**Why Spark when Hadoop is already there?**

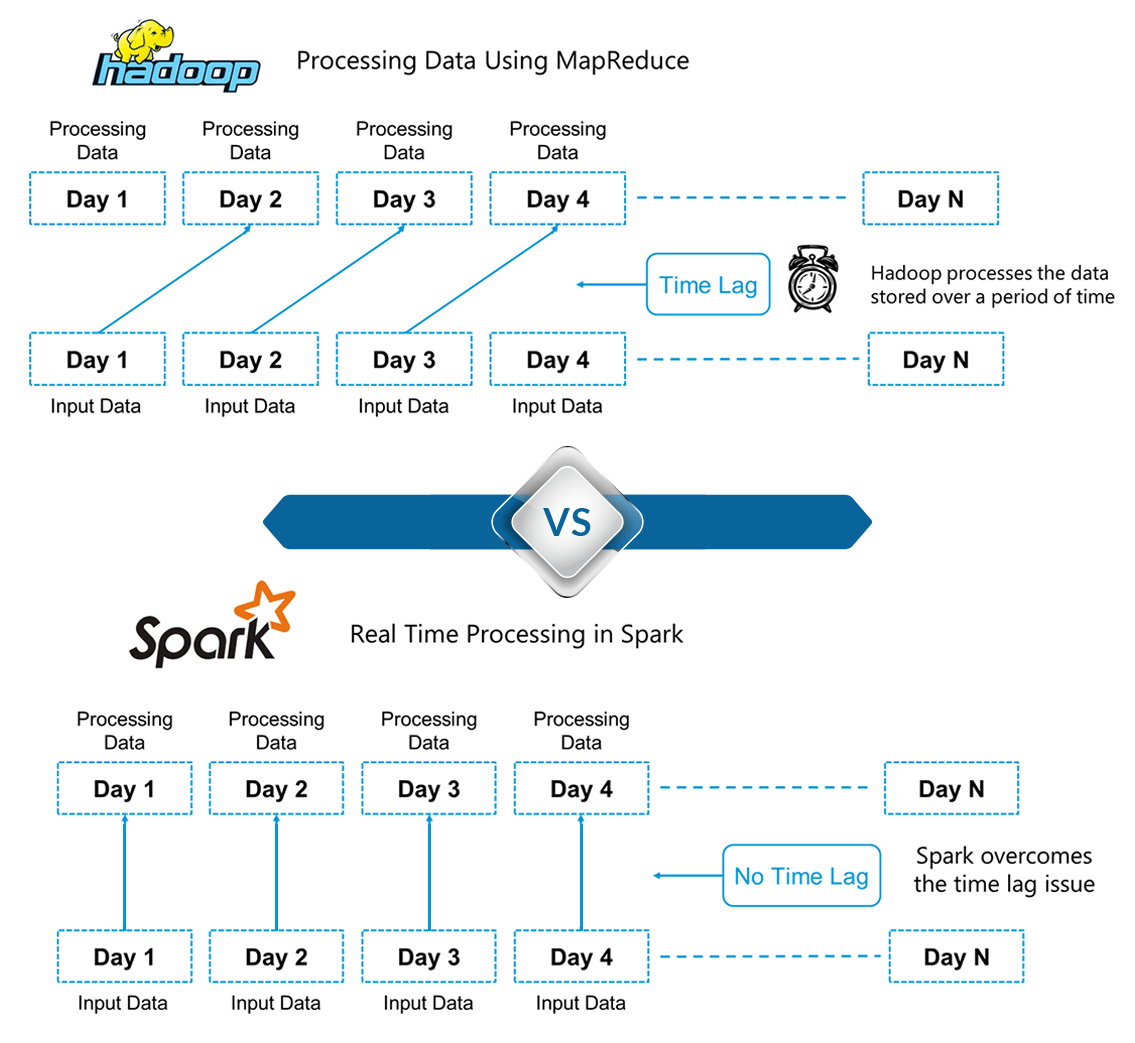
The first of the many questions everyone asks when it comes to Spark is, “***Why Spark when we have Hadoop already?*** “.

batch and real-time processing.

*Hadoop* is based on the concept of *batch processing* where the processing happens of blocks of data that have already been stored over a period of time.

This went on until 2014, till Spark overtook Hadoop. The USP for Spark was that it could*process data* in *real time* and was about 100 times faster than Hadoop MapReduce in batch processing large data sets.

The following figure gives a detailed explanation of the differences between processing in Spark and Hadoop.

****

**Figure:***Spark Tutorial – Differences between Hadoop and Spark*

Hadoop is based on batch processing of big data. This means that the data is stored over a period of time and is then processed using Hadoop. Whereas in Spark, processing can take place in real-time. This real-time processing power in Spark helps us to solve the use cases of Real Time Analytics

Alongside this, Spark is also able to do batch processing 100 times faster than that of Hadoop MapReduce (Processing framework in Apache Hadoop). *Therefore, Apache Spark**is the go-to tool for big data processing in the industry.*