Advanced Big Data Analytics

Dr. Salahuddin Shaikh

Big Data

01 — What is big data?

02 — What is big data analytics?

03 — Types of big data analytics

04 — Characteristics of big data analytics

05 — Process of big data analytics?

06 — Big data application domains
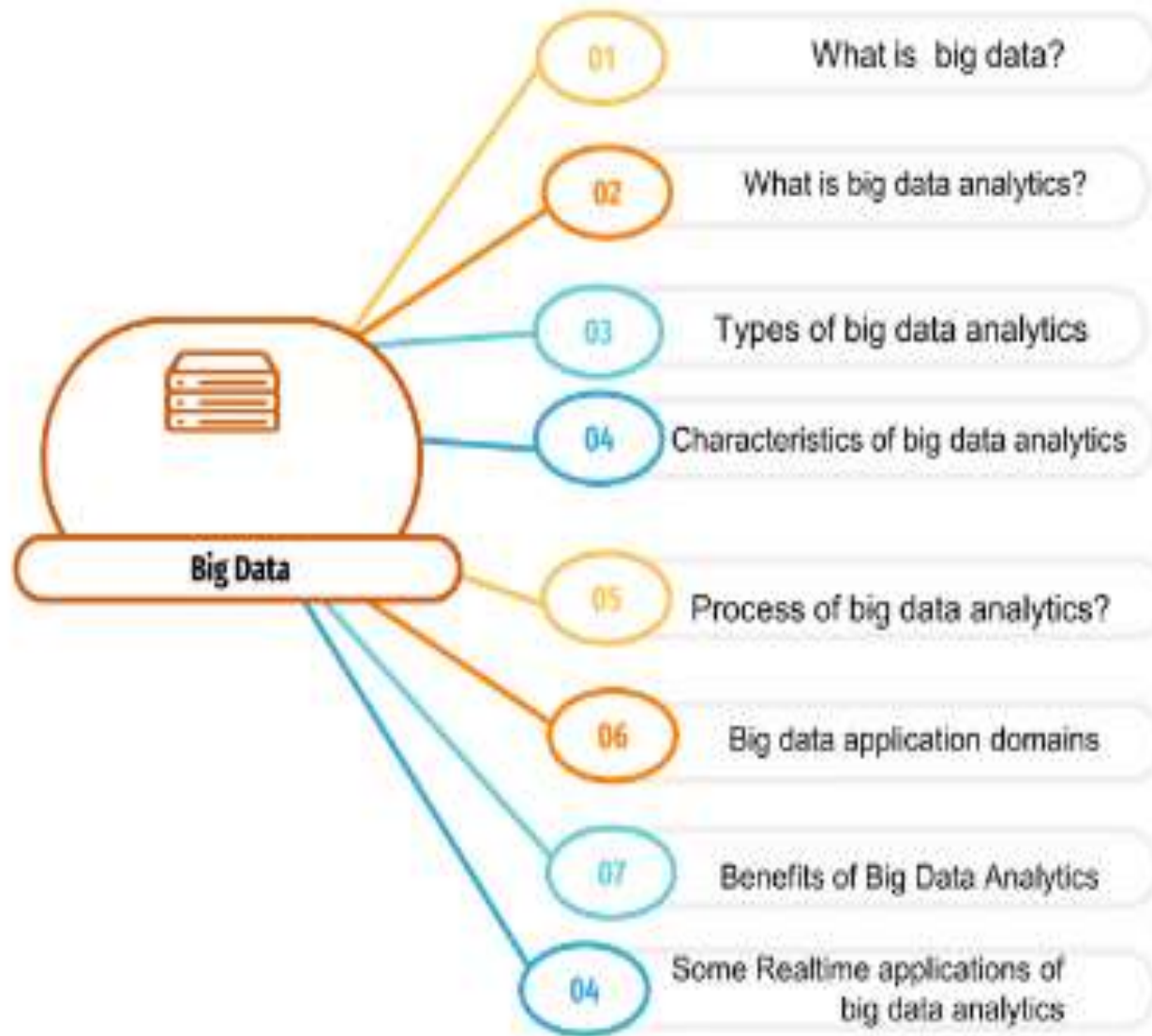
07 — Benefits of Big Data Analytics

04 — Some Realtime applications of big data analytics

What is big data?

Big Data refers to a huge volume of data, that cannot be stored and processed using the traditional computing approach within a given time frame.
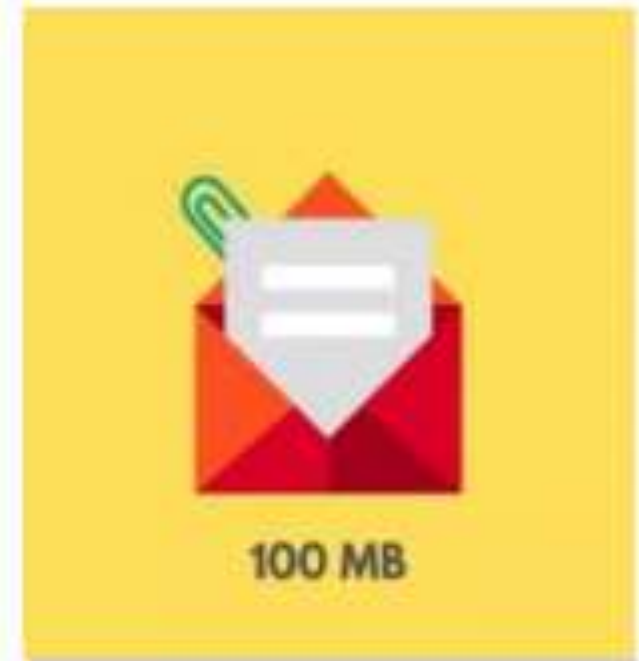
BIG DATA

Example of big data?

For example, if we try to attach a document that is of 100 megabytes in size to an email we would not be able to do so. As the email system would not support an attachment of this size.

Therefore this 100 megabytes of attachment with respect to email can be referred to as Big Data.

100 MB

What is big data Analytics?

Big data Analytics is a process to extract meaningful insight from big such as hidden patterns, unknown correlations, market trends and customer preferences

Types of Big Data Analytics

Structured Data

Unstructured Data

Semi-Structured Data

Structured Data

Structured Data refers to the data that has a proper structure associated with it. For example, the data that is present within the databases, the CSV files, and the excel spreadsheets can be referred to as Structured Data.

STRUCTURED

Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

Unstructured Data

Un-Structured Data refers to the data that does not have any structure associated with it at all. For example, the image files, the audio files, and the video files can be referred to as Un-Structured Data.

Semi-structured Data

Semi-Structured Data refers to the data that does not have a proper structure associated with it. For example, the data that is present within the emails, the log files, and the word documents can be referred to as Semi-Structured Data.

SEMI-STRUCTURED

<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>
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>

Characteristics of Big Data

Big Data is categorized into 3 important characteristics.
- Volume
- Velocity
- Variety

Three Characteristics of Big Data V3s

Volume
- Data quantity

Velocity
- Data Speed

Variety
- Data Types

Volume

Volume refers to the amount of data that is getting generated.

Velocity

Velocity refers to the speed at which the data is getting generated.

Variety

Variety refers to the different types of data that is getting generated.

1st Character of Big Data
Volume

• A typical PC might have had 10 gigabytes of storage in 2000.

• Today, Facebook ingests 500 terabytes of new data every day.

• Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.

• The smart phones, the data they create and consume; sensors embedded into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including video.
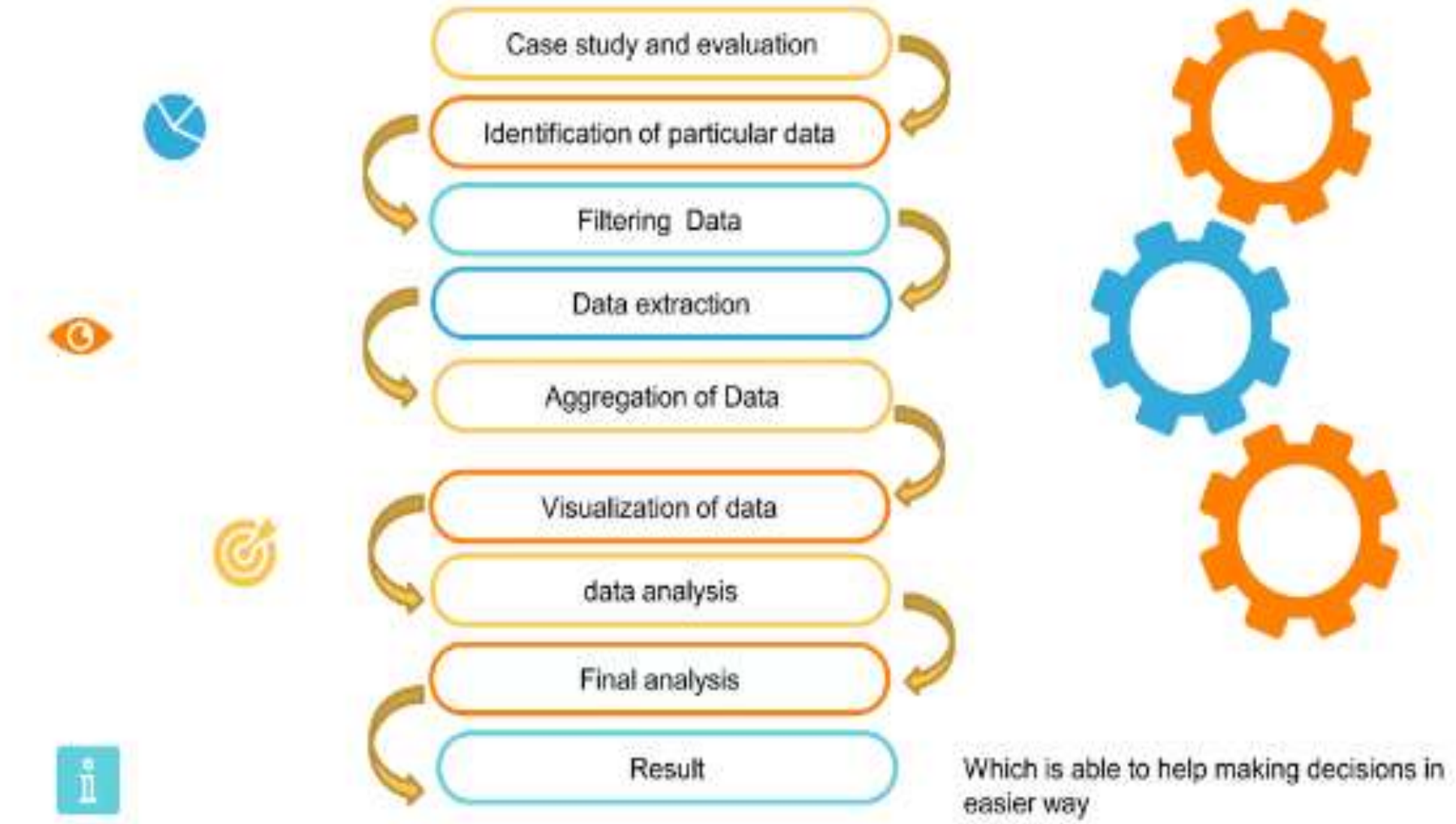
2nd Character of Big Data
Velocity

- Clickstreams and ad impressions capture user behavior at millions of events per second

- high-frequency stock trading algorithms reflect market changes within microseconds

- machine to machine processes exchange data between billions of devices

- infrastructure and sensors generate massive log data in real-time

- on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.

3rd Character of Big Data
Variety

- Big Data isn't just numbers, dates, and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.

- Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.

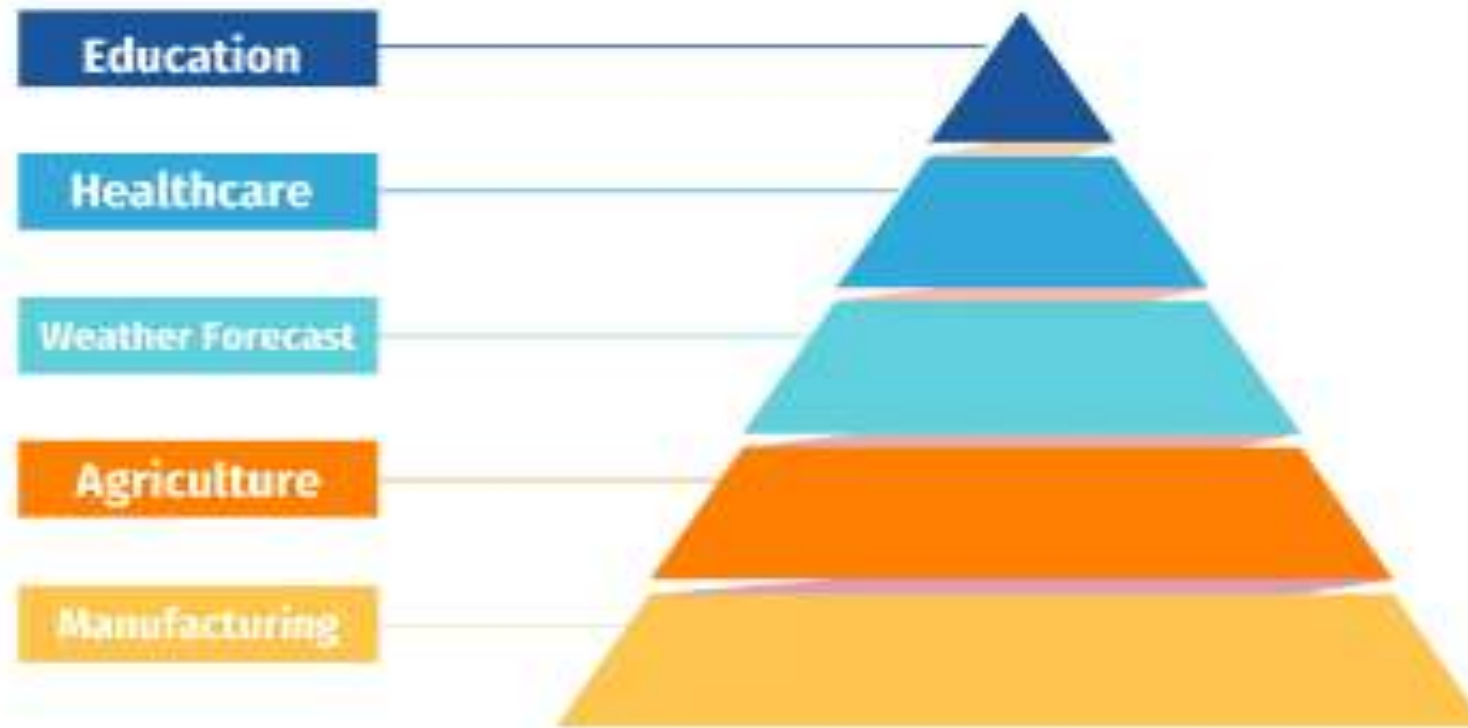- Big Data analysis includes different types of data

Process of Big Data Analytics

- Case study and evaluation
- Identification of particular data
- Filtering Data
- Data extraction
- Aggregation of Data
- Visualization of data
- data analysis
- Final analysis
- Result

Which is able to help making decisions in easier way

Tools used in Big Data Analytics

Tools

MongoDB — MongoDB is used on datasets that change frequently. Tools used in big data analytics

Hadoop — Hadoop helps in storing and analyzing big data Tools used in big data analytics

Talend — Talend is a tool used for data integration and management Tools used in big data analytics

Cassandra — It is a distributed database that is used for handling chunks of data

Storm — It is an open source real time computational system

Spark — It is used for real time processing and analyzing large amount of data

Education

Benefits of Big Data in Education
- Personalized Education
- Choosing a Future Career
- Enhancing Quality of Teaching
- Analyzing Student's Participation
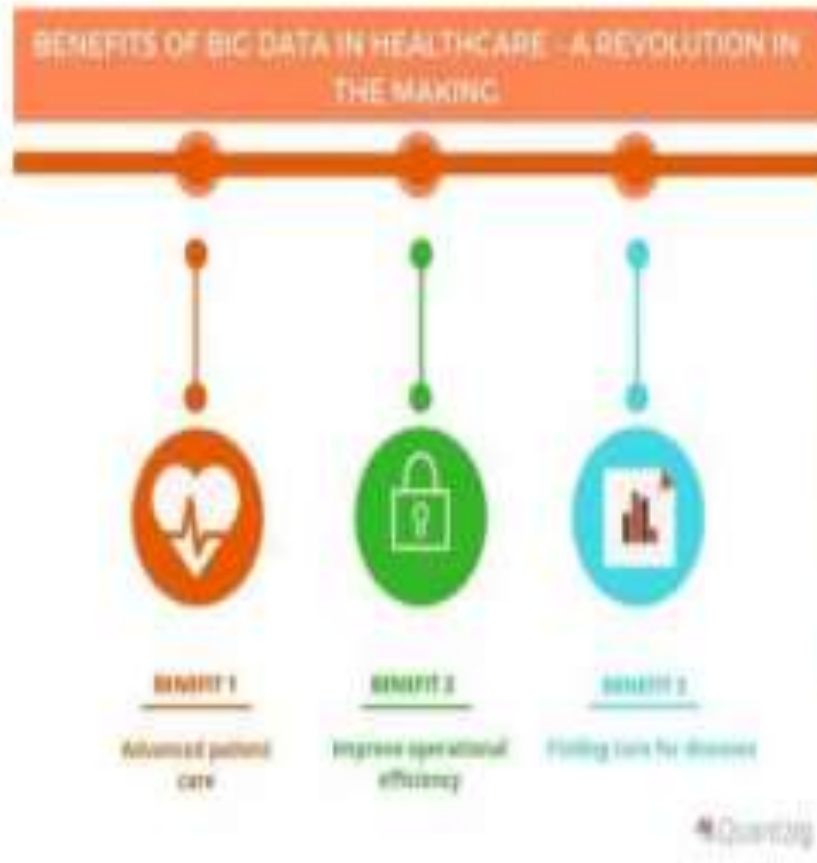
Enhancing Student Results

A better Grading System

Gaining Attention

Customized Programs
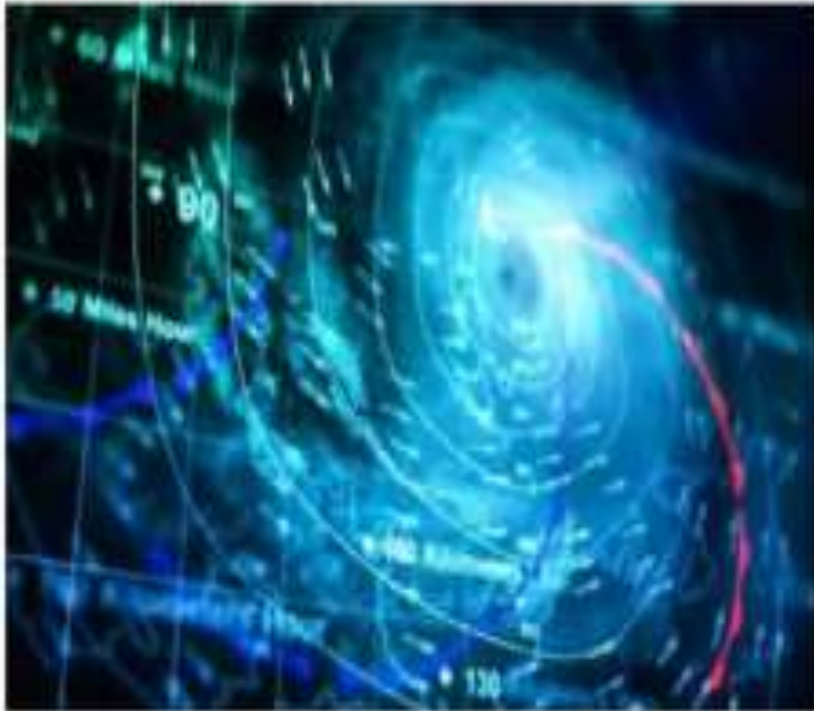
Reducing The Number of Dropouts

HealthCare

Health Tracking

Improve the care delivery system / machinery

Fraud detection and prevention

Real-time alerts

Weather Forecast

- Estimates of areas where flooding is likely to be most severe
- The strength and direction of tropical storms
- The most likely amount of snow or rain that will fall in a specific area
- The most likely locations of downed power lines
- Estimates of areas where wind speeds are likely to be greatest
- The locations where bridges and roads most likely to be damaged by storms
- The likelihood of flights being cancelled at specific airports

Agriculture

Big Data in Agriculture

- Boosting productivity
- Access to plant genome information
- Predicting yields
- Risk management
- Food safety
- Savings

Manufacturing

Product quality

Defects tracking

Supply planning

Manufacturing process defect tracking

Testing and simulation of new manufacturing processes

Benefits Big Data Analytics

Big data analytics is used for risk management

Big data analytics is used to improve customer experience

Big data analytics is used for product development and innovations

Big data analytics helps in quicker and better decision making in organizations

Google has mastered the domain of big data analytics and it has developed several tools and techniques to capture the data of users which includes their preference, their likes, dislikes, the area of specialization, their requirement etc.

Some Real Time Applications of Big Data

Google Big Data

Google developed several open source tools and techniques that are extensively used in big data ecosystem. With the help of different big data tools and techniques, Google is now capable of exploring millions of websites and **fetch you the right answer or information within milliseconds.**

Google Big Data

The first question that comes to our mind is how can Google perform such complex operations so efficiently?

The answer is Big data analytics.

Big Data tools and techniques to understand our requirements based on several parameters like search history, locations, trends etc.
Google effortlessly displays the complex calculations which are designed to match the user's requirement.

Google always wanted to develop a search engine that has the ability to think like a human and understand the phrase, logic, and goal of any search query. Semantics has helped Google to accomplish this task to look beyond the literal meaning of any phrase of a search query.

Google Has adopted the following techniques using Big Data Analytics

IBM's Weather Forecasting

One example of an application of big data to weather forecasting is IBM's Deep Thunder. Unlike many weather forecasting systems, which give general information about a broad geographical region, Deep Thunder provides forecasts for extremely specific locations, such as a single airport, so that local authorities can get critically important information in real time.

New York Stock Exchange

The New York Stock Exchange generates about one terabyte of new trade data per day.

Social Media

500TB+ data is generated everyday only on Facebook .This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

Air Industry

A single Jet engine can generate 10+terabytes of data in 30 minutes of flight time. With many thousand flights per day, generation of data reaches up to many Petabytes.

Team Mates