**What is Data?**

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

Now, let's learn Big Data introduction

**What is Big Data?**

**Big Data** is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

In this, you will learnesting Principles Software Testing

* [What is Data?](https://www.guru99.com/what-is-big-data.html#5)
* [What is Big Data?](https://www.guru99.com/what-is-big-data.html#6)
* [Examples Of Big Data](https://www.guru99.com/what-is-big-data.html#1)
* [Types Of Big Data](https://www.guru99.com/what-is-big-data.html#2)
* [Characteristics Of Big Data](https://www.guru99.com/what-is-big-data.html#3)
* [Advantages Of Big Data Processing](https://www.guru99.com/what-is-big-data.html#4)

**Examples Of Big Data**

Following are some of the Big Data examples-

The **New York Stock Exchange** is an example of Big Data that generates about ***one terabyte*** of new trade data per day.

[](https://cdn.guru99.com/images/Big_Data/061114_0759_WhatIsBigDa2.jpg)

**Social Media**

The statistic shows that ***500+terabytes*** of new data get ingested into the databases of social media site **Facebook**, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

[](https://cdn.guru99.com/images/Big_Data/061114_0759_WhatIsBigDa3.jpg)

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.***

[](https://cdn.guru99.com/images/Big_Data/061114_0759_WhatIsBigDa4.jpg)

**Types Of Big Data**

Following are the types of Big Data:

1. **Structured**
2. **Unstructured**
3. **Semi-structured**

**Structured**

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data. Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving value out of it. However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the rage of multiple zettabytes.

***Do you know? 1021 bytes*** equal to ***1 zettabyte*** or ***one billion terabytes*** forms ***a zettabyte***.

Looking at these figures one can easily understand why the name Big Data is given and imagine the challenges involved in its storage and processing.

***Do you know?*** Data stored in a relational database management system is one example of a **'structured'** data.

**Examples Of Structured Data**

An 'Employee' table in a database is an example of Structured Data

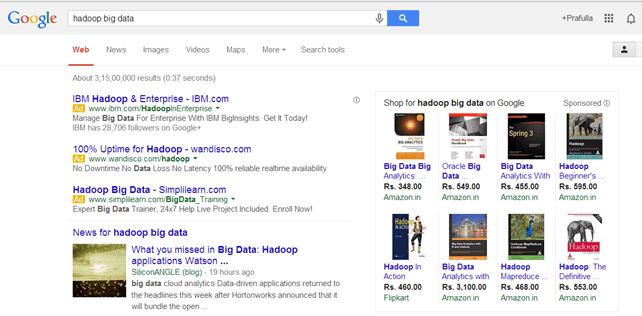
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Employee\_ID** | **Employee\_Name** | **Gender** | **Department** | **Salary\_In\_lacs** |
| 2365 | Rajesh Kulkarni | Male | Finance | 650000 |
| 3398 | Pratibha Joshi | Female | Admin | 650000 |
| 7465 | Shushil Roy | Male | Admin | 500000 |
| 7500 | Shubhojit Das | Male | Finance | 500000 |
| 7699 | Priya Sane | Female | Finance | 550000 |

**Unstructured**

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc. Now day organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

**Examples Of Un-structured Data**

The output returned by 'Google Search'

[](https://cdn.guru99.com/images/Big_Data/061114_0759_WhatIsBigDa5.png)

Example Of Un-structured Data

**Semi-structured**

Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational [DBMS](https://www.guru99.com/what-is-dbms.html). Example of semi-structured data is a data represented in an XML file.

Examples Of Semi-structured Data

Personal data stored in an XML file-

<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>

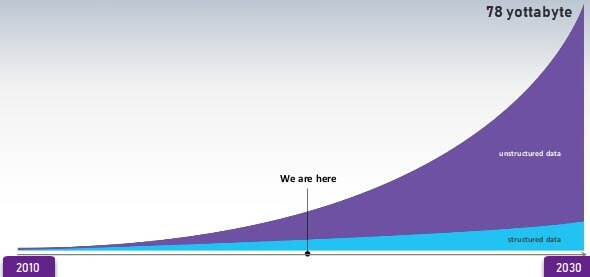
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>

<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>

**Data Growth over the years**

[](https://cdn.guru99.com/images/1/big-data-growth.jpg)

Data Growth over the years

 Please note that [web application](https://www.guru99.com/difference-web-application-website.html) data, which is unstructured, consists of log files, transaction history files etc. OLTP systems are built to work with structured data wherein data is stored in relations (tables).

**Characteristics Of Big Data**

Big data can be described by the following characteristics:

* Volume
* Variety
* Velocity
* Variability

***(i) Volume –*** The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, **'Volume'** is one characteristic which needs to be considered while dealing with Big Data solutions.

***(ii) Variety –*** The next aspect of Big Data is its **variety**.

Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analyzing data.

***(iii) Velocity –*** The term **'velocity'** refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.

Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors,[Mobile](https://www.guru99.com/mobile-testing.html)devices, etc. The flow of data is massive and continuous.

***(iv) Variability –*** This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

**Advantages Of Big Data Processing**

Ability to process Big Data in DBMS brings in multiple benefits, such as-

* + Businesses can utilize outside intelligence while taking decisions

Access to social data from search engines and sites like facebook, twitter are enabling organizations to fine tune their business strategies.

* + Improved customer service

Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.

* + Early identification of risk to the product/services, if any
  + Better operational efficiency

Big Data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the [data warehouse](https://www.guru99.com/data-warehousing.html). In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.

**Summary**

* Big Data definition : Big Data meaning a data that is huge in size. Bigdata is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.
* Big Data analytics examples includes stock exchanges, social media sites, jet engines, etc.
* Big Data could be 1) Structured, 2) Unstructured, 3) Semi-structured
* Volume, Variety, Velocity, and Variability are few Big Data characteristics
* Improved customer service, better operational efficiency, Better Decision Making are few advantages of Bigdata