ETL
- Tipos de DataWarehouse.
- Comparativa con Bases de Datos Relacionales.
- **Ejemplos prácticos / Casos de Éxito.**
- Resumen.

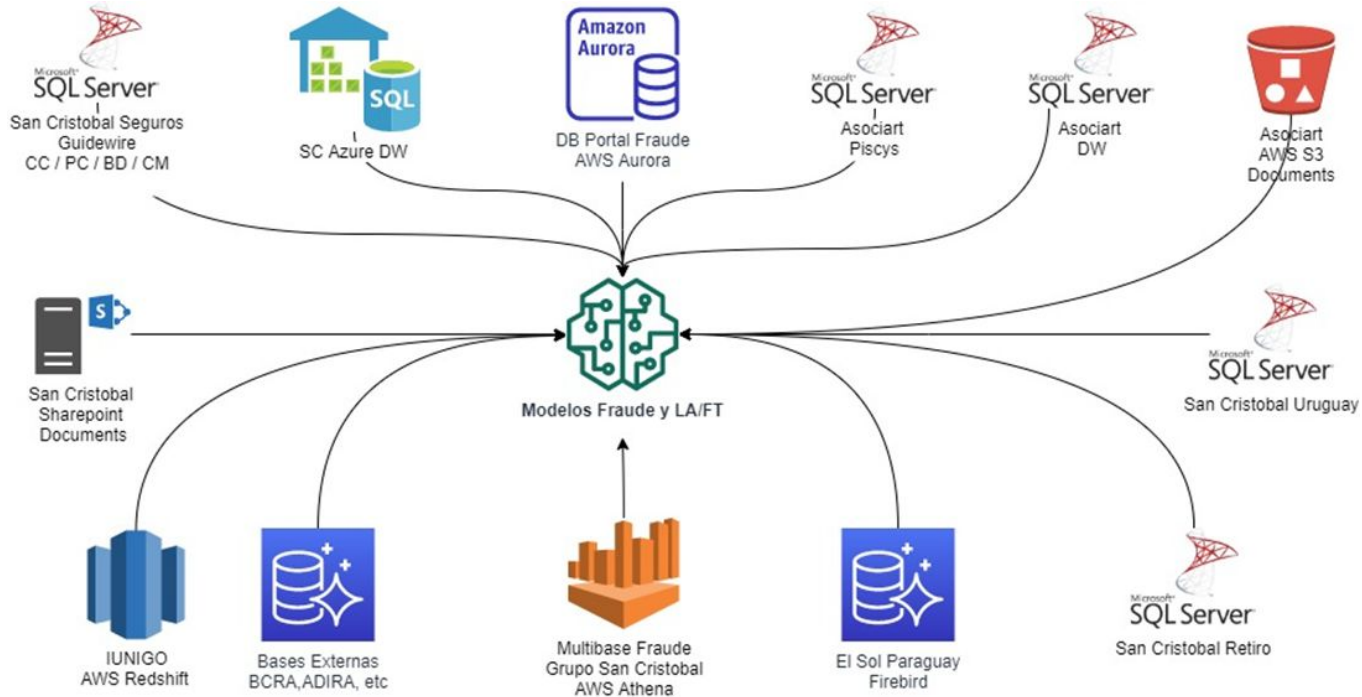
Introducción a DataWarehouse

Ejemplos Prácticos

- **Bancos:** Identificar clientes más rentables.
- **Telecomunicaciones:** Identificar potenciales clientes que quieran cambiar de compañía.
- **Seguros:** Reclamos y patrones de clientes de riesgo.
- **Fabricantes:** comparativas de costos de las líneas de producto.
- **Comercio Minorista:** Características demográficas de los clientes y patrones de compra.

Introducción a DataWarehouse

Ejemplo Práctico de DW Empresarial



Introducción a DataWarehouse

Casos de éxito - Amazon



Reducción de fraudes y agilización de precios dinámicos

Amazon, el gigante minorista online y que recientemente implementó tiendas físicas, ha sabido aprovechar las bondades del Business Intelligence para **optimizar la detección de fraudes y además la personalización en la atención al cliente.**

Con la integración Big Data han reforzado su seguridad con un sistema de detección de fraude de tarjetas de crédito y débito, reduciendo hasta un 50% los actos fraudulentos.

También han mejorado su estrategia comercial utilizando **precios dinámicos** que cambian cada 2 minutos, en comparación a otros minoristas que tardan semanas y meses para actualizar sus precios.

Introducción a DataWarehouse

Casos de éxito - Staples



Procesamiento de data y pronósticos de ventas

Staples, es otro caso exitoso del **Business Intelligence**, esta gran cadena minorista que tiene presencia en América latina, maneja aproximadamente 10 millones de transacciones por semana, lo que anteriormente resultaba muy difícil de manejar. Aprovechando esta tecnología han logrado lanzar **campañas de marketing dirigidas y pronosticar sus ventas diarias**.

Introducción a DataWarehouse

Casos de éxito - CVS



Gestión de Inventario

El retailer del área farmacéutica **CVS**, con casi 1000 tiendas en los Estados Unidos de América, experimentó problemas en cuanto a los artículos que vendía, los cuales presentaban errores en cuanto a su clasificación debido a la mala gestión del inventario.

A través del Business Intelligence y el big data, se hizo posible **rastrear los movimientos de productos en el inventario** utilizando para ello 160 indicadores clave, lo que permitió una **gestión más eficiente de inventario**.

Introducción a DataWarehouse

Casos de éxito - Waitrose



Previsión de Ventas

Waitrose, es la cadena de supermercados más grande del Reino Unido, pero su demanda varía según la temporada, lo que representó un desafío para la compañía que pudo resolver a través del business intelligence.

Al integrar esta solución se facilita (tanto en la tienda física como en su plataforma online) el **identificar las demandas a nivel de cada tienda y pronosticar qué productos** deben tener disponibles en función de las necesidades del cliente en cada temporada.

Introducción a DataWarehouse

Casos de éxito - Tesco



Marketing Personalizado

Este caso de éxito de business intelligence corresponde al multinacional **Tesco**, la cual se caracteriza por adoptar tecnologías de vanguardia para mejorar la calidad de sus servicios.

Pensando en optimizar sus campañas de marketing personalizadas, implementaron el Big Data a su sistema obteniendo un incremento en el uso de sus tarjetas de crédito propias y en el uso de cupones distribuidos a sus clientes mediante el e-mail marketing.

Introducción a DataWarehouse

Ejemplos prácticos / Casos de Éxito

Dudas / Preguntas



Introducción a DataWarehouse

Contenido de la Unidad

- **Conceptos, motivación y marco histórico.**
- **Utilidad y funciones.**
- **Generación y construcción de DW: ETL**
- **Tipos de DataWarehouse.**
- **Comparativa con Bases de Datos Relacionales.**
- **Ejemplos prácticos / Casos de Éxito.**
- **Resumen.**

Introducción a DataWarehouse

Resumen de la Unidad

Durante esta unidad hemos aprendido:

- La evolución de los sistemas de información y cómo coexisten en una organización cualquiera.
- Algunos problemas típicos que las organizaciones modernas enfrentan al trabajar con datos.
- Conceptos y definiciones de Business Intelligence, Data Warehouse y Data Warehousing.
- Esquemas resumidos de arquitectura que las empresas construyen o podrían construir; sus diferencias y ventajas.

Introducción a DataWarehouse

Resumen de la Unidad

Durante esta unidad hemos aprendido:

- Approaches, Arquitecturas y Metodologías de DW difundidas mundialmente.
- Pasos comunes de diferentes metodologías para la construcción de un DW.
- Algunos roles críticos en proyectos de DW.
- Resumen de metodologías ágiles aplicables a proyectos de DW.

Introducción a DataWarehouse

Bibliografía Recomendada



- **Inmon, W. H., Imhoff Claudia and Sousa Ryan.** The Corporate Information Factory. 2nd ed. John Wiley and Sons, 2000. Print.
- **Kimball, Ralph, and Ross Margy .** The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3d ed. Wiley, 2013. Print.
- **Laudon, Kenneth, and Laudon, Jane.** Sistemas de Información Gerencial. 14th ed. Pearson, 2016. Print

