Module - 01

Introduction to Big Data Analytics

BIG DATA OVERVIEW

STATE OF PRACTICE IN ANALYTICS

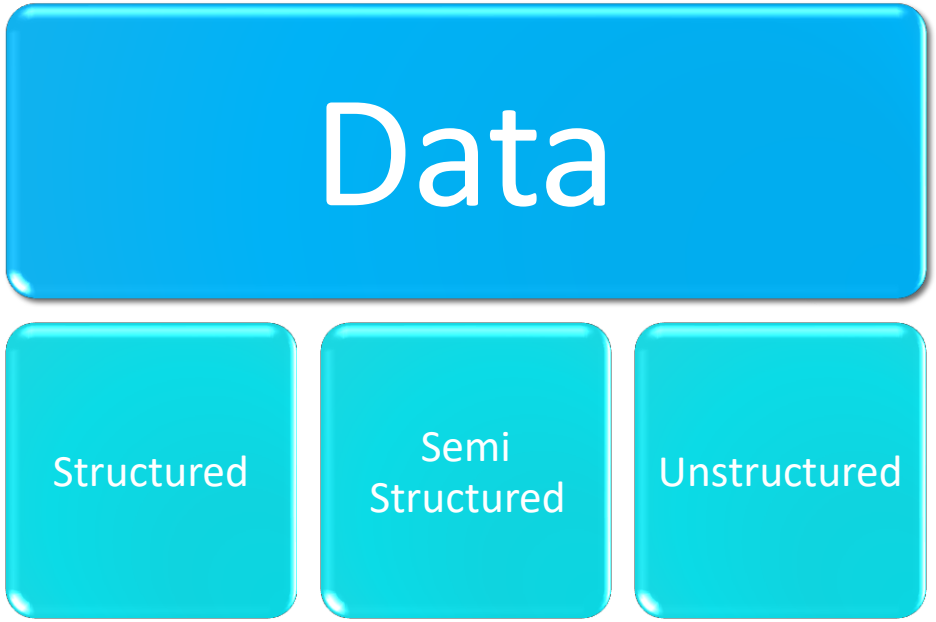
ROLE OF DATA SCIENTISTS

EXAMPLES OF BIG DATA ANALYTICS

DATA ANALYTICS LIFE CYCLE

Introduction to Big Data

Big data is data that contains greater variety, arriving in increasing volumes and with more velocity.



Name	Equal to:	Size in Bytes
Bit	1 bit	1/8
Nibble	4 bits	1/2 (rare)
Byte	8 bits	1
Kilobyte	1,024 bytes	1,024
Megabyte	1,024 kilobytes	1,048,576
Gigabyte	1,024 megabytes	1,073,741,824
Terrabyte	1,024 gigabytes	1,099,511,627,776
Petabyte	1,024 terrabytes	1,125,899,906,842,624
Exabyte	1,024 petabytes	1,152,921,504,606,846,976
Zettabyte	1,024 exabytes	1,180,591,620,717,411,303,424
Yottabyte	1,024 zettabytes	1,208,925,819,614,629,174,706,176

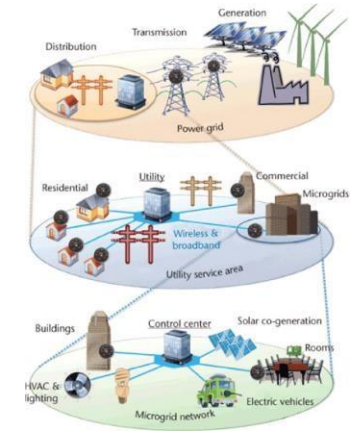
Types of Big Data

Evolution of big data

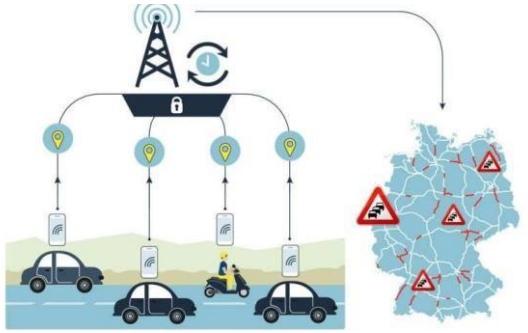
Big Data

Social Media

Stock Exchange



Power Grid



Transport

Healthcare

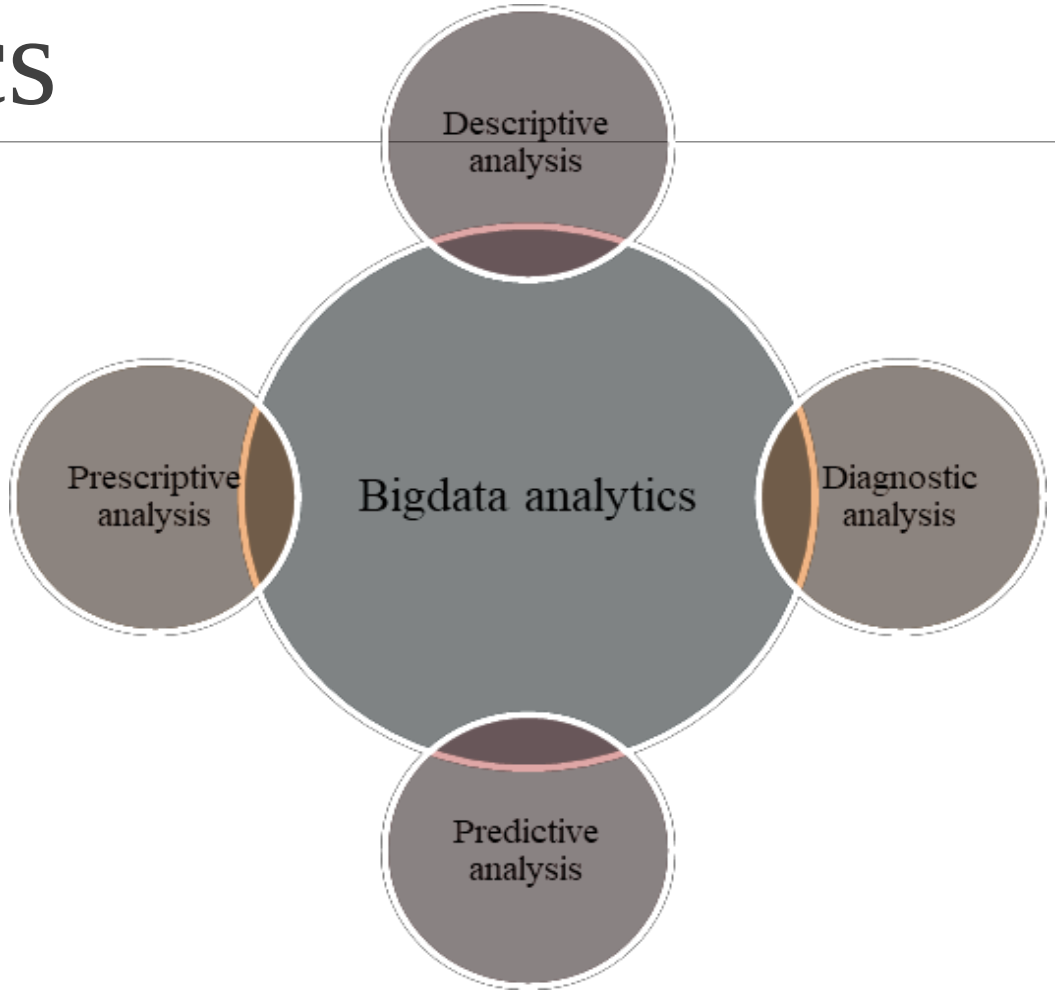
Introduction to Big Data

- Volume - Ever growing nature.
- Velocity - Speed of data generating.
- Variety - Different types of data.
- Veracity - Accuracy and quality.
- Value - Economic value of data

Big Data Challenges

- Lack of Understanding How to Work with Big Data.
- Poor Data Quality and Data Silos.
- Big Data Issues in Scaling.
- Variety of Big Data Technologies.
- Incorrect Integration.
- Heavy Expenses.
- Real-Time Big Data Problems.
- Data Verification.

Big Data Analytics



Big Data Analytics Tools

- Hadoop - helps in storing and analyzing data
- MongoDB - used on datasets that change frequently
- Talend - used for data integration and management
- Cassandra - a distributed database used to handle chunks of data
- Spark - used for real-time processing and analyzing large amounts of data
- STORM - an open-source real-time computational system
- Kafka - a distributed streaming platform that is used for fault-tolerant storage

Big Data Industry Applications

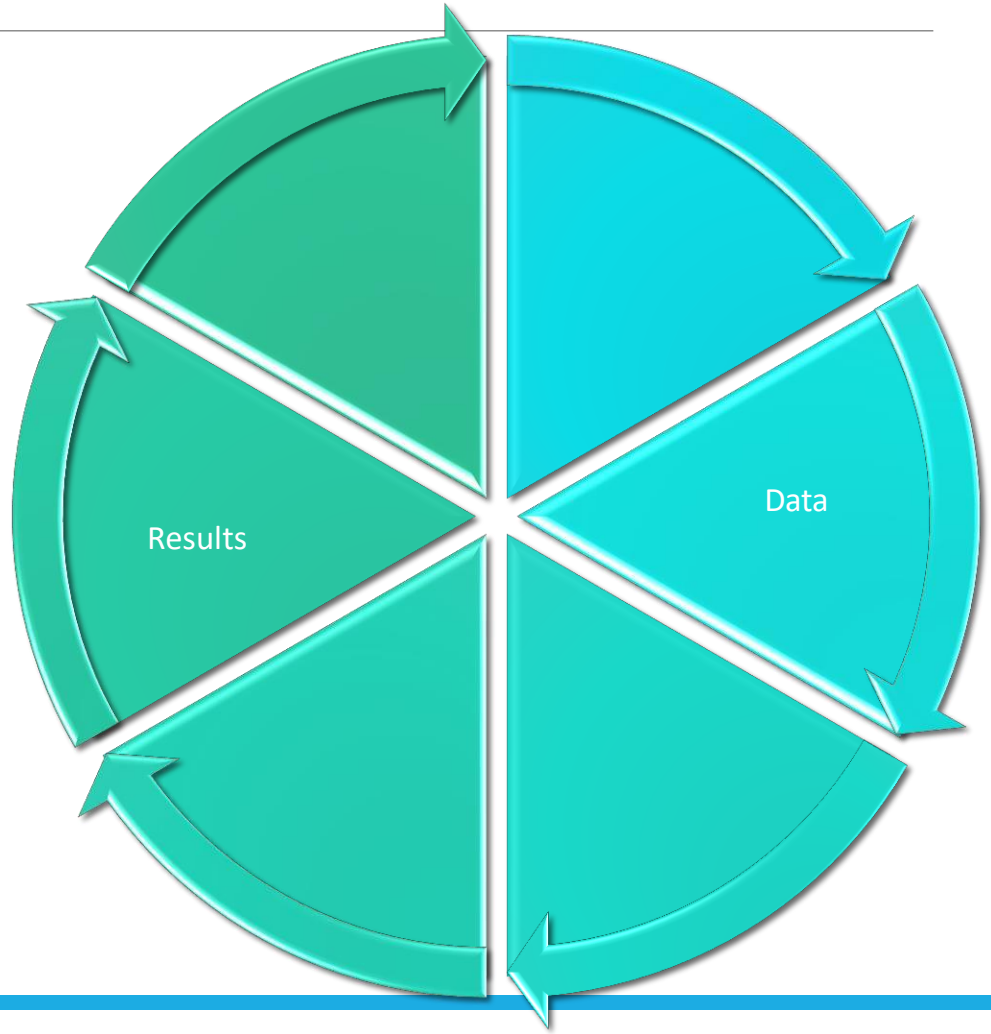
- **Ecommerce -** Predicting customer trends and optimizing prices are a few of the ways e-commerce uses Big Data analytics
- **Marketing -** Big Data analytics helps to drive high ROI marketing campaigns, which result in improved sales
- **Education -** Used to develop new and improve existing courses based on market requirements
- **Healthcare -** With the help of a patient's medical history, Big Data analytics is used to predict how likely they are to have health issues

Big Data Industry Applications

- **Media and entertainment -** Used to understand the demand of shows, movies, songs, and more to deliver a personalized recommendation list to its users
- **Banking -** Customer income and spending patterns help to predict the likelihood of choosing various banking offers, like loans and credit cards
- **Telecommunications -** Used to forecast network capacity and improve customer experience
- **Government -** Big Data analytics helps governments in law enforcement, among other things

Customer Relationship Management - Style 1

Data Analytics Lifecycle



Role of Data Scientists

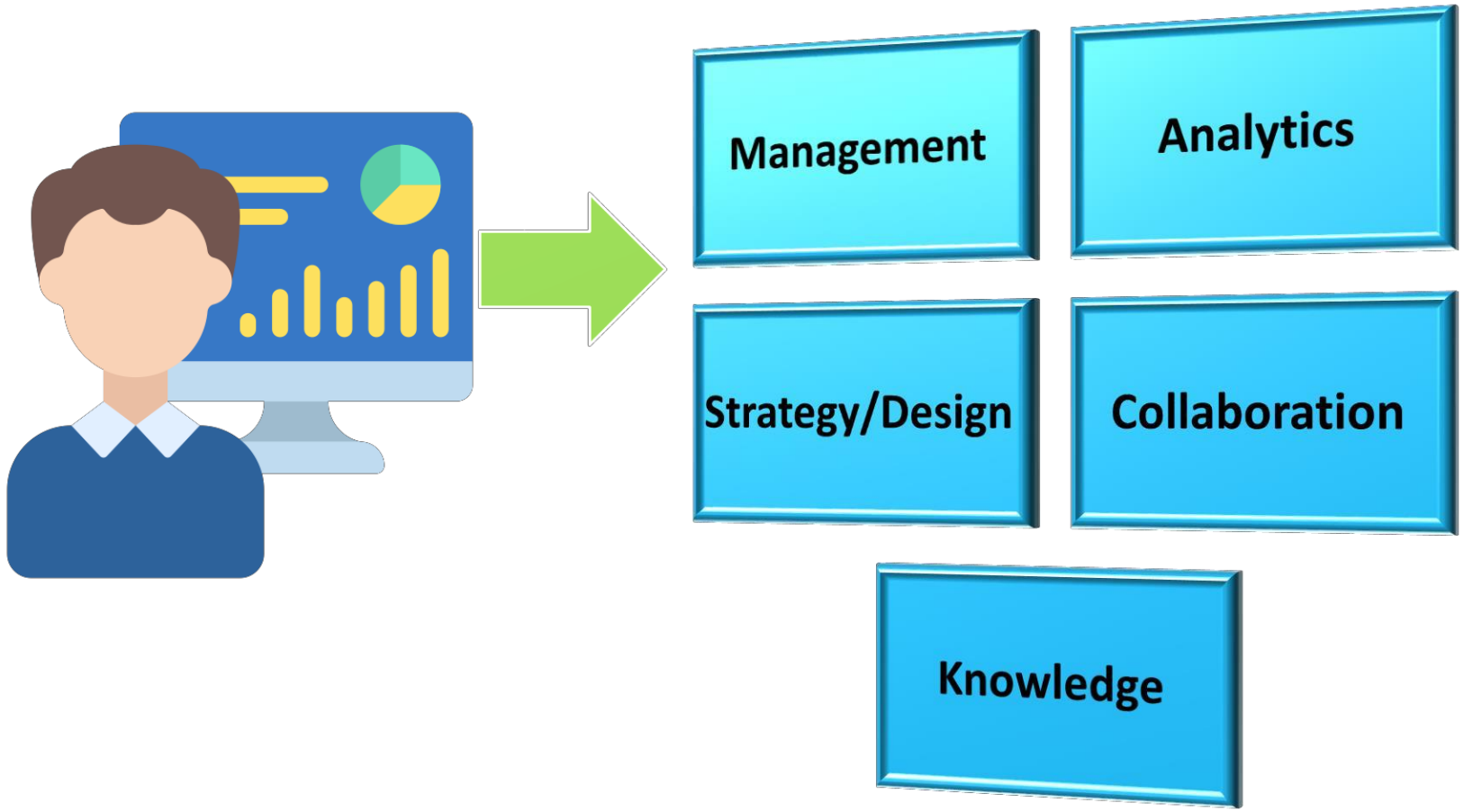
Introduction

- Big Data emerged when organizations are dealing with petabytes and exabytes of data
- For Organizations, storage is difficult without the proper technologies like Hadoop.
- ***Data science is the science of analyzing raw data using statistics and machine learning techniques with the purpose of drawing conclusions about that information.***
- A data architect defines the tools and the architecture the data would be stored at, whereas a data scientist uses this architecture.

Required skills for a data scientist

- Programming in a statistical package such as: R, Python, SAS, SPSS, or Julia
- Able to clean, extract, and explore data from different sources
- Research, design, and implementation of statistical models
- Deep statistical, mathematical, and computer science knowledge

Role of Data Scientists



Role of Data Scientists

Identifying the data-analytics problems that offer the greatest opportunities to the organization

Determining the correct data sets and variables

Collecting large sets of structured and unstructured data from disparate sources

Cleaning and validating the data to ensure accuracy, completeness, and uniformity

Devising and applying models and algorithms to mine the stores of big data

Analyzing the data to identify patterns and trends

Interpreting the data to discover solutions and opportunities

Communicating findings to stakeholders using visualization and other means

Data Analyst Vs Data Scientist

Data Analyst	Data Engineer	Data Scientist
Data Warehousing	Data Warehousing & ETL	Statistical & Analytical skills
Adobe & Google Analytics	Advanced programming knowledge	Data Mining
Programming knowledge	Hadoop-based Analytics	Machine Learning & Deep learning principles
Scripting & Statistical skills	In-depth knowledge of SQL/ database	In-depth programming knowledge (SAS/R/ Python coding)
Reporting & data visualization	Data architecture & pipelining	Hadoop-based analytics
SQL/ database knowledge	Machine learning concept knowledge	Data optimization
Spread-Sheet knowledge	Scripting, reporting & data visualization	Decision making and soft skills

Examples of Big Data Analytics

- Discovering consumer shopping habits
- Personalized marketing
- Finding new customer leads
- Fuel optimization tools for the transportation industry
- User demand prediction for ridesharing companies
- Monitoring health conditions through data from wearables
- Live road mapping for autonomous vehicles
- Streamlined media streaming
- Predictive inventory ordering
- Personalized health plans for cancer patients
- Real-time data monitoring and cybersecurity protocols

References

https://www.simplilearn.com/what-is-big-data-analytics-article

https://nix-united.com/blog/12-big-data-issues-growing-companies-face/

https://www.javatpoint.com/what-is-big-data

https://www.investopedia.com/terms/b/big-data.asp

https://www.designerwomen.co.uk/big-data-analytics-in-agriculture-market-overview-and-analysis-by-top-manufacturers-share-regional-demand-with-growth-forecast/

https://www.oracle.com/in/big-data/what-is-big-data/?source=:ad:pas:go:dg:a_apac:71700000087135169-58700007772107138-p70598024870:RC_WWMK220512P00035C0002:&gclid=Cj0KCQjwzLCVBhD3ARIsAPKYTcQD9GnB7VqWCzp15NWSfGEgcNPC4j0V6FgEi7_KLoEdmy_estBIhRQaAo3tEALw_wcB&gclsrc=aw.ds

References

https://www.educba.com/big-data-analytics-examples/

https://www.geeksforgeeks.org/what-are-the-roles-and-responsibilities-of-a-data-scientist/

https://www.edureka.co/community/29859/what-are-the-roles-of-a-data-scientist

https://www.edureka.co/blog/data-analyst-vs-data-engineer-vs-data-scientist/

https://www.simplilearn.com/tutorials/data-analytics-tutorial/data-analyst-vs-data-scientist

https://builtin.com/big-data/big-data-examples-applications