

**Table Of Contents****UNIT I****CHAPTER 1 : Introduction to Big Data 1-1 to 1-25****Syllabus :**

Introduction to Big Data, Big Data characteristics, types of Big Data, Traditional vs. Big Data business approach, Big Data Challenges, Examples of Big Data in Real Life, Big Data Applications.

**Self-learning Topics : Identification of Big Data applications and its solutions.**

1.1	Introduction to Big Data and Hadoop .....	1-1
1.1.1	What is Big Data ? .....	1-2
1.1.2	Sources of Big Data .....	1-2
1.2	Big Data Characteristics .....	1-3
1.2.1	Volume .....	1-4
1.2.2	Variety .....	1-5
1.2.3	Veracity .....	1-5
1.2.4	Velocity .....	1-6
1.2.5	Value .....	1-7
1.2.6	Visualization .....	1-7
1.2.7	Virality .....	1-8
1.3	Types of Big Data .....	1-8
1.3.1	Type #1 : Unstructured .....	1-8
1.3.1(A)	Characteristics of Unstructured Data .....	1-9
1.3.1(B)	Sources of Unstructured Data .....	1-9
1.3.1(C)	Advantages and Disadvantages of Unstructured Data .....	1-10
1.3.2	Type #2 : Structured .....	1-10
1.3.2(A)	Characteristics of Structured Data .....	1-11
1.3.2(B)	Sources of Structured Data .....	1-11

1.3.2(C) Advantages of Structured Data.....	1-12
1.3.3 Type #3 : Semi Structured.....	1-12
1.3.3(A) Characteristics of Semi-structured Data .....	1-13
1.3.3(B) Sources of semi-structured Data.....	1-14
1.3.3(C) Advantages and Disadvantages of Semi-structured Data .....	1-14
1.4 Difference between Structured, Semi-Structured and Un-Structured Data .....	1-15
1.5 Traditional vs. Big Data business approach.....	1-16
1.6 Examples of Big Data Applications .....	1-18
1.7 Big Data Challenges .....	1-20
1.8 Examples of Big Data in Real Life.....	1-23
• Chapter Ends .....	1-25

## **UNIT I**

### **CHAPTER 1**

# **Introduction to Big Data**

#### **University Prescribed Syllabus**

Introduction to Big Data, Big Data characteristics, types of Big Data, Traditional vs. Big Data business approach, Big Data Challenges, Examples of Big Data in Real Life, Big Data Applications.

**Self-learning Topics :** Identification of Big Data applications and its solutions.

### **► 1.1 INTRODUCTION TO BIG DATA AND HADOOP**

**GQ.** Firstly, We need to know "what is data" ?

- (1) Now a day the amount of data created by various advanced technologies like Social networking sites, E-commerce etc. is very large. It is really difficult to store such huge data by using the traditional data storage facilities.
- (2) Until 2003, the size of data produced was 5 billion gigabytes. If this data is stored in the form of disks it may fill an entire football field. In 2011, the same amount of data was created in every two days and in 2013 it was created in every ten minutes. This is really tremendous rate.

- (3) In this topic, we will discuss about big data on a fundamental level and define common concepts related to big data. We will also see in deep about some of the processes and technologies currently being used in this field.

### ❖ **1.1.1 What is Big Data ?**

**GQ. What is Big Data ?**

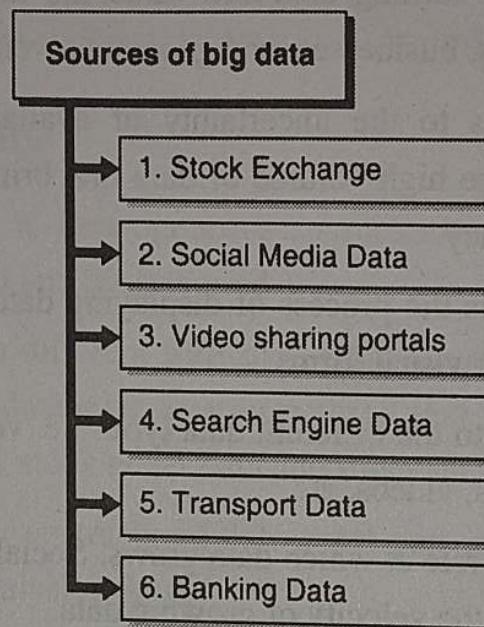
1. Big Data is a massive collection of data that continues to grow dramatically over time.
2. It is a data set that is so huge and complicated that no typical data management technologies can effectively store or process it.
3. Big Data is like regular data, but it is much larger. A data which are very large in size.
4. Normally we work on data of size MB(WordDoc ,Excel) or maximum GB(Movies, Codes) but data in Peta bytes i.e.  $10^{15}$  byte size is called Big Data.
5. It is stated that almost 90% of today's data has been generated in the past 3 years.

### ❖ **1.1.2 Sources of Big Data**

There are various sources of big data. Now a days in number of fields such huge data get created. Following are the some of fields.

1. **Stock Exchange :** The data in the share market regarding information about prices and status details of shares of thousands of companies is very huge.
2. **Social Media Data :** The data of social networking sites contains information about all the account holders, their posts, chat history, advertisements etc. On topmost sites like facebook and whatsapp, there are literally billions of users.

3. **Video sharing portals** : Video sharing portals like youtube, Vimeo etc. contains millions of videos each of which requires lots of memory to store.



**Fig. 1.1.1 : Sources of big data**

4. **Search Engine Data** : The search engines like Google and Yahoo holds lot much of metadata regarding various sites.
5. **Transport Data** : Transport data contains information about model, capacity, distance and availability of various vehicles.
6. **Banking Data** : The big giants in banking domain like SBI or ICICI hold large amount of data regarding huge transactions of account holders.

## ► 1.2 BIG DATA CHARACTERISTICS

**GQ.** What are Characteristics of Big Data ?

**UQ.** Describe any five characteristics of Big Data.

(MU - Dec. 17, 5 Marks)

**UQ.** Explain what characteristic of Social Networks make it Big Data.

(MU - May 18, 5 Marks)

**GQ.** Explain Big data along with 5V's.

- (1) **Volume** represents the volume i.e. amount of data that is growing at a high rate i.e. data volume in Petabytes.
- (2) **Value** refers to turning data into value. By turning accessed big data into values, businesses may generate revenue.
- (3) **Veracity** refers to the uncertainty of available data. Veracity arises due to the high volume of data that brings incompleteness and inconsistency.
- (4) **Visualization** is the process of displaying data in charts, graphs, maps, and other visual forms.
- (5) **Variety** refers to the different data types i.e. various data formats like text, audios, videos, etc.
- (6) **Velocity** is the rate at which data grows. Social media contributes a major role in the velocity of growing data.
- (7) **Virality** describes how quickly information gets spread across people to people (P2P) networks.

### 1.2.1 **Volume**

- As it follows from the name, big data is used to refer to enormous amounts of information.
- We are talking about not gigabytes but terabytes and petabytes of data.
- The IoT (Internet of Things) is creating exponential growth in data.
- The volume of data is projected to change significantly in the coming years.
- Hence, 'Volume' is one characteristic which needs to be considered while dealing with Big Data.

### **Volume**

[Data at Rest ]

Terabytes, Petabytes      Records/Arch      Table/Files      Distributed

### **1.2.2 Variety**

- Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.
- Data comes in different formats – from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data and financial transactions.
- This variety of unstructured data poses certain issues for storage, mining and analysing data.
- Organizing the data in a meaningful way is no simple task, especially when the data itself changes rapidly.
- Another challenge of Big Data processing goes beyond the massive volumes and increasing velocities of data but also in manipulating the enormous variety of these data.

### **Variety**

[ Data in many Forms ]

Structured      Unstructured      Text      Multimedia

### **1.2.3 Veracity**

- Veracity describes whether the data can be trusted. Veracity refers to the uncertainty of available data.
- Veracity arises due to the high volume of data that brings incompleteness and inconsistency.
- Hygiene of data in analytics is important because otherwise, you cannot guarantee the accuracy of your results.

- Because data comes from so many different sources, it's difficult to link, match, cleanse and transform data across systems.
- However, it is useless if the data being analysed are inaccurate or incomplete.
- Veracity is all about making sure the data is accurate, which requires processes to keep the bad data from accumulating in your systems.

### **Veracity**

[Data in Doubt ]

Trustworthiness      Authenticity      Accurate      Availability

### **1.2.4 Velocity**

- Velocity is the speed in which data is grows, process and becomes accessible.
- A data flows in from sources like business processes, application logs, networks, and social media sites, sensors, Mobile devices, etc.
- The flow of data is massive and continuous.
- Most data are warehoused before analysis, there is an increasing need for real-time processing of these enormous volumes.
- Real-time processing reduces storage requirements while providing more responsive, accurate and profitable responses.
- It should be processed fast by batch, in a stream-like manner because it just keeps growing every years.

### **Velocity**

[Data in Motion ]

Streaming      Batch      Real / Near Time      Processes

### **1.2.5 Value**

- It refers to turning data into value. By turning accessed big data into values, businesses may generate revenue.
- Value is the end game. After addressing volume, velocity, variety, variability, veracity, and visualization – which takes a lot of time, effort and resources – you want to be sure your organization is getting value from the data.
- For example, data that can be used to analyze consumer behavior is valuable for your company because you can use the research results to make individualized offers.

### **Value**

[Data into Money]

Statistical                      Events                      Correlations

### **1.2.6 Visualization**

- Big data visualization is the process of displaying data in charts, graphs, maps, and other visual forms.
- It is used to help people easily understand and interpret their data at a glance, and to clearly show trends and patterns that arise from this data.
- Raw data comes in a different formats, so creating data visualizations is process of gathering, managing, and transforming data into a format that's most usable and meaningful.
- Big Data Visualization makes your data as accessible as possible to everyone within your organization, whether they have technical data skills or not.

## ☛ **Visualization**

[Data Readable ]

Readable      Accessible      Presentation      Visual Forms

### ☛ **1.2.7 Virality**

- Virality describes how quickly information gets spread across people to people (P2P) networks.
- It measures how quickly data is spread and shared to each unique node.
- Time is a determinant factor along with rate of spread.

## ☛ **Virality**

[Data Spread ]

- P2P
- Shared
- Rate of Spread

## ► **1.3 TYPES OF BIG DATA**

**GQ.** What are different Types of Big Data ?

There are three types of Big Data Analytics :

1. Unstructured
2. Structured
3. Semi-structured

### ☛ **1.3.1 Type #1 : Unstructured**

- Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, unstructured data poses multiple challenges in terms of its processing for deriving value out of it.

- Typical example of unstructured data is, a heterogeneous data source containing a combination of simple text files, images, videos like search in Google Engine.
- Now a day organizations have wealth of data available with them but unfortunately they don't know how to derive value out of it since this data is in its raw form or unstructured format.
- Human Generated Data Machine Generated Data.
- Unstructured – Example : The output returned by 'Google Search'

### **❖ 1.3.1(A) Characteristics of Unstructured Data**

- (1) Data neither conforms to a data model nor has any structure.
- (2) Data can not be stored in the form of rows and columns as in Databases.
- (3) Data does not follows any semantic or rules.
- (4) Data lacks any particular format or sequence.
- (5) Data has no easily identifiable structure.
- (6) Due to lack of identifiable structure, it can not used by computer programs easily.

### **❖ 1.3.1(B) Sources of Unstructured Data**

- (1) Web pages
- (2) Images (JPEG, GIF, PNG, etc.)
- (3) Videos
- (4) Memos
- (5) Reports
- (6) Word documents and PowerPoint presentations
- (7) Surveys

### 1.3.1(C) Advantages and Disadvantages of Unstructured Data

#### Advantages

1. Its supports the data which lacks a proper format or sequence.
2. The data is not constrained by a fixed schema.
3. Very Flexible due to absence of schema.
4. Data is portable.
5. It is very scalable.
6. It can deal easily with the heterogeneity of sources.
7. These type of data have a variety of business intelligence and analytics applications.

#### Disadvantages

1. It is difficult to store and manage unstructured data due to lack of schema and structure.
2. Indexing the data is difficult and error prone due to unclear structure and not having pre-defined attributes. Due to which search results are not very accurate.
3. Ensuring security to data is difficult task.

### 1.3.2 Type #2 : Structured

- Any data that can be stored, accessed and processed in the form of fixed format is termed as a "Structured" data.
- Over the period of time, talent in computer science have achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also determining value out of it.
- When size of such data grows to a huge extent, typical sizes are being in the range of multiple zettabyte. Data stored in a relational database management system is one example of a structured data.



- **Structured data** is the data which conforms to a data model, has a well define structure, follows a consistent order and can be easily accessed and used by a person or a computer program.
- Structured data is usually stored in well-defined schemas such as Databases. It is generally tabular with column and rows that clearly define its attributes.
- SQL (Structured Query language) is often used to manage structured data stored in databases.

### **1.3.2(A) Characteristics of Structured Data**

- Data conforms to a data model and has easily identifiable structure.
- Data is stored in the form of rows and columns.

#### **Example : Database**

- Data is well organised so, Definition, Format and Meaning of data is explicitly known.
- Data resides in fixed fields within a record or file.
- Similar entities are grouped together to form relations or classes.
- Entities in the same group have same attributes.
- Easy to access and query, So data can be easily used by other programs.
- Data elements are addressable, so efficient to analyse and process.

### **1.3.2(B) Sources of Structured Data**

- (1) SQL Databases
- (2) Spreadsheets such as Excel
- (3) OLTP Systems
- (4) Online forms
- (5) Sensors such as GPS or RFID tags
- (6) Network and Web server logs
- (7) Medical devices

### 1.3.2(C) Advantages of Structured Data

1. Structured data have a well defined structure that helps in easy storage and access of data.
2. Data can be indexed based on text string as well as attributes. This makes search operation hassle-free.
3. Data mining is easy i.e. knowledge can be easily extracted from data.
4. Operations such as Updating and deleting is easy due to well structured form of data.
5. Business Intelligence operations such as Data warehousing can be easily undertaken.
6. Easily scalable in case there is an increment of data.
7. Ensuring security to data is easy.

#### Structured - Example

Employee\_Table

Employee_ID	Employee_Name	Gender	Department	Salary_In_Lacs
1	XYX	MALE	FINANCE	850000
2	ABC	MALE	ADMIN	250000
3	PQR	FEMALE	SALES	350000
4	MNR	FEMALE	FINANCE	600000

### 1.3.3 Type #3 : Semi Structured

- Semi structured is the third type of big data. Semi-structured data can contain both the forms of data.
- Semi-structured data pertains to the data containing both the formats mentioned above, that is, structured and unstructured data.

- To be precise, it refers to the data that although has not been classified under a particular repository (database), yet contains vital information or tags that segregate individual elements within the data.
- Web application data, which is unstructured, consists of log files, transaction history files etc.
- Online transaction processing systems are built to work with structured data wherein data is stored in relations (tables).
- Semi-structured data is data that does not conform to a data model but has some structure. It lacks a fixed or rigid schema. It is the data that does not reside in a rational database but that have some organizational properties that make it easier to analyze. With some processes, we can store them in the relational database.

#### **1.3.3(A) Characteristics of Semi-structured Data**

1. Data does not conform to a data model but has some structure. Data can not be stored in the form of rows and columns as in Databases
2. Semi-structured data contains tags and elements (Metadata) which is used to group data and describe how the data is stored.
3. Similar entities are grouped together and organized in a hierarchy. Entities in the same group may or may not have the same attributes or properties.
4. Does not contain sufficient metadata which makes automation and management of data difficult.
5. Size and type of the same attributes in a group may differ.
6. Due to lack of a well-defined structure, it can not be used by computer programs easily.

### ☞ **1.3.3(B) Sources of semi-structured Data**

- (1) E-mails
- (2) XML and other markup languages
- (3) Binary executables
- (4) TCP/IP packets
- (5) Zipped files
- (6) Integration of data from different sources
- (7) Web pages

### ☞ **1.3.3(C) Advantages and Disadvantages of Semi-structured Data**

#### ☛ **Advantages**

- 1. The data is not constrained by a fixed schema.
- 2. Flexible i.e. Schema can be easily changed.
- 3. Data is portable.
- 4. It is possible to view structured data as semi-structured data.
- 5. It supports users who can not express their need in SQL.
- 6. It can deal easily with the heterogeneity of sources.

#### ☛ **Disadvantages**

- 1. Lack of fixed, rigid schema make it difficult in storage of the data.
- 2. Interpreting the relationship between data is difficult as there is no separation of the schema and the data.
- 3. Queries are less efficient as compared to structured data.

#### ☛ **Semi-structured - Example**

- User can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS.

- Personal data stored in a XML file :

```

<rec><name>Prashant
Rao</name><sex>Male</sex><age>35</age></rec><rec>
<name>Seema
R.</name><sex>Female</sex><age>41</age></rec><rec>
<name>Satish
Mane</name><sex>Male</sex><age>29</age></rec>

```

## ► 1.4 DIFFERENCE BETWEEN STRUCTURED, SEMI-STRUCTURED AND UN-STRUCTURED DATA

**GQ.** What is difference between structured, semi-structured and Un-Structured Data ?

Properties	Structured data	Semi-structured data	Unstructured data
Technology	It is based on Relational database table	It is based on XML/RDF(Resource Description Framework).	It is based on character and binary data
Transaction management	Matured transaction and various concurrency techniques	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 a whole
Flexibility	It is schema dependent and less flexible	It is more flexible than structured data but less flexible than unstructured data	It is more flexible and there is absence of schema

Properties	Structured data	Semi-structured data	Unstructured data
Scalability	It is very difficult to scale DB schema	It's scaling is simpler than structured data	It is more scalable.
Robustness	Very robust	New technology, not very spread	-
Query performance	Structured query allow complex joining	Queries over anonymous nodes are possible	Only textual queries are possible

## ► 1.5 TRADITIONAL VS. BIG DATA BUSINESS APPROACH

**GQ.** What is Traditional Data & Big Data ?

**GQ.** Explain in detail Traditional vs. Big Data Business Approach.

(5 Marks)

### 1. Traditional Data

- Traditional data is the structured data which is being majorly maintained by all types of businesses starting from very small to big organizations.
- In traditional database system a centralized database architecture used to store and maintain the data in a fixed format or fields in a file. For managing and accessing the data Structured Query Language (SQL) is used.

### 2. Bigdata

- We can consider big data an upper version of traditional data. Big data deal with too large or complex data sets which is difficult to manage in traditional data-processing application software.

- It deals with large volume of both structured, semi structured and unstructured data. Volume, Velocity and Variety, Veracity and Value refer to the 5'V characteristics of big data.
- Big data not only refers to large amount of data it refers to extracting meaningful data by analyzing the huge amount of complex data sets.

**UQ.** Compare big data analytics with traditional data mining.

(MU - Dec. 18, 5 Marks)

Sr. No.	Traditional Data	Big Data
1.	Traditional data is generated in enterprise level.	Big data is generated in outside and enterprise level.
2.	Its volume ranges from Gigabytes to Terabytes.	Its volume ranges from Petabytes to Zettabytes or Exabytes.
3.	Traditional database system deals with structured data.	Big data system deals with structured, semi structured and unstructured data.
4.	Traditional data is generated per hour or per day or more.	But big data is generated more frequently mainly per seconds.
5.	Traditional data source is centralized and it is managed in centralized form.	Big data source is distributed and it is managed in distributed form.
6.	Data integration is very easy.	Data integration is very difficult.
7.	Normal system configuration is capable to process traditional data.	High system configuration is required to process big data.

Sr. No.	Traditional Data	Big Data
8.	The size of the data is very small.	The size is more than the traditional data size.
9.	Traditional data base tools are required to perform any data base operation.	Special kind of data base tools are required to perform any data base operation.
10.	Normal functions can manipulate data.	Special kind of functions can manipulate data.
11.	Its data model is strict schema based and it is static.	Its data model is flat schema based and it is dynamic.
12.	Traditional data is stable and inter relationship.	Big data is not stable and unknown relationship.
13.	Traditional data is in manageable volume.	Big data is in huge volume which becomes unmanageable.
14.	It is easy to manage and manipulate the data.	It is difficult to manage and manipulate the data.
15.	Its data sources includes ERP transaction data, CRM transaction data, financial data, organizational data, web transaction data etc.	Its data sources includes social media, device data, sensor data, video, images, audio etc.

## 1.6 EXAMPLES OF BIG DATA APPLICATIONS

GQ. List the examples of big data.

(2 Marks)

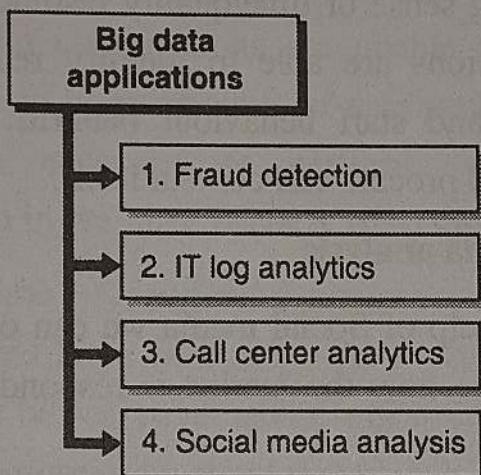
GQ. Explain the examples of big data.

(6 Marks)

There are various big data applications as shown in Fig 1.6.1

► **1. Fraud detection**

- Fraud detection is a Big Data application example for businesses which has operations like any type of claims or transaction processing.
- Number of times the detection of fraud is concluded long after the fact. At this point the damage has been already done all that's left is to decrease the harm and revise policies to prevent it in future.



**Fig. 1.6.1 : Big data applications**

- The Big Data platforms can analyze claims and transactions of businesses. They identify large-scale patterns across many transactions or detect anomalous behaviour of a some user. This helps to avoid the fraud.

► **2. IT log analytics**

- An enormous quantity of logs and trace data is generated in IT solutions and IT departments. Many times such data go unexamined: organizations simply don't have the manpower or resource to go through all such information.

- Big data has the ability to quickly identify large-scale patterns to help in diagnosing and preventing problems. It helps the organization with a large IT department.

### ► 3. Call center analytics

- Now we turn to the customer-facing Big Data application examples, of which call center analytics are particularly powerful. Without a Big Data solution, much of the insight that a call center can provide will be ignored or exposed later.
- By making sense of time/quality resolution metrics, the Big Data solutions are able to identify recurring problems or customer and staff behaviour patterns. Big data can also capture and process call content itself.

### ► 4. Social media analysis

- With the help of Social media we can observe the real-time insights into how the market is responding to products and campaigns.
- With the help of these insights, it is possible for companies to adjust their pricing, promotion, and campaign placement to get optimal results.

---

## ► 1.7 BIG DATA CHALLENGES

---

### 1. Sharing and Accessing Data

- Perhaps the most frequent challenge in big data efforts is the inaccessibility of data sets from external sources.
- Sharing data can cause substantial challenges.
- It include the need for inter and intra- institutional legal documents.
- Accessing data from public repositories leads to multiple difficulties.

- It is necessary for the data to be available in an accurate, complete and timely manner because if data in the companies information system is to be used to make accurate decisions in time then it becomes necessary for data to be available in this manner.

## **2. Privacy and Security**

- It is another most important challenge with Big Data. This challenge includes sensitive, conceptual, technical as well as legal significance.
- Most of the organizations are unable to maintain regular checks due to large amounts of data generation. However, it should be necessary to perform security checks and observation in real time because it is most beneficial.
- There is some information of a person which when combined with external large data may lead to some facts of a person which may be secretive and he might not want the owner to know this information about that person.
- Some of the organization collects information of the people in order to add value to their business. This is done by making insights into their lives that they're unaware of.

## **3. Analytical Challenges**

- There are some huge analytical challenges in big data which arise some main challenges questions like how to deal with a problem if data volume gets too large?
- Or how to find out the important data points?
- Or how to use data to the best advantage?
- These large amount of data on which these type of analysis is to be done can be structured (organized data), semi-structured (Semi-organized data) or unstructured (unorganized data).

There are two techniques through which decision making can be done :

1. Either incorporate massive data volumes in the analysis.
2. Or determine upfront which Big data is relevant.

#### **4. Technical challenges**

##### **Quality of data**

1. When there is a collection of a large amount of data and storage of this data, it comes at a cost. Big companies, business leaders and IT leaders always want large data storage.
2. For better results and conclusions, Big data rather than having irrelevant data, focuses on quality data storage.
3. This further arise a question that how it can be ensured that data is relevant, how much data would be enough for decision making and whether the stored data is accurate or not.

##### **Fault tolerance**

1. Fault tolerance is another technical challenge and fault tolerance computing is extremely hard, involving intricate algorithms.
2. Nowadays some of the new technologies like cloud computing and big data always intended that whenever the failure occurs the damage done should be within the acceptable threshold that is the whole task should not begin from the scratch.

##### **Scalability**

1. Big data projects can grow and evolve rapidly. The scalability issue of Big Data has lead towards cloud computing.
2. It leads to various challenges like how to run and execute various jobs so that goal of each workload can be achieved cost-effectively.

3. It also requires dealing with the system failures in an efficient manner. This leads to a big question again that what kinds of storage devices are to be used.

## **► 1.8 EXAMPLES OF BIG DATA IN REAL LIFE**

### **(1) In the Education Industry**

The University of Alabama has more than 38,000 students and an ocean of data. In the past when there were no real solutions to analyze that much data, some of them seemed useless. Now, administrators can use analytics and data visualizations for this data to draw out patterns of students revolutionizing the university's operations, recruitment, and retention efforts.

### **(2) In the Healthcare**

Wearable devices and sensors have been introduced in the healthcare industry which can provide real-time feed to the electronic health record of a patient. One such technology is Apple.

Apple has come up with Apple HealthKit, CareKit, and ResearchKit. The main goal is to empower iPhone users to store and access their real-time health records on their phones.

### **(3) In Government Sector**

Food and Drug Administration (FDA) which runs under the jurisdiction of the Federal Government of the USA leverages the analysis of big data to discover patterns and associations to identify and examine the expected or unexpected occurrences of food-based infections.

### **(4) In Media and Entertainment Industry**

Spotify, on-demand music-providing platform, uses Big Data Analytics, collects data from all its users around the globe, and then uses the analyzed data to give informed music recommendations and suggestions to every individual user.

Amazon Prime which offers, videos, music, and Kindle books in a one-stop shop is also big on using big data.

#### **(5) In Weather Patterns**

IBM Deep Thunder, which is a research project by IBM, provides weather forecasting through high-performance computing of big data. IBM is also assisting Tokyo with improved weather forecasting for natural disasters or predicting the probability of damaged power lines.

#### **(6) In Transportation Industry**

Uber generates and uses a huge amount of data regarding drivers, their vehicles, locations, every trip from every vehicle, etc. All this data is analyzed and then used to predict supply, demand, location of drivers, and fares that will be set for every trip.

#### **(7) In Banking Sector**

Various anti-money laundering software such as SAS AML uses Data Analytics in Banking to detect suspicious transactions and analyze customer data. Bank of America has been a SAS AML customer for more than 25 years.

#### **(8) In Marketing**

Amazon collected data about the purchase done by millions of people around the world. They analyzed the purchase patterns and payment methods used by the customers and used the results to design new offers and advertisements.

#### **(9) In Business Insights**

Netflix is using Big Data to understand the user behavior, the type of content they like, popular movies on the website, similar content that can suggest to the user, and which series or movies should they invest in.

### **(10) In Space Sector**

NASA is collecting data from different satellites and rovers about the geography, atmospheric conditions, and other factors of mars for their upcoming mission. It uses big data to manage all that data and analyzes that to run simulations.

---

*Chapter Ends...*

