

BACS3183  
Advanced Database Management

Chapter 1  
Information Models

# BACS3183

## Assessment Criteria

- Examination Component 30%
- Course Work Assessment 70%
  - Test 40%
  - Assignment 60%

# Chapters

**Chapter 1: Information Models**

**Chapter 2: Database Systems**

**Chapter 3: Data Modelling**

**Chapter 4: Relational algebra**

**Chapter 5: Relational Database Design**

**Chapter 6: Indexing**

**Chapter 7: Physical Database Design**

**Chapter 8: Database Security and Transaction Management**

**Chapter 9: Distributed Database**

# Introduction

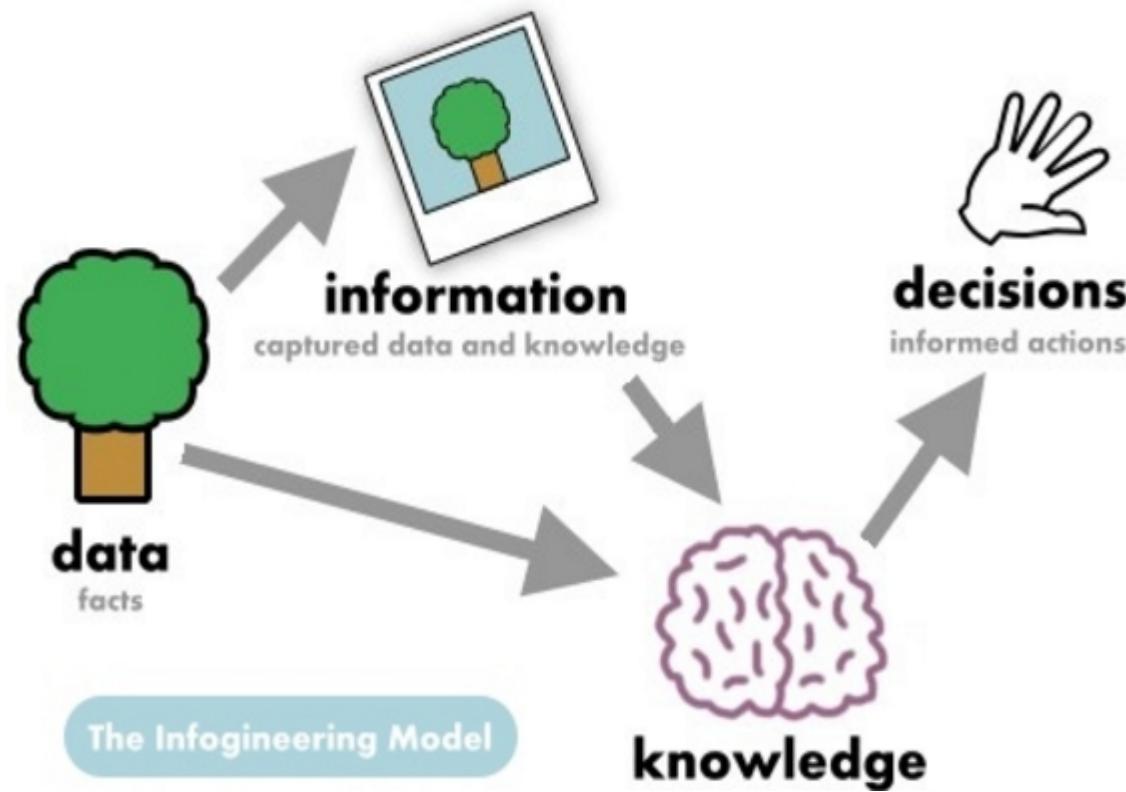
**Information Management plays a critical role in almost all areas where computers are used. IM includes the**

- capture, digitization, representation, organization, transformation, and presentation of information
- algorithms for efficient and effective access and updating of stored information [ ch 6 & 7 & practical ]
- data modelling and abstraction [ ch 3 & 4 & 5 ]
- physical file storage techniques [ ch 7 ]
- information security, privacy, integrity, and protection in a shared environment. [ ch 8 & 9 ]

# Introduction

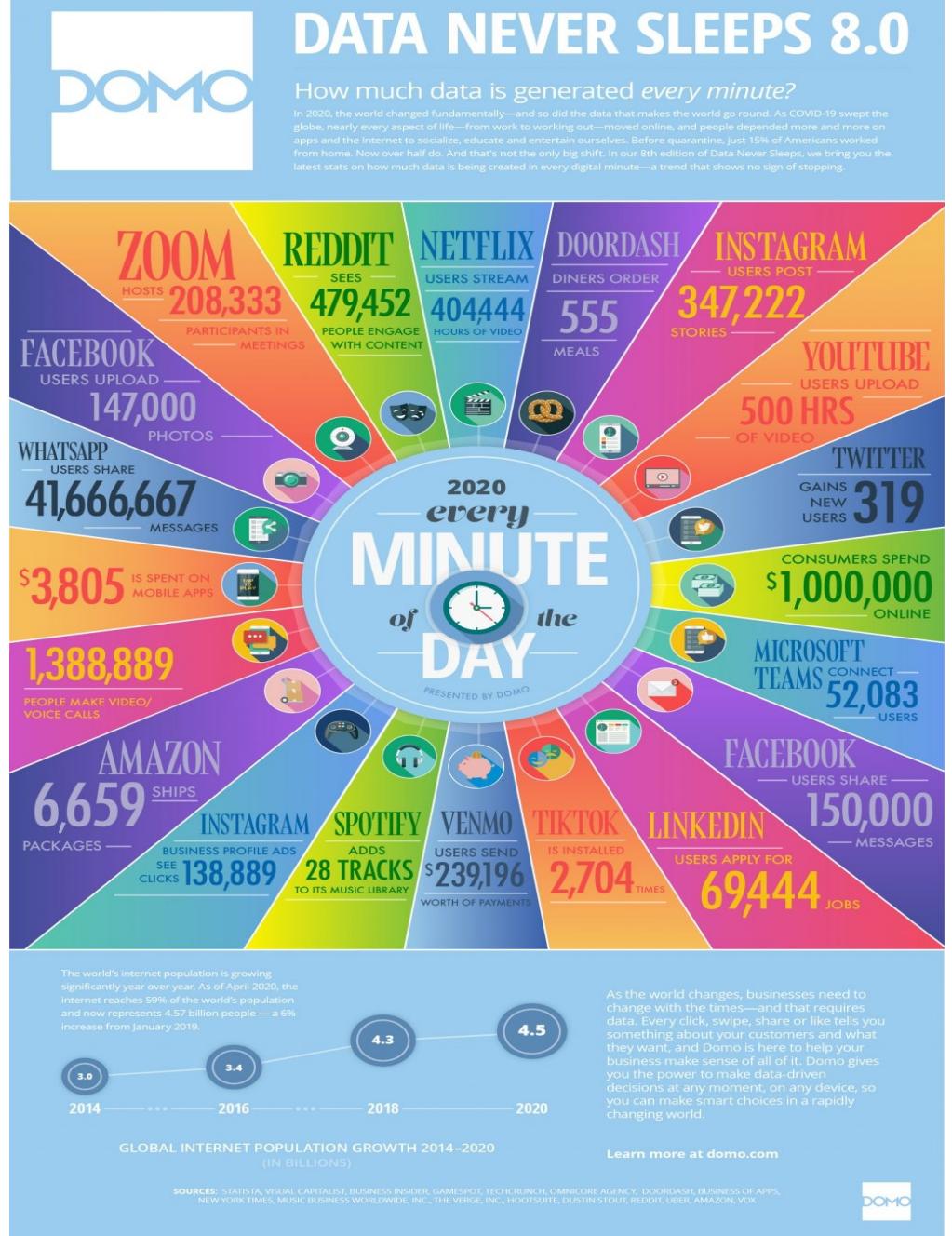
The goal of information systems is to transform **data** into **information** in order to generate **knowledge** that can be used for **decision making**.

ie the system must be able to take data, put the data into context, and provide tools for aggregation and analysis.



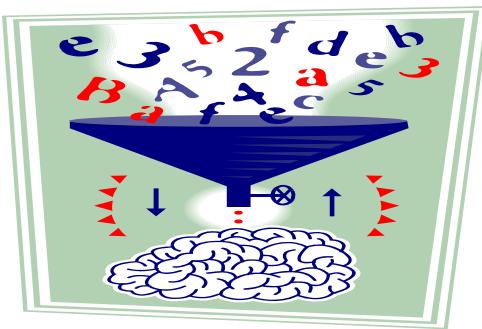
# Lots of Data

# How much of this **data** is useful **information??**

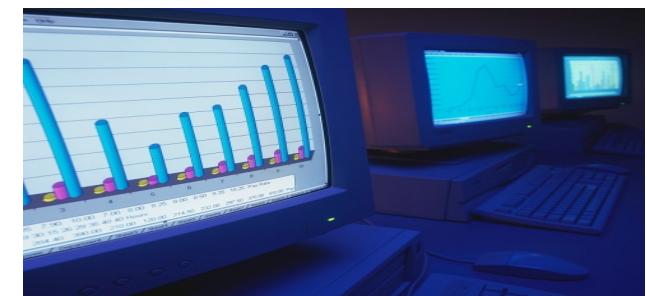


# Data, Information & knowledge

**Data-** Raw, non-summarized , and unanalyzed facts and figures



**Information-** Data that have been converted into a meaningful and useful context for the receiver



We tend to gain **knowledge** from information and we use that information to **make decisions**.

# Data, Information & knowledge

**Knowledge:** Provide answers to “how” questions

**Information :**  
Provide answers to “who”, “what”, “where” and “when” questions

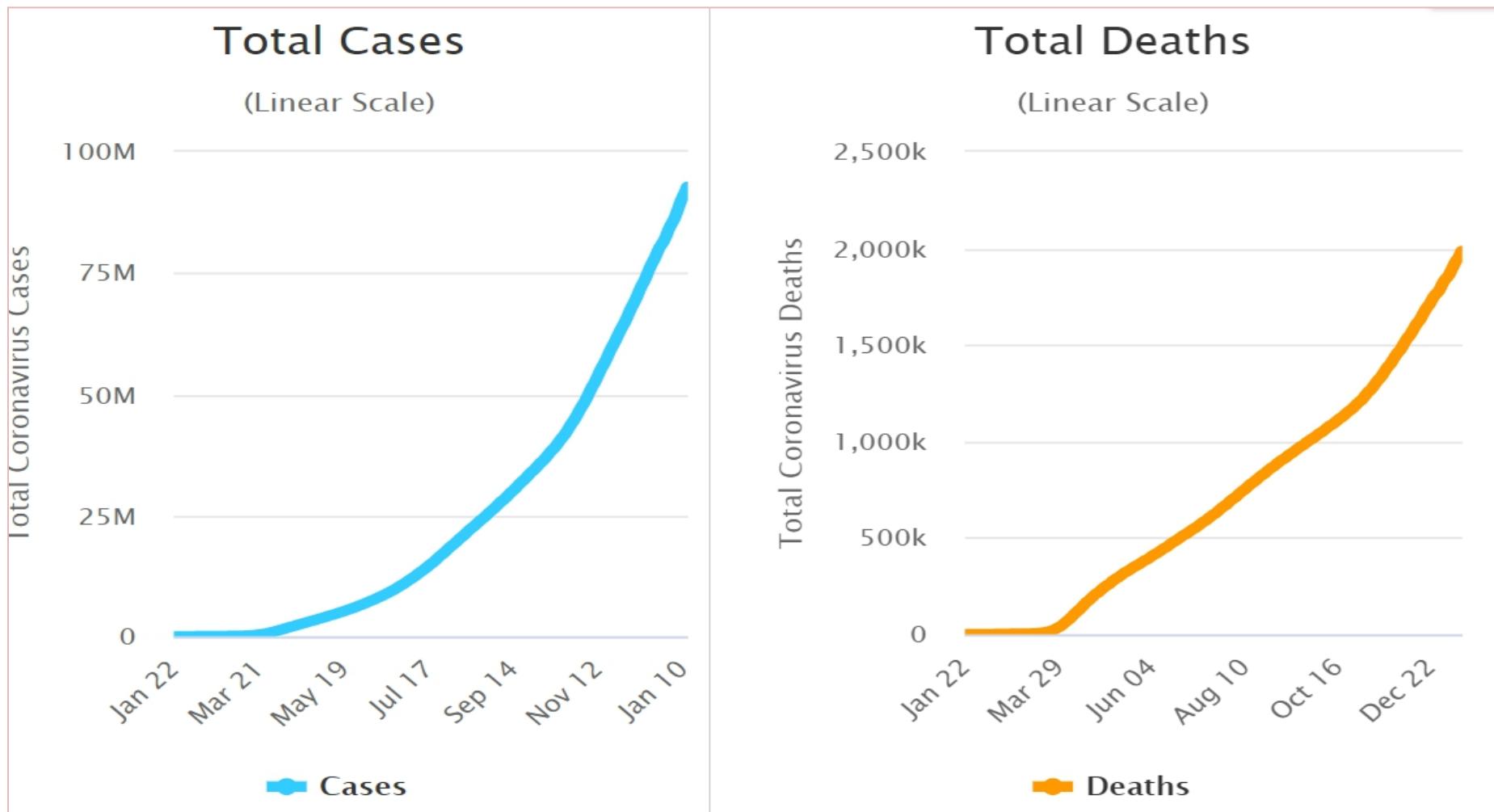


# Data or Information?

#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop
60	<a href="#">Ireland</a>	159,144		2,400		25,504	133,520	170	32,050	495	2,005,124	550,252
61	<a href="#">Kuwait</a>	155,874		946		149,809	5,119	48	36,214	220	1,359,421	315,830
62	<a href="#">Egypt</a>	152,719		8,362		120,312	24,045	90	1,478	81	1,000,000	9,675
63	<a href="#">Moldova</a>	150,946		3,193		140,217	7,536	214	37,466	793	584,518	145,081
64	<a href="#">Palestine</a>	149,769		1,658		134,977	13,134	93	29,001	321	944,333	182,857
65	<a href="#">Malaysia</a>	147,855	+3,337	578	+15	113,288	33,989	195	4,537	18	3,735,103	114,617
66	<a href="#">Greece</a>	146,688		5,354		9,989	131,345	340	14,111	515	3,689,863	354,947
67	<a href="#">Qatar</a>	146,480		246		143,094	3,140	28	52,169	88	1,301,053	463,370
68	<a href="#">Guatemala</a>	145,986		5,117		132,507	8,362	5	8,069	283	684,552	37,837
69	<a href="#">Slovenia</a>	145,472	+1,767	3,093	+23	119,149	23,230	199	69,969	1,488	735,437	353,730
70	<a href="#">Myanmar</a>	132,260		2,902		115,621	13,737		2,422	53	2,055,757	37,647
71	<a href="#">Oman</a>	131,264	+178	1,509	+1	123,593	6,162	21	25,363	292	571,472	110,421
72	<a href="#">Honduras</a>	131,009	+1,204	3,320	+26	59,647	68,042	220	13,116	332	336,727	33,711
73	<a href="#">Ethiopia</a>	129,455		2,006		114,567	12,882	210	1,111	17	1,864,567	16,009

<https://www.worldometers.info/coronavirus/#countries>

# Data or Information?

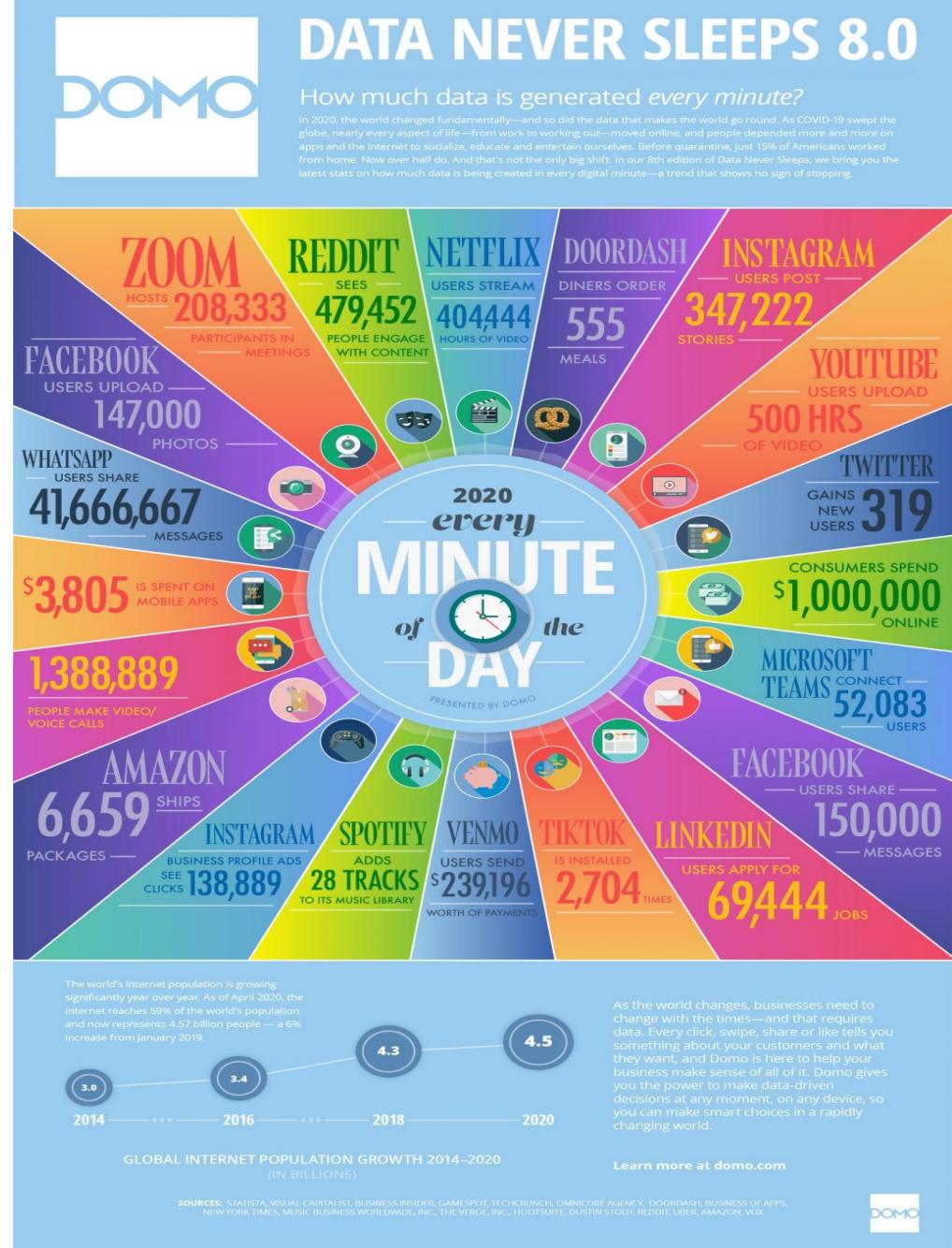


<https://www.worldometers.info/coronavirus/#countries>

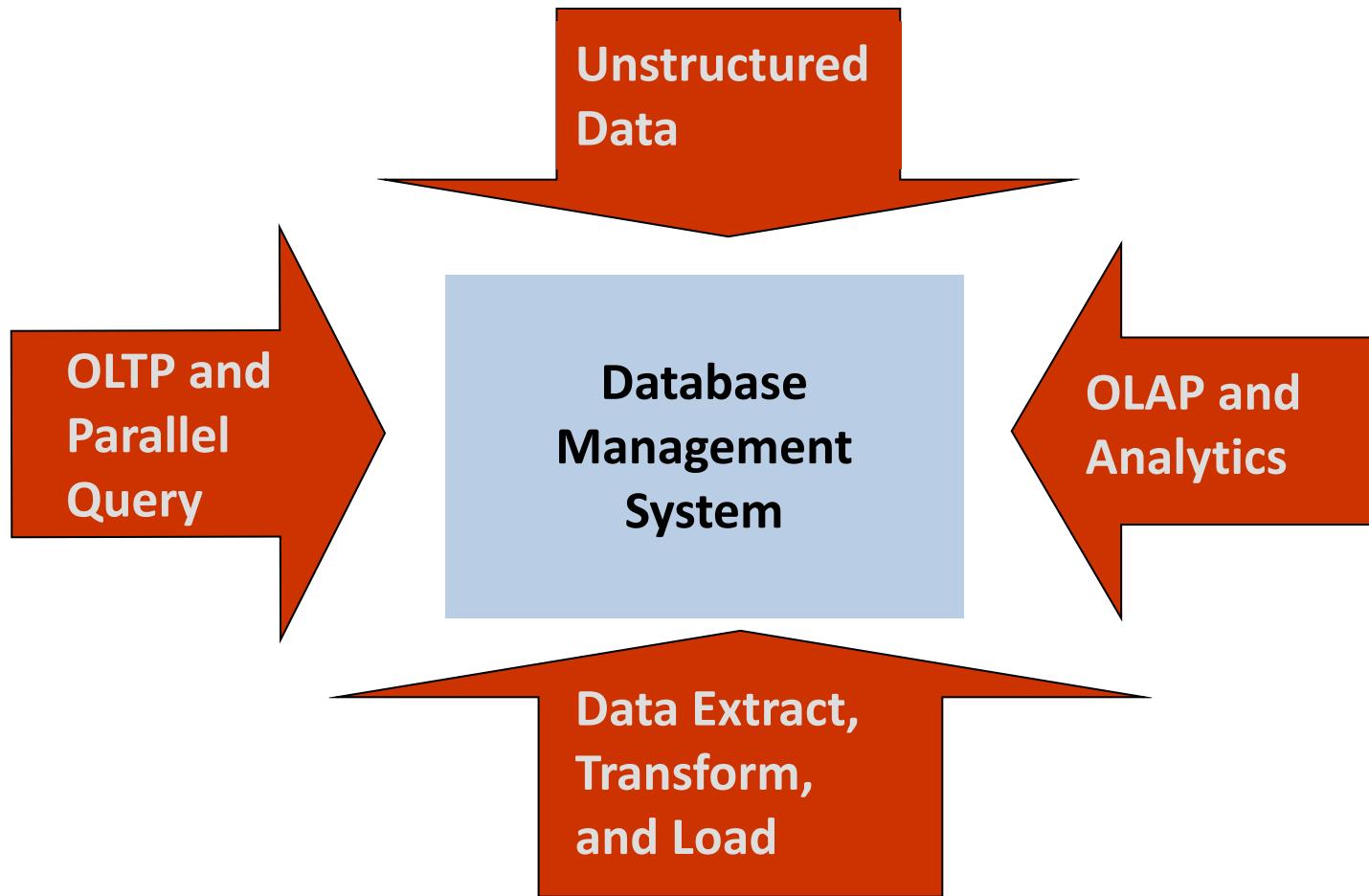
# Lots of Data

By 2025, it's estimated humanity will have produced a total of 175 zettabytes of data. For context, that's 175,000,000,000 terabytes. (1 TB =  $2^{10}$  GB)

<https://chartio.com/learn/business-intelligence/how-to-use-data-warehouses-in-business-intelligence/>



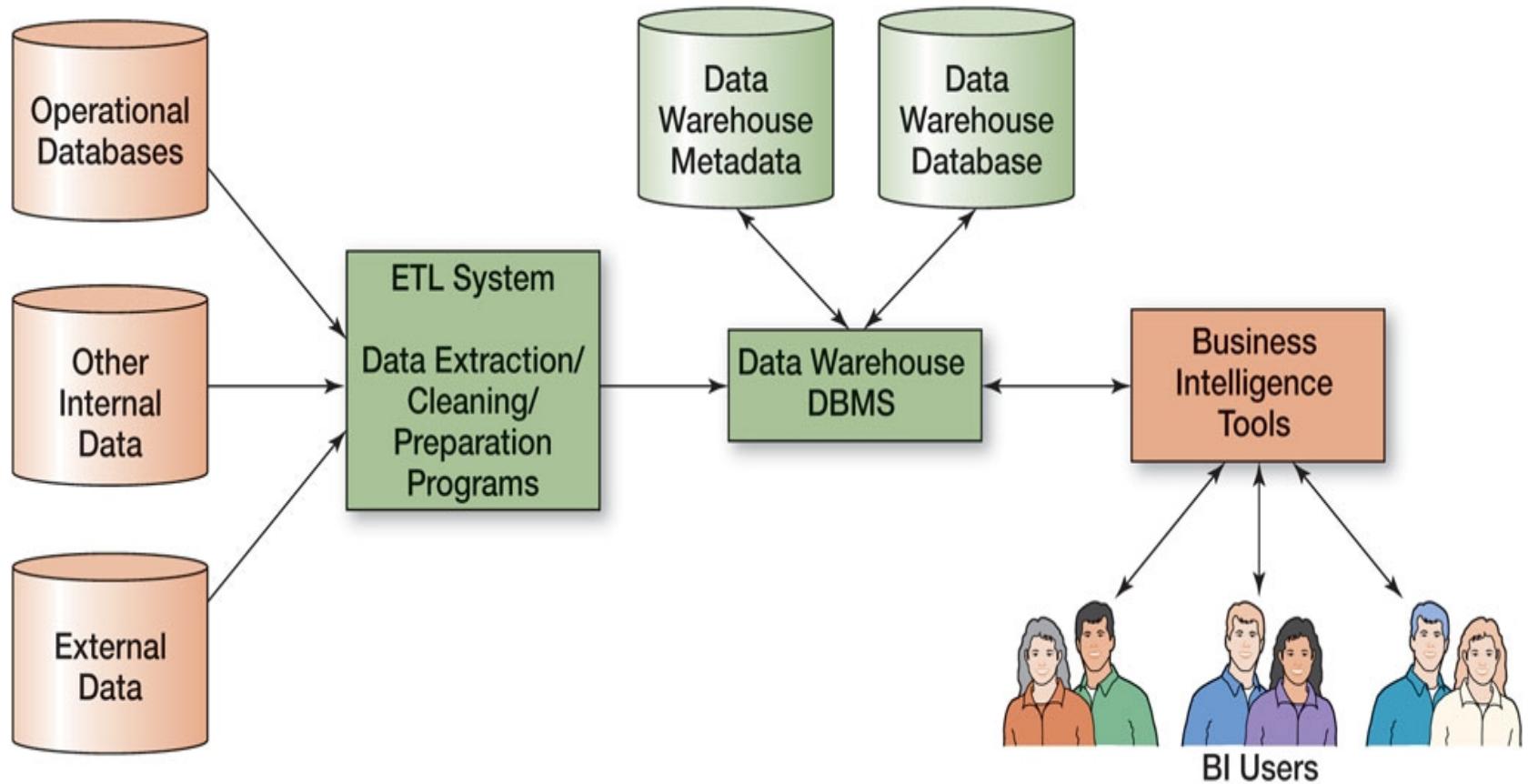
# Information Systems & DBMS



**Online transaction processing (OLTP)** captures, stores, and processes data from transactions in real time.

**Online analytical processing (OLAP)** uses complex queries to analyze aggregated historical data.

# Data Warehouse & Business Intelligence



A **data warehouse** is a database designed to enable **business intelligence** activities to help users make faster, more informed business decisions.

ETL systems take large volumes of raw data from multiple sources, converts it for analysis, and loads that data into the data warehouse.

# Categorization of information

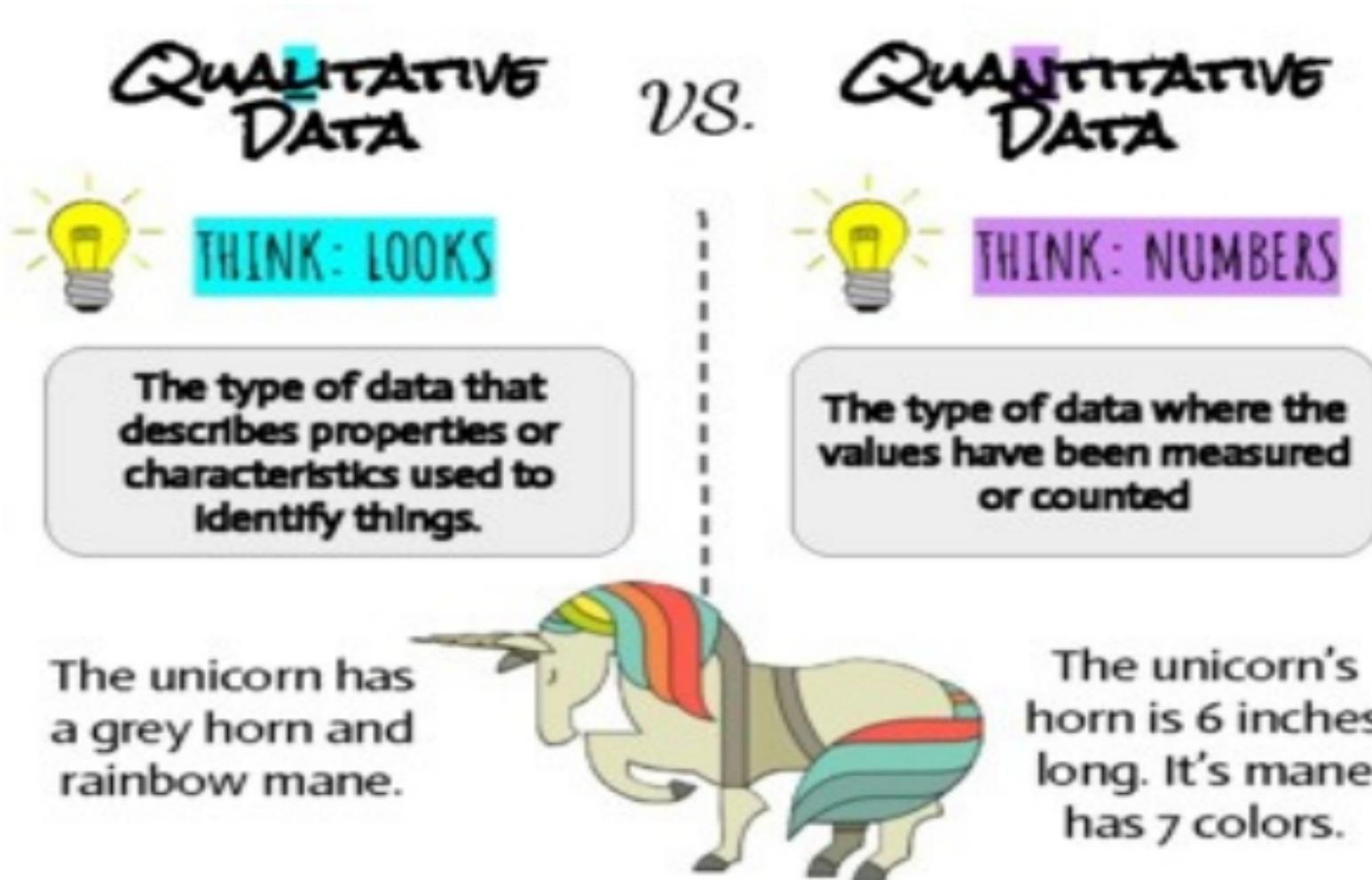
- Source
- Nature
- Level
- Time

# Categorization of information

- **Source**
  - **Primary information**
    - original source document
  - **Secondary information**
    - processed primary sources, second-hand versions
  - **Internal information**
    - Internal information could come from variety of source **within the company** such as different departments.
  - **External information**
    - External information can be gathered **outside the company**, either by interviewing customers or examining published data.

# Categorization of information

- Nature - Qualitative or Quantitative



## **Example 1:**

### ***Oil Painting***

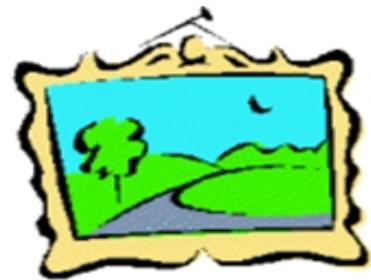


#### **Qualitative data:**

- blue/green color, gold frame
- smells old and musty
- texture shows brush strokes of oil paint
- peaceful scene of the country
- masterful brush strokes

## **Example 1:**

### ***Oil Painting***



#### **Quantitative data:**

- picture is 10" by 14"
- with frame 14" by 18"
- weighs 8.5 pounds
- surface area of painting is 140 sq. in.
- cost \$300

## **Example 2:**

### ***Latte***



#### **Qualitative data:**

- robust aroma
- frothy appearance
- strong taste
- burgundy cup

## **Example 2:**

### ***Latte***



#### **Quantitative data:**

- 12 ounces of latte
- serving temperature 150° F.
- serving cup 7 inches in height
- cost \$4.95

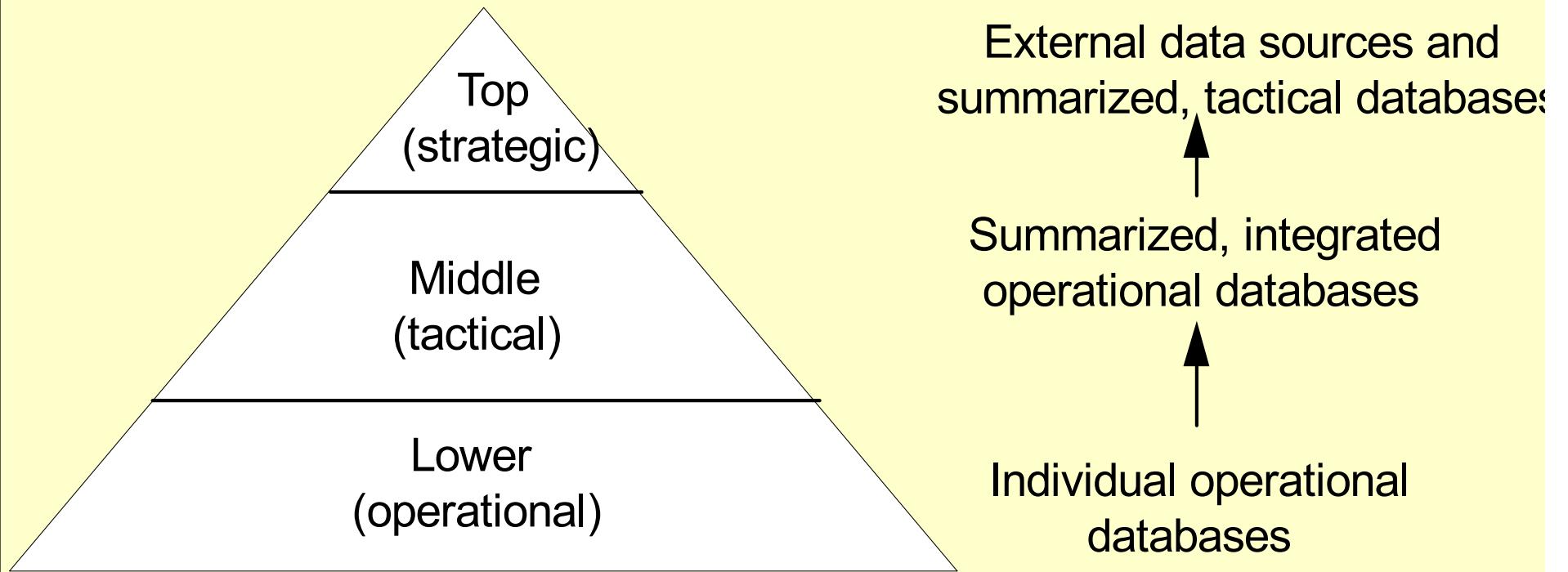
# Categorization of information

## – Level

- **Strategic information:** for **long-term decisions** (eg identify new markets and products; plan growth; reallocate resources across divisions.)
- **Tactical information:** used for **tactical planning and decision-making** within the guidelines set by the strategic plan. (eg choose suppliers; revise staffing levels; forecast sales, inventory and cash; prepare budgets.)
- **Operational information:** for **operational planning** based on the **tactical plans**. (eg correct order delays; schedule employees; find production bottlenecks; monitoring resource usage.)

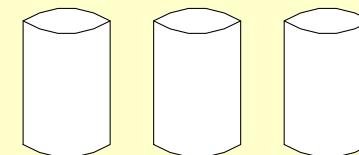
# Level of Information

Management Hierarchy



Transaction databases eg human resource db, inventory db, production db etc

Operational databases



# Categorization of information

## – Time

- **Historic information** : gathered and stored over a period of time; allows decision makers to draw comparisons between previous and present activities; can be used to identify trends over a period of time.
- **Present** : information created from activities during the current work window.
- **Future** : information that is created using present and historic information to try to predict the future activities, trends and events relating to the operation of an organization.

# MALAYSIA WEATHER EXTREMES

	Today (Wed)	Thu	Fri	Sat	Sun	Mon
HOTTEST	Klang max 32 °C	Sungai Petani max 33 °C	Sungai Petani max 34 °C	Alor Setar max 35 °C	Alor Setar max 35 °C	Sungai Petani max 35 °C
COLDEST	Ipoh min 25 °C	Sibu min 24 °C	Ipoh min 25 °C	Ipoh min 26 °C	Ipoh min 26 °C	Sibu min 25 °C
WETTEST	Ipoh 16 mm	Kota Kinabalu 51 mm	Sandakan 19 mm	Sandakan 18 mm	Kota Kinabalu 10 mm	Kota Kinabalu 23 mm
WINDIEST	Kota Bahru 15 km/h	Kuala Terengganu 16 km/h	Johore Bharu 17 km/h	George Town 24 km/h	George Town 23 km/h	Kuala Terengganu 15 km/h

Historic, Present or Future Information?

# Information Capturing and Representation

## Information Capturing:

- OCR, Barcode
  - RFID
  - Document Imaging
  - The WWW (unstructured)
    - Twitter
    - Blogs
    - Social Networks
    - Web pages
- (Requires **Big Data Analytics**)

**Big Data** is the current term for the enormous data sets generated by Web and mobile applications. Nonrelational databases (**No SQL**) are used to store Big Data

# Big Data Analytics

- **Politics** (the prediction of election results)
- **Business** (targeted social media advertising etc)
- **Health care** (the identification of epidemics and market efficiencies in delivery)
- **etc**

The screenshot shows the homepage of HealthITAnalytics.com. The header features the site's logo, "HEALTH IT ANALYTICS", with "xtelligent HEALTHCARE MEDIA" underneath. A navigation bar includes links for Home, News, Features, Interviews, Podcasts, and Research. Below the header is a secondary navigation bar with links for Population Health, Precision Medicine, Quality & Governance, Tools & Strategies, Analytics in Action, and Focus on AI. A section titled "ANALYTICS IN ACTION NEWS" contains an article with the title "Intersection of Big Data Analytics, COVID-19 Top Focus of 2020". The article's text reads: "The top stories on HealthITAnalytics in 2020 focused on the industry's use of big data analytics tools to understand, track, and reduce the spread of COVID-19."

**ANALYTICS IN ACTION NEWS**

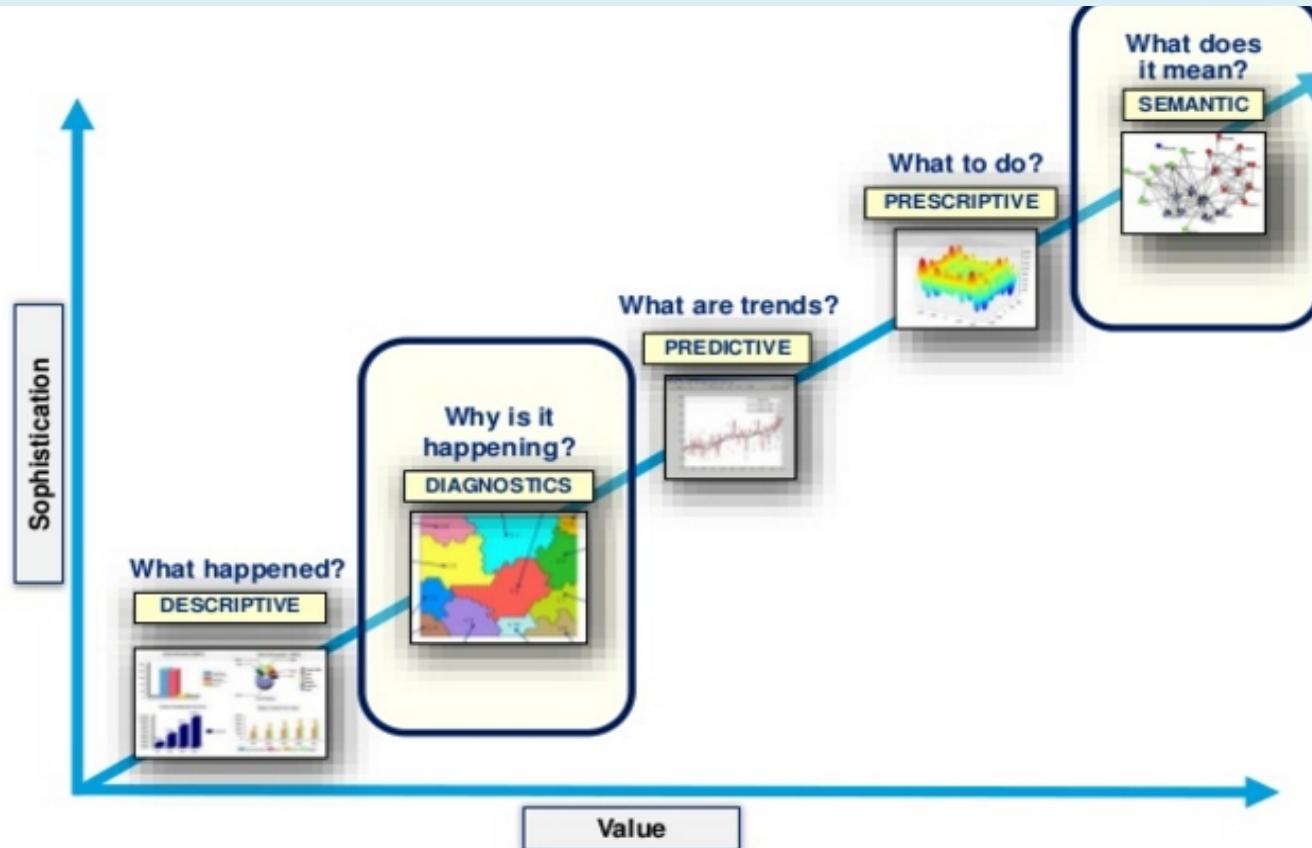
**Intersection of Big Data Analytics, COVID-19 Top Focus of 2020**

The top stories on HealthITAnalytics in 2020 focused on the industry's use of big data analytics tools to understand, track, and reduce the spread of COVID-19.

<https://healthitanalytics.com/news/intersection-of-big-data-analytics-covid-19-top-focus-of-2020>

# Big Data Analytics

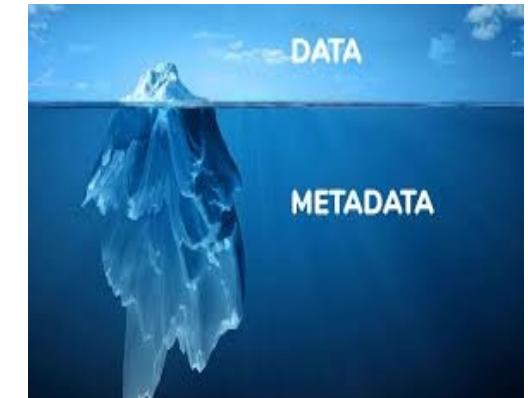
**Big data** - data that exist in very large volumes and many different varieties (data types) and that need to be processed at a very high velocity (speed).



**Analytics** - Systematic analysis and interpretation of data—typically using mathematical, statistical, and computational tools—to improve our understanding of a real-world domain

# Metadata/schema association with data

**Metadata** (Greek: meta- + Latin: data “information”), literally “**data about data**,” ie **information about another set of data.**



A common example is a **library catalog**, which contains data about the contents and location of a book.

ie they are data about the data in the book

Sample Catalog Record
Author: Kesey, Ken.
Title: One flew over the cuckoo's nest, a novel.
Published: New York, Viking Press [1962]
LC Call No.: PZ4.K42On
Subjects: Psychiatric hospital patients--United States--fiction
Control No.: 62008602

# Metadata/schema association with data

## Data dictionary

Name	Type	Length	Min	Max	Description	Source
Course	Alphanumeric	30			Course I D and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student I S
I D	Integer	9			Student I D (S S N)	Student I S
Major	Alphanumeric	4			Student major	Student I S
G P A	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

# Metadata/schema association with data

- How a database sees data so that it can be found
  - by storing metadata/schema associated with data
- DBMS schema are stored in **data dictionary**
  - In Oracle, the views such as **ALL\_USERS**, **ALL\_TABLES**, **ALL\_COLS\_PRIVS**, **ALL\_CONS\_COLUMNS** enable one to retrieve the data dictionary

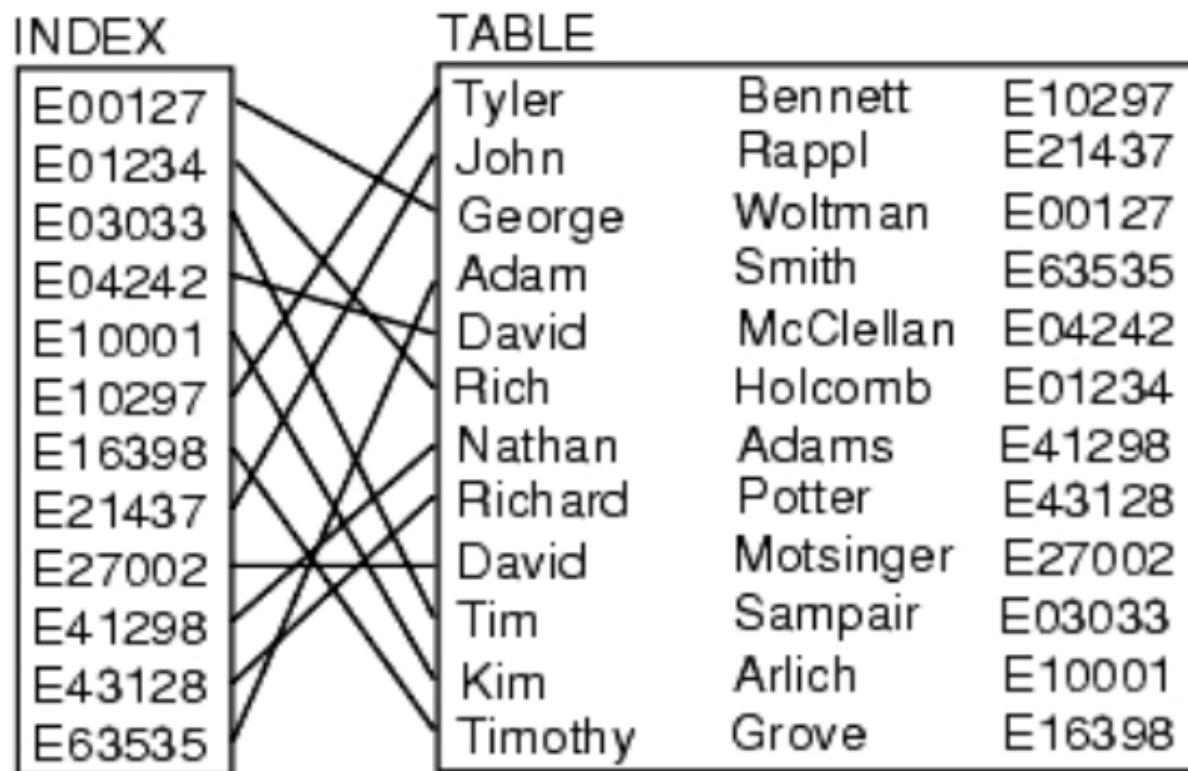
```
SQL> desc all_users
```

Name	Null?	Type
-----	-----	-----
USERNAME	NOT NULL	VARCHAR2(30)
USER_ID	NOT NULL	NUMBER
CREATED	NOT NULL	DATE

# Indexing for fast access

- Types of indexing
  - Hashed files, B<sup>+</sup>-trees, bit-map etc
  - Database efficiency and tuning

Select \* from employee where empId = 'E03033';



# Access through queries

- **Procedural queries** is cumbersome and prone to error
  - programmer must provide the right sequence of instructions
  - Requires some technical knowledge
- **Declarative queries**
  - Only need to state **WHAT** you need not **HOW** to get it
  - E.g SQL
- **Navigational queries**
  - where the searcher knows where he wants to go to find something.
  - Navigational queries come in two types:
    - a word, name or brand strongly or uniquely associated with one particular web site: HP, ebay, hotmail, new york times, yahoo.
    - a partial or complete web address: ebay.com, www.hotmail, yahoo.com

# Information Security

- Information security is designed to protect the
  - confidentiality
  - integrity
  - availability of information



# Information Security

**Confidentiality** – preserving authorized restrictions on information access and disclosure

**Integrity** – guarding against improper information modification or destruction, including ensuring **information authenticity** and **non-repudiation**

**Availability** – ensuring timely and reliable access to and use of information.



# Data Breaches Expose 4.1 Billion Records In First Six Months Of 2019



Davey Winder Senior Contributor @  
Cybersecurity

I report and analyse breaking cybersecurity and privacy stories

According to **Risk Based Security** research newly published in the [2019 MidYear QuickView Data Breach Report](#), the first six months of 2019 have seen more than **3,800 publicly disclosed breaches** exposing an incredible 4.1 billion compromised records.

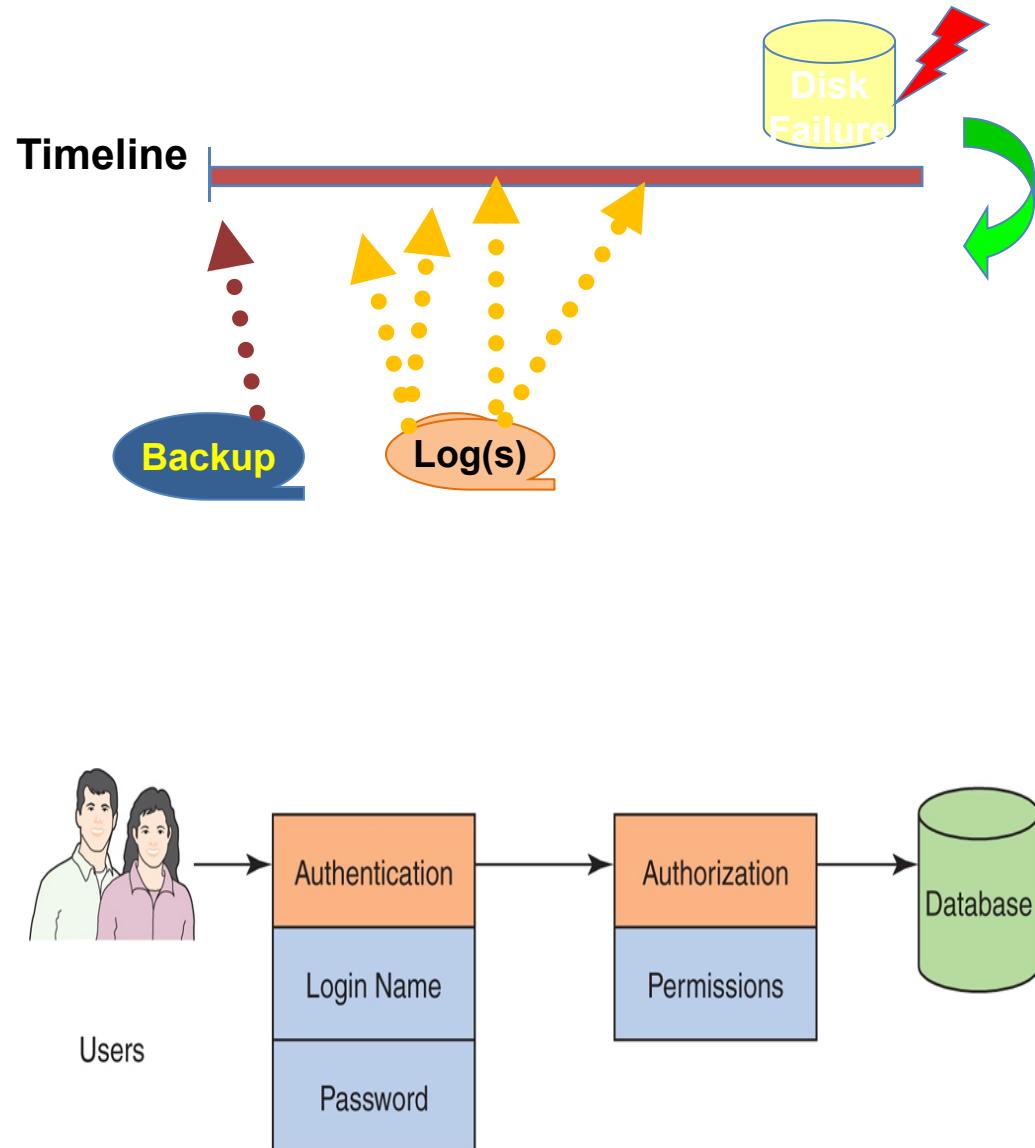
Perhaps even more remarkable is the fact that **3.2 billion** of those records were exposed by just **8 breaches**. As for the exposed data itself, the report has **email (contained in 70% of breaches)** and **passwords (65%)** at the top of the pile.

# Threats to Information Security

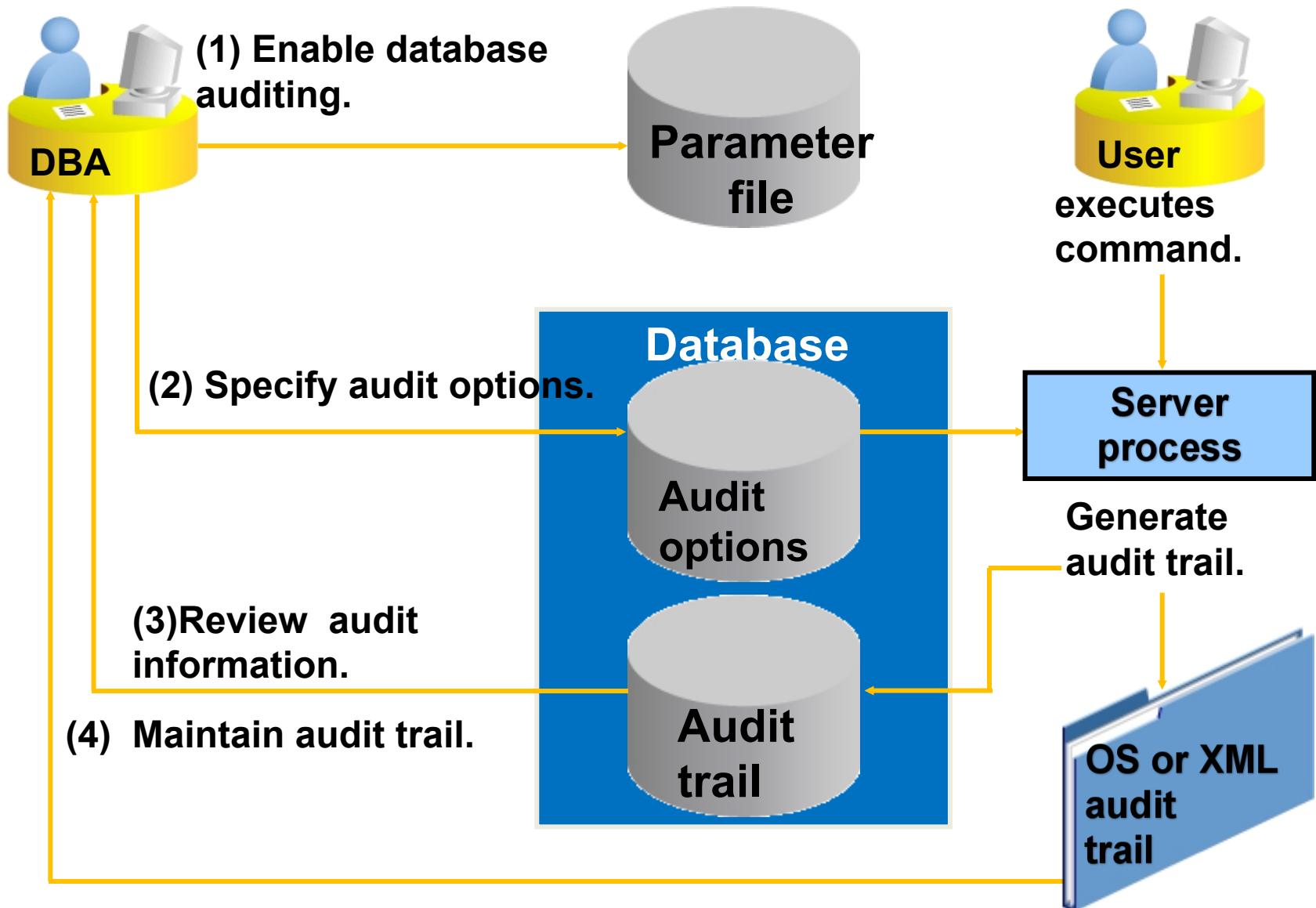
- **Accidental losses attributable to:**
  - Human error
  - Software failure
  - Hardware failure
- **Theft and fraud**
- **Loss of privacy or confidentiality**
  - Loss of privacy (personal data)
  - Loss of confidentiality (corporate data)
- **Loss of data integrity**
- **Loss of availability**
  - disruption of access to information or use of information
    - e.g., through sabotage

# Information Security

- **Backup & recovery (ch 8)**
- **Physical security** - computerized & manual policies and procedures, fence, lock & key, security guards, etc
- **Technical security** – encryption, authentication, authorization, audit trace etc



# Database Audit Trace



# Information Assurance

- Information Assurance (IA) is defined as "measures that protect and defend information and information systems by ensuring their **availability, integrity, authentication, confidentiality, and non-repudiation.**"

**Nonrepudiation is the assurance that someone cannot deny something**

# References

- *Database Systems: A Practical Approach to Design, Implementation and Management.* Connolly, T. M. and Begg, C. E. .
- *Modern Database Management.* Hoffer, J.A., Prescott, M., and McFadden,F.
- *Database System Concepts* . Silberschatz, A., Korth, H,. and Sudarshan, S.
- *Fundamentals of Database Systems.* Elmasri, R. and Navathe, S.B.