Fundamentos de la Informática I
2022

PARAGUAYO
ALEMANA

25/08/2022

Business Intelligence

Prof. Carlos Núñez, PhD
carlos.nunez@upa.edu.py

UPA
UNIVERSIDAD PARAGUAYO ALEMANA
DE CIENCIAS APLICADAS
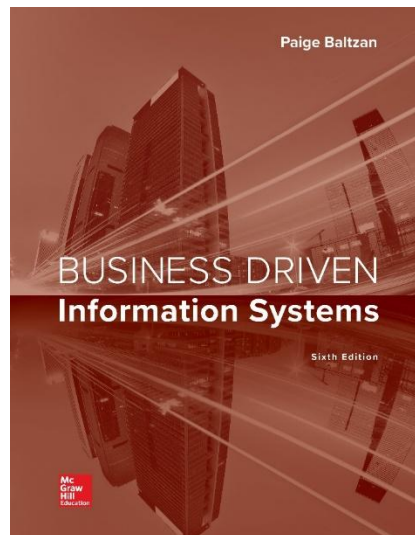
Business Intelligence

SECTION 6.2 – Business Intelligence

- Supporting Decisions with Business Intelligence
- The Business Benefits of Data Warehousing
- The Power of Big Data Analytics

LEARNING OUTCOMES

❖ Identify the advantages of using business intelligence to support managerial decision making

❖ Define data warehousing and data marts and explain how they support business decisions

❖ Describe the three organizational methods for analyzing big data.

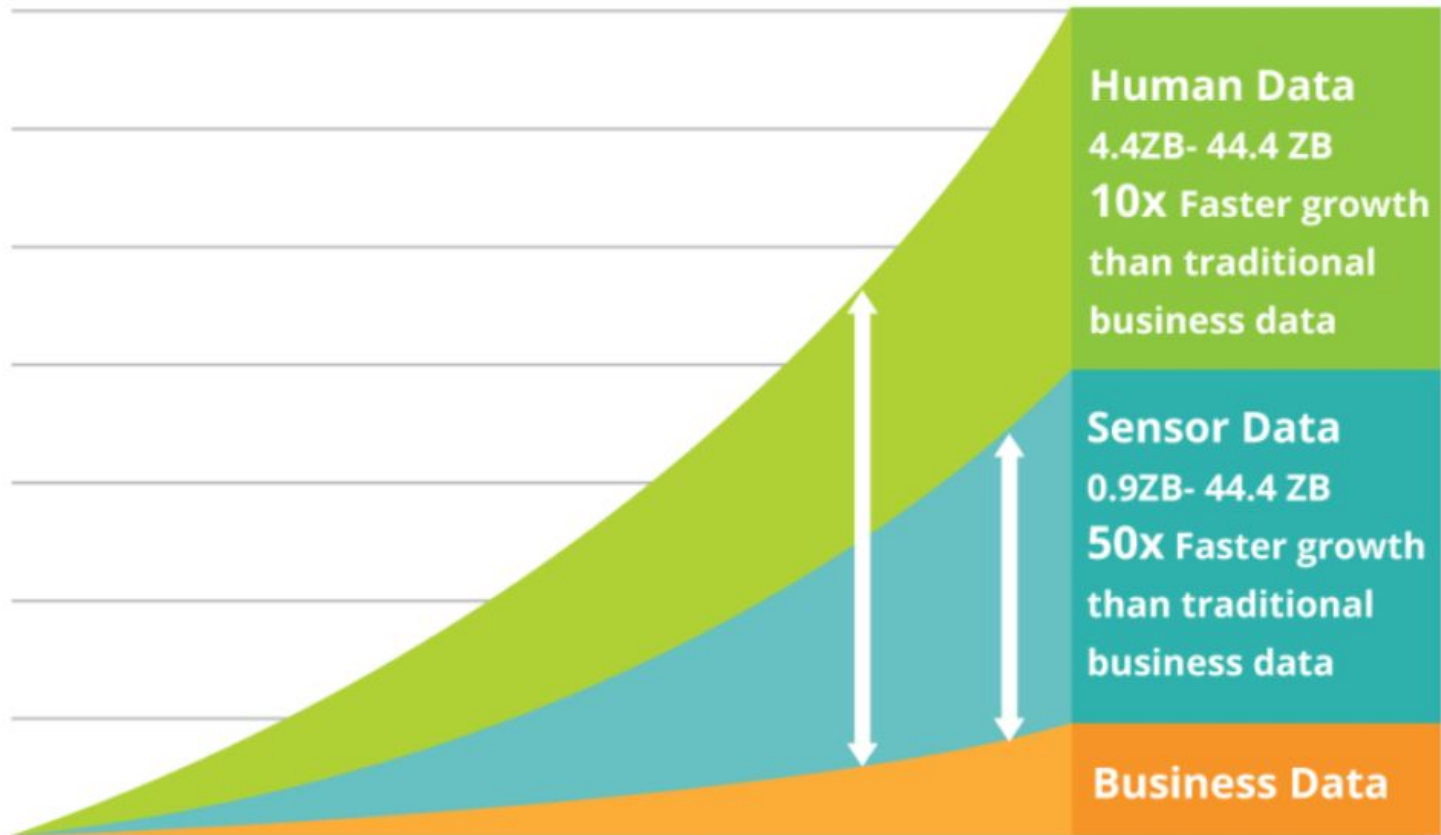# SUPPORTING DECISIONS WITH BUSINESS INTELLIGENCE

For Perspective:

The amount of information in the digital universe would fill a stack of iPad Air tablets reaching 2/3 of the way to the moon (157,674 miles/253,704 kilometers). By 2020, there will be 6.6 stacks.

Today, the average household creates enough data to fill 65 iPhones (32gb) per year. In 2020, this will grow to 318 iPhones.

Today, if a byte of data were a gallon of water, in only 10 seconds there would be enough data to fill an average house. In 2020, it will only take 2 seconds

The digital universe is doubling in size every two years and will multiply 10-fold between 2013 and 2020 – from 4.4 trillion gigabytes to 44 trillion gigabytes

The growth of human and machine-generated data

Human Data
4.4ZB- 44.4 ZB
10x Faster growth than traditional business data

Sensor Data
0.9ZB- 44.4 ZB
50x Faster growth than traditional business data

Business Data

Source: Inside big data

SUPPORTING DECISIONS WITH BUSINESS INTELLIGENCE

¿Toda esta cantidad de información digital cómo la manejamos? Es imposible hacerlo de forma manual o con herramientas tradicionales.

- Organizational data is difficult to access

- Organizational data contains structured data in database

- Organizational data contains unstructured data such as voice mail, phone calls, text messages, and video clips
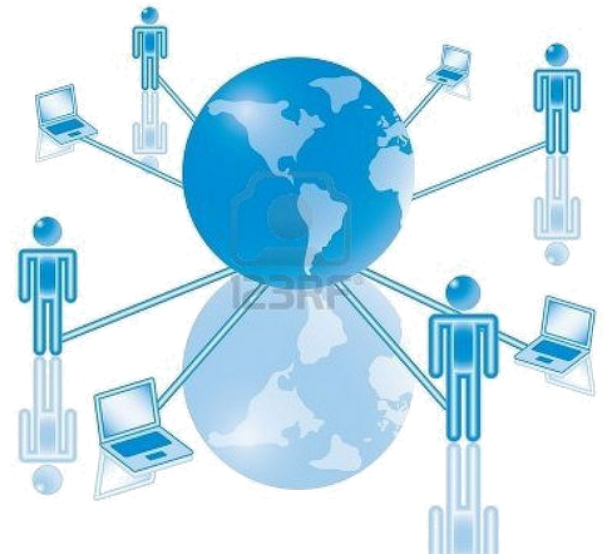
Reasons Business Analysis is Difficult from Operational Databases

Inconsistent Data Definitions
- Every department had its own method for recording data so when trying to share information, data did not match and users did not get the data they really needed.

Lack of Data Standards
- Managers needed to perform cross-functional analysis using data from all departments, which differed in granularities, formats, and levels.

Poor Data Quality
- The data, if available, were often incorrect or incomplete. Therefore, users could not rely on the data to make decisions.

Inadequate Data Usefulness
- Users could not get the data they needed; what was collected was not always useful for intended purposes.

Ineffective Direct Data Access
- Most data stored in operational databases did not allow users direct access; users had to wait to have their queries or questions answered by MIS professionals who could code SQL.

THE BUSINESS BENEFITS OF DATA WAREHOUSING

- Data Warehouses extend the transformation of data into information

- In the 1990's executives became less concerned with the day-to-day business operations and more concerned with overall business functions

- The data warehouse provided the ability to support decision making without disrupting the day-to-day operations

THE BUSINESS BENEFITS OF DATA WAREHOUSING

❖ **Data warehouse** – A logical collection of information – gathered from many different operational databases – that supports business analysis activities and decision-making tasks

Tool that enables business users, typically managers, to be more effective in many ways, including:

❖ Developing customer profiles.

❖ Identifying new-product opportunities.

❖ Improving business operations.

❖ Identifying financial issues.

❖ Analyzing trends.

❖ Understanding competitors.

❖ Understanding product performance.

THE BUSINESS BENEFITS OF DATA WAREHOUSING

Q: ¿Cuál es la diferencia principal entre una base de datos (Database) y un almacén de datos (Data warehouse)?

❖ The primary purpose of a data warehouse is to aggregate information throughout an organization into a single repository for decision-making purposes.

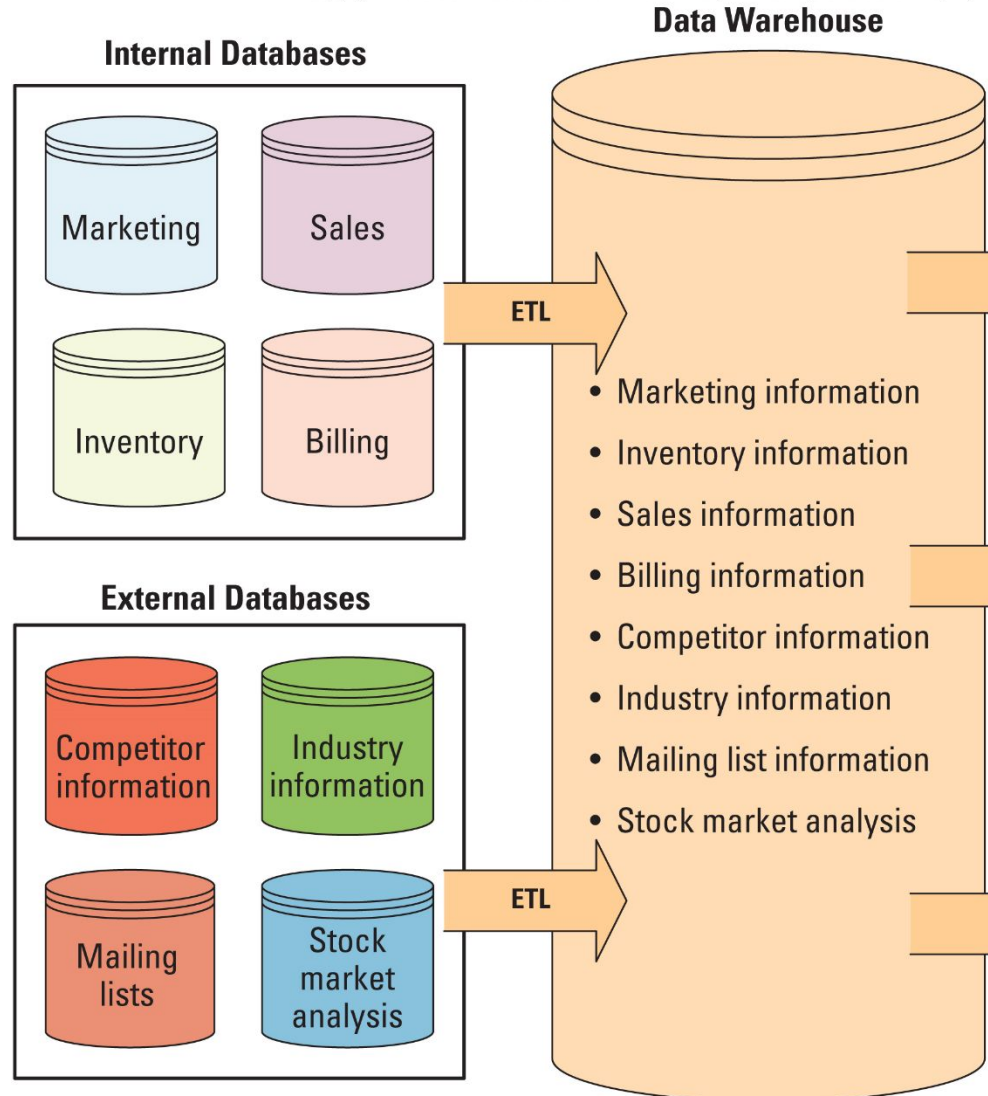❖ **Data aggregation –** Collection of data from for the purpose of data processing

Extraction, Transformation, and Loading (ETL)

A process that extracts information from internal and external databases, transforms the information using a common set of enterprise definitions, and loads the information into a data warehouse
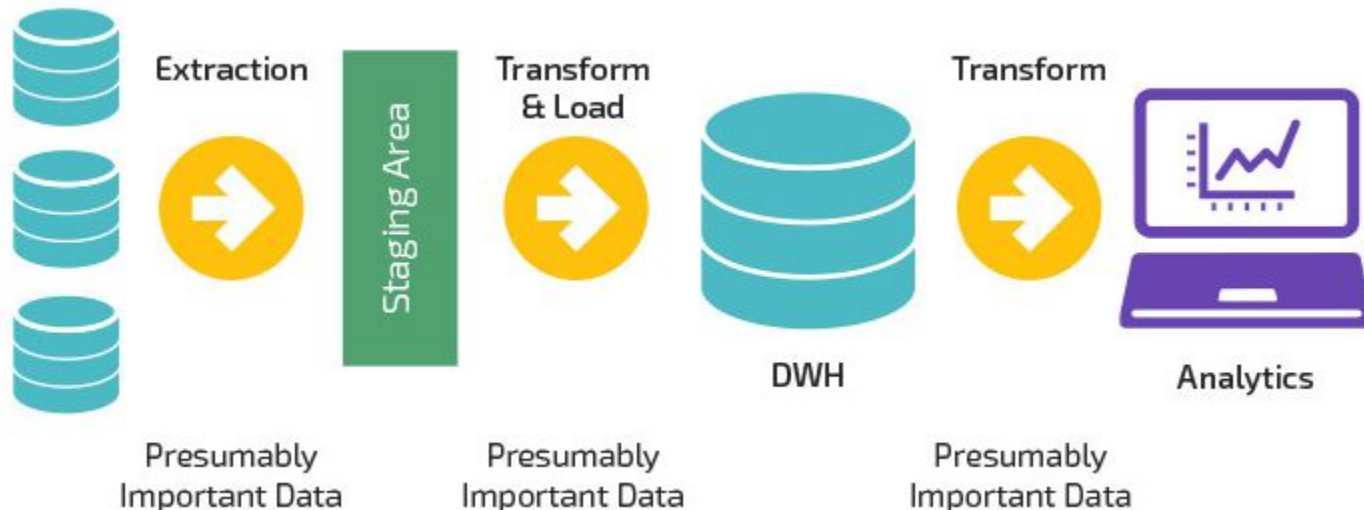
DATA WAREHOUSE

Copyright © McGraw-Hill Education. Permission required for reproduction or display.

a) Data Mart

b) Information Cleansing

c) Business Intelligence

Data Warehousing Components

a) Data Mart

❖ **Extraction, Transformation, and Loading (ETL)** – A process that extracts information from internal and external databases, transforms the information using a common set of enterprise definitions, and loads the information into a data warehouse
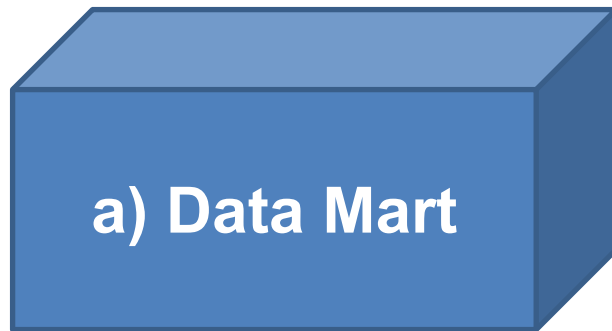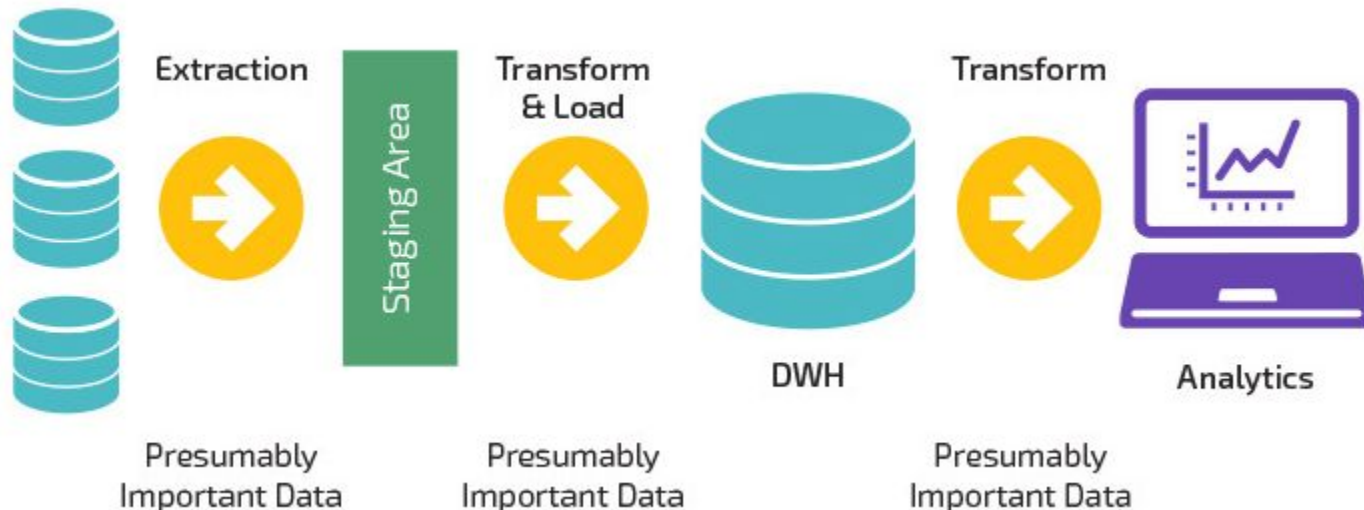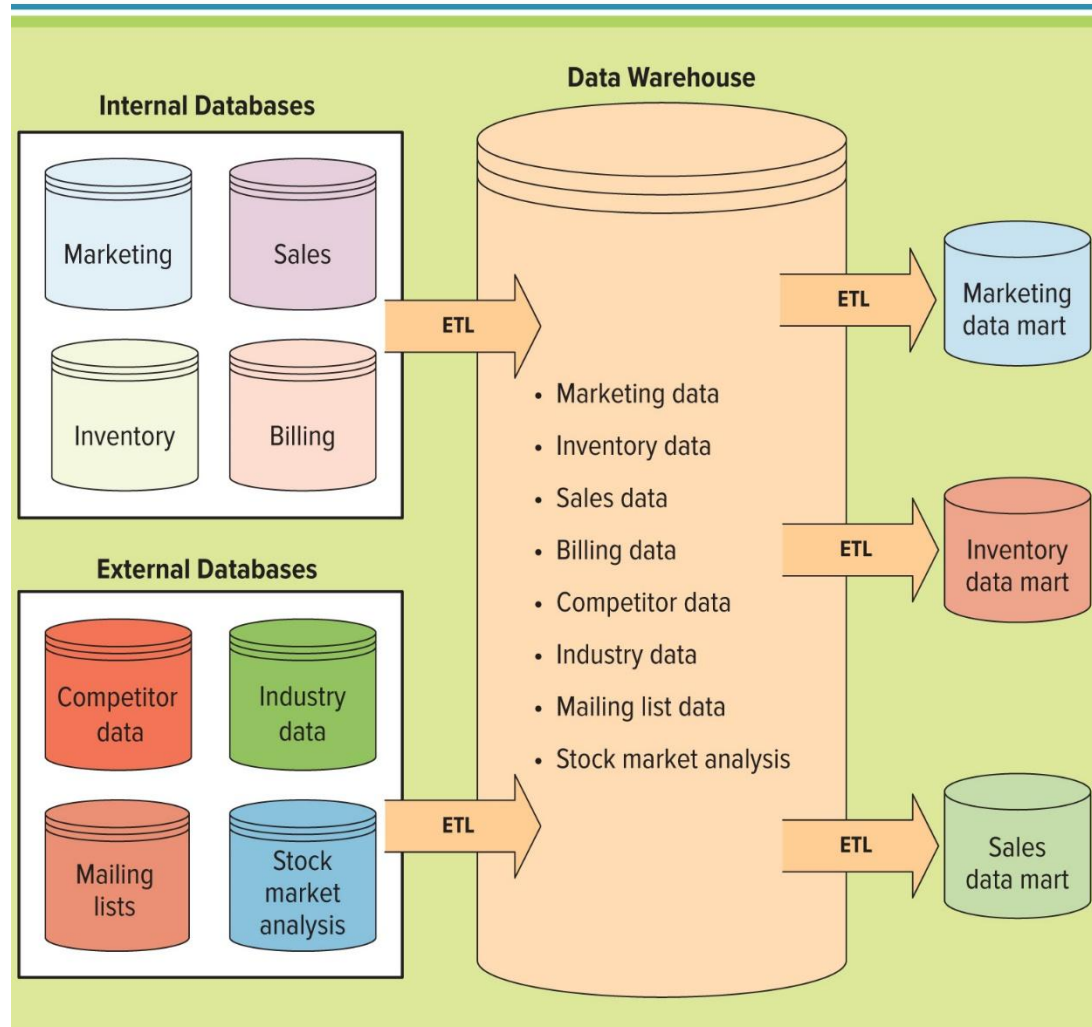
a) Data Mart

❖ **Data mart** – Contains a subset of data warehouse information

Data Warehouse Model

b) Information Cleansing or Scrubbing

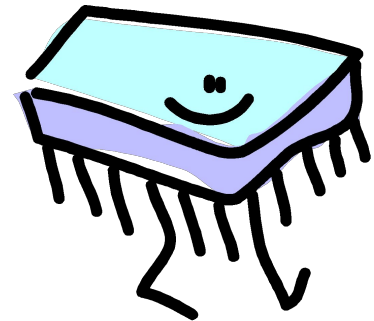An organization must maintain high-quality data in the data warehouse
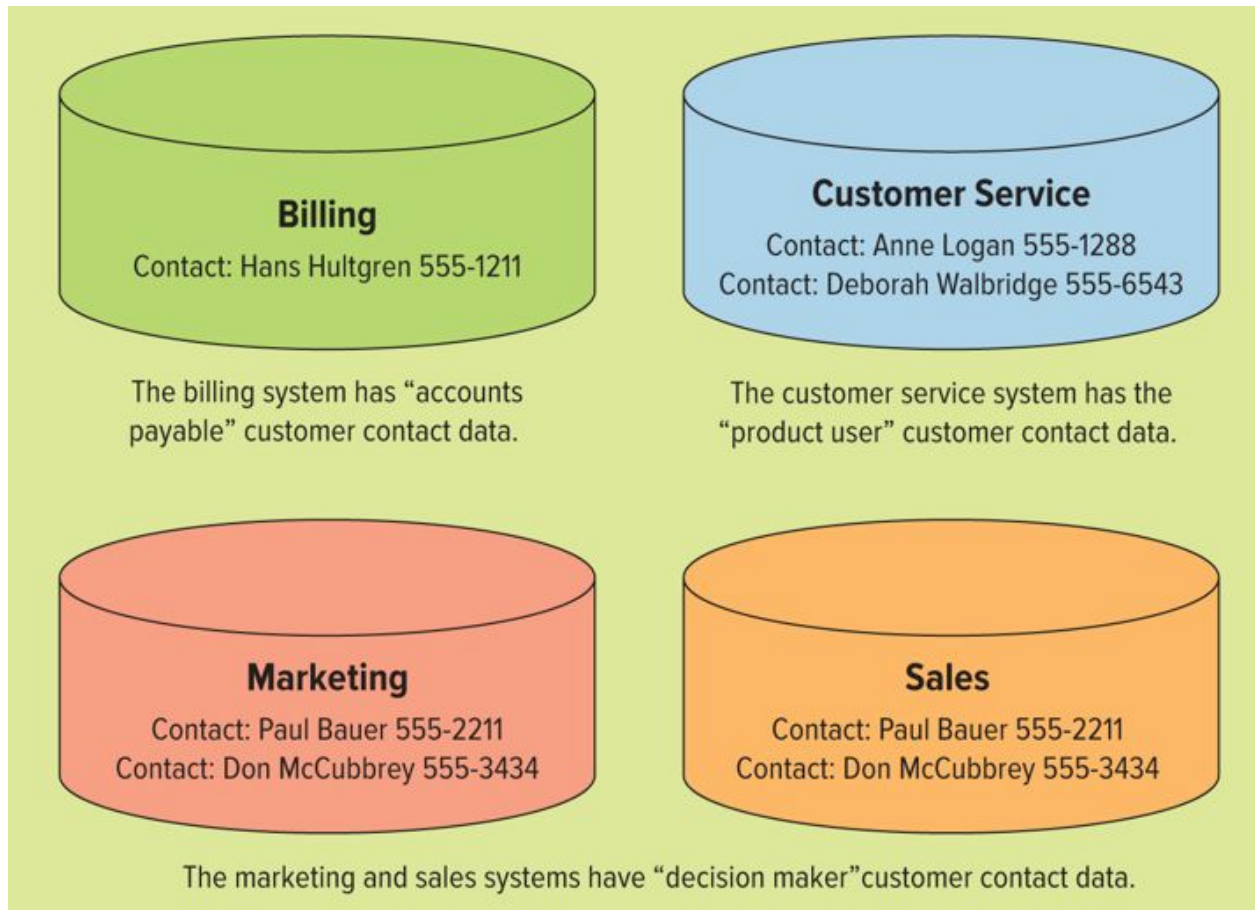

Alex Slobodkin/Getty Images RF

Dirty Data Problems

b) Information Cleansing or Scrubbing

❖ An organization must maintain high-quality data in the data warehouse

❖ **Information cleansing or scrubbing** – A process that weeds out and fixes or discards inconsistent, incorrect, or incomplete information

b) Information Cleansing or Scrubbing

Billing
Contact: Hans Hultgren 555-1211

The billing system has "accounts payable" customer contact data.

Customer Service
Contact: Anne Logan 555-1288
Contact: Deborah Walbridge 555-6543

The customer service system has the "product user" customer contact data.

Marketing
Contact: Paul Bauer 555-2211
Contact: Don McCubbrey 555-3434

Sales
Contact: Paul Bauer 555-2211
Contact: Don McCubbrey 555-3434

The marketing and sales systems have "decision maker" customer contact data.

Contact Information in an Operational System

b) Information Cleansing or Scrubbing

Sales

Customers:
JD0021 *Jane Doe*
BL0557 *Bob Lake*
JS0288 *Judy Smith*
PB0092 *Pat Burton*

Customer Service

Customers:
10622FA *Susan Brown*
10472FB *Judie R Smithe*
10772FA *Patti Burten*
10922MC *Larry Trump*

Billing

Customers:
000980 *Burton, Tricia*
02670 *Smith, Judie*
000466 *Burton, Patricia*
006777 *Lake, Robert P.*

Customer Data

Customers:
10001 *Jane Doe*
10002 *Robert P. Lake*
10003 *Judie R. Smith*
10004 *Patricia Burton*

Standardizing Customer Name from Operational Systems

b) Information Cleansing or Scrubbing

Cleansing

- Missing records or attributes
- Redundant records
- Missing keys or other required data
- Erroneous relationships or references
- Inaccurate or incomplete data
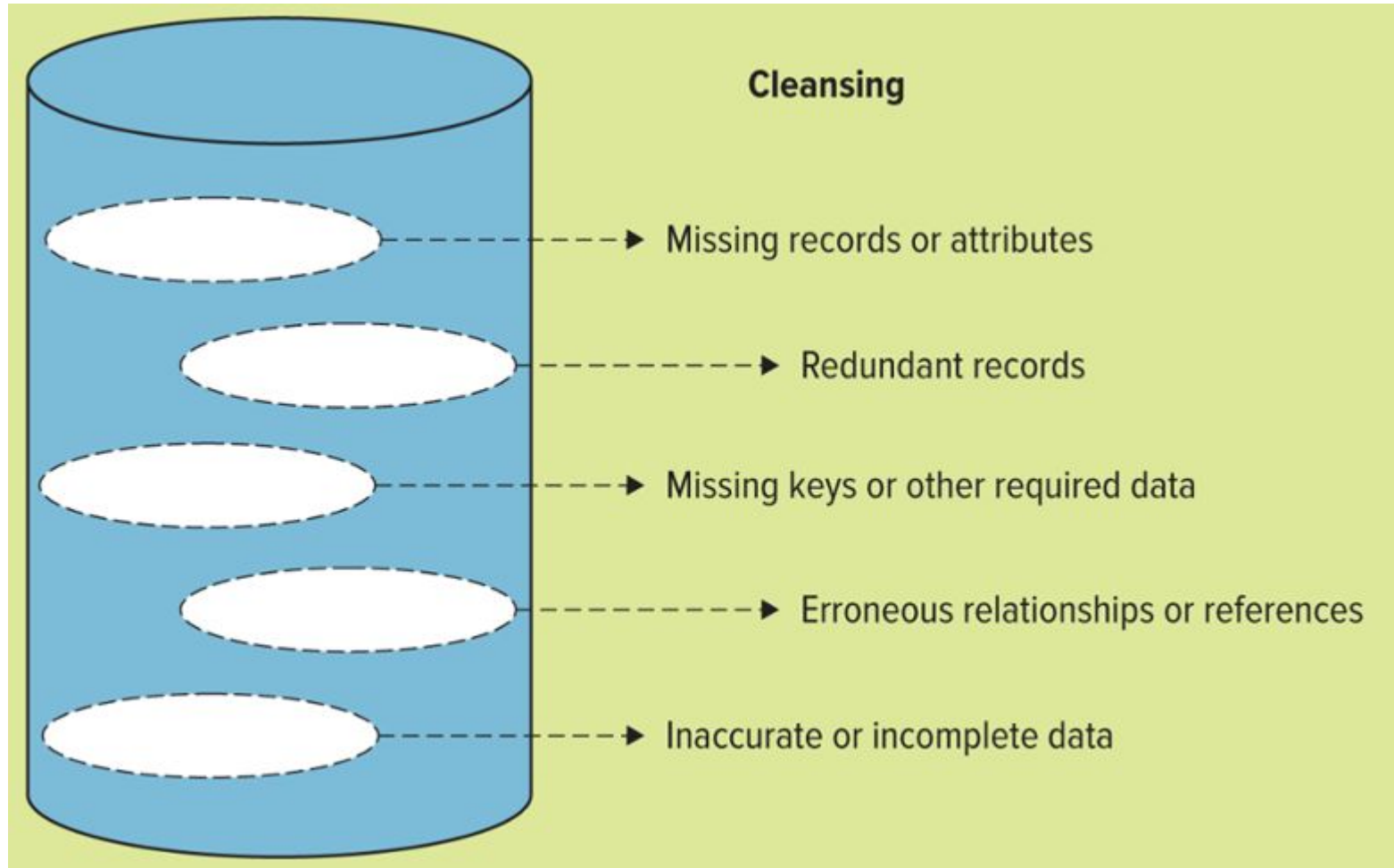
Information Cleansing Activities

Q: Puedes dar ejemplos de cómo los errores pueden afectar a una organización?

The butterfly effect, an idea from chaos theory in mathematics, refers to the way a minor event—like the movement of a butterfly's wing—can have a major impact on a complex system like the weather. Dirty data can have the same impact on a business as the butterfly effect. Organizations depend on the movement and sharing of data throughout the organization, so the impact of data quality errors are costly and far-reaching. Such data issues often begin with a tiny mistake in one part of the organization, but the butterfly effect can produce disastrous results, making its way through MIS systems to the data warehouse and other enterprise systems. When dirty data or low-quality data enters organizational systems, a tiny error such as a spelling mistake can lead to revenue loss, process inefficiency, and failure to comply with industry and government regulations. Explain how the following errors can affect an organization:

- A cascading spelling mistake
- Inaccurate customer records
- Incomplete purchasing history
- Inaccurate mailing address
- Duplicate customer numbers for different customers

c) Business Intelligence

The Problem: Data Rich, Information Poor

- Businesses face a data explosion as digital images, email in-boxes, and broadband connections doubles by 2010

- The amount of data generated is doubling every year

- Some believe it will soon double monthly

- As businesses increase their reliance on enterprise systems such as CRM, they are rapidly accumulating vast amounts of data.

muchos DATOS – Poca INFORMACION

c) Business Intelligence

The Solution: Business Intelligence

Business Data into Business Intelligence

Improving the quality of business decisions has a direct impact on costs and revenue

BI enables business users to receive data for analysis that is:

- Reliable
- Consistent
- Understandable
- Easily manipulated

EJEMPLOS:

Venta minorista y ventas: predicción de ventas; determinación de niveles de inventario correctos y cronogramas de distribución entre puntos de venta; y prevención de pérdidas.

Banca: Pronostique los niveles de préstamos incobrables y el uso fraudulento de tarjetas de crédito, el gasto de tarjetas de crédito por parte de nuevos clientes y qué tipos de clientes responderán mejor (y calificarán) para nuevas ofertas de préstamos.

Gestión de operaciones: predicción de fallas de maquinaria; Encontrar factores clave que controlan la optimización de la capacidad de fabricación.

Corretaje y comercio de valores: predecir cuándo cambiarán los precios de los bonos; pronosticar el rango de fluctuaciones de stock para problemas particulares y el mercado general; determinar cuándo comprar o vender acciones.

Seguro: Previsión de montos de reclamos y costos de cobertura médica; clasificar los elementos más importantes que afectan la cobertura
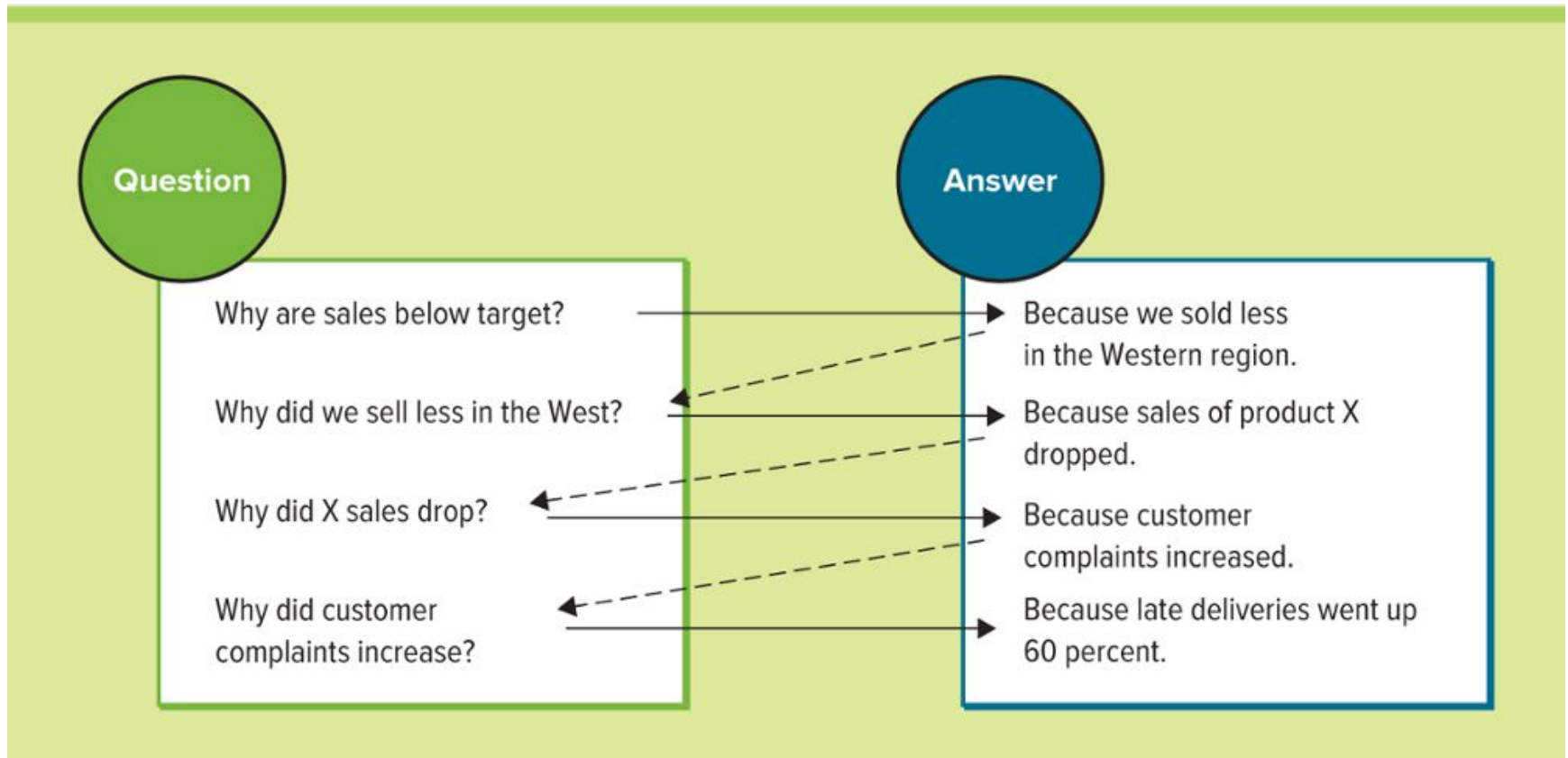
¿Puedes mencionar otros ejemplos?

c) Business Intelligence

BI Can Answer Tough Questions:

Where the business has been?

Where is now?

Where it will be in the near future?

**Key? -> single source
of trusted data!**

Big data - A collection of large, complex data sets, including structured and unstructured data, which cannot be analyzed using traditional database methods and tools.

The four common characteristics of big data:

- **Variety**
- **Veracity**
- **Volume**
- **Velocity**

https://youtu.be/CcDCqW88uNY

VARIETY

- Different forms of structured and unstructured data
- Data from spreadsheets and databases as well as from email, videos, photos, and PDFs, all of which must be analyzed

VERACITY

- The uncertainty of data, including biases, noise, and abnormalities
- Uncertainty or untrustworthiness of data
- Data must be meaningful to the problem being analyzed
- Must keep data clean and implement processes to keep dirty data from accumulating in systems

THE POWER OF BIG DATA ANALYTICS

VOLUME

- The scale of data
- Includes enormous volumes of data generated daily
- Massive volume created by machines and networks
- Big data tools necessary to analyze zettabytes and brontobytes

VELOCITY

- The analysis of streaming data as it travels around the Internet
- Analysis necessary of social media messages spreading globally

THE POWER OF BIG DATA ANALYTICS

❖ **Structured data** – Contains a defined length, type, and format and includes numbers, dates, or strings

- ▪ **Machine-generated data:** created by a machine without human intervention -> sensor data, point-of-sale data, and web log (blog) data.
- ▪ **Human-generated data:** is data that humans, in interaction with computers, generate -> input data, click-stream data, or gaming data.

❖ **Unstructured data** – Not defined, does not follow a specified format, and is typically freeform text

- ▪ **Machine-generated unstructured data:** satellite images, scientific atmosphere data, and radar data.
- ▪ **Human-generated unstructured data:** text messages, social media data, and emails.

THE POWER OF BIG DATA ANALYTICS

## Structured and Unstructured Data Examples

Structured Data	Unstructured Data
Sensor data	Satellite images
Weblog data	Photographic data
Financial data	Video data
Click-stream data	Social media data
Point of sale data	Text message
Accounting data	Voice mail data

Two primary computing models that have shaped the collection of big data include:

Distributed computing & Virtualization

Distributed computing – Processes and manages algorithms across many machines in a computing environment

31
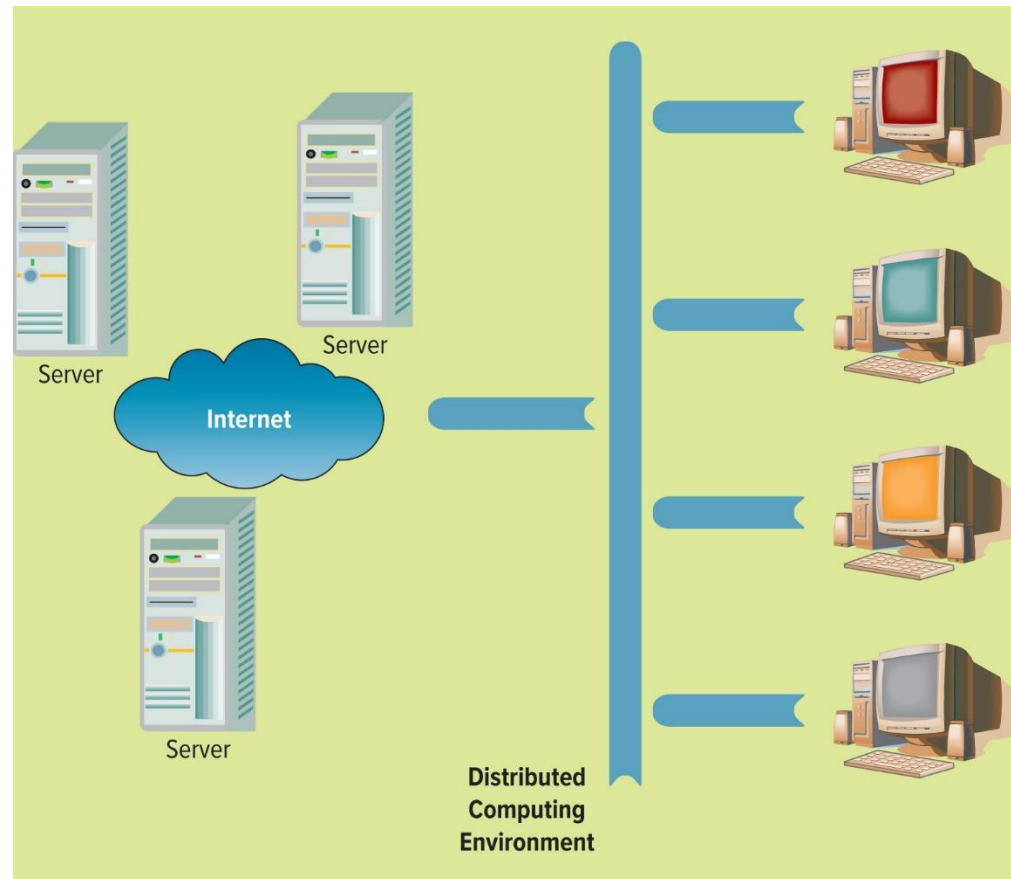
THE POWER OF BIG DATA ANALYTICS

Virtualization – The creation of a virtual (rather than actual) version of computing resources, such as an operating system, a server, a storage device, or network resources
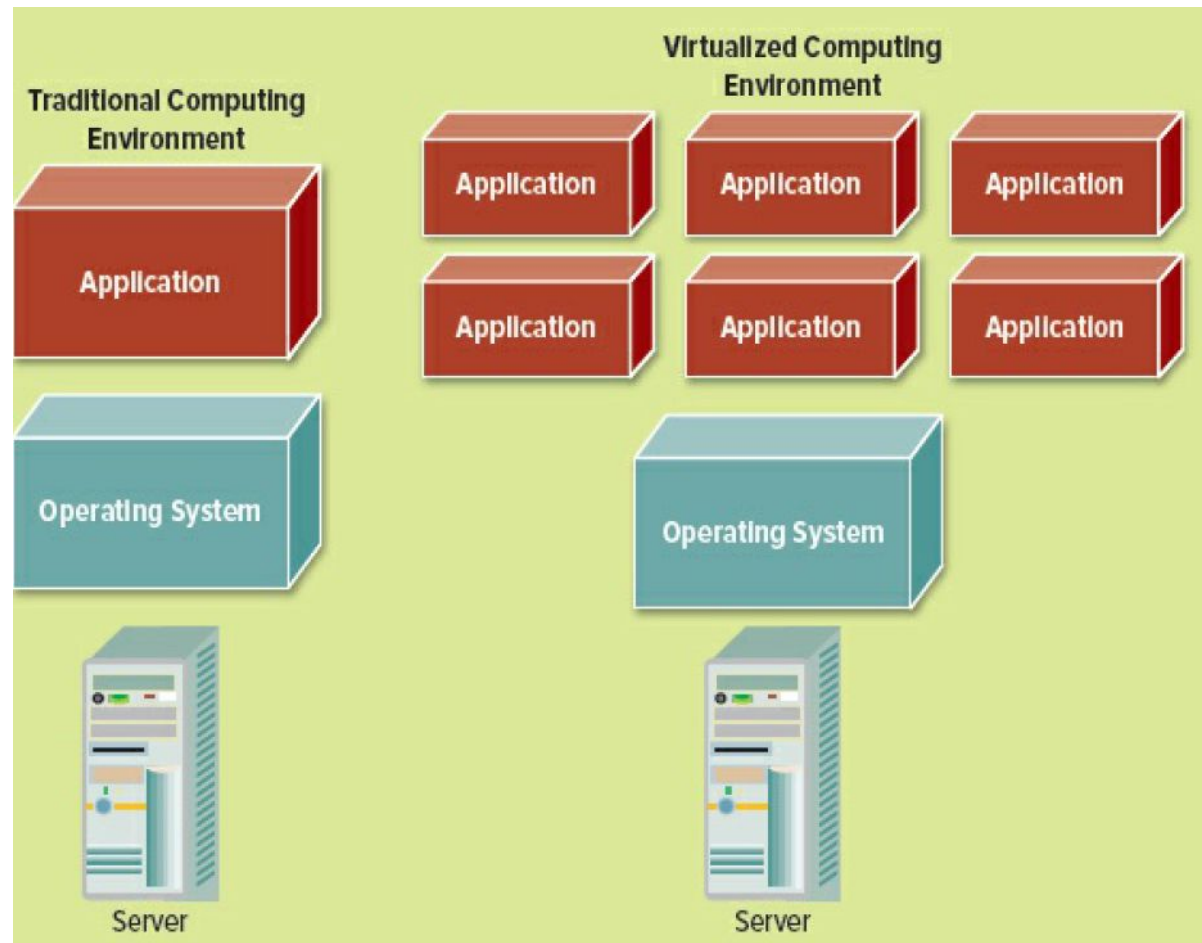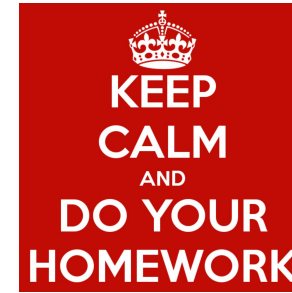
¿PREGUNTAS?

¿Qué aprendimos hoy?

TAREA

TEC#10 - Guía de Trabajo en el Classroom

Individual

Entrega: Martes 24/08/2021 – 23:59