

El camino de la Analítica de Datos a Big Data

Fernando Grados, Analista Senior
Dominio Consultores

Agenda

- Conceptos previos
- Analítica de datos
- Big Data
- Retos
- Comparación entre Data Mining, BI y Big Data

Midiendo el almacenamiento

Nombre	Medida Binaria	Cantidad de Bytes	Equivalente
Kilobyte (KB)	2^{10}	1,024	1024 bytes
Megabyte (MB)	2^{20}	1,048,576	1024 KB
Gigabyte (GB)	2^{30}	1,073,741,824	1024 MB
Terabyte (TB)	2^{40}	1,099,511,627,776	1024 GB
Petabyte (PB)	2^{50}	1,125,899,906,842,620	1024 TB
Exabyte (EB)	2^{60}	1,152,921,504,606,840,000	1024 PB
Zettabyte (ZB)	2^{70}	1,180,591,620,717,410,000,000	1024 EB
Yottabyte (YB)	2^{80}	1,208,925,819,614,620,000,000,000	1024 ZB

Un Yottabyte



93 370 681 605 975 723, 62
Fotografías digitales de 16
megapíxeles en cámara Casio EX-
ZR1000, con calidad fina



18 518 518 518 518, discos Blu-ray
de doble capa de 54 GB



1 333 333 333 333,33 discos duros
de 750 GB

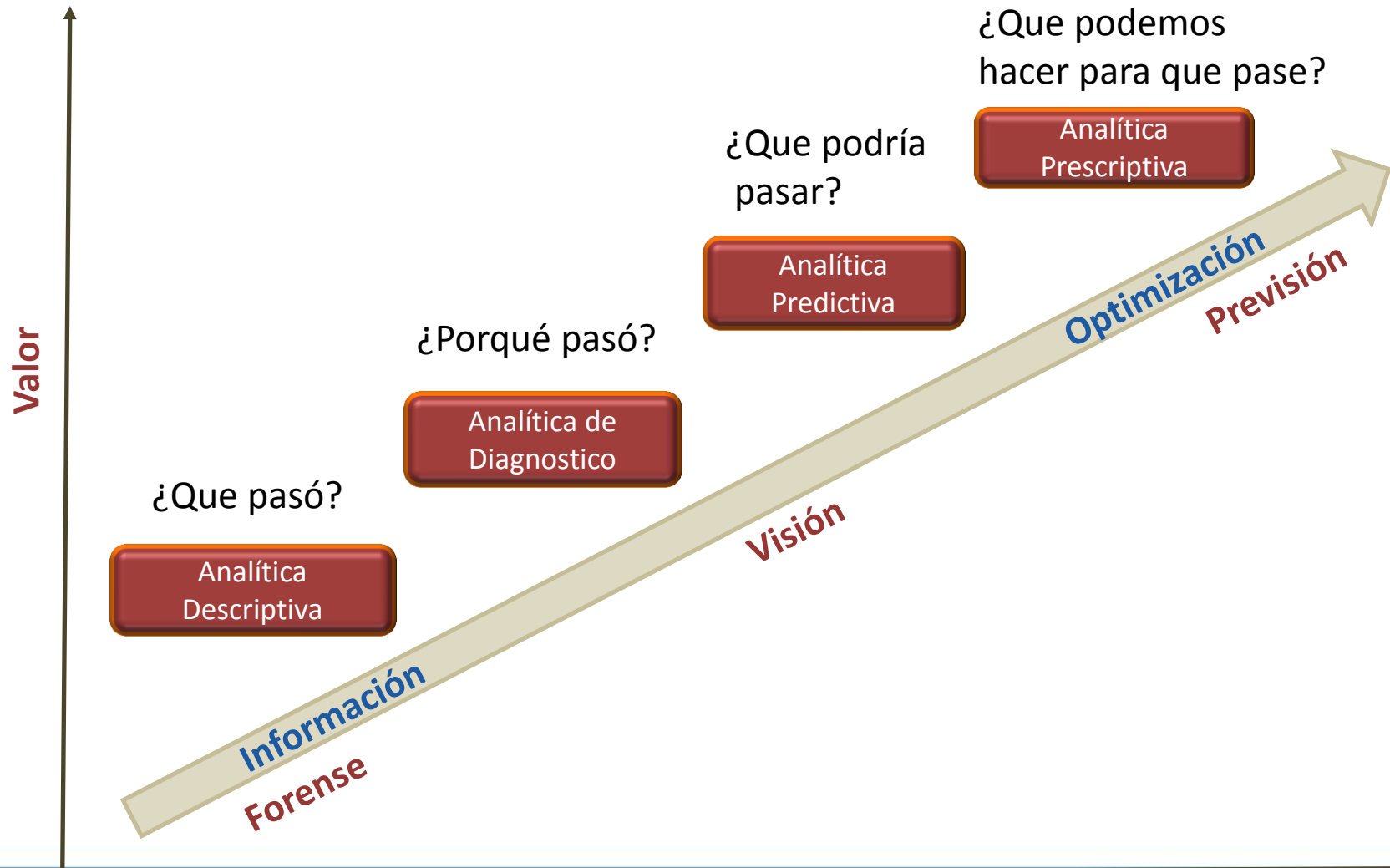


Concepto

Analítica de datos es la ciencia que examinar los datos en bruto con el fin de sacar conclusiones acerca de esa información.

El análisis de datos se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador.

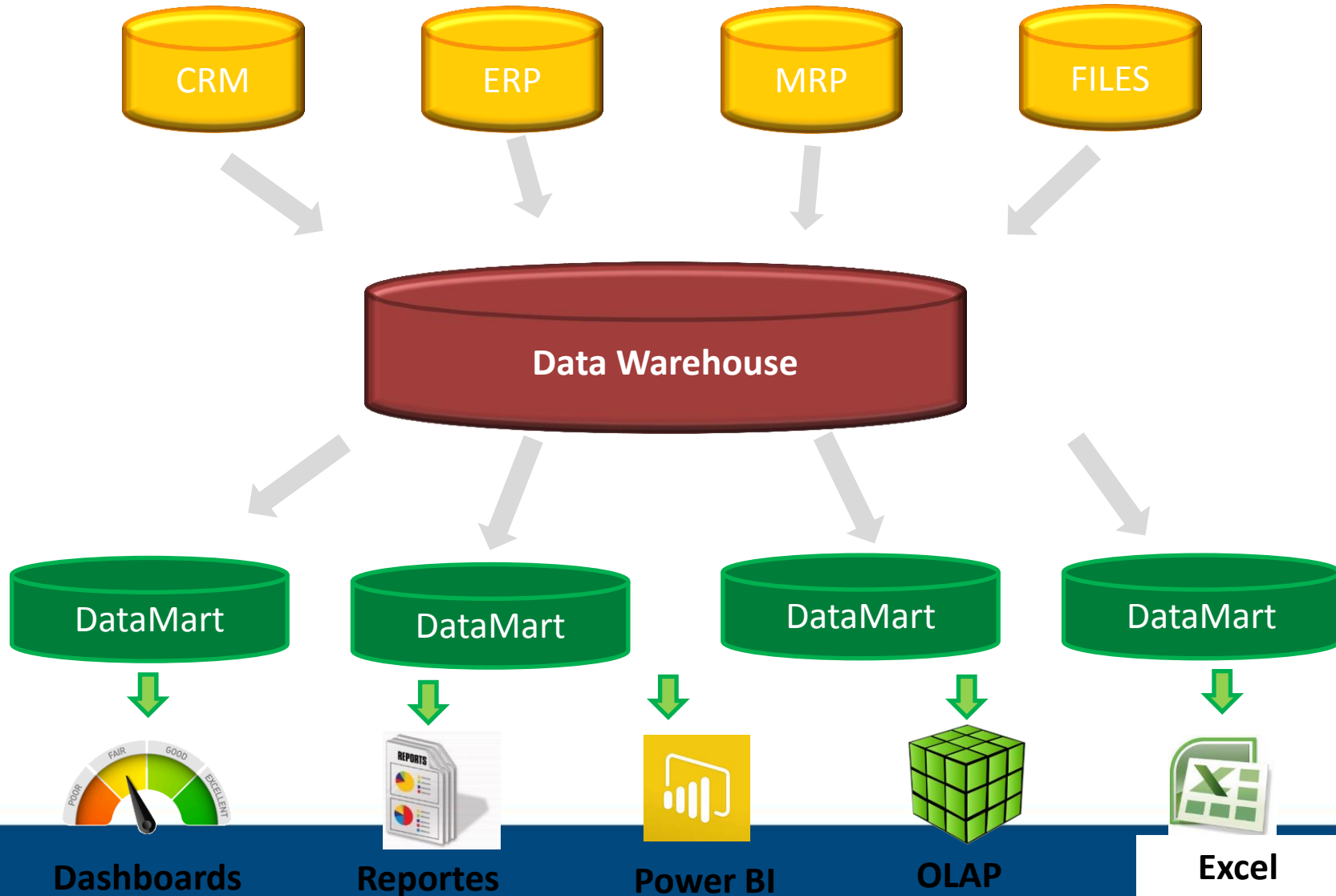
Desarrollo



Atributos

- Predicción
- Análisis rápido
- Aplicación en el negocio de los conocimientos resultantes
- Facilidad de uso, con herramientas accesibles a los usuarios de negocios.

Proceso



Herramientas



Oracle Discoverer Desktop

Oracle Developer Suite 10g

ORACLE

Copyright © 1996, 2003, Oracle. All rights reserved.



Power BI

IBM Watson
Analytics



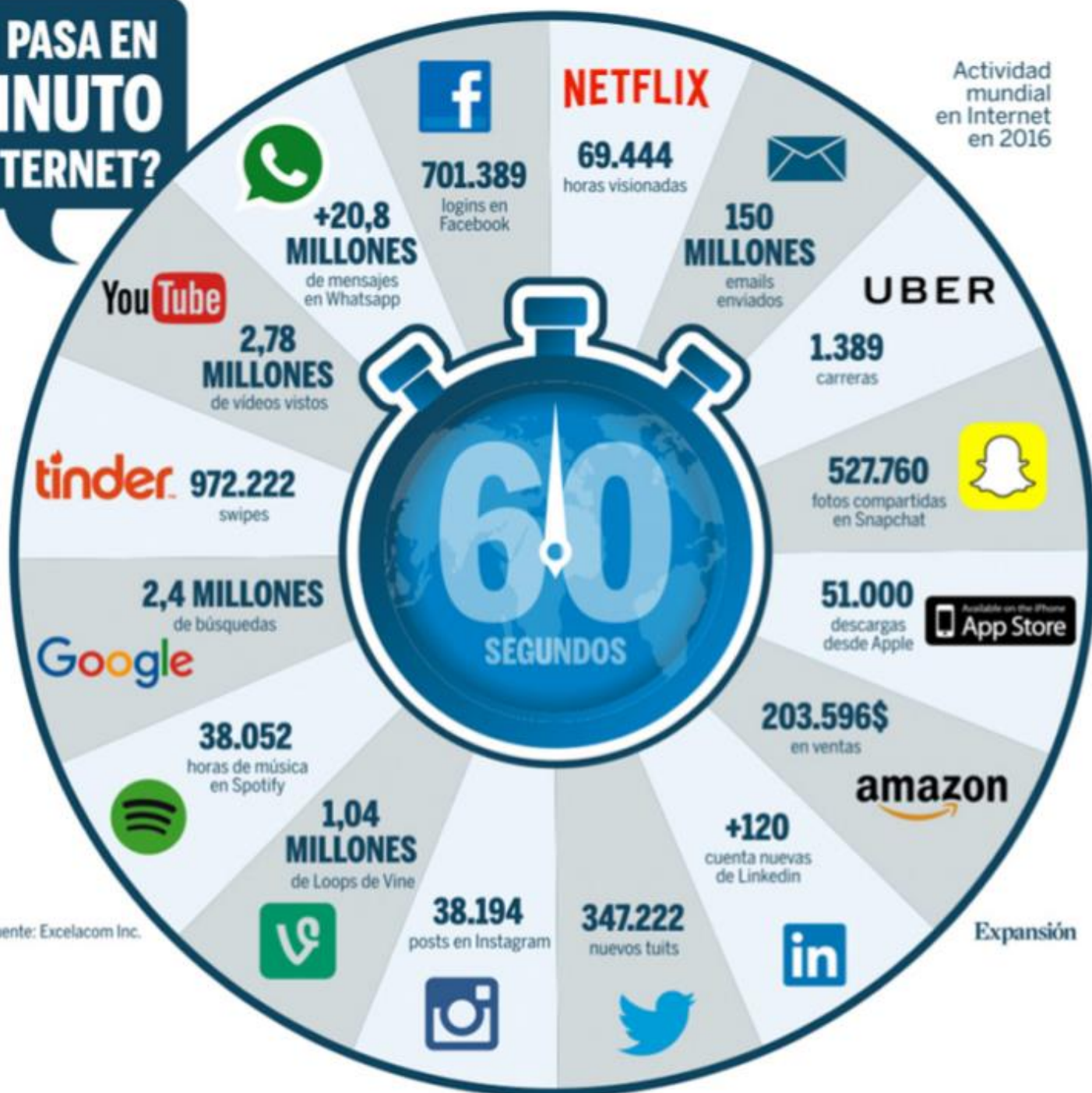
+ a b l e a u
S O F T W A R E

Qlik Q



¿QUÉ PASA EN 1 MINUTO EN INTERNET?

Actividad mundial
en Internet
en 2016



Fuente: Excelscom Inc.

Expansión

Concepto

Son datos de gran **variedad**,
en alto **volumen**,
distribuidos a alta **velocidad**,
que exigen formas
automatizadas e innovadoras
de procesamiento de la
información que permiten
una visión mejorada, toma
de decisiones, y la
automatización de procesos.



Gartner

Big Data requiere "nuevas formas de tratamiento de la información para el descubrimiento de una visión mejorada, toma de decisiones y la automatización de procesos."

Las Tres V del Big Data

Volumen

Exxabytes
Registros
Transacciones
Tablas



Velocidad

A tiempo real
Continuo



Variedad

Estructurados
No Estructurados
Semiestructurados



Tipos de Datos

10% Información Estructurada

Cifras, Bases de datos

80% Información NO Estructurada

Correos electrónicos, videos, imágenes, voz, tweets, post de Facebook,



etc

Fuentes

- Computadoras
- Dispositivos móviles
- Equipos de sonido, grabación, etc.
- Equipos de video, video vigilancia
- Cámaras fotográficas, scanner
- Sistemas GPS,
- Sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, Woreables etc., (IOT y M2M)
- Web y Redes Sociales
- Equipos biométricos
- Dispositivos de RFID y de Código de barras
- Otros



Propósito

- Análisis Actualizado
- Información en tiempo real
- Predicciones
- Generar mejor conocimiento
- Mejorar de la toma de decisiones

Catalizadores



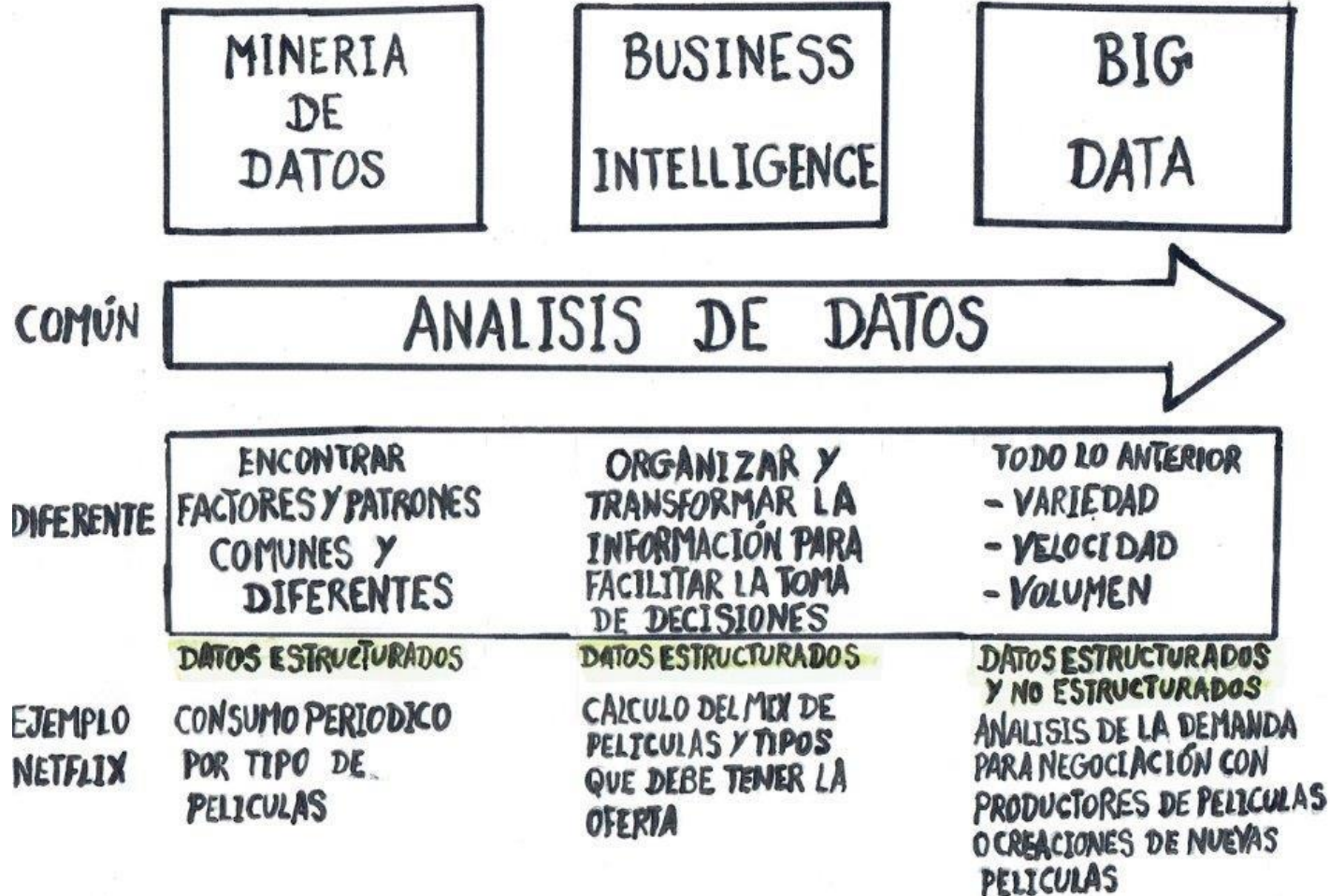
Soluciones



Retos

- El usuario empresarial esta mucho más empoderado de las herramientas de gestión de información.
- El CIO tiene que acelerar su proceso de actualización en gestión de datos y orientar la administración de infraestructura a terceros.
- Muchas decisiones de adquisición de soluciones de tecnologías de la información se están tomando en las áreas funcionales de la empresa.

Comparativa



Dominio
■ ■ ■ ■

www.dominio-consultores.com

Fernando Grados: fgrados@dominio-consultores.com



Dominio Consultores