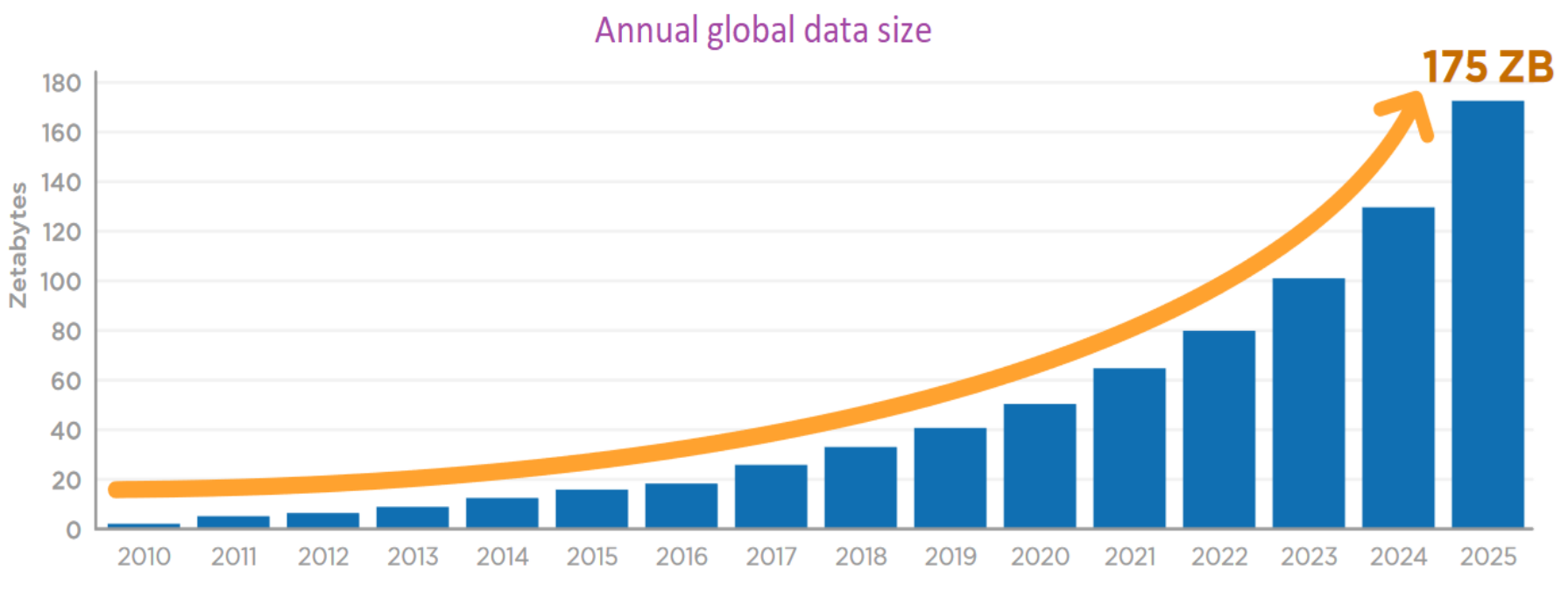
BIG DATA ANALYTICS

DATA EXPANSION



International Data Corporation report

- How many data in a day?
  - 7 TB (Twitter)
  - 10 TB (Facebook)
- 90% of world's data: generated over last two years!

UNITS OF DATA

Unit	Bytes
Kilobyte (KB)	10^3 (1,000)
Megabyte (MB)	10^6 (1,000,000)
Gigabyte (GB)	10^9 (1,000,000,000)
Terabyte (TB)	10^{12} (1,000,000,000,000)
Petabyte (PB)	10^{15} (1,000,000,000,000,000)
Exabyte (EB)	10^{18} (1,000,000,000,000,000,000)
Zettabyte (ZB)	10^{21} (1,000,000,000,000,000,000,000)
Yottabyte (YB)	10^{24} (1,000,000,000,000,000,000,000,000)

BIG DATA

- **Big Data** is nothing but lots of data consisting of varieties of data.
- **Analytics** is the discovery and communication of meaningful patterns in data.
- **Big data analytics** It is the concept of gathering useful insights from such voluminous amounts of structured, semi-structured and unstructured data that can be used for effective decision making in the business environment.

TYPES OF BIG DATA

Structured data –

It concerns all data which can be stored in database SQL in a table with rows and columns.

Example: Relational data

Semi-Structured data –

Semi-structured data is information that does not reside in a relational database but that have some organizational properties that make it easier to analyze. *Example*: XML data

Unstructured data –

Unstructured data is not organized in a pre-defined manner or does not have a pre-defined data model.

Example: Word, PDF, Text, Media logs.

PROPERTIES	STRUCTURED DATA	SEMI-STRUCTURED DATA	UNSTRUCTURED DATA
Technology	It is based on Relational database table	It is based on XML/RDF	It is based on character and binary data
Transaction management	Matured transaction and various concurrency technique	Transaction is adapted from DBMS not matured	No transaction management and no concurrency
Version management	Versioning over tuples,row,tables	Versioning over tuples or graph is possible	Versioned as whole
Flexibility	It is sehema dependent and less flexible	It is more flexible than structuded data but less than flexible than unstructured data	it very flexible and there is abbsence of schema
Scalability	It is very difficult to scale DB schema	It's scaling is simpler than sstructured data	It is very scalable
Robustness	Very robust	New technology, not very spread	–
Query performance	Structured query allow complex joining	Queries over anonymous nodes are possible	Only textual query are possible

EVOLUTION OF BIG DATA

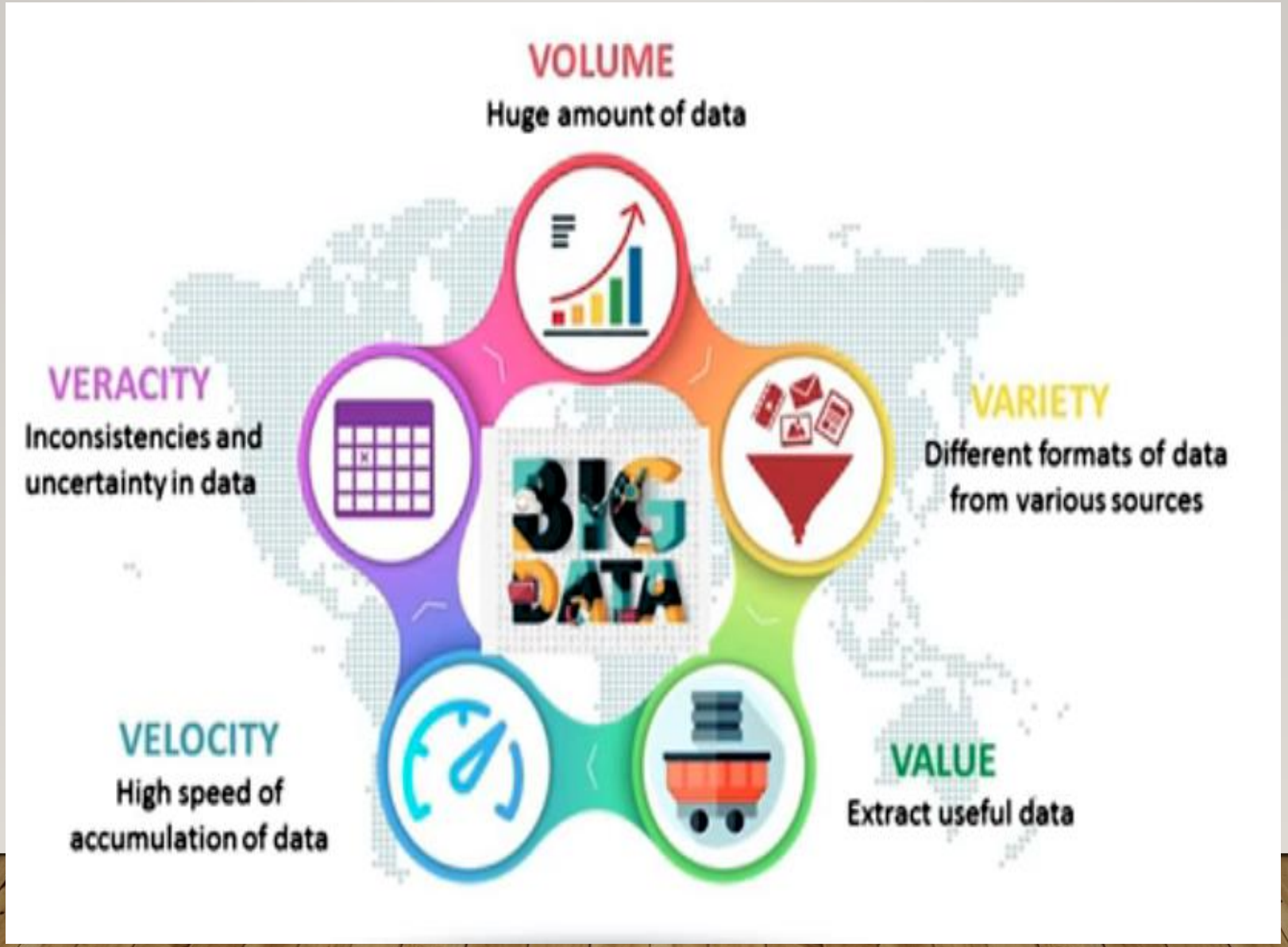
- Google search engine -> crawls thousands of web pages-> into File system-> Apply Page Rank ->Displays result.
- Google File System (GFS)

  Distributed File System created by google
- Google Map Reduce

  Distributed computing framework
- Google Big Table

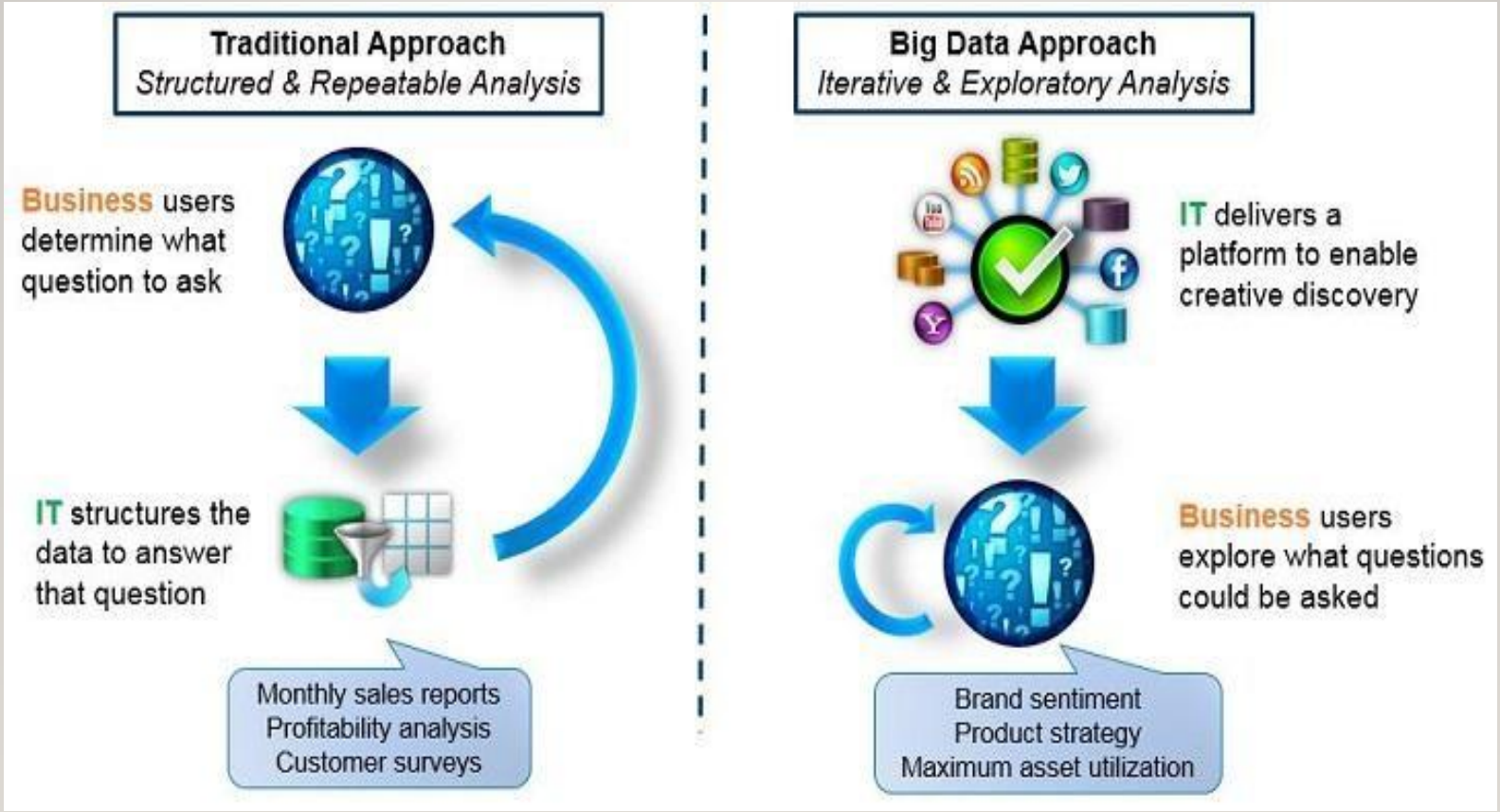
  Data structure, key-value pair concept

CHARACTERISTICS OF BIG DATA



Sl. No.	Traditional Data	Big Data
1.	Here the data is "Structured" data	Here the data is "Unstructured or Semi structured" data
2.	The size of the data is very small	The size is more than the traditional data size
3.	Here the data is Centralized	Here the data are distributed
4.	It is easy to work or manipulate	It is difficult to handle the data
5.	Normal system configuration is sufficient to process	High system configuration is required to process the data
6.	A traditional database tools is enough	Special kind of tools are required
7.	Normal functions are enough to manipulate the data	Requires special kind of functions to manipulate the data

TRADITIONAL VS. BIG DATA BUSINESS APPROACH



NEED OF BIG DATA ANALYTICS

1. Making smarter and more efficient organisation.
2. Optimize business operations by analysising customer behaviour.
3. Cost reduction.
4. Next Generation products.

1 Making Smarter and More Efficient Organisations

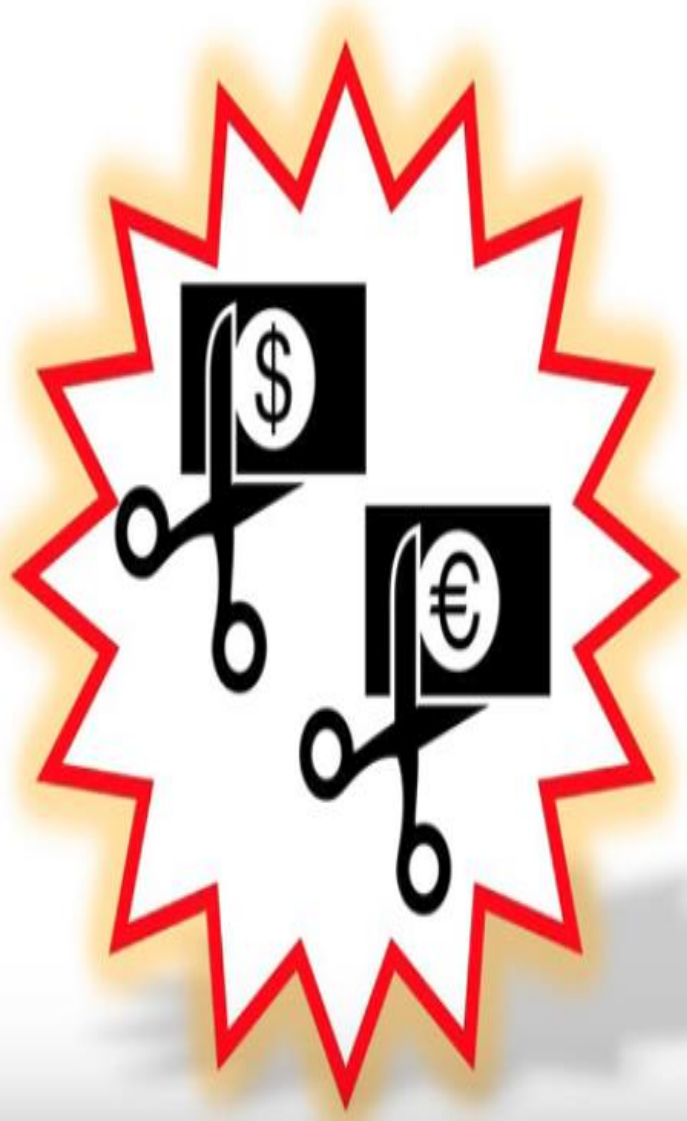
New York Police Department is utilizing data patterns, scientific analysis, and technological tools to prevent the occurrence of crime

2 Optimize Business Operations by analysing customer behaviour

- Analysing all the clicks of every visitor on a website
- Studying the paths leading them to buy products
- Customer Satisfaction

Amazon uses customer click-stream data and historical purchase data of more than 300 million customers and each user is shown customized results on customized web pages.

3 Cost Reduction



Parkland Hospital uses analytics and predictive modelling to identify high-risk patients and predict likely outcomes once patients are sent home. As a result, Parkland reduced 30-day readmissions for patients with heart failure, by 31 percent, saving $500,000 annually.

4 Next Generation Products

Big Data tools are used to operate Google's Self Driving Cars. The Toyota Prius is fitted with cameras, GPS as well as powerful computers and sensors to safely drive on the road without the intervention of human beings.

Netflix launched the seasons of its TV show House of Cards based on the user reviews, ratings and viewership.

NETFLIX

A smart yoga mat has sensors embedded in the mat will be able to provide feedback on your postures, score your practice, and even guide you through an at-home practice.

APPLICATIONS OF BIG DATA

- Entertainment Industry
- Insurance Industry
- Financial and Banking System
- Healthcare Industry
- Telecommunication and Media sector
- Security Enforcement

ADVANTAGES OF BIG DATA

- Predictive analysis which can save organizations from operational risks.
- Organizations grow business by analyzing customer needs.
- Enabled many multimedia platforms to share data.(ex. YouTube, Instagram, Facebook etc.)
- Big Data changed the face of customer based companies worldwide market.

- 1) How big data problems are handled by Hadoop system. 5M
- 2) What characteristics of social networks makes it Big data? 5M
- 3) What is Big data? What is Hadoop? How they are linked? 5M
- 4) How Big data can be useful in developing Digital India? 5M
- 5) Mention the 4 characteristics of bigdata. Elaborate these characteristics w.r.to social media websites. 5M
- 6) Give difference between Traditional data management and analytics approach Versus Big data Approach. 5M
- 7) Explain Hadoop Ecosystem with core components and explain its physical architecture. State limitations of Hadoop. 10M
- 8) Why is HDFS is more suited for applications having large datasets and not when there are small files? 5M

THANK YOU!