UNTAR
Universitas Tarumanagara

Terakreditasi BAN PT A (Unggul)

UNTAR untuk INDONESIA

Introduction Big Data

Bagus Mulyawan

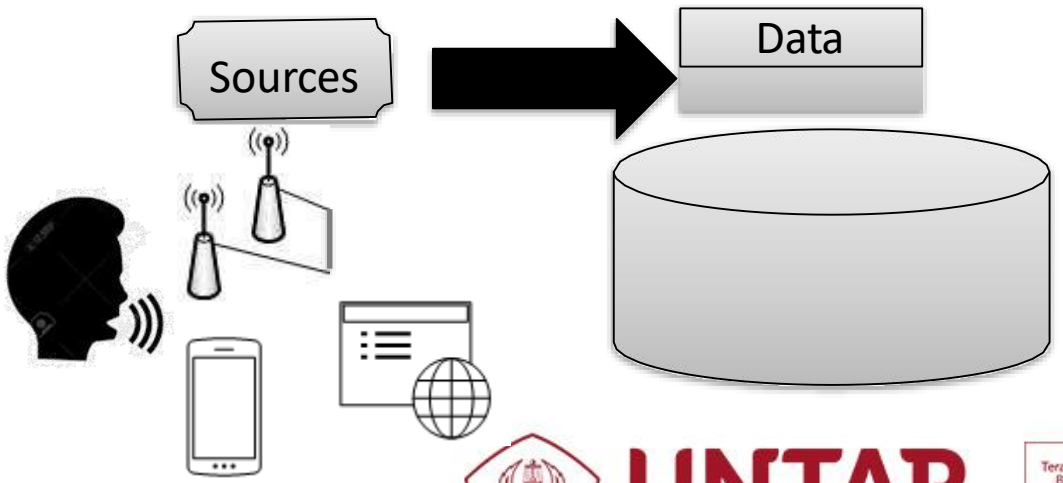
What is 'Data' ?!

5min to answer…

Data

Are the raw facts (or descriptions of facts) that were taken, observed, recorded, agreed, such as words, numbers, observations, surveys, etc.

- Data are unprocessed facts, figures, schemas, etc.
- In Information Systems (Computerized), Data is the input in the computer system.
- Data doesn't have a meaning !



Data

- 2 kinds of Data :
  - **Qualitative:** textual or symbolic
  - **Quantitative:** numerical

	Quantitative	Qualitative
Concept/Definition	Valuated facts	Describted facts
Methodology	Collected by measurement tools	Collected by observation
Analysis	Performed by statistical and numerical methods	Perfomed by specific adapted methods of classification, quantification, etc.
Results	Reported through statistic methods	Reported through a specific format/language

- **Data**

Example

~~Meaning~~ ~~What context ?~~ ~~Temperature of what ?~~

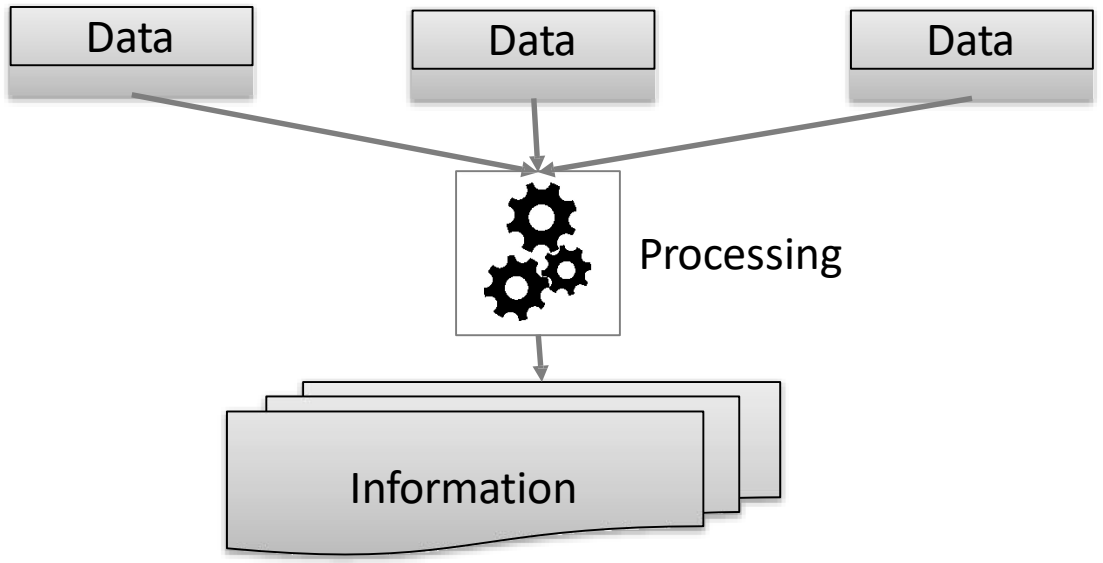
24°

Data

Information

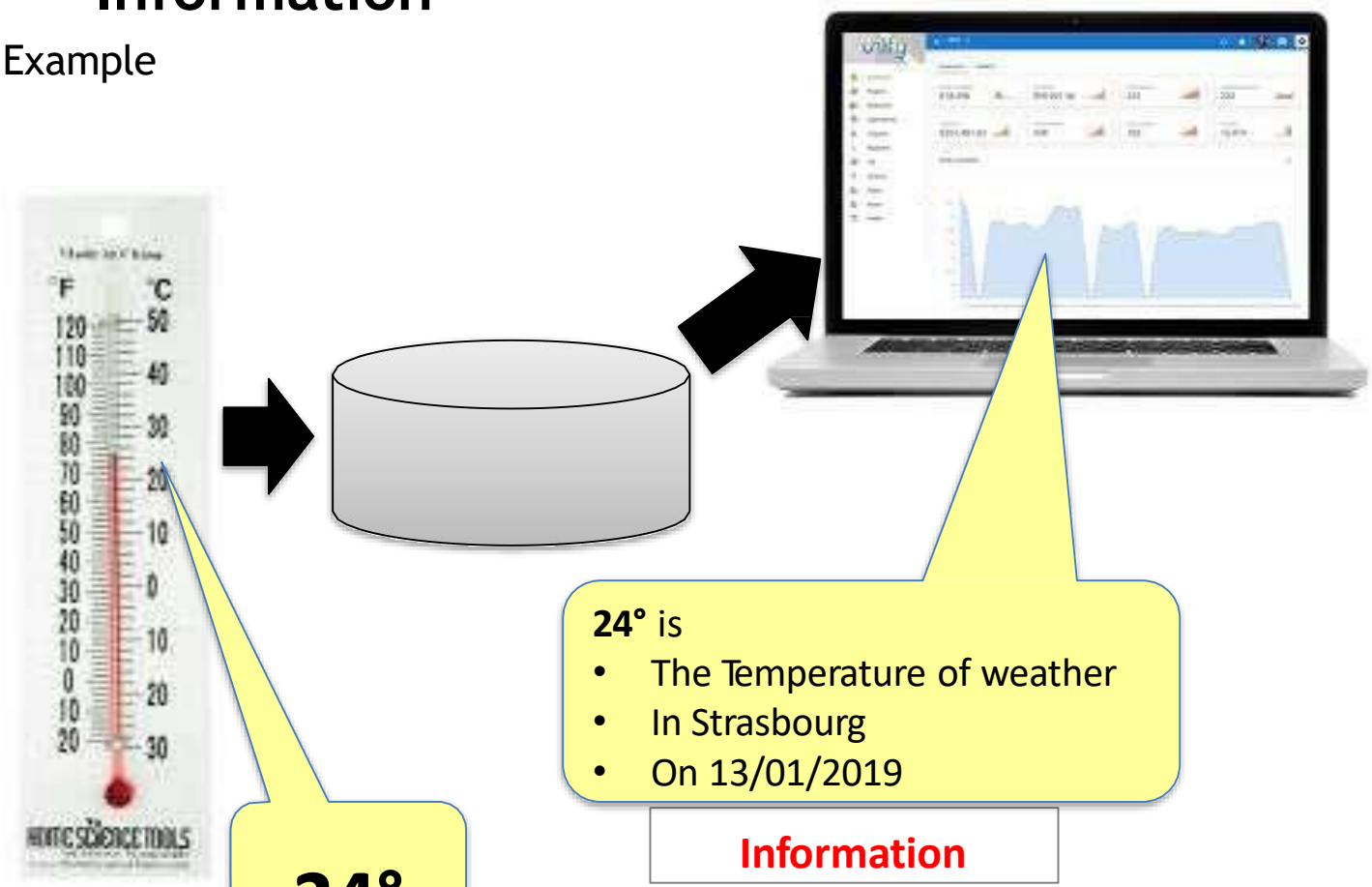
Is the raw fact that was taken, observed, recorded, agreed.

- Information is processed
- Processed Data become information.
- Information is based on Data



Information

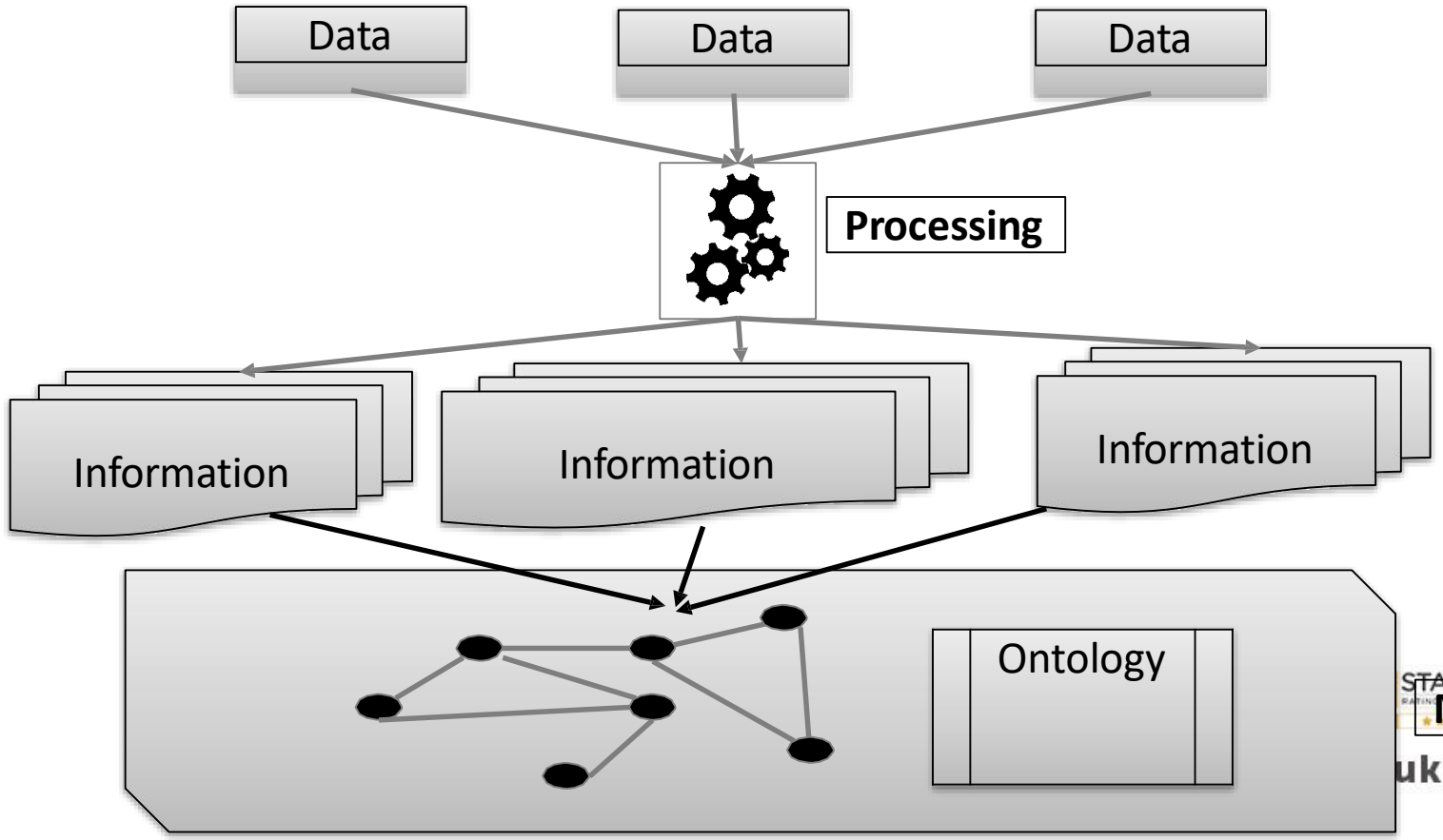
Example



Knowledge

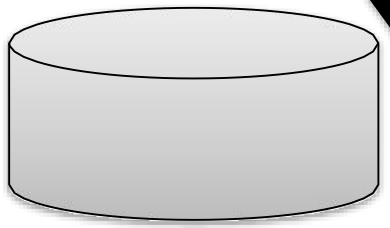
Is the set of relationships between information elements following an ontology

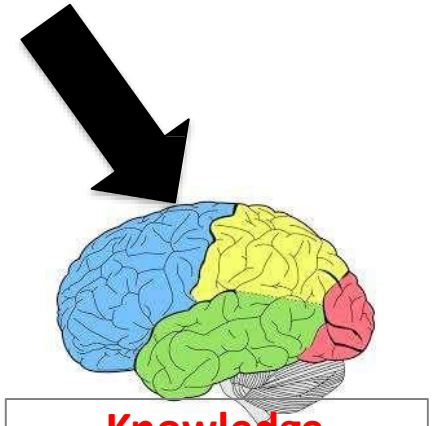
- Mapped information: how they are related, what is compound in what, where ?
- There is ontology (definition of meanings) that frame the set of information



- **Knowledge**

Example





24°

Data

24° is
- The Temperature of weather
- In Strasbourg
- On 13/01/2019

Information

Knowledge

Decision:
Go out for a cruise

What is 'Big Data' ?!

5min to answer…

Big Data → Big Data Concept

Definition

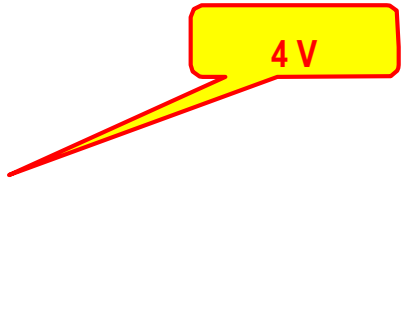
Big Data is the field that gathers all activities and functions of :

- **Acquisition** of Data,
- **Storage** of Data.

from multiple sources that cannot be processed by common and **traditional** systems (for example ERP, Excel, etc.).

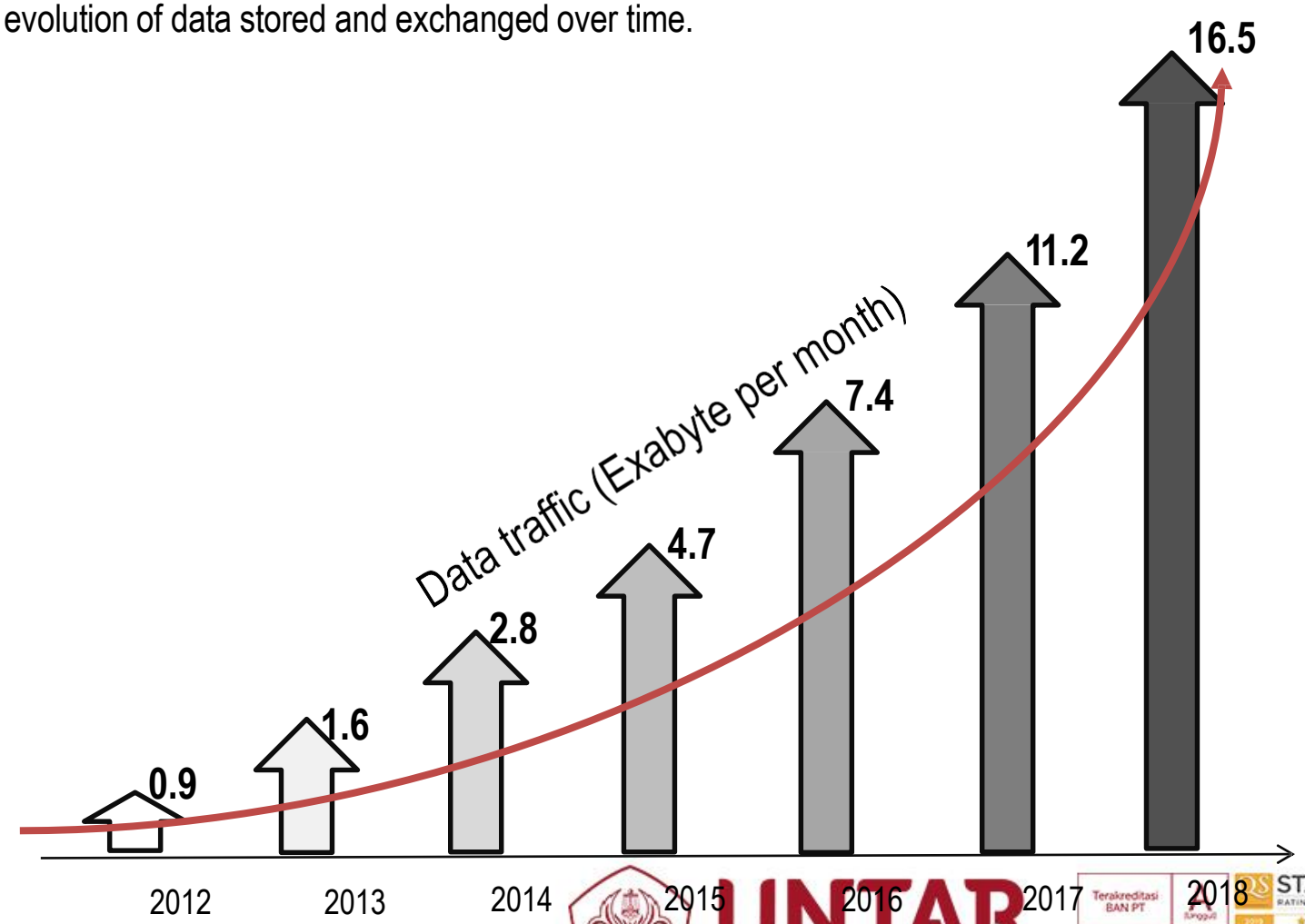
In Big Data, Data are :

- **Huge(Volume),**
- **hetereougenious (Variety),**
- **Dynamic (Velocity),**
- **Uncertain (Veracity).**



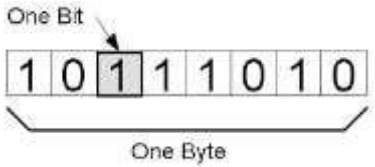
- **Volume**

The evolution of data stored and exchanged over time.



Volume

The evolution of data stored and exchanged over time.



8 bits

X 1024

Byte

X 1024

KB

X 1024

MB

X 1024

GB

Volume

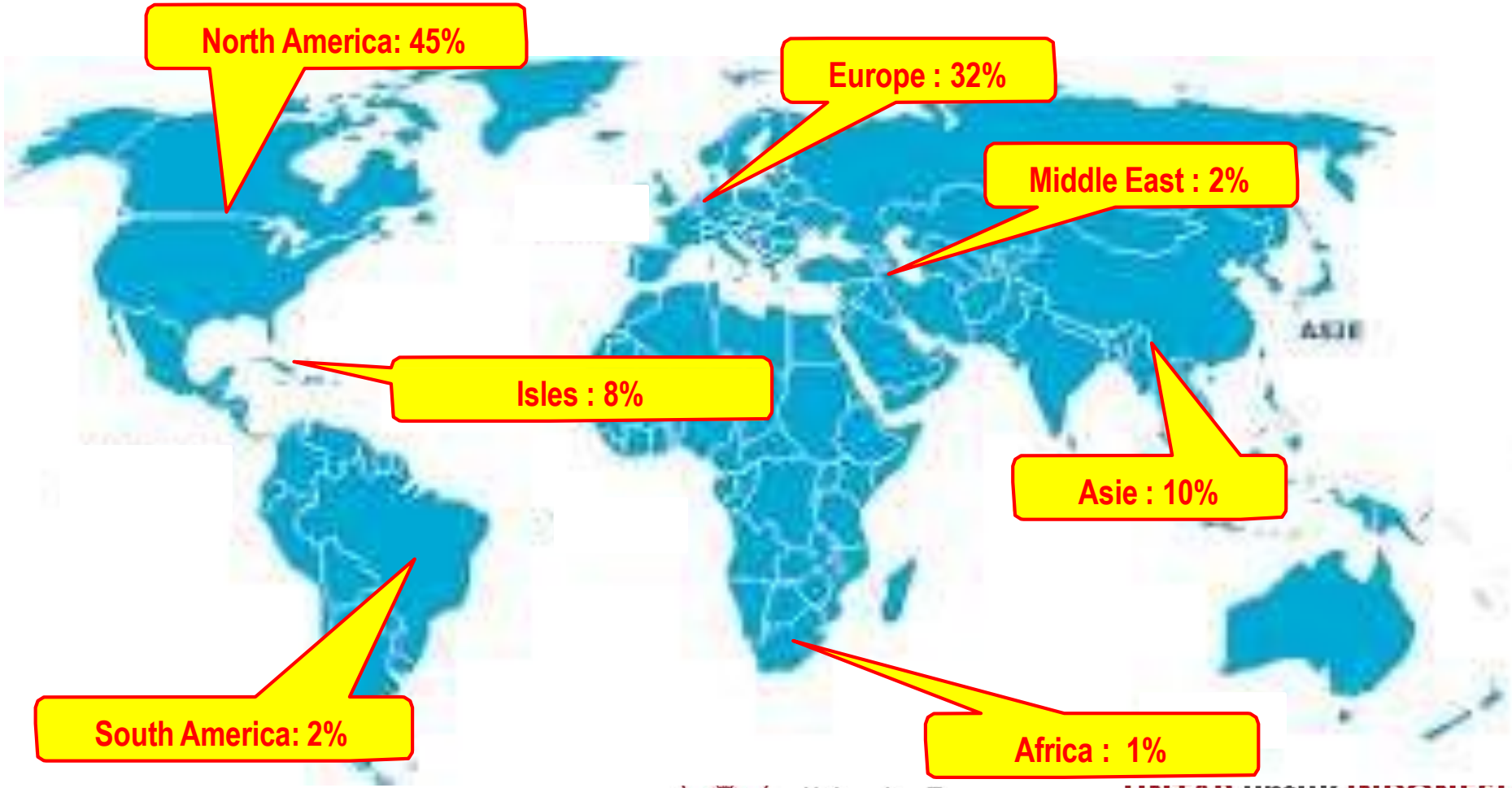
The evolution of data stored and exchanged over time.

- Data Measurement Units

Unit	Abbreviation	Decimal	Binary	Size
bit	b	0 or 1	0 or 1	1/8 of a byte
byte	B	8 bits	8 bits	1 byte
kilobyte	KB	$1,000^1$ bytes	$1,024^1$ bytes	1,000 bytes
megabyte	MB	$1,000^2$ bytes	$1,024^2$ bytes	1,000,000 bytes
gigabyte	GB	$1,000^3$ bytes	$1,024^3$ bytes	1,000,000,000 bytes
terabyte	TB	$1,000^4$ bytes	$1,024^4$ bytes	1,000,000,000,000 bytes
petabyte	PB	$1,000^5$ bytes	$1,024^5$ bytes	1,000,000,000,000,000 bytes
exabyte	EB	$1,000^6$ bytes	$1,024^6$ bytes	1,000,000,000,000,000,000 bytes
zettabyte	ZB	$1,000^7$ bytes	$1,024^7$ bytes	1,000,000,000,000,000,000,000 bytes
yottabyte	YB	$1,000^8$ bytes	$1,024^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

- **Volume**

Distribution of Data centers around the world.



- **Volume**

Multiple Data centers around the world.

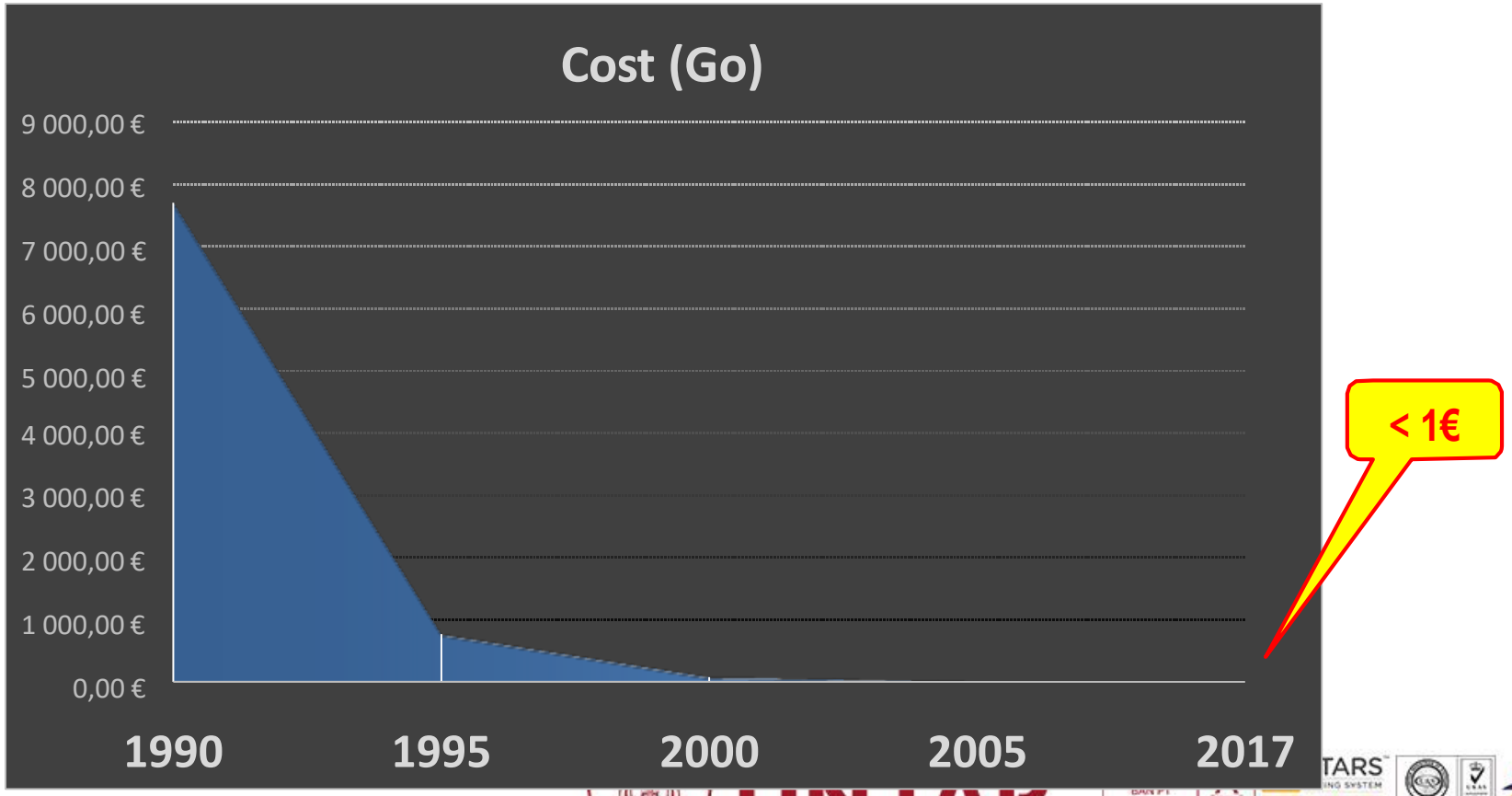
Data Center Utah (USA)

Data Center Vitry (France)

Data Center Busan (South Korea)

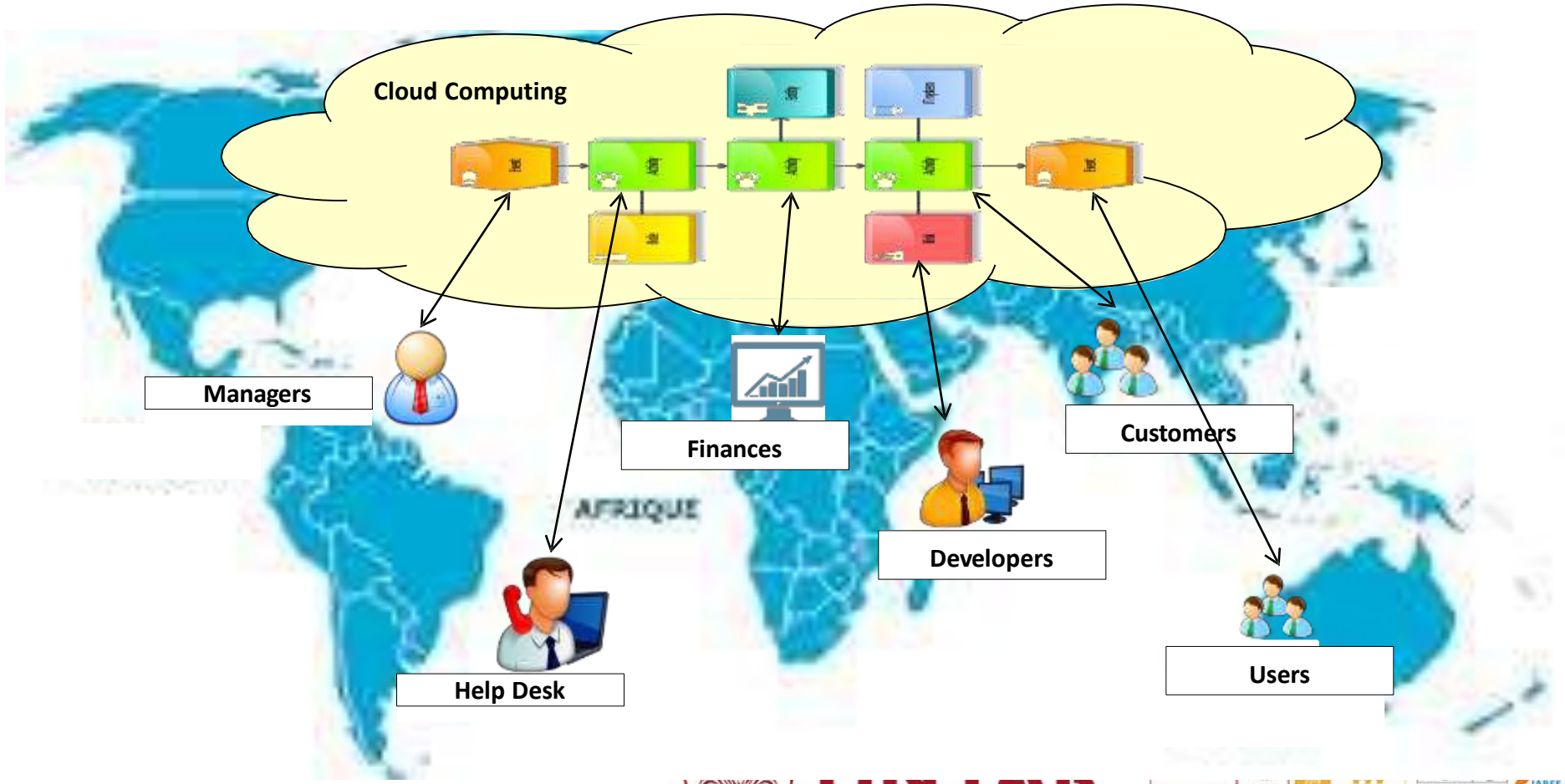
- **Volume**

Storage price decreasing.



- **Volume**

Virtualization.

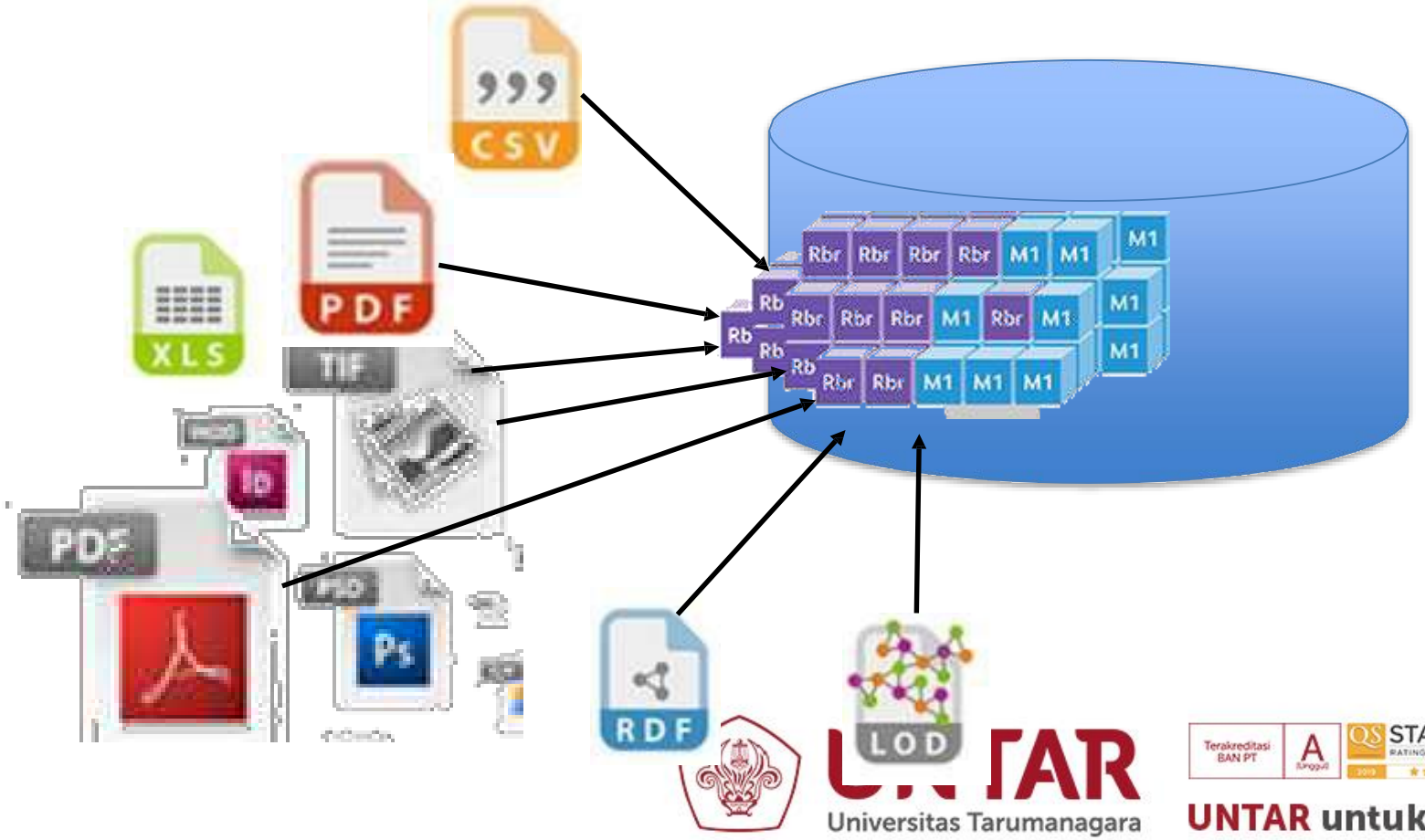


- **Volume**

Internet Of Things.

Variety

No standard format of data in storages. Meta-Data of structures and organizations, semantic Data, images, videos, texts, XML, text formats, etc.

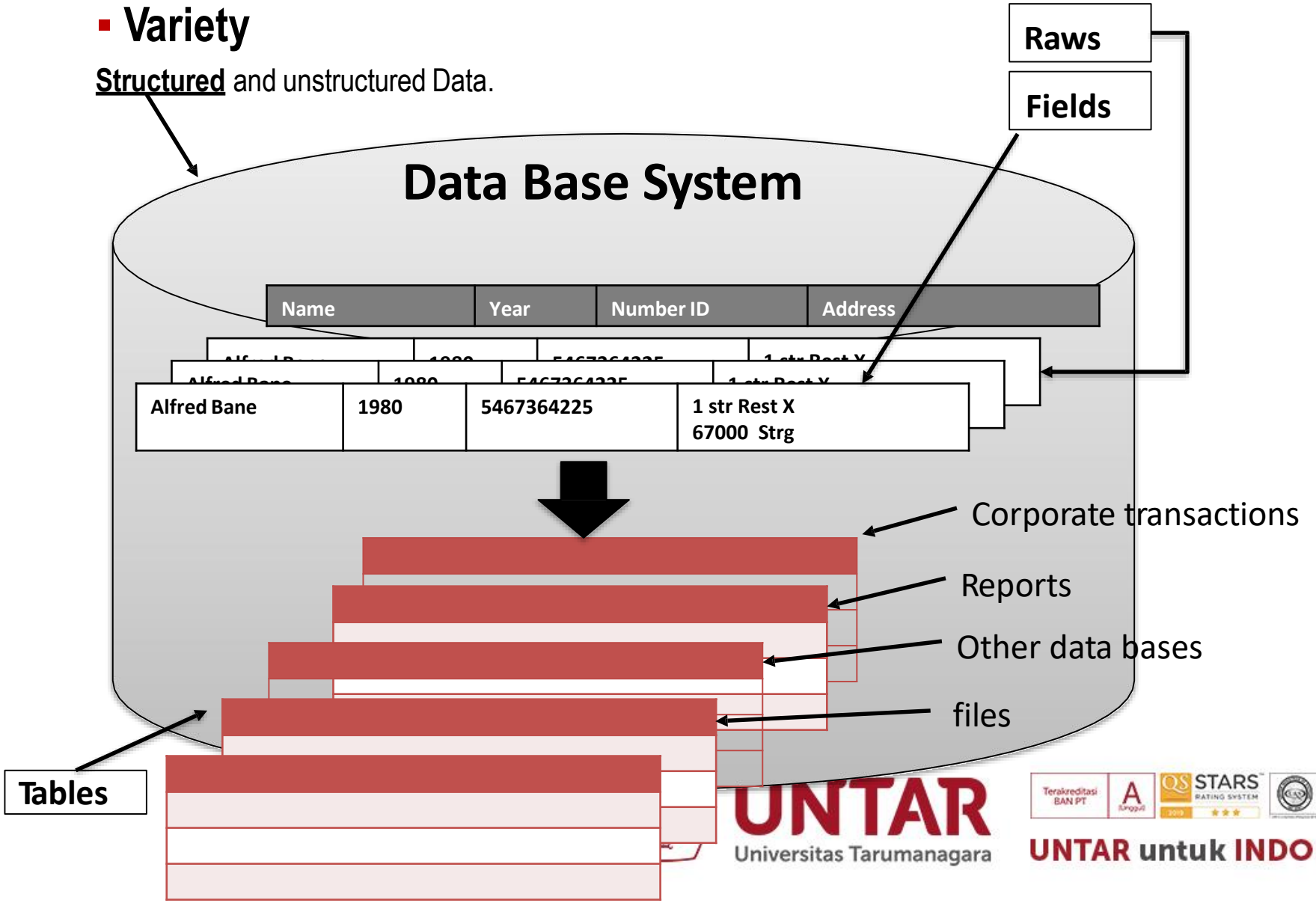


Variety

Different sources.

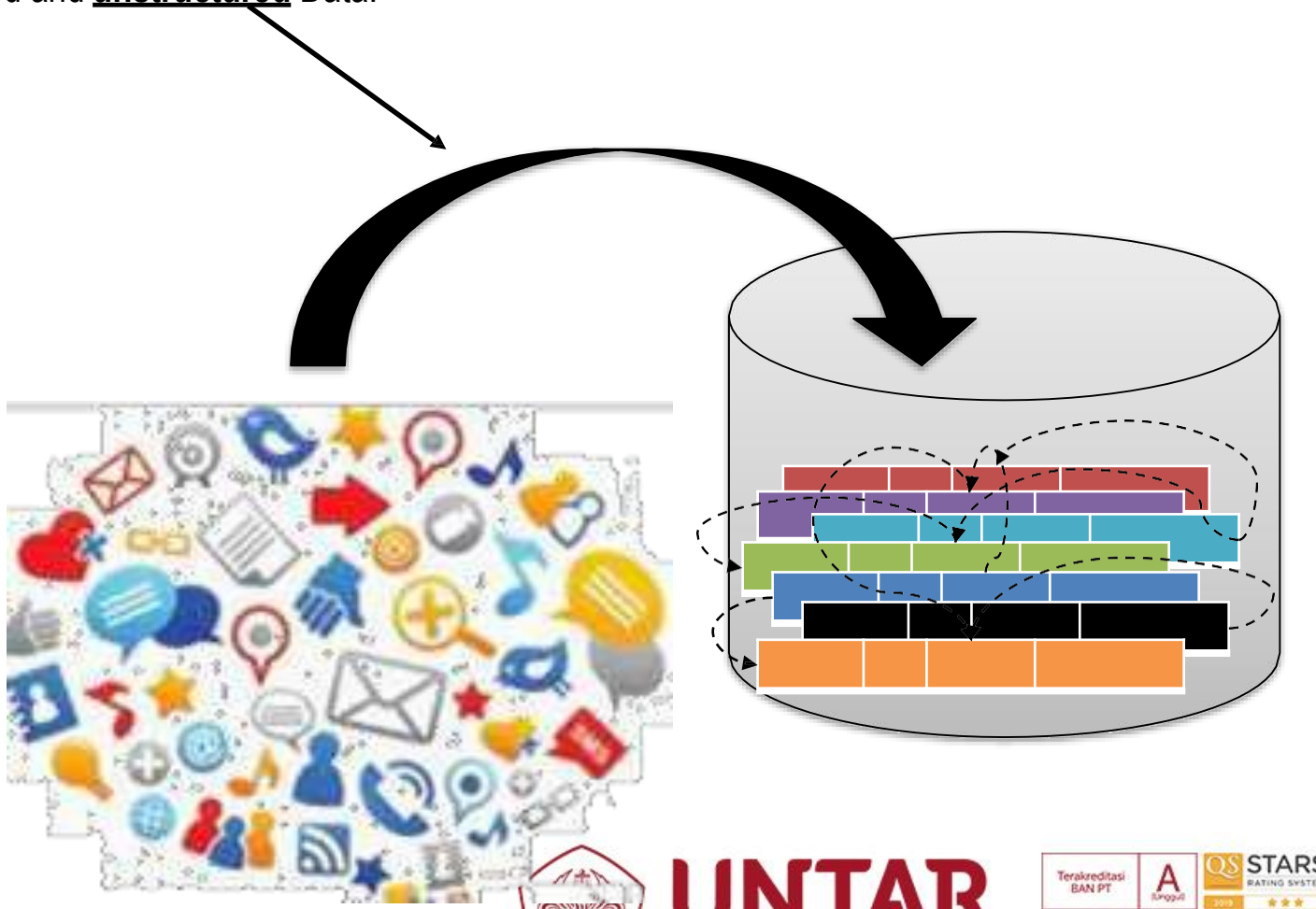
Variety

Structured and unstructured Data.



- **Variety**

Structured and **unstructured** Data.



Understanding Big Data

A. Concepts and Terminology

- Datasets
- Data Analysis
- Data Analytics

B.Big Data Characteristics

Figure 1.1 Datasets can be found in many different formats.

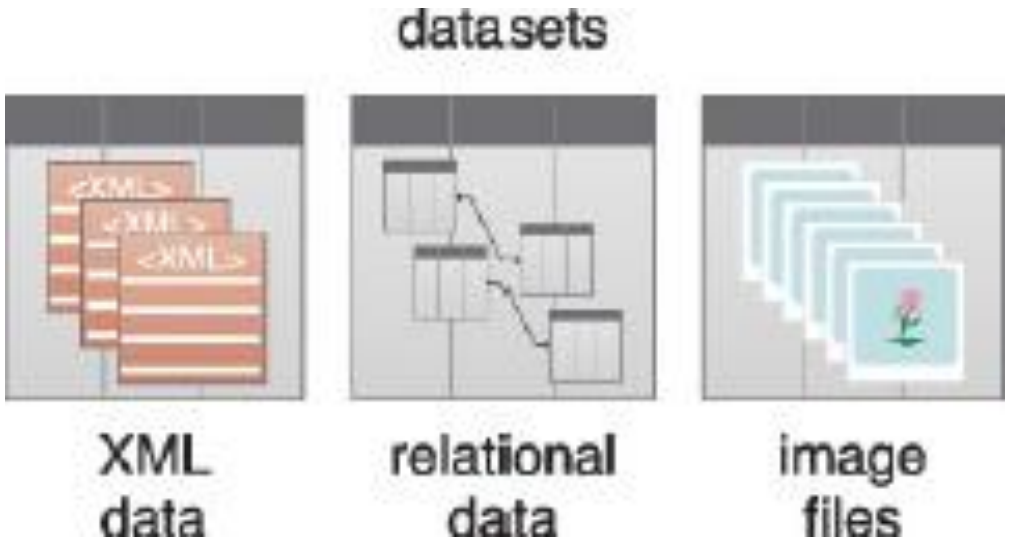


Figure 1.2 The symbol used to represent data analysis.

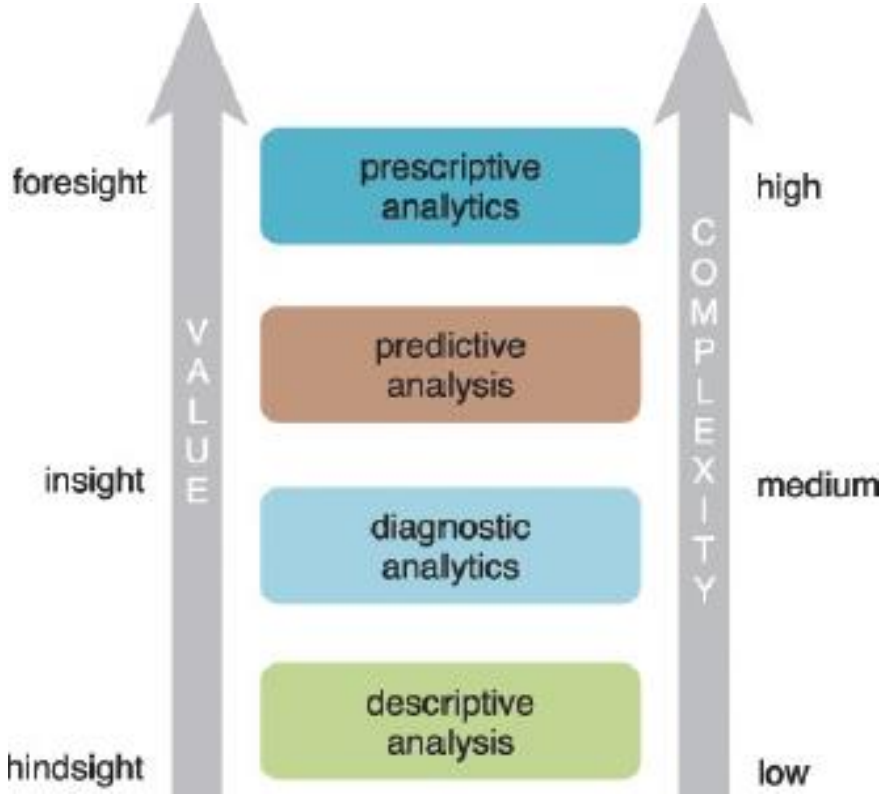
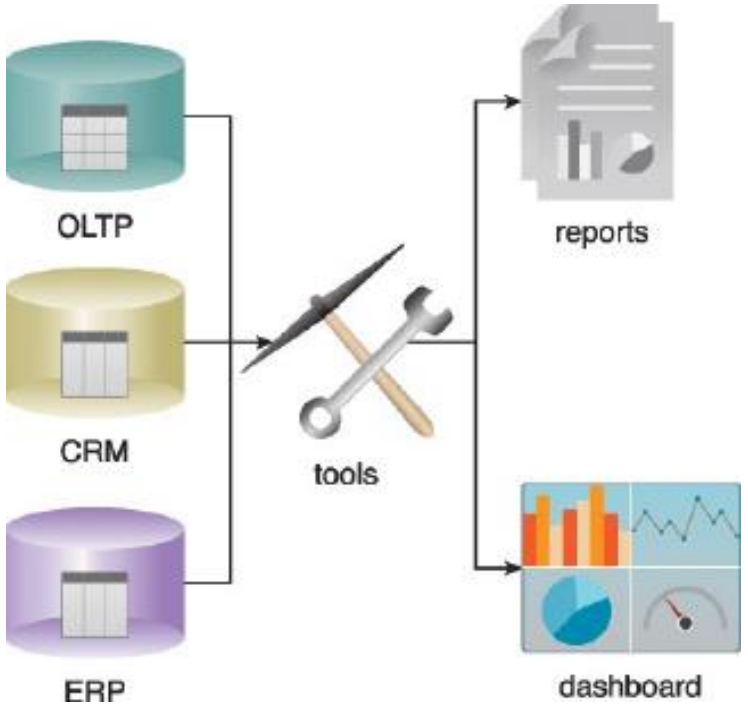
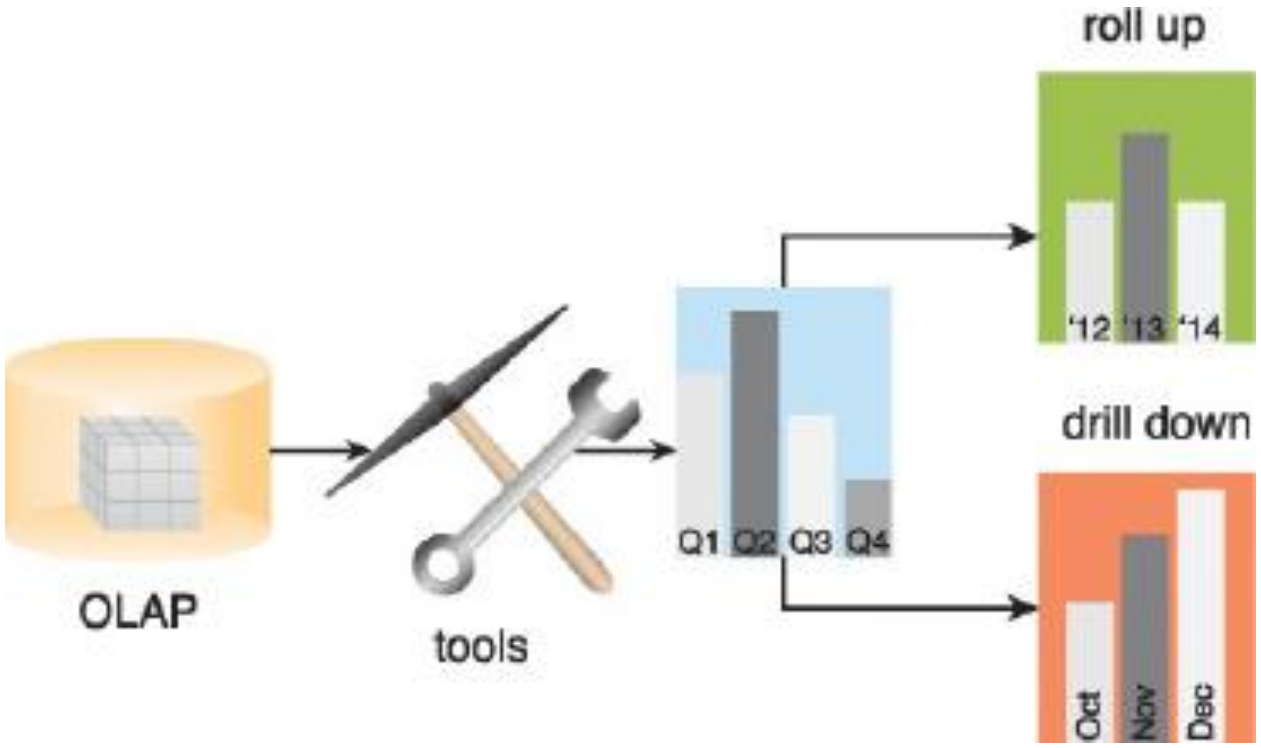


Figure 1.4 Value and complexity increase from descriptive to prescriptive analytics.

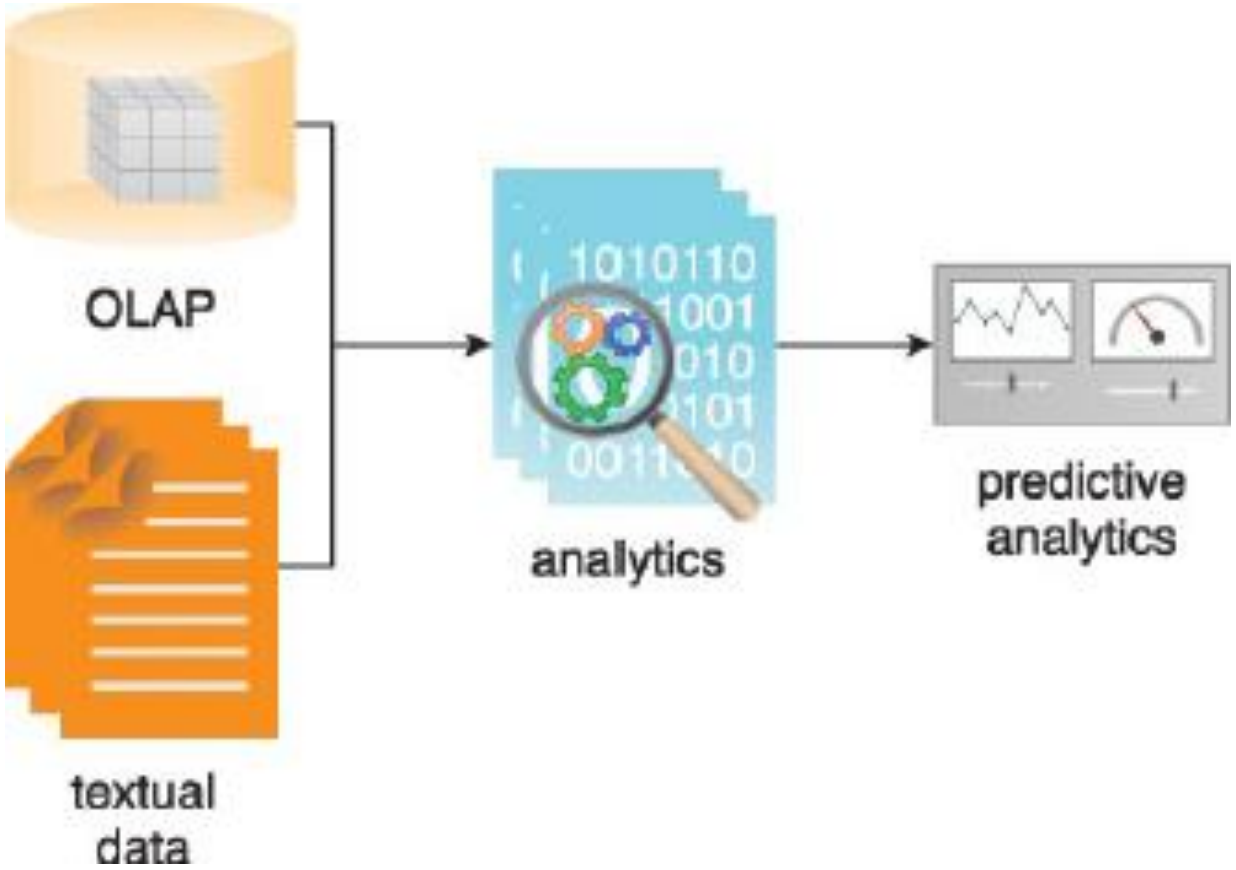
The operational systems, pictured left, are queried via descriptive analytics tools to generate reports or dashboards, pictured right.

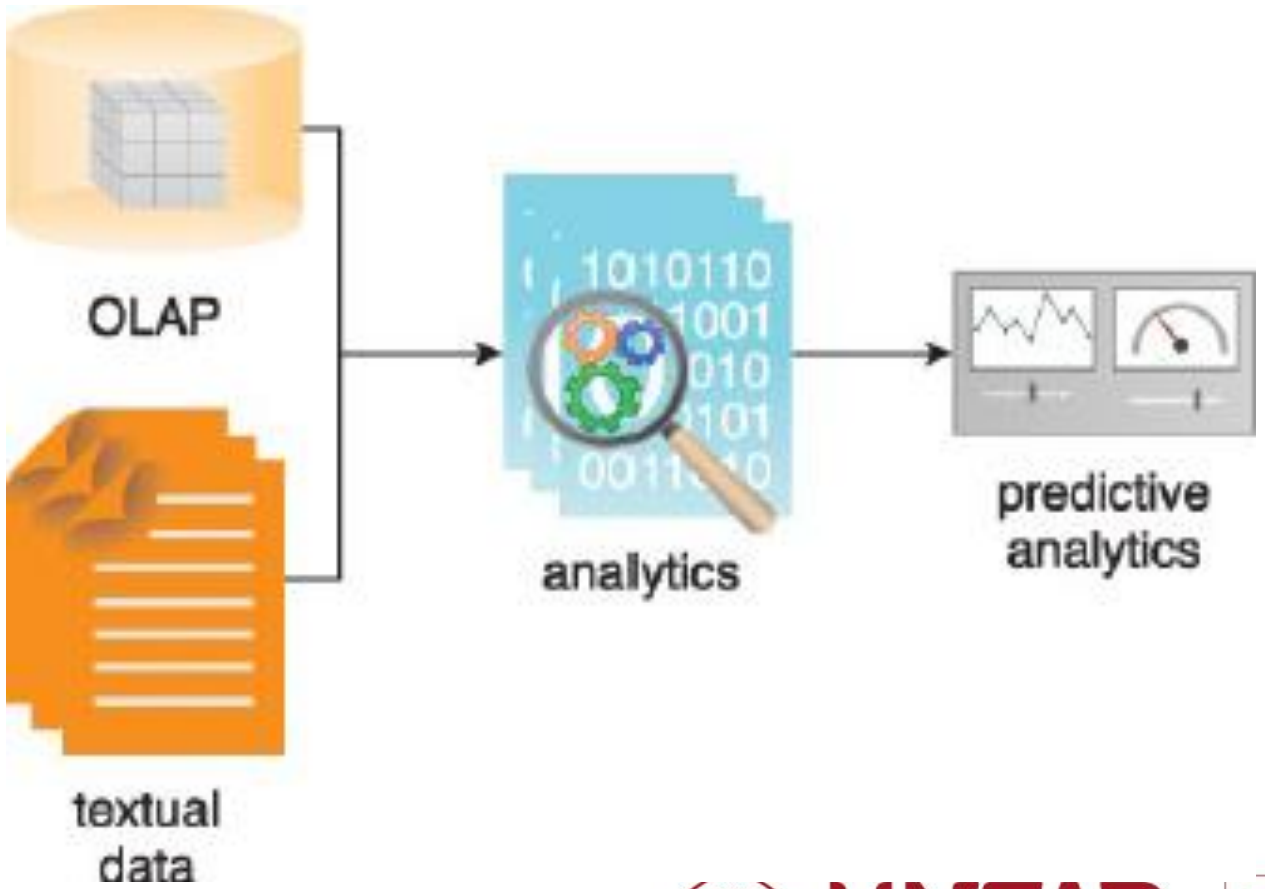


Diagnostic analytics can result in data that is suitable for performing drill-down and roll-up analysis.

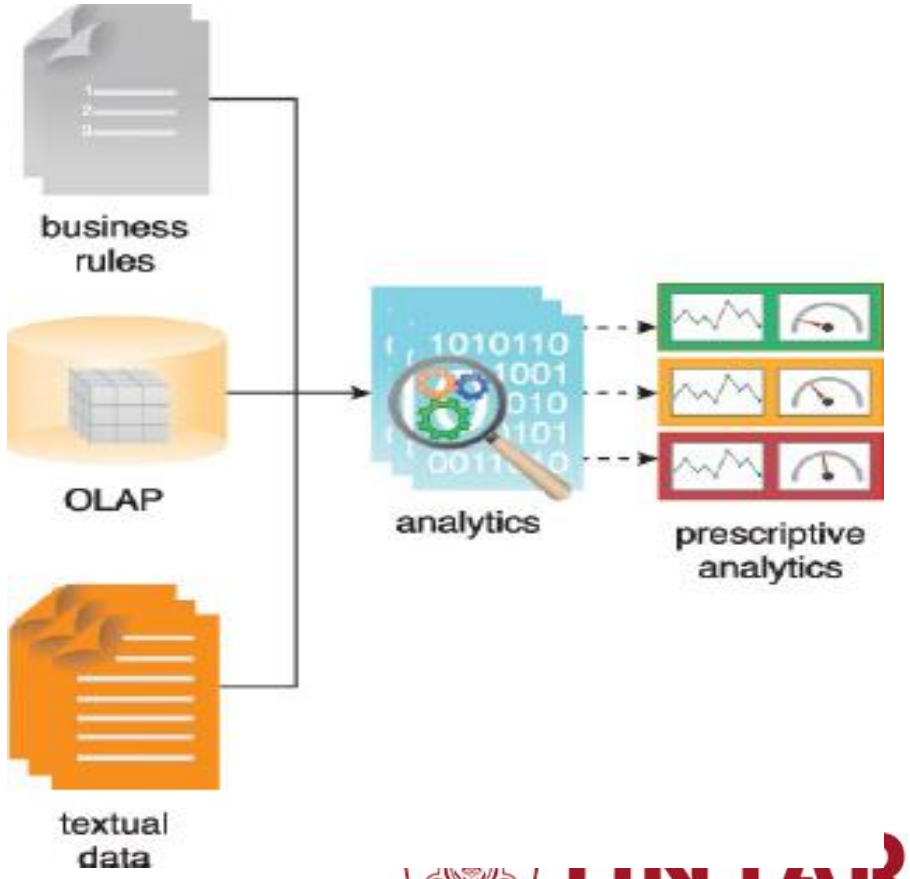


Predictive analytics tools can provide user-friendly front-end interfaces.

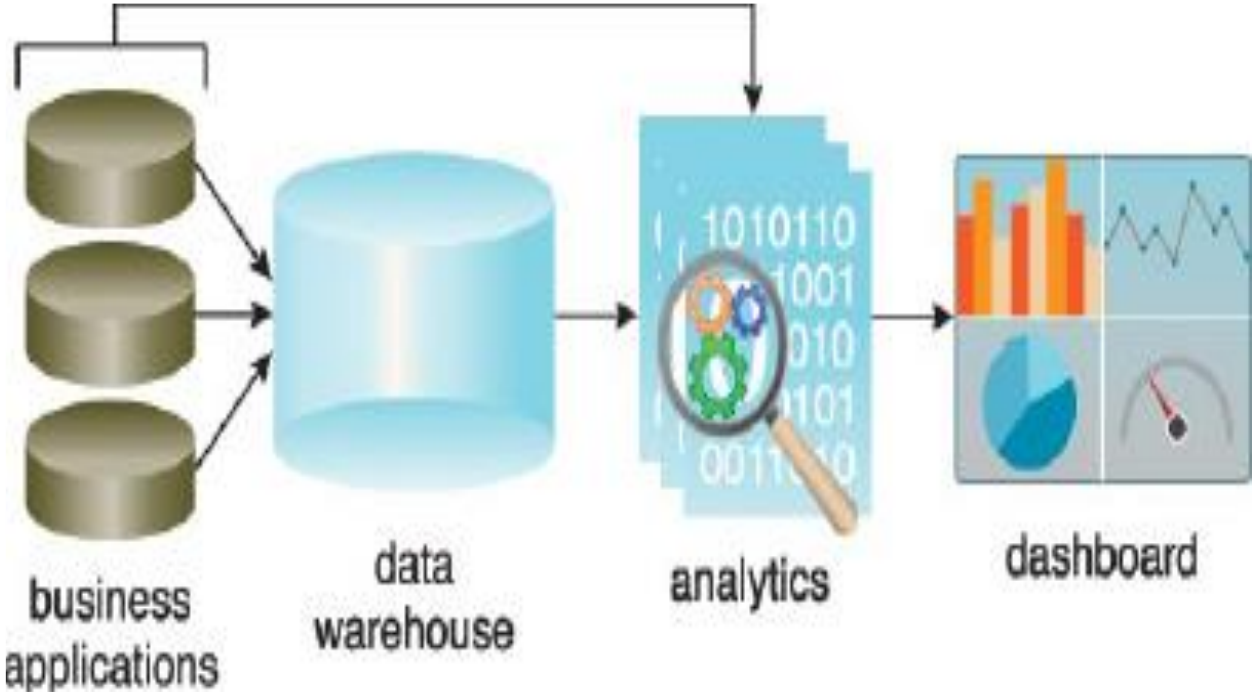



OLAP
textual data
analytics
predictive analytics

Prescriptive analytics involves the use of business rules and internal and/or external data to perform an in-depth analysis.



BI can be used to improve business applications, consolidate data in data warehouses and analyze queries via a dashboard.

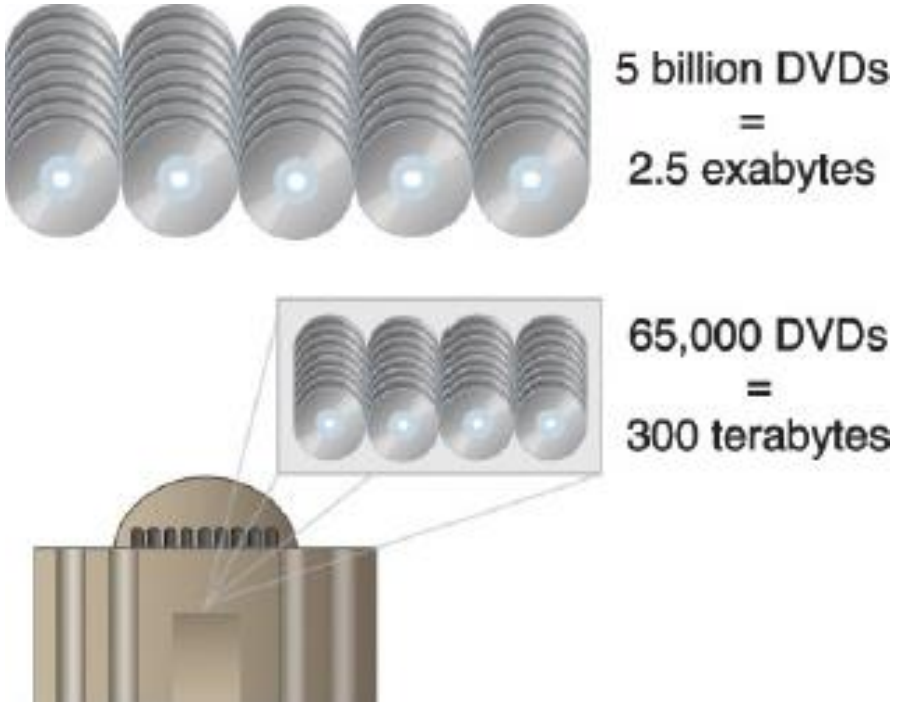


A KPI dashboard acts as a central reference point for gauging business performance.

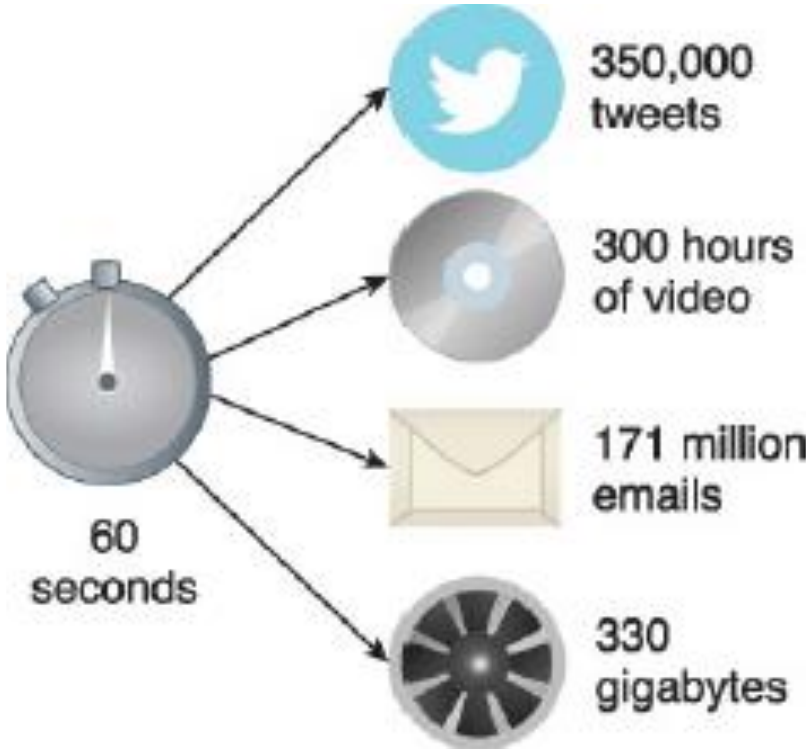
KPI dashboard

Big Data Characteristics

Volume



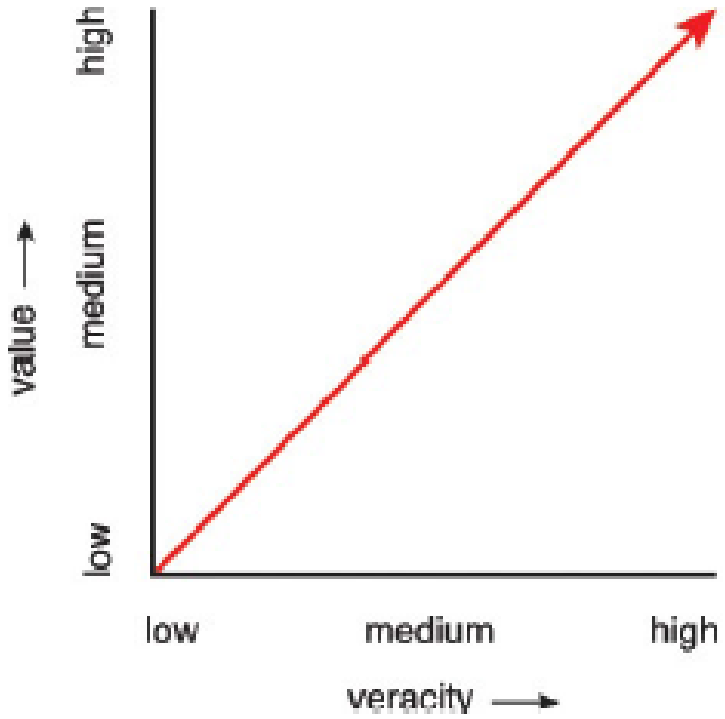
Velocity

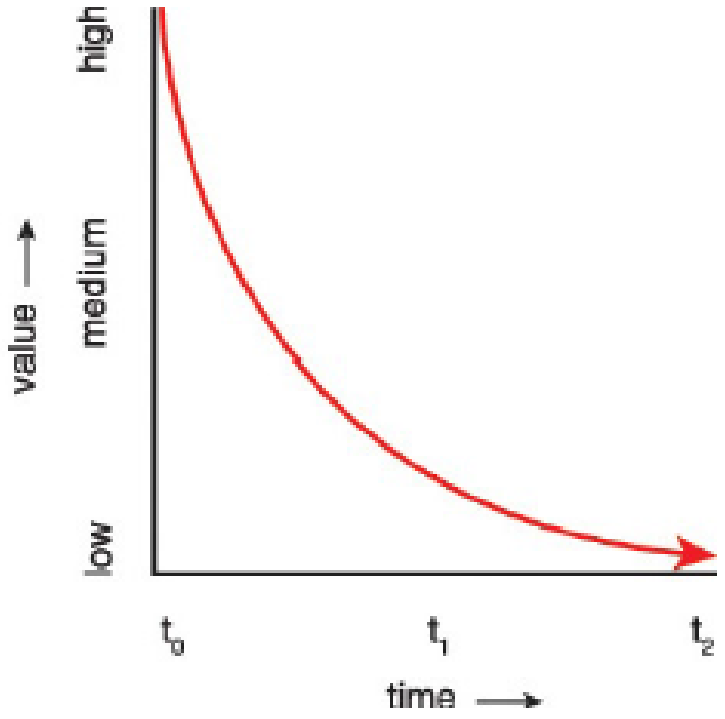


Variety

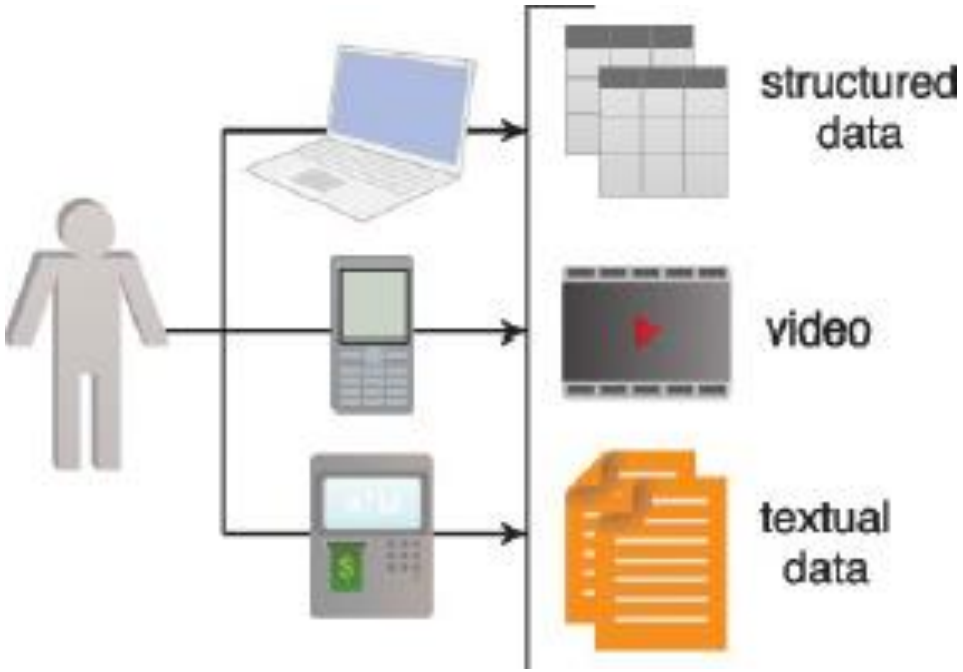

structured data
texual data
image files
video
audio
<XML>
JSON
XML data
JSON data
sensor data
metadata

Value





Different Types of Data



Thank You

Reference: Database Systems A Practical Approach to Design, Implementation, and Management Fourth Edition.

Thomas M. Connolly and Carolyn E. Begg

UNTAR
Universitas Tarumanagara

UNTAR untuk INDONESIA