



Modulo DWH

Diseño de una solución de BI



■ Módulo Data Warehouse

- Acerca del profesor
- Conceptos BI
 - Definiciones y arquitectura general
 - Diseño DWH: Kimball vs Inmon
 - Software
- Caso práctico
 - Entender el problema de negocio
 - Diseño del modelo de datos
 - Diseño de los procesos ETL
 - Kettle
 - Diseño de la solución





■ Acerca del profesor



■ Acerca del profesor

<https://es.linkedin.com/in/israelvizcainolopez>



Israel Vizcaíno López

Más de 500
contactos

Responsable de Inteligencia de Negocio en
Self Bank

Madrid y alrededores, España
| Servicios y tecnologías de la información

Actual

Self Bank

Anterior

Pragsis, AGS - Alpama Global Services,
Tecnocom

Educación

Universidad Complutense de Madrid

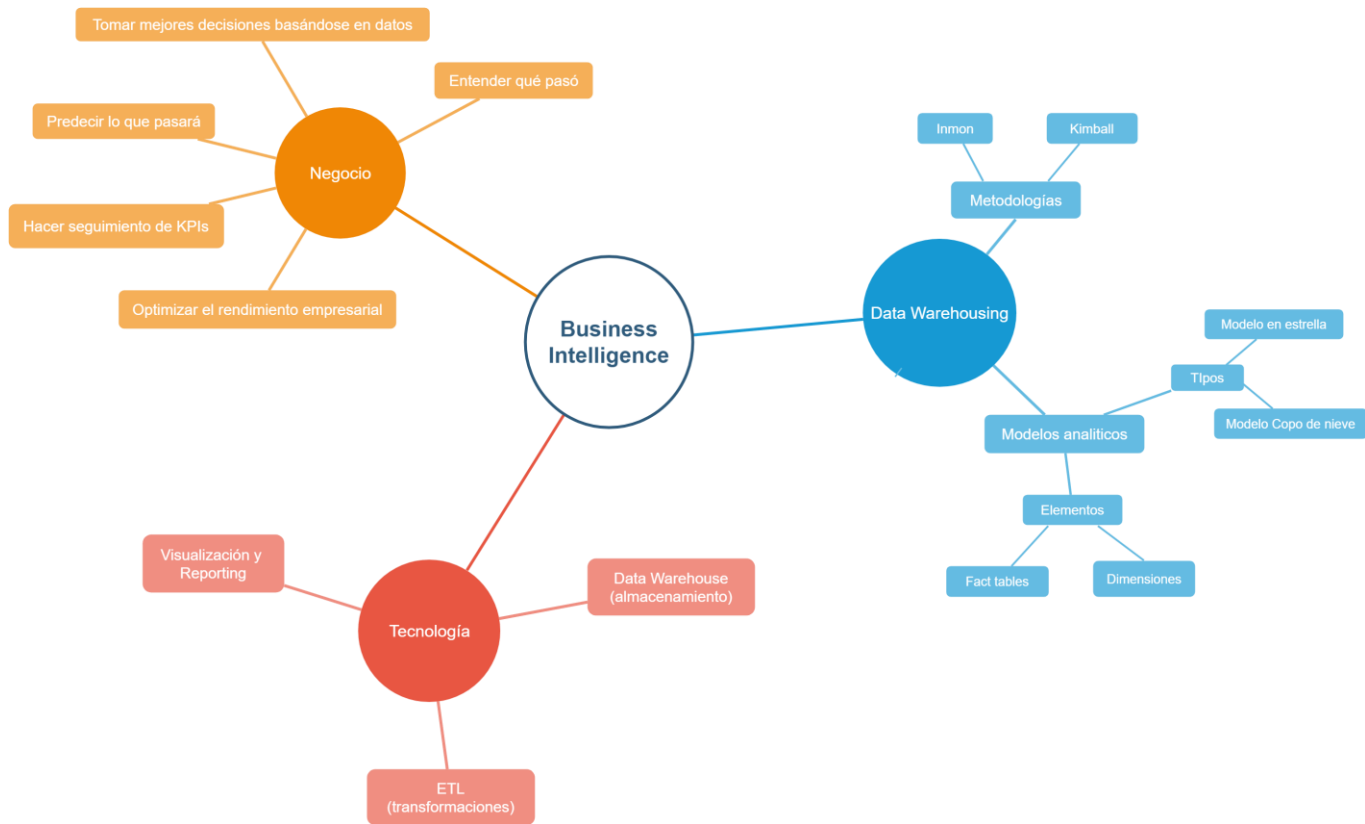




■ Conceptos BI



Mind Map



■ Definiciones BI: Sistemas OLTP vs OLAP

- **OLTP (On-line Transaction Processing)** is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second. In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).
- **OLAP (On-line Analytical Processing)** is characterized by relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems a response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques. In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

<https://www.datawarehouse4u.info/OLTP-vs-OLAP.html>



■ Definiciones BI

<https://tdwi.org/pages/topics/analytics-data-science-business-intelligence.aspx>

The technologies, techniques, and algorithms for analyzing all kinds of data to derive insights and take action for better decision making and enterprise success.

<https://www.gartner.com/it-glossary/business-intelligence-bi/>

Business intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance.



■ Definición ETL

https://es.wikipedia.org/wiki/Extract,_transform_and_load

Extract, Transform and Load (“extraer, transformar y cargar”, frecuentemente abreviado **ETL**) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.



■ Extracción de datos (E)

Obtener información de diversos orígenes de datos:

- Excel
- Ficheros texto plano (csv, txt, tsv, etc.)
- Bases de datos relacionales
- Sistemas NoSQL
- Ecosistema Hadoop
- Aplicaciones Cloud (APIs)

Pensad siempre en impactar lo menos posible al sistema origen.



■ Transformación de datos (T)

Realizar una secuencia de acciones sobre los datos obtenidos en la fase anterior:

- Analizar
- Validar
- Limpiar
- Transformar
- Enriquecer

Puede que, en esta fase, decidamos no ejecutar la última (L), si detectamos errores importantes en los datos.



■ Carga de datos (L)

Volcar los datos modificados en la fase anterior en algún sistema de almacenamiento:

- Excel
- Ficheros texto plano (csv, txt, tsv, etc.)
- Bases de datos relacionales
- Sistemas NoSQL
- Ecosistema Hadoop
- Aplicaciones Cloud (APIs)

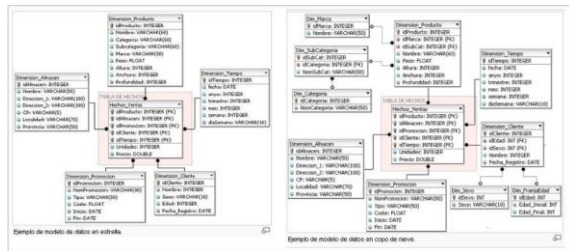
Los datos finales quedan limpios, completos, enriquecidos o agregados como se necesite para la capa de explotación.



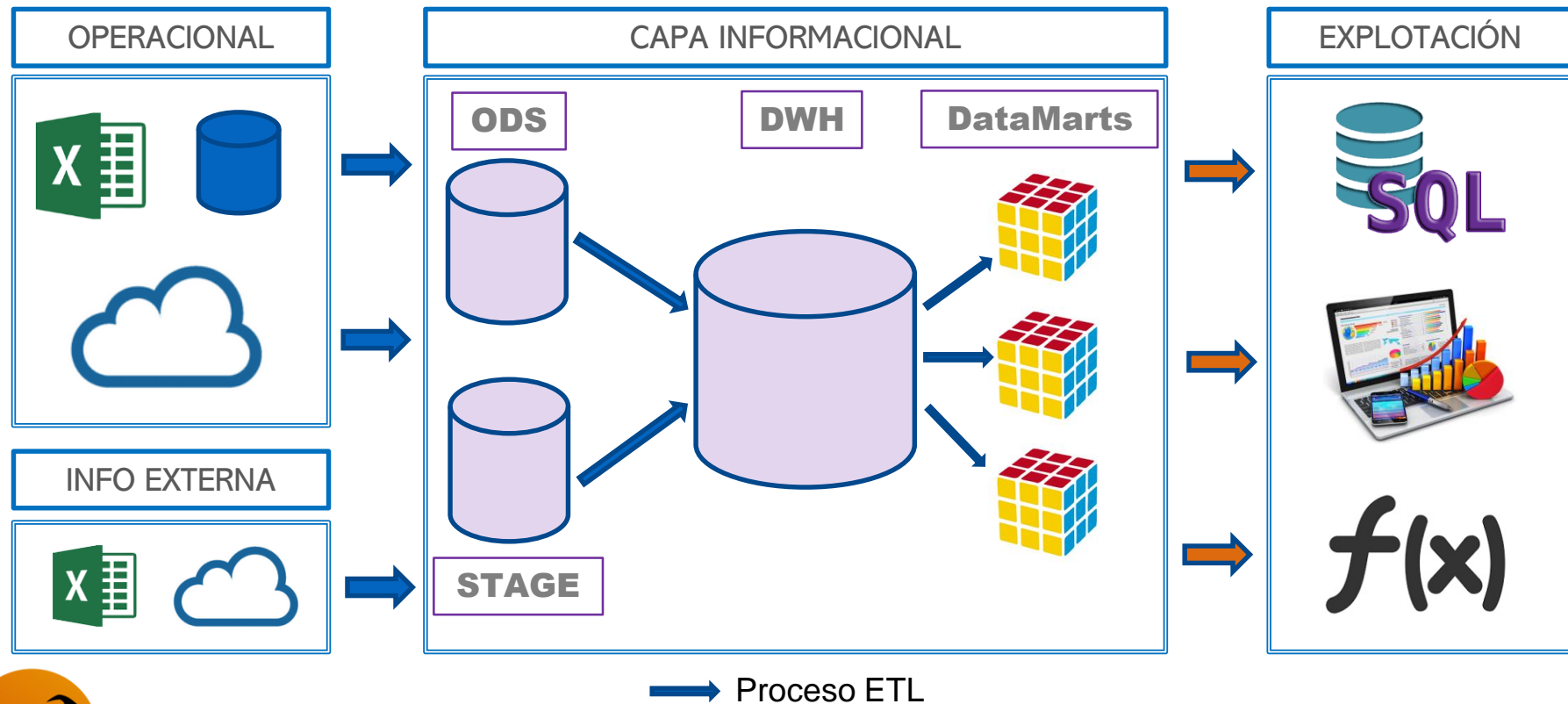
Definiciones BI

Conceptos modelos analíticos:

- Tablas de hechos (son las medidas del proceso de negocio)
 - Aditivas, Semiaditivas, No aditivas
 - Fotos periódicas, fotos acumulativas
 - Factless tables
- Tablas de dimensiones (los atributos a través de los cuales queremos explicar las medidas)
 - Dimensiones estáticas, dimensiones que cambian en el tiempo (SCD)
 - Dimensiones conformadas
- Modelos analíticos
 - Modelo en estrella
 - Modelo copo de nieve



Arquitectura BI



Kimball vs Inmon

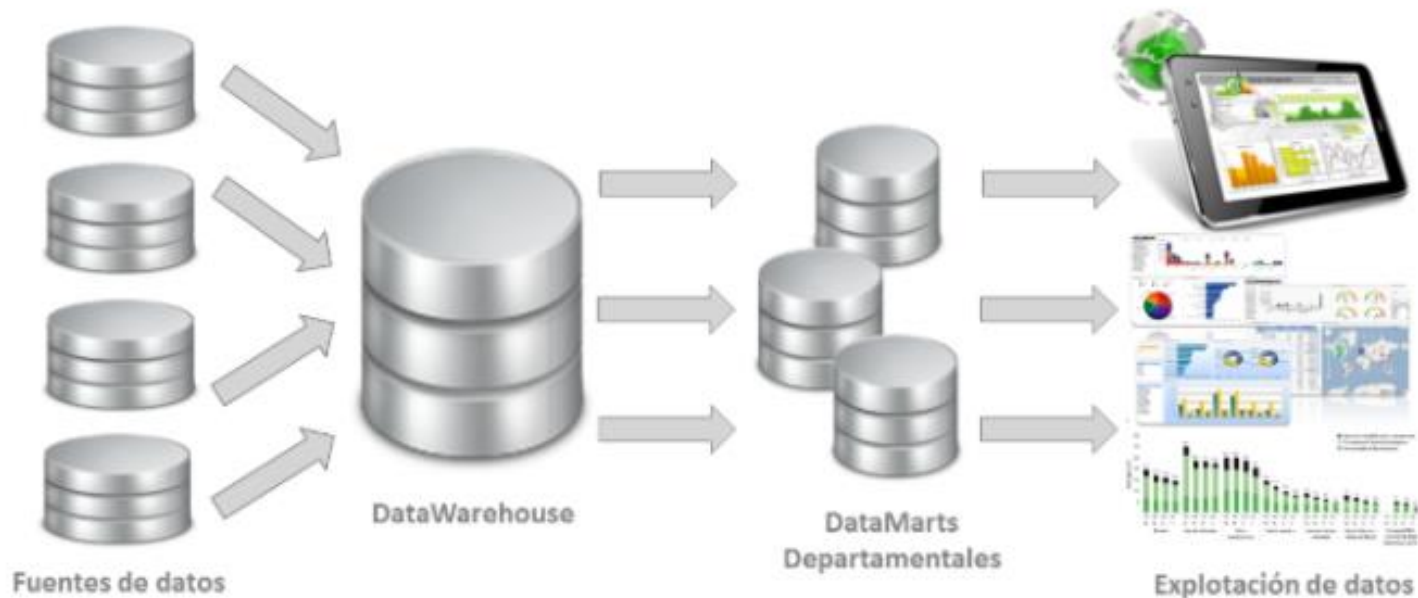
	Inmon	Kimball
Presupuesto	Coste inicial alto	Coste inicial bajo
Plazos	Requiere más tiempo de desarrollo	Tiempo de desarrollo inferior
Expertise	Equipo con especialización alta	Equipo con especialización media
Alcance	Toda la compañía	Departamentos individuales
Mantenimiento	Fácil mantenimiento	Mantenimiento más complejo

<https://blog.bi-geek.com/arquitectura-comparativa-inmon-y-kimball/>

<http://sistemasdecisionales.blogspot.com/2006/11/kimball-1-inmon-1-usabilidad-vs.html>



■ Diseño DWH según Inmon (Top down)



■ Diseño DWH según Kimball (Bottom up)



<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/>



Software BI: plataformas BI

A conocer:

- Microsoft Power BI
- Tableau
- Qlik
- Tibco Spotfire
- MicroStrategy
- Looker
- Chartio

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Software BI: soluciones ETL

A conocer:

- Informatica Powercenter
- IBM DataStage
- Talend
- Pentaho (Hitachi Vantara)
- SSIS (SQL Server)
- Denodo (virtualización)
- Alteryx

Figure 1. Magic Quadrant for Data Integration Tools



Software BI: DWH

A conocer:

- Teradata
- Oracle exadata
- Amazon Redshift
- Google BigQuery
- Azure Datawarehouse
- Snowflake
- Cloudera (Hadoop)

Figure 1. Magic Quadrant for Data Management Solutions for Analytics





■ Caso práctico



■ Entender el problema de negocio

Adventure Works, Inc. es un mayorista de productos de bicicletas ficticio que vende y fabrica bicicletas y vende ropa y otros accesorios relacionados a otros distribuidores en diferentes zonas geográficas.

Las personas en la compañía realizan actividades empresariales habituales como:

- Ventas
- Configuración de ventas y otros objetivos de la empresa
- Cumplimiento
- Administración de inventario
- Administración de comercio de muestra
- Administración de personal
- Presupuestos



■ Entender el problema de negocio

Se pone en contacto con nosotros el director del departamento de ventas porque actualmente le es muy costoso obtener la información que necesita para hacer seguimiento de cómo va su negocio y necesita que implantemos una solución de BI para dar respuesta a sus necesidades de información.

Nos comenta inicialmente que quiere ver datos y evolución de las ventas y entender el comportamiento de determinados productos, clientes y vendedores.

No tiene mucho presupuesto, pero quiere algo que pueda poner en marcha rápidamente.



■ Entender el problema de negocio

¿Por dónde empezamos?

Con lo poco que sabemos, ¿Inmon o Kimball?



■ Entender el problema de negocio

Tendremos que tener una reunión inicial donde definiremos el listado de preguntas que quiere responder con el objetivo de sacar información de los hechos y dimensiones que necesitamos.



■ Entender el problema de negocio

Visión de evolución global

1. Evolución del importe total de ventas y márgenes en el tiempo, pudiendo verlo de manera anual, trimestral, mensual, semanal, diario u horario. Se debe poder ver esta evolución de manera global o desglosada por categoría/subcategoría de producto.



■ Entender el problema de negocio

Análisis de producto

2. Cuáles es el top de productos de los que he vendido más unidades en un cierto período de tiempo. Cuáles me han generado más ingresos. Cuáles me han dejado más margen. Todo visto globalmente y por región, por categoría, subcategoría, línea del producto y nivel de calidad.
3. Para un producto concreto, ¿cuál es la evolución en el tiempo de la distribución de ventas por color, talla y estilo (si es para hombres, mujeres o universal)? Globalmente y por región en caso de ventas.
4. Para un producto concreto, ¿qué otros productos suelen comprarse al mismo tiempo?



■ Entender el problema de negocio

Análisis de cliente minorista

5. Perfil de cliente minorista: distribución por sexo, estudios, ocupación, nivel de ingresos y geografía (estado, ciudad, código postal) por año de primera compra.
6. Evolución de la distribución de las ventas minoristas totales y el margen generados por mujeres vs hombres y por grupo editario (desde los 15 de 10 en 10).
7. ¿Cuántos clientes me han comprado una sola vez? ¿Cuántos repiten?
¿Cuántas veces al año me compra un cliente?
8. ¿Cuánto es el ticket medio por cliente? ¿Cuál es la distribución por número de productos? ¿Cuál es la relación entre ambos? ¿Ticket medio hombre vs mujer?
¿Y por grupo editario?



■ Entender el problema de negocio

Análisis de vendedor

- 9.Cuál es el top 3 de vendedores en cada región/territorio por trimestre según el importe total de ventas y según margen.
10. Qué vendedores están por debajo del ticket mediano en el último período (mes o trimestre).
11. ¿Hay diferencias significativas entre ventas generadas por vendedores según su sexo y grupo editario?
12. Para un producto concreto, ¿se reparten sus ventas de igual forma entre todos los vendedores?
13. Para un vendedor concreto, ¿qué tipos de producto (categoría, subcategoría, línea y clase) son los que más vende? ¿Está alineado con la distribución de ventas totales de la empresa por esos mismos parámetros?



■ Entender el problema de negocio

Análisis de fecha

15. Cual es la distribución del número de clientes comprando por mes, día de la semana y hora.
16. Encontrar qué fechas tienen ventas significativamente mayores del resto.
17. Encontrar qué fechas tienen ventas significativamente menores del resto.



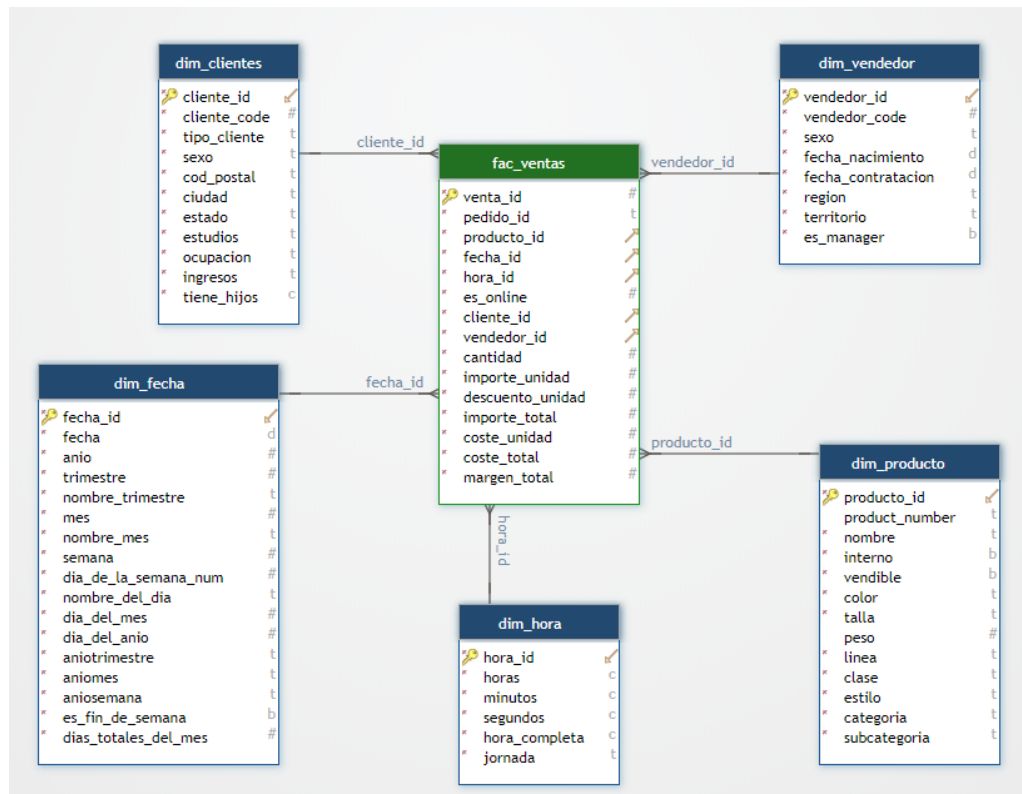
EJERCICIO

Con esta información, tenemos que definir dimensiones y hechos.

Lo hacemos en una hoja Excel.



■ Diseño del modelo de datos analítico



■ Diseño de los procesos ETL

Una vez modelizado el DWH, tenemos que:

- Bucear en los modelos de datos operacionales para mapear la información que necesitamos cargar en el modelo analítico.
- Comprobar la calidad de los datos procedente de los modelos operacionales y tomar medidas si no es suficiente:
 - ¿Tenemos todos los datos que necesitamos?
 - ¿Qué porcentajes de nulos tenemos en cada campo? ¿Son razonables?
 - ¿El formato es adecuado? ¿Hay valores que no tengan sentido?
- Definir las transformaciones que necesitan los datos.
- Implementar el proceso ETL en la herramienta seleccionada.



EJERCICIO

Hay que localizar cada campo de cada tabla del DWH de donde los podemos obtener en la BBDD operacional de AdventureWorks.

Lo hacemos en la misma hoja Excel.



CONCLUSION

Esta es una tarea difícil, por regla general.

- Las empresas no suelen tener buena documentación de los modelos operacionales.
- El conocimiento recae sobre las personas.
- Necesitaremos ayuda de expertos técnicos y del negocio.



■ Nivel de calidad de los datos

EJERCICIO

Mirad en las tablas localizadas del operacional:

- Qué porcentaje de nulos hay.
- Si responden a una lógica de negocio.
- Si hay valores extraños o atípicos.
- Decidir si lo anómalo:
 - se transforma y se carga
 - se descarta y se informa.



■ Definir las transformaciones a realizar

EJERCICIO

Para cada campo definido en el modelo analítico:

- ¿Se puede cargar tal cual viene de origen?
- ¿Hay que hacer alguna transformación?
- ¿Es un campo que hay que calcular a través de otros?

Lo ideal es dejarlo documentado en un análisis previo.



■ Desarrollar los procesos ETL

Hemos elegido Pentaho Data Integration para la implementación de los procesos ETL.

- Leeremos los datos de la base de datos PostgreSQL operacional de AdventureWorks (ejercicio: diseñar consultas).
- Realizaremos las transformaciones oportunas sobre la herramienta.
- Volcaremos los datos en la base de datos PostgreSQL que hemos elegido como DWH (AdventureWorksDWH).





Kettle

Pentaho Data Integration

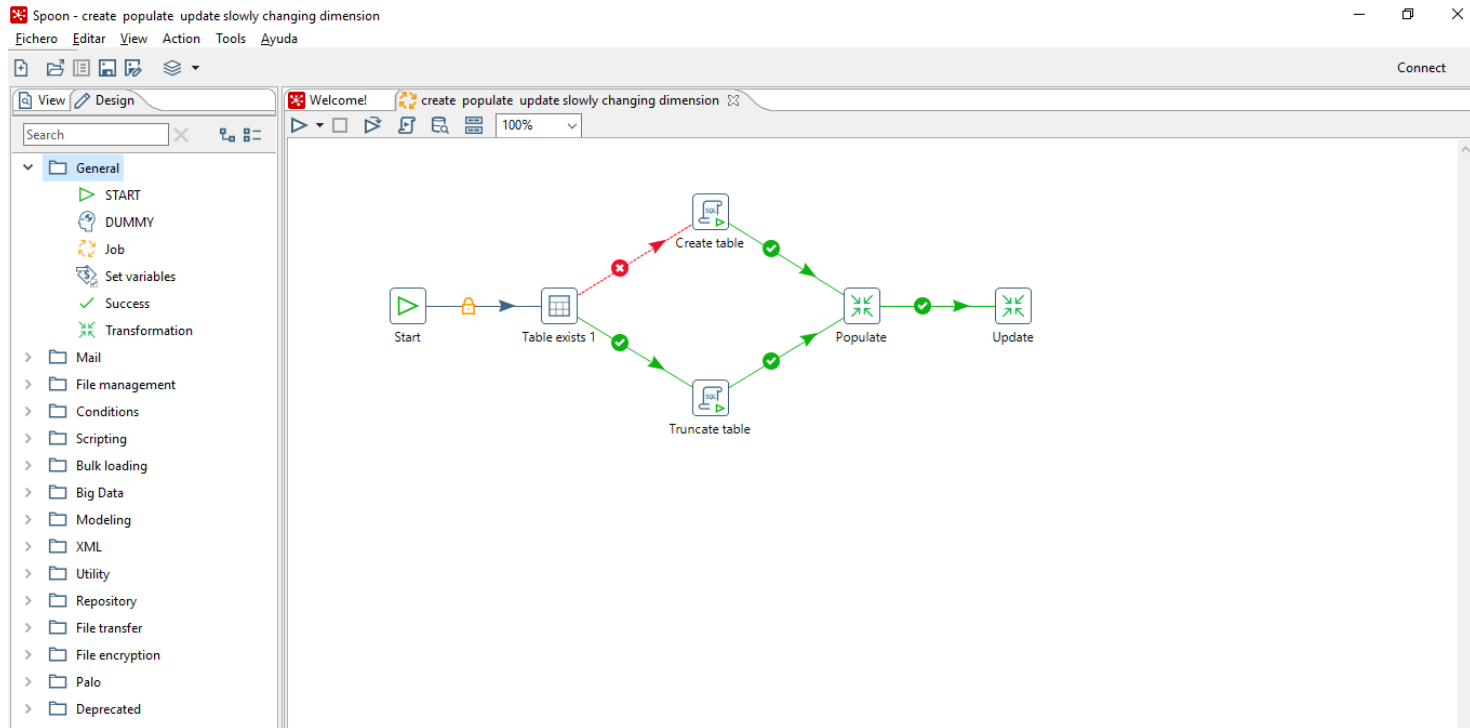


■ Elementos de Kettle

- Spoon:
 - entorno gráfico (IDE) de desarrollo de ETLs.
- Pan:
 - motor de transformación de datos.
- Kitchen:
 - programa para ejecutar los trabajos diseñados con spoon.



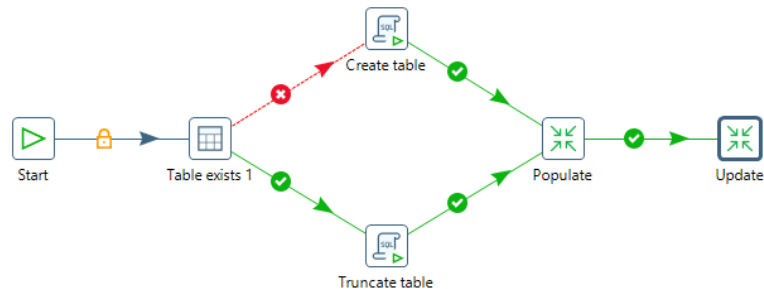
Entorno Spoon



Trabajos y transformaciones

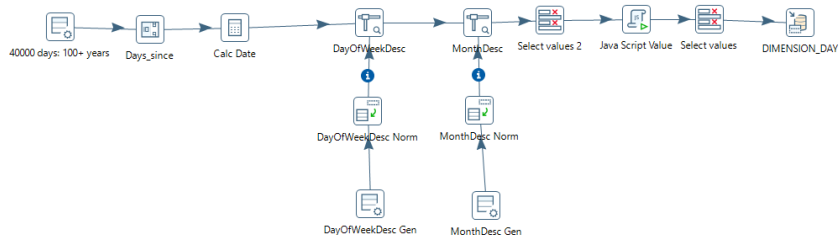
Trabajos: 

- Diseño de alto nivel
- Control del flujo del proceso
- Pueden ejecutar otros Jobs o transformaciones



Transformaciones: 

- Operativa de bajo nivel
- Acciones con los datos



Trabajos: utilidades principales

General

- START
- DUMMY
- Job
- Set variables
- Success
- Transformation

Mail

- Get mails (POP3/IMAP)
- Mail
- Mail validator

File management

- Add filenames to result
- Compare folders
- Convert file between Windows and Unix
- Copy Files
- Create a folder
- Create file
- Delete file
- Delete filenames from result
- Delete files
- Delete folders
- File Compare
- HTTP
- Move Files
- Process result filenames
- Unzip file
- Wait for file
- Write to file
- Zip file

Conditions

- Check Db Connections
- Check files locked
- Check if a folder is empty
- Check webservice availability
- Checks if files exist
- Columns exist in a table
- Evaluate files metrics
- Evaluate rows number in a table
- File Exists
- Simple evaluation
- Table exists
- Wait for

Scripting

- JavaScript
- Shell
- SQL



Trabajos: utilidades principales

Utility

- Abort job
- Display Msgbox Info
- HL7 MLLP Acknowledge
- HL7 MLLP Input
- Ping a host
- Send information using Syslog
- Send Nagios passive check
- Send SNMP trap
- Talend Job Execution
- Telnet a host
- Truncate tables
- Wait for SQL
- Write To Log

Big Data

- Amazon EMR Job Executor
- Amazon Hive Job Executor
- Google BigQuery Loader
- Hadoop Copy Files
- Hadoop Job Executor
- Oozie Job Executor
- Pentaho MapReduce
- Pig Script Executor
- Spark Submit
- Sqoop Export
- Sqoop Import

Bulk loading

- BulkLoad from Mysql into file
- BulkLoad into MSSQL
- BulkLoad into Mysql

File transfer

- FTP Delete
- Get a file with FTP
- Get a file with FTPS
- Get a file with SFTP
- Put a file with FTP
- Put a file with SFTP
- Upload files to FTPS

Repository

- Check if connected to repository
- Export repository to XML file

Modeling

- Build Model
- Publish Model



Transformaciones: utilidades principales

▼ Entrada

- CSV file input
- Data Grid
- Des-Serialization desde Fichero
- Email messages input
- Entrada Excel
- Entrada Fichero de Texto
- Entrada Tabla
- Entrada XBase
- Fixed file input
- GZIP CSV Input
- Generar Filas
- Generate random credit card numbers
- Generate random value

- Get Files Rows Count
- Get SubFolder names
- Get repository names
- Get table names
- Google Analytics
- HL7 Input
- Información de Sistema
- LDAP Input
- LDIF Input
- Load file content in memory
- Microsoft Access Input
- Mondrian Input
- OLAP Input
- Obtener nombres de fichero
- Property Input
- RSS Input
- SAS Input
- Salesforce Input
- Yaml Input

▼ Scripting

- Ejecutar script SQL
- Execute row SQL script
- Formula
- Regex Evaluation
- Rules Accumulator
- Rules Executor
- User Defined Java Class
- User Defined Java Expression
- Valor Java Script Modificado

▼ Streaming

- Get records from stream
- JMS Consumer
- JMS Producer
- Kafka Consumer
- Kafka Producer
- MQTT Consumer
- MQTT Producer



Transformaciones: utilidades principales

Transformar

- Add a checksum
- Add value fields changing sequence
- Añadir constantes
- Añadir secuencia
- Calculadora
- Closure Generator
- Concat Fields
- Des-Normalización de Fila
- Filas Unicas
- Get ID from slave server
- Mapeo de Valores
- Normalización de Fila
- Number range
- Ordenar filas

Partir campos

Replace in string

Row flattener

Selecciona/Renombra valores

Set field value

Set field value to a constant

Split field to rows

String operations

Strings cut

Unique rows (HashSet)

Uniones

- Fundir filas
- Juntar Filas (producto cartesiano)
- Multiway Merge Join
- Unión Ordenada
- Unión por Clave

Salida

- Actualizar
- Automatic Documentation Output
- Eliminar
- Insertar / Actualizar
- LDAP Output
- Microsoft Excel Writer
- Pentaho Reporting Output
- Properties Output
- RSS Output
- S3 File Output
- SQL File Output
- Salesforce Delete
- Salesforce Insert
- Salesforce Update
- Salesforce Upsert
- Salida Access
- Salida Excel
- Salida Fichero de Texto
- Salida Tabla
- Serialización a Fichero
- Synchronize after merge
- Text file output (legacy)



Transformaciones: utilidades principales

Flow

- Abort
- Annotate Stream
- Append streams
- Block this step until steps finish
- Blocking Step
- Detect empty stream
- ETL Metadata Injection
- Filtrar filas
- Identify last row in a stream
- Java Filter
- Job Executor
- Prioritize streams
- Shared Dimension
- Single Threader
- Switch / Case
- Transformación Simulada (no hace nada)
- Transformation Executor

Utility

- Change file encoding
- Clone row
- Delay row
- Edi to XML
- Execute a process
- If field value is null
- Mail
- Metadata structure of stream
- Nulo si...
- Process files
- Run SSH commands
- Send message to Syslog
- Table Compare
- Write to log
- Zip file

Big Data

- Avro Input
- Avro Output
- Cassandra Input
- Cassandra Output
- CouchDb Input
- HBase Input
- HBase Output
- HBase Row Decoder
- Hadoop File Input
- Hadoop File Output
- MapReduce Input
- MapReduce Output
- MongoDB Input
- MongoDB Output
- ORC Input
- ORC Output
- Parquet Input
- Parquet Output
- SSTable Output



Transformaciones: utilidades principales

Trabajo

- Asignar valor a variables
- Copia filas a resultado
- Copia nombres de fichero al resultado
- Obtener Variables
- Obtener ficheros de resultado anterior
- Obtener filas de resultado anterior

Cryptography

- PGP Decrypt stream
- PGP Encrypt stream
- Secret key generator
- Symmetric Cryptography

Bulk loading

- ElasticSearch Bulk Insert
- Greenplum Load
- Infobright Loader
- Ingres VectorWise Bulk Loader
- MonetDB Bulk Loader
- MySQL Bulk Loader
- Oracle Bulk Loader
- PostgreSQL Bulk Loader
- Teradata Fastload Bulk Loader
- Teradata TPT Bulk Loader
- Vertica Bulk Loader

Almacén de Datos

- Búsqueda/Actualización en Combinación
- Búsqueda/Actualización en Dimensión

Statistics

- Agrupar por
- Analytic Query
- Memory Group by
- Output steps metrics
- Reservoir Sampling
- Sample rows
- Univariate Statistics

Validation

- Credit card validator
- Data Validator
- Mail Validator
- XSD Validator





■ Ejemplos



■ Ejemplos: (E) Extracción de datos



files



output



01 CSV Input - Reading customer data.ktr



02 CSV Input - Reading customer data with error logging.ktr



03 Excel Input - XLS - XLSX - ODS reading.ktr



04 Fixed Input - fixed length reading .ktr



05 JsonInput - read a file.ktr



06 JSON - read nested fields.ktr



07 JsonInput - read a dynamic file.ktr



08 JsonInput - read incoming stream.ktr



09 HTTP Client - simple retrieval example.ktr



10_Lectura_BBDD.kjb



10_Lectura_BBDD.ktr



11_Lectura_MongoDB.ktr



Entrada Excel



CSV file input



Entrada Fichero de Texto



Entrada Tabla



Json input



MongoDB Input



Hadoop File Input



HBase Input



■ Ejemplos: (T) Transformaciones de datos

- Validaciones (BBDD, ficheros y datos)
- Generación de datos
- Cálculos
 - Numéricos
 - Fechas
 - Cadenas
- Filtrado de registros
- Pivotar filas columnas
- Agregaciones y queries analíticas
- Específicos Datawarehouse
- Ficheros

- 00 Block this step until steps finish - basic example - dialog.ktr
- 01 01 01 Validaciones BBDD Jobs.kjb
- 01 01 02 Validaciones ficheros Jobs.kjb
- 01 01 Validaciones.ktr
- 01 02 Validar datos contrastando otra fuente.ktr
- 01 03 Data Validator - all usecases with error handling.ktr
- 01 04 Check if a column exists - Basic example.ktr
- 01 05 Comprobar si existe una tabla.kjb
- 01 06 Comprobar si existe una columna.kjb
- 02 01 Add a sequence - Basic example.ktr
- 02 02 Add sequence - specify a common counter.ktr
- 02 03 Add value fields changing sequence - basic example.ktr
- 02 04 Append streams - Basic example.ktr
- 02 05 Clone row - Basic example.ktr
- 03 01 Calculadora - Operaciones con Numeros.ktr
- 03 02 Calculadora - Operaciones con Fechas.ktr
- 03 03 Operaciones con Strings.ktr
- 03 04 Replace in string - Simple example.ktr
- 03 05 Calculos con Formula.ktr
- 04 01 Expresiones_Regulares.ktr
- 04 02 01 Process changelog.kjb
- 04 02 02 Change log processing.ktr



■ Ejemplos: (L) Carga de datos

- files
- output
- 01 Salida Excel.ktr
- 02 Salida Fichero.ktr
- 03 Salida Json.ktr
- 04 Salida MongoDB.ktr
- 05 Carga BBDD.kjb
- 05 Carga BBDD.ktr
- 06 Borra Tabla BBDD.kjb



Salida Excel



Salida Fichero de Texto



Salida Tabla



Microsoft Excel Writer



Json Output



MongoDB Output



Hadoop File Output



HBase Output



■ Buenas prácticas

- Diseña el proceso pensando en los posibles fallos que puedan ocurrir e introduce controles (falta de conectividad, ficheros que no llegan, tablas o campos inexistentes).
- Envía notificaciones cuando el proceso no termine correctamente.
- Divide tu proceso en tareas sencillas siempre que se pueda.
- Permite que se pueda ejecutar desde un punto de fallo sin tener que procesar pasos anteriores.
- Documenta lo mejor posible los procesos.





■ Diseño de la solución

Dar respuestas a las preguntas planteadas por el cliente



Visión de evolución global

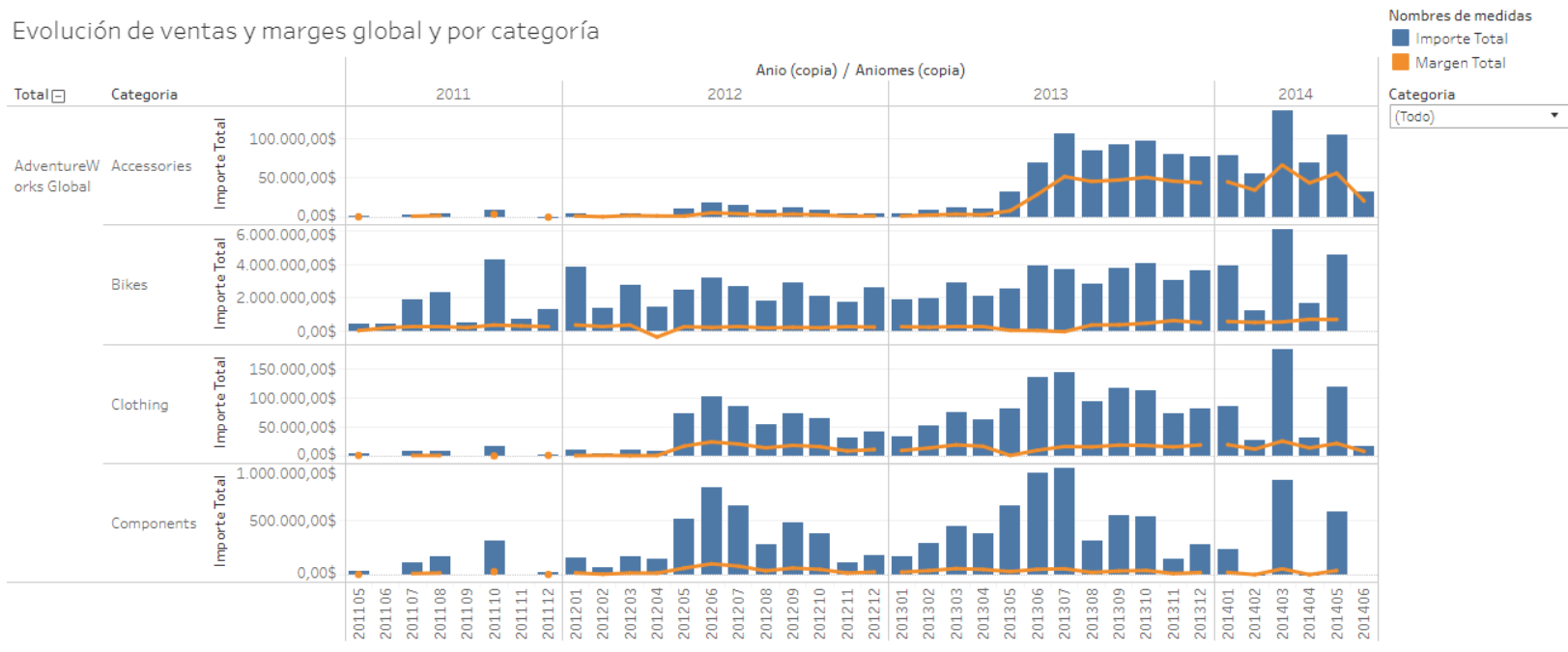
1. Evolución del importe total de ventas y márgenes en el tiempo, pudiendo verlo de manera anual, trimestral, mensual, semanal, diario u horario. Se debe poder ver esta evolución de manera global o desglosada por categoría/subcategoría de producto, así como por región/territorio de venta.



Solución BI

Visión de evolución global

Evolución de ventas y márgenes global y por categoría





■ Diseño de la solución

Realizar como ejercicio las consultas sobre el modelo analítico o los cuadros de mando necesarios para dar respuesta a las preguntas de negocio de nuestro cliente.



■ Solución BI

Análisis de producto

2. Cuáles son los diez productos de los que he vendido más unidades en un cierto período de tiempo. Cuáles me han generado más ingresos. Cuáles me han dejado más margen. Todo visto globalmente y por categoría, subcategoría, línea del producto y nivel de calidad.
3. Para un producto concreto, ¿cuál es la evolución en el tiempo de la distribución de ventas por color, talla y estilo (si es para hombres, mujeres o universal)? Globalmente y por región.
4. Para un producto concreto, ¿qué otros productos suelen comprarse al mismo tiempo?



Análisis de cliente minorista

5. Perfil de cliente minorista: distribución por sexo, estudios, ocupación, nivel de ingresos y geografía (estado, ciudad, código postal) por año de primera compra.
6. Evolución de la distribución de las ventas totales y el margen generados por mujeres vs hombres y por grupo editario (desde los 15 de 10 en 10).
7. ¿Cuántos clientes me han comprado una sola vez? ¿Cuántos repiten?
¿Cuántas veces al año me compra un cliente?
8. ¿Cuánto es el ticket medio por cliente? ¿Cuál es la distribución por número de productos? ¿Cuál es la relación entre ambos? ¿Ticket medio hombre vs mujer?
¿Y por grupo editario?



■ Solución BI

Análisis de vendedor

- 9.Cuál es el top 3 de vendedores en cada región/territorio por trimestre según el importe total de ventas y según margen.
10. Qué vendedores están por debajo del ticket mediano en el último período (mes o trimestre).
11. ¿Hay diferencias significativas entre ventas generadas por vendedores según su sexo y grupo editario?
12. Para un producto concreto, ¿se reparten sus ventas de igual forma entre todos los vendedores?
13. Para un vendedor concreto, ¿qué tipos de producto (categoría, subcategoría, línea y clase) son los que más vende? ¿Está alineado con la distribución de ventas totales de la empresa por esos mismos parámetros?



■ Entender el problema de negocio

Análisis de fecha

15. Cual es la distribución de clientes comprando por mes, día de la semana y hora.
16. Encontrar qué fechas tienen ventas significativamente mayores del resto.
17. Encontrar qué fechas tienen ventas significativamente menores del resto.





GRACIAS

www.keepcoding.io

