ADC 405

Internet of Things

Module – 5

IoT & Data Analytics

BY

Mrs. Kanchan Chavan

Data Analytics for IoT

- act of analyzing data generated and collected from IoT devices by utilizing a specific set of data analytics tools and techniques

- process of analyzing unstructured data to give meaningful conclusions

- to turn vast quantities of unstructured data from various devices and sensors within the Internet of Things ecosystem, which is heterogeneous, into valuable and actionable insights for
    - driving sound business decision-making
    - further data analysis
    - enables identifying the patterns in data sets
    - make predictions and adjustments about future events - healthcare, retail, and eCommerce to manufacturing, transportation, and more

Linked Analytics Data Sets

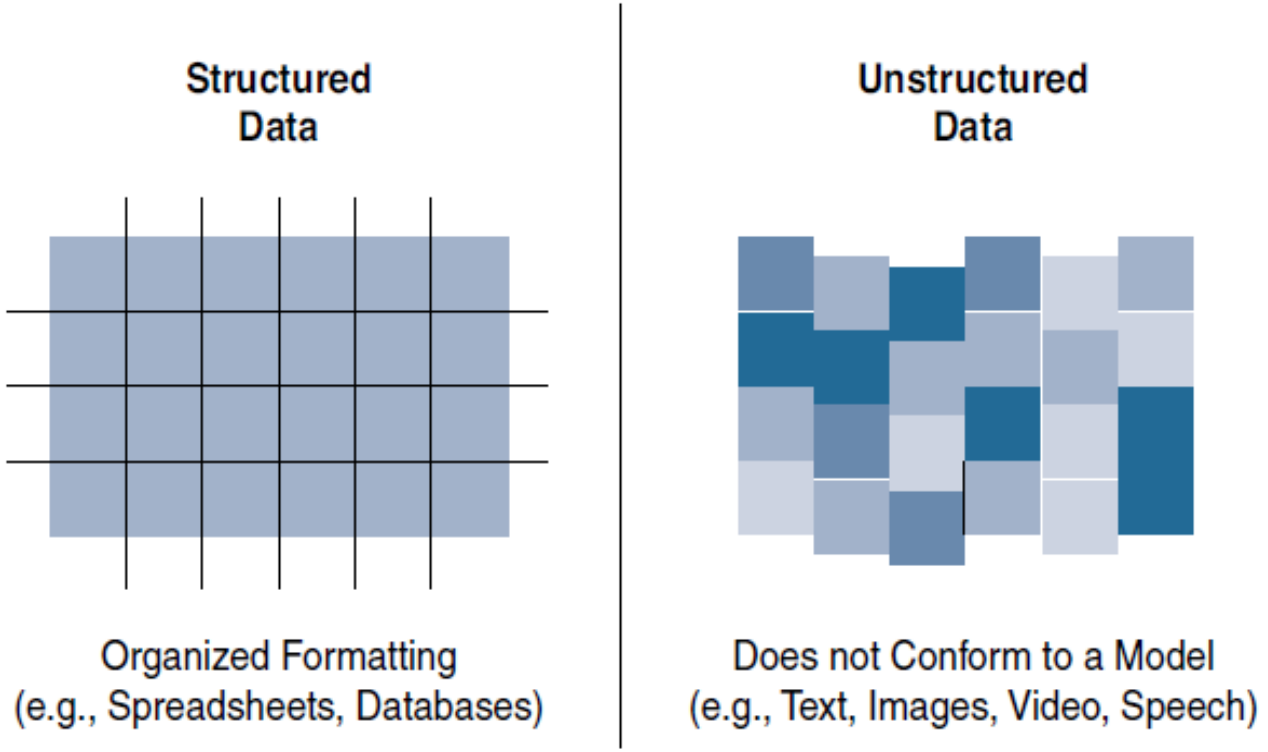

Figure 7-2 *Comparison Between Structured and Unstructured Data*

Different Types of IoT Data Analytics

- **Descriptive Analytics**
  - focus on what happened in the past
  - historical data collected from devices are processed and analyzed to generate a report
  - describes what took place, when it occurred, and how often it did

- **Diagnostic Analytics**
  - drilling down into the data to identify the root cause of a specific issue
  - make use of techniques like data mining and statistical analysis
  - offer actionable insights into the causes of specific problems

- **Predictive Analytics**
  - predict future events by analyzing historical data and trends
  - makes use of various statistical and machine learning algorithms to build models
  - plays a significant role in supporting business decisions

- **Prescriptive Analytics**
  - predicts what will happen in the future
  - provides recommendations on what should be done to achieve the desired business outcomes
  - makes use of optimization algorithms to identify the best course of action

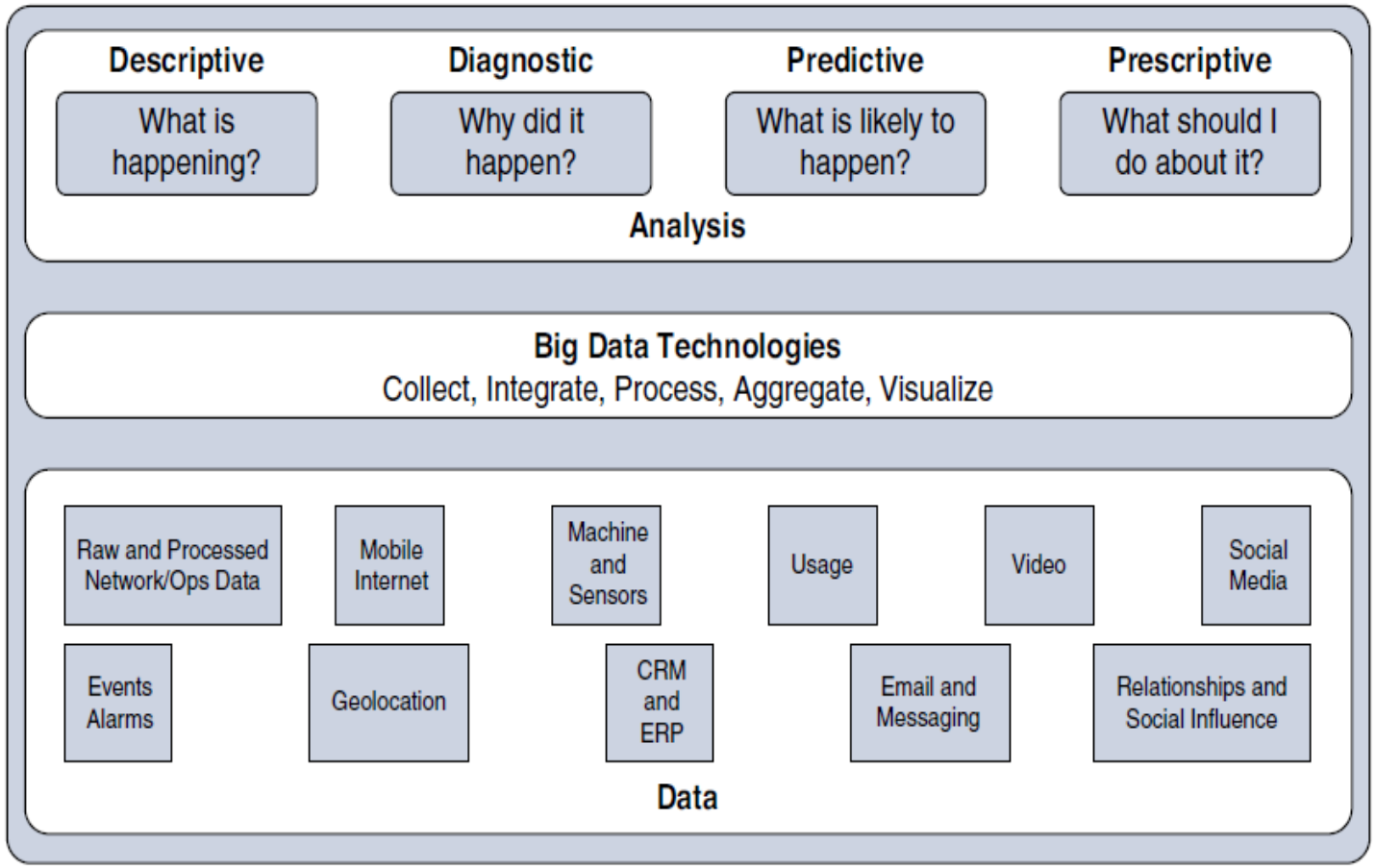

Figure 7-3 *Types of Data Analysis Results*

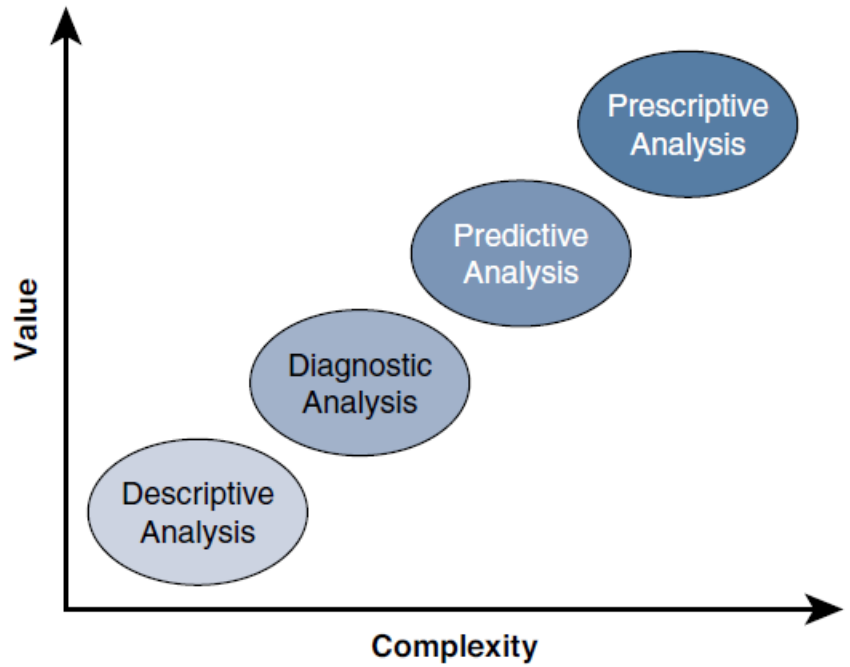

Figure 7-4 *Application of Value and Complexity Factors to the Types of Data Analysis*

Business Benefits of IoT Analytics

- **Optimizing Operational Efficiency**
  to identify issues and problems that lead to inefficiencies
  food & beverage company - track the temperature of its refrigerators in real-time & prevent food spoilage
  smart power grid

- **Reducing Costs**
  monitor the performance of its production line and make adjustments to avoid wastage of materials

- **Enhancing Customer Experience**
  to collect and analyze customer data in order to understand their needs and preferences

- **Improving Safety**
  identify potential safety hazards and take preventive measures

Technical Benefits of IoT Analytics

- **Real-time Data Analysis**
  ability to analyze real-time data points

- **Improved Scalability**
  IoT data analytics can be deployed on the cloud
  businesses can scale up their operations quickly and easily without incurring any additional costs

- **Increased Accuracy**

- **Enhanced Security**
  identify and track potential threats and then take measures to avoid them

Implement IoT Analytics

- Determine The Use Cases
- Data Collection
- Data Storage
- Data Visualization
- Data Analysis

IoT Analytics Tools

- Microsoft Azure IoT Analytics
- AWS IoT Analytics
- IBM Watson IoT Platform
- Google Cloud IoT Core and Cloud IoT Analytics
- Thing Speak
- Predix(by GE Digital)

Challenges of IoT Data Analytics

- **Data Security**
  vulnerable to cyber-attacks

- **Data Privacy**
  collect personal data such as location, health, and behaviour

- **Data Quality**
  noisy and inconsistent

- **Scalability**
  data is generated at a massive scale

- **Interoperability**
  IoT devices come from different manufacturers and have different protocols & standards

Managing Data lakes

- A data lake is a central location that holds a large amount of data in its native, raw format
- A data lake is a centralized repository for storing a wide variety of data types, including structured, semi-structured, and unstructured data
- **data warehouse, which stores data in files or folders**
- 1. Data Ingestion **-** The process of bringing data into the data lake **,** batch data uploads, streaming data from IoT devices, or data from external sources like APIs
- 2. Storage - use distributed storage systems to store data, Hadoop Distributed File System (HDFS), cloud-based storage services like Amazon S3, Azure Data Lake Storage, and Google Cloud Storage
- 3. Data Processing - data processing frameworks like Apache Spark, Apache Hadoop, and Apache Flink, enable you to perform data transformations, analytics, and other processing tasks
- 4. Metadata Management – means “data about data” - tools like Apache Atlas, AWS Glue, and Azure Data Catalog help organize and annotate data

Managing Data lakes

- 5. Data Governance and Security - to ensure data quality and regulatory compliance, Tools like Apache Ranger and cloud-native identity
- 6. Data Analytics and Visualization - Apache Zeppelin, Jupyter Notebooks, Tableau, and Power BI
- 7. Machine Learning and AI
- 8. ETL (Extract, Transform, Load)

Data Retention Strategy

- refers to the practice of storing data for a specific period of time
- Data retention policies concern what data should be stored or archived, where that should happen, and for exactly how long
- Once the retention time period for a particular data set expires, it can be deleted or moved as historical data to secondary or tertiary storage, depending on the requirements

IoT Data Visualization

- represents information through visual elements like charts, graphs, and diagrams.
- to quickly identify patterns, trends, and insights
- enables us to make informed decisions, solve problems