

## Introducción al Business Intelligence – 1. 8/1 pág 7

¿Para qué sirve BI?

Toma dediciones de negocio.

Dar valor agregado al dato.

Tener repositorio de dato ordenado y organizado (normalizado).

Pág 8 se la salta

## Introducción al Business Intelligence – 1. Pág 9 rápid

Davenport “Las empresas invierten en BI per generar una aventatge competitiva.”

Kimbal “Como dimensionar las estrucutras de datos para que puedan escalar y sean óptimas.”

## Introducción al Business Intelligence – 1. Pág 10

Ciclo virtuoso del dato. Como se van aportando valor a los datos.

Ejemplo. Empresa de venda de perfumes, aplicaciones o servicios. En cada operación de negocio genera datos. Los agrega y genera información al comprarlo con los otros y meterlo en contexto.

Objetivo final del BI: es crear conocimiento → generar conocimiento es entender porqué suceden las cosas.

El conocimiento provocará unas acciones (pivotar el negocio o variar algo) y eso creará nuevos datos.

## Introducción al Business Intelligence – 1. Pág 11

Zara. Es referencia en el fast fashion. Fueron muy rápidos en captar datos y monetizar. Mentre els competidors intentaven convèncer els consumidors per comprar lo seu, Zara intentava ser els primers en fer els productes que la gent vol.

Blackberry. Tenia moltes dades de l'appel i no eren capaços d'entendre perquè la gent no volia un teclat al mòbil. Van ser molt lents en girar la roda i entendre les dades.

Pág 12 rápid

Pág 13 se la salta

Pág 14

Zara volia ↑ la repetividad. Més compres. Baixant una mica el Ticket medio, el valor que es gasta de mitjana la gent.

Cogian datos → Lo diseñaban → Fabrica → Venda. De forma rápida que en 2 semana tenian el diseño que la gente quería. Si funcionaba se quedaba el diseño. Si no funcionaba, lo quitaban. Fa 8 anys ho va fer.

## Introducción al Business Intelligence – 1. Pág 15

¿Quines dades agafar?

Classe diu:

Xarxes socials

Motors de búsqueda. Búsquedes fallides.

Queixes i reclamacions

Compres que es fan

Vendes que es fan

Profe diu:

No hi havia tanta analítica de web. Van mirar els valors del tickets. Què son bons què no. Què se'ns compren i què no. Els responsables presencials de la botiga havien d'encarregar-se de parlar amb tots els staff de la botiga per apuntar el que els havien demanat i no tenien i registrar-ho.

Creuar dades del que es ven i què no VS el que es vol i el que no.

Es van adonar que hi havia un comportament molt estacional.

Fabricaven poc i el que veien que tenien èxit en feien més. Analitzant l'stock.

El que funciona perquè funciona. El que no funciona perquè no funciona.

Quan vols créixer, t'oblides una mica de les mètriques de devolucions, stock i rendiment. No et fixes en l'error i en optimitzar i fer eficiència. Quan les empreses diuen que volen menjar-se el món, s'endeuten i tiren cap a un objectiu clar sense fer massa cas al nivell tàctic, sinó que tiren cap a una estratègia.

Introducción al Business Intelligence – 1. Pág 16

ERP. Plataforma que opera una empresa: es generen les factures, s'operen clients, es gestiona la logística.

SCM. Cadena de suministro.

CRM. Gestió clients.

HRM. Gestió recursos humans.

Nivel táctico → Decisiones de curts plaços (1 mes o així). El BI sobretot cau a aquí. Normalment el big data està aquí.

Nivel estratégico → Decisiones a llarg plaç (1-2 anys). KPI més agregats, més centrats en una àrea (logística, venda o consumidor). Es gestionen de forma diferent.

Introducción al Business Intelligence – 1. Pág 17 salta

Pág 18 estava empanat i va a idees molt per sobre

Pág 19 va a idees molt per sobre

Pág 20

Riscs. Business Analyst: Expert que sap molt de finances o RRHH i sap algo de dates i fa de pont.

Aventatges. Pressupost independent. Amb aquell diner pot fer el que vol.

Pág 21 se la salta

Pág 22

Pág 23

Potencia el backend del BI. Es desprèn del frontend. Genera data pipelines. Hi ha dos perfils:

- Perfil SQL:

- Sap molt de SQL i algo de python

- Perfil NoSQL:

- Sistema Hadoop + Spark

- Avançat en llenguatges de programació com Java, Scala o Python

Es raro tenir un perfil mixt. Normalment els Data Engineers tenen un perfil o altra.

Quan va sortir Redshift,... Va donar molta més importància a necessitar analitzar les dades de SQL

Introducción al Business Intelligence – 1. Pág 24 no explica molt

### Introducción al Business Intelligence – 1. Pág 25

Eren empreses que tenien una plataforma comprada (Oracle, SAP) que es gestiona de forma visual. I tens consultors que necessites que t'ho gestionin. No són perfils tant de programar. Sap de SQL.

### Introducción al Business Intelligence – 1. Pág 26

No tant de construir datos. Sinó que un cop s'ha construït, fer anàlisis de datos. Sap de python, R. Explorar les dades. Fer histogrames. Usuari que consumeix les dades i l'explora. Fa presentacions. Story telling...

### Introducción al Business Intelligence – 1. Pág 27

No és perfil tècnic. Està basat en una àrea. Sap definir mètriques perquè entén el negoci. Fa propostes de projectes (hauríem d'anàlitzar això)...

### Introducción al Business Intelligence – 1. Pág 28

A on se sobreposen?

L'objectiu de data science i BI poden coincidir. Ambdós volen generar coneixement per l'empresa i crear oportunitats de negoci. Molts cops conviuen.

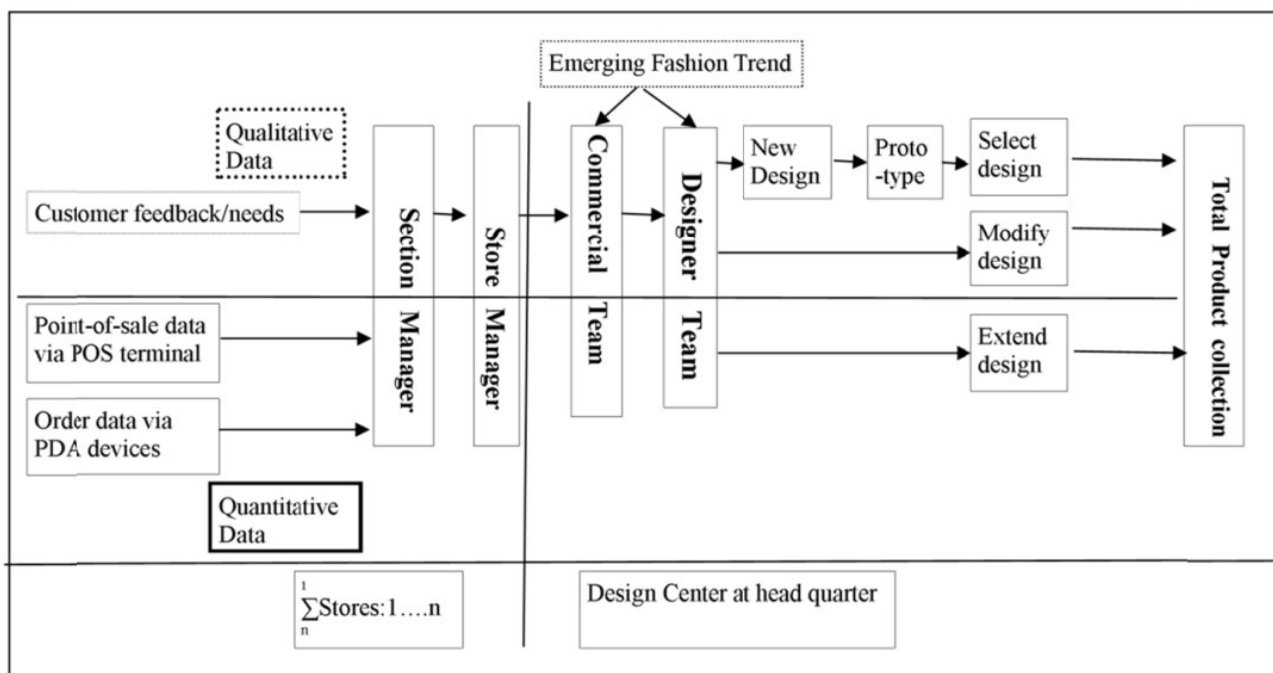
Molts cops el BI treballa les dades. Data Science agafen les dades i treballen a través d'algoritmes.

---

La pregunta de què analitzar normalment ve de l'àrea de negoci. BI connecta l'àrea de negoci amb l'execució de la intel·ligència de les dades.

---

Zara case of study



## Introducción al Business Intelligence – 2. Pág 5

**KPI.** Indicador de rendimiento que se usa para medir como se estaba antes y como se está ahora. Para comprarar. Indicador clave de desempeño. Permite medir si una decisión funciona o no. Es lo que permite definir si las cosas van bien o no → Reinforcement learning. Comparación contra año anterior o contra un plan o ambición.

La clave del Business analyst es definir bien los KPI. Definirlos bien te permitirá generar ese conocimiento.

Ejemplos de KPI

- EBITDA
- Churn rate. Cantidad de clientes que se han dado de baja / número totales de clientes
- CTOR % de usuarios que han clicado a un link concreto.

## Introducción al Business Intelligence – 2. Pág 6

Característiques de un KPI:

- Numèric amb unitat de mesura.
- Agregable → que es pot fer operació matemàtica.
- Hauria de ser comparable amb un objectiu (vs LY-*last year*-último año, BUD-*Budget*-presupost, PLAN-planificació)
- Tiene que estar bien definido. Definición consensuada y conocida en toda la unidad de negocio. Normalment és difícil posar-se d'acord en fer la mateixa definició entre les diferents àrees. Llavors és important documentar-la. Dir quina definició s'està fent.

## Introducción al Business Intelligence – 2. Pág 7

KPI molt bons o KPI bàsics. Es comencen sent bàsics i poc a poc es va creixent en maduresa.

Número de vendes i quantitat d'ingressos → Potser no dona molta informació

PPI → Quant ingreso per usuari vs quant m'ha costat l'usuari.

## Introducción al Business Intelligence – 2. Pág 8 molt ràpid

## Introducción al Business Intelligence – 2. Pág 9

Volent analitzar quant de bé va Glovo, quins KPI utilitzaríem?

- N.º de pedidos entregados. Comparado contra el mes anterior o contra otra región.
- Cantidad de entrada de personas que entran en la app / Cantidad de pedidos. → Mide la confianza del usuario.
- Número de pedidos no entregados debido a errores. → Mide eficacia del servicio.
- Cantidad de usuarios nuevos. Comparado con el mes anterior.
- Número de descargas de la app. Comparado con las ventas realizadas.
- Margen ganado por pedido. Comparado con el mes anterior.

## Introducción al Business Intelligence – 2. Pág 10

Dimensión o característica. Analizado desde distintos puntos de vista:

- Por tipología de producto
- Por país o región
- Por momento temporal

Podríamos encontrar por ejemplo que un tipo de producto va mejor en una región en concreto.

## Introducción al Business Intelligence – 2. Pág 11

Diferenciar què és dimesió de què és característica.

## Introducción al Business Intelligence – 2. Pág 12

KPI – Dimensión – Característica – Tipo de gráfico

N.º de pedidos entregados – tiempo – meses – Línea

Beneficio – producto – tipología

Número de usuarios únicos – Geográfico – Por país – Geomap

Introducción al Business Intelligence – 2. Pág 13

OLAP Online Analytical Processing

OLTP Sistema operativo. Transaccional

- Falta integración de la información. Pot passar que es vulgui consultar les dades del ecommerce ahora creuat amb les dades de la publicitat que es fan. Però com que les dades estan a llocs diferents, costa consultar-les ahora.

Introducción al Business Intelligence – 2. Pág 14a18

Cub OLAP.

Cada creuament es el valor que té el KPI en aquell valor.

MBD (2024) - Introducción al Business Intelligence – 2. 10/1 pág 21

Drill down consiste en seleccionar tipología de producto de forma jerárquica. O sea: Ropa hombre o mujer? Hombre. Torso, piernas o zapatos? Torso. Abrigos, chaquetas, sudaderas o camisetas? Sudaderas. XL, L, M, S, XS? M.

Roll up. Al tener un análisis, por ejemplo venta por meses, se puede haver un roll up y subir a venta por días.

Introducción al Business Intelligence – 2. Pág 22

Pregunta: Què és un sistema transaccional?

- Sistema transaccional. És per exemple la plataforma de Estudy de la Salle que digatilitza els exàmens i entregues i material. És qualsevol sistema que opera d'una empresa i li serveix d'ajuda per fer operacions de negoci. Relacionat amb les dades pot ser un sistema que agafi les dades i les presenti.

- Sistema analític. És un sistema que fa un anàlisis d'unes dades. Per exemple La Salle analitza quins graus o màsters tenen més èxit. Aquest sistema normalment està desacoplat del transaccional perquè si es penja una analítica no volem que es pengi el EStudy.

Estos sistemas pueden colgar de OLAP o OLTP. Són sistemas muy distintos.

Pregunta: Què és OLAP i OLTP?

- OLAP.
- OLTP.

Introducción al Business Intelligence – 2. Pág 23

	OLAP	OLTP
Nivel almacenamiento		Mínimo detalle pero cruzado con las distintas áreas de negocio.
Número de transacciones:	Real time	Batch nocturno o cada 15 min

Introducción al Business Intelligence – 2. Pág 24

EDW (Enterprise Data Warehouse. Almacén de datos corporativo) → Solo hay 1.

Cualquier solución analítica pasará por allí. Si se tiene un detalle de una área de negocio (marketing, negocios., finanzas, comercial..) normalmente se pasa por los Data Marts (Almacén de datos departamental).

Si se pasa por un repositorio central por donde se pasan todos los datos es fácil detectar si hay algún error. Permite que los KPI sean rehusables y crear los dashboards desde allí.

Centralizarlo en un sitio para normalizarlo todo y que todos los datos estén allí iguales y usables. Después ya se puede particularizar por cada área de negocio y redefinir los KPI o gráficas o lo que sea. Así no tienen los problemas de que los departamentos hablan con distintos datos. Todos parten del mismo conjunto de datos.

Introducción al Business Intelligence – 3. Pág 16

Garner se dedica a entrevistar a distintos usuarios de las plataformas y clasifican los vendedores según dos ejes:

↑ Cantidad de clientes (activity of commerce)

Apuesta de futuro con BI (completeness of vision) →

Introducción al Business Intelligence – 3. Pág 23

ETL → data factory

Para no instalar jupyter notebooks → Synapse Data Science

Introducción al Business Intelligence – 3. Pág 24

Salesforce compró Tableau (focalizada en dashboards)

Salesforce en el futuro tendrá una base de datos seguramente.

Tableau, Snowflake, SAAP.

Introducción al Business Intelligence – 3. Pág 25

Qlik. Empezó dashboard y compró backend. Compró 2009 Attunity que le da una base de datos. Compró Talend.

No funciona porque no suele ser escalable. Empezó para subir exceles. No es un software adecuado para BigData.

Introducción al Business Intelligence – 3. Pág 26

- Google es de las mejores en cuanto a datos. Fue de las primeras para BigData. Tiene de las tecnologías más nuevas y de calidad es muy superior al resto por lo que hace a los datos (big data). Es la única que puede analizar PetaBytes con BigQuery de Google.

Compró Looker que era como un Tableau.

- Microsoft de las mejores en ofimática.
- AWS nació de necesidad de la marketplace.

### **Herramientas más potentes según Francesc:**

- BigQuery
- Synapse
- Redshift
- Snowflake

- Power BI. Para Big Data va flojo porque empezó que tenías que importar los datos. Al conectar con una base de datos difícil. Muy ágil cuando no tienes grandes cantidades de datos.

- Looker
- Tableau

Introducción al Business Intelligence – 4. Pág 7

Para dar apoyo a unas ventas apoyado por un comercial. Clientes, comerciales (employees) y en que oficinas están.

Introducción al Business Intelligence – 4. Pág 9

Como sería diseño orientada a negocio. Diseño centrado en indicadores y dimensiones. Se quiere ventas analizada en función de tiempo, de producto, de consumidor, de promoción...

Se tienen menos tablas. Se va desnormalizando.

Introducción al Business Intelligence – 4. Pág 10

Es más corto y entonces es más rápido. Eso permite a motores OLAP ir más rápidos.

Introducción al Business Intelligence – 4. Pág 11

Fotos: Un dato orientado a aplicación ←

Un dato orientado a negocio (+rápido consulta) →

Introducción al Business Intelligence – 4. Pág 12

Crear un almacén de datos corporativo: Copiar el dato y metiéndolo en todos los sitios y tener un sistema que agregue todos los datos es al final lo que acaba y necesita hacer toda empresa que crece.

Introducción al Business Intelligence – 4. Pág 14

Cambiar los nombres según la definición que el área de negocios entiende.

Introducción al Business Intelligence – 4. Pág 15i16

Origen del DataLake. DataLake surgió después del DataWarehouse y no lo sustituye, conviven.

En el inicio, el dato de hace 10 años era estructurado. En los últimos 10 años ha habido boom con internet con muchos datos de carácter distintos: audio, foto, sensores... Un JSON, XSMML es un dato no estructurado. Y allí es donde salió la idea de hacer un bolcado de datos no estructurado para después irlo estructurado.

Introducción al Business Intelligence – 4. Pág 17

Uber tiene todo (90%) de datos no estructurados. Cuando se necesiten los datos ya lo estructuraremos. Un dato para ser analizado siempre tiene que acabar siendo estructurado. Pregunta: Se guarda estructurado o en la consulta se estructura? Estructurar el dato requiere computación así que es difícil estructurarlo a tiempo real. Ya se guarda y se accede después.

Introducción al Business Intelligence – 4. Pág 18

Tradicionalmente era: ←

Normalmente se trabaja ahora: →

DataLake no reemplaza DataWarehouse. Eliminar el Data Warehouse de la estructura de → no.

ETL es un equipo haciendo el desarrollo para normalizar el dato, estructurarlo y cargarlo a DataWarehouse.

### Quando usan los Data Science datos de DataWarehouse y de Data Lake?

- Proyectos de data science te puede interesar usar datos de Data Warehouse. Son más fáciles de usar y más rápidos. Se acostumbran a usar cuando se hacen análisis típicos y regulares:
  - Segmentar los consumidores. Por ingresos, por frec de compra, por dinero gastado, por última compra realizada.
  - Recomendar productos.
  - ...

- Después hay proyectos de data science que son exploratorios. Este tipo de proyectos no harás trabajar un equipo en el ETL para que te estructure el dato. Lo cojerás directamente de data lake.

Introducción al Business Intelligence – 4. Pág 24. La explica molt

4 ítems importants.

Per exemple:

- Location: Botiga de Barcelona
- Time: hoy a las 19h
- Branch: área de perfumes
- Ítem: Ha vendido 200 de este producto
- Todo de información adicional: Dolares, units sold, av sales del dia...

Capa semántica dice como se enlazan las tablas.

#### Introducción al Business Intelligence – 4. Pág 27

Diseño copito de nieve. Permite dar más contexto a los departamentos. Se estructura un poco más el dato para suplir las particularidades de más departamentos. Para no tener que poner el dato disponible para todos ni tampoco tener que hacer de nuevo otro esquema.

#### Introducción al Business Intelligence – 4. Pág 28

Se compatibilizan las estructuras para distintos departamentos. Departamento ventas y shipping comparten los datos de tiempo y ítem. El de ventas solo tiene el de branch y el de shipping solo tiene el de shipper etc.

#### Introducción al Business Intelligence – 4. Pág 29

ETL (Extraction, Transformation and Load). Son los Flujos de datos o canales de datos.

Suelen ser los más críticos.

Estos procesos son los que más cuesta. Incorporar un KPI

En cambio al equivocarse en un dashboard no pasa nada, se corrige rápido.

Por esto es importante que los data engineer no sean solo muy buenos en SQL o solo sean buenos en negocio. Sinó que es importante un perfil que sepa SQL y también a la vez entienda las soluciones de negocio que está construyendo. Porque sino esta bien hecho el ETL después hacer un cambio (como por ejemplo incorporar un KPI) es muy costoso.

**Qué perfiles técnicos tienen mejor o peor salario?** Es según la oferta y la demanda. Es difícil encontrar ese perfil? Qué aportas?

Los Business intelligence aportan mucho a la empresa sobretodo en el marketing

Según las tecnologías que estan implementando es difícil encontrar algunos perfiles. Looker y Tableau hay poca gente que sepa usarla, PowerBI más. Entonces se premia los que saben lo que necesita la empresa que sepa.

Backend → + críticos y difíciles de encontrar. → +sueldo

Frontend → solo dashboard, no prospera tanto → - sueldo

#### Introducción al Business Intelligence – 4. Pág 32

### 15/1 Introducción al Business Intelligence – 4. Pág 24 Fa repàs

#### Introducción al Business Intelligence – 4. Pág 31

Uno de los errores que se cometen mucho es que al tener una base de datos que escale, no resulta nada ágil que el visualizador de datos tenga que importar todos los datos.

Normalmente se quiere permitir que el usuario pueda explorar los datos (*journey*). Si se quiere hacer un *journey* muy largo, no es factible que el power BI importe todos los datos (normalmente se hace por la noche 1h) que se necesitan explorar y acabe almacenando 10 TB (8h de carga).

Las arquitecturas de datos que funcionan bien son arquitecturas que se acoplan a los campos o tablas.

En la imagen vemos: Bases de datos, desnormaliza, se ponen en la base de datos corporativo (data mart) y lo que se hace normalmente es que se cargan los datos en el frontend en el gráfico. Al empezar no hay problema pero al escalar, hay límite en la capacidad de los sistemas de frontend para poder tener los datos guardados.

Crear conexión online al Data Mart para poder acceder a todos los datos.

En vez de importar el dato, los dashboards lanzan queries online para importar los datos que necesiten. Arquitectura headless, sin que tenga que persistir el dato en el frontend.



Els processos ETL poden arribar a carregar molt a la xarxa. El que es pot fer és carregar-los i que treballi en la data base.

Introducción al Business Intelligence – 4. Pág 35

Procesos de ETL

Introducción al Business Intelligence – 4. Pág 36

Capa semántica facilita la relación entre las distintas dimensiones de las tablas de la base de datos. Permite acceder a los distintos campos de la base de datos. Permite que la interfaz visual entienda como funcionan las tablas para que haga las queries necesarias para agregar los datos deseados en los gráficos.

Data base - Capa semántica - Visualización

Seria muy costoso que cada gráfico tenga que escribir la SQL a la base de datos. La capa semántica ahorra esto.

Introducción al Business Intelligence – 5. Pág 5

BigQuery compite con: Snowflake, Redshift (amazon), Synapse (microsoft)

Las diferencias:

- BigQuery es más autogestionada para el escalamiento. No hace falta preocuparse por escalar. Se puede captar el escalamiento pero funciona automáticamente.
- Con Snowflake Redshift escala según lo que se le vaya diciendo

Introducción al Business Intelligence – 5. Pág 6

Una base de datos tiene que:

- Procesar (CPU, Cores). → En BigQuery hay los **Dremel**.

Root server:

Mixer: Agregan la información de los leaf nodes.

Leaf Nodes: Leen los datos, son los únicos que acceden a los datos.

- Almacenar datos. → **Colossus**
- Comunicar por red el almacenamiento y la computación → **Jupiter**
- Orquestar todas las tareas → **Borg**

Introducción al Business Intelligence – 5. Pág 6

Query analítica es más óptima a nivel de columnas que a nivel de filas. Ya que si son de columnas hay menos ficheros pero más densos.

Introducción al Business Intelligence – 5. Pág 8

Normalmente hay muchos datos repetidos. Normalmente al graficar lo que se hace es agregar los datos (tabla de columnas de pocas columnas y más frecuencias).

Una de las maneras de comprimir los datos es en vez de guardar los valores tal cual, encontrar un carácter o codificación que pese menos para los datos que están muy repetidos y hacer un diccionario asociado. Así se logra que los datos pesen menos.

<https://console.cloud.google.com/>

Introducción al Business Intelligence – 5. Pág 15

El caso práctico que se hará será de una empresa de fashion que quiere ver datos. Tendremos que modelar dato, capa semántica y hacer dashboard.

<https://console.cloud.google.com/bigquery?referrer=search&project=lasalle-albert-ripoll&ws=!1m0>

Introducción al Business Intelligence – 5. Pág 20

Con el with creas bloques reutilizables. Bloque “q1”, bloque “q2”...

Se usa para explorar el dato.

Se usa para realizar el pipeline más fácil de mantener con cajitas.

```
--obtener información de la cabecera de ordenes
with q1 as (
select order_id, user_id, status, created_at
from bigquery-public-data.thelook_ecommerce.orders
where status = 'Complete'),
--obtener información de las líneas
q2 as (
select order_id, product_id, id as line_id, cast(sale_price as numeric) as sale_price
from bigquery-public-data.thelook_ecommerce.order_items
),
```

Introducción al Business Intelligence – 5. Pág 25

Mejor hacer left outer join que no inner join para no perder identificadores que no tengan datos comunes

Introducción al Business Intelligence – 5. Pág 27

Agregaciones por productos únicos vendidos o por clientes únicos.

Count(distinct product\_id) Conteo de valores únicos

Introducción al Business Intelligence – 5. Pág 28

La capa de *staging* (transformaciones) es en la que más se invierte. Para normalizar. Cambiar el formato del dato. Operaciones matemáticas (quitar decimales...). Operaciones fechas (comparar fechas...). Operaciones con string (quitar strings innecesarias).

Introducción al Business Intelligence – 5. Pág 30

Las funciones ventana o funciones analíticas. Permite poder comparar por ejemplo la diferencia de gasto entre el 1o y 2o cliente, entre el 2o y el 3o, entre el 3o y el cuarto... Siempre utilizan la sentencia over y comparan sobre una ventana o partición y comprar con otra ventana.

(ORDER BY q6.total\_price desc) as next\_customer\_total\_price

Ordena por precio y acceder a la siguiente fila para comparar ya con el siguiente.

Introducción al Business Intelligence – 5. Pág 32

Funciones aproximadas. Evitan hacer scanning o conteos de toda la base de datos para sacar un valor rápido y aproximado. Se estan implementando en Snowfake y BigQuery.

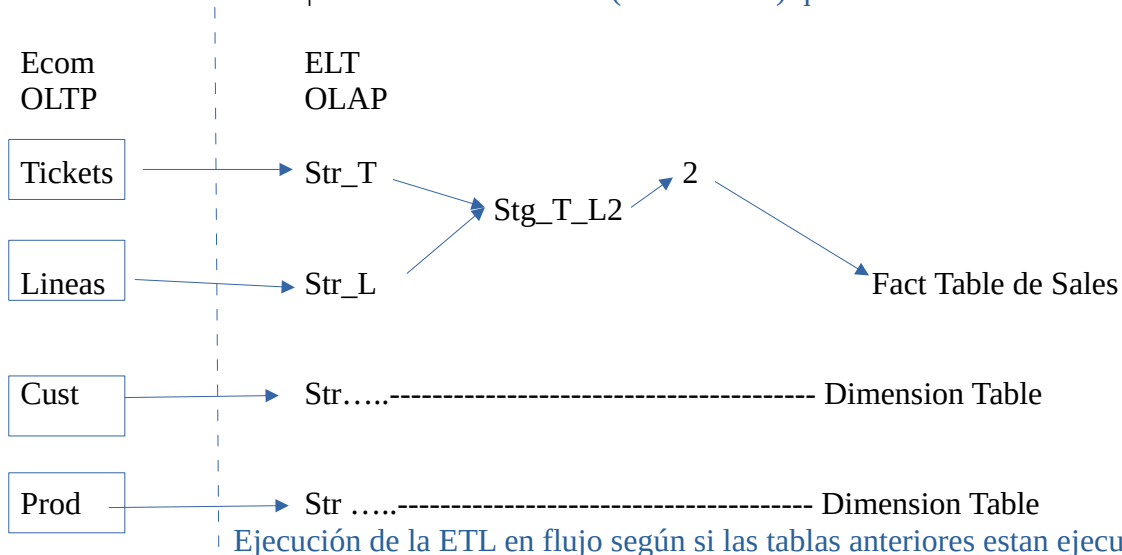
Por ejemplo análisis de venta por quantiles con el fin de segmentar los clientes para enviar un correo a cada tipo de cliente.

Introducción al Business Intelligence – 5. Pág 39

DAG (Direct Acyclic Graph) es un grafo (nodos y flechas) que nunca tendrá un loop cerrado. Los procesos ETL van de la capa transaccional a la derecha, nunca se puede crear un loop. Una tabla se crea a partir de otra y ya.

Sistema transaccional |

Cada nodo (tabla o vista) que es un select de lo otro.



dbt airflow orquestran

Introducción al Business Intelligence – 6. Pág 1 a 15 no escolto.

Introducción al Business Intelligence – 6. Pág 16

17/1 Introducción al Business Intelligence – 7. No estic a classe

Introducción al Business Intelligence – 8. Pág 14

Lo mejor es que el pipeline de un dato corporativo en el data mart lo mejor es que se realice dentro de la base de datos. Pero igualmente hay algunos procesos que se tienen que hacer fuera de la base de datos. Una solución para eso es Apache Airflow para llamar API y orquestrar datos.

Sequencializar unos procesos. El flujo de datos a veces es más difícil.

Introducción al Business Intelligence – 8. Pág 15 a 31 bastant rapid

Introducción al Business Intelligence – 8. Pág 32

En vez de llevarte el dato con spark y trabajarlo en local con Python, trabajar con machine learning con SQL con cosas básicas y no muy parametrizables.

Caso práctico. Pág 2