



BD GUIDANCE

GLOBAL EXPANSION EXPERTS



MATERIAL DE
REPASO PARA
ESTUDIANTES DE
BIG DATA



BD GUIDANCE

BIG DATA

Modulo 1

Big Data

```
graph TD; A[Big Data] --> B[Dedicated To]; A --> C[Requirements]; A --> D[Data Processed used by]; B --> E[analysis]; B --> F[Storage of large collections of data]; C --> G[Combining multiple unrelated datasets]; C --> H[Processing unstructured data]; D --> I[Enterprise Apps]; D --> J[Another DW to enrich existing data];
```

Dedicated To

Processing

analysis

Storage of **large** collections of data

Requirements

Combining multiple unrelated datasets

Processing unstructured data

Data Processed used by

Enterprise Apps

Another DW to enrich existing data

Big Data

```
graph TD; A[Big Data] --> B[BD type of Data]; A --> C[Benefits]; B --> D[Human generated]; B --> E[Machine generated]; C --> F[Improved decision-making]; C --> G[Scientific discoveries]; C --> H[More detailed records]; C --> I[Fault and fraud detection]; C --> J[Operational optimisation]; C --> K[ID of new markets];
```

The diagram is a hierarchical tree structure. At the top is a dark blue box with the text 'Big Data'. A line from this box splits into two branches. The left branch leads to a dark blue box labeled 'BD type of Data'. From this box, two lines lead down to 'Human generated' and 'Machine generated'. The right branch from the top box leads to a dark blue box labeled 'Benefits'. From this box, a line leads down to a horizontal line, which then branches into seven separate lines leading to: 'Improved decision-making', 'Scientific discoveries', 'More detailed records', 'Fault and fraud detection', 'Operational optimisation', and 'ID of new markets'. All boxes are dark blue with white text. The background is a blurred image of a desk with a keyboard, a spiral notebook, and a pen.

BD type of Data

Human generated

Machine generated

Benefits

Improved decision-making

Scientific discoveries

More detailed records

Fault and fraud detection

Operational optimisation

ID of new markets

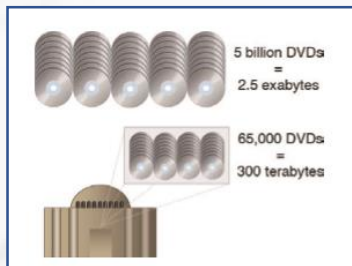


CHARACTERISTICS OF DATA

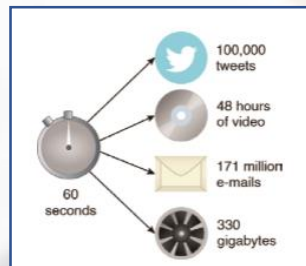
Modulo 1

Characteristics of Data in Big Data Environments

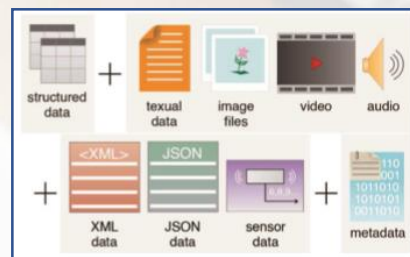
Volume



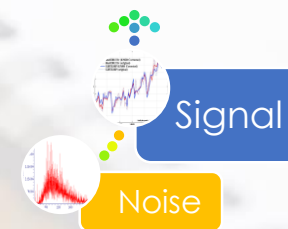
Velocity



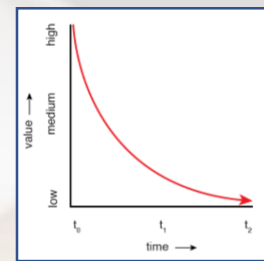
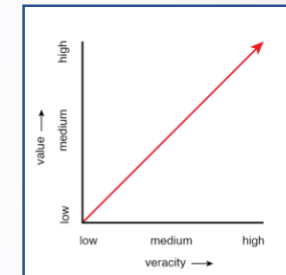
Variety



Veracity



Value



Fundamental Terminology

Data Sets

Data Analysis

Analytics

BI

KPI

Big Data Business & Technology Drivers

Analytics & Data Science

Digitalization

Affordable Tech & HW

Social Media

Hyper-connected communities & devices

Cloud Computing

OLTP & OLAP

OLTP: Transaction Data

OLAP: Analysed Data

Data definitions

Data Warehouse

Data Marts

Hadoop



TYPES OF DATA

Modulo 1

Structured Data

- Conforms to a data model
- Tabular form
- Can be relational
- Not need of preprocessing of storage

Unstructured Data

- Does not conform to data model
- Inconsistent & Non-relational
- Textual or binary form
- Faster growth than structured
- Need logic to pre-process
- NoSQL is used to store Unstructured data

Semi Structured Data

- Defined level of structure and consistency
- Cannot be relational
- Mostly textuals
- Sources:
 - Emails
 - Spreadsheets
 - RSS
 - Sensor Data
- Special Pre-processing and storage requirements

Types of Data Analysis

Quantitative Analysis

- Statistical Techniques
- Large amount of data
- Applied in a generalised manner to the data set
- Results are numerical

Qualitative Analysis

- Small samples in great depth
- Cannot be generalised
- Cannot be measured numerically or for numerical comparison
- Results are descriptive

Data Mining

- Based on techniques that sift through massive datasets
- Seeks to find trends and patterns
- Extracts hidden or unknown patterns
- Basis for predictive analysis and BI.



TYPES OF ANALYTICS

Modulo 1

Descriptive

Answer questions about events that have already happened

Via Ad-hoc reporting and dashboards

Diagnostic

Descriptive + Using questions that focus on the reason behind the event

Multiple sources

Structure Lets

Interactive visualization tools

More complex queries to an OLAP

Predictive

Determine the outcome of an event that might happen in the future

Base on

- Patterns
- Trends
- Exceptions

ID of

- Risks
- Opportunities

Large DataSet

- Internal
- External

Statistical techniques

Quantitative analysis

Machine Learning

Data Mining

Prescriptive

Predictive + prescribing actions that should be taken

From exploratory to advisory

Datasets

- Internal
- External
- Rules of business



MACHINE LEARNING

Modulo 1

Machine Learning

Process of teaching computers to **Learn from existing Data** and apply the acquired knowledge to **formulate predictions about unknown Data**:

- 1. Identify Patterns
- 2. Classify new or unseen data on known patterns
- Adjust using a FeedBack Loop

Machine Learning

Supervised

- Categories are already known
- Develops the understanding of categories
- Categories unknown data

Unsupervised

- Categories are unknown
- Categorise data by similar attributes



BUSSINES INTELLIGENCE AND BIG DATA

Modulo 1

Traditional BI

Descriptive & Diagnostic Analysis

Historical and current events

Not intelligent

Reports KPIs Throught

- Ad Hoc Rporting
- Dashboards

Big Data BI

Traditional BI + Semistructured and Unstructured

Predictive + Prescriptive Analytics

Understands the way a business Works

Focus on multiple Business Processes

Algunos Consejos

- Hacer los Test del Sistema
 - Tener lista la Web Cam
 - Pearson Ingles talks
- Estar Completamente SOLO
 - Escritorio VACIO
- Tener la Cédula a la mano
 - Ojo con los LABS