

MapReduce is a programming model designed for processing large volumes of data in parallel by dividing the work into a set of independent tasks.

Hadoop is capable of running MapReduce programs written in various languages: Java, Ruby, Python, and C++.

The whole process of Map reduce goes through four phases of execution:

- 1. Splitting
- 2. Mapping
- 3. Shuffling
- 4. Reducing

For example, take following data as input for the Map Reduce.

```
Easy exam notes
Notes are good
Notes are bad
```

The final output of the MapReduce task is:

Words	Counts
Easy	1
Exam	1
Notes	3

Words	Counts
Are	2
Good	1
Bad	1

The process of Map Reduce is:

1. Splitting: Input divided into fixed-size pieces called input splits.
2. Mapping: In this phase data in each split is passed to a mapping function to produce output values.
3. Shuffling: The same words are clubed together along with their respective frequency.
4. Reducing: This phase summarizes the complete dataset.