What is Big Data? – A Beginner's Guide to the World of Big Data

Anushree Subramaniam
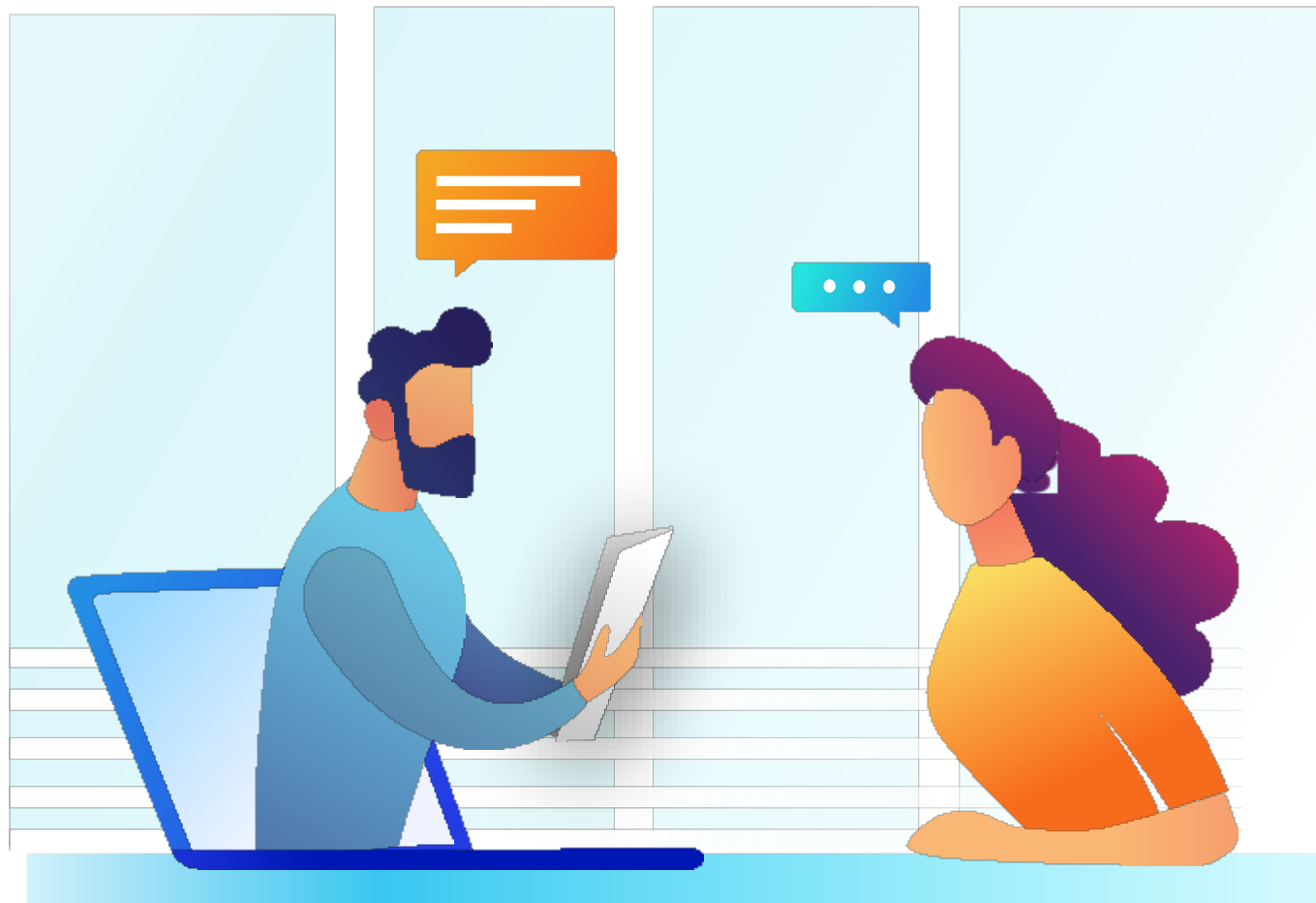
There is no place where Big Data does not exist! The curiosity about what is Big Data has been soaring in the past few years. Let me tell you some mind-boggling facts! Forbes reports that every minute, users watch *4.15 million YouTube videos*, send *456,000 tweets* on Twitter, post *46,740 photos* on Instagram and there are *510,000 comments* posted and *293,000 statuses* updated on Facebook!

Just imagine the huge chunk of data that is produced with such activities. This constant creation of data using social media, business applications, telecom and various other domains is leading to the formation of Big Data.

In order to explain **what is Big Data**, I will be covering the following topics:

- Evolution of Big Data
- Big Data Defined
- Characteristics of Big Data
- Big Data Analytics
- Industrial Applications of Big Data
- Scope of Big Data

Evolution of Big Data

Before exploring what is Big Data, let me begin by giving some insight into why the term Big Data has gained so much importance.

When was the last time you guys remember using a floppy or a CD to store your data? Let me guess, had to go way back in the early 21st century right? The use of manual paper records, files, floppy and discs have now become obsolete. The reason for this is the exponential growth of data. People began storing their data in relational database systems but with the hunger for new inventions, technologies, applications with quick response time and with the introduction of the internet, even that is insufficient now. This generation of continuous and massive data can be referred to as Big Data. There are a few other factors that characterize Big Data which I will be explaining later in this blog.

Forbes reports that there are 2.5 quintillion bytes of data created each day at our current pace, but that pace is only accelerating. Internet of Things(IoT) is one such technology which plays a major role in this acceleration. 90% of all data today was generated in the last two years.

What is Big Data | Big Data Analytics | Edureka

Big Data Definition

What is Big Data?

So before I explain what is Big Data, let me also tell you what it is not! The most common myth associated with Big Data is that it is just about the size or volume of data. But actually, it's not just about the "big" amounts of data being collected. **Big Data** refers to the large amounts of data which is pouring in from various data sources and has different formats. Even previously there was huge data which were being stored in databases, but because of the varied nature of this Data, the traditional relational database systems are incapable of handling this Data. Big Data is much more than a collection of datasets with different formats, it is an important asset which can be used to obtain enumerable benefits.

The three different formats of big data are:

1. **Structured:** Organised data format with a fixed schema. Ex: RDBMS
2. **Semi-Structured:** Partially organised data which does not have a fixed format. Ex: XML, JSON
3. **Unstructured:** Unorganised data with an unknown schema. Ex: Audio, video files etc.

Characteristics of Big Data

These are the following characteristics associated with Big Data:
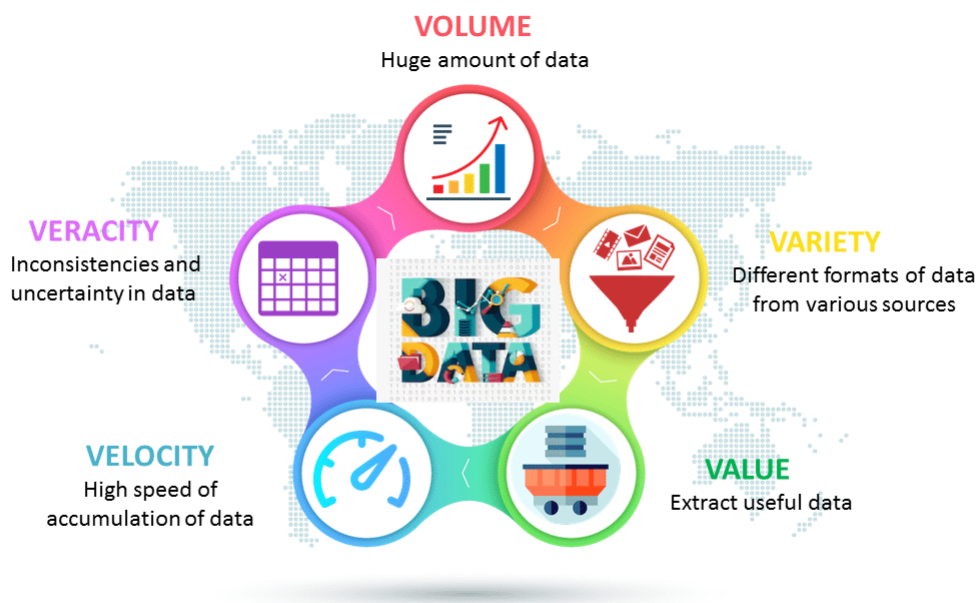
The above image depicts the five V's of Big Data but as and when the data keeps evolving so will the V's. I am listing five more V's which have developed gradually over time:

- Validity: correctness of data
- Variability: dynamic behaviour
- Volatility: tendency to change in time
- Vulnerability: vulnerable to breach or attacks
- Visualization: visualizing meaningful usage of data

Big Data Analytics

Now that I have told you what is Big Data and how it's being generated exponentially, let me present to you a very interesting example of how *Starbucks*, one of the leading coffeehouse chain is making use of this Big Data.

I came across this article by Forbes which reported how *Starbucks* made use of Big Data to analyse the preferences of their customers to enhance and personalize their experience. They analysed their member's coffee buying habits along with their preferred drinks to what time of day they are usually ordering. So, even when people visit a "new" Starbucks location, that store's point-of-sale system is able to identify the customer through their smartphone and give the barista their preferred order. In addition, based on ordering preferences, their app will suggest new products that the customers might be interested in trying. This my friends is what we call Big Data Analytics.

Basically, Big Data Analytics is largely used by companies to facilitate their growth and development. This majorly involves applying various data mining algorithms on the given set of data, which will then aid them in better decision making.

There are multiple tools for processing Big Data such as **Hadoop**, **Pig**, **Hive**, **Cassandra**, **Spark**, **Kafka**, etc. depending upon the requirement of the organisation.

Big Data Applications

These are some of the following domains where *Big Data Applications* has been revolutionized:

- **Entertainment:** Netflix and Amazon use Big Data to make shows and movie recommendations to their users.
- **Insurance:** Uses Big data to predict illness, accidents and price their products accordingly.
- **Driver-less Cars:** Google's driver-less cars collect about one gigabyte of data per second. These experiments require more and more data for their successful execution.
- **Education:** Opting for big data powered technology as a learning tool instead of traditional lecture methods, which enhanced the learning of students as well aided the teacher to track their performance better.
- **Automobile:** Rolls Royce has embraced Big Data by fitting hundreds of sensors into its engines and propulsion systems, which record every tiny detail about their operation. The changes in data in real-time are reported to engineers who will decide the best course of action such as scheduling maintenance or dispatching engineering teams should the problem require it.
- **Government:** A very interesting use of Big Data is in the field of politics to analyse patterns and influence election results. Cambridge Analytica Ltd. is one such organisation which completely drives on data to change audience behaviour and plays a major role in the electoral process.

Scope of Big Data

- **Numerous Job opportunities:** The career opportunities pertaining to the field of Big data include, Big Data Analyst, Big Data Engineer, Big Data solution architect etc. According to IBM, 59% of all Data Science and Analytics (DSA) job demand is in Finance and Insurance, Professional Services, and IT.
- **Rising demand for Analytics Professional:** An article by Forbes reveals that "IBM predicts demand for Data Scientists will soar by 28%". By 2020, the number of jobs for all US data professionals will increase by 364,000 openings to 2,720,000 according to IBM.
- **Salary Aspects:** Forbes reported that employers are willing to pay a premium of $8,736 above median bachelor's and graduate-level salaries, with successful applicants earning a starting salary of $80,265
- **Adoption of Big Data analytics:** Immense growth in the usage of big data analysis across the world.
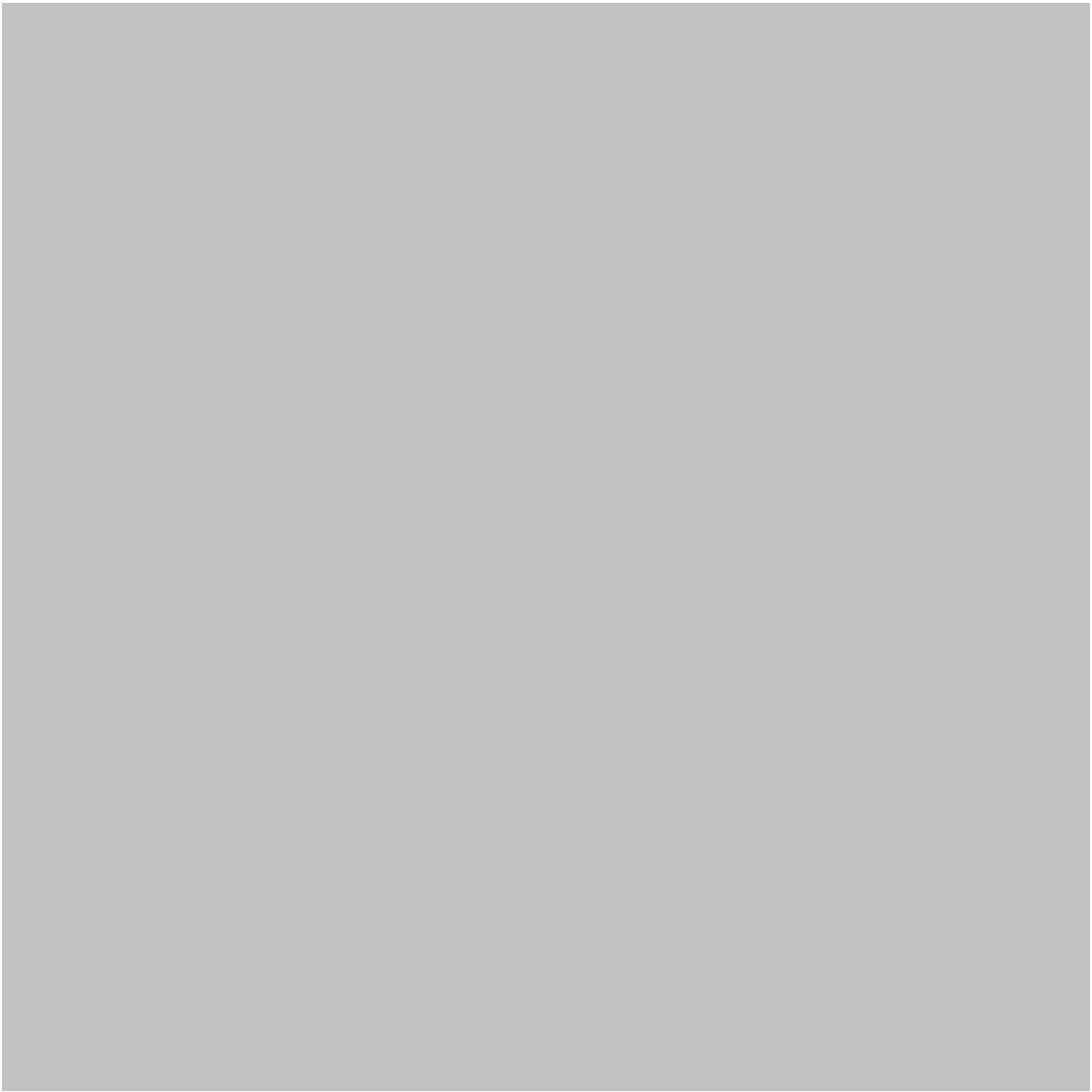
The above image depicts the growing market revenue of Big Data in billion U.S. dollars from the year 2011 to 2027. So that was all about What is Big Data and I hope this blog was helpful.

Got a question for us? Please mention it in the comments section and we will get back to you.

Recommended videos for you

Recommended blogs for you

Comments

Trending Courses in Big Data

Browse Categories

You are here:

edureka!

TRENDING CERTIFICATION COURSES

DevOps Certification Training

AWS Architect Certification Training

Big Data Hadoop Certification Training

Tableau Training & Certification

Python Certification Training for Data Science

Selenium Certification Training

PMP® Certification Exam Training

Robotic Process Automation Training using UiPath

Apache Spark and Scala Certification Training

Microsoft Power BI Training

Online Java Course and Training

Python Certification Course

CATEGORIES ⌄

Cloud Computing | DevOps | Big Data | Data Science | BI and Visualization | Programming & Frameworks | Software Testing | Project Management and Methodologies | Robotic Process Automation | Frontend Development | Data Warehousing and ETL | Artificial Intelligence | Blockchain | Databases | Cyber Security | Mobile Development | Operating Systems | Architecture & Design Patterns | Digital Marketing

TRENDING BLOG ARTICLES ⌄

Selenium tutorial | Selenium interview questions | Java tutorial | What is HTML | Java interview questions | PHP tutorial | JavaScript interview questions | Spring tutorial | PHP interview questions | Inheritance in Java | Polymorphism in Java | Spring interview questions | Pointers in C | Linux commands | Android tutorial | JavaScript tutorial | jQuery tutorial | SQL interview questions | MySQL tutorial | Machine learning tutorial | Python tutorial | What is machine learning | Ethical hacking tutorial | SQL injection | AWS certification career opportunities | AWS tutorial | What Is cloud computing | What is blockchain | Hadoop tutorial | What is artificial intelligence | Node Tutorial | Collections in Java | Exception handling in java | Python Programming Language | Python interview questions | Multithreading in Java | ReactJS Tutorial | Data Science vs Big Data vs Data Analyt... | Software Testing Interview Questions | R Tutorial | Java Programs | JavaScript Reserved Words and Keywords | Implement thread.yield() in Java: Exampl... | Implement Optical Character Recognitio... | All you Need to Know About Implements...

Subscribe to our Newsletter, and get personalized recommendations. ✕

Sign up with Google

Signup with Facebook

Already have an account? Sign in.