Small Data

Feature	Small Data	Big Data
Variety	Data is typically structured and uniform	Data is often unstructured and heterogeneous
Veracity	Data is generally high quality and reliable	Data quality and reliability can vary widely
Processing	Data can often be processed on a single machine or in-memory	Data requires distributed processing frameworks such as MapReduce or Spark
Technology	Traditional	Modern
Analytics	Traditional statistical techniques can be used to analyze data	Advanced analytics techniques such as

Feature	Small Data	Big Data
		machine learning are often require
Collection	Generally, it is obtained in an organized manner than is inserted into the database	The Big Data collection is done by using pipelines having queues like AWS Kinesis or Google Pub / Sub to balance high-speed data
Volume	Data in the range of tens or hundreds of Gigabytes	Size of Data is more than Terabytes
Analysis Areas	Data marts(Analysts)	Clusters(Data Scientists), Data marts(Analysts)
Quality	Contains less noise as data is less collected in a controlled manner	Usually, the quality of data is not guaranteed

Feature	Small Data	Big Data
Processing	It requires batch-oriented processing pipelines	It has both batch and stream processing pipelines
Database	SQL	NoSQL
Structure	Structured data in tabular format with fixed schema(Relational)	Numerous variety of data set including tabular data, text, audio, images, video, logs, etc.(Non Relational)
Hardware	A single server is sufficient	Requires more than one server
Optimization	Data can be optimized manually(human powered)	Requires machine learning techniques for data optimization

Feature	Small Data	Big Data
Storage	Storage within enterprises, local servers etc.	Usually requires distributed storage systems on cloud or in external file systems
People	Data Analysts, Database Administrators and Data Engineers	Data Scientists, Data Analysts, Database Administrators and Data Engineers
Nomenclature	Database, Data Warehouse, Data Mart	Data Lake
Infrastructure	Predictable resource allocation, mostly vertically scalable hardware.	More agile infrastructure with horizontally scalable hardware
Applications	Small-scale applications, such as personal or small business data management	Large-scale applications, such as enterprise-level data management, internet

Feature	Small Data	Big Data
		of things (IoT), and social media analysis

What is Data

data is a distinct piece of information that is gathered and translated for some purpose. If data is not formatted in a specific way, it does not valuable to computers or humans. Data can be available in terms of different forms

- Sound
- Video
- Single character
- Number (integer or floating-point)
- Picture
- Boolean (true or false)
- Text (string)

Types of Data

There are two types of data

. Qualitative Data: Qualitative data is information that represents some characteristics or attributes.

It depicts descriptions that cannot be counted, measured, or easily expressed with the help of numbers.

It can be collected from audio, text, and pictures. It is shared via data visualization tools, such as concept maps, clouds, infographics, timelines, and databases.

Qualitative data

Qualitative data can be analyzed through being either deductive or inductive approach.

In the deductive technique, the analyst starts with a question and evaluate data subjectively in terms of the question.

In the inductive technique, he or she simply evaluates the data to look for patterns as in this approach; the analyst has no agenda.

Quantitative data

Quantitative data is information that can be counted or measured—or, in other words, quantified—and given a numerical value.

Quantitative data is anything that **can be counted in definite units and numbers**. So, among many, many other things, some examples of quantitative data include:

- Revenue in dollars
- Weight in kilograms or pounds
- Age in months or years
- Distance in miles or kilometers