

## Lecture 1.2

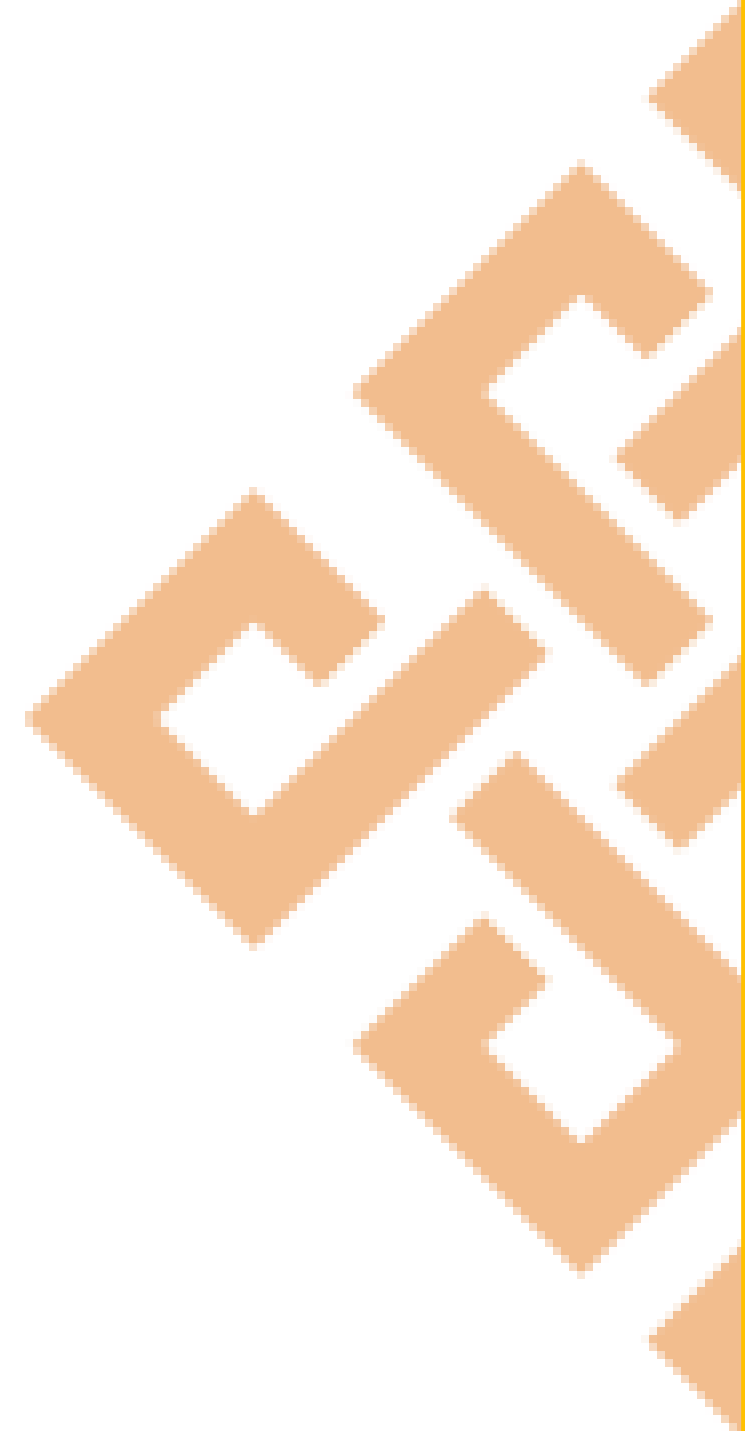
# Classification of digital data

**School of Computer Science & Engineering**

AY: 2021-2022

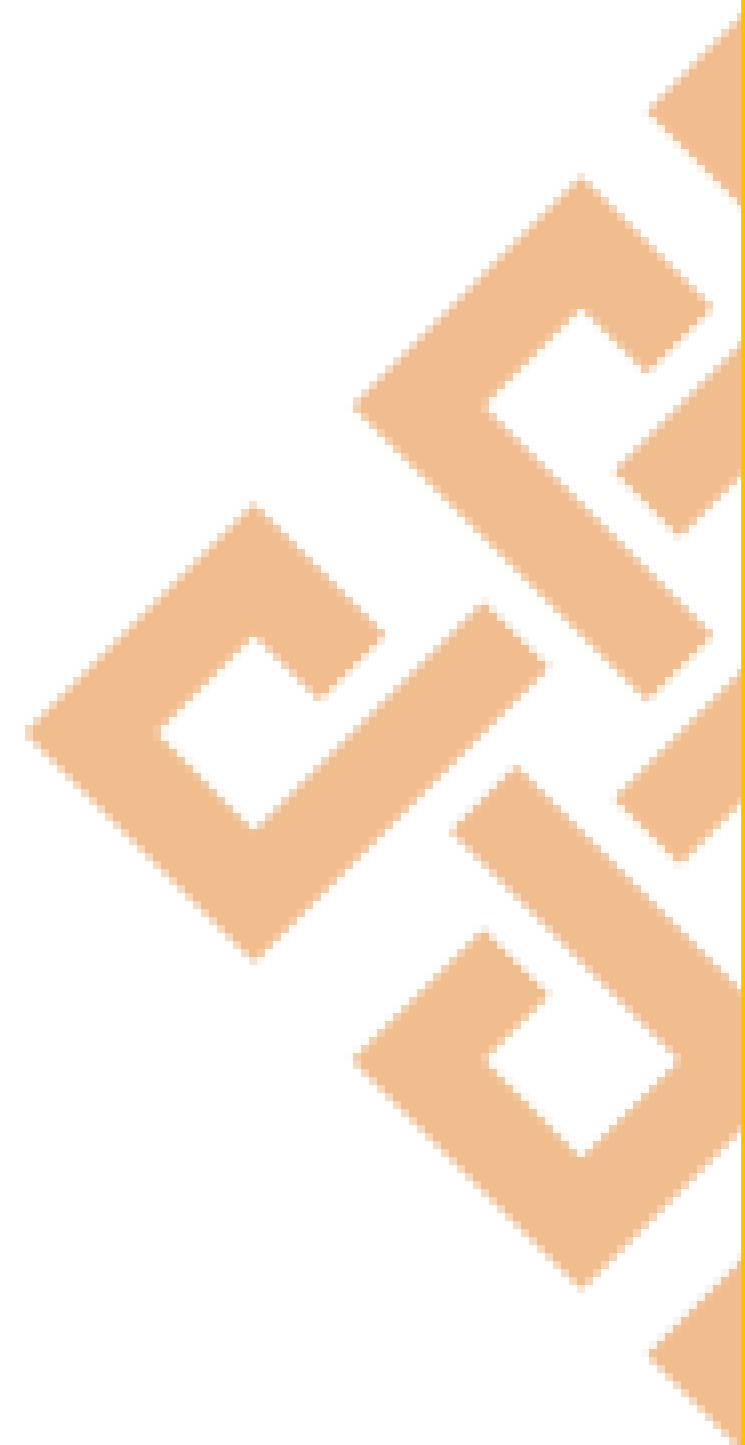
# **Classification of digital data**

**Recap of previous Lecture**



# **Classification of digital data**

**Topic of the Lecture**



# TOPIC OF THE LECTURE

Introduction to Digital Data

Structured Data

Semi-structured Data

Unstructured Data

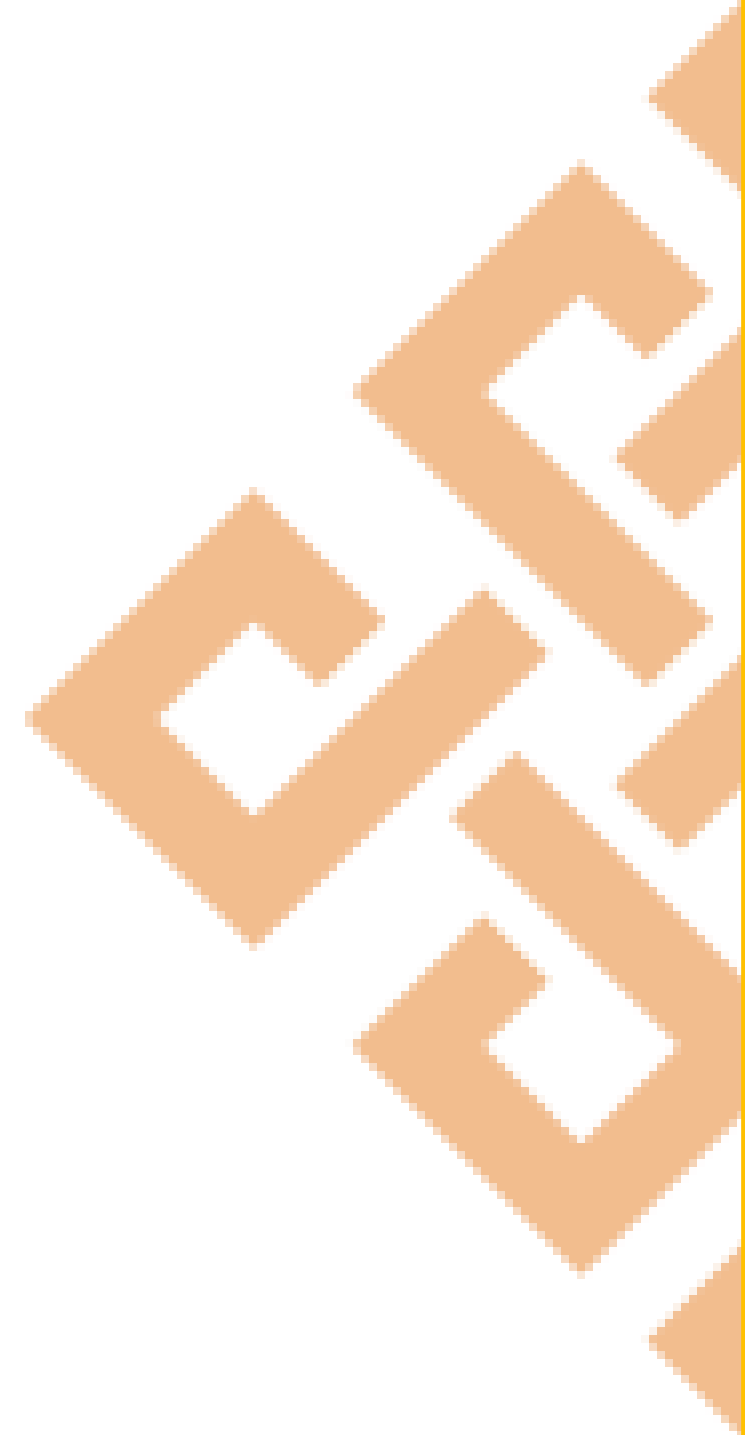
Structured Vs. Unstructured Data

Characteristics of Data



# **Classification of digital data**

## **Introduction to Digital Data**



# CLASSIFICATION OF DIGITAL DATA

## Introduction to Digital Data

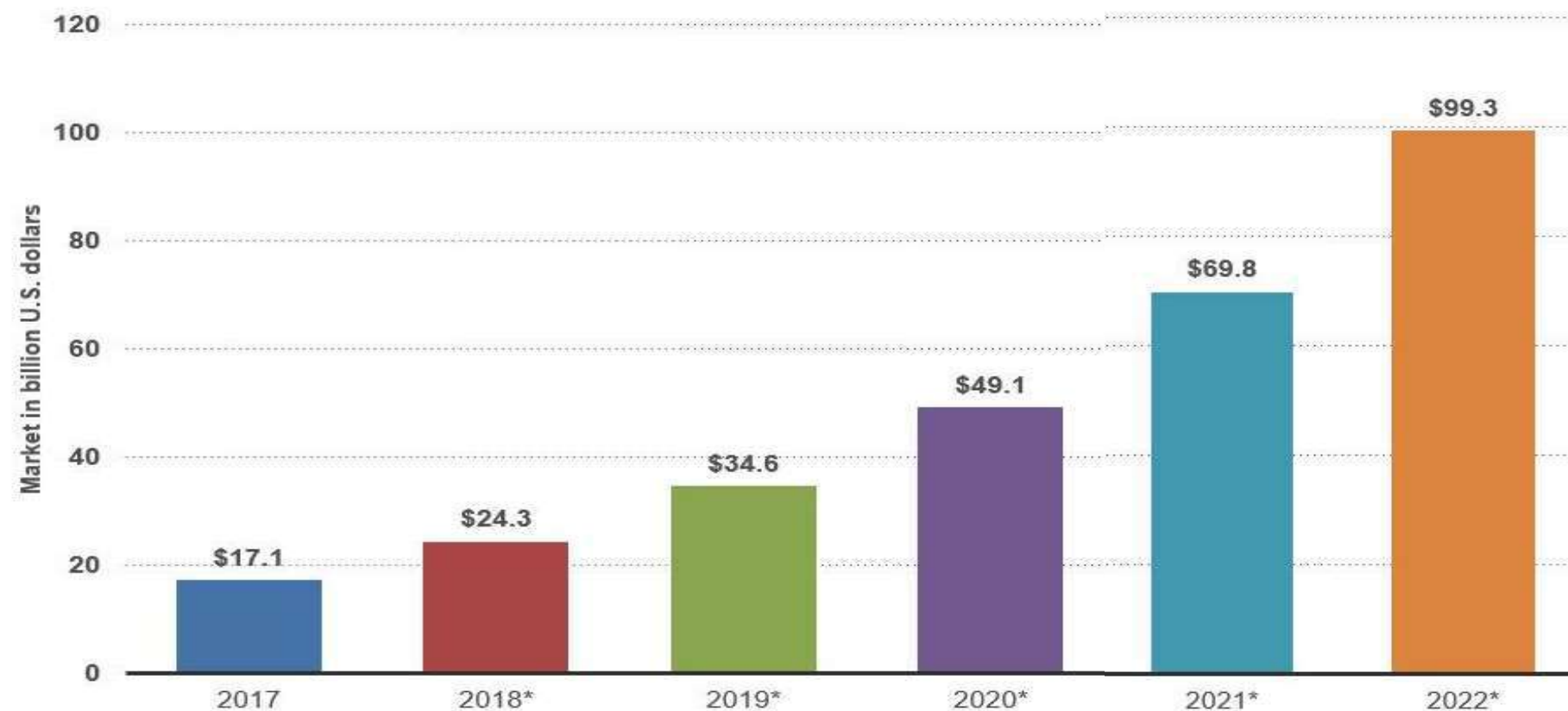
NAME	SYMBOL	VALUE	EQUAL VALUE
byte	b	8 bits	1 byte
kilobyte	Kb	1024 bytes	1 024 bytes
megabyte	MB	1024 KB	1 048 576 bytes
gigabyte	GB	1024 MB	1 073 741 824 bytes
terabyte	TB	1024 GB	1 099 511 627 776 bytes
Petabyte	PB	1024 TB	1 125 899 906 842 624 bytes
Exabyte	EB	1024 PB	1 152 921 504 606 846 976 bytes
Zetabyte	ZB	1024 EB	1 180 591 620 717 411 303 424 bytes
Yottabyte	YB	1024 ZB	1 208 925 819 614 629 174 706 176 bytes
Brontobyte	BB	1024 YB	1 237 940 039 285 380 274 899 124 224 bytes
Geopbyte	GB	1024 BB	1 267 650 600 228 229 401 496 703 205 376 bytes



# CLASSIFICATION OF DIGITAL DATA

## Introduction to Digital Data (contd..)

Size of Big Data Market Worldwide in U.S. Billion Dollars (2017 to 2022)



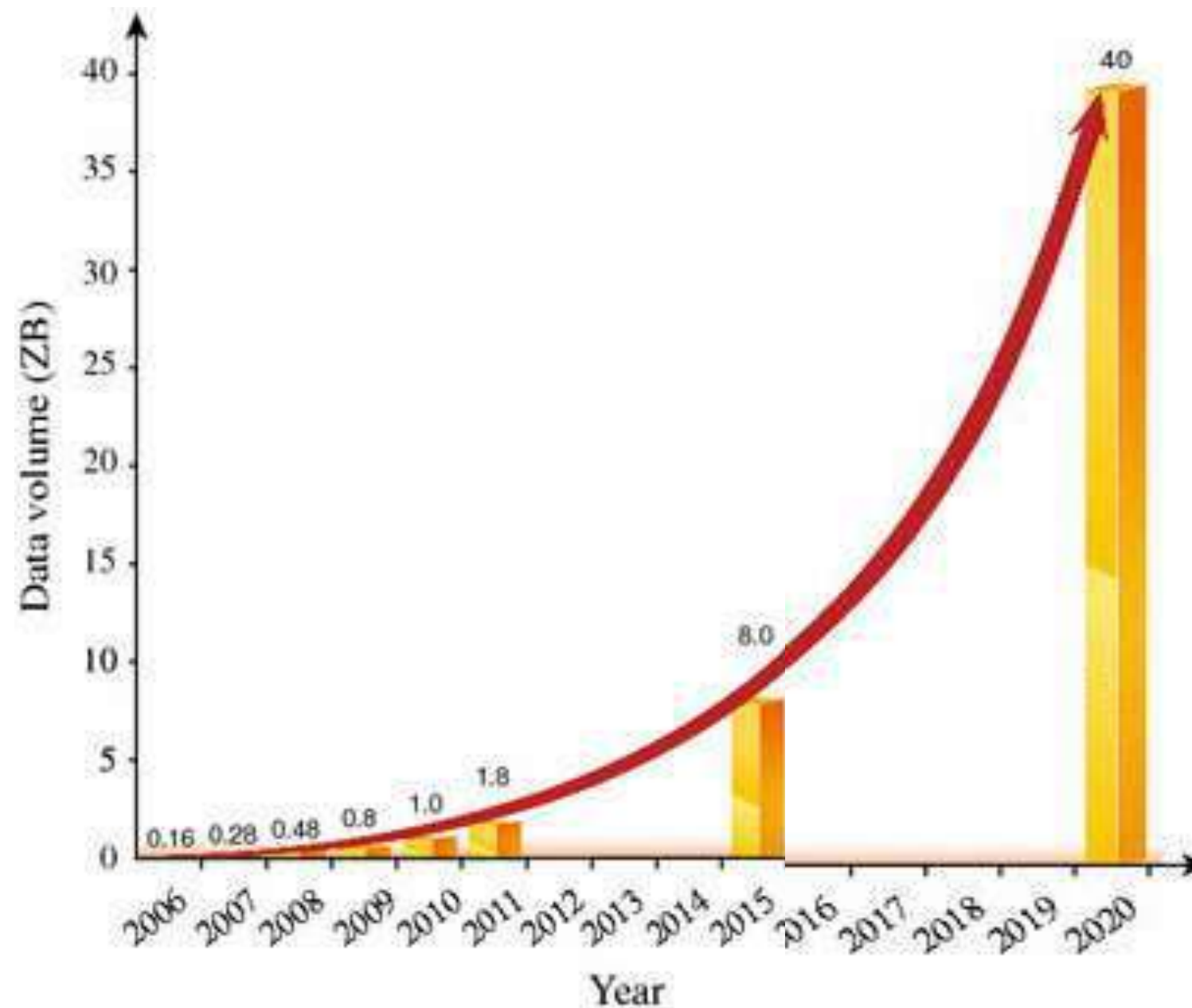
Worldwide Big Data market revenues for software and services are projected to increase from attaining a Compound Annual Growth Rate (CAGR) of 10.48%.

Source: [Wikibon](#) and [reported by Statista](#).



# CLASSIFICATION OF DIGITAL DATA

## Introduction to Digital Data (contd..)



Data growth --- exponential acceleration --- advent of the computer and internet

Defined as the data stored in digital format

Ex. A picture, a document or a video etc.

Not physical --- but stored in digital form





# CLASSIFICATION OF DIGITAL DATA

## Introduction to Digital Data (contd..)

### Structured data

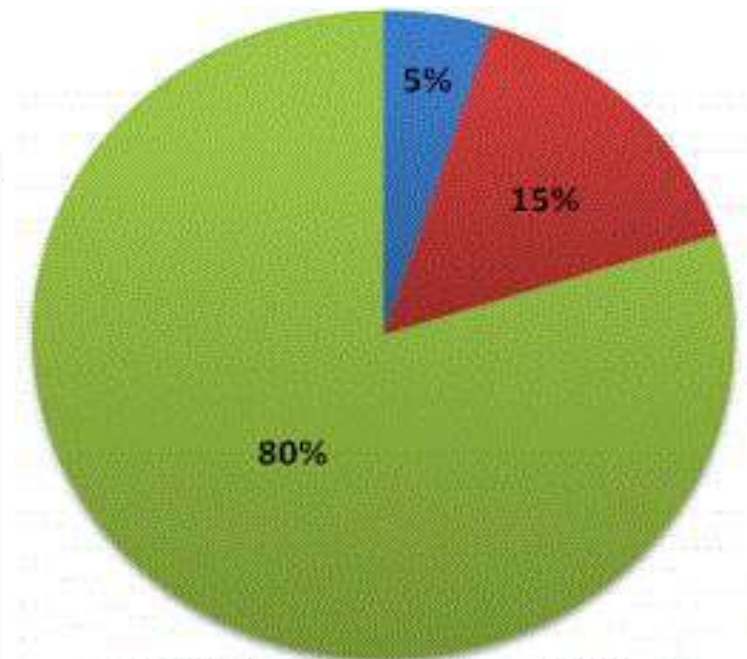
ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

### Semi-structured data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

### Unstructured data

The university has 5600 students.  
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.  
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

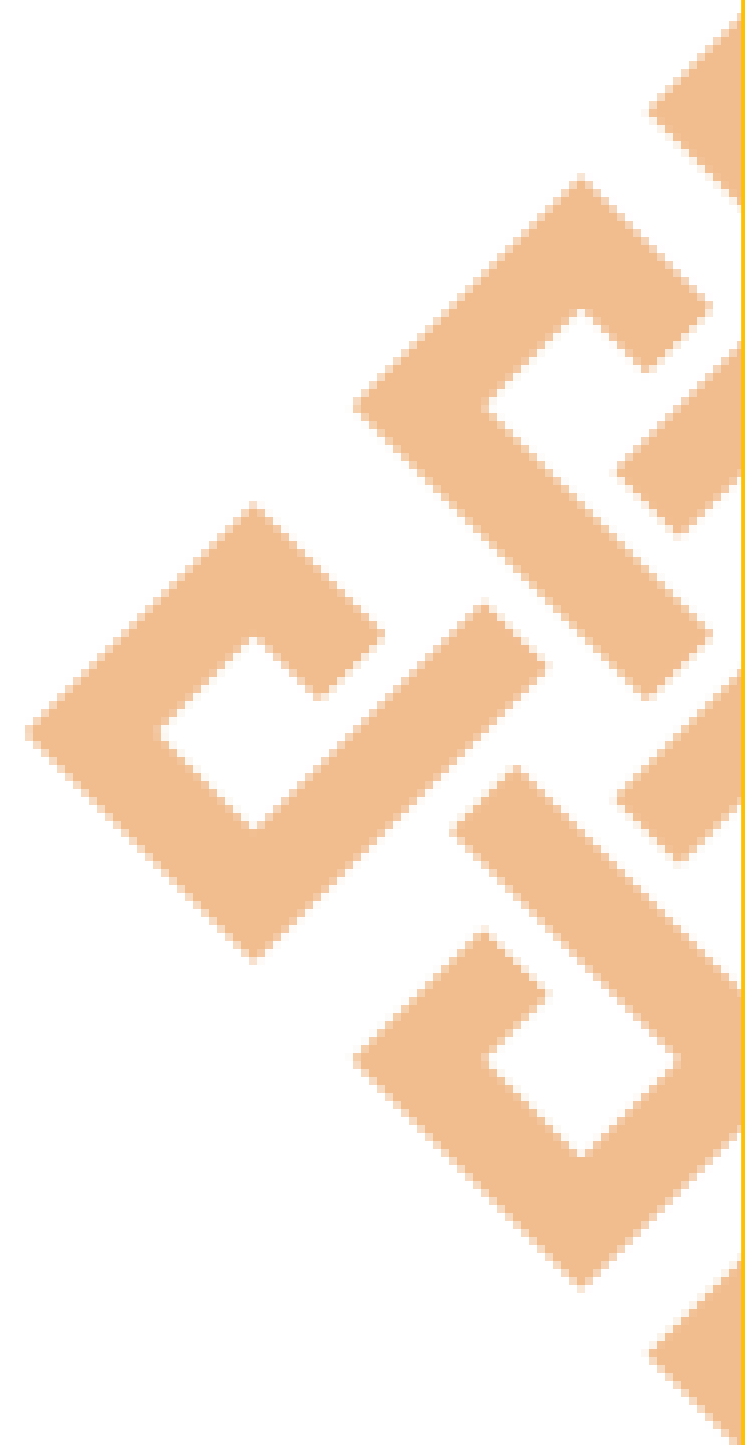


- Structured
- Semi-Structured
- Un-Structured



# **Classification of digital data**

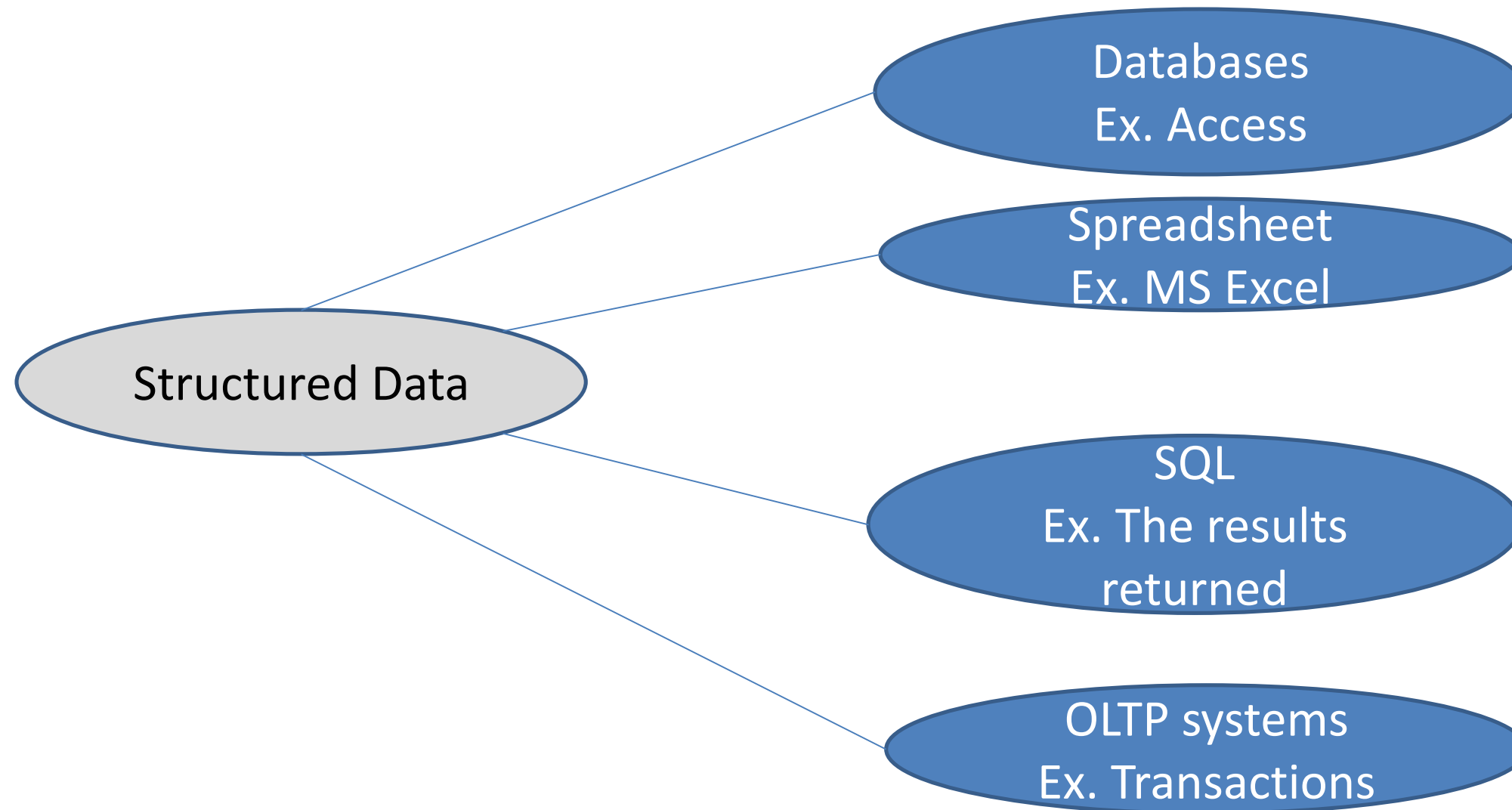
## **Structured Data**



# CLASSIFICATION OF DIGITAL DATA

## Structured Data

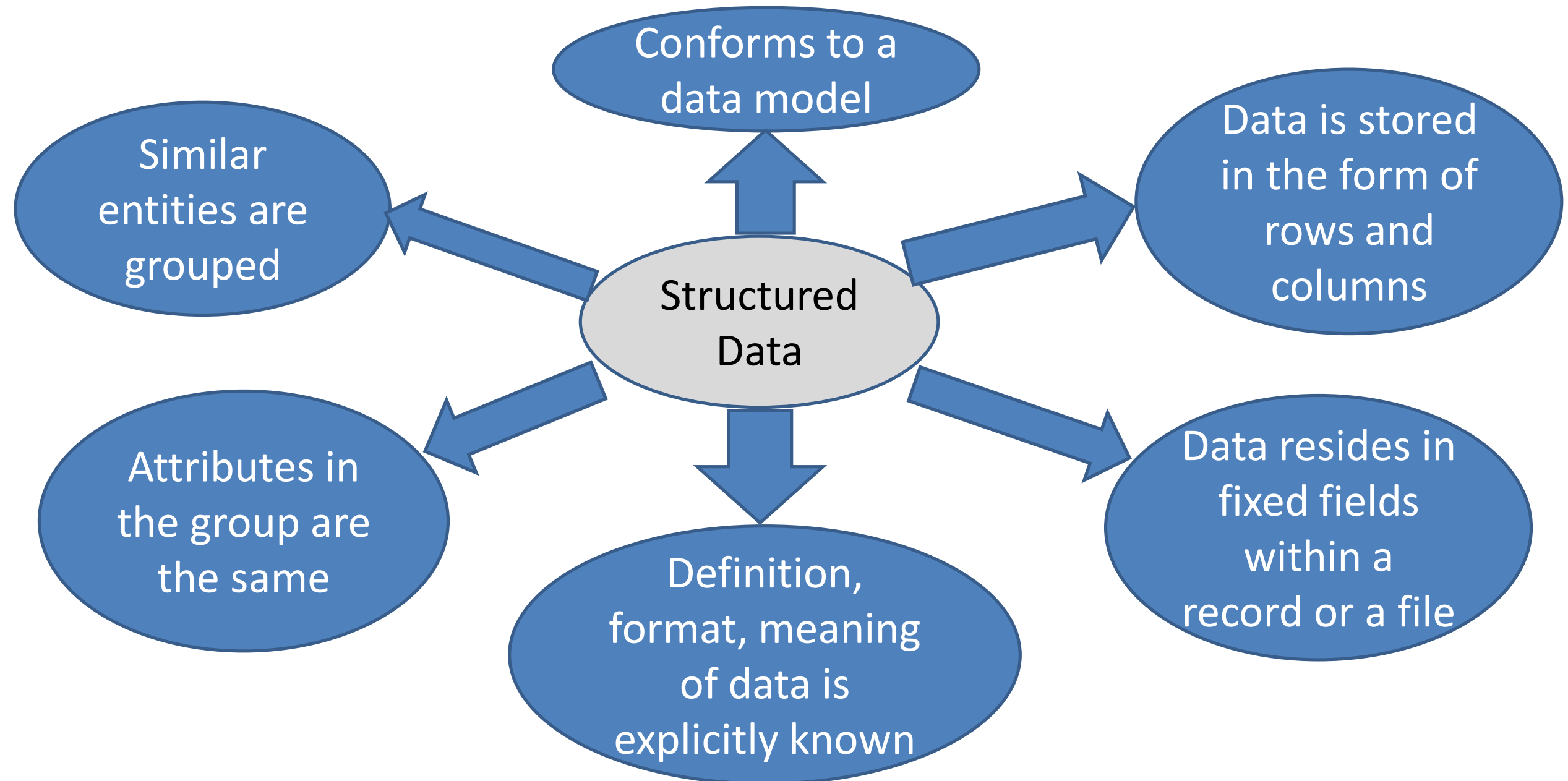
The sources of Structured Data:



# CLASSIFICATION OF DIGITAL DATA

## Structured Data (contd..)

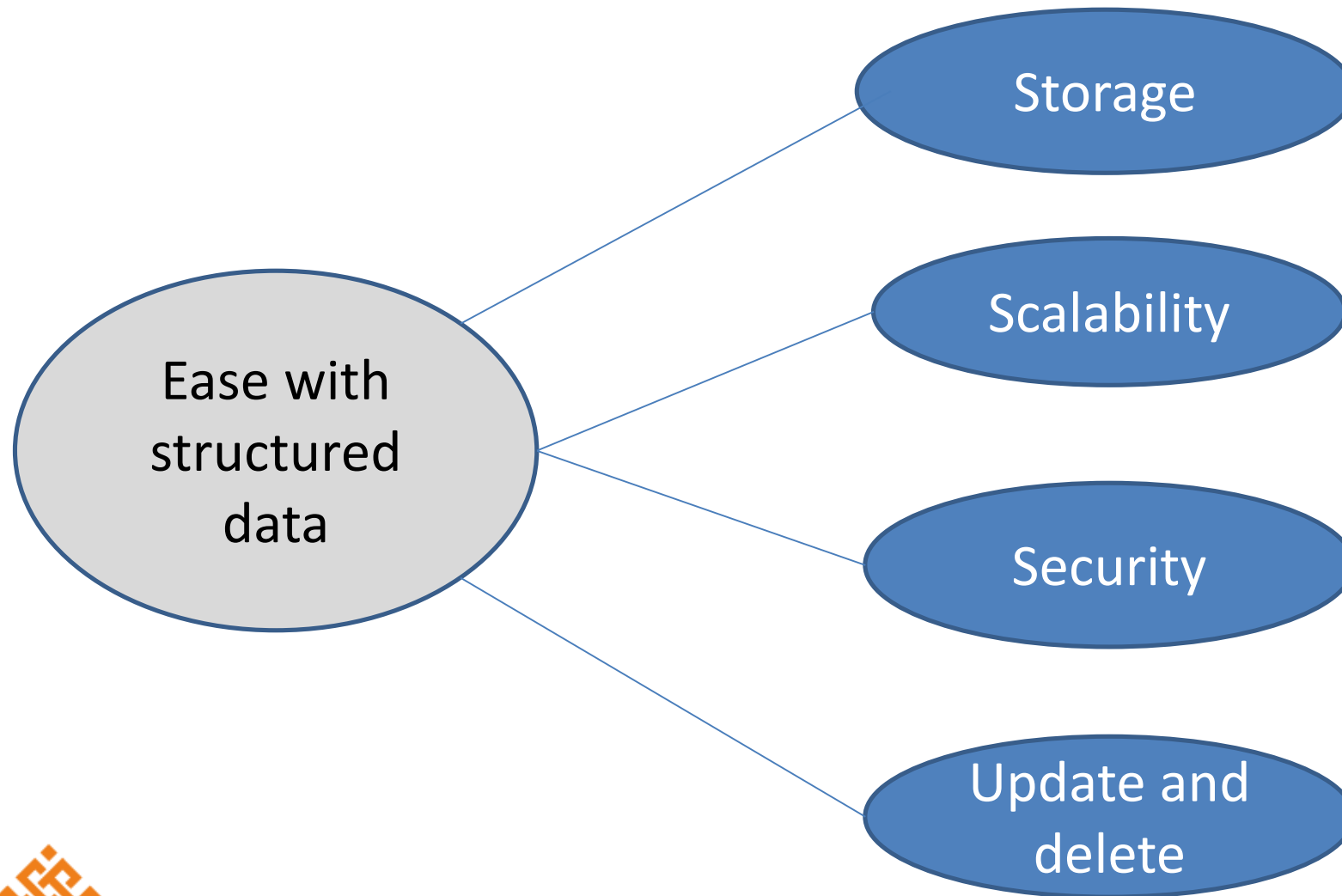
The characteristics of Structured Data:



# CLASSIFICATION OF DIGITAL DATA

## Structured Data (contd..)

The ease of dealing with Structured Data:



➤ Retrieval of structured data is totally hassle free.

➤ **Indexing and searching**

➤ **Mining Data**

➤ **BI operations**



# CLASSIFICATION OF DIGITAL DATA

## Structured Data (contd..)

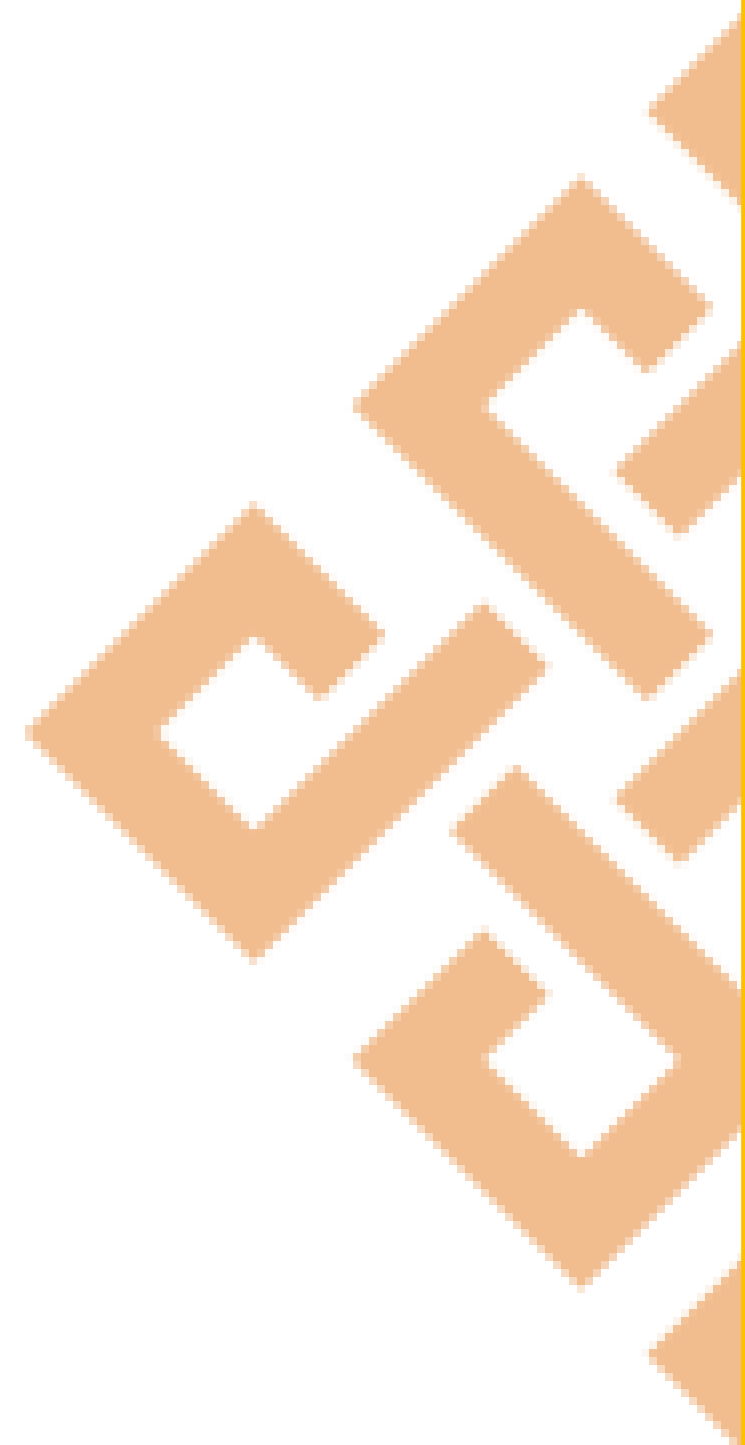
### The summary of Structured Data:

- Consists of fully described data sets.
- Has clearly defined categories and sub-categories.
- Is placed neatly in rows and columns.
- Goes into records and hence the database is regulated by a well-defined structure.
- Can be indexed easily by the Database System itself or manually.



# Classification of digital data

## Semi-structured Data

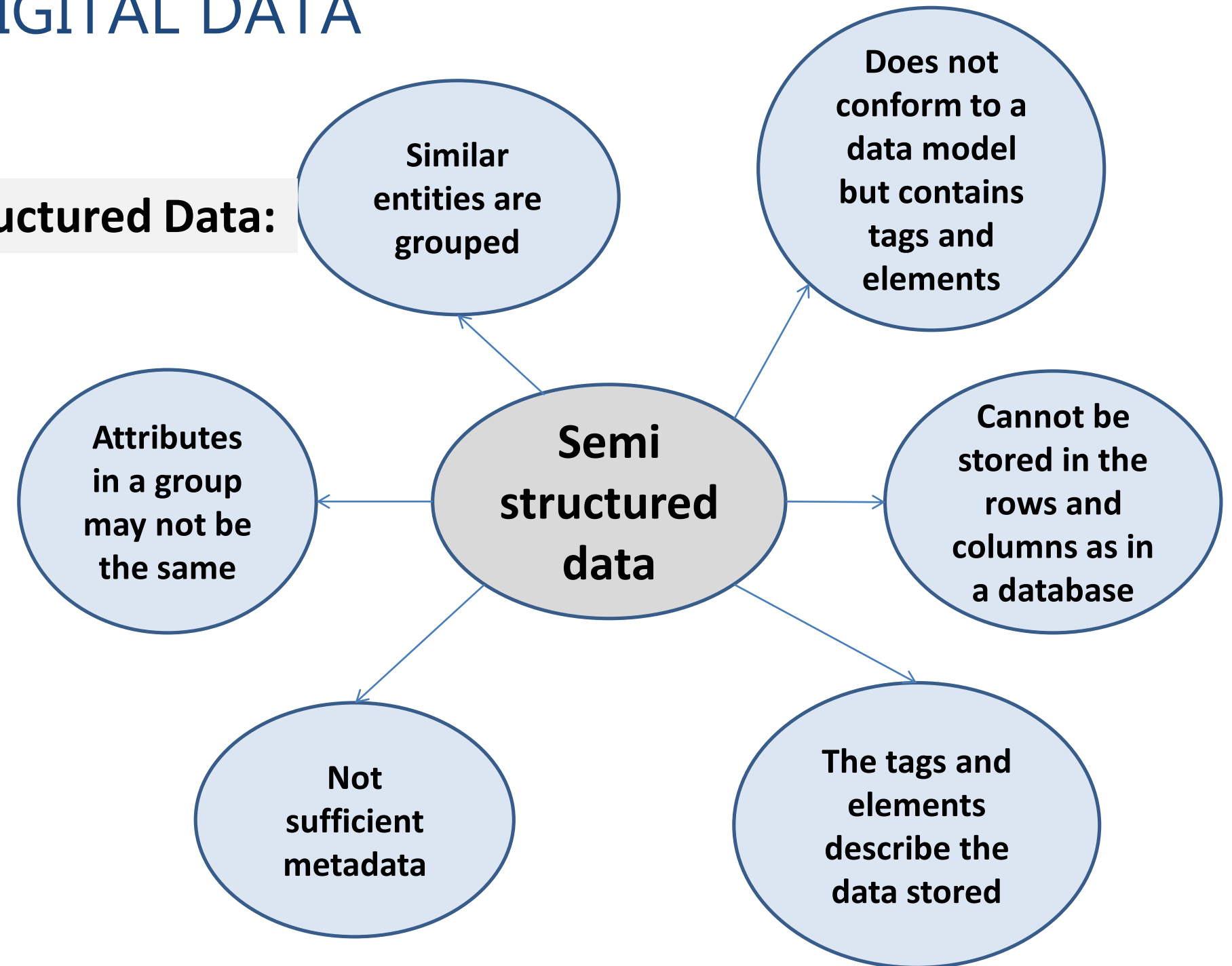


# CLASSIFICATION OF DIGITAL DATA

## Semi-Structured Data

### The characteristics of Semi-Structured Data:

- Only about 10% of data in any organization is semi-structured
- Comes from heterogeneous sources.





# CLASSIFICATION OF DIGITAL DATA

## Semi-Structured Data (contd..)

Email Standard format:

To : <NAME>

From : <NAME>

Subject : <TEXT>

CC : <NAME>

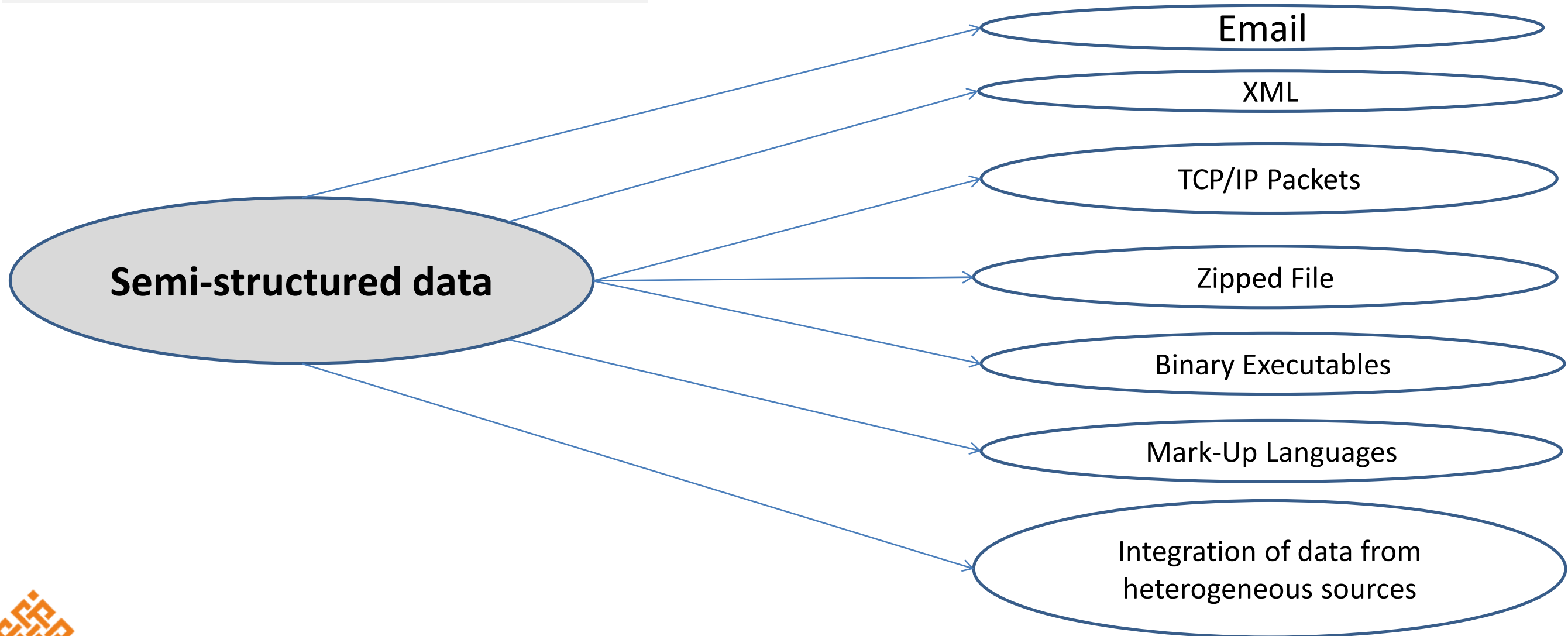
Body : <TEXT,GRAPHICS,IMAGES,ETC>



# CLASSIFICATION OF DIGITAL DATA

## Semi-Structured Data (contd..)

### Sources of Semi-structured Data:



# CLASSIFICATION OF DIGITAL DATA

## Unstructured Data

- Cannot be stored in the form of rows and columns
- Does not conform to any data model
- Difficult to determine the meaning of the data
- Does not follow any rules
- Can be of any type
- Unpredictable



# CLASSIFICATION OF DIGITAL DATA

## Unstructured Data (contd..)

### Major sources of Unstructured data:









- Anything in a non-database form
- It can be divided into two broad categories:
  - **Bitmap objects:** For e.g. Image, video or audio files.
  - **Textual objects:** For e.g. Microsoft word documents, emails or MS Excel.
- A noisy text such as chats, emails and SMS texts.



# CLASSIFICATION OF DIGITAL DATA

## Unstructured Data (contd..)

### Major sources of Unstructured data:

 Text files and documents	 Server, website and application logs	 Sensor data	 Images
 Video files	 Audio files	 Emails	 Social media data

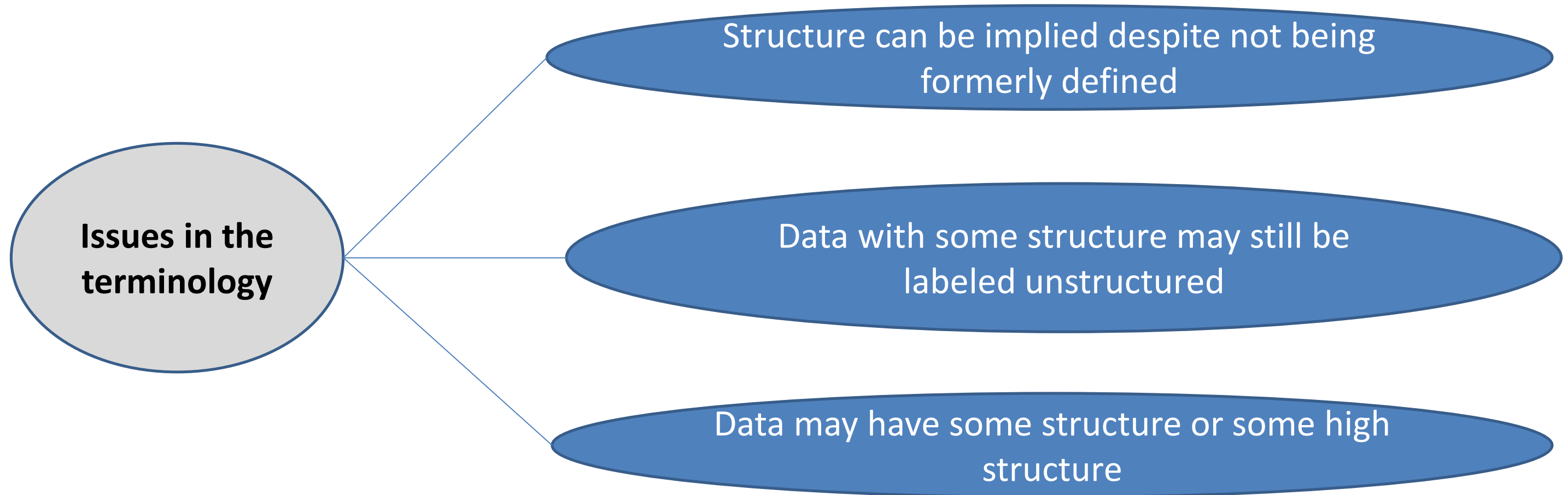
- ❖ Web pages,
- ❖ Memos,
- ❖ Videos (MPEG, etc.),
- ❖ Images (JPEG, GIF, etc.),
- ❖ body of an email,
- ❖ Word document,
- ❖ PowerPoint presentation,
- ❖ Chats, Reports,
- ❖ White papers,
- ❖ Surveys etc.



# CLASSIFICATION OF DIGITAL DATA

## Unstructured Data (contd..)

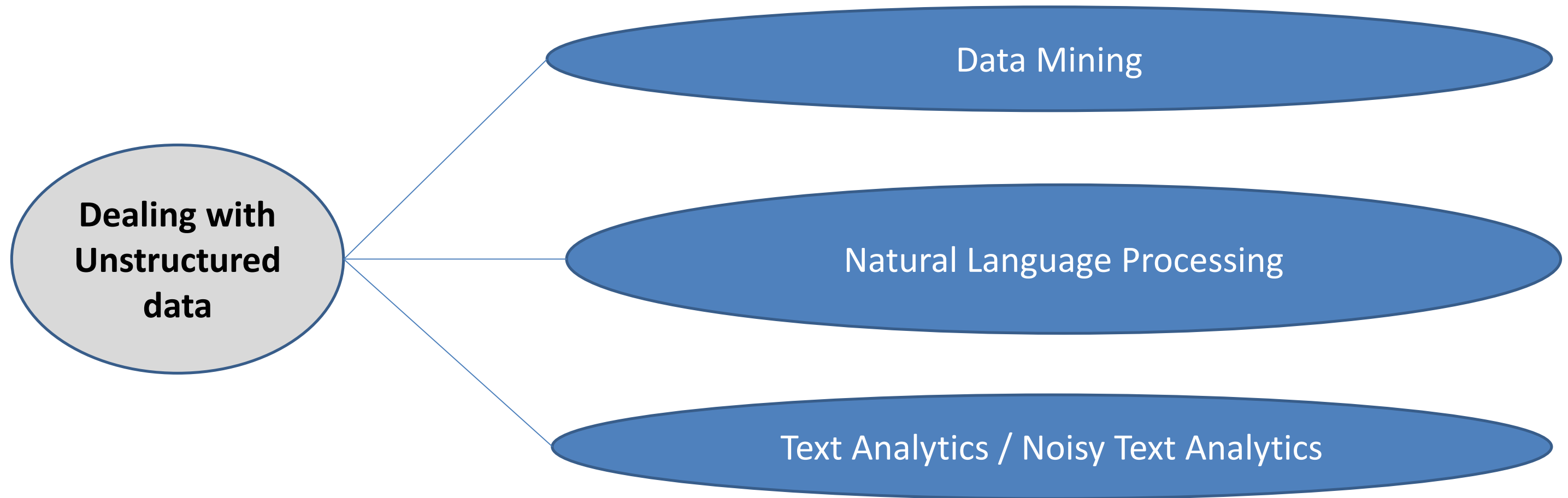
### Issues in the terminology of Unstructured data:



# CLASSIFICATION OF DIGITAL DATA

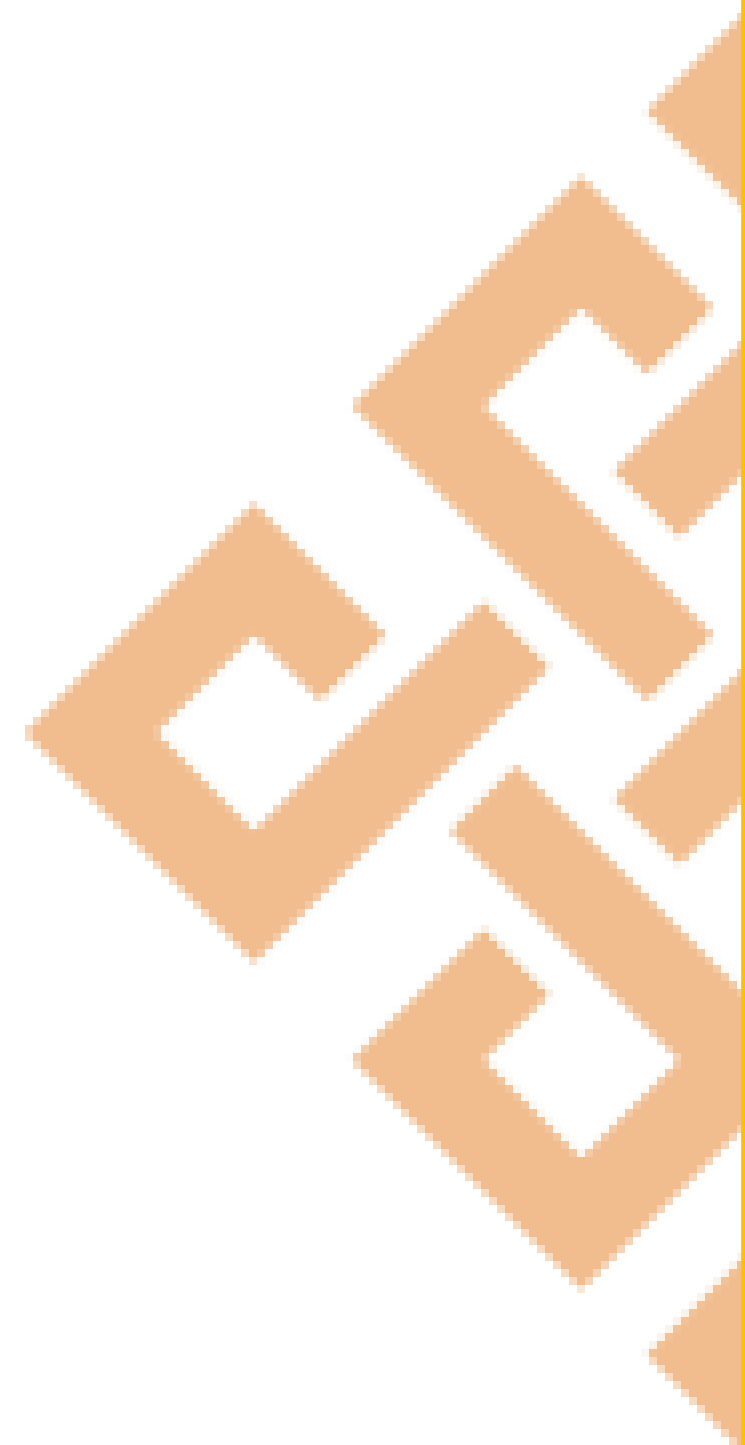
## Unstructured Data (contd..)

Dealing with Unstructured data:



# **Classification of digital data**

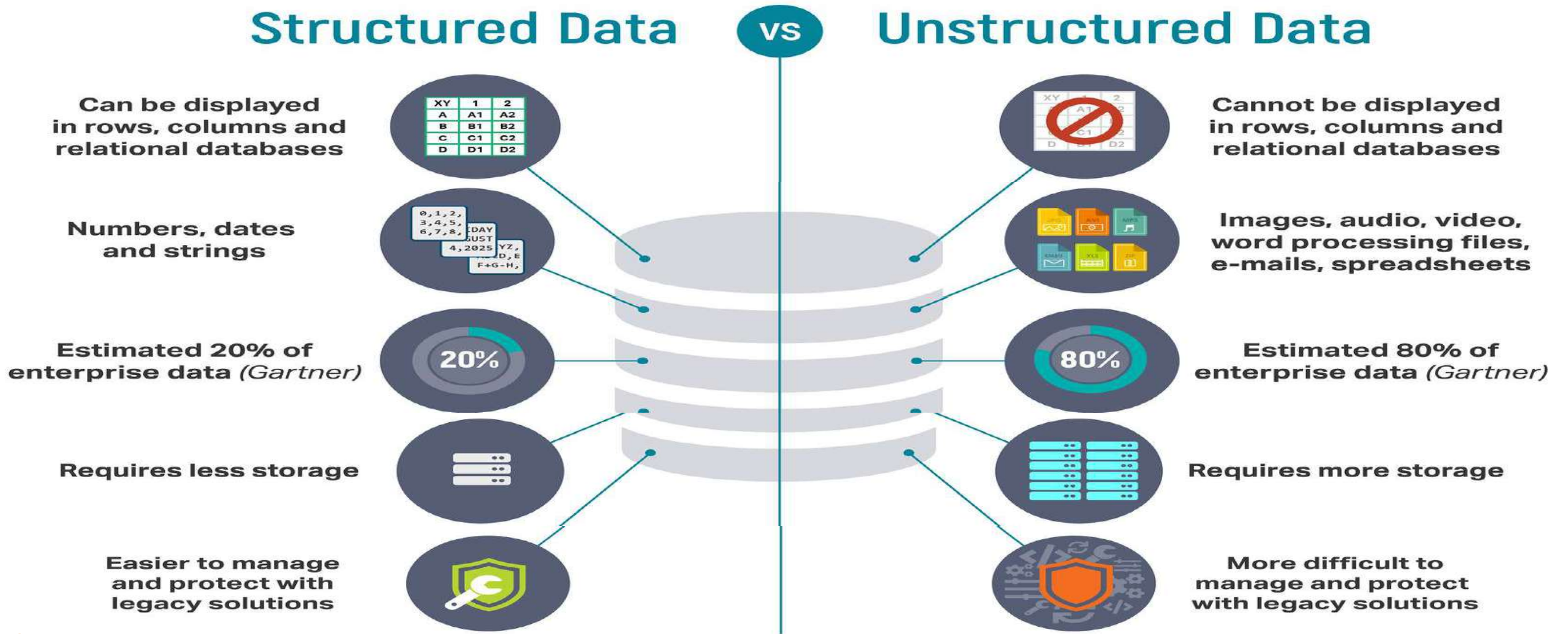
## **Structured Data Vs. Unstructured Data**





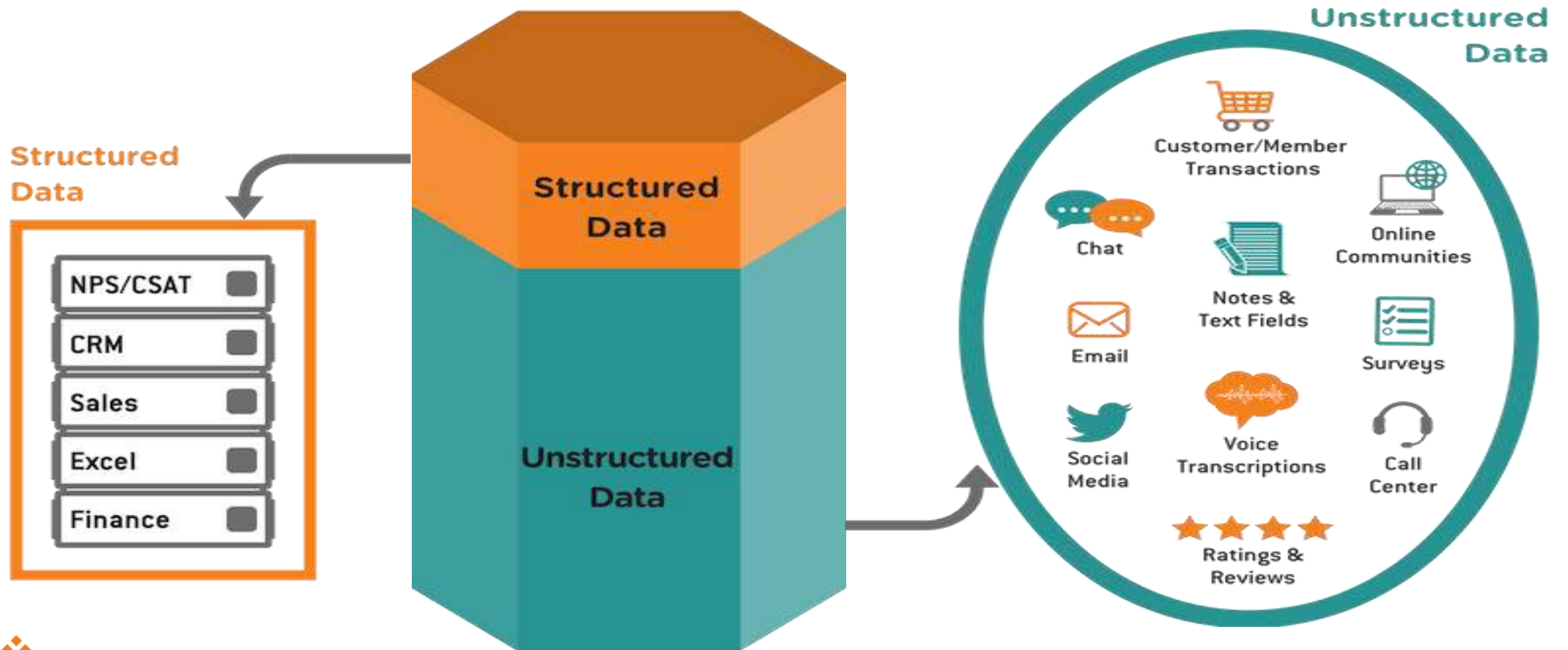
# CLASSIFICATION OF DIGITAL DATA

## Structured Data Vs. Unstructured Data



# CLASSIFICATION OF DIGITAL DATA

## Structured Data Vs. Unstructured Data (contd..)



# CLASSIFICATION OF DIGITAL DATA

## Structured Data Vs. Unstructured Data (contd..)

Structured data

Databases

Semi-structured data

XML / JSON data

Email

Web pages

Unstructured data

Audio

Video

Image data

Natural language

Documents





# CLASSIFICATION OF DIGITAL DATA

## Structured Data Vs. Unstructured Data (contd..)

	Structured Data	Unstructured Data
<b>Characteristics</b>	<ul style="list-style-type: none"><li>• Pre-defined data models</li><li>• Usually text only</li><li>• Easy to search</li></ul>	<ul style="list-style-type: none"><li>• No pre-defined data model</li><li>• May be text, images, sound, video or other formats</li><li>• Difficult to search</li></ul>
<b>Resides in</b>	<ul style="list-style-type: none"><li>• Relational databases</li><li>• Data warehouses</li></ul>	<ul style="list-style-type: none"><li>• Applications</li><li>• NoSQL databases</li><li>• Data warehouses</li><li>• Data lakes</li></ul>
<b>Generated by</b>	Humans or machines	Humans or machines
<b>Typical applications</b>	<ul style="list-style-type: none"><li>• Airline reservation systems</li><li>• Inventory control</li><li>• CRM systems</li><li>• ERP systems</li></ul>	<ul style="list-style-type: none"><li>• Word processing</li><li>• Presentation software</li><li>• Email clients</li><li>• Tools for viewing or editing media</li></ul>
<b>Examples</b>	<ul style="list-style-type: none"><li>• Dates</li><li>• Phone numbers</li><li>• Social security numbers</li><li>• Credit card numbers</li><li>• Customer names</li><li>• Addresses</li></ul>	<ul style="list-style-type: none"><li>• Text files</li><li>• Reports</li><li>• Email messages</li><li>• Audio files</li><li>• Video files</li><li>• Images</li></ul>



# CLASSIFICATION OF DIGITAL DATA

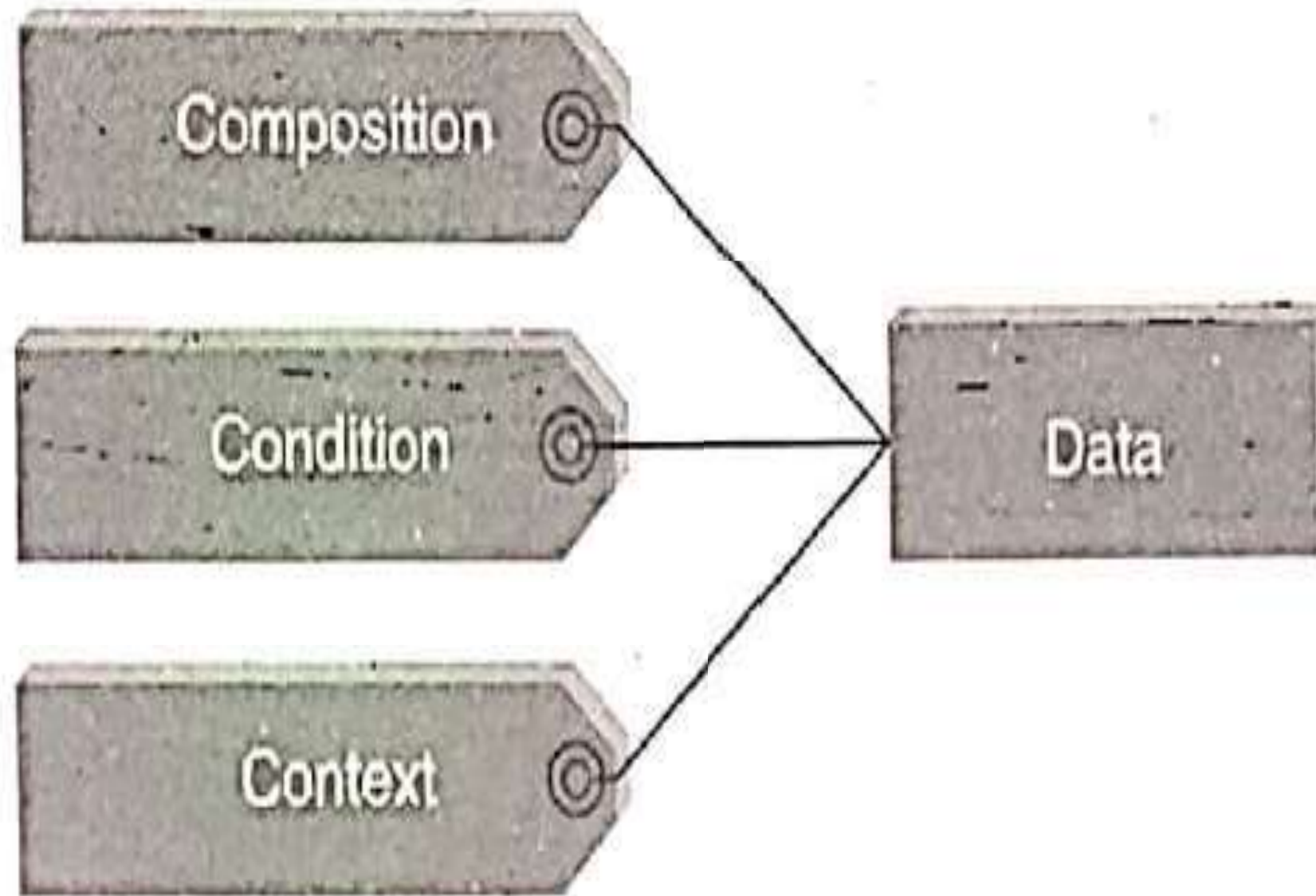
## Structured Data Vs. Unstructured Data (contd..)

	<i>Data Types</i>	<i>Source</i>	<i>Examples in e-Businesses</i>
<i>Structured</i>	Transaction or business activity data	Retail transactions, customer profiles, product consumption, customer complaints	Amazon revealed at one point that 30% of sales were generated through its recommendation engine.
	Click-stream data	social media content, online advertisements	eBay conducts thousands of experiments with different aspects of its website to determine optimal layout.
<i>Unstructured</i>	Video data	Video data from retail and other settings	Netflix uses video data to predict viewing habits and evaluate the quality of customers experiences.
	Voice data	Voice data from phone calls, call centers, customer service	Credit card companies can make personalized offers in milliseconds and to optimize offers by tracking responses.



# CLASSIFICATION OF DIGITAL DATA

## Characteristics of Data





# CLASSIFICATION OF DIGITAL DATA

## Characteristics of Data (contd..)

### ➤ **Composition:**

- structure, source, granularity, nature of data - check if the data is static or real-time processing

### ➤ **Condition:**

- State of the data - is the data clean or needs cleaning?

### ➤ **Context:**

- Data source, associated events – understand the data



# SUMMARY OF THE LECTURE

Introduction  
to digital  
data

Structured  
data

Semi-  
structured  
data

Unstructur  
ed data

Characteris  
tics of data



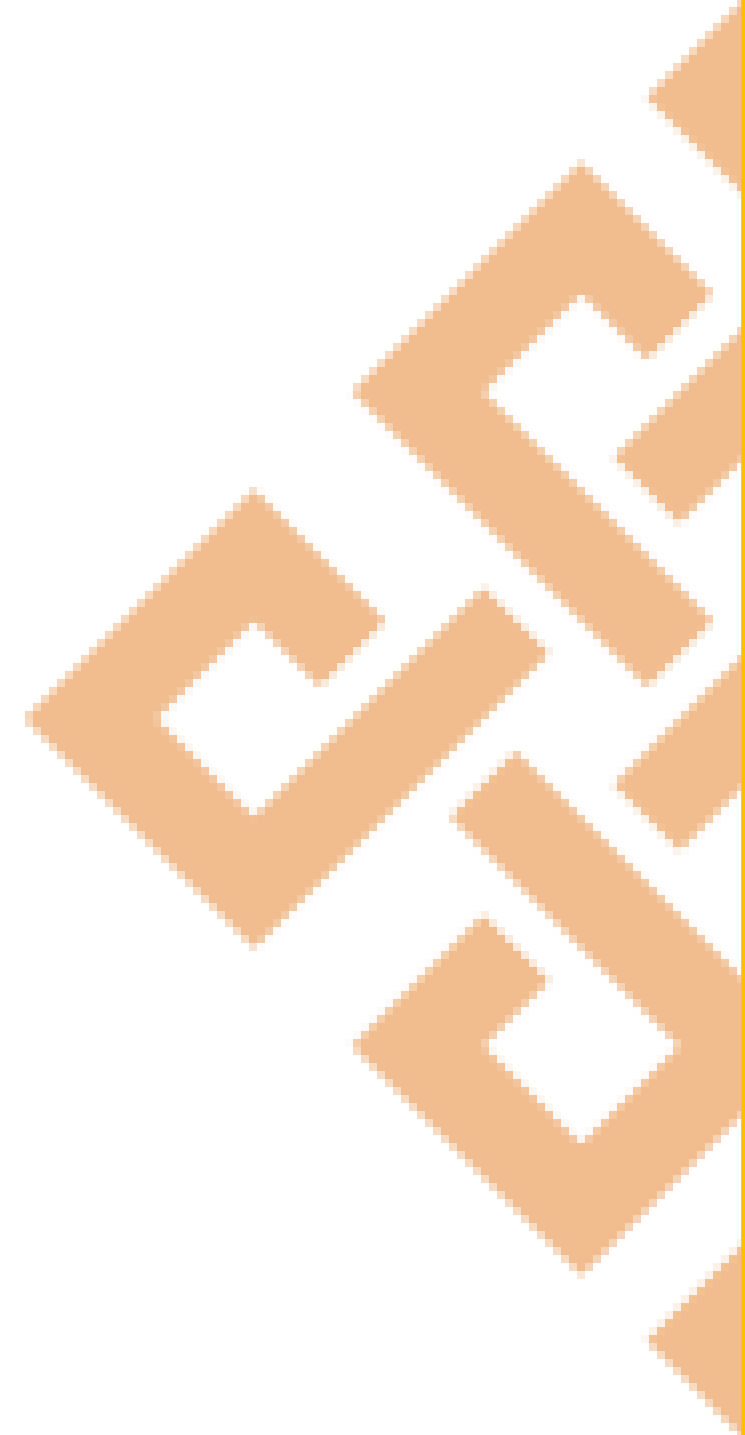


# Lecture 1.4

## Evolution of Big Data

# **Evolution of Big Data**

**Topic of the Lecture**



# EVOLUTION OF BIG DATA

## Introduction



# EVOLUTION OF BIG DATA

## Introduction (contd..)

Big data --- an enigma to many people.  
new term --- coined during the latter part of the last decade.

Ambiguous to many people--- since it's inception

Not just enormous amounts of data --- whole process of gathering, storing and analyzing that data.



# EVOLUTION OF BIG DATA

## Introduction (contd..)

Big data --- big business tool --- past  
Ambiguous to many people--- since it's inception

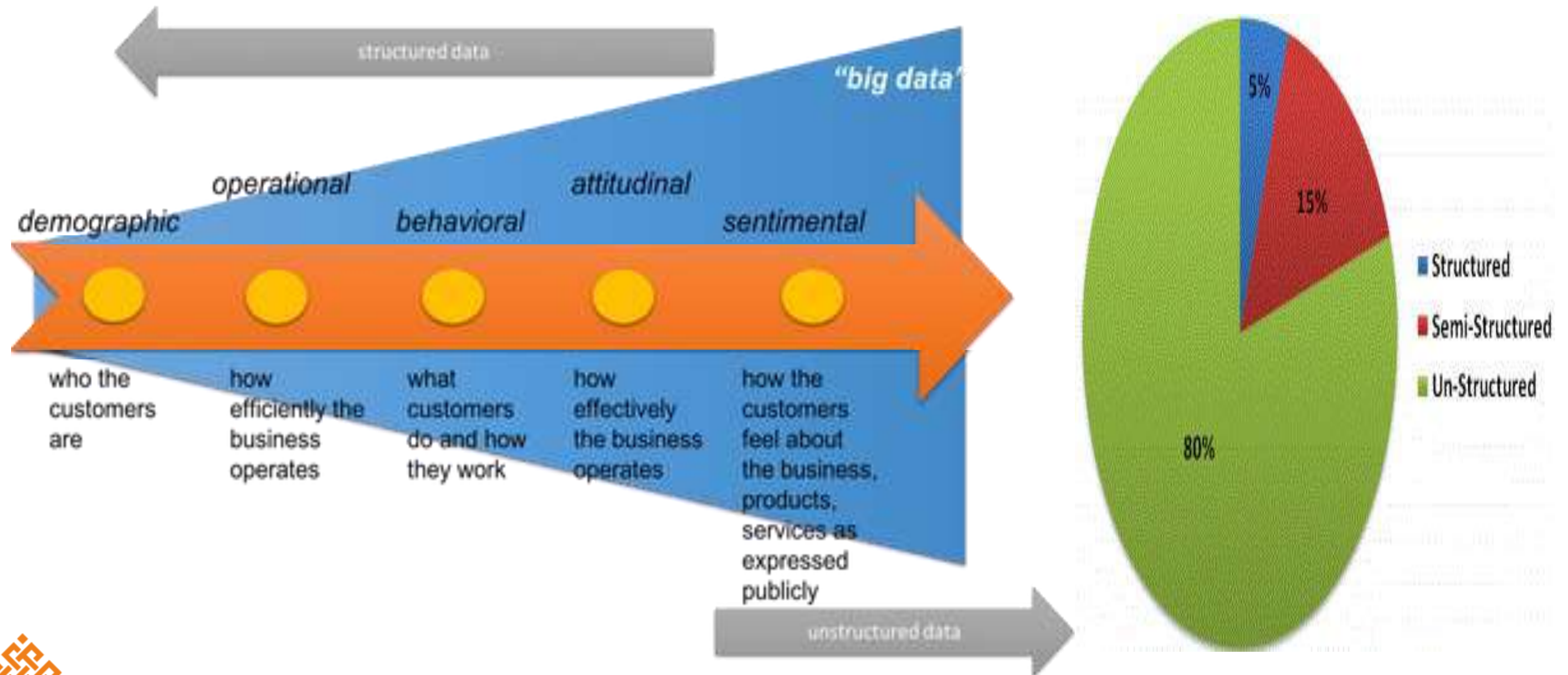
Increasingly clear about what and why big data is important --- to so many different companies.

In 1998, **John Mashey**, who was Chief Scientist at SGI presented a paper titled “Big Data... and the Next Wave of Infrastrass.” at a USENIX meeting.



# EVOLUTION OF BIG DATA

## Introduction (contd..)



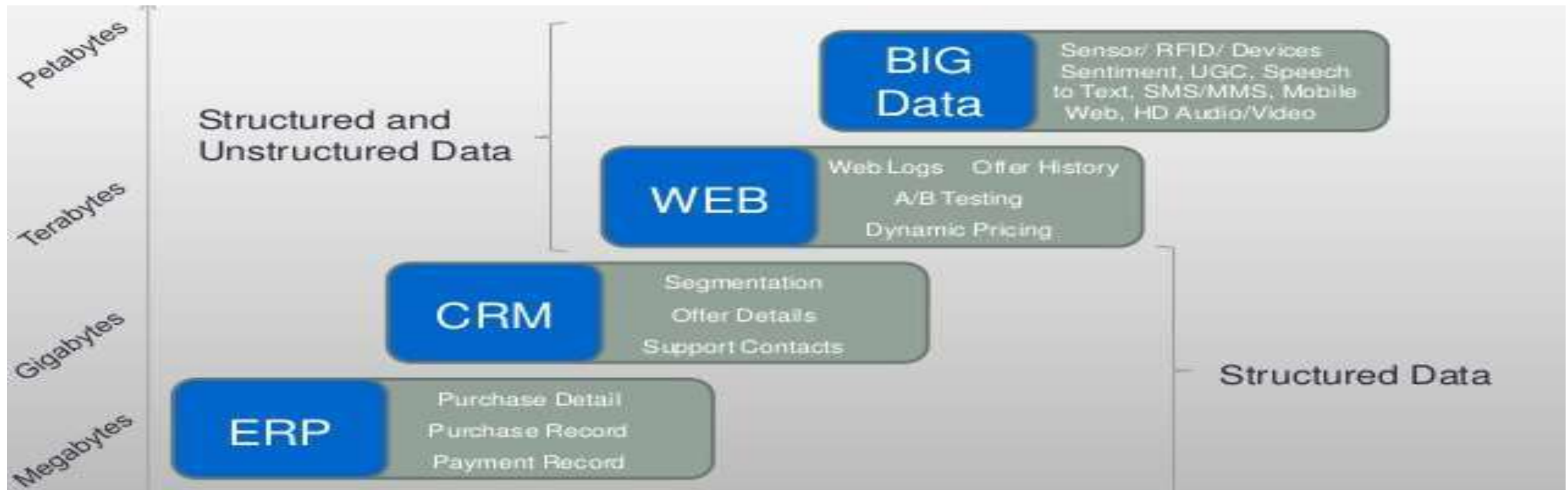


# EVOLUTION OF BIG DATA

## Introduction (contd..)

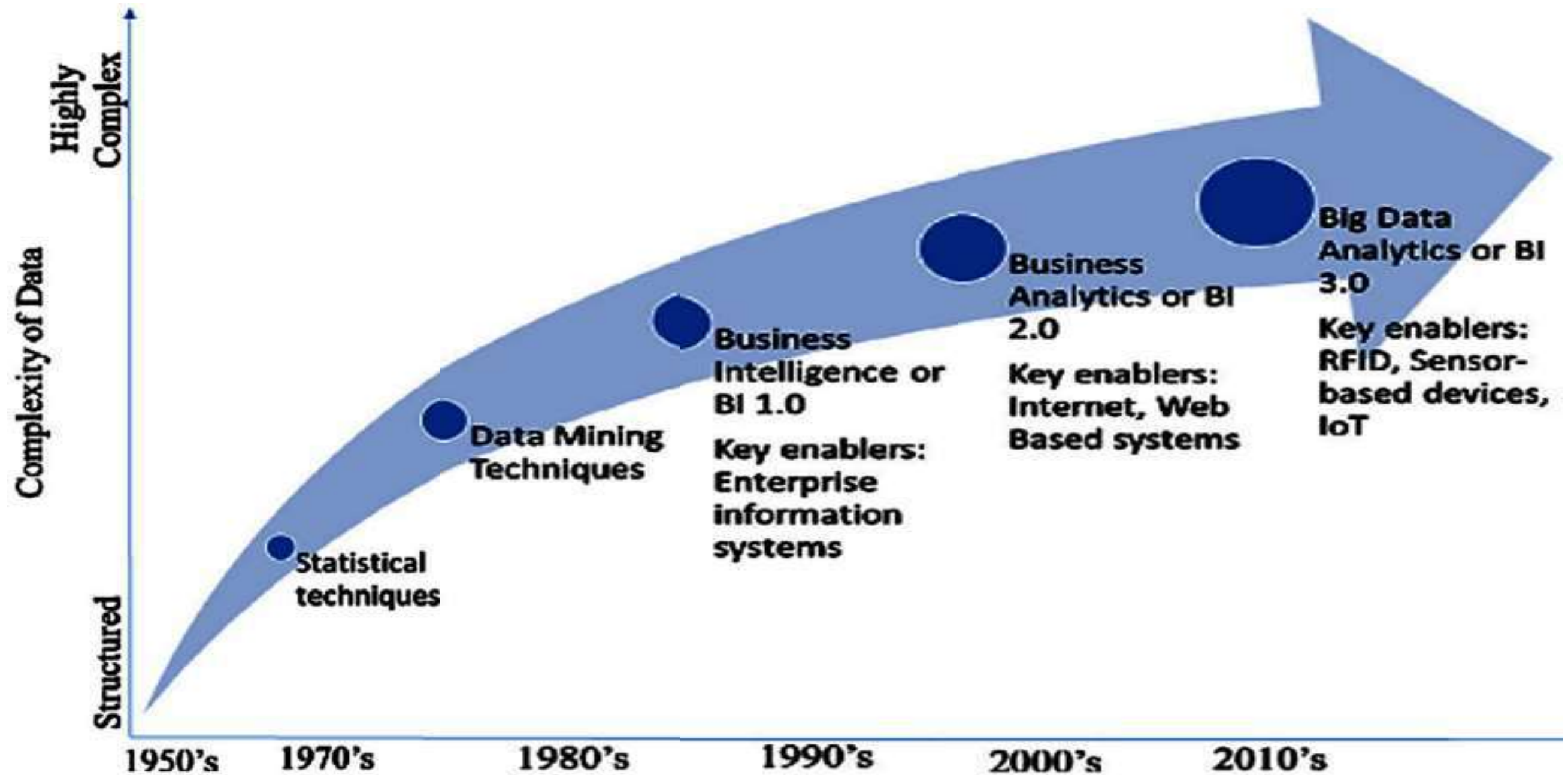
### Data Evolution .....

10% are structured and 90% are unstructured like emails, videos, facebook posts, website clicks etc.



# EVOLUTION OF BIG DATA

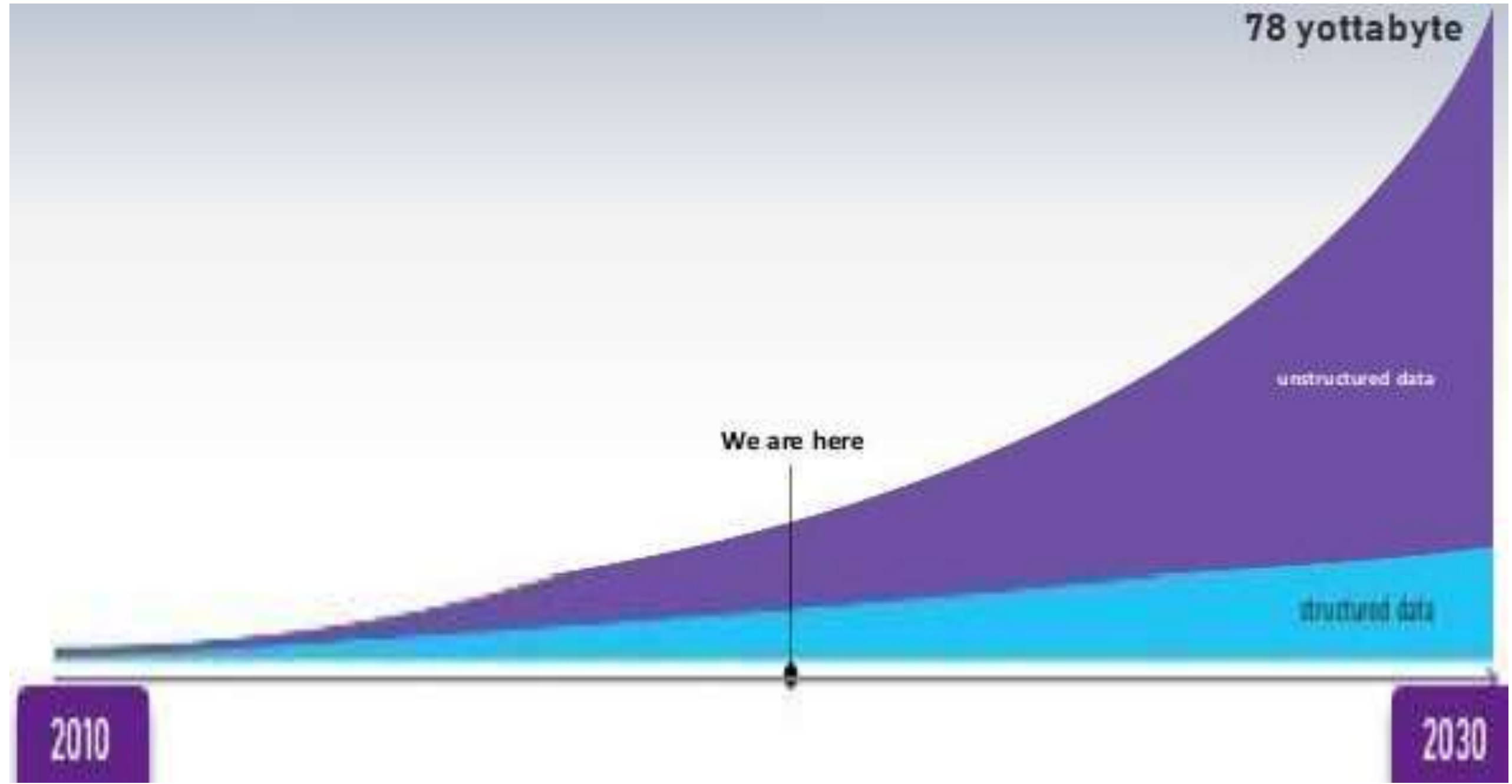
## Introduction (contd..)





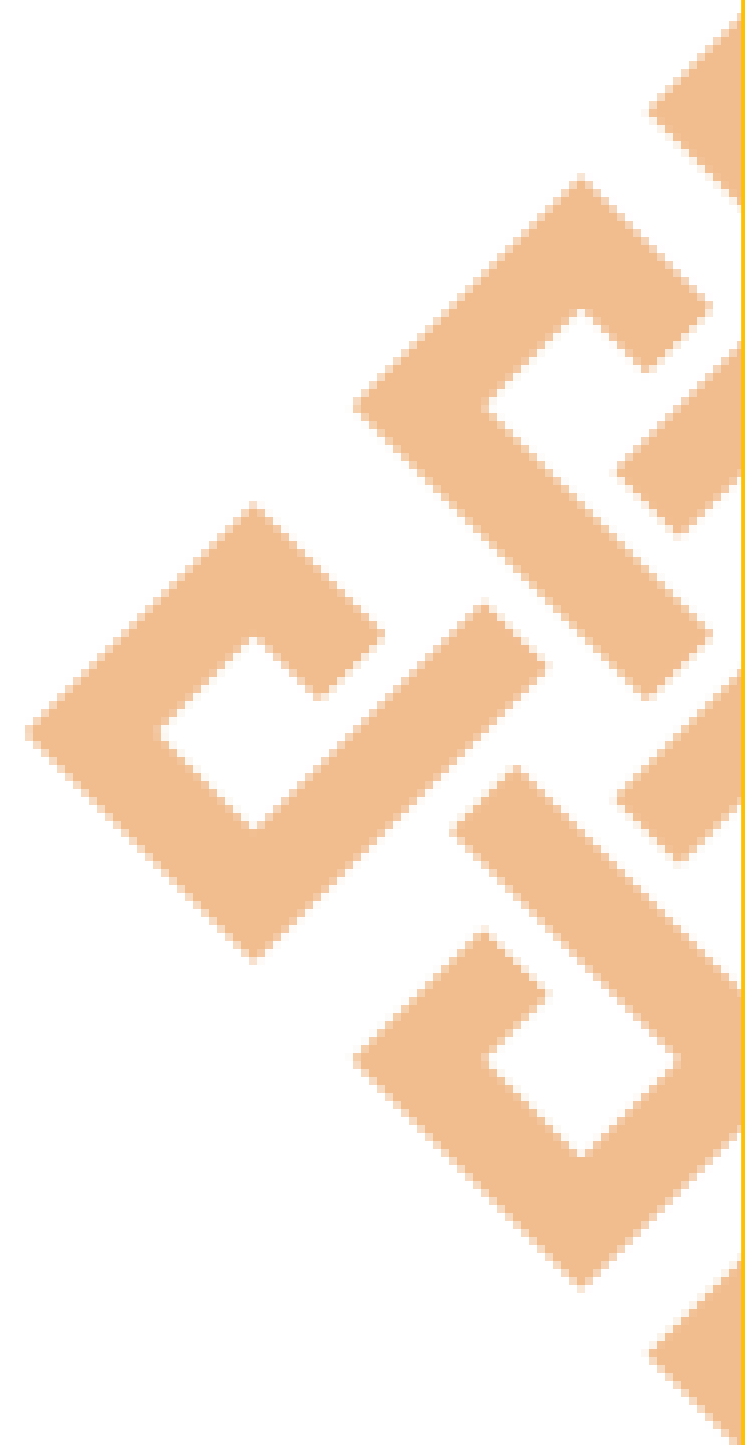
# EVOLUTION OF BIG DATA

## Introduction (contd..)



# Evolution of Big Data

## Big Data Use Cases



# EVOLUTION OF BIG DATA

## Big Data Use Cases

### Banking Sector:



# EVOLUTION OF BIG DATA

## Big Data Use Cases (contd..)

### Health Care Sector:





# EVOLUTION OF BIG DATA

## Big Data Use Cases (contd..)

### Retail Sector:



# EVOLUTION OF BIG DATA Technologies

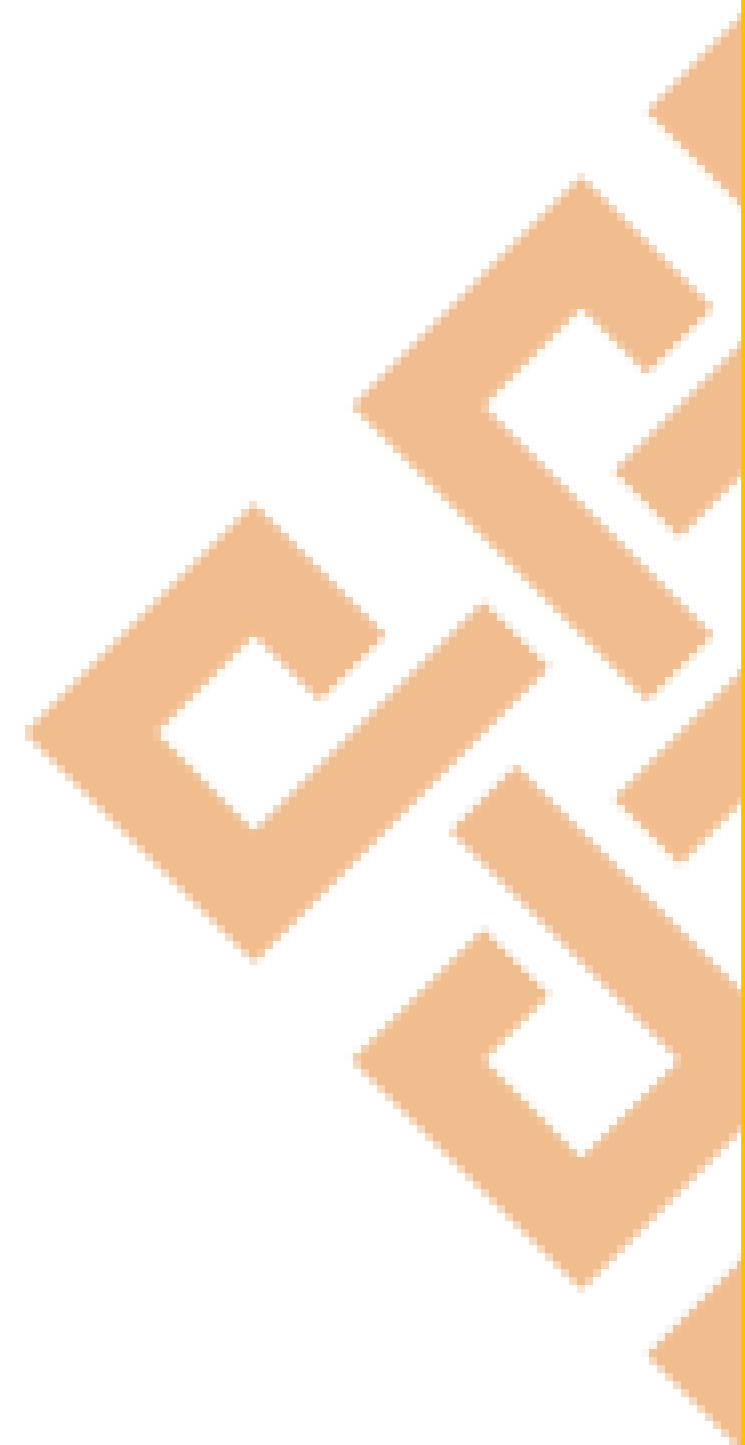


Apache Flink



# Evolution of Big Data

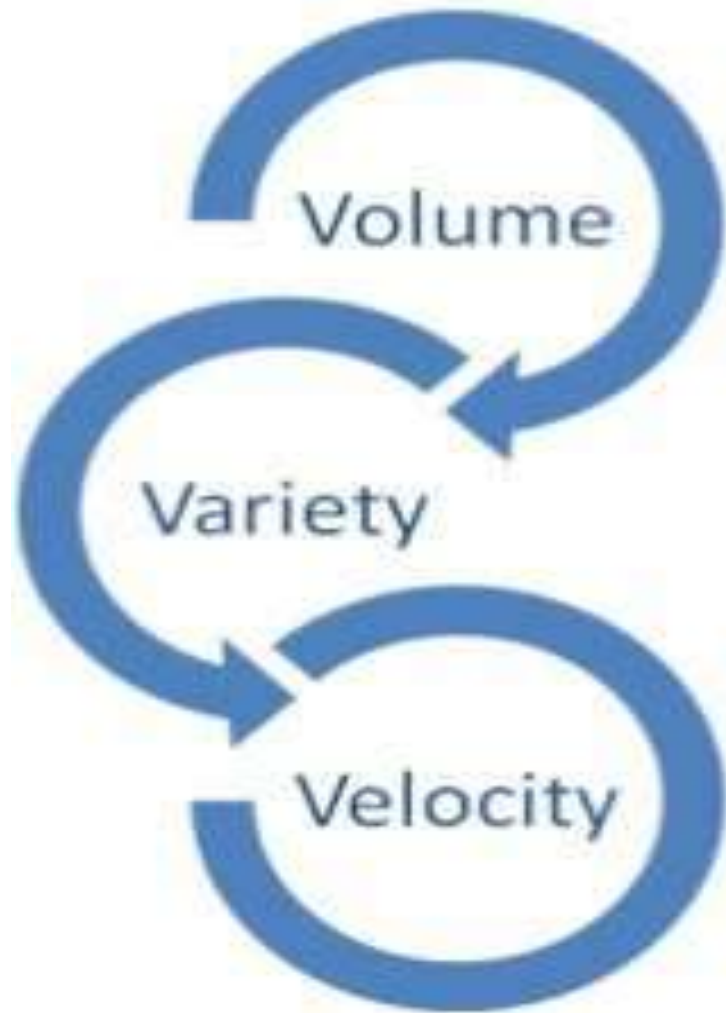
## Definition of Big Data





# EVOLUTION OF BIG DATA

## Definition of Big Data



**2001:** Doug Laney first uses “Volume, Velocity & Variety” to describe Big Data<sup>2</sup>

**2012:** Gartner updates the definition to:

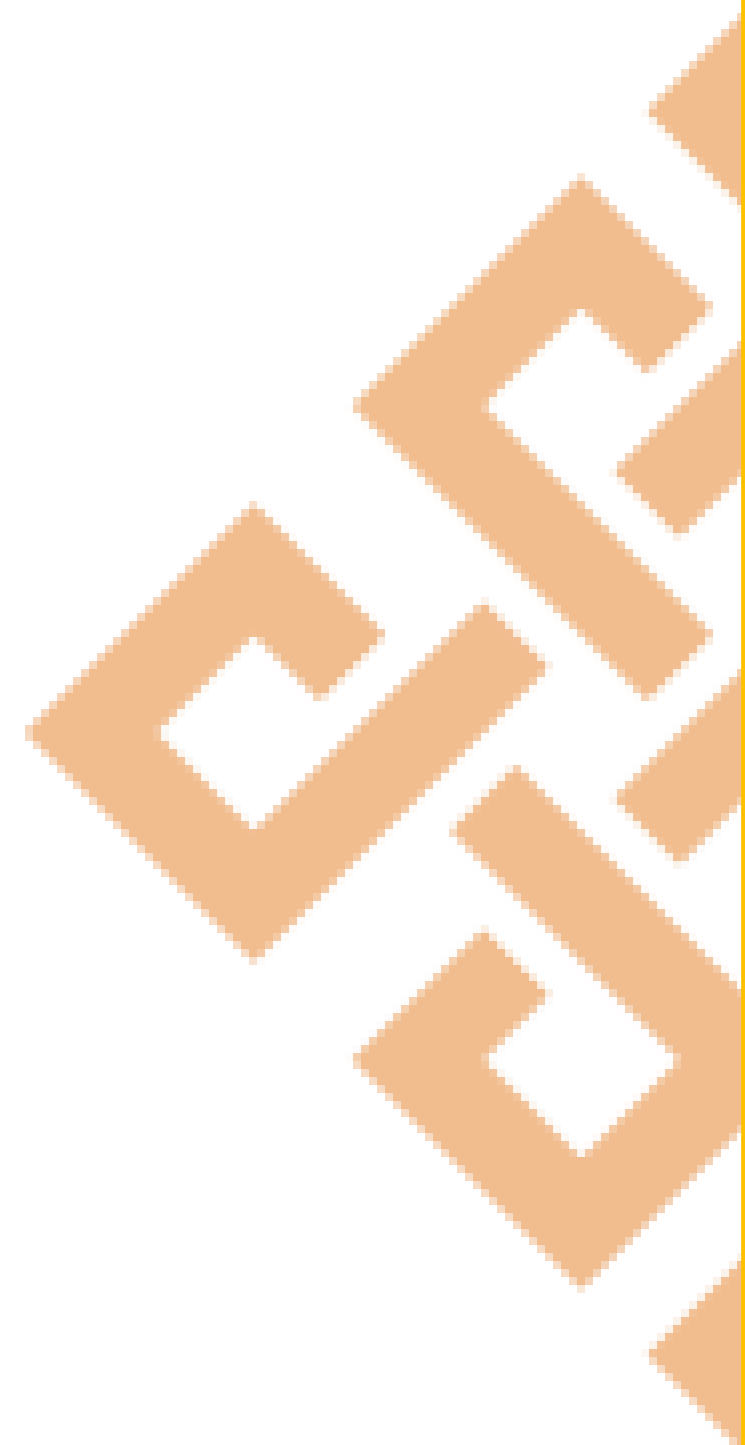
“Big data are high volume, high velocity and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process automation”





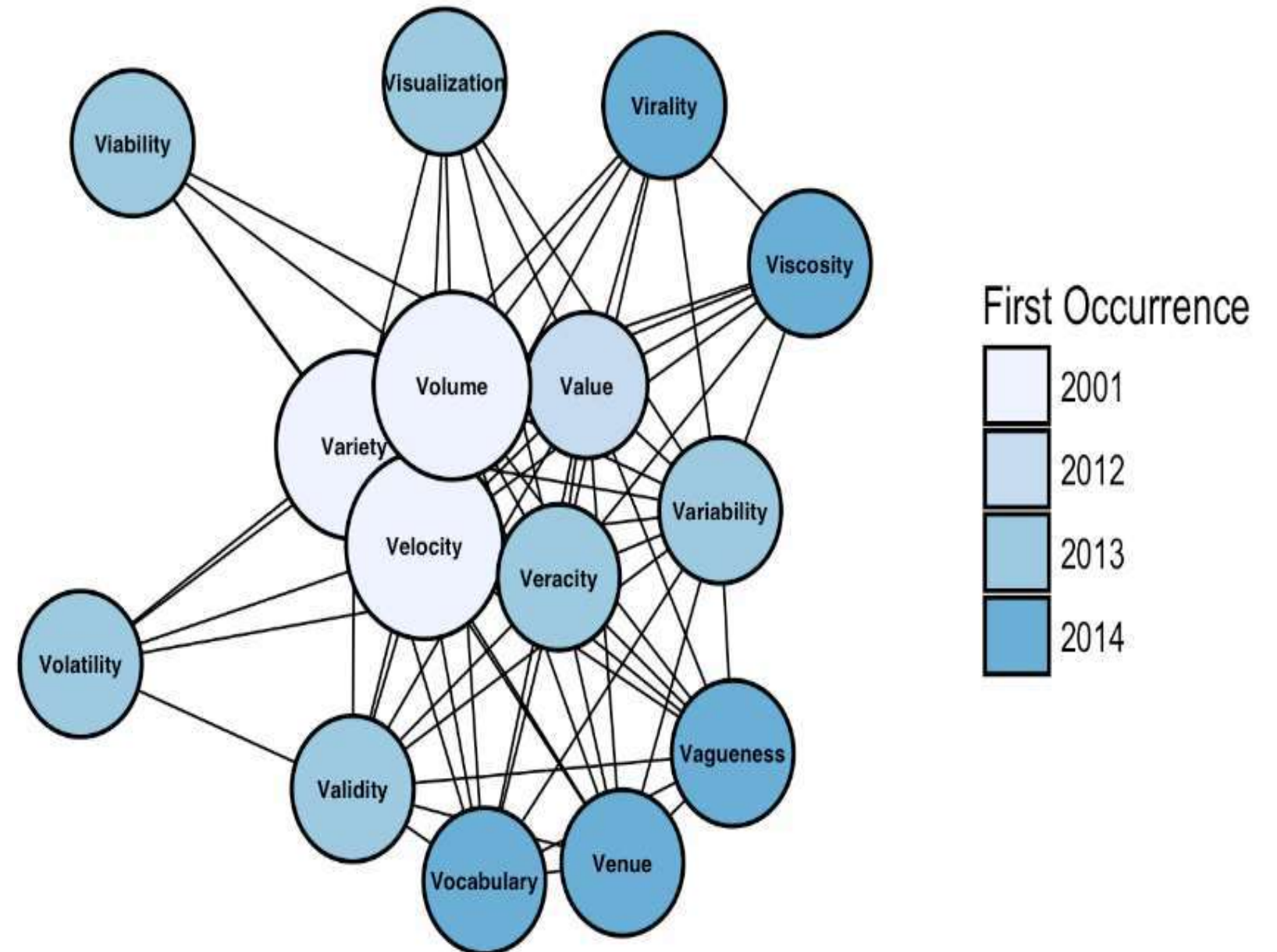
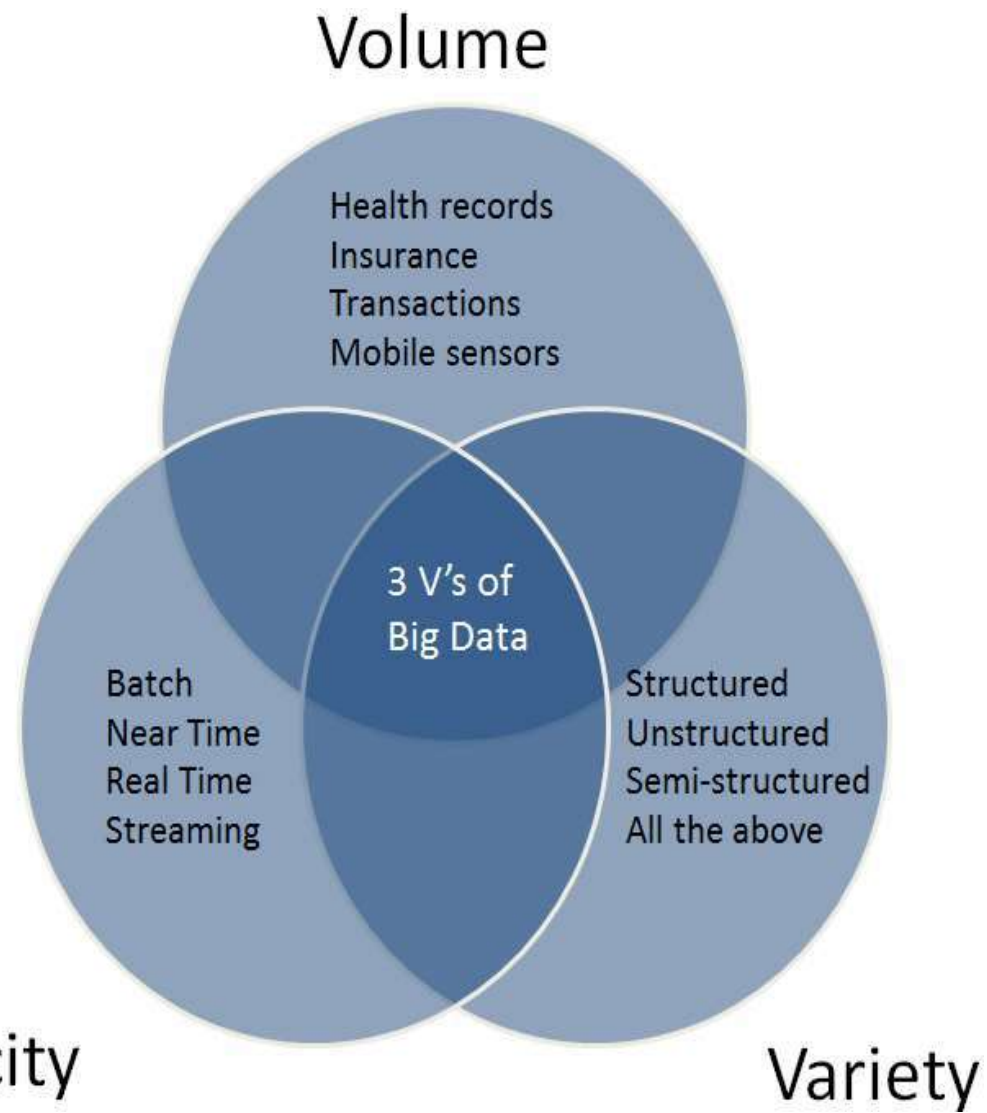
# Evolution of Big Data

**V's of Big Data**



# EVOLUTION OF BIG DATA

## V's of Big Data



# SUMMARY OF THE LECTURE

Introduction  
to Evolution  
of Big Data

Big Data  
Use Cases

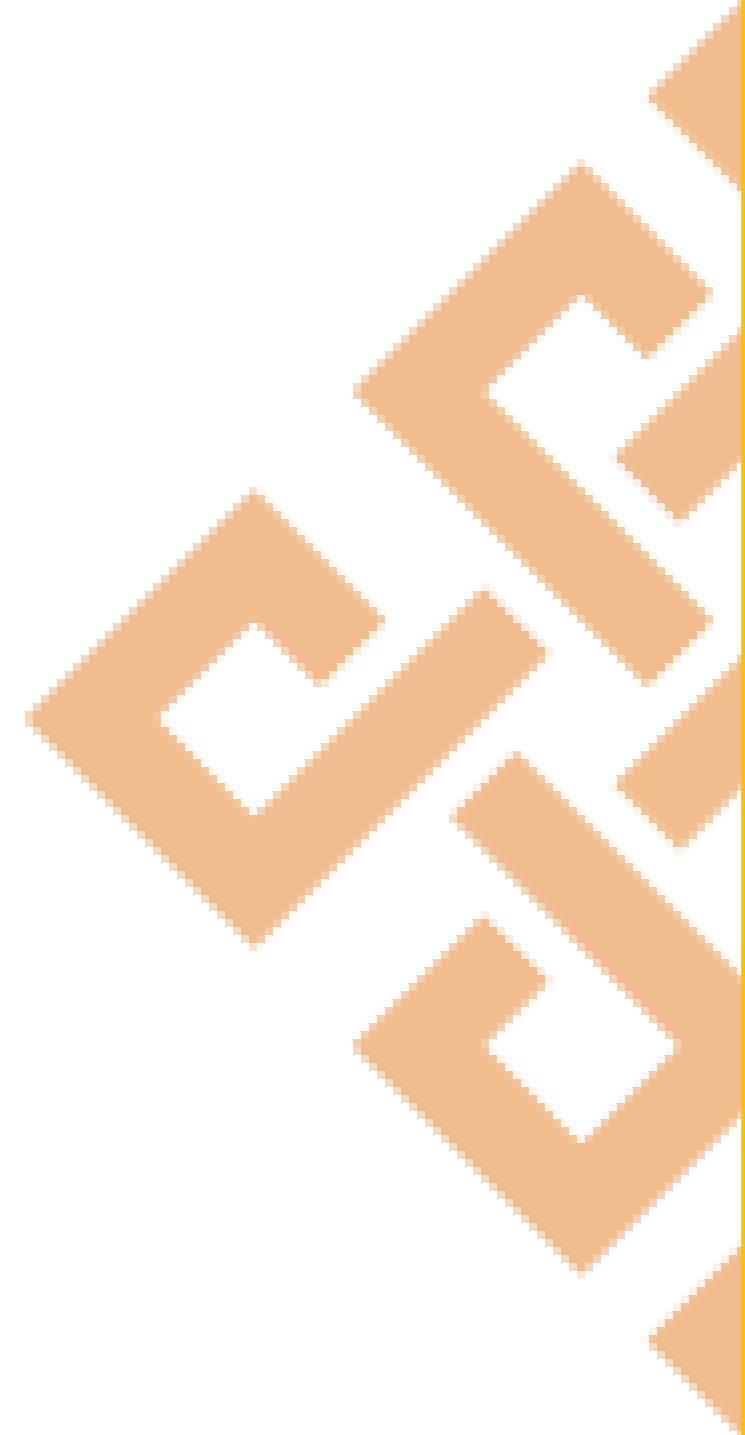
Technologies  
used for Big  
Data

Definition  
of Big Data



# **Evolution of Big Data**

**Resources and Tasks to be completed**



## Lecture 1.5

# Challenges with Big Data

School of Computer Science and Engineering

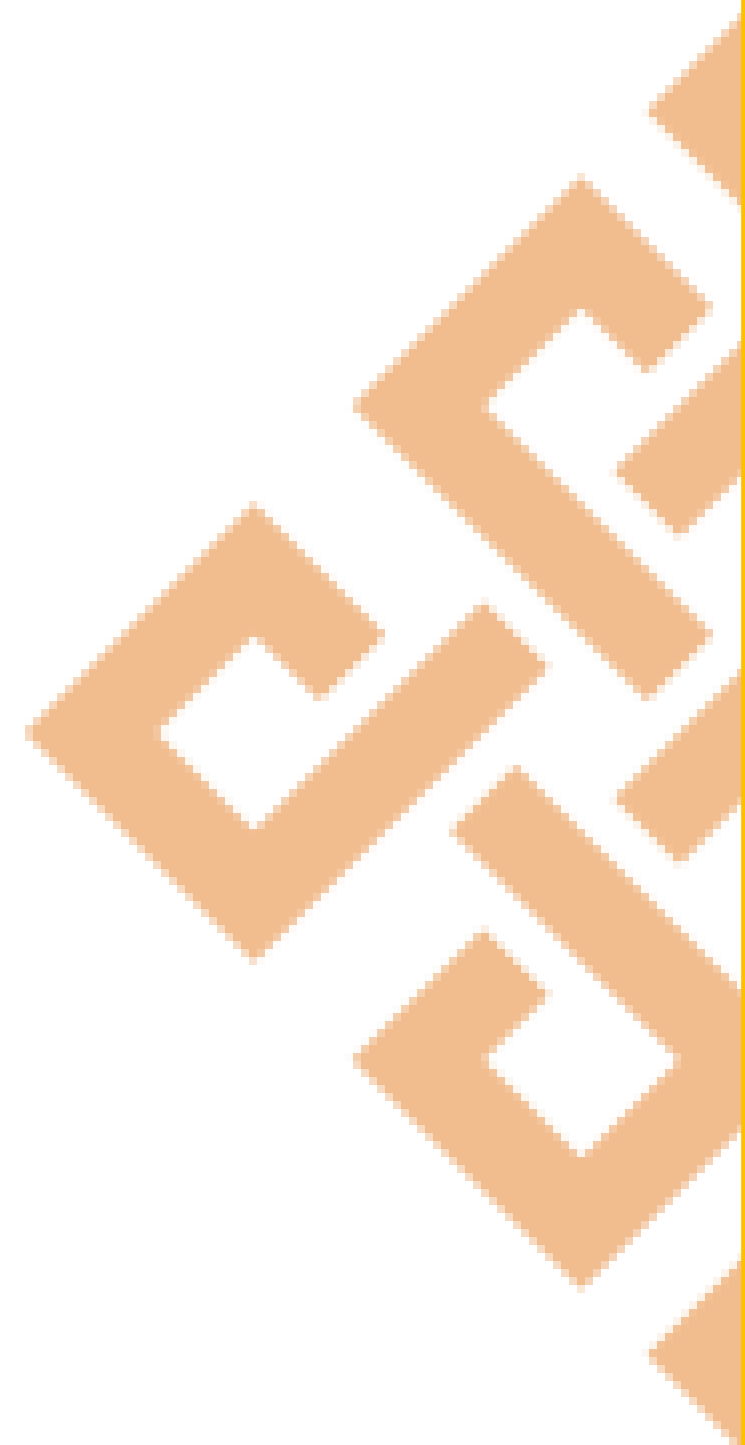
# Challenges with Big Data

**Recap of previous Lecture**



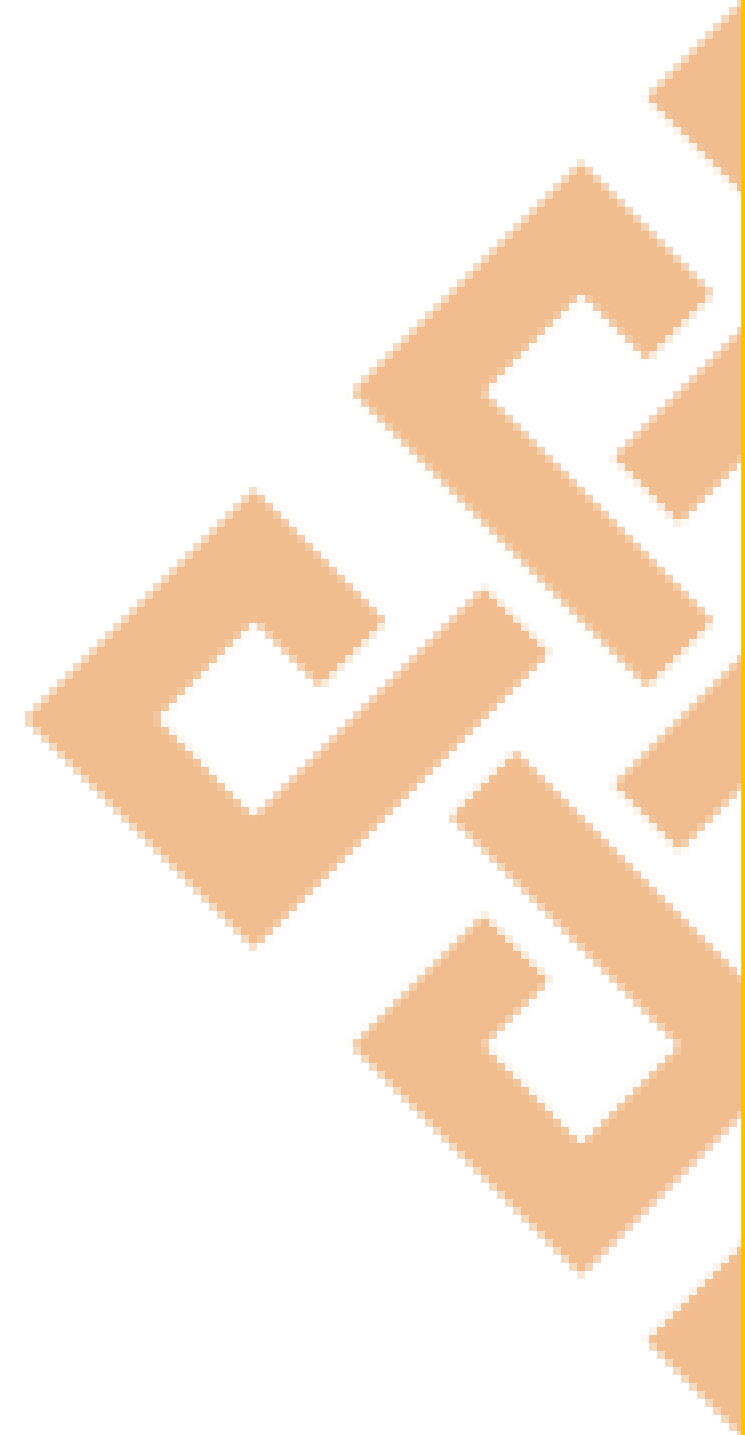
# Challenges with Big Data

**Topic of the Lecture**



# Challenges with Big Data

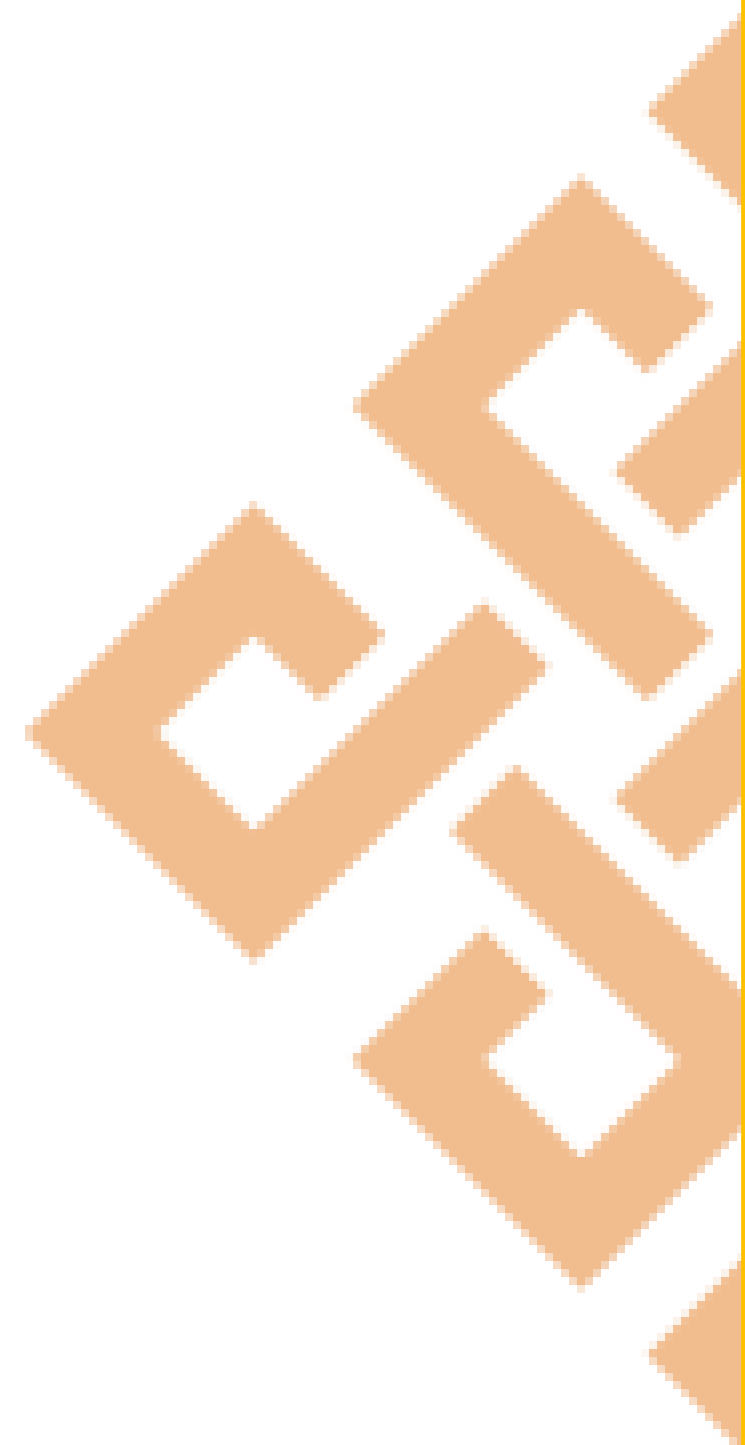
**Objective and Outcome of Lecture**





# Challenges with Big Data

## Introduction



# CHALLENGES WITH BIG DATA

## Introduction

The challenges involved with Big Data:

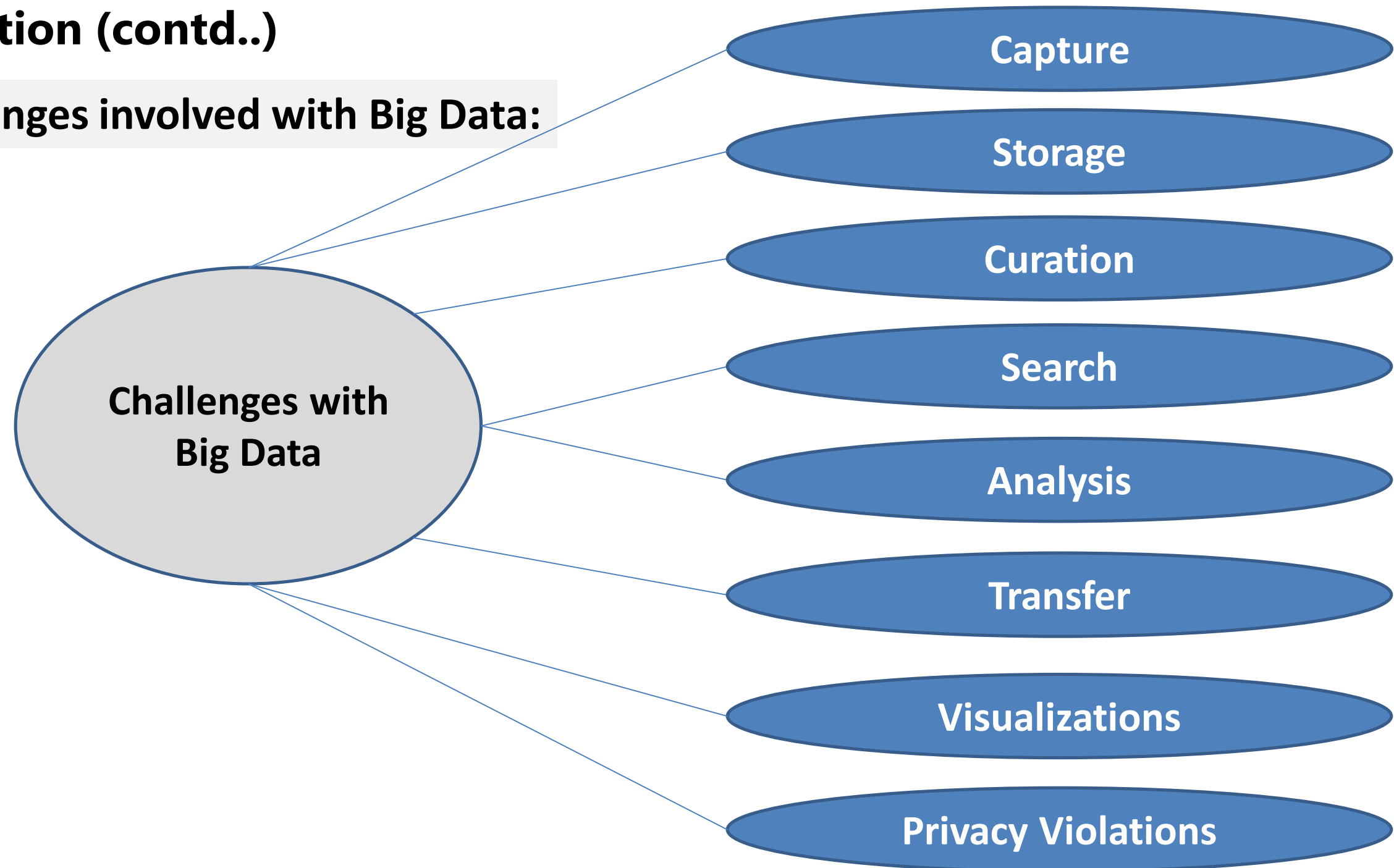
- 1) Picking the Right NoSQL Tools
- 2) Scaling up and down Big Data according to Current Demand
- 3) Overcoming Big Data Talent and Resource Constraints
- 4) Collecting and Integrating Massive and Diverse Datasets
- 5) Maintaining Data Integrity, Security, and Privacy



# CHALLENGES WITH BIG DATA

## Introduction (contd..)

The challenges involved with Big Data:



# CHALLENGES WITH BIG DATA

## Introduction (contd.) Summary

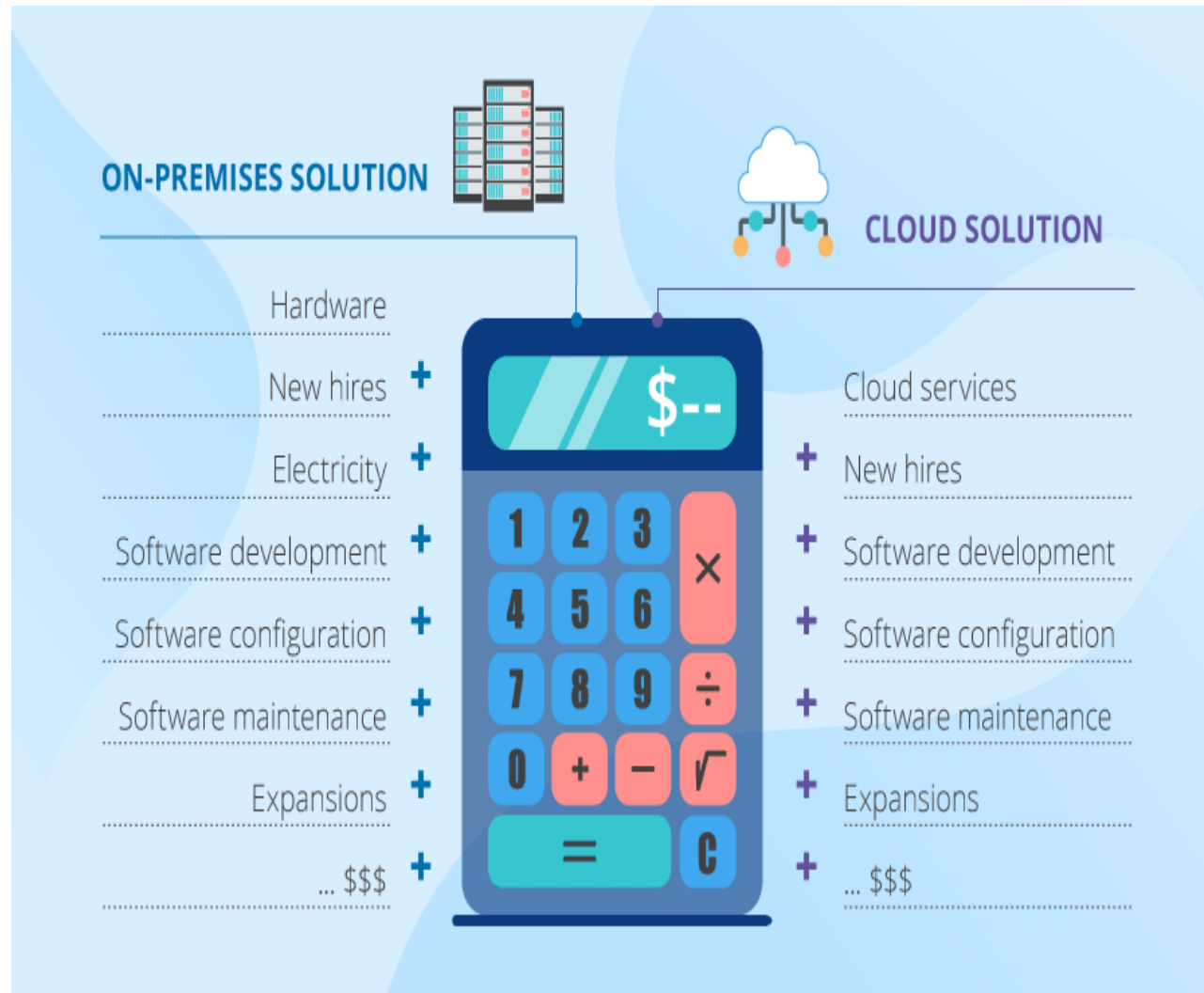
- Insufficient understanding and acceptance of big data
- Confusing variety of big data technologies
- Paying loads of money
- Complexity of managing data quality
- Dangerous big data security holes
- Tricky process of converting big data into valuable insights
- Troubles of upscaling



# CHALLENGES WITH BIG DATA

## Introduction (contd..)

### ➤ Paying loads of money



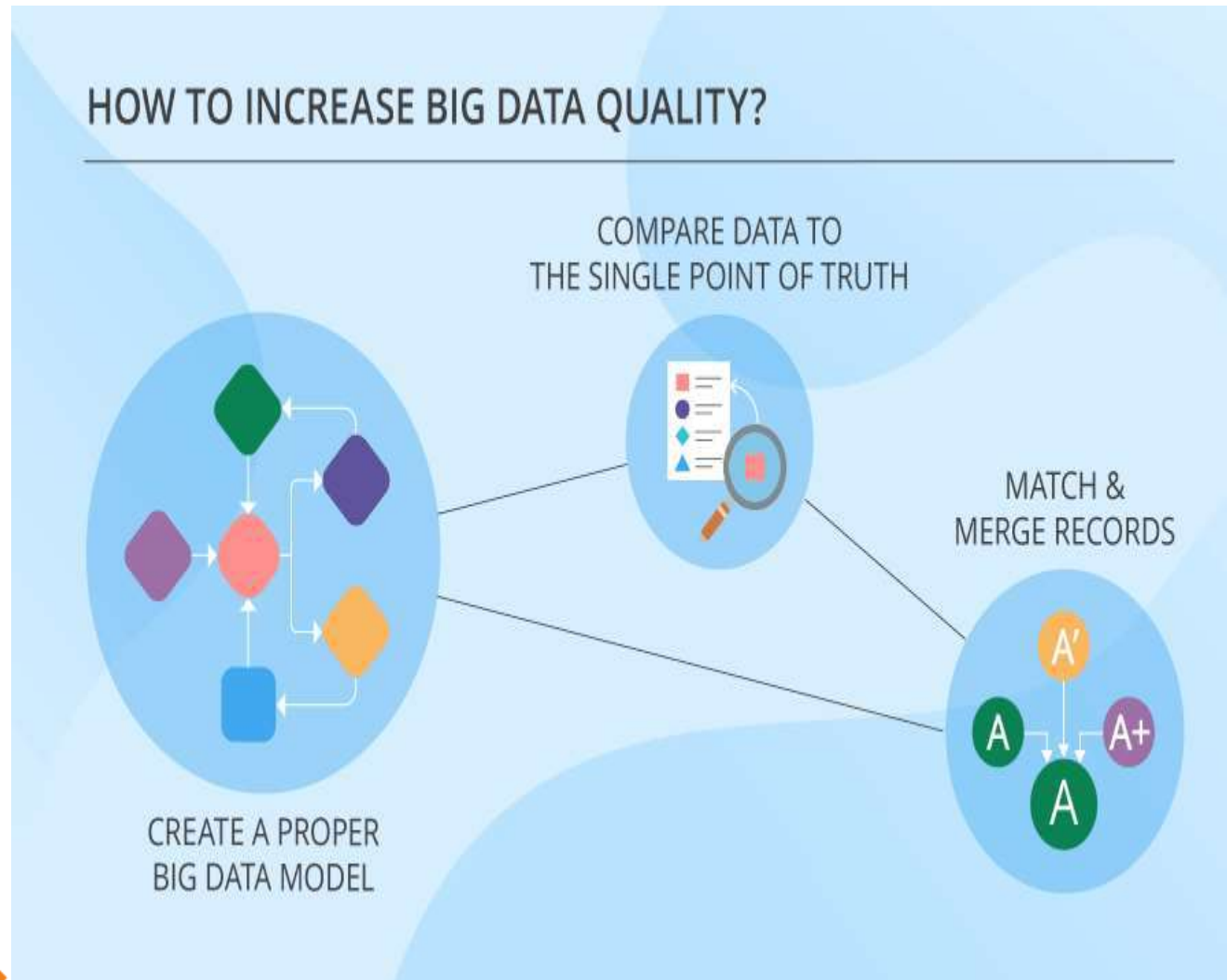
### ➤ Confusing variety of big data technologies



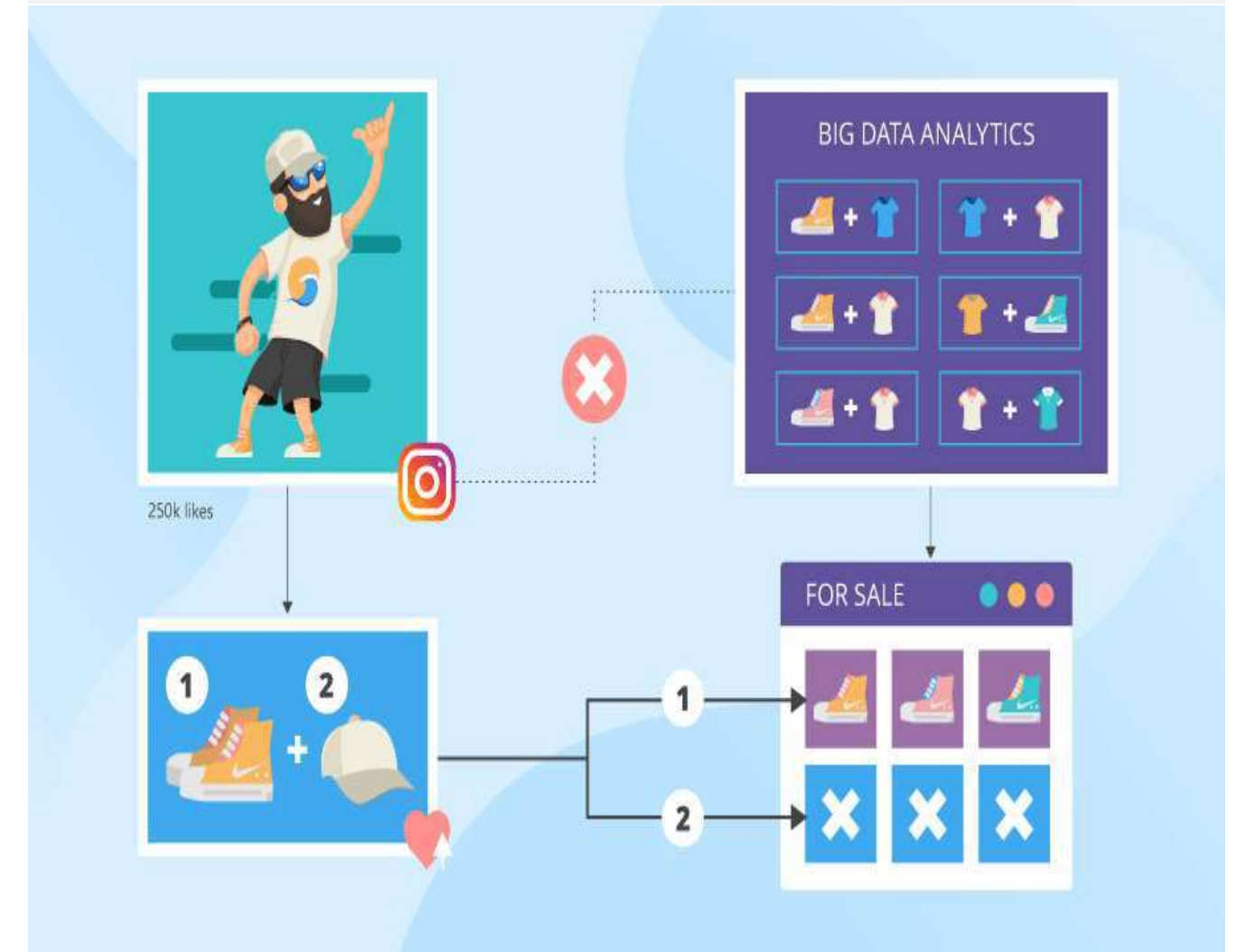
# CHALLENGES WITH BIG DATA

## Introduction (contd..)

- Complexity of managing data quality



- Tricky process of converting big data into valuable insights





# CHALLENGES WITH BIG DATA

What is Big Data?

What happens  
**online**  
in 60 seconds?



## 2020 This Is What Happens In An Internet Minute





# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

### What is Data?

- the quantities, characters, or symbols, ...
- operations are performed by a computer on them
- may be stored and transmitted in the form of electrical signals
- recorded on magnetic, optical, or mechanical recording media.

### Then, What is Big Data?

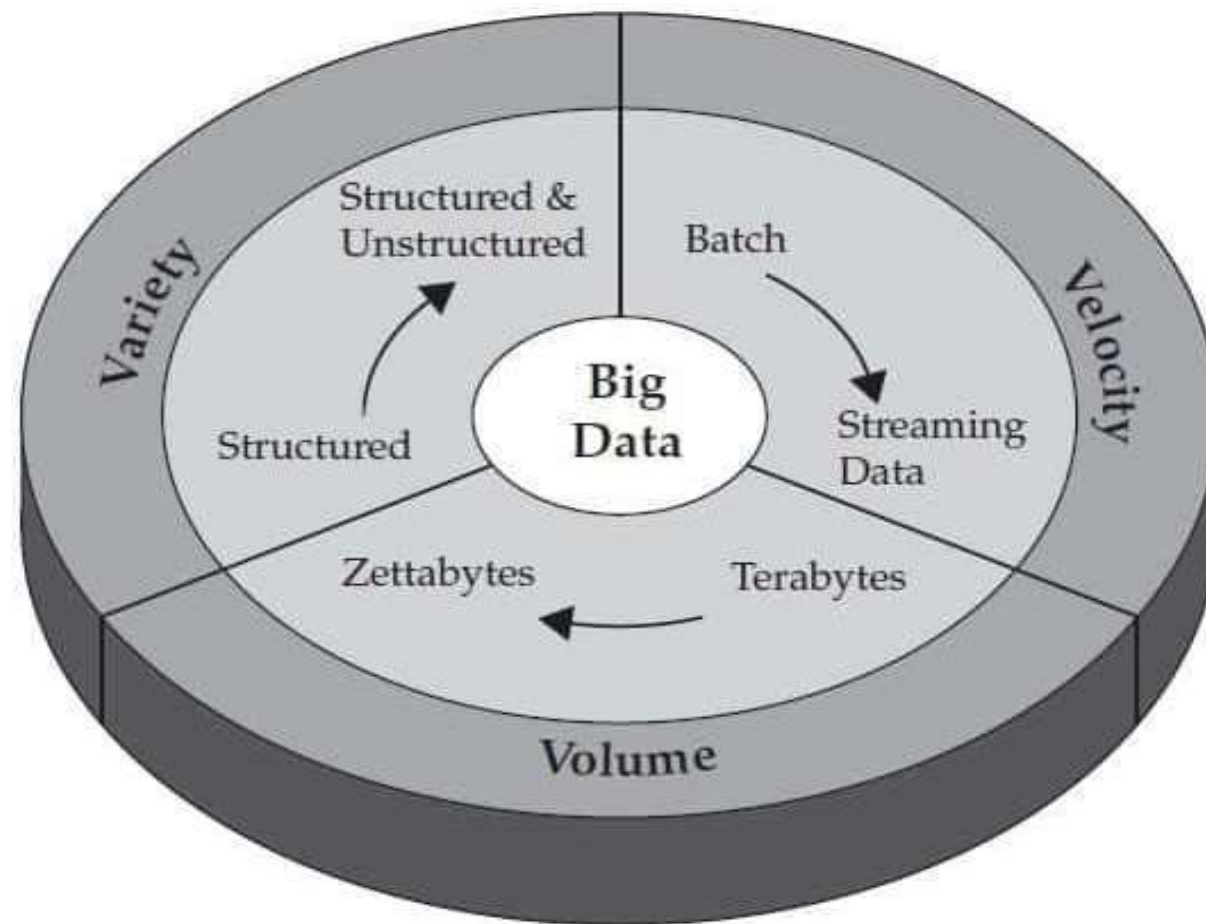
- **Data with a huge size**
- A term used to describe a collection of huge volume of data --- yet growing exponentially with time
- **“In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently”.**



# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

### What?? Definition?



Large amounts of data



collected passively from digital interactions



with great variety and a high rate of velocity.

- Gartner's definition:
- “Big data is data that contains greater variety arriving in increasing volumes and with ever-higher velocity”. --- 3Vs.
- Larger, more complex, voluminous data sets from new data sources

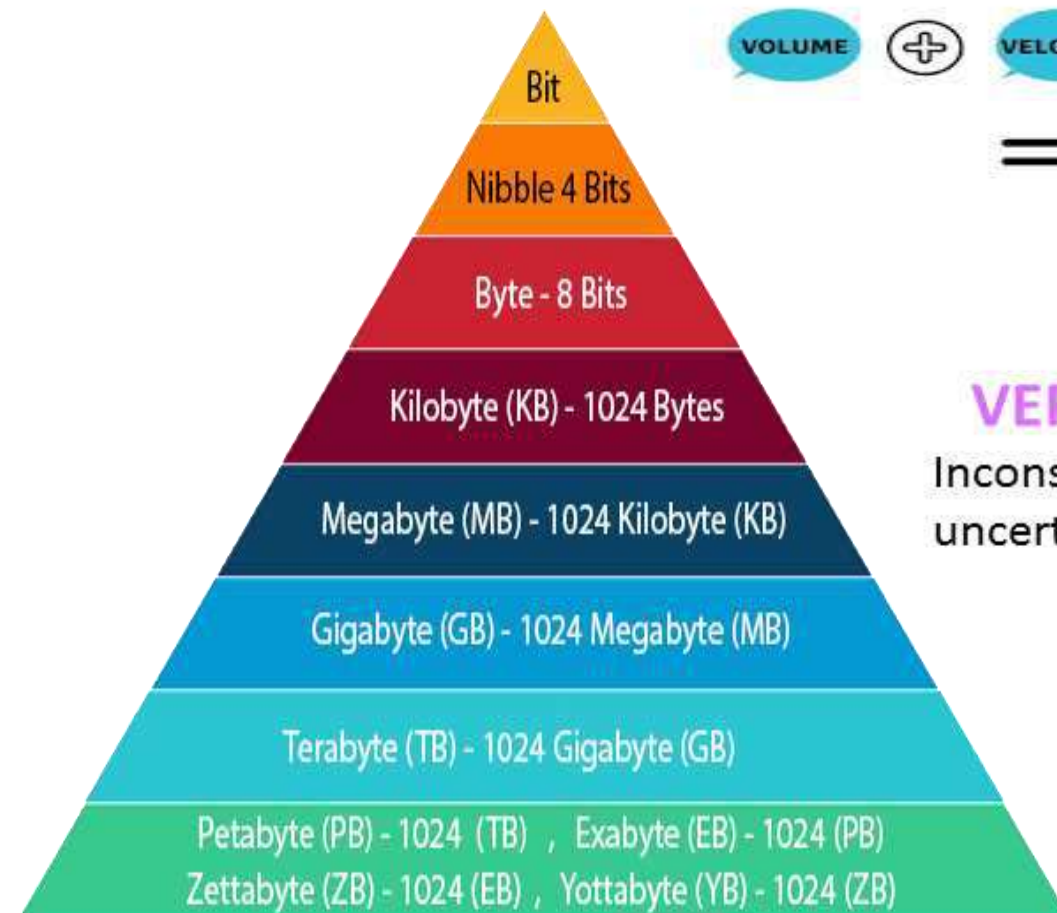


- “But it's not the amount of data that's important”.
- Can be analyzed for insights --- better decisions and strategic business moves

# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

Lets see Big Data with V's



**VERACITY**  
Inconsistencies and  
uncertainty in data

**VELOCITY**  
High speed of  
accumulation of data

**VOLUME**  
Huge amount of data



**VARIETY**  
Different formats of data  
from various sources



**VALUE**  
Extract useful data

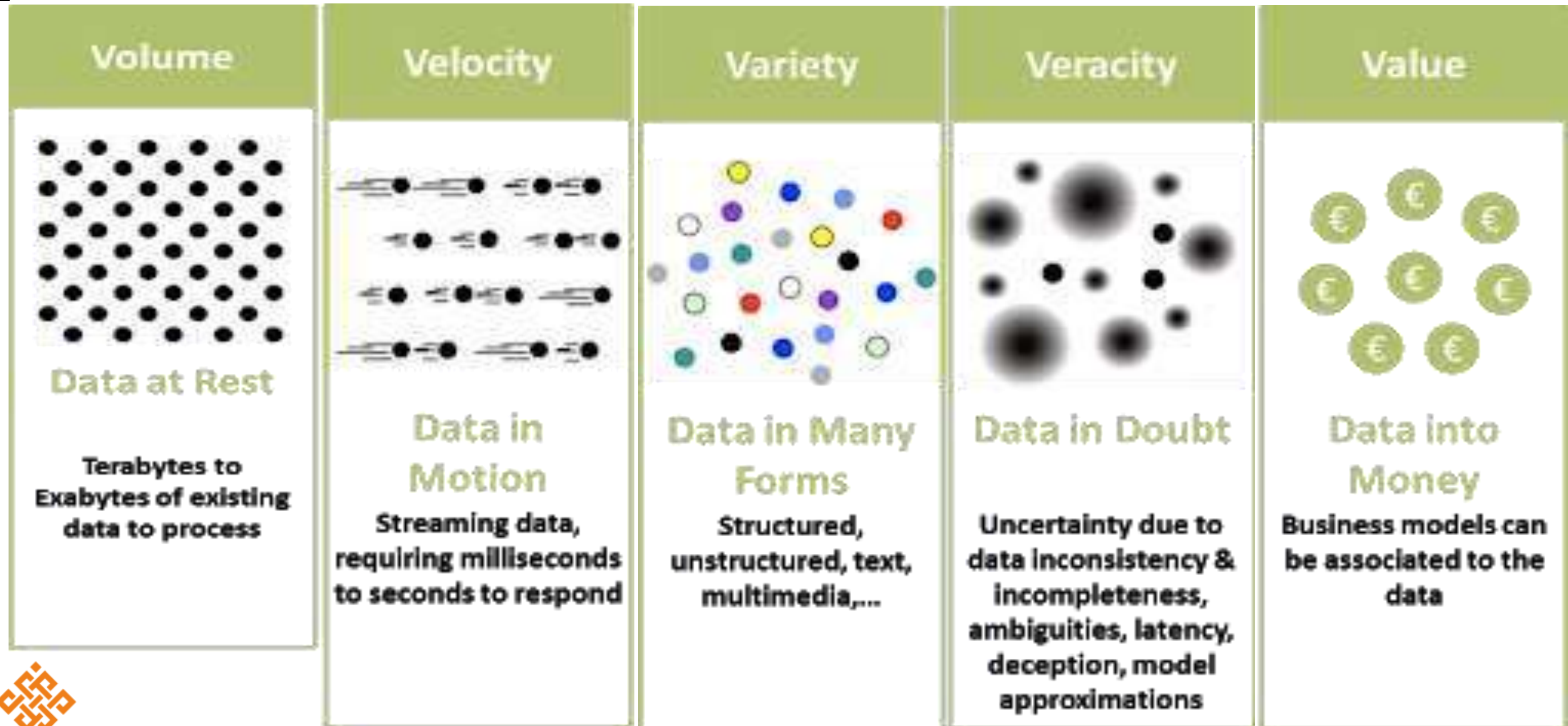




# CHALLENGES WITH BIG DATA

More on Big Data V's

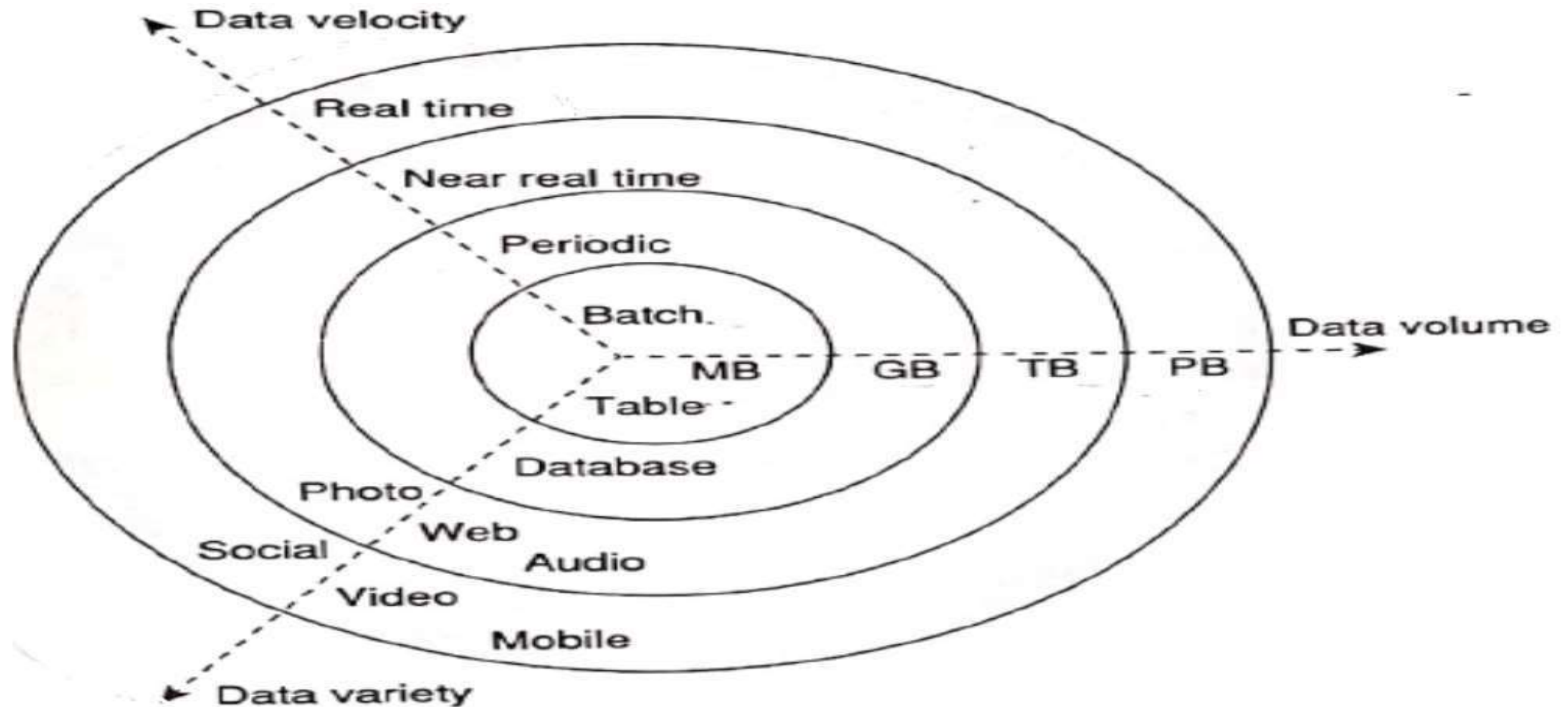
## What is Big Data? (contd..)



# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

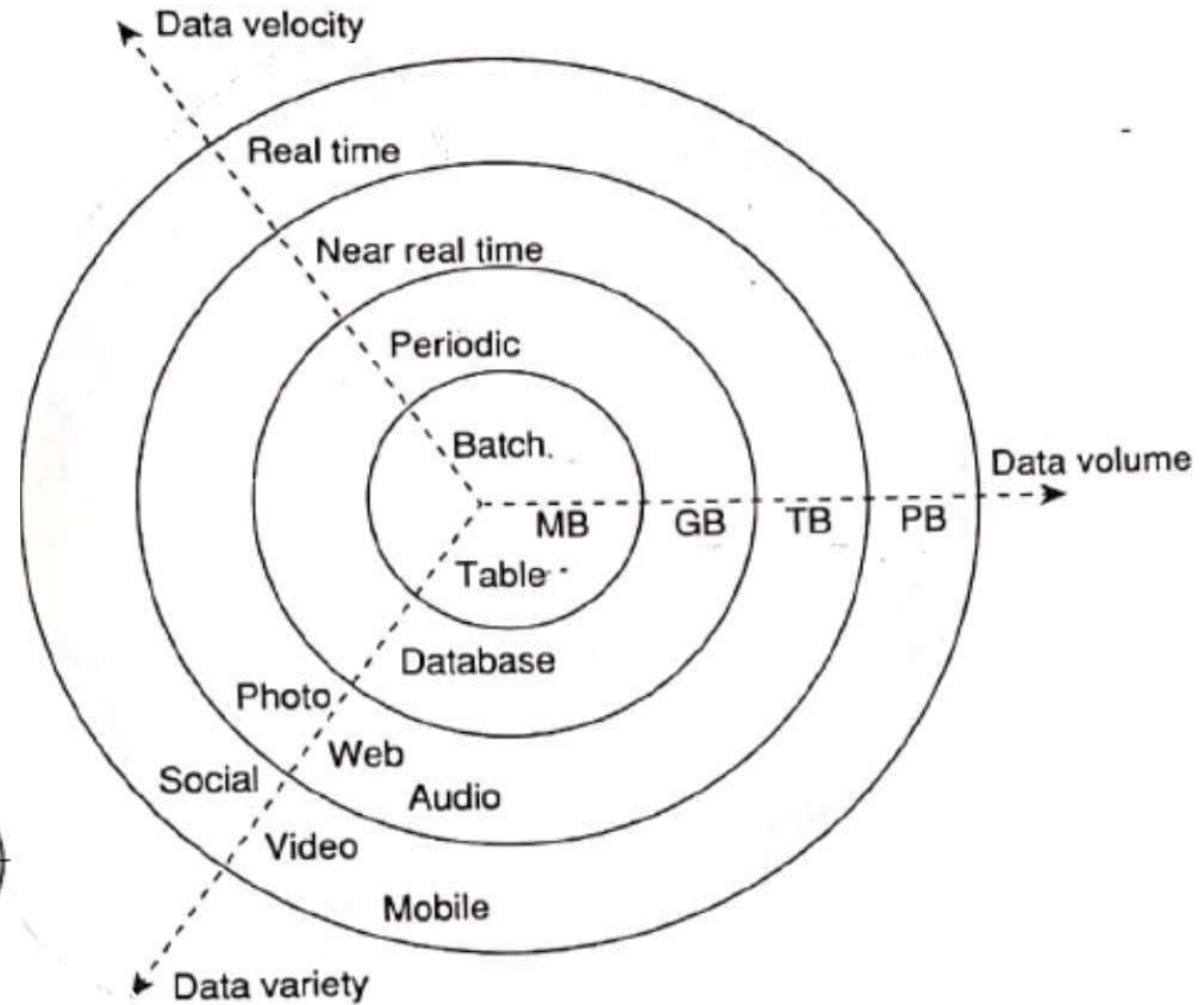
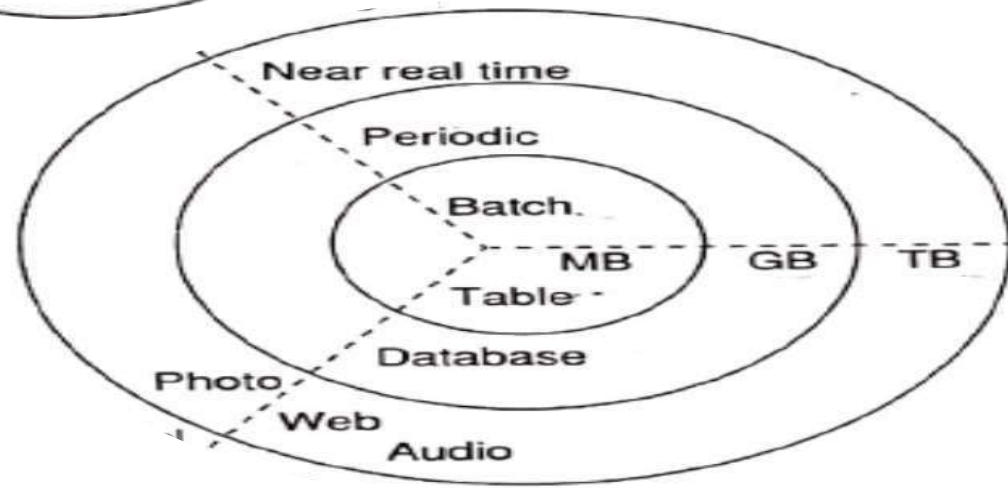
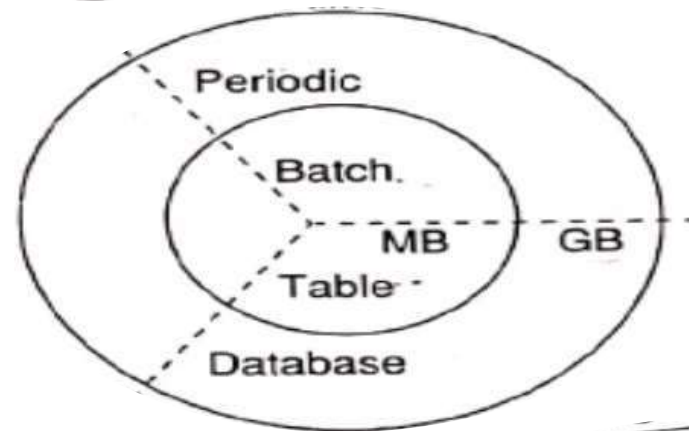
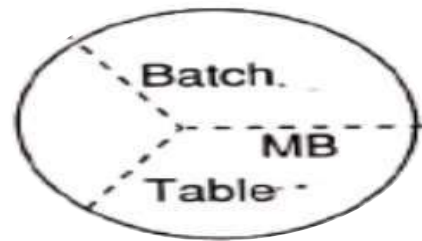
Comprehensive view



# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

Velocity



Batch → Periodic → Near real time → Real-time processing





# CHALLENGES WITH BIG DATA

## What is Big Data? (contd..)

Volume

Data Source

Internal



### Structured



#### Human-Generated

- Survey ratings
- Aptitude testing

#### Machine-Generated

- Web metrics from Web logs
- Product purchase from sales Records
- Process control measures

### Unstructured



#### Human-Generated

- Emails, letters, text messages
- Audio transcripts
- Customer comments
- Voicemails
- Corporate video/communications
- Pictures, illustrations
- Employee reviews

External



#### Human-Generated

- Number of Retweets, Facebook likes, Google Plus +1s
- Ratings on Yelp
- Patient ratings ratings

#### Machine-Generated

- GPS for tweets
- Time of tweet/updates/postings

#### Human-Generated

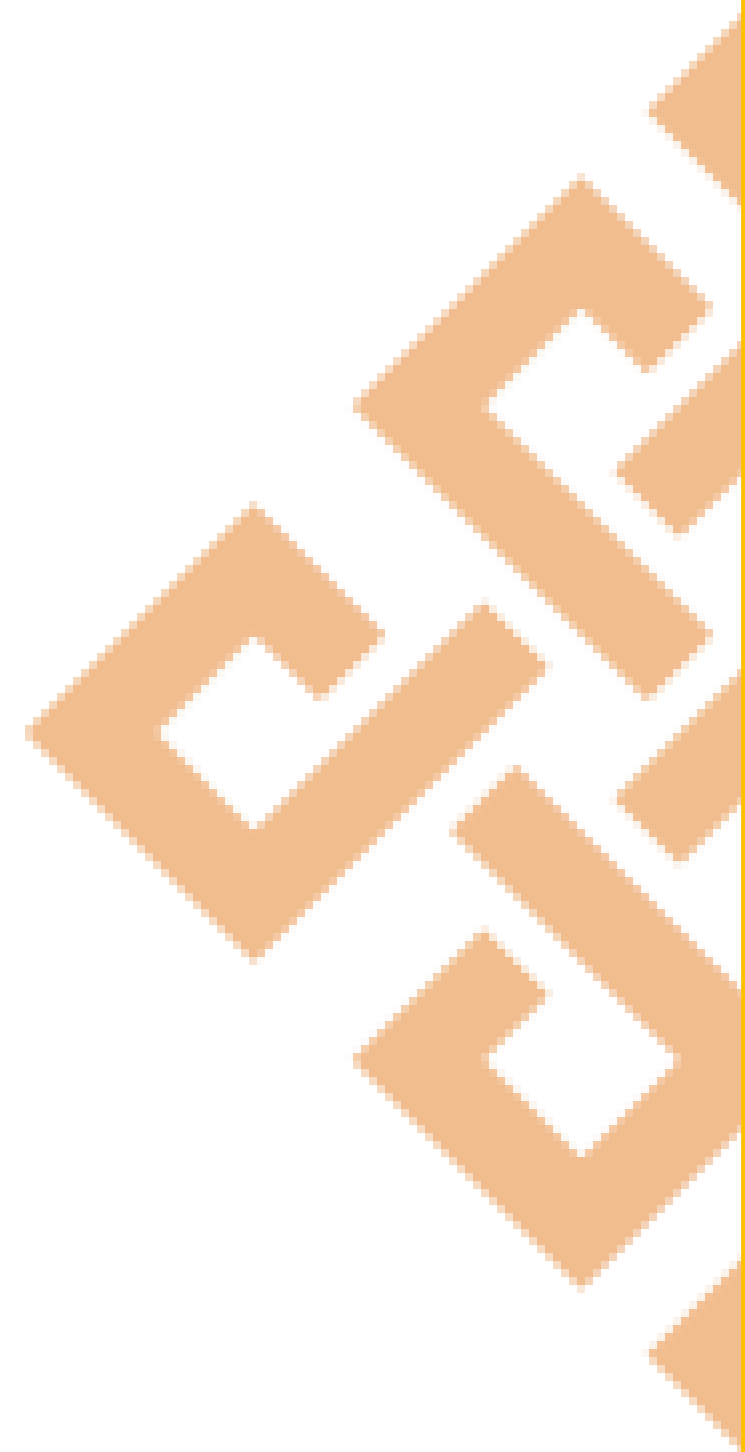
- Content of social media updates
- Comments in online forums
- Comments on Yelp
- Video reviews
- Pinterest images
- Surveillance video





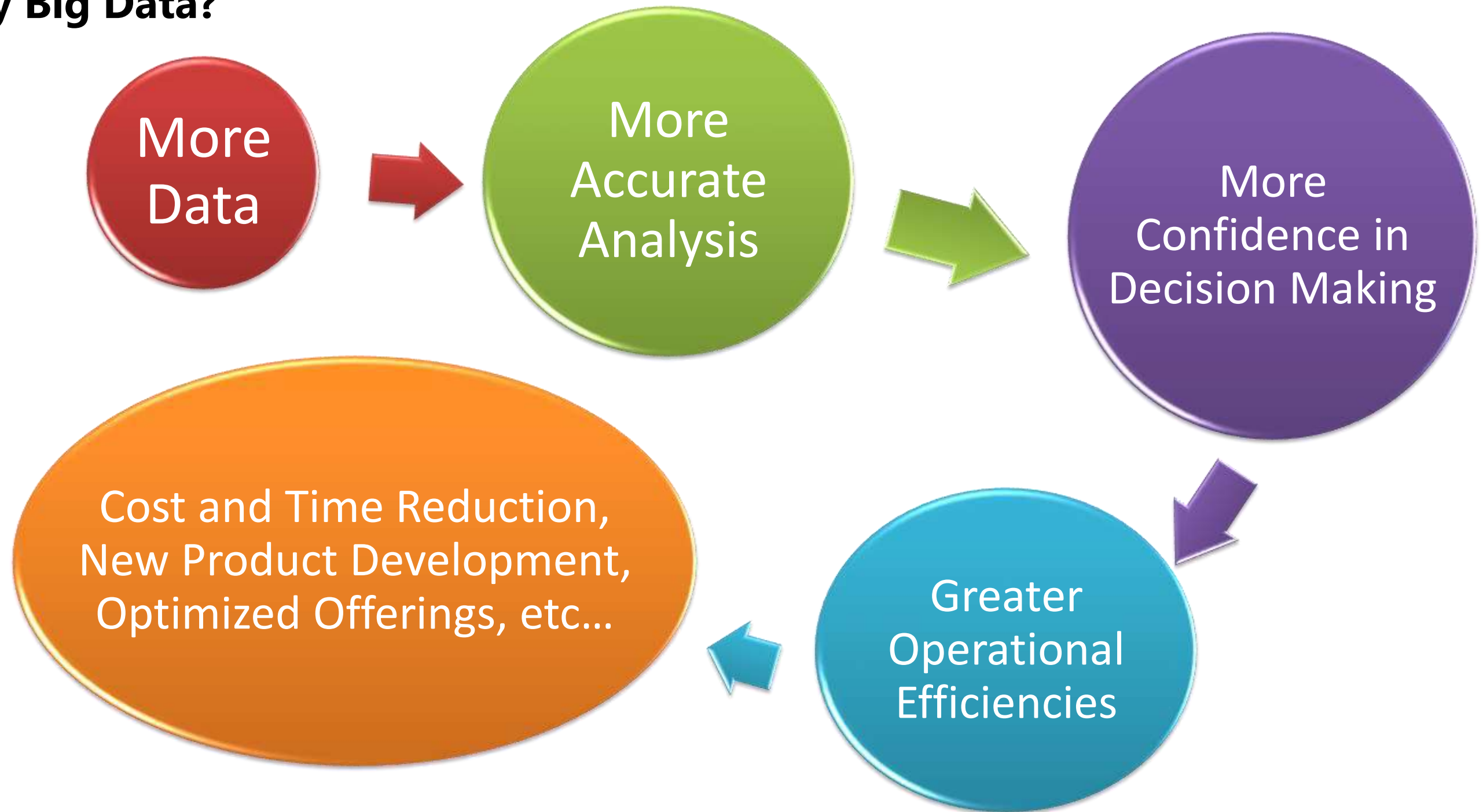
# Challenges with Big Data

**Why Big Data?**



# CHALLENGES WITH BIG DATA

## Why Big Data?



# CHALLENGES WITH BIG DATA

## Why is Big Data? (contd..)

### SALIENT FEATURES

TB's → PB's → EB's → ZB's → YB's →



**Data Volumes**



**Data Sources**



**Non-traditional  
data types**

- Unstructured data
- Blogs, Text, chats
- Images, Videos
- System Logs
- Weak relational schema

- Sensors
- RFID
- Devices
- Traditional applications
- Web Servers

**Big  
Insights**



**Technologies**

- Distributed Parallel Processing architectures
- Highly Scalable commodity hardware
- ACID free approach
- MapReduce-style programming models



**Smart  
Business  
Queries**

- Which region should I increase my marketing /sales efforts in?
- Who are my top paying customers?
- How to increase my customer loyalty?



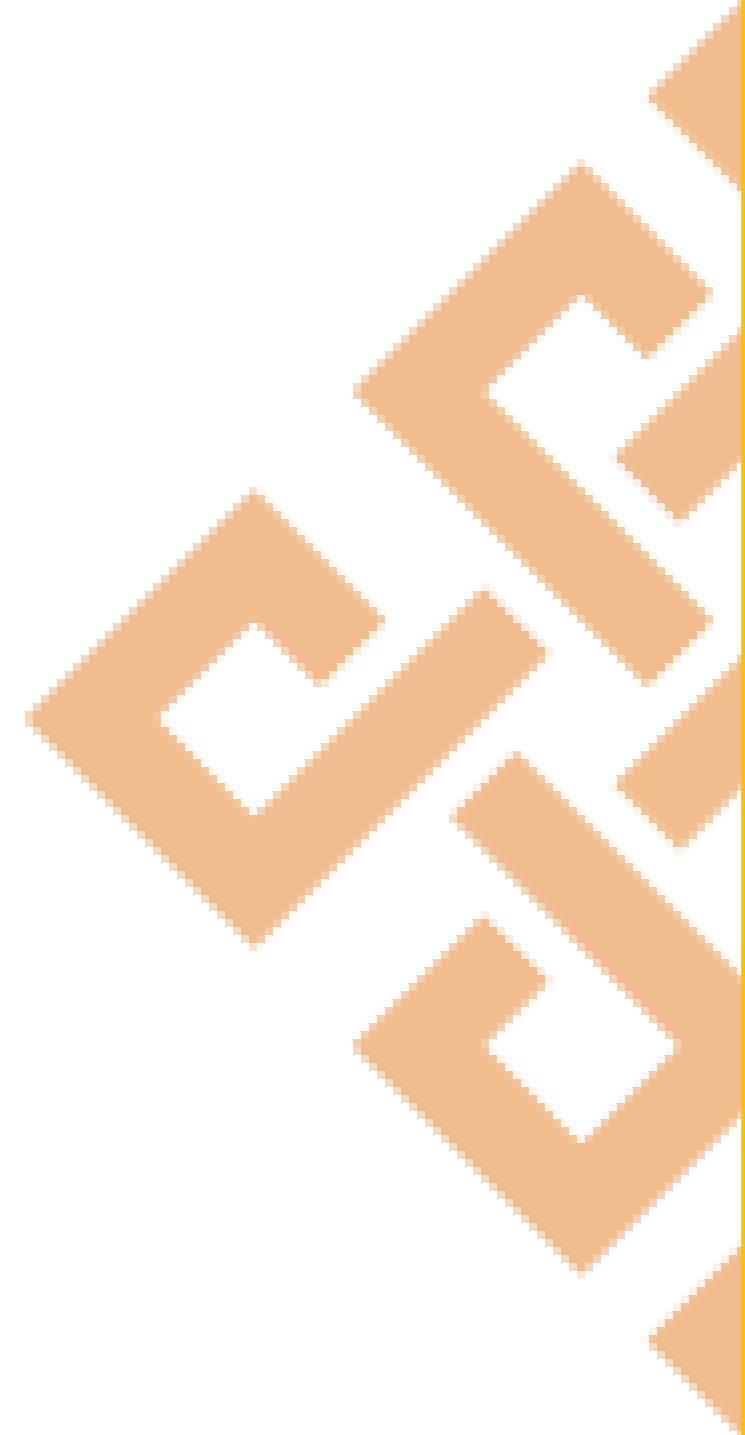
**Economics**

- Performance and price optimized business analytics solutions (includes hardware and software)



# Challenges with Big Data

**Other Characteristics of Big Data**



# CHALLENGES WITH BIG DATA

## Other Characteristics of Big Data

*Veracity* refers to biases, noise, and abnormality in data.

*Validity* refers to the accuracy and correctness of the data.

Volatility of data deals with, how long is the data valid? And how long should it be stored?

**Variability:** Data flows can be highly inconsistent with periodic peaks.





# SUMMARY OF THE LECTURE

Introduction  
to  
Challenges  
of Big Data

Definition  
of Big Data

What is Big  
Data

Why is Big  
Data

Other  
Characteristics  
of Big Data



# Traditional BI vs Big Data Data Warehouse Environment

School of Computer Science and Engineering

AY: 2021-2022



# OUTLINE

## Recap of previous Lecture

---

Topic for the Lecture

---

Objective and Outcome of Lecture

---

## Lecture Discussion

---

**Traditional Business Intelligence**

---

**Features of Business Intelligence**

---

**Business Intelligence Applications**

---

**Small Data v/s Big Data**

---

**Business Intelligence v/s Big Data**

---

Data Warehouse - Concept

---

Need for Data Warehouse

---

A typical Data Warehouse Environment

---

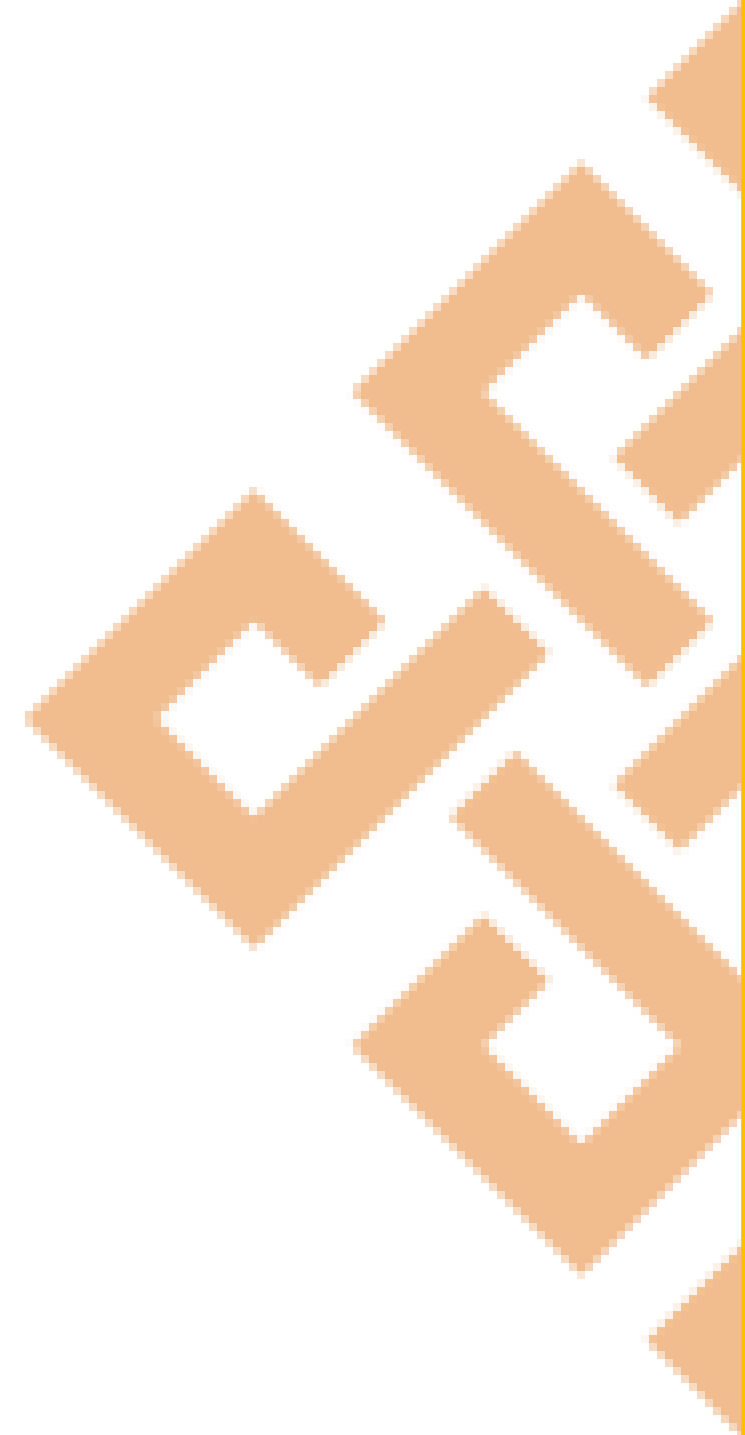
Data Warehouse Tools

---



# **Traditional BI vs Big Data**

**Recap of previous Lecture**



# RECAP OF PREVIOUS LECTURE

## Big Data

What is Big Data?

Why Big Data?

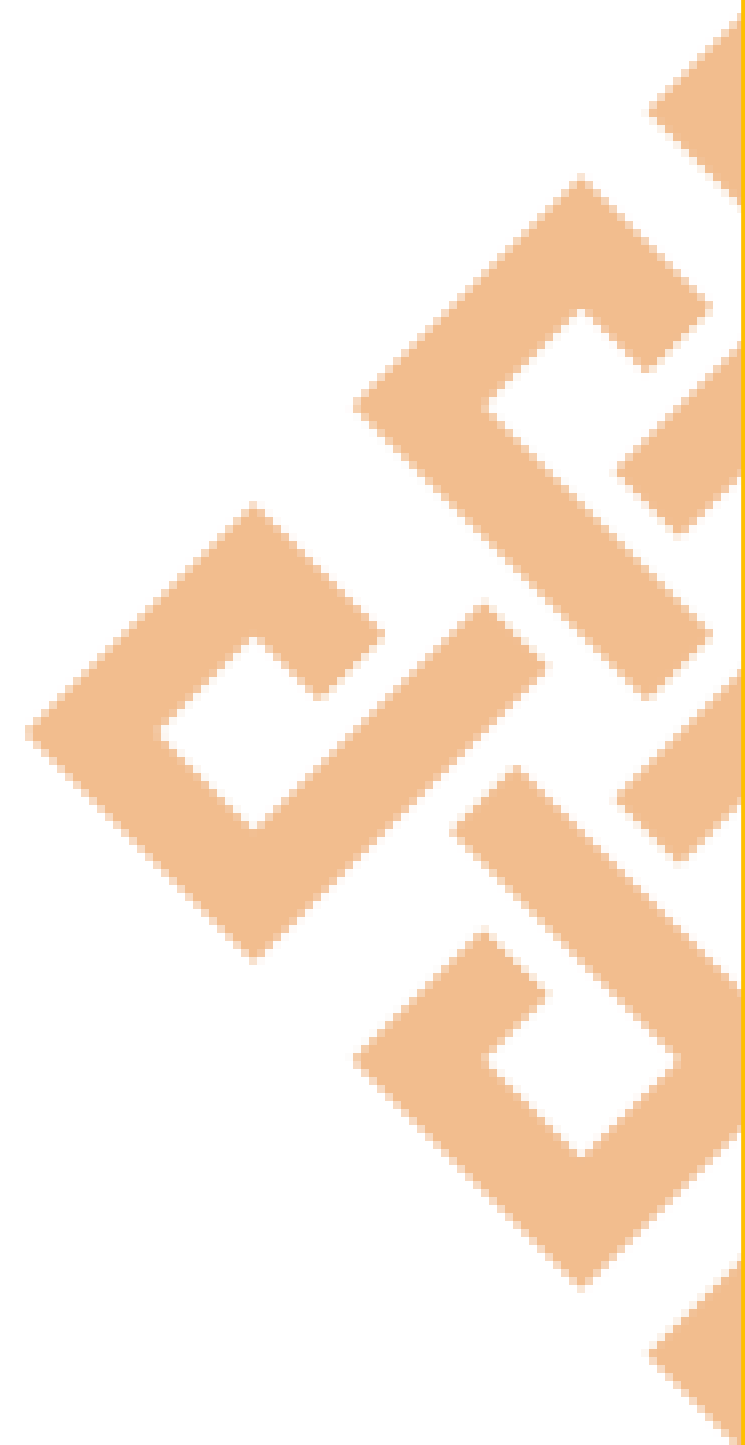
Characteristics of Big Data

Challenges of Big Data



# **Traditional BI vs Big Data**

**Business Intelligence**



# TOPIC OF THE LECTURE

## Business Intelligence

Traditional Business Intelligence

Features of Business Intelligence

Business Intelligence Applications

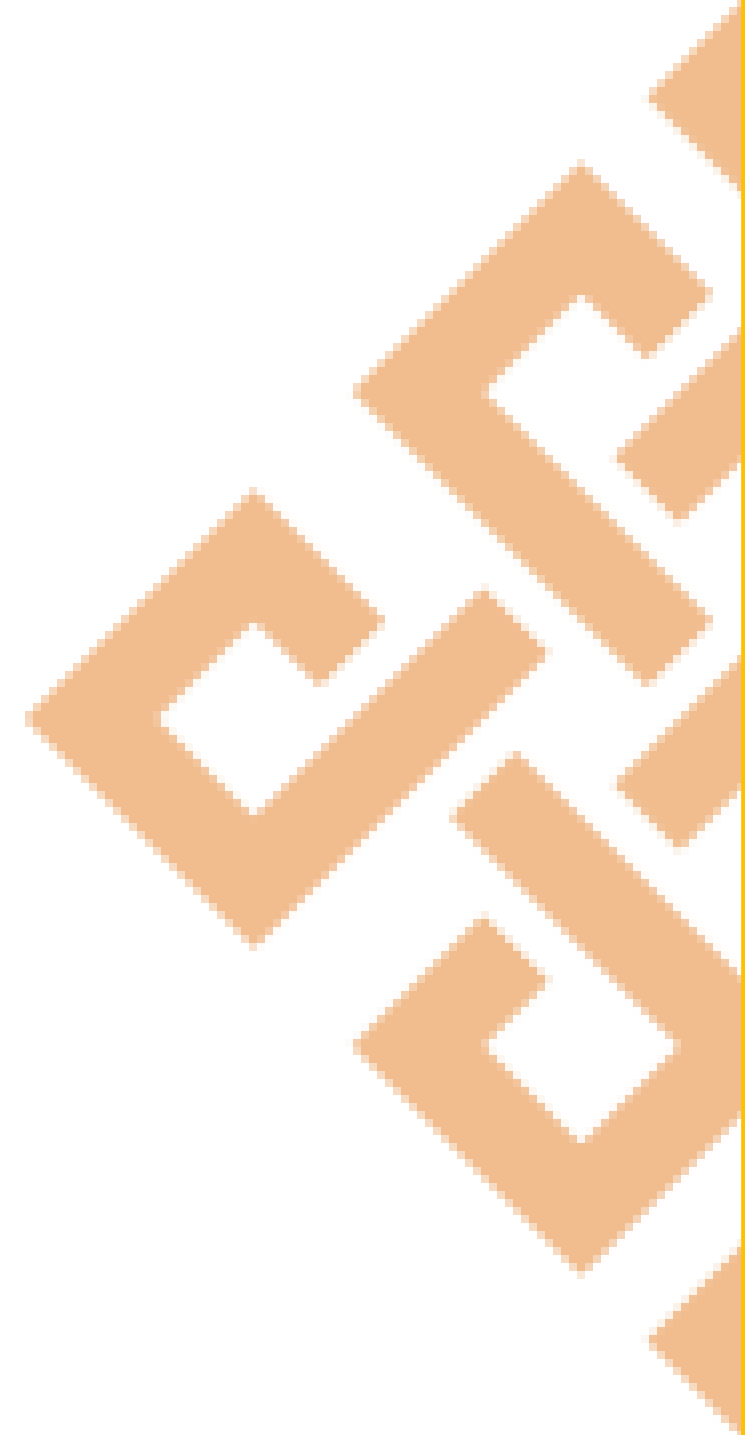
Small Data v/s Big Data

Business Intelligence v/s Big Data



# **Traditional BI vs Big Data**

**Objective and Outcome of Lecture**



# OBJECTIVE AND OUTCOME OF LECTURE

## Lecture Objective

Explain the different aspects of Big Data, Business Intelligence

## Lecture Outcome

Paraphrase the aspects of Business Intelligence, Contrast BI and Big Data





# **Traditional BI vs Big Data**

**Traditional Business Intelligence**



# TRADITIONAL BI VS BIG DATA

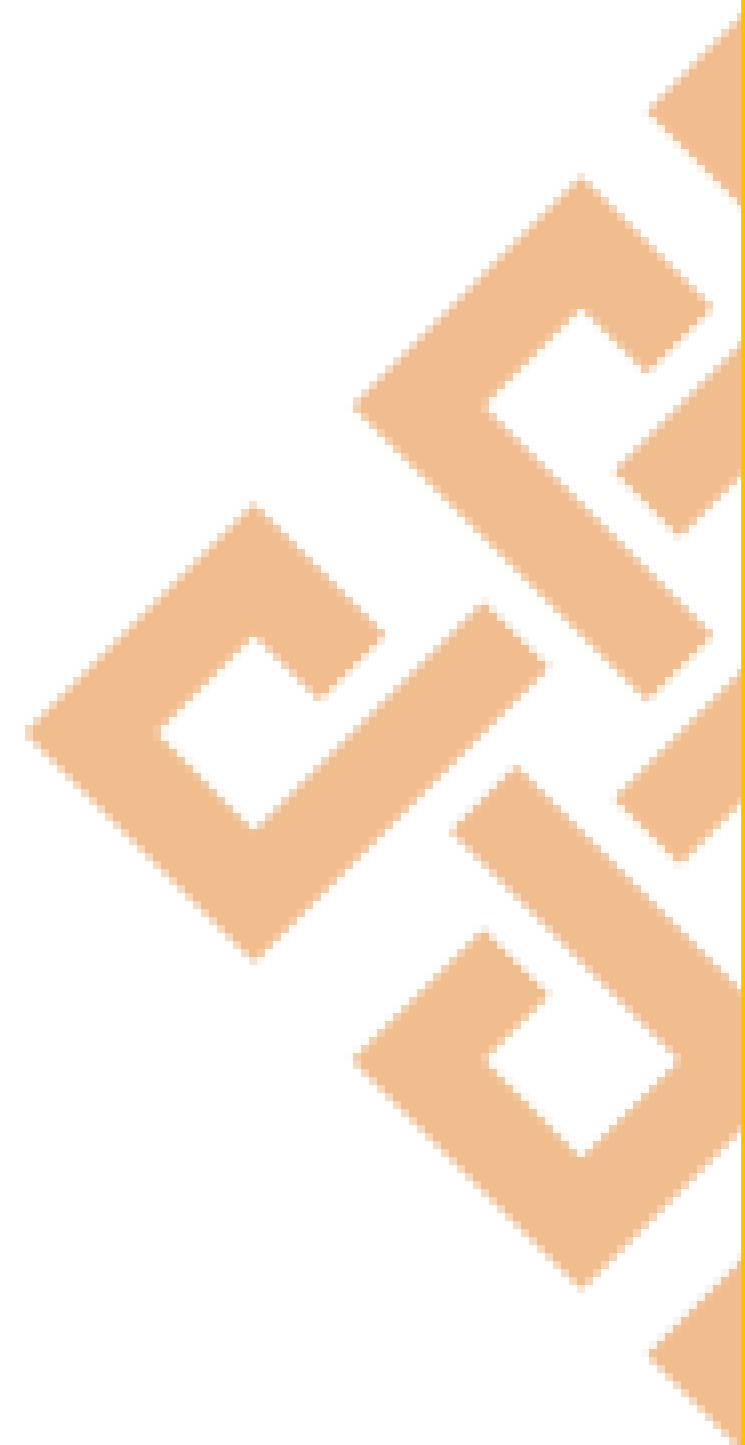
## BI Definition

Business Intelligence (BI) uses a set of processes, technologies, and tools to transform raw data into meaningful information and then transform information to provide knowledge.



# **Traditional BI vs Big Data**

**BI Features**



# TRADITIONAL BI VS BIG DATA

## BI Features



Enables users to predict customer behavior, forecast demand, and prepare strategies using modeling, statistics, machine learning, and data mining tools.



# TRADITIONAL BI VS BIG DATA

## BI Features



**Online analytical processing** helps users view a data slice from different viewpoints and improves reporting capabilities.



# TRADITIONAL BI VS BIG DATA

## BI Features



**Trend Indicators**

Help users spot patterns in production, sales, and distribution lines and identifies anomalies.





# TRADITIONAL BI VS BIG DATA

## BI Features

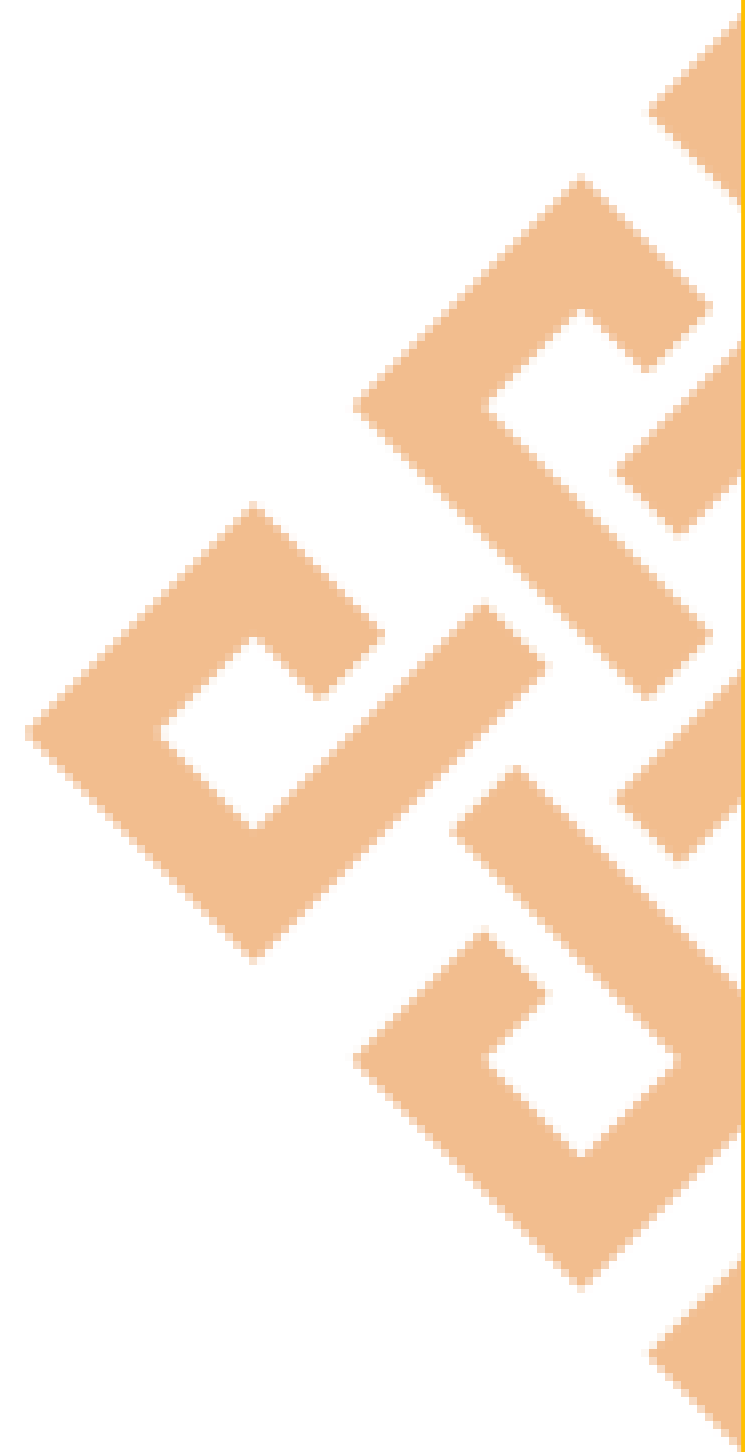


Allow users to create high-level financial and operational plans based on past performances and future goals.



# **Traditional BI vs Big Data**

**BI Applications**



# TRADITIONAL BI VS BIG DATA

## BI Applications



### Sales Intelligence

A key application of BI focuses on where your business meets the customer.



# TRADITIONAL BI VS BIG DATA

## BI Applications



### Visualization

Utilizes a range of data analytic tools to visualize



# TRADITIONAL BI VS BIG DATA

## BI Applications



### Reporting

A crucial business application of BI is reporting.



# TRADITIONAL BI VS BIG DATA

## BI Applications



### Performance management

organizations can monitor goal progress based on pre-defined or customizable timeframes.





# TRADITIONAL BI VS BIG DATA

## BI Applications



### Knowledge management

It is concerned with the creation, distribution, use, and management of business intelligence



# Traditional BI vs Big Data

## Small Data v/s Big Data



# TRADITIONAL BI VS BIG DATA

## Small Data v/s Big Data

	Small Data	Big Data
Data Condition	Ready for analysis, Flat files, no need of merging tables	Always Unstructured, not ready for analysis, many relational database tables that need merging
Location	Database, Local PC	Cloud, offshore, SQL Server etc.
Data Size	File that is in a spreadsheet, files that can be viewed on a few sheets of paper	Over 50K Variables, over 50K individual random samples, unstructured
Data Purpose	Intended purpose for Data Collection	No intended purpose



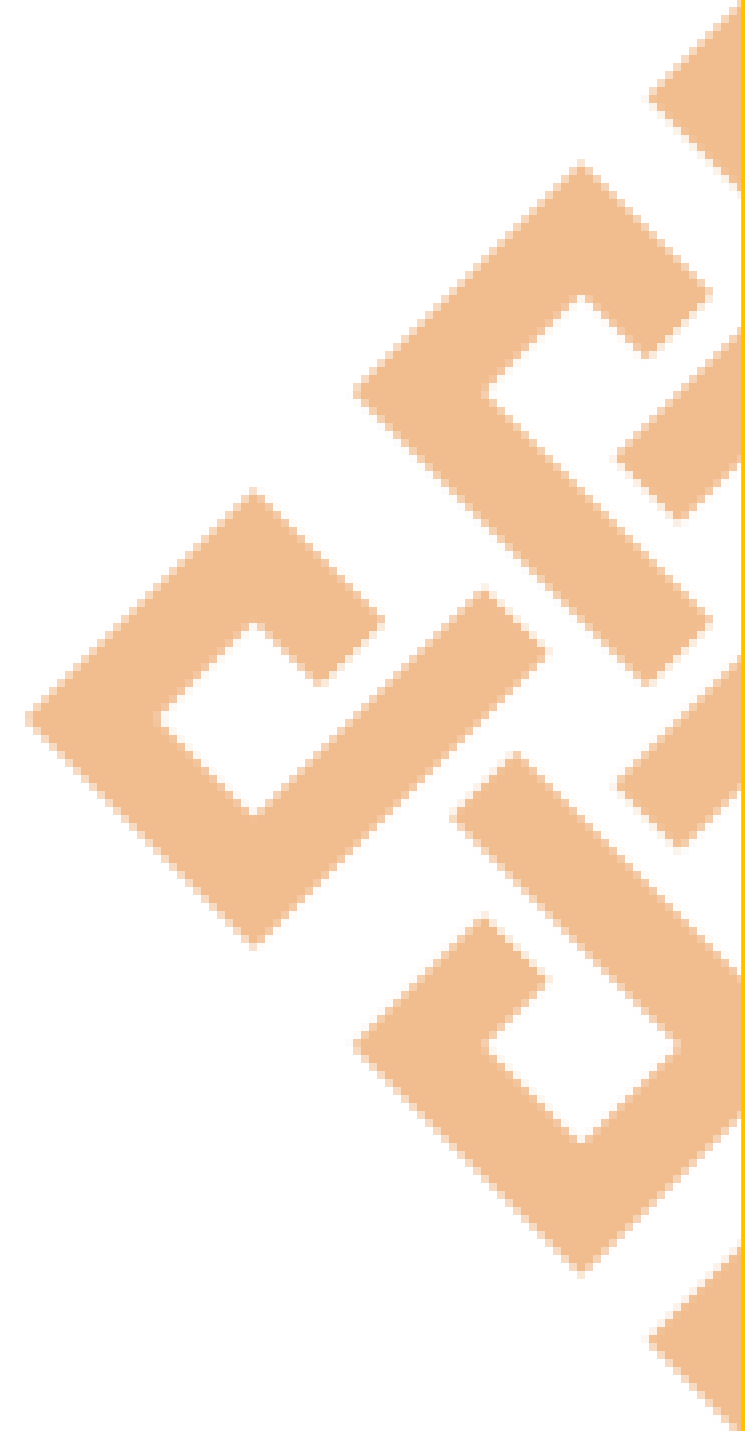
# Traditional BI vs Big Data

**Business Intelligence v/s Big Data**

**Business  
Intelligence**

vs

**Big Data**



# TRADITIONAL BI V/S BIG DATA

## Business Intelligence v/s Big Data

	Business Intelligence	Big Data
Data Storage	All the enterprise's data is stored in a central server	Data resides in a distributed file
Data Analysis	Data is generally analysed in an offline mode	Data is analysed in both in real times as well as offline mode
Data type	It is about structured	It is about structured, semi-structured and unstructured
Data Processing	data is taken to processing functions (move data to code)	processing functions are taken to data (move code to data)



# DATA WAREHOUSE ENVIRONMENT

## Data Warehouse - Concept



**Data Warehouse**

A **Data Warehouse** is different from DBMS, it stores huge amount of data, which is typically collected from multiple heterogeneous source like files, DBMS, etc.

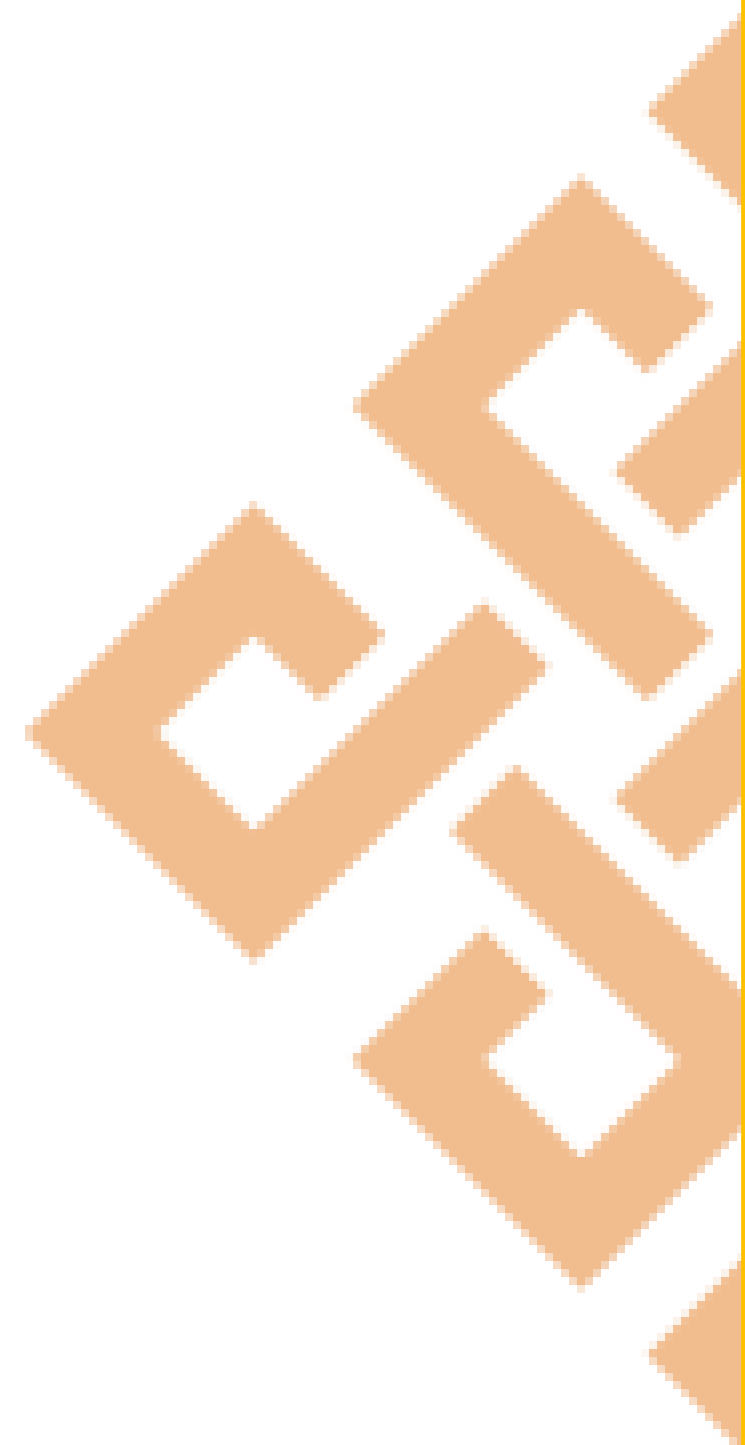
The **goal** is to produce statistical results that may help in decision making.





# **Data Warehouse Environment**

**Need for Data Warehouse**



# DATA WAREHOUSE ENVIRONMENT

## Need for Data Warehouse

**Goal of any business:** To make better Decisions

Eg: E-Commerce

We maintain data like

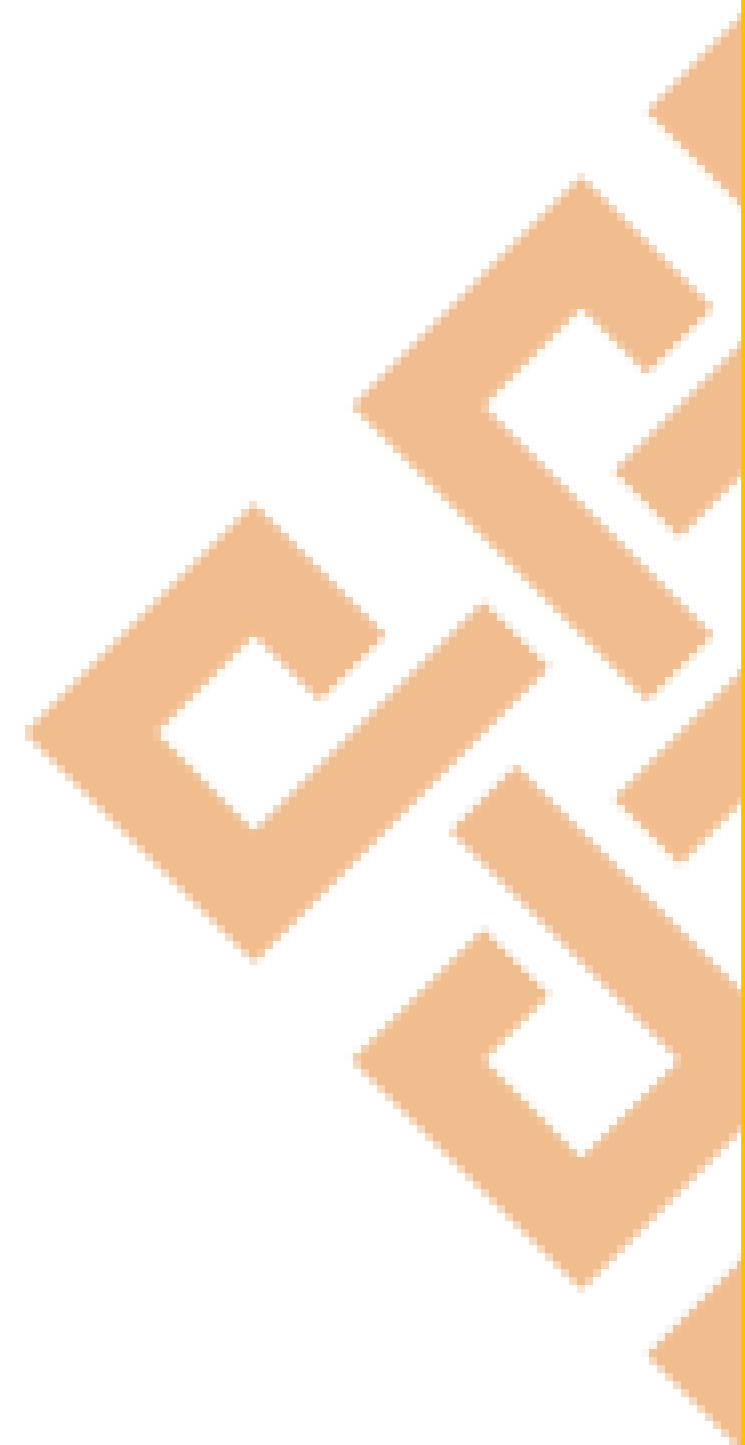
- Product details
- Customer Login Credentials
- Checkout details
- Merchant Account
- Other Information



**Need:** Concepts to be Extracted on a periodic basis, Formatted, Summarized and Supplemented.

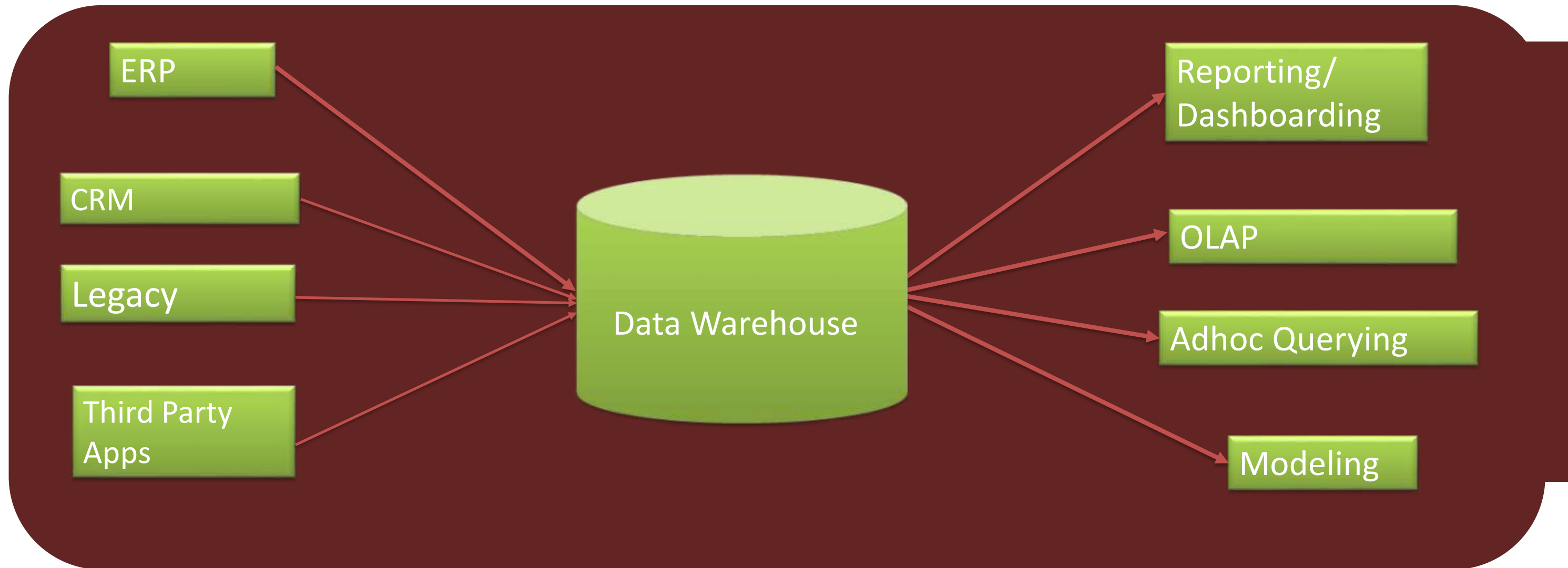
# **Data Warehouse Environment**

**A typical Data Warehouse Environment**



# DATA WAREHOUSE ENVIRONMENT

## A typical Data Warehouse Environment



# DATA WAREHOUSE ENVIRONMENT

## A typical Data Warehouse Environment

ERP: enterprise resource planning, the management of all the information and resources involved in a company's operations by means of an integrated computer system.

CRM: customer relationship management, denoting strategies and software that enable a company to optimize its customer relations.

Legacy: denoting or relating to software or hardware that has been superseded but is difficult to replace because of its wide use.

A third-party apps : Is an application created by a developer that isn't the manufacturer of the device the app runs on or the owner of the website that offers it.



# DATA WAREHOUSE ENVIRONMENT

## A typical Data Warehouse Environment

Reporting: Is used to generate human-readable reports from various data sources..

OLAP: Online analytical processing is a computer-based technique of analyzing data to look for insights.

Ad hoc query: ad hoc query is a loosely typed command/query whose value depends upon some variable

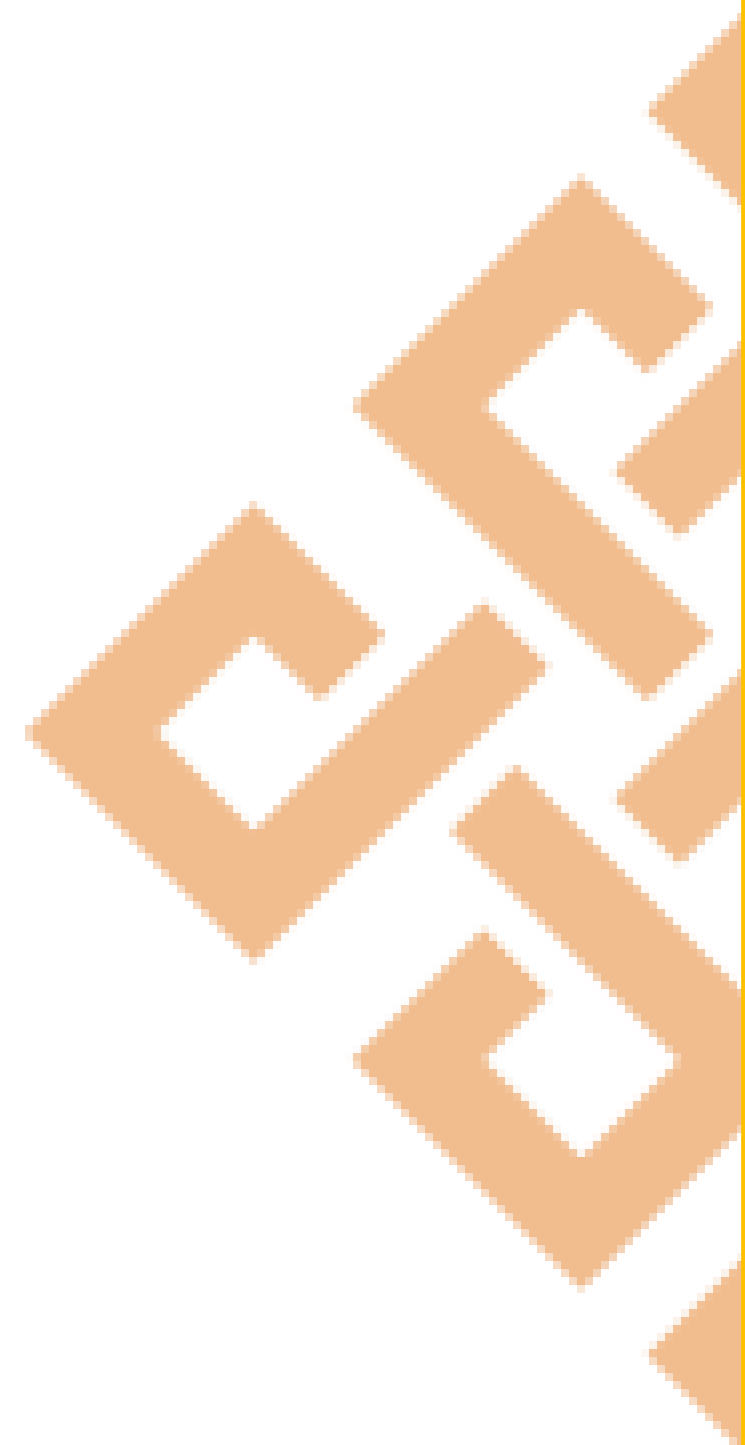
Modeling: is a graphical view of data created for analysis and design purposes.





# **Data Warehouse Environment**

**Data Warehouse Tools**



# DATA WAREHOUSE ENVIRONMENT

## Data Warehouse Tools



# SUMMARY OF THE LECTURE

Business  
Intelligence

Big Data

Business  
Intelligence  
v/s  
Big Data

Data  
Warehouse

Need for  
Data  
Warehouse,  
Architecture  
Tools





# THANK YOU

