

IoT Architecture and Technologies

IRE 211



Big Data Analytics

Presented By

Mumith Chowdhury (2101019)

Sifatul Islam Rafi (2101021)

Kamrunnahar Keya (2101025)

Presented To

Sourav Chandra Das

Lecturer

Dept. of BDU,IRE

What is Big Data?

- ❑ Big Data is a massive collection of data that is growing exponentially over time.
- ❑ It is a data set that is so large and complex that traditional data management tools cannot store or process it efficiently.
- ❑ Big data is a type of data that is extremely large in size.



Types of big-data

Big Data is generally categorized into three different varieties.

- Structured Data
- Semi-Structured Data
- Unstructured Data



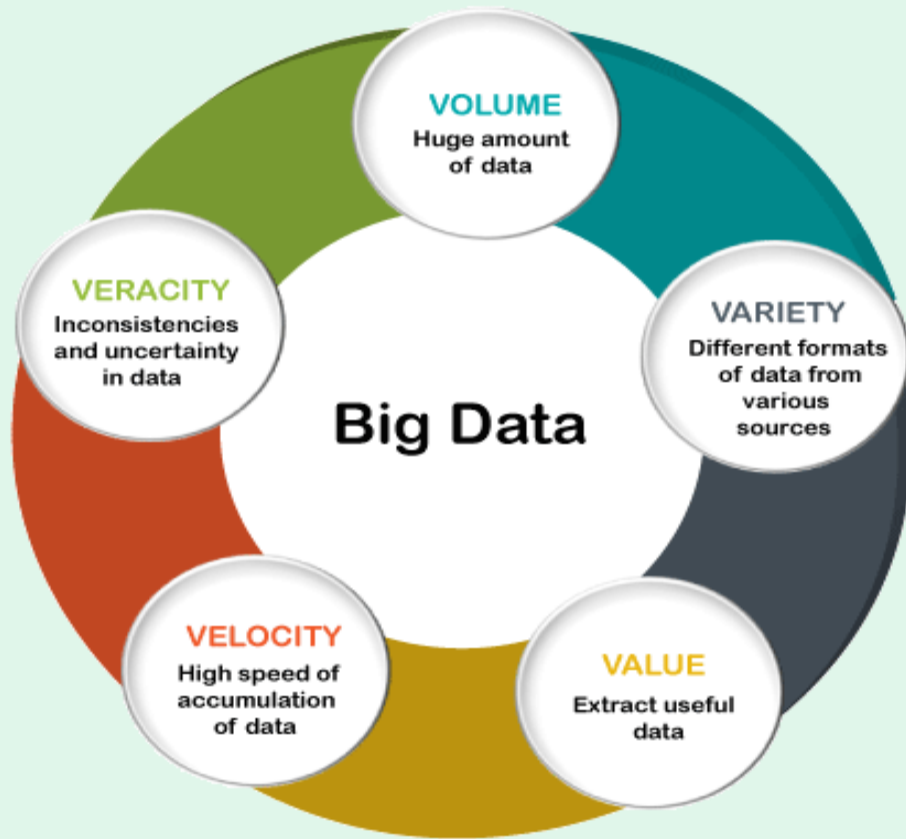
Types of big-data

- **Structured Data:** Has a clear structure and consistent format, stored in well-defined columns or databases.
Example: Databases (DBMS).
- **Semi-Structured Data:** Partially structured, lacks a strict format but includes some organization.
Example: CSV files.
- **Unstructured Data:** No defined structure or consistent format; varies greatly.
Example: Audio files, images.

Characteristics of Big Data

- **Volume**
- **Velocity**
- **Variety**
- **Veracity**
- **Value**

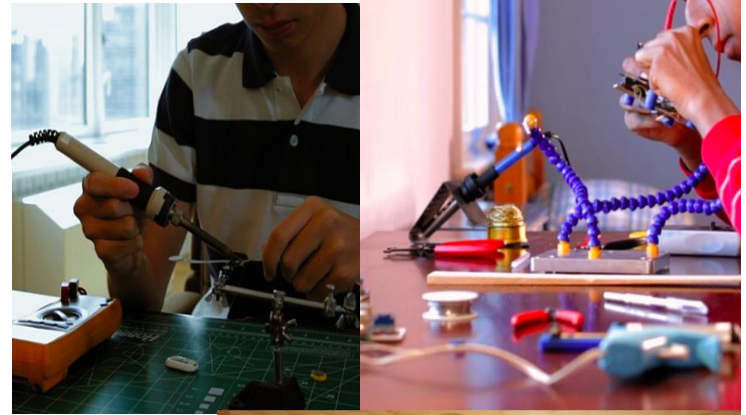
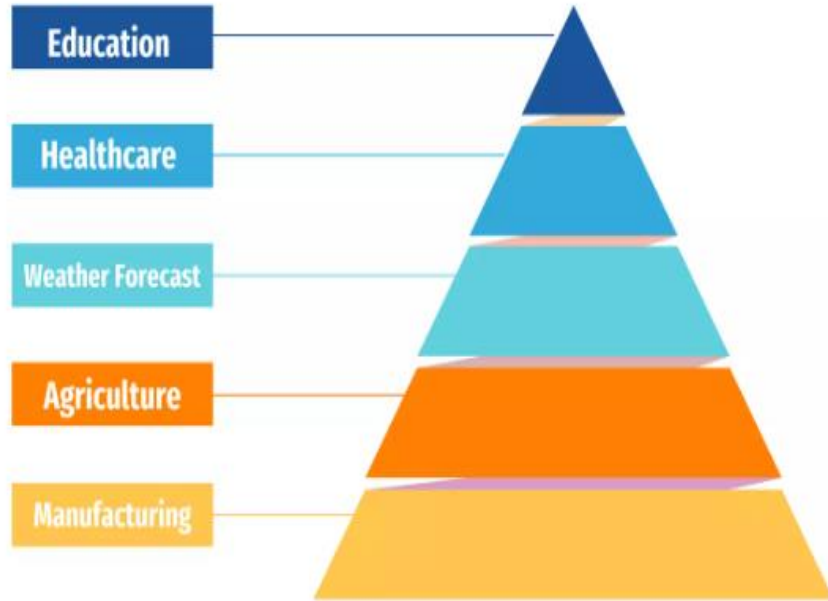




Importance of Big Data in Modern Industries



Applications of Big Data



Application of Big Data

Education



Enhancing Student Results

A better Grading System

Gaining Attention

Customized Programs

Reducing The Number of Dropouts

Application of Big Data

Agriculture



Boosting productivity

Access to plant genome information

Predicting yields

Risk management

Food safety

Savings

Application of Big Data Analysis

Manufacturing



Product quality

Defects tracking

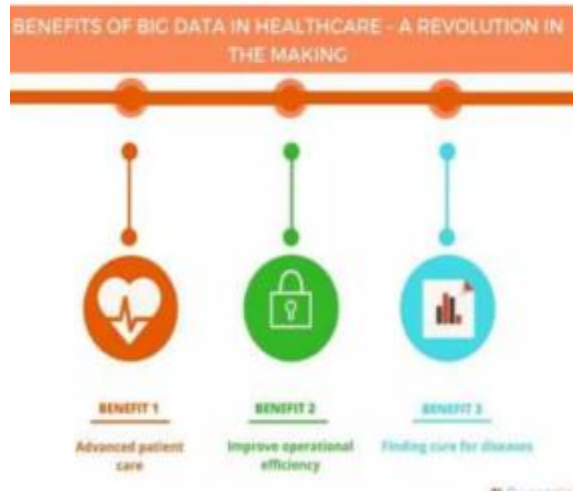
Supply planning

Manufacturing process defect tracking

Testing and simulation of new
manufacturing processes

Application of Big Data Analysis

HealthCare



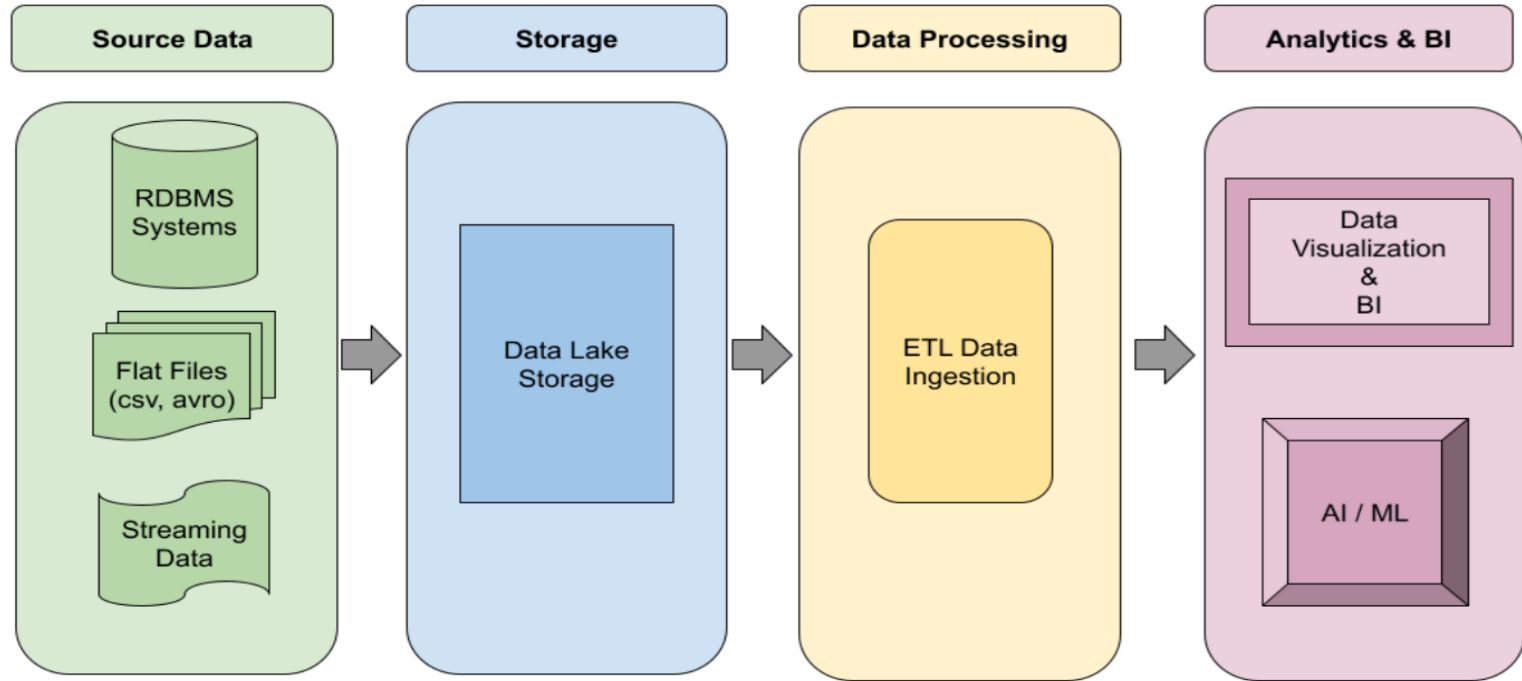
Health Tracking

Improve the care delivery system / machinery

Fraud detection and prevention

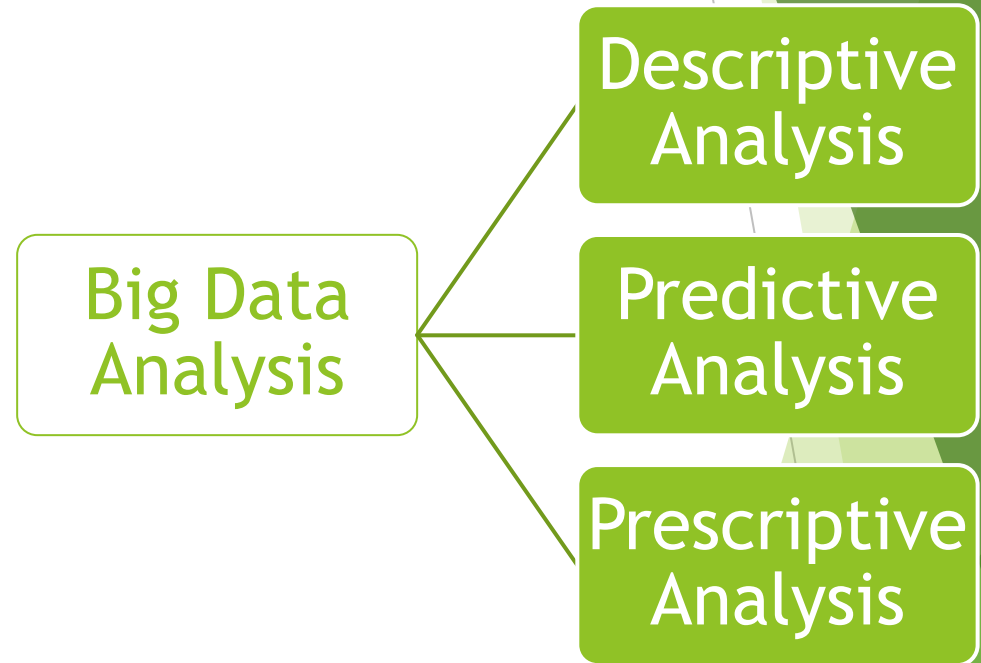
Real-time alerts

Architectural Diagram:



What is Big data Analysis?

Big Data Analysis involves examining large, complex datasets to uncover patterns, trends, and insights. Using tools like machine learning, AI, and statistical models, it processes vast volumes of structured and unstructured data quickly. Applied in fields like healthcare, finance, and marketing, it aids in data-driven decision-making and improving efficiency. Platforms like Hadoop and Spark are key to managing and analyzing big data.



Types of analysis:

- ▶ Descriptive Analysis
- ▶ Predictive Analysis
- ▶ Prescriptive Analysis



Predictive Analysis:

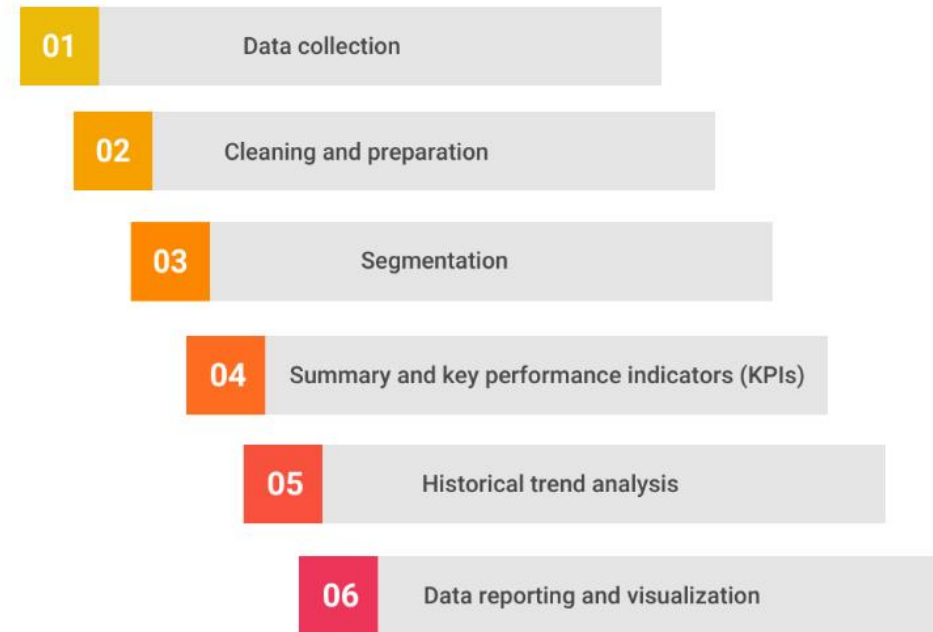
- ▶ **Focus:** Uses statistical models and machine learning to forecast future outcomes based on historical data.
- ▶ **Purpose:** Predicts *what is likely to happen*.
- ▶ **Example:** Predicting customer churn or future sales based on past behaviors.



Descriptive Analysis:

- ▶ **Focus:** Summarizes historical data to identify trends and patterns.
- ▶ **Purpose:** Provides insights into *what happened* using tools like dashboards, charts, and reports.
- ▶ **Example:** Analyzing sales data to understand seasonal trends.

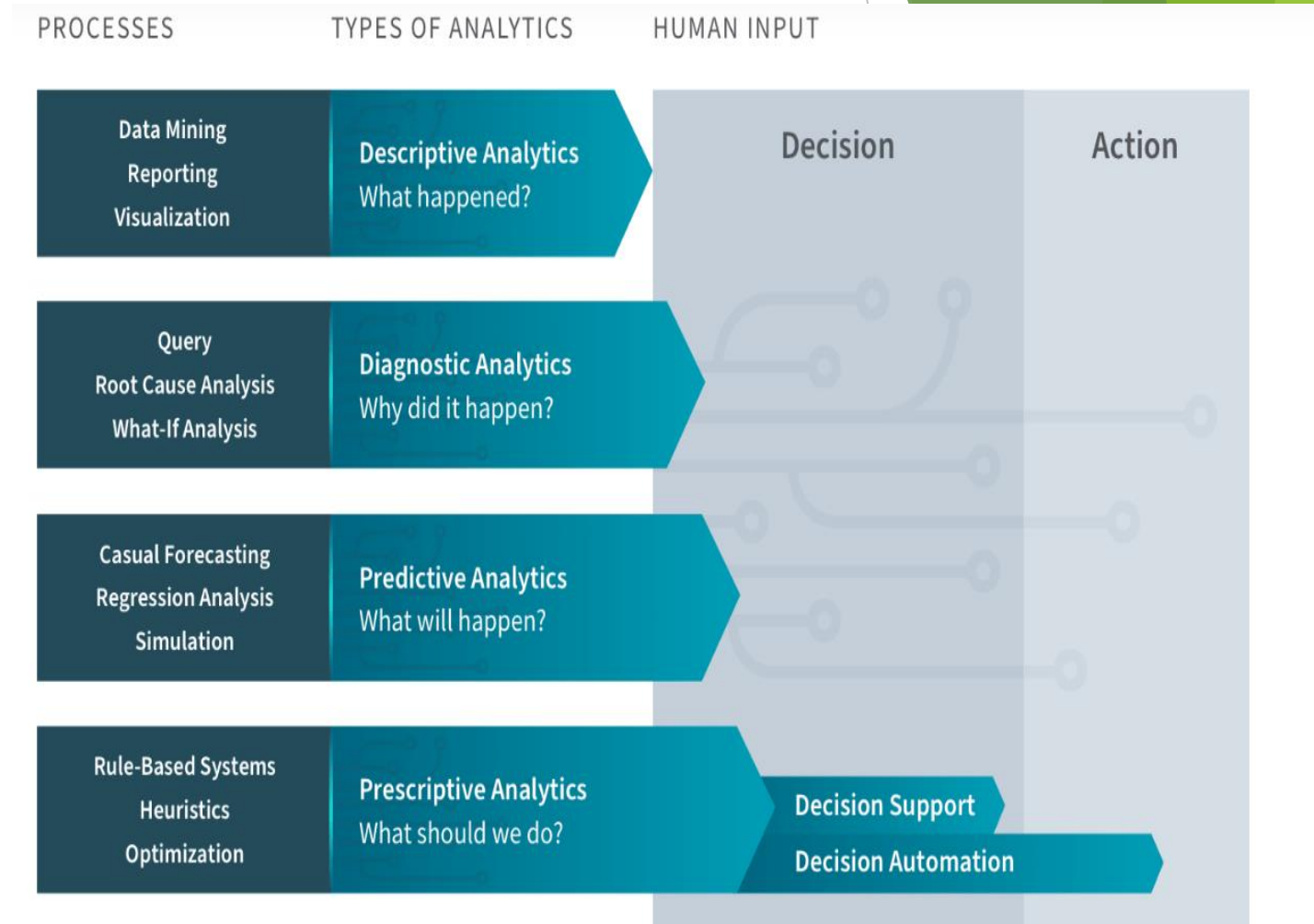
Descriptive Analytics Process



Descriptive Analytics Process

Prescriptive Analysis:

- **Focus:** Recommends actions based on predictions and optimization algorithms.
- **Purpose:** Suggests *what should be done* to achieve desired outcomes.
- **Example:** Recommending the best supply chain strategy to minimize costs and maximize efficiency.



Big Data Tools & Technologies:

- Data Storage Tools.
- Data Processing Tools.
- Data Analytics Tools.
- Visualizations Tools.



Data Storage Tools:

Hadoop

- An open-source framework for distributed storage and processing of large datasets.
- Key Components:
 - **HDFS (Hadoop Distributed File System):** Stores data across multiple nodes.
 - **MapReduce:** Processes large-scale data in parallel.
- **Use Case:** Batch processing of massive datasets in industries like finance and retail.

Apache HBase

- A distributed, open-source NoSQL database designed for real-time read and write operations on large datasets.
- Key Features:
 - Built on top of Hadoop for seamless integration with HDFS.
 - Provides random, real-time access to structured data.
 - Suitable for sparse data with billions of rows and millions of columns.
- **Use Case:** Real-time analytics on massive datasets, such as tracking user behavior through clickstream data.

Amazon S3

- A scalable, cloud-based object storage service designed for storing and retrieving any amount of data from anywhere.
- **Key Features:**
 - Provides high durability and availability for data storage.
 - Supports a wide range of data formats, including unstructured data.
 - Integrates with various big data tools for seamless data processing.
- **Use Case:** Efficiently storing and archiving massive amounts of unstructured data, such as multimedia files, logs, and backups.

Data Processing Tools:

Apache Spark

- A fast, in-memory data processing engine for large-scale data analysis.
- Features:** Supports batch processing, real-time data streaming, and machine learning.
- Use Case:** Fraud detection in financial transactions.



Apache Storm

- A distributed real-time processing system.
- Use Case:** Analyzing social media feeds in real time.



APACHE
HTTP SERVER

Apache Kafka

- A distributed messaging system for real-time data pipelines and streaming.
- Use Case:** Streaming sensor data in IoT applications.

Data Analytics Tools:

R and Python

- Programming languages with extensive libraries for statistical analysis and machine learning.
- **Use Case:** Predictive modeling, statistical testing, and machine learning applications.

TensorFlow and PyTorch

- Frameworks for building and deploying machine learning and deep learning models.
- **Use Case:** Image recognition, natural language processing, and recommendation systems.

RapidMiner

- A data science platform offering end-to-end machine learning workflows.
- **Use Case:** Predictive analytics in marketing campaigns

Data Visualization Tools:

Tableau

- A powerful tool for creating interactive dashboards and visualizations.
- Use Case:** Analyzing and presenting sales performance data.

Power BI

- A Microsoft tool for integrating and visualizing big data from multiple sources.
- Use Case:** Business performance monitoring.

D3.js

- A JavaScript library for creating custom, interactive data visualizations in web browsers.
- Use Case:** Visualizing complex network data

Big Data Challenges

- No organization can function without huge amount of data these days. But, there are some challenges of Big Data encountered by companies.
- These include data quality, storage, lack of data science professionals, validating data, and accumulating data from different sources.

Lack of knowledge Professionals

These professionals will include data scientists, data analysts and data engineers who are experienced in working with the tools and making sense out of huge data sets.

Solution

Investing more money in the recruitment of skilled professionals.

Lack of proper understanding of Massive Data

- Companies fail in their Big Data initiatives, all thanks to insufficient understanding.
- For example, if employees do not understand the importance of data storage, they might not keep the backup of sensitive data.

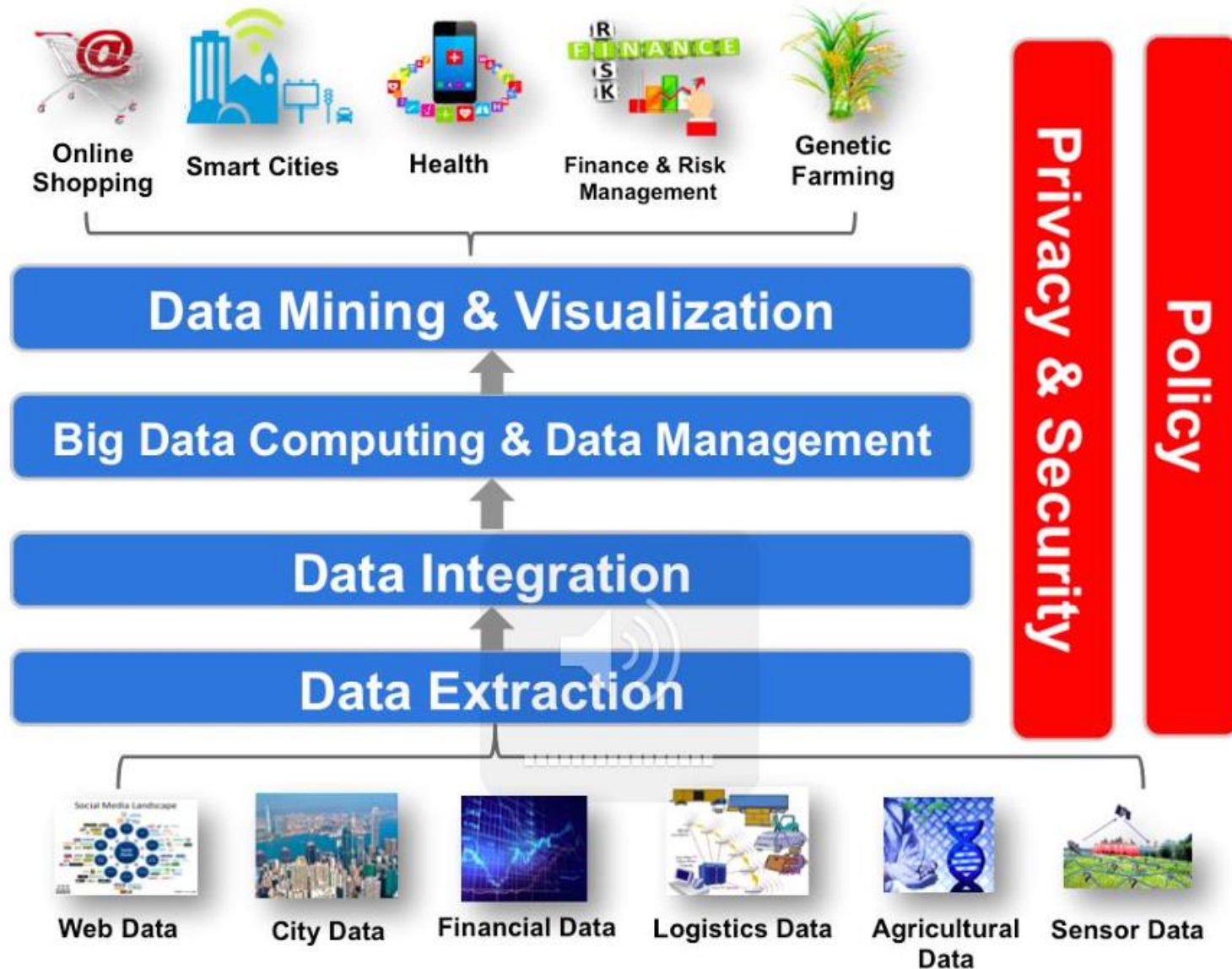
Solution

Basic training programs must be arranged for all the employees who are handling data regularly

Data Growth Issues

- As these data sets grow exponentially with time, it gets challenging to handle.
- Data and analytics fuels digital business and plays a major role in the future survival of organizations worldwide. Source: Gartner, Inc

Big Data Analytics Objectives



How big data analytics works

Once the data is ready, it can be analyzed with the software commonly used for advanced analytics processes. That includes tools for:

- **data mining**, which sift through data sets in search of patterns and relationships;
- **predictive analytics**, which build models to forecast customer behavior and other future developments;
- **machine learning**, which taps algorithms to analyze large data sets; and
- **deep learning**, a more advanced offshoot of machine learning.

Confusion while Big Data tool selection

These confusions bother companies and sometimes they are unable to find the answers. They end up making poor decisions and selecting an inappropriate technology.

Solution

The best way to go about it is to seek professional help.

Data security

- Securing these huge sets of knowledge is one of the daunting challenges of massive data.
- Companies can lose up to \$3.7 million for a stolen record or a knowledge breach.

Solution

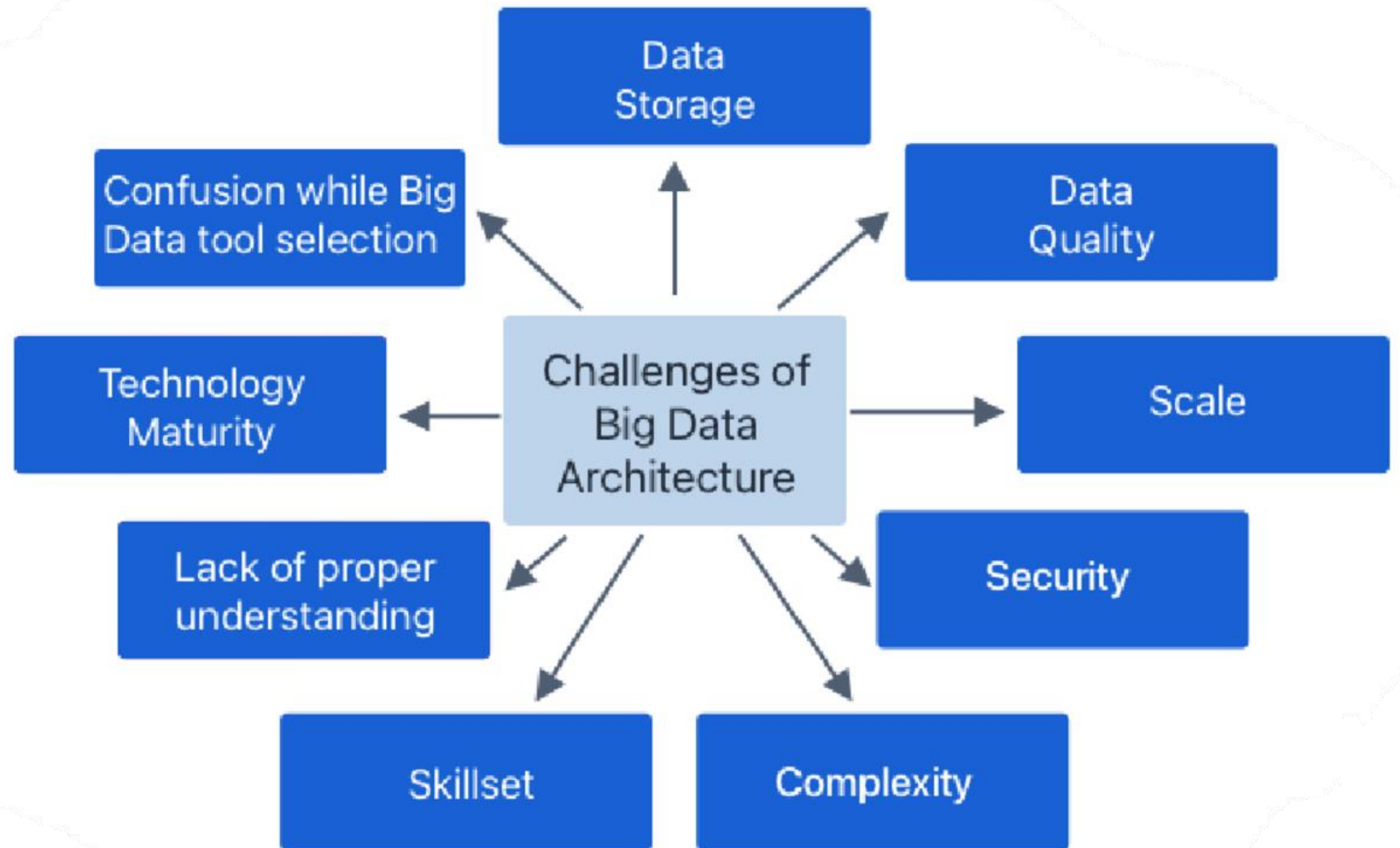
Companies should recruit more cybersecurity professionals to guard their data.

Integrating data from a variety of sources

- Data in a corporation comes from various sources, like social media pages, customer logs, financial reports, e-mails, and reports created by employees.
- Combining all this data to organize reports may be a challenging task.

Solution

Data Integration problems solves by purchasing the proper tools.



Future of Big Data Analysis

Trends

AI and Machine Learning:

AI algorithms analyze vast datasets to identify patterns, improve predictions, and support decision-making. Examples include recommendation systems (Netflix) and fraud detection.

Edge Computing:

Processing data closer to the source (e.g., IoT devices) reduces latency, enhances real-time analytics, and complements cloud-based systems.

Opportunities

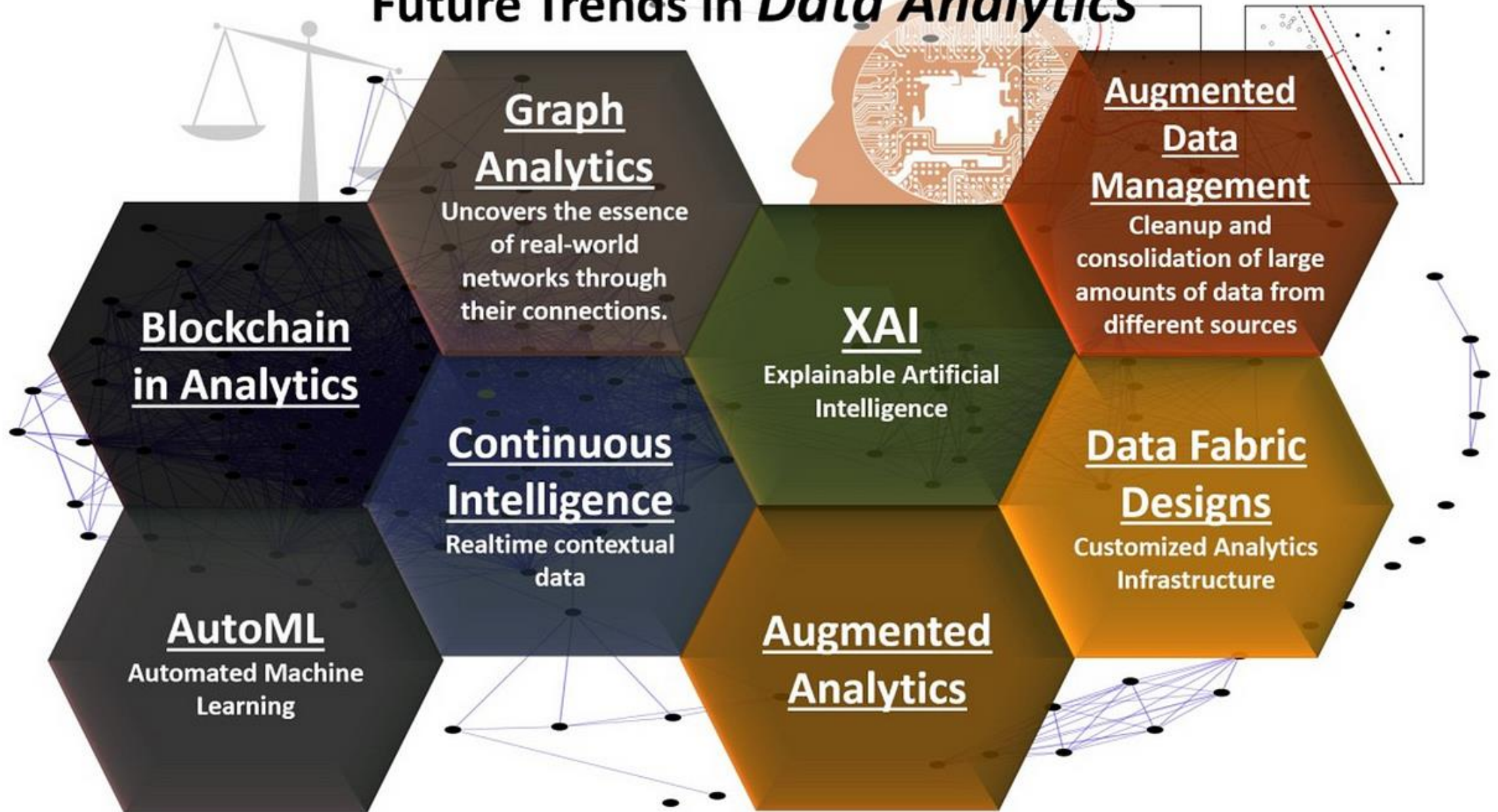
Improved Decision-making:

Big data drives efficiency and cost reduction across industries like healthcare, finance, and logistics.

Enhanced Customer Experience:

Personalization powered by big data (e.g., tailored Netflix recommendations or Amazon shopping suggestions) boosts satisfaction and retention.

Future Trends in *Data Analytics*



Traditional Data Vs Big Data

Components	Big data	Traditional Data
Queries	Largely Abandoned SQL	Traditional SQL
Architecture	Distributed	Centralized
Data Types	Structured, Semi- Structured and Unstructured	Structured
Data Model	No Schema	Fixed Schema
Data Relationship	Unknown or complex	Known Relationship
Data volume	Petabytes or Exabytes	Terabytes
Data Traffic	More	Less
Data Integrity	Less	High

Conclusion

Challenges: Privacy, integration, scalability, and skill gaps.

Future Trends: AI, machine learning, and edge computing.

Opportunities: Enhanced decision-making and improved customer experiences.

Importance of Big Data:

1. Drives business strategies and innovation.
2. Solves real-world problems across industries.

Opportunities:

1. Enhanced Decision-Making
2. Improved Customer Experiences



Thank You



Question1: A healthcare provider collects vast amounts of patient data through electronic health records (EHR), wearable devices, and lab reports. The data includes structured formats like patient demographics and lab results, as well as unstructured formats like doctor's notes and medical images. The provider aims to improve patient outcomes by predicting potential health risks (like heart attacks or diabetes) and personalizing treatment plans. They also want to ensure data privacy and comply with regulations like GDPR.

How can predictive analysis using big data tools help the healthcare provider improve patient outcomes? Include the steps involved in processing structured and unstructured data, and discuss the importance of maintaining data privacy.

Question 2: A multinational corporation operates in various industries, such as manufacturing, retail, and logistics. They face challenges like understanding global customer preferences, identifying seasonal trends in sales, and optimizing supply chain operations. To address these, they decide to utilize big data, focusing on its characteristics and leveraging descriptive and predictive analysis to gain insights.

How can the characteristics of big data (Volume, Velocity, Variety, Veracity, and Value) be applied to solve business challenges in global supply chain optimization? Illustrate your answer using examples of descriptive and predictive analysis to address customer trends and seasonal demands.