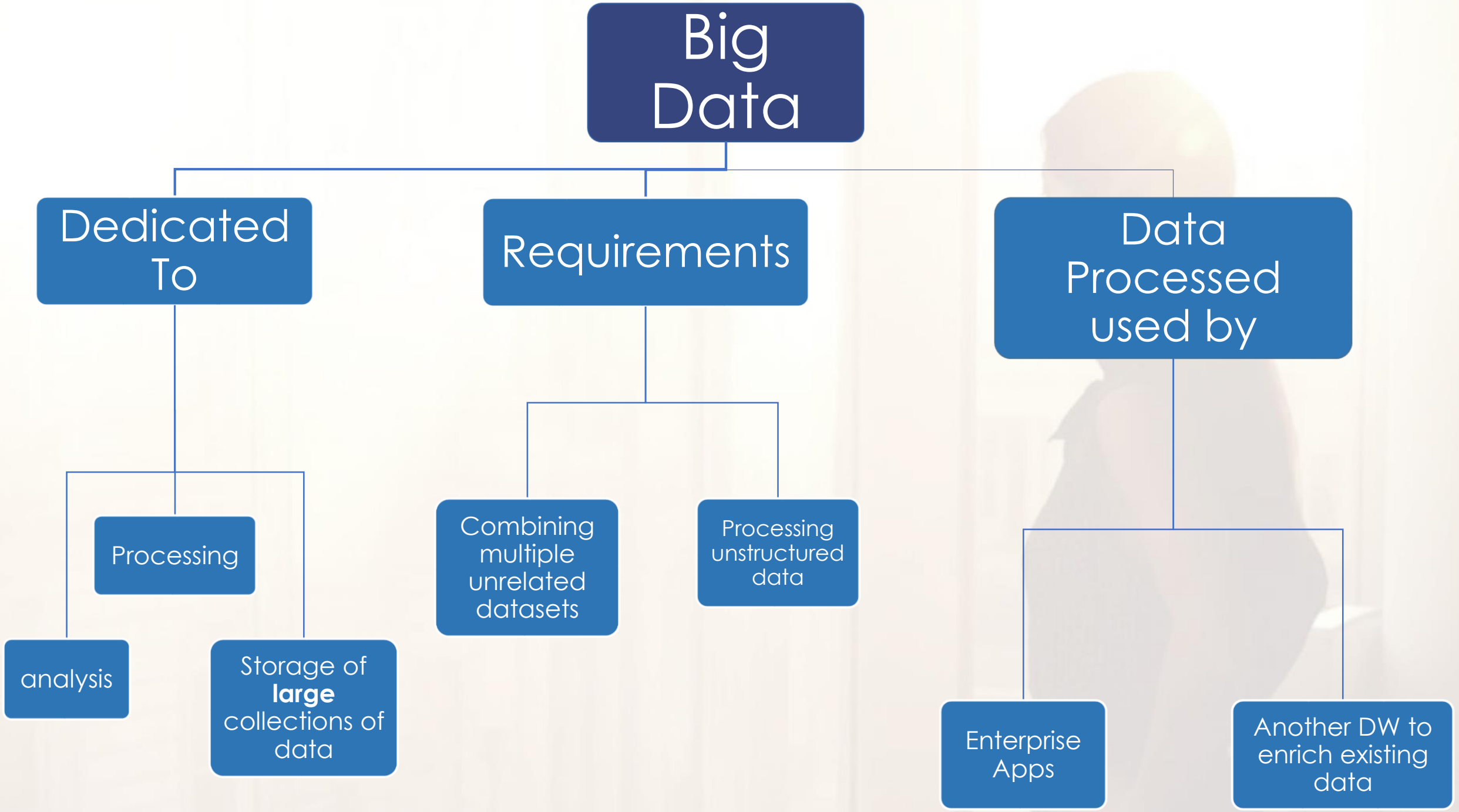

BD GUIDANCE
GLOBAL EXPANSION EXPERTS

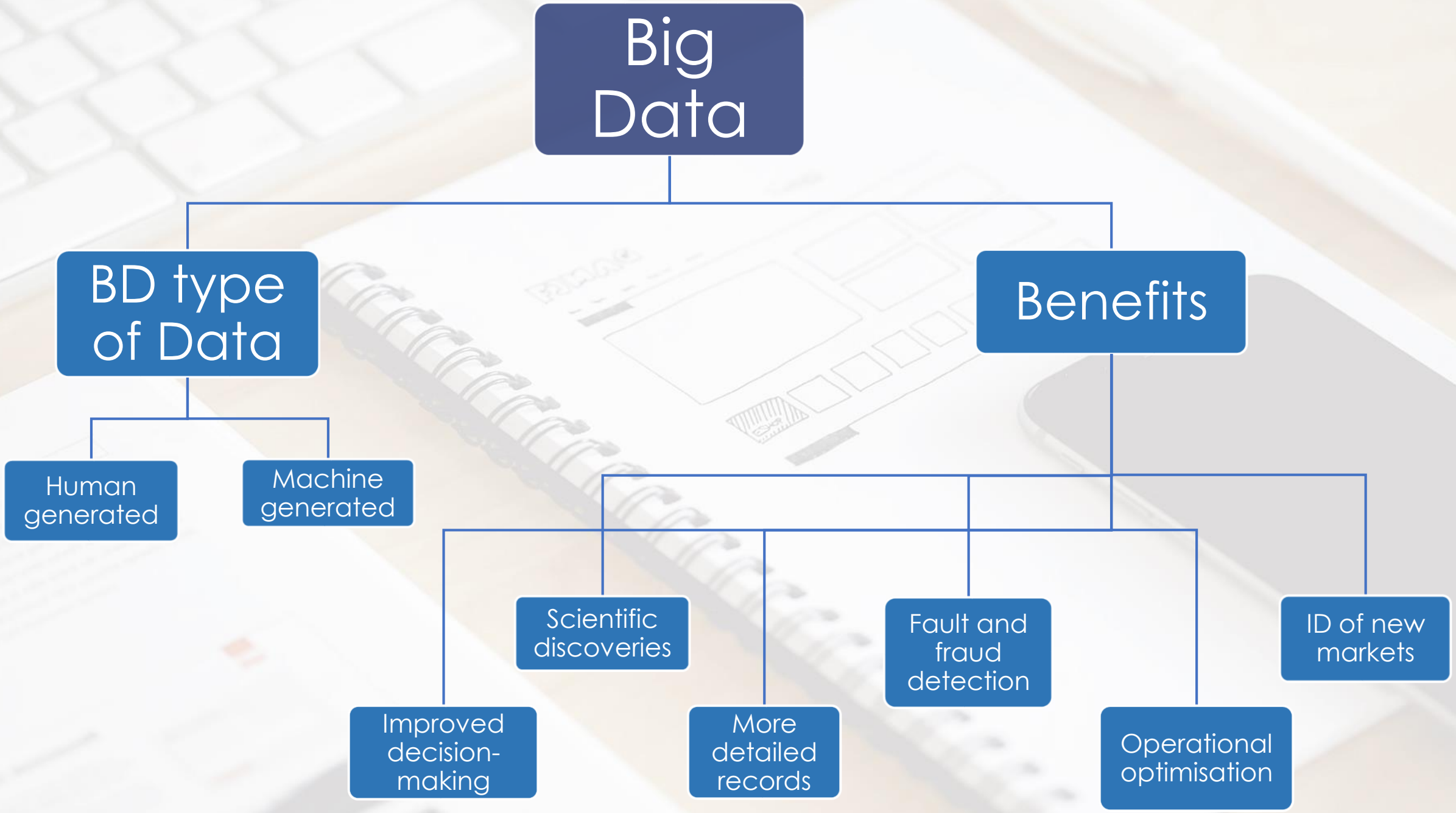
BD GUIDANCE

MATERIAL DE REPASO PARA ESTUDIANTES DE BIG DATA

BIG DATA

Modulo 1


Big Data
Dedicated To
Requirements
Data Processed used by
Processing
analysis
Storage of **large** collections of data
Combining multiple unrelated datasets
Processing unstructured data
Enterprise Apps
Another DW to enrich existing data


Big Data
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

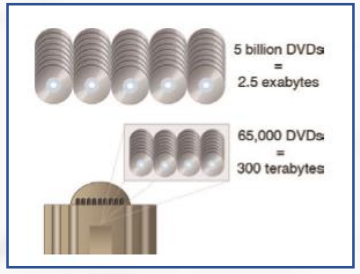
BD GUIDANCE

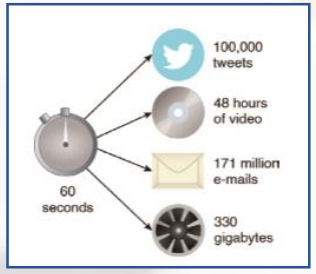
CHARACTERISTICS OF DATA

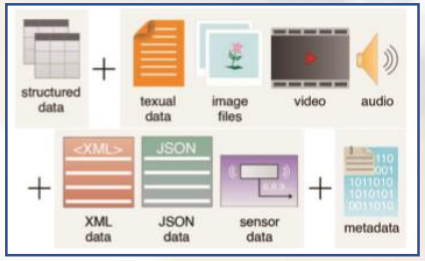
Modulo 1

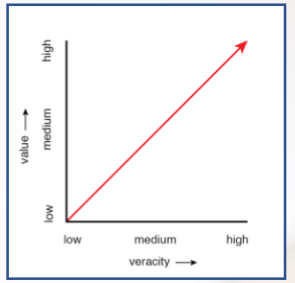
Characteristics of Data in Big Data Environments

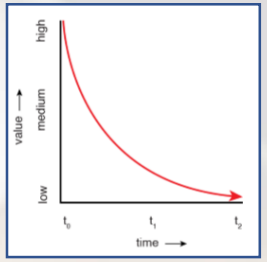
- Volume
- Velocity
- Variety
- Veracity
- Value











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: Analised Data

Data definitions

- Data Warehouse
- Data Marts
- Hadoop

BD GUIDANCE

TYPES OF DATA

Modulo 1

Structured Data

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

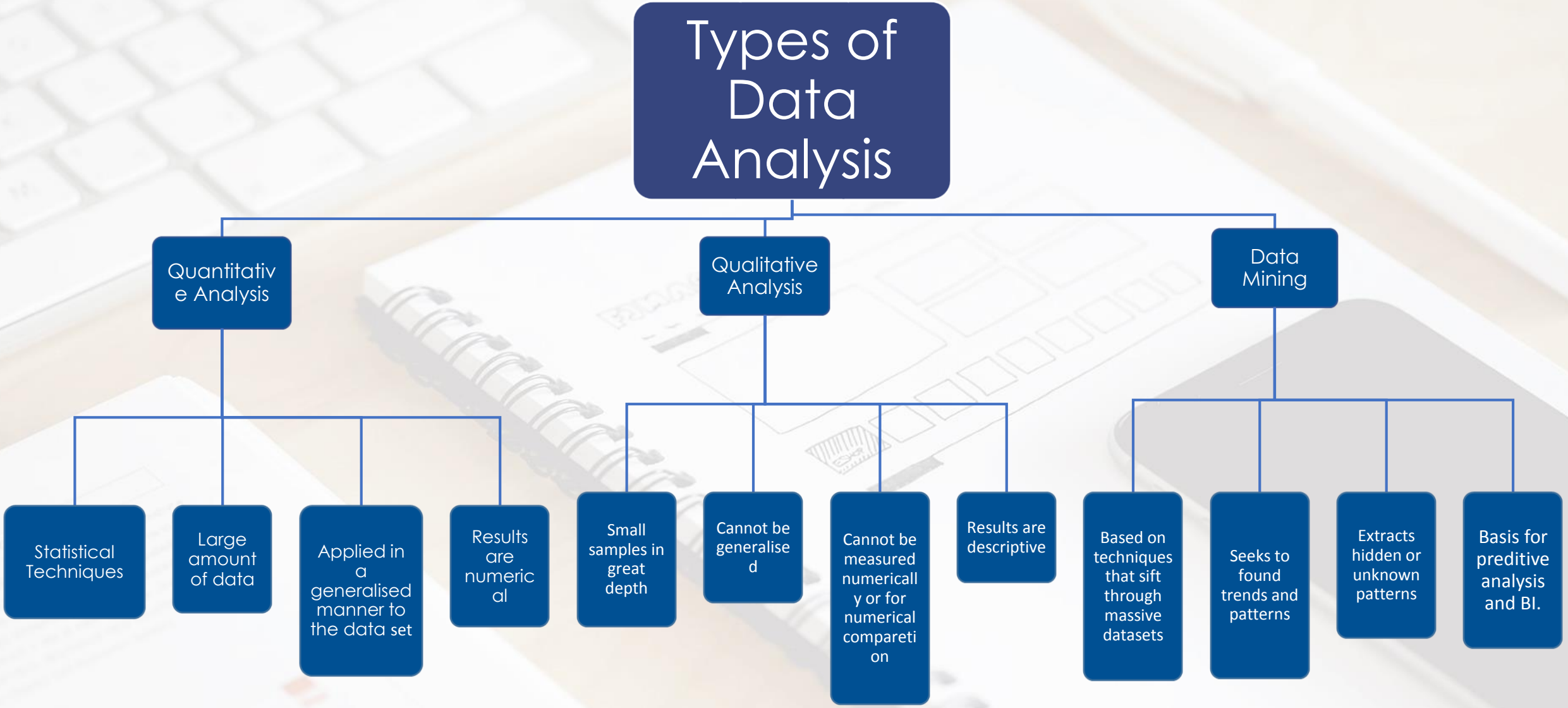
Unstructured Data

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

Semi Structured Data

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

Types of Data Analysis



BD GUIDANCE

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 tolos

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 exporatory to advisory

Datsets
- Internal
- External
- Rules of business

BD GUIDANCE

MACHINE LEARNING

Modulo 1

Machine Learning

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

- 1. Identify Paterns
- 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

BD GUIDANCE

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