UNIT-1

SYLLABUS

Data science in a big data world: Benefits and uses of data science and big data

Facets of data: Structured data, Unstructured data, Natural language, Machine-generated data, Graph-based or network data, Audio, image, and video, Streaming data

The data science process: Setting the research goal, Retrieving data, Data preparation, Data exploration, Data modeling or model building, Presentation and automation

Data science in a big data world: -

Big data is a blanket term for any collection of data sets so large or complex that it becomes difficult to process them using traditional data management techniques such as, for example, the RDBMS (relational database management systems). The widely adopted RDBMS has long been regarded as a one-size-fits-all solution, but the demands of handling big data have shown otherwise. Data science involves using methods to analyze massive amounts of data and extract the knowledge it contains. You can think of the relationship between big data and data science as being like the relationship between crude oil and an oil refinery. Data science and big data evolved from statistics and traditional data management but are now considered to be distinct disciplines.

The characteristics of big data are often referred to as the three Vs:

- Volume—How much data is there?
- Varieties—How diverse are different types of data?
- Velocity—At what speed is new data generated?

Often these characteristics are complemented with a fourth V, veracity: How accurate is the data?

These four properties make big data different from the data found in traditional data management tools. Consequently, the challenges they bring can be felt in almost every aspect: data capture, curation, storage, search, sharing, transfer, and visualization. In addition, big data calls for specialized techniques to extract the insights.

Data science is an evolutionary extension of statistics capable of dealing with the massive amounts of data produced today. It adds methods from computer science to the repertoire of statistics. The main things that set a data scientist apart from a statistician are the ability to work with big data and experience in machine learning, computing, and algorithm building. Their tools tend to differ too, with data scientist job descriptions more frequently mentioning the ability to use Hadoop, Pig, Spark, R, Python, and Java, among others. Python is a great language for data science because it has many data science libraries available, and it's widely supported by specialized software. For instance, almost every popular NoSQL database has a Python-specific API. Because of these features and the ability to prototype quickly with Python while keeping acceptable performance, its influence is steadily growing in the data science world.

As the amount of data continues to grow and the need to leverage it becomes more important, every data scientist will come across big data projects throughout their career.

BENEFITS AND USES OF DATA SCIENCE AND BIG DATA: -

Data science and big data are used almost everywhere in both commercial and noncommercial settings.

Commercial companies in almost every industry use data science and big data to gain insights into their customers, processes, staff, completion, and products. Many companies use data science to offer customers a better user experience, as well as to cross-sell, up-sell, and personalize their offerings. A good example of this is Google AdSense, which collects data from internet users so relevant commercial messages can be matched to the person browsing the internet. MaxPoint (http://maxpoint.com/us) is another example of real-time personalized advertising.

Human resource professionals use people analytics and text mining to screen candidates, monitor the mood of employees, and study informal networks among coworkers. People analytics is the central theme in the book Moneyball: The Art of Winning an Unfair Game. In the book (and movie) we saw that the traditional scouting process for American baseball was random, and replacing it with correlated signals changed everything. Relying on statistics allowed them to hire the right players and pit them against the opponents where they would have the biggest advantage.

Financial institutions use data science to predict stock markets, determine the risk of lending money, and learn how to attract new clients for their services. At the time of writing this book, at least 50% of trades worldwide are performed automatically by machines based on algorithms developed by quants, as data scientists who work on trading algorithms are often called, with the help of big data and data science techniques.

Governmental organizations are also aware of data's value. Many governmental organizations not only rely on internal data scientists to discover valuable information, but also share their data with the public. You can use this data to gain insights or build data-driven applications. Data.gov is but one example; it's the home of the US Government's open data. A data scientist in a governmental organization gets to work on diverse projects such as detecting fraud and other criminal activity or optimizing project funding. A well-known example was provided by Edward Snowden, who leaked internal documents of the American National Security Agency and the British Government Communications Headquarters that show clearly how they used data science and big data to monitor millions of individuals. Those organizations collected 5 billion data records from widespread applications such as Google Maps, Angry Birds, email, and text messages, among many other data sources. Then they applied data science techniques to distill information.

Nongovernmental organizations (NGOs) are also no strangers to using data. They use it to raise money and defend their causes. The World Wildlife Fund (WWF), for instance, employs data scientists to increase the effectiveness of their fundraising efforts. Many data scientists devote part of their time to helping NGOs, because NGOs often lack the resources to

collect data and employ data scientists. DataKind is one such data scientist group that devotes its time to the benefit of mankind.

Universities use data science in their research but also to enhance the study experience of their students. The rise of massive open online courses (MOOC) produces a lot of data, which allows universities to study how this type of learning can complement traditional classes. MOOCs are an invaluable asset if you want to become a data scientist and big data professional, so definitely look at a few of the better-known ones: Coursera, Udacity, and edX. The big data and data science landscape changes quickly, and MOOCs allow you to stay up to date by following courses from top universities.

FACETS OF DATA: -

In data science and big data you'll come across many different types of data, and each of them tends to require different tools and techniques. The main categories of data are these:

- Structured
- Unstructured
- Natural language
- Machine-generated
- Graph-based
- Audio, video, and images
- Streaming

Structured data: -

Structured data is data that depends on a data model and resides in a fixed field within a record. As such, it's often easy to store structured data in tables within databases or Excel files (figure 1.1). SQL, or Structured Query Language, is the preferred way to manage and query data that resides in databases. You may also come across structured data that might give you a hard time storing it in a traditional relational database. Hierarchical data such as a family tree is one such example.

	Indicator ID	Dimension List	Timeframe	Numeric Value	Missing Value Flag	Confidence Inte
2	214390830	Total (Age-adjusted)	2008	74.6%		73.8%
3	214390833	Aged 18-44 years	2008	59.4%		58.0%
4	214390831	Aged 18-24 years	2008	37.4%		34.6%
5	214390832	Aged 25-44 years	2008	66.9%		65.5%
6	214390836	Aged 45-64 years	2008	88.6%		87.7%
7	214390834	Aged 45-54 years	2008	86.3%		85.1%
8	214390835	Aged 55-64 years	2008	91.5%		90.4%
9	214390840	Aged 65 years and over	2008	94.6%		93.8%
10	214390837	Aged 65-74 years	2008	93.6%		92.4%
11	214390838	Aged 75-84 years	2008	95.6%		94.4%
12	214390839	Aged 85 years and over	2008	96.0%		94.0%
13	214390841	Male (Age-adjusted)	2008	72.2%		71.1%
14	214390842	Female (Age-adjusted)	2008	76.8%		75.9%
15	214390843	White only (Age-adjusted)	2008	73.8%		72.9%
16	214390844	Black or African American only (Age-adjusted)	2008	77.0%		75.0%
17	214390845	American Indian or Alaska Native only (Age-adjusted)	2008	66.5%		57.1%
18	214390846	Asian only (Age-adjusted)	2008	80.5%		77.7%
19	214390847	Native Hawaiian or Other Pacific Islander only (Age-adjusted)	2008	DSU		
20	214390848	2 or more races (Age-adjusted)	2008	75.6%		69.6%

Figure 1.1 An Excel table is an example of structured data.

The world isn't made up of structured data, though; it's imposed upon it by humans and machines. More often, data comes unstructured.

Unstructured data: -

Unstructured data is data that isn't easy to fit into a data model because the content is context-specific or varying. One example of unstructured data is your regular email (figure 1.2). Although email contains structured elements such as the sender, title, and body text, it's a challenge to find the number of people who have written an email complaint about a specific employee because so many ways exist to refer to a person, for example. The thousands of different languages and dialects out there further complicate this.

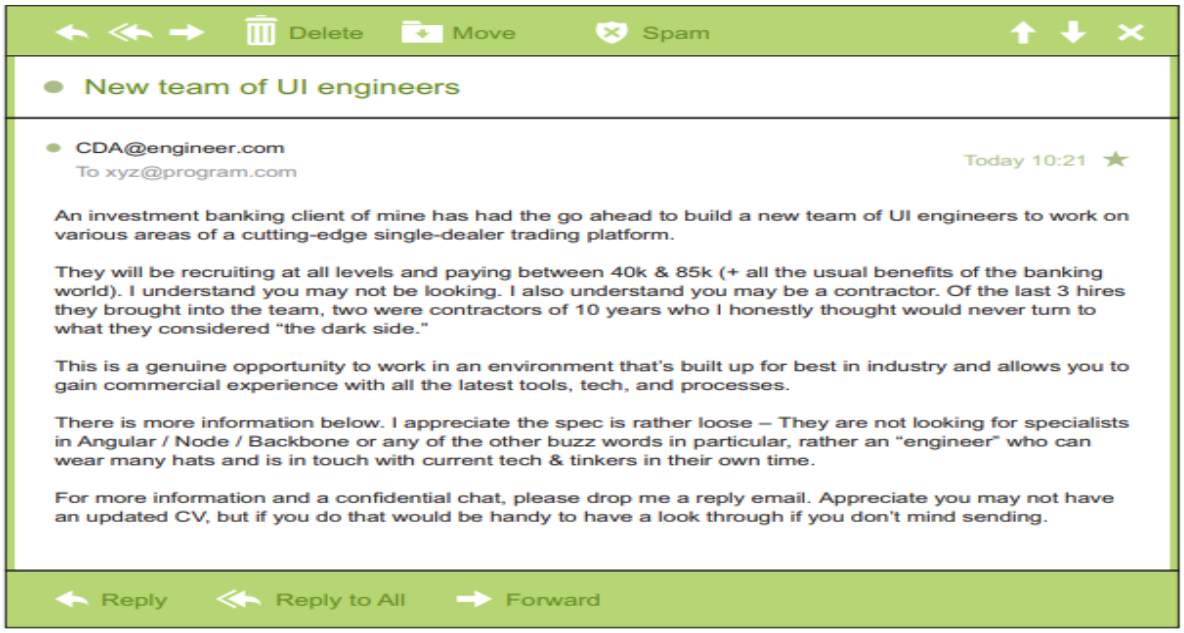


Figure 1.2 Email is simultaneously an example of unstructured data and natural language data.

A human-written email, as shown in figure 1.2, is also a perfect example of natural language data.

Natural language: -

Natural language is a special type of unstructured data; it's challenging to process because it requires knowledge of specific data science techniques and linguistics.

The natural language processing community has had success in entity recognition, topic recognition, summarization, text completion, and sentiment analysis, but models trained in one domain don't generalize well to other domains. Even state-of-the-art techniques aren't able to decipher the meaning of every piece of text. This shouldn't be a surprise though: humans struggle with natural language as well. It's ambiguous by nature. The concept of meaning itself is questionable here. Have two people listen to the same conversation. Will they get the same meaning? The meaning of the same words can vary when coming from someone upset or joyous.

Machine-generated data

Machine-generated data is information that's automatically created by a computer, process, application, or other machine without human intervention. Machine-generated data is

becoming a major data resource and will continue to do so. Wikibon has forecast that the market value of the industrial Internet will be approximately $540 billion in 2020. IDC (International Data Corporation) has estimated there will be 26 times more connected things than people in 2020. This network is commonly referred to as the internet of things.

The analysis of machine data relies on highly scalable tools, due to its high volume and speed. Examples of machine data are web server logs, call detail records, network event logs, and telemetry (figure 1.3).

```
CSIPERF:TXCOMMIT;313236
2014-11-28 11:36:13, Info                  CSI    00000153 Creating NT transaction (seq
69), objectname [6]"(null)"
2014-11-28 11:36:13, Info                  CSI    00000154 Created NT transaction (seq 69)
result 0x00000000, handle @0x4e54
2014-11-28 11:36:13, Info                  CSI    00000155@2014/11/28:10:36:13.471
Beginning NT transaction commit...
2014-11-28 11:36:13, Info                  CSI    00000156@2014/11/28:10:36:13.705 CSI perf
trace:
CSIPERF:TXCOMMIT;273983
2014-11-28 11:36:13, Info                  CSI    00000157 Creating NT transaction (seq
70), objectname [6]"(null)"
2014-11-28 11:36:13, Info                  CSI    00000158 Created NT transaction (seq 70)
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:13, Info                  CSI    00000159@2014/11/28:10:36:13.764
Beginning NT transaction commit...
2014-11-28 11:36:14, Info                  CSI    0000015a@2014/11/28:10:36:14.094 CSI perf
trace:
CSIPERF:TXCOMMIT;386259
2014-11-28 11:36:14, Info                  CSI    0000015b Creating NT transaction (seq
71), objectname [6]"(null)"
2014-11-28 11:36:14, Info                  CSI    0000015c Created NT transaction (seq 71)
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:14, Info                  CSI    0000015d@2014/11/28:10:36:14.106
Beginning NT transaction commit...
2014-11-28 11:36:14, Info                  CSI    0000015e@2014/11/28:10:36:14.428 CSI perf
trace:
CSIPERF:TXCOMMIT;375581
```

Figure 1.3 Example of machine-generated data

The machine data shown in figure 1.3 would fit nicely in a classic table-structured database. This isn't the best approach for highly interconnected or "networked" data, where the relationships between entities have a valuable role to play.

Graph-based or network data: -

"Graph data" can be a confusing term because any data can be shown in a graph. "Graph" in this case points to mathematical graph theory. In graph theory, a graph is a mathematical structure to model pair-wise relationships between objects. Graph or network data is, in short, data that focuses on the relationship or adjacency of objects. The graph structures use nodes, edges, and properties to represent and store graphical data. Graph-based data is a natural way to represent social networks, and its structure allows you to calculate specific metrics such as the influence of a person and the shortest path between two people.

Examples of graph-based data can be found on many social media websites (figure 1.4). For instance, on LinkedIn you can see who you know at which company. Your follower list on Twitter is another example of graph-based data. The power and sophistication comes from multiple, overlapping graphs of the same nodes. For example, imagine the connecting edges

here to show “friends” on Facebook. Imagine another graph with the same person which connects business colleagues via LinkedIn. Imagine a third graph based on movie interests on Netflix. Overlapping the three different-looking graphs makes more interesting questions possible.

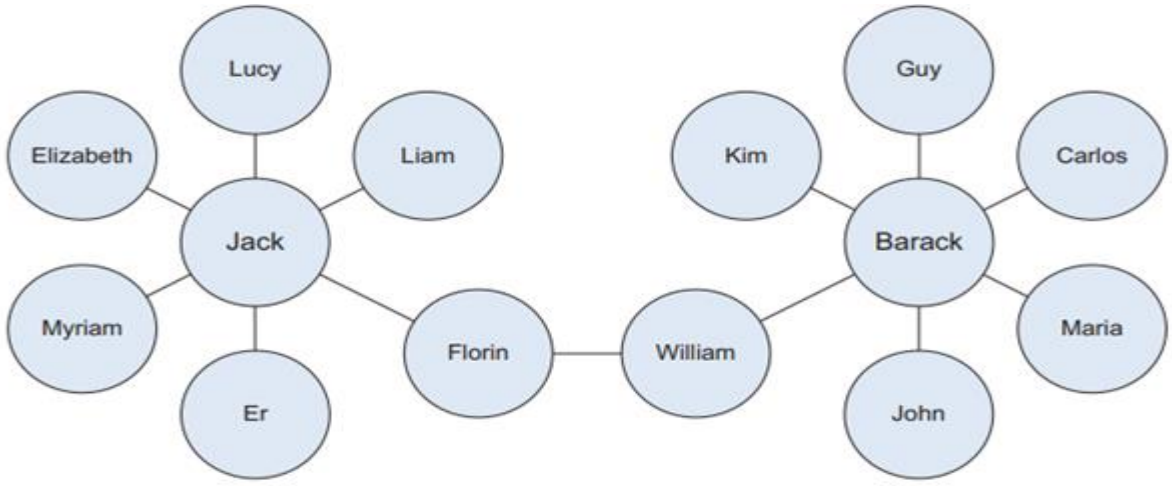


Figure 1.4 Friends in a social network are an example of graph-based data.

Graph databases are used to store graph-based data and are queried with specialized query languages such as SPARQL. Graph data poses its challenges, but for a computer interpreting additive and image data, it can be even more difficult.

Audio, image, and video: -

Audio, image, and video are data types that pose specific challenges to a data scientist. Tasks that are trivial for humans, such as recognizing objects in pictures, turn out to be challenging for computers. MLBAM (Major League Baseball Advanced Media) announced in 2014 that they’ll increase video capture to approximately 7 TB per game for the purpose of live, in-game analytics. High-speed cameras at stadiums will capture ball and athlete movements to calculate in real time, for example, the path taken by a defender relative to two baselines.

Recently a company called DeepMind succeeded at creating an algorithm that’s capable of learning how to play video games. This algorithm takes the video screen as input and learns to interpret everything via a complex process of deep learning. It’s a remarkable feat that prompted Google to buy the company for their own Artificial Intelligence (AI) development plans. The learning algorithm takes in data as it’s produced by the computer game; it’s streaming data.

Streaming data: -

While streaming data can take almost any of the previous forms, it has an extra property. The data flows into the system when an event happens instead of being loaded into a data store in a batch. Although this isn’t really a different type of data, we treat it here as such because you need to adapt your process to deal with this type of information.

Examples are the “What’s trending” on Twitter, live sporting or music events, and the stock market.

THE DATA SCIENCE PROCESS: -

The data science process typically consists of six steps, as you can see in the mind map in figure 1.5.

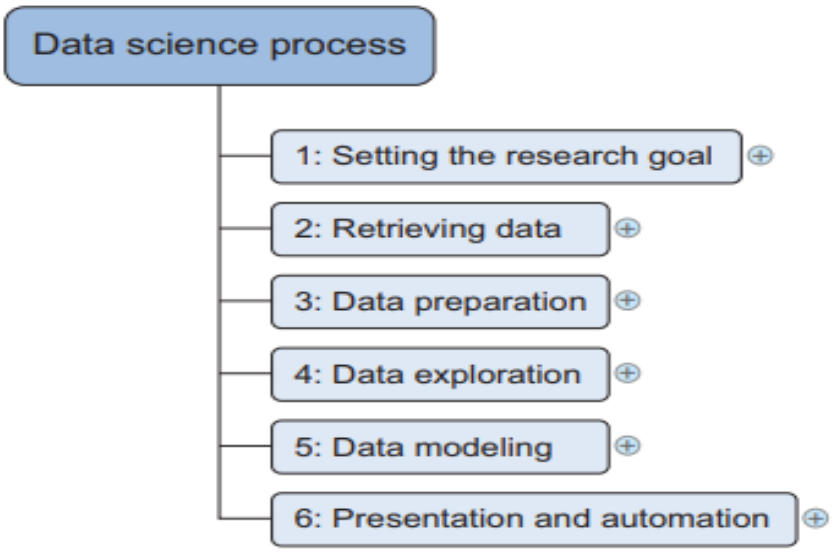


Figure 1.5 The data science process

Setting the research goal: -

Data science is mostly applied in the context of an organization. When the business asks you to perform a data science project, you'll first prepare a project charter. This charter contains information such as what you're going to research, how the company benefits from that, what data and resources you need, a timetable, and deliverables.

Retrieving data: -

The second step is to collect data. You've stated in the project charter which data you need and where you can find it. In this step you ensure that you can use the data in your program, which means checking the existence of, quality, and access to the data. Data can also be delivered by third-party companies and takes many forms ranging from Excel spreadsheets to different types of databases.

Data preparation: -

Data collection is an error-prone process; in this phase you enhance the quality of the data and prepare it for use in subsequent steps. This phase consists of three subphases: **data cleansing** removes false values from a data source and inconsistencies across data sources, **data integration** enriches data sources by combining information from multiple data sources, and **data transformation** ensures that the data is in a suitable format for use in your models.

Data exploration: -

Data exploration is concerned with building a deeper understanding of your data. You try to understand how variables interact with each other, the distribution of the data, and whether there are outliers. To achieve this you mainly use descriptive statistics, visual

techniques, and simple modeling. This step often goes by the abbreviation EDA, for Exploratory Data Analysis.

Data modeling or model building: -

In this phase you use models, domain knowledge, and insights about the data you found in the previous steps to answer the research question. You select a technique from the fields of statistics, machine learning, operations research, and so on. Building a model is an iterative process that involves selecting the variables for the model, executing the model, and model diagnostics.

Presentation and automation: -

Finally, you present the results to your business. These results can take many forms, ranging from presentations to research reports. Sometimes you'll need to automate the execution of the process because the business will want to use the insights you gained in another project or enable an operational process to use the outcome from your model.

AN ITERATIVE PROCESS The previous description of the data science process gives you the impression that you walk through this process in a linear way, but in reality you often have to step back and rework certain findings. For instance, you might find outliers in the data exploration phase that point to data import errors. As part of the data science process you gain incremental insights, which may lead to new questions. To prevent rework, make sure that you scope the business question clearly and thoroughly at the start.

UNIT WISE IMPORTANT QUESTIONS: -

1. What is Data Science? Explain the benefits and uses of data science and big data.
2. Explain different types of data that come across data science and big data.
3. Differentiate structured data and unstructured data
4. Explain in detail about Graph based data
5. Discuss different steps involved in data science process
6. Explain in detail about Network data
7. Explain in detail about Machine generated data
8. Describe life cycle of Data Science with neat diagram
9. Define data? Explain different forms of data.
10. What is Data Science process? Explain
11. Compare Big Data with Data Science
12. List out the areas in which Data Science can be applied

Differentiate structured data and unstructured data

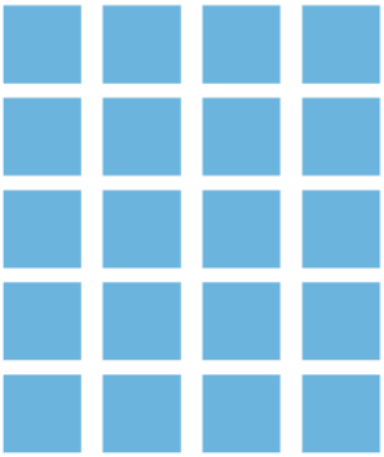
Structured Data: -

The data which is to the point, factual, and highly organized is referred to as structured data. It is quantitative in nature, i.e., it is related to quantities that means it contains measurable numerical values like numbers, dates, and times.

It is easy to search and analyze structured data. Structured data exists in a predefined format. Relational database consisting of tables with rows and columns is one of the best

examples of structured data. Structured data generally exist in tables like excel files and Google Docs spreadsheets. The programming language SQL (structured query language) is used for managing the structured data. SQL is developed by IBM in the 1970s and majorly used to handle relational databases and warehouses.

Structured data is highly organized and understandable for machine language. Common applications of relational databases with structured data include sales transactions, Airline reservation systems, inventory control, and others.



Structured Data

Unstructured Data

Unstructured Data: -

All the unstructured files, log files, audio files, and image files are included in the unstructured data. Some organizations have much data available, but they did not know how to derive data value since the data is raw.

Unstructured data is the data that lacks any predefined model or format. It requires a lot of storage space, and it is hard to maintain security in it. It cannot be presented in a data model or schema. That's why managing, analyzing, or searching for unstructured data is hard. It resides in various different formats like text, images, audio and video files, etc. It is qualitative in nature and sometimes stored in a non-relational database or NO-SQL.

It is not stored in relational databases, so it is hard for computers and humans to interpret it. The limitations of unstructured data include the requirement of data science experts and specialized tools to manipulate the data.

The amount of unstructured data is much more than the structured or semi-structured data. Examples of human-generated unstructured data are Text files, Email, social media, media, mobile data, business applications, and others. The machine-generated unstructured data includes satellite images, scientific data, sensor data, digital surveillance, and many more.

Here, we are tabulating the difference between both terms based on some characteristics.

On the basis of	**Structured data**	**Unstructured data**
Technology	It is based on a relational database.	It is based on character and binary data.
Flexibility	Structured data is less flexible	There is an absence of schema, so

	and schema-dependent.	it is more flexible.
Scalability	It is hard to scale database schema.	It is more scalable.
Robustness	It is very robust.	It is less robust.
Performance	Here, we can perform a structured query that allows complex joining, so the performance is higher.	While in unstructured data, textual queries are possible, the performance is lower than semi-structured and structured data.
Nature	Structured data is quantitative, i.e., it consists of hard numbers or things that can be counted.	It is qualitative, as it cannot be processed and analyzed using conventional tools.
Format	It has a predefined format.	It has a variety of formats, i.e., it comes in a variety of shapes and sizes.
Analysis	It is easy to search.	Searching for unstructured data is more difficult.

Compare Big Data with Data Science

Big data:

Big data is huge, large, or voluminous data, information, or the relevant statistics acquired by large organizations that are difficult to process by traditional tools. It is referred to as the study of collecting and analyzing the huge volume of data sets to find a hidden pattern that helps in stronger decision-making for the firms using specialized software and analytical tools. Big data can be structured, unstructured, or semi-structured.

Big Data is used to store, analyze and organize the huge volume of structured as well as unstructured datasets. Big Data can be described mainly with 5 V's such as Volume, Variety, velocity, value, and Veracity.

Data Science:

Data science is the study of working with a huge volume of data and enables data for prediction, prescriptive, and prescriptive analytical models. It helps to discriminate useful and raw data/insights from the vast amount of data sets using various scientific methods, algorithms, tools, and processes. It includes digging, capturing, analyzing, and utilizing the data from a vast volume of datasets.

It is a combination of various filed such as computer science, machine learning, AI, Mathematics, business, and statistics.

Major differences between Data Science and Big Data in the below table.

Data Science	Big data
Data science is the study of working with a huge volume of data and enables data for prediction, prescriptive, and prescriptive analytical models.	Big data is the study of collecting and analyzing a huge volume of data sets to find a hidden pattern that helps in stronger decision-making.

It is a combination of various concepts of computer science, statistics, and applied mathematics.	It is a technique to extract meaningful insights from complex data sets.
The main aim of data science is to build data-based products for firms.	The main goal of big data is to extract useful information from the huge volume of data and use it for building products for firms.
It requires strong knowledge of Python, R, SAS, Scala, as well as hands-on knowledge of SQL databases.	It requires tools like Apache Hadoop MongoDB.
It is used for scientific or research purposes.	It is used for businesses and customer satisfaction.
It broadly focuses on the science of the data.	It is more involved with the processes of handling voluminous data.
It includes various data operations such as cleaning, collection, manipulation, etc.	It includes analysis of data stored in a structured format such as stock market analysis, etc.

List out the areas in which Data Science can be applied

Data Science is the deep study of a large quantity of data, which involves extracting some meaningful from the raw, structured, and unstructured data. The extracting out meaningful data from large amounts use processing of data and this processing can be done using statistical techniques and algorithm, scientific techniques, different technologies, etc. It uses various tools and techniques to extract meaningful data from raw data. Data Science is also known as the Future of Artificial Intelligence.

Applications of Data Science

In Search Engines

The most useful application of Data Science is Search Engines. As we know when we want to search for something on the internet, we mostly used Search engines like Google, Yahoo, Safari, Firefox, etc. So Data Science is used to get Searches faster.

For Example, when we search something suppose "Data Structure and algorithm courses" then at that time on the Internet Explorer we get the first link of Infosys Spring board Courses. This happens because the Infosys Spring board website is visited most in order to get information regarding Data Structure courses and Computer related subjects. So this analysis is done using Data Science, and we get the topmost visited Web Links.

In Transport

Data Science also entered into the Transport field like Driverless Cars. With the help of Driverless Cars, it is easy to reduce the number of Accidents.

For Example, In Driverless Cars the training data is fed into the algorithm and with the help of Data Science techniques, the Data is analyzed like what is the speed limit in Highway, Busy Streets, Narrow Roads, etc. And how to handle different situations while driving etc.

In Finance

Data Science plays a key role in Financial Industries. Financial Industries always have an issue of fraud and risk of losses. Thus, Financial Industries needs to automate risk of loss analysis in order to carry out strategic decisions for the company. Also, Financial Industries uses Data Science Analytics tools in order to predict the future. It allows the companies to predict customer lifetime value and their stock market moves.

For Example, In Stock Market, Data Science is the main part. In the Stock Market, Data Science is used to examine past behavior with past data and their goal is to examine the future outcome. Data is analyzed in such a way that it makes it possible to predict future stock prices over a set timetable.

In E-Commerce

E-Commerce Websites like Amazon, Flipkart, etc. uses data Science to make a better user experience with personalized recommendations.

For Example, when we search for something on the E-commerce websites we get suggestions similar to choices according to our past data and also we get recommendations according to most buy the product, most rated, most searched, etc. This is all done with the help of Data Science.

In Health Care

In the Healthcare Industry data science acts as a boon. Data Science is used for:

- Detecting Tumor.
- Drug discoveries.
- Medical Image Analysis.
- Virtual Medical Bots.
- Genetics and Genomics.
- Predictive Modeling for Diagnosis etc.

Image Recognition

Currently, Data Science is also used in Image Recognition. **For Example,** when we upload our image with our friend on Facebook, Facebook gives suggestions tagging who is in the picture. This is done with the help of machine learning and Data Science. When an Image is Recognized, the data analysis is done on one's Facebook friends and after analysis, if the faces which are present in the picture matched with someone else profile then Facebook suggests us auto-tagging.

Targeting Recommendation

Targeting Recommendation is the most important application of Data Science. Whatever the user searches on the Internet, he/she will see numerous posts everywhere. This can be explained properly with an example: Suppose I want a mobile phone, so I just Google search it and after that, I changed my mind to buy offline. Data Science helps those companies who are paying for Advertisements for their mobile. So everywhere on the internet in the social media, in the websites, in the apps everywhere I will see the

recommendation of that mobile phone which I searched for. So this will force me to buy online.

Airline Routing Planning: -

With the help of Data Science, Airline Sector is also growing like with the help of it, it becomes easy to predict flight delays. It also helps to decide whether to directly land into the destination or take a halt in between like a flight can have a direct route from Delhi to the U.S.A or it can halt in between after that reach at the destination.

Data Science in Gaming: -

In most of the games where a user will play with an opponent i.e. a Computer Opponent, data science concepts are used with machine learning where with the help of past data the Computer will improve its performance. There are many games like Chess, EA Sports, etc. will use Data Science concepts.

Medicine and Drug Development: -

The process of creating medicine is very difficult and time-consuming and has to be done with full disciplined because it is a matter of Someone's life. Without Data Science, it takes lots of time, resources, and finance or developing new Medicine or drug but with the help of Data Science, it becomes easy because the prediction of success rate can be easily determined based on biological data or factors. The algorithms based on data science will forecast how this will react to the human body without lab experiments.

In Delivery Logistics: -

Various Logistics companies like DHL, FedEx, etc. make use of Data Science. Data Science helps these companies to find the best route for the Shipment of their Products, the best time suited for delivery, the best mode of transport to reach the destination, etc.

Autocomplete: -

AutoComplete feature is an important part of Data Science where the user will get the facility to just type a few letters or words, and he will get the feature of auto-completing the line. In Google Mail, when we are writing formal mail to someone so at that time data science concept of Autocomplete feature is used where he/she is an efficient choice to auto-complete the whole line. Also in Search Engines in social media, in various apps, AutoComplete feature is widely used.