INICIO GRABACIÓN

MINERIA DE DATOS//DATAWAREHOUSE NECESIDADES QUE IMPLICAN EL BIGDATA

Bibliografía: Unipymes 2021

SANJOSÉ
FUNDACIÓN DE EDUCACIÓN SUPERIOR

SANJOSÉ
FUNDACIÓN DE EDUCACIÓN SUPERIOR

INDICE

1 Nuevas necesidades y requerimientos que implican el uso del Big Data

2 CONCEPTOS

3 EJERCICIOS Y EJEMPLOS APLICADOS

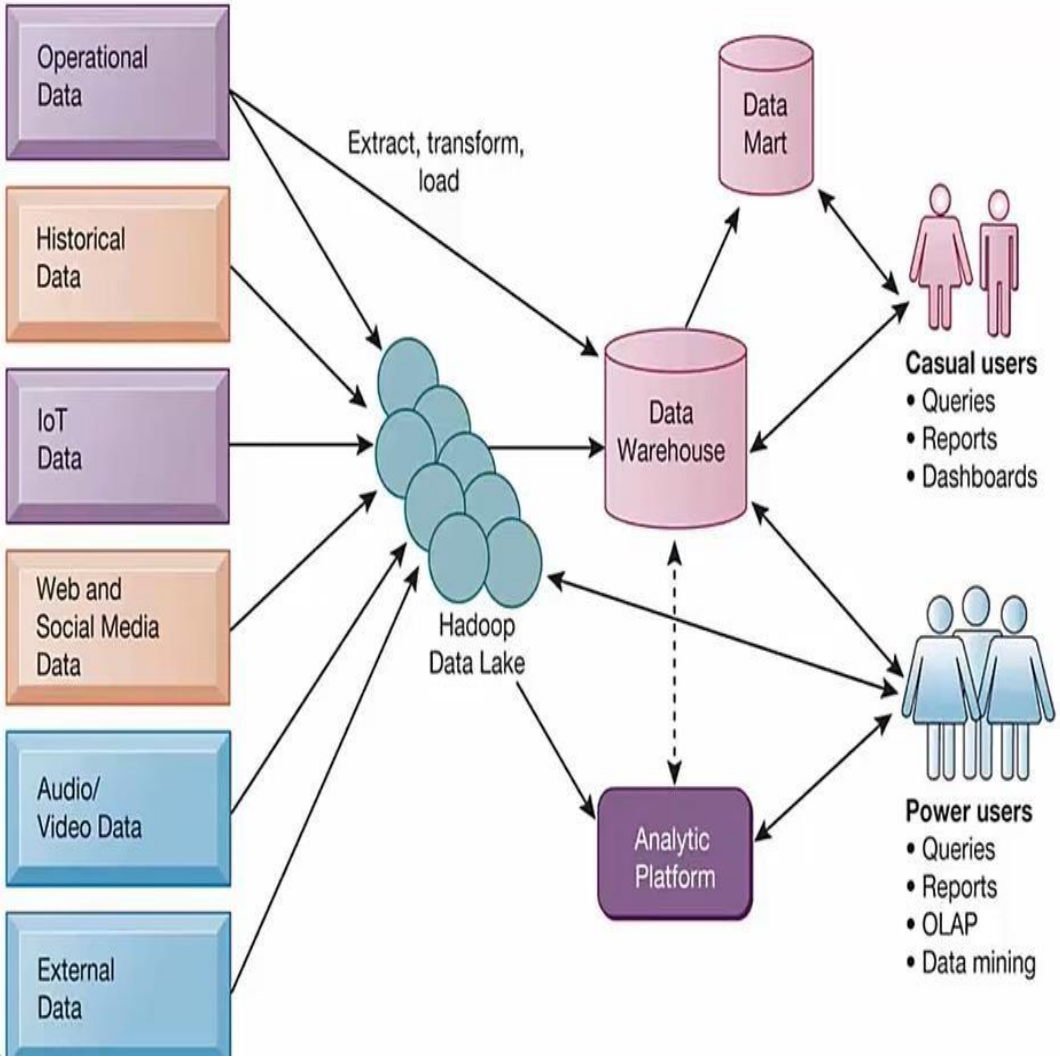
4 CONCLUSIONES

SANJOSÉ

FUNDACIÓN DE EDUCACIÓN SUPERIOR

CONCEPTOS

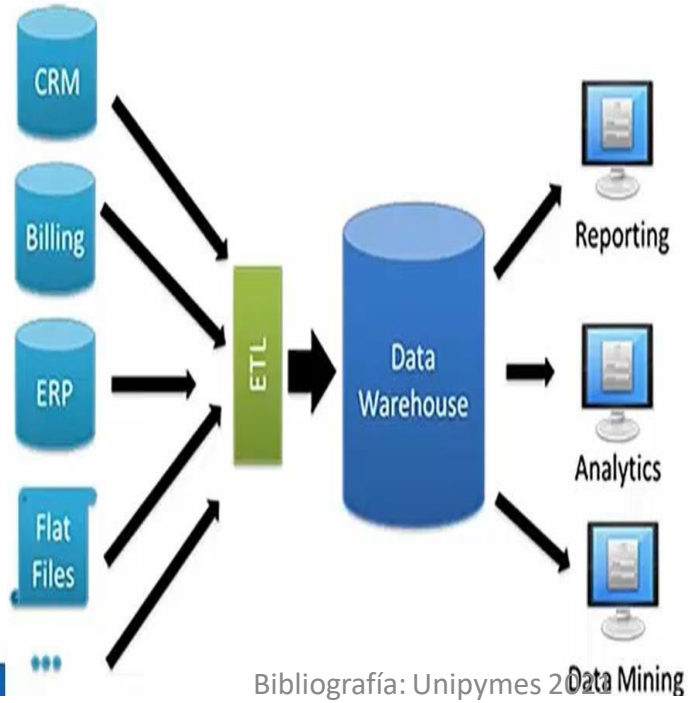
Ecosistema Big Data

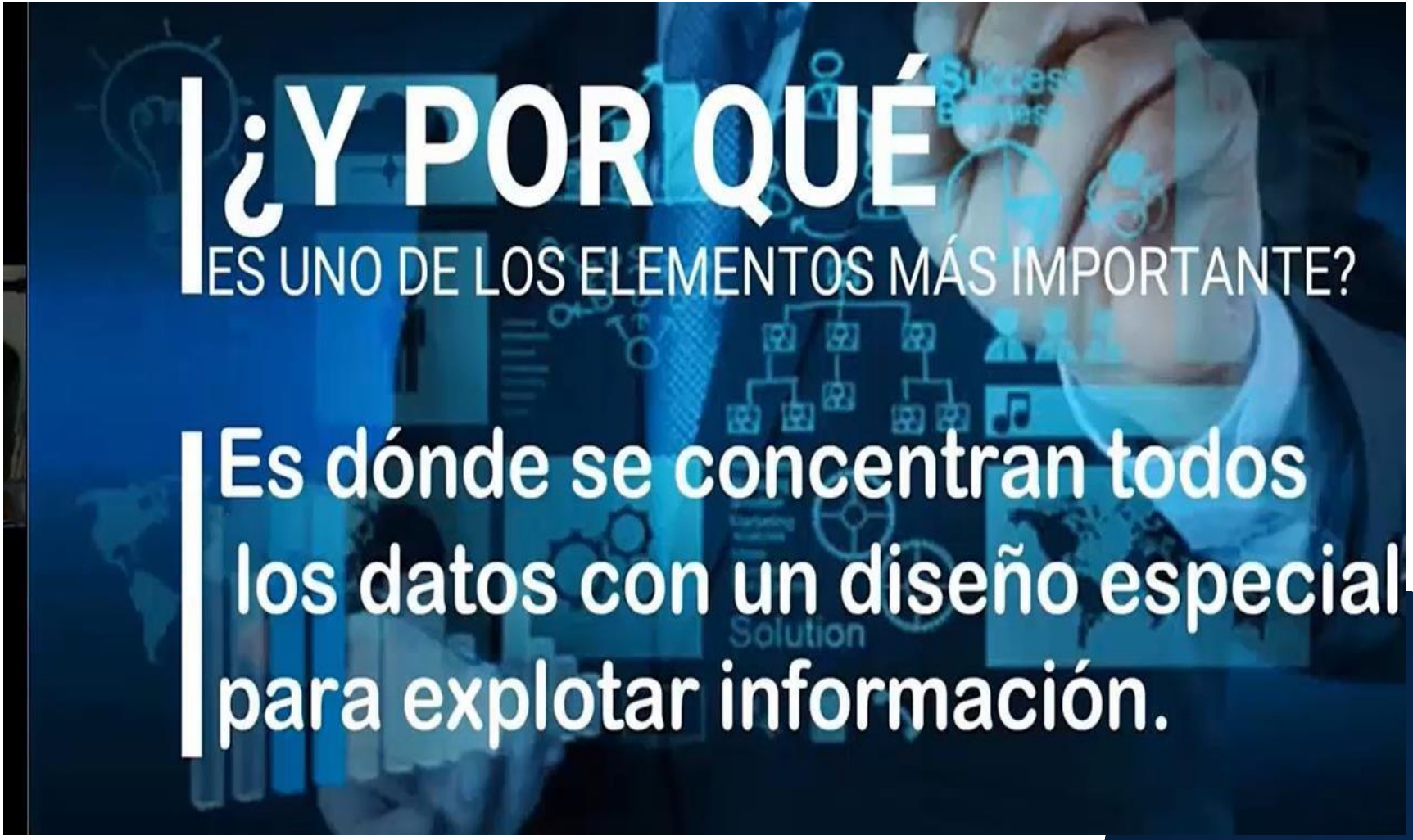


Bibliografía: Unipymes 2021

¿Qué es un Analytic Data Warehouse? (ADW)

- Esta evolución tiene su razón de ser en que no es lo mismo aplicar determinados algoritmos complejos sobre grandes volúmenes de datos, que aplicar estos mismos sobre un conjunto reducido de ellos.

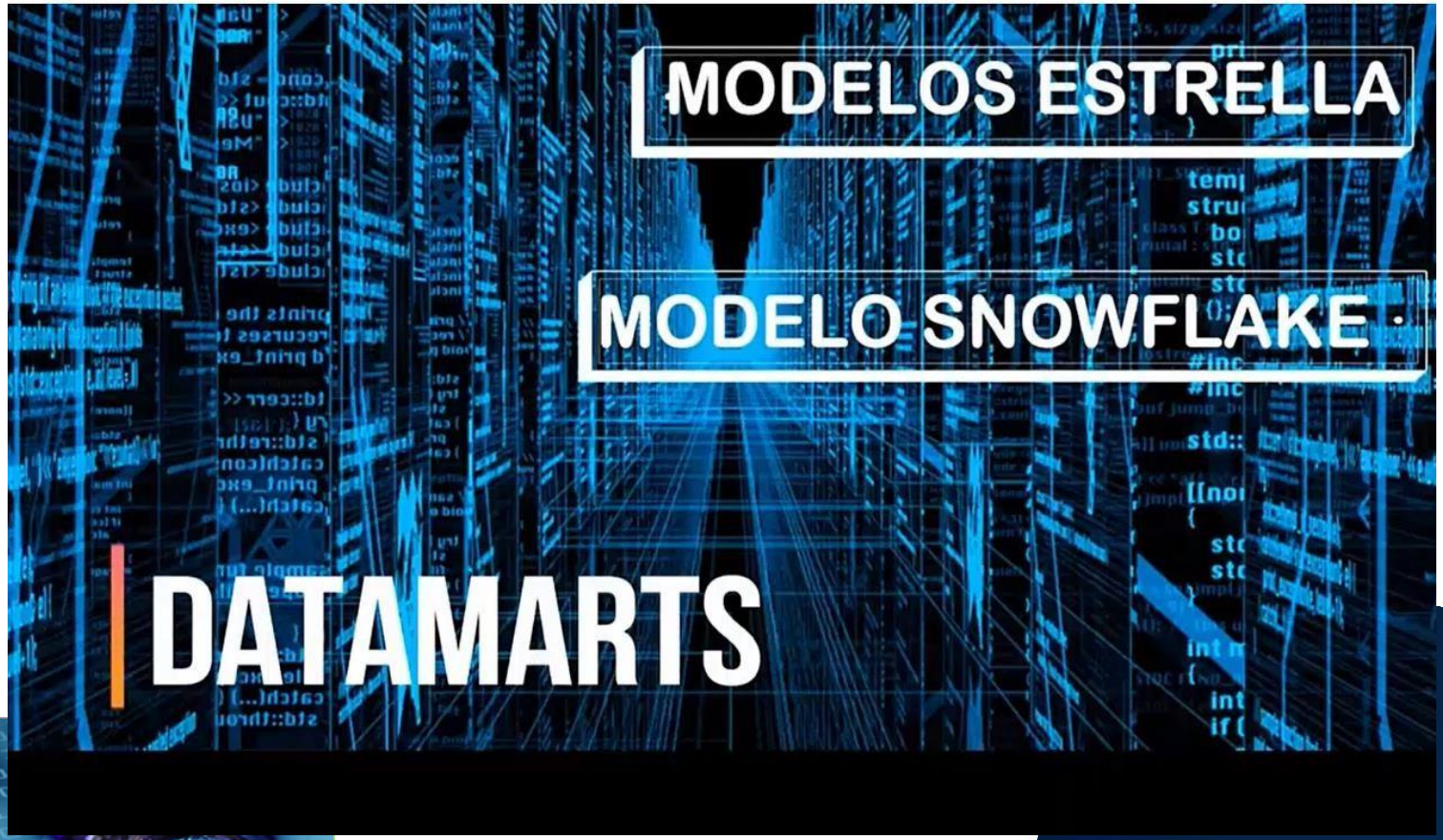




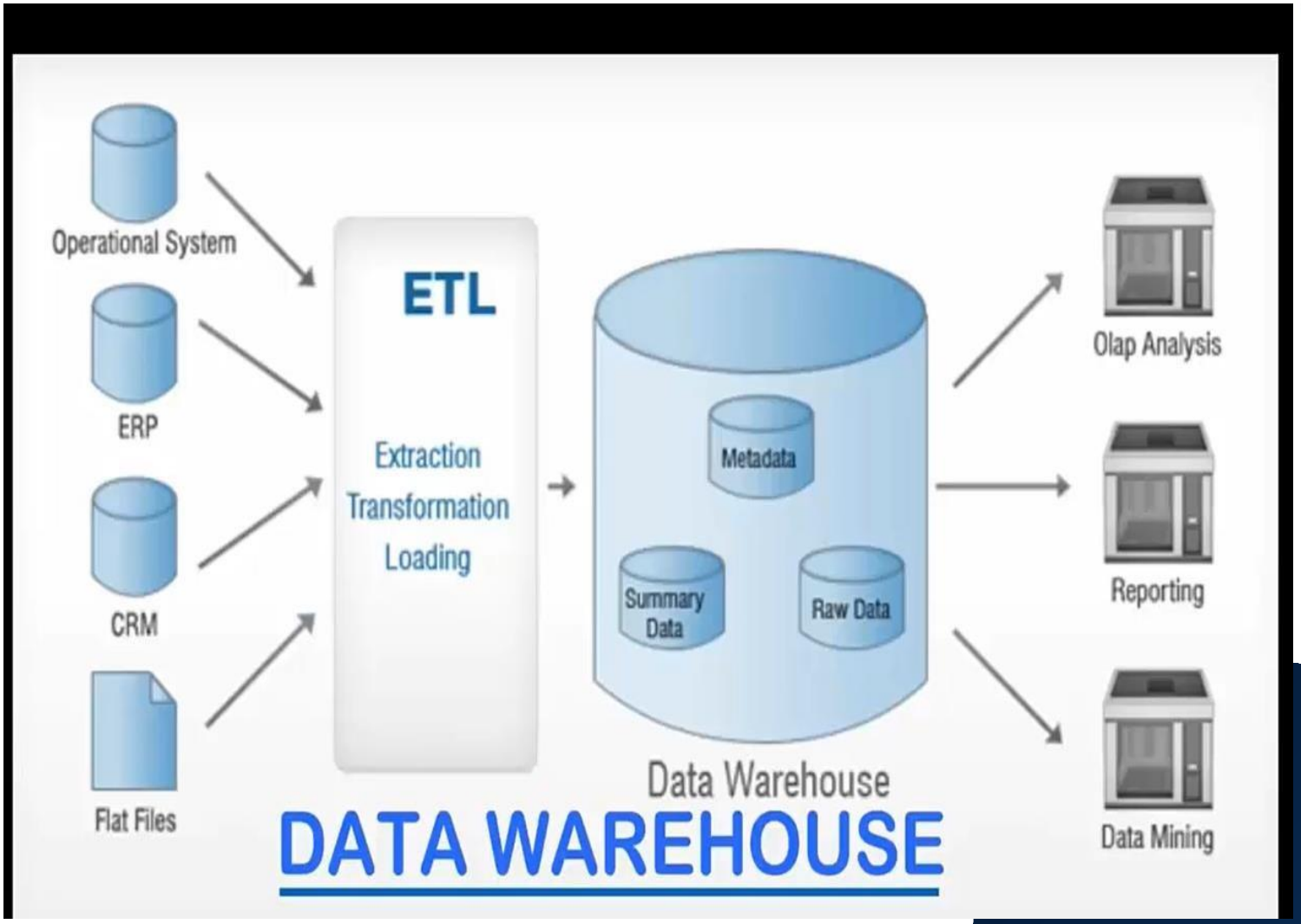
Bibliografía: Unipymes 2021

Se componen por fragmentos derivados del DATA WAREHOUSE conocidos como **DATAMARTS**

Bibliografía: Unipymes 2021



Bibliografía: Unipymes 2021



Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Bibliografía: Unipymes 2021

Base de datos corporativa

Datawarehouse

- Un Datawarehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un datawarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.
- La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales... etc). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales).

Datawarehouse

- El término Datawarehouse fue acuñado por primera vez por Bill Inmon, y se traduce literalmente como almacén de datos. No obstante, y como cabe suponer, es mucho más que eso. Según definió el propio Bill Inmon, un datawarehouse se caracteriza por ser:
- **Integrado**: los datos almacenados en el datawarehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Datawarehouse

Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del datawarehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Bibliografía: Unipymes 2021

Datawarehouse

- **Histórico**: el tiempo es parte implícita de la información contenida en un datawarehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el datawarehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el datawarehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- **No volátil**: el almacén de información de un datawarehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del datawarehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía

Datamart

- El Data Mart es un subconjunto de un Data Warehouse orientado al análisis, almacenamiento e integración de los datos de un área de la empresa. Esto quiere decir que el Data Mart posee la misma funcionalidad y complejidad que el del Data Warehouse. Generalmente los datos están estructurados en modelos estrellas o copo de nieve. Los departamentos de una empresa, por ejemplo el departamento de ventas, tiene necesidades distintas a las necesidades de la organización, por lo que la información también tiene un nivel de complejidad distinto.
- Una de las principales desventajas de esta herramienta es que no considera otras fuentes de datos de la empresa lo cual podría ser un complemento a la riqueza de la información.

Tipos de Datamart

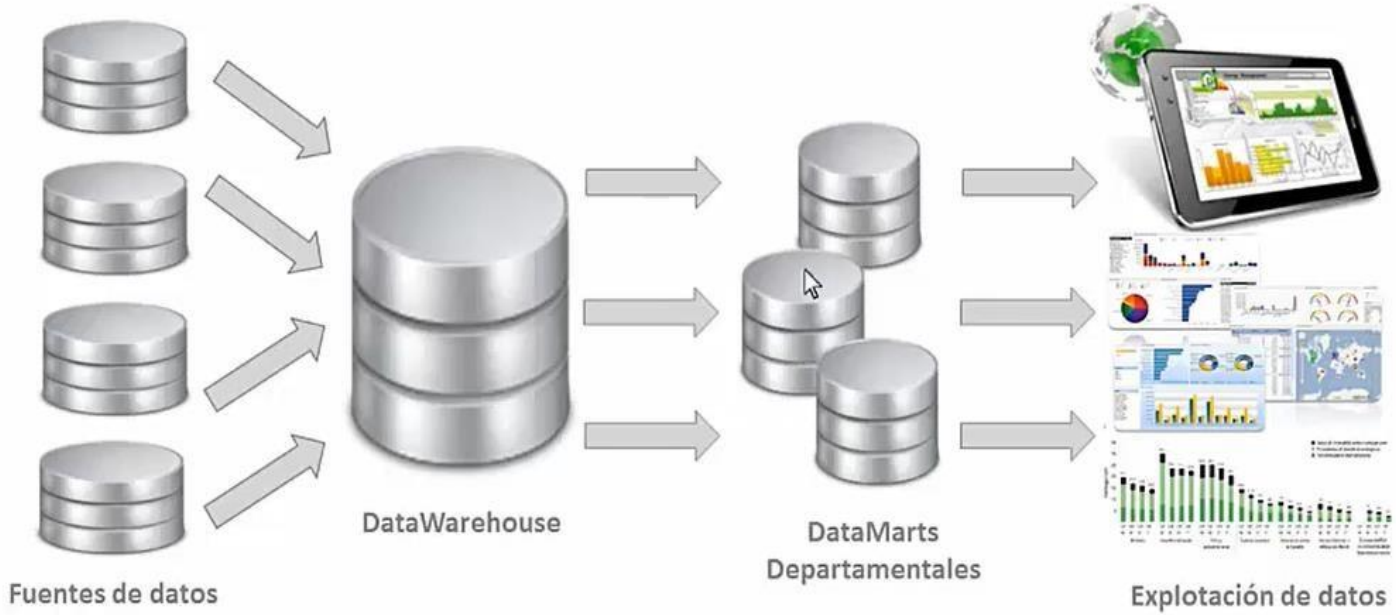
- **Datamart OLAP**

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

- **Datamart OLTP**

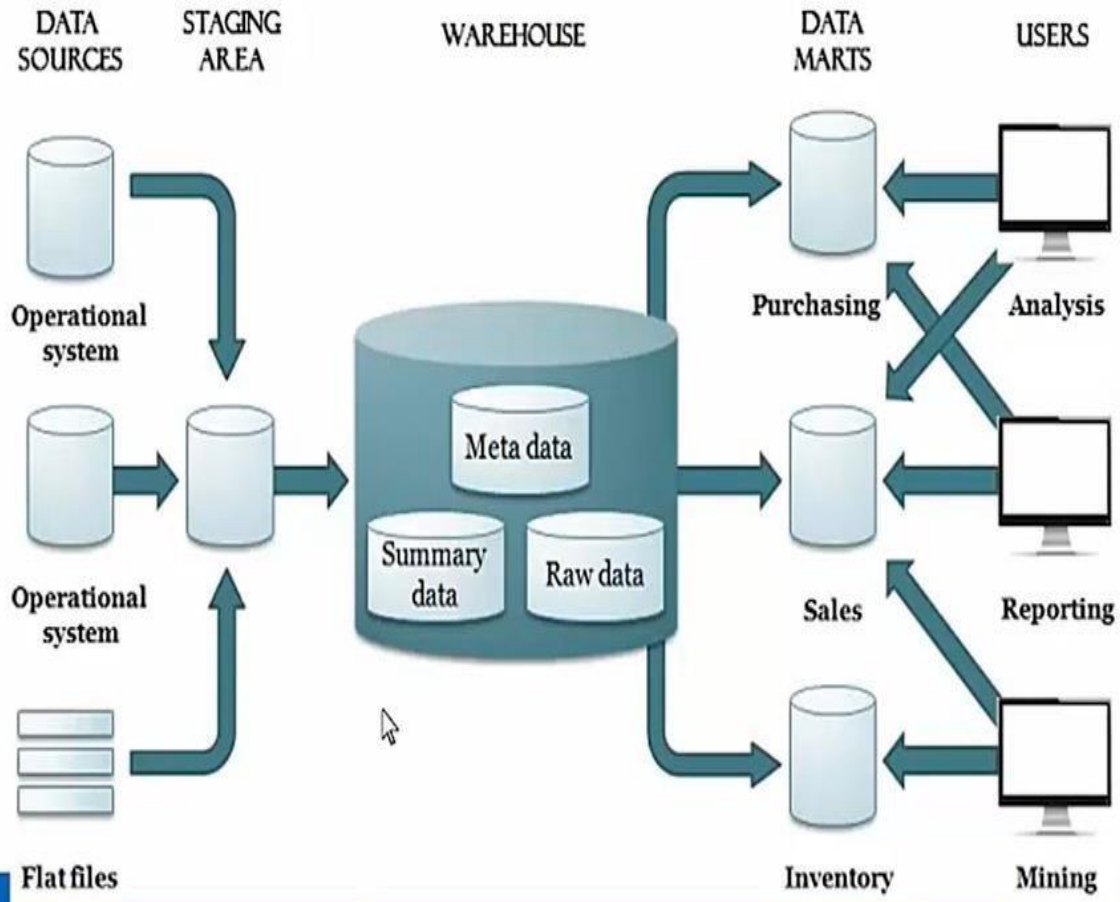
Pueden basarse en un simple extracto del datawarehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser fact-tables reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries.

Datamart



Bibliografía: Unipymes 2021

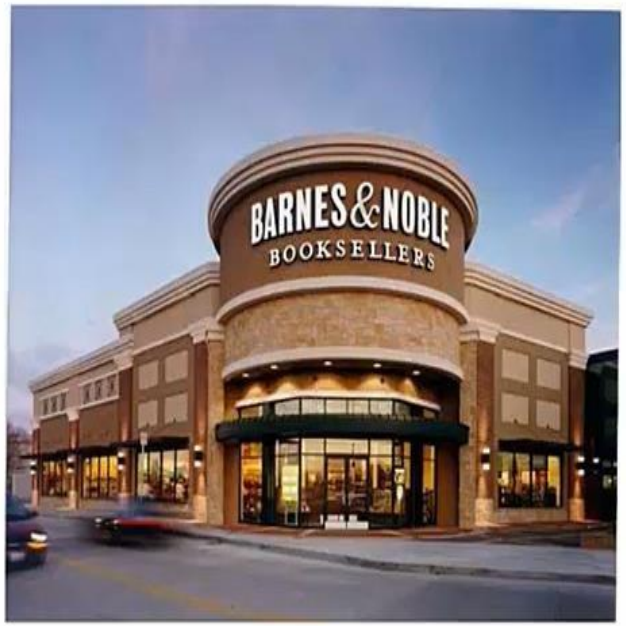
Datamart: Diferencias con Datawarehouse



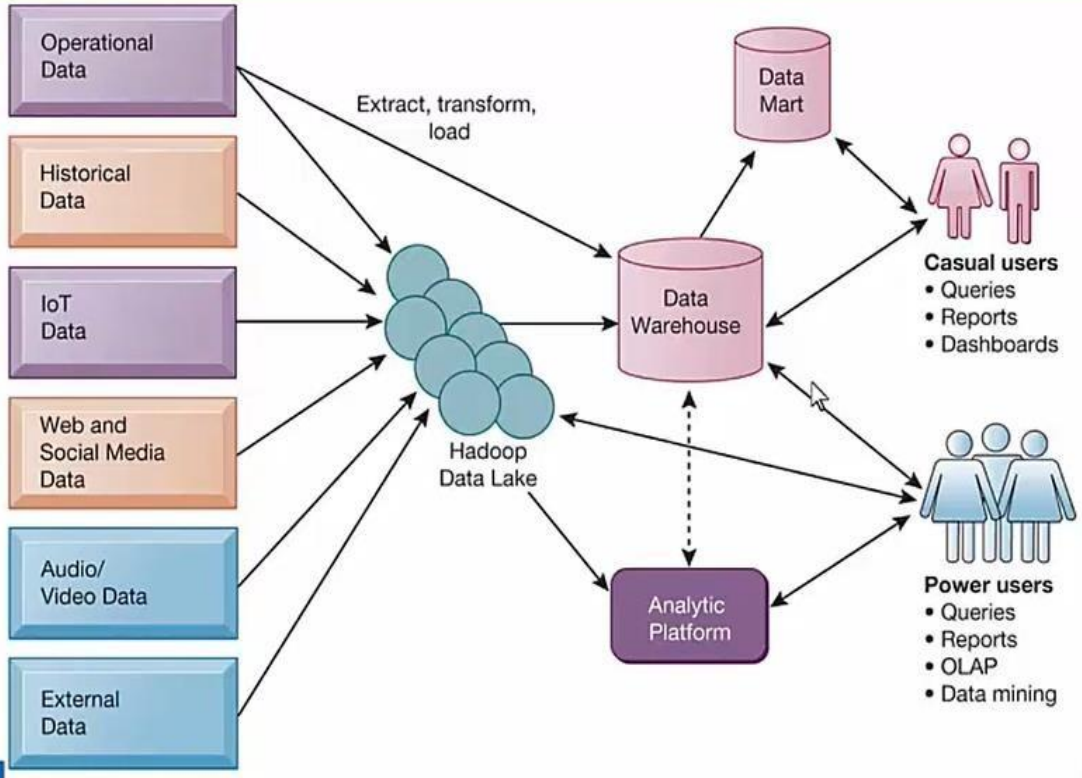
Bibliografía: Unipymes 2021

Datawarehouse y Datamarts

Por ejemplo, una empresa podría desarrollar mercados de datos de marketing y ventas para tratar con la información del cliente. La librería Barnes & Noble suele mantener una serie de Datamarts: Uno para datos de puntos de venta en tiendas minoristas, otro para ventas en tiendas de libros universitarios y un tercero para ventas en línea.

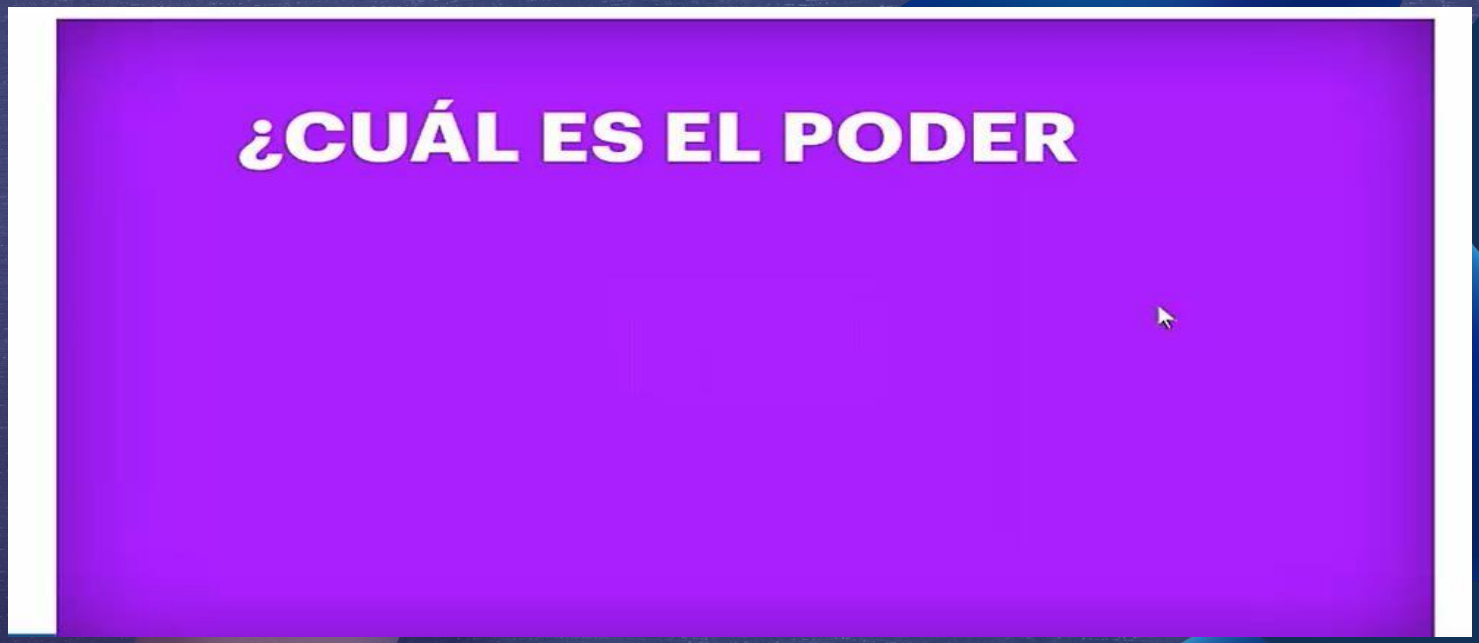


Infraestructura de Business Intelligence Contemporánea



Bibliografía: Unipymes 2021

Bibliografía
Bibliografía



Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

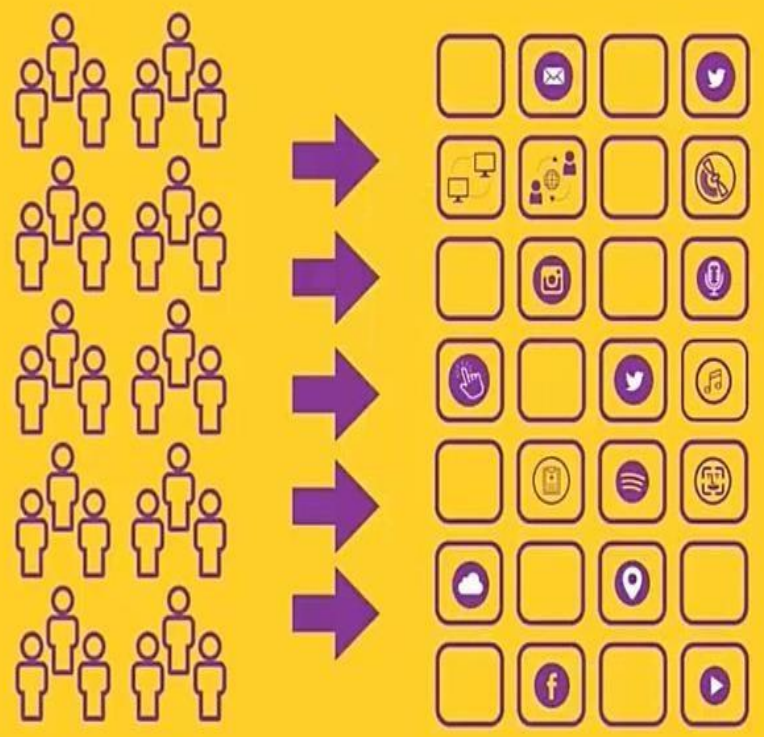
Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

Repositorio de datos no estructurados- también son transaccionales- no extracción – Internet de las cosas- se procesa a Hadoop y luego a un Data Ware House

Bibliografía: Unipymes 2021

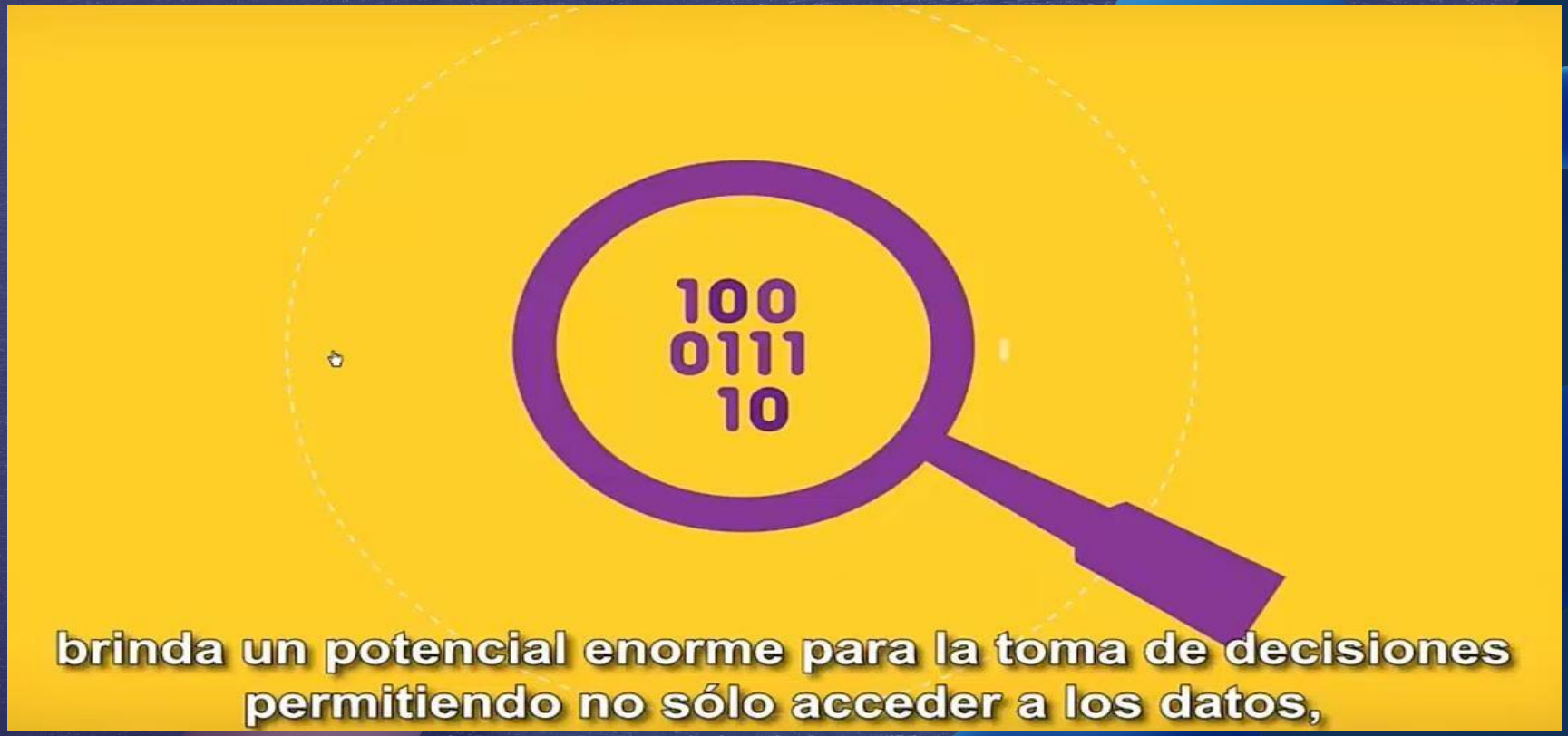
Bibliografía
Bibliografía



las cuales interaccionan con la Data Lake, que ya contiene datos, tanto básicos como procesados,

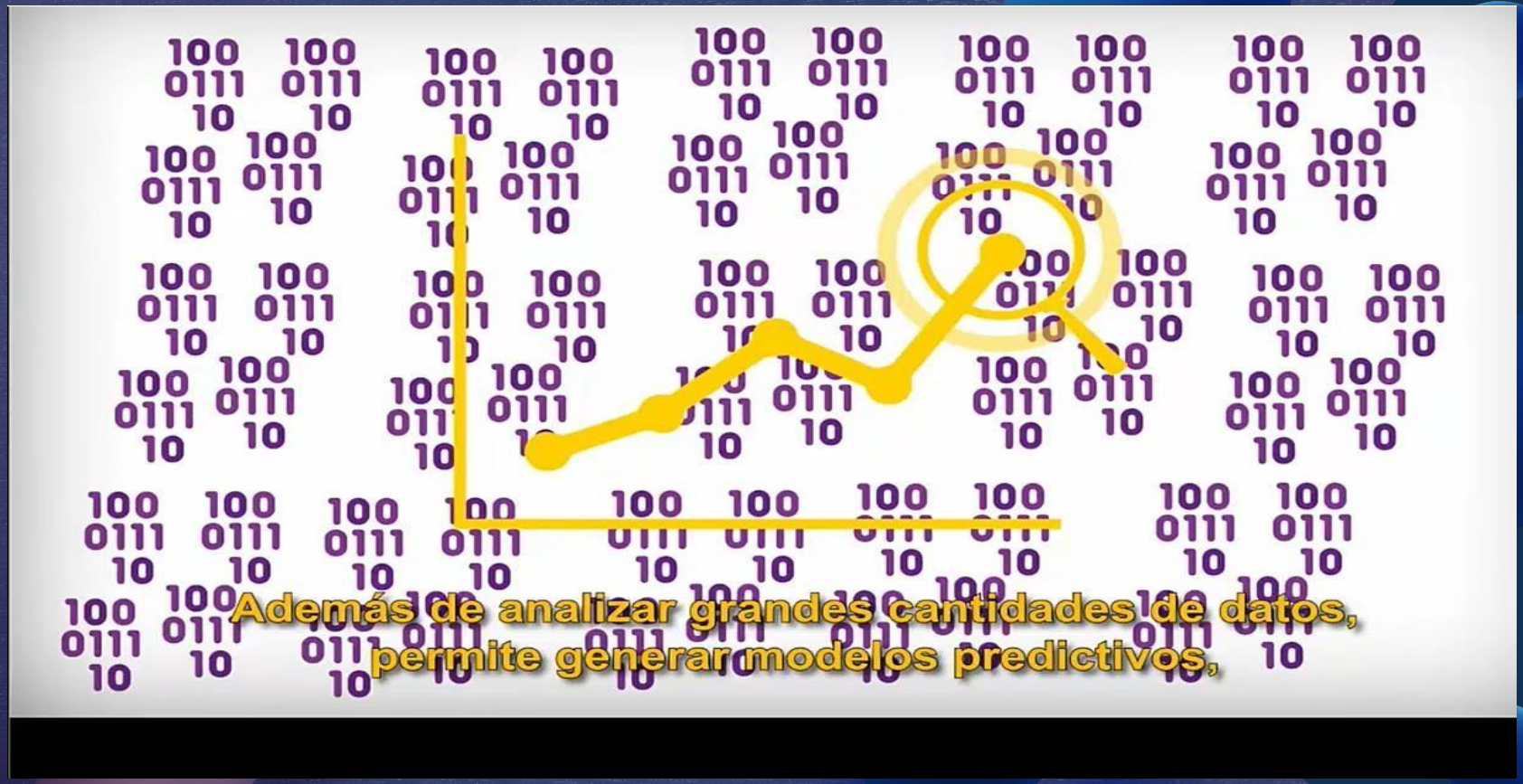
Bibliografía: Unipymes 2021

Bibliografía
Bibliografía



Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

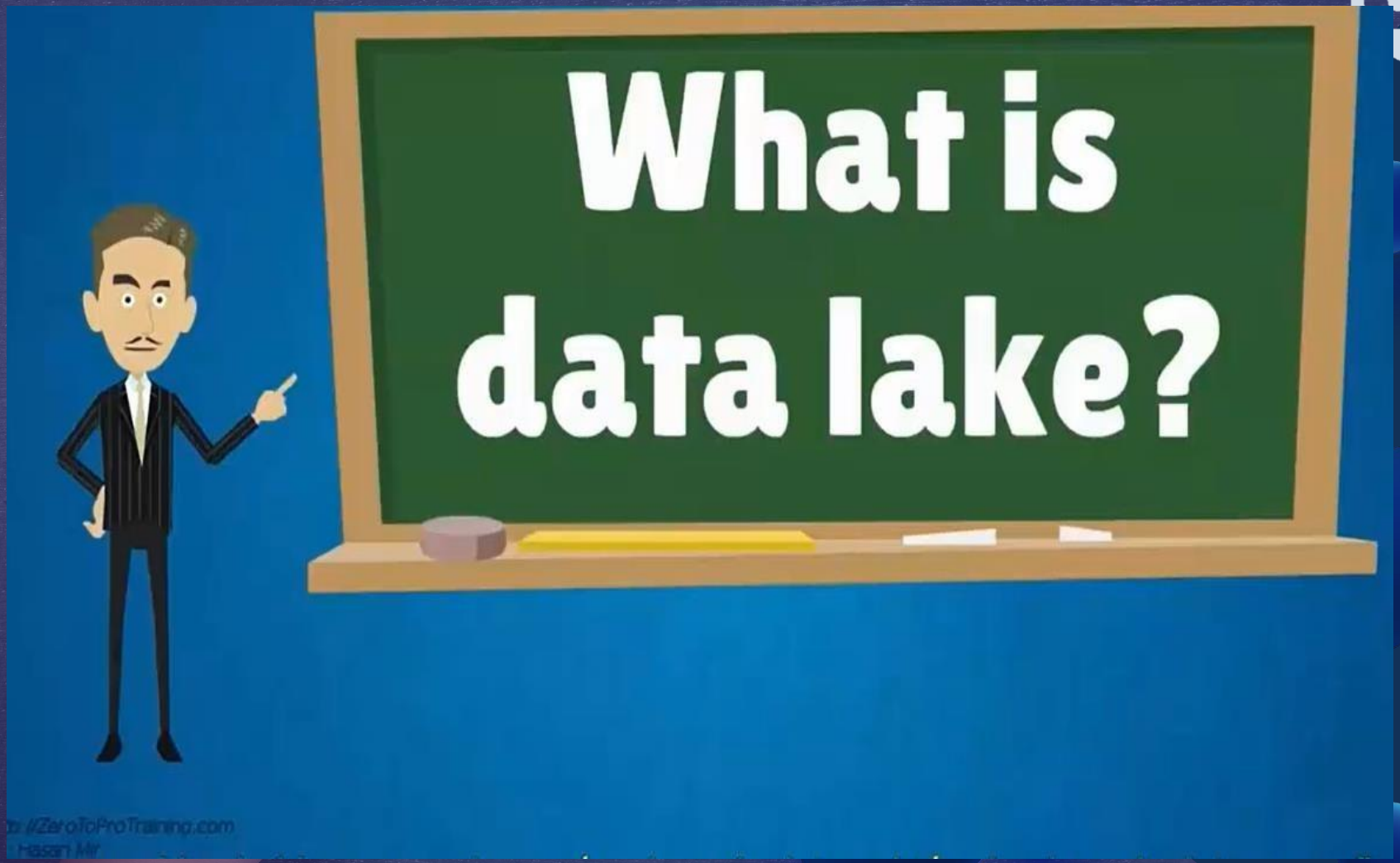


Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

Bibliografía: Unipymes 2021

Bibliografía
Bibliografía



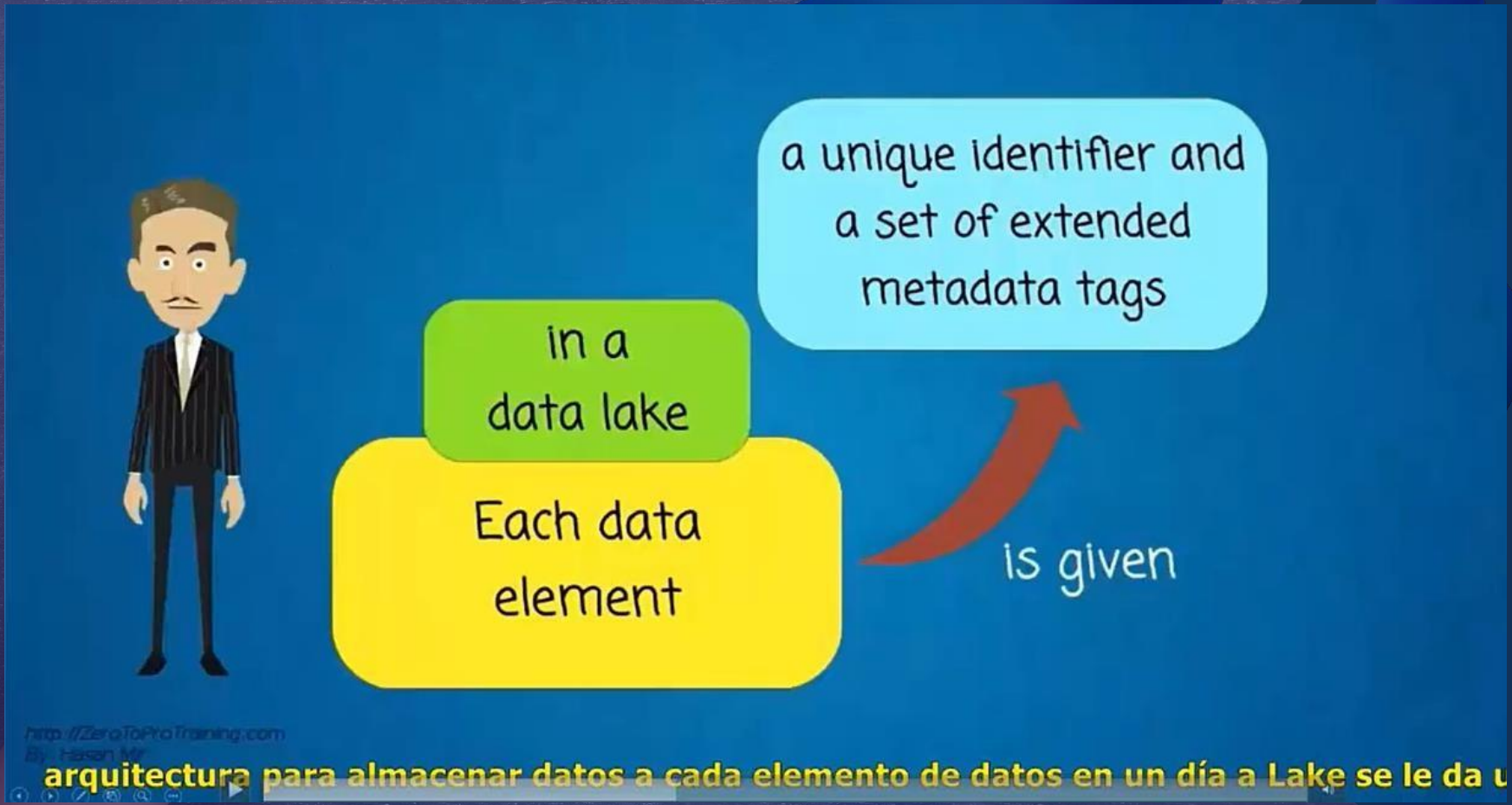
Bibliografía: Unipymes 2021

Bibliografía
Bibliografía

Bibliografía
Bibliografía

Bibliografía: Unipymes 2021

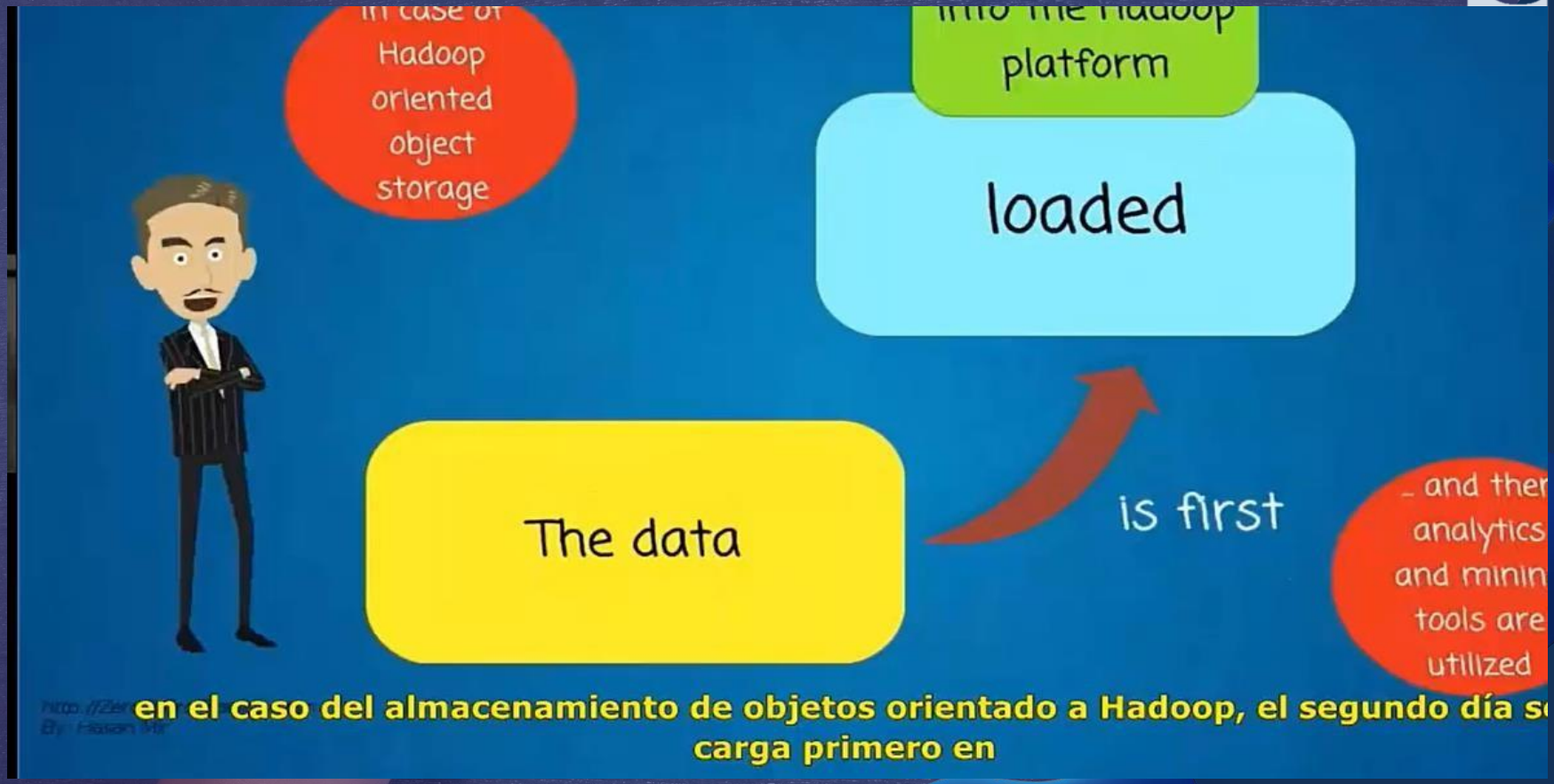
Bibliografía
Bibliografía



Bibliografía
Bibliografía

surge el lago de datos se puede consultar y el conjunto de resultados se analiza el término

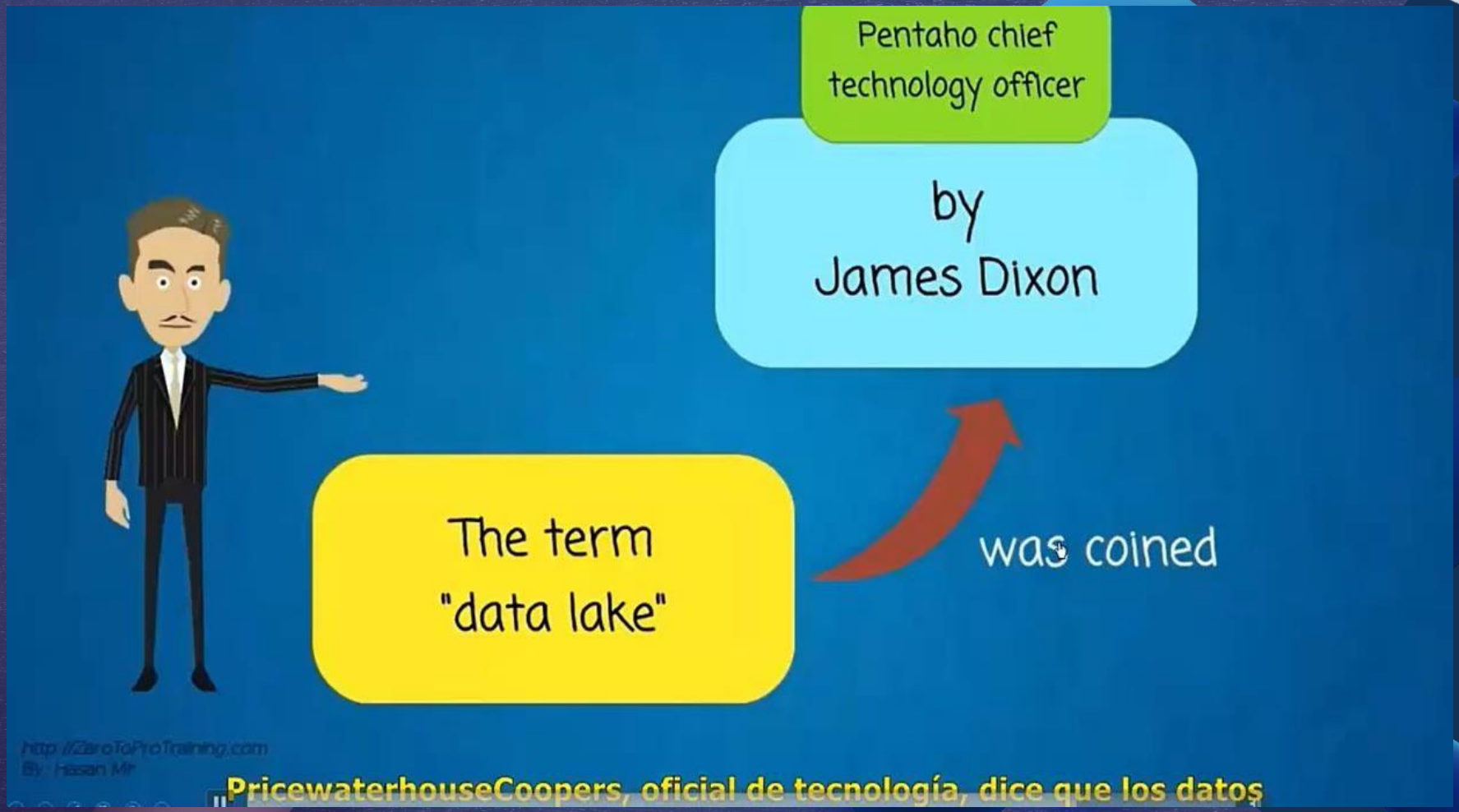
Bibliografía
Bibliografía


In case of Hadoop oriented object storage
into the Hadoop platform
loaded
The data
is first
and then analytics and mining tools are utilized
en el caso del almacenamiento de objetos orientado a Hadoop, el segundo día se carga primero en

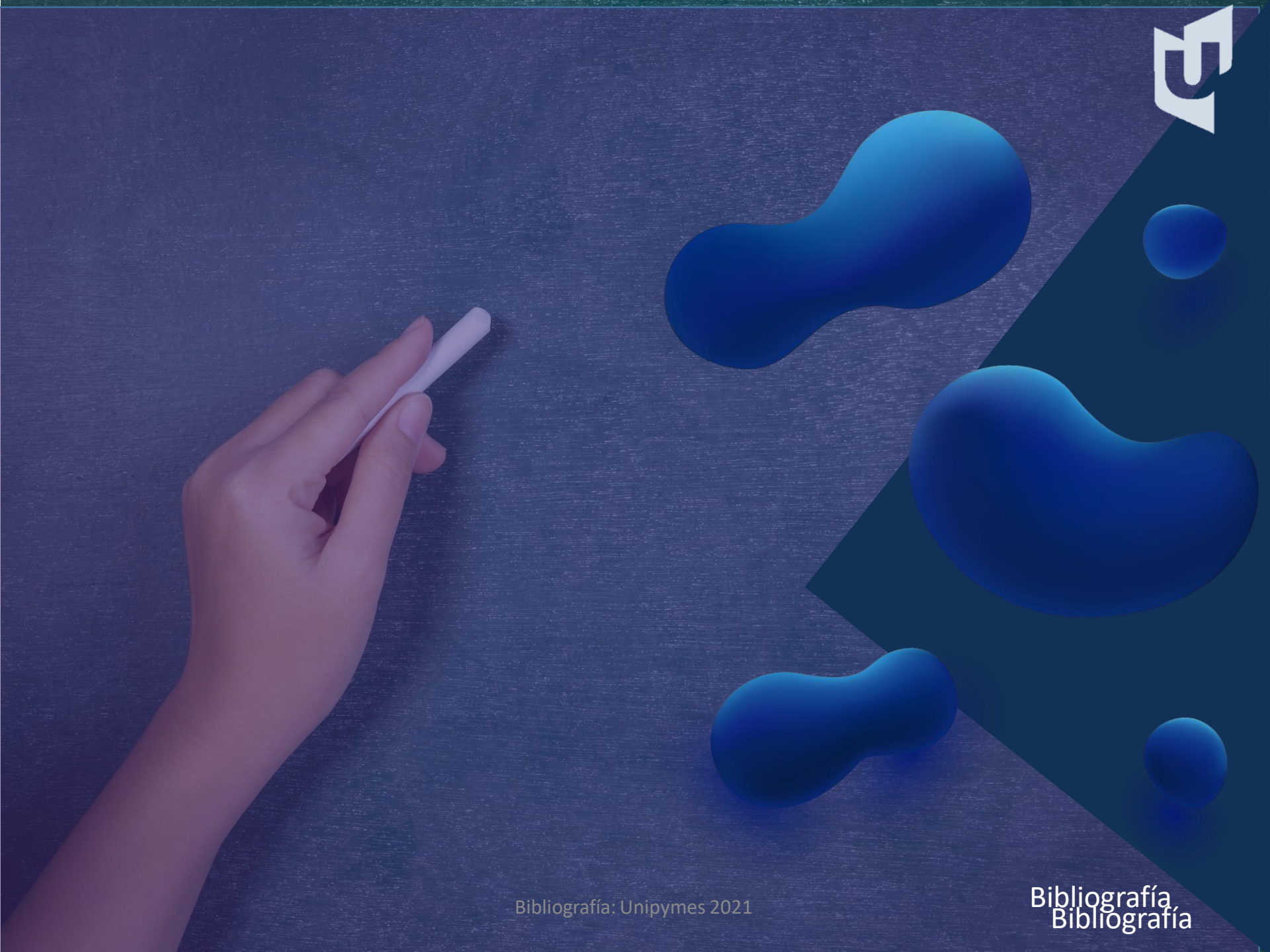
Similar a Pentaho

Bibliografía: Unipymes 2021

Bibliografía
Bibliografía



"PricewaterhouseCoopers, oficial de tecnología, dice que los datos



Bibliografía

Bibliografía: Unipymes 2021

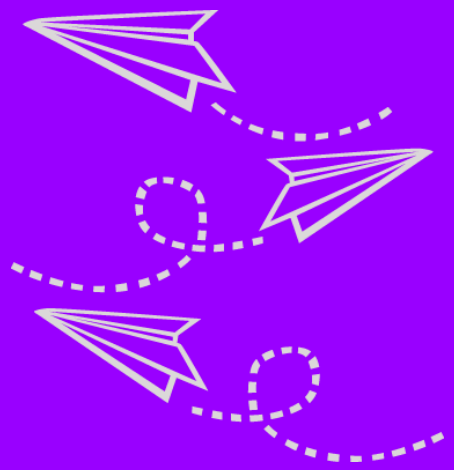
Bibliografía

las filtraciones podrían poner fin a los silos de datos

CONCLUSIONES

- Podemos gestionar con Big Data a través de diferentes herramientas información estructurada y no estructurada

- El Big Data es más simple en la medida que tengamos una estructura adecuada en infraestructura en software y hardware



FRASE

Vale la pena invertir tiempo en el análisis de los datos.
"El precio de la luz es menor que el costo de la oscuridad." – Arthur C. Nielsen

El genio se informa por los datos.
"Con demasiada frecuencia nos olvidamos de que el genio, también, depende de los datos dentro de su alcance, que incluso Arquímedes no podría haber ideado los inventos de Edison." – Ernest Dimnet

INICIO RECESO

Bibliografía: Unipymes 2021

FIN DE RECESO

Bibliografía: Unipymes 2021

FUNDACIÓN DE EDUCACIÓN SUPERIOR
SANJOSÉ
INSTITUCIÓN TECNOLÓGICA

FIN DE GRABACIÓN

Bibliografía: Unipymes 2021