Mapreduce was not sufficient
It takes lots of time for processing data as it ~~fo~~ works on batch processing
So, the Spark came to overcome all the problems

Spark uses in memory computation i.e. It takes all the data in RAM & process there where as in Hadoop first of all data is stored in hard-disk and then it is loaded to RAM for processing, which takes lots of time. If the size of RAM is less than data then most of the data is stored in RAM & ~~rest~~ rest in the harddisk.
In hadoop if we want to do different tasks we have to install different tool where as Spark is can do different processing like

1- Batch processing
2- machine learning
3- Streaming data
4- Graph data processing

Spark can take data from almost all the platform & can process data of any type.

Components of Spark.

* Spark Core :- Contains the basic functionality of Spark like task scheduling, fault recovery etc
* Spark sql :- It is Spark package for working with structured data
* Spark Streaming :- Enable processing of live streams
* MLlib :- Spark library which provide multiple machine-learning algo
* Graphx :- Library for manipulating graphs

Basics of Spark

RDD :- It is a Resilient Distributed Dataset. RDD is a Dataset which is distributed among clusters & have the power of recovery of the failed.

Page :
Date: / /

Features of RDD :-

1.) Fault tolerant, If the RDD is lost then we can gain data from previous RDD
2.) RDD is immutable i.e. we can not change the Data of RDD
3.) Lazy Evaluation :- Spark uses Lazy Evaluation it does not execute anything until is it asked to do so, it make only DAG