

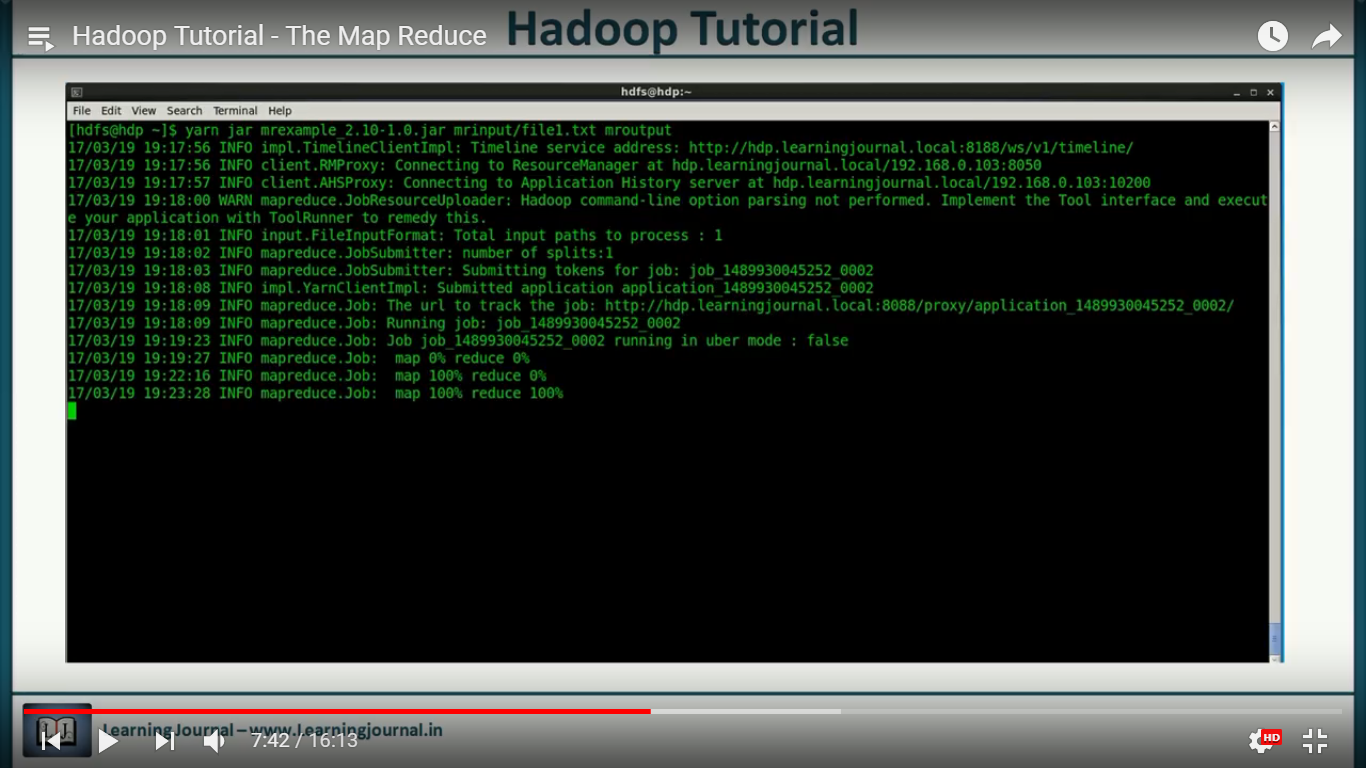
**What is Map Reduce?**

**How will a map function read lines from a block?**

**Who will collect data from each map computer and send it to the reduce computer?**

The execution engine takes care of all these things.

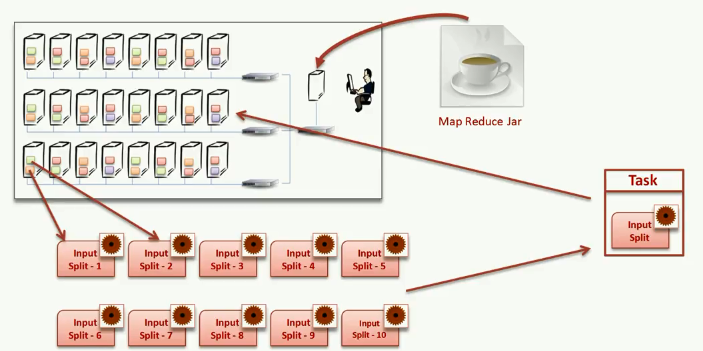
To run a MR Job we need to compile and package it in Jar using build tools like maven and sbt. Then we need to submit it to YARN for execution.



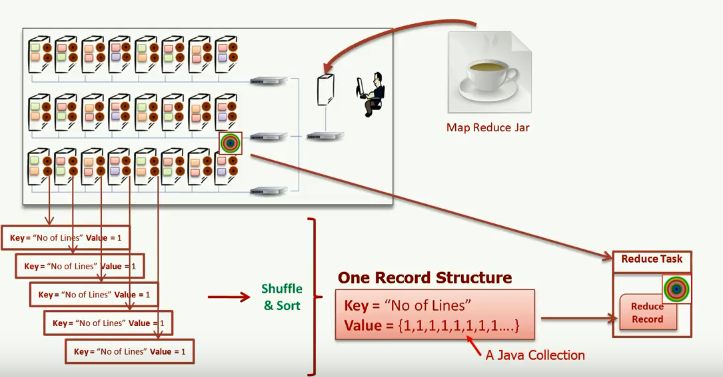
The YARN and MR execution framework take care of everything.

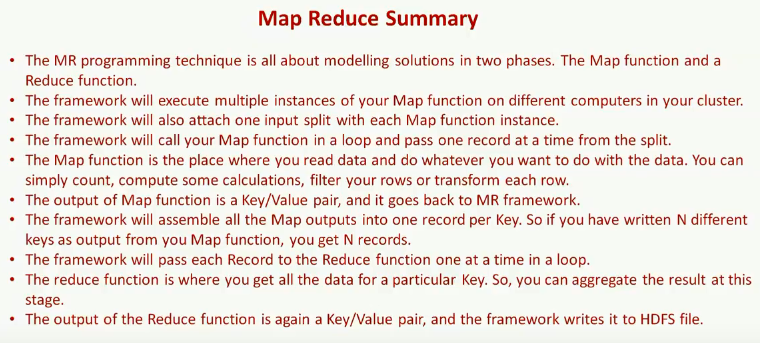


**How will a map function read lines from a block?**

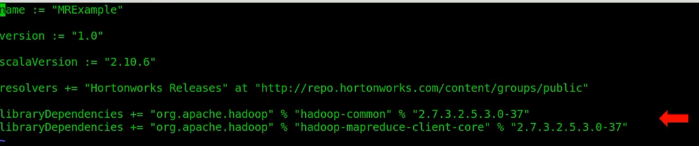


**Who will collect data from each map computer and send it to the reduce computer?**





Build File Example



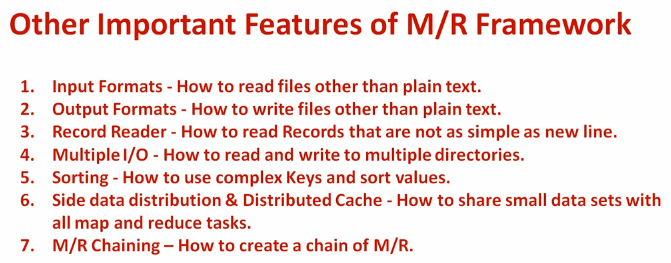
How and where to use log4j with Map reduce jobs?

Creating Multiple reducer.

<https://www.youtube.com/watch?v=Mc-kRwfaZj0&index=10&list=PLkz1SCf5iB4dw3jbRo0SYCk2urRESUA3v&pbjreload=10>

We don’t need reducer always.

How to use Tool Runner?



#### **40.What are the basic parameters of a Mapper?**

The basic parameters of a Mapper are

* LongWritable and Text
* Text and IntWritable

*Hadoop and Spark are the two most popular big data frameworks. But there is a commonly asked question – do we need Hadoop to run Spark? Watch this video to find the answer to this question.*

#### **How do Hadoop MapReduce works?**

There are two phases of MapReduce operation.

* Map phase – In this phase, the input data is split by map tasks. The map tasks run in parallel. These split data is used for analysis purpose.
* Reduce phase- In this phase, the similar split data is aggregated from the entire collection and shows the result.

#### **37. What is MapReduce? What is the syntax you use to run a MapReduce program?**

MapReduce is a programming model in Hadoop for processing large data sets over a cluster of computers, commonly known as HDFS. It is a parallel programming model.

The syntax to run a MapReduce program is – *hadoop\_jar\_file.jar /input\_path /output\_path****.***

#### **26. What are the configuration parameters in a “MapReduce” program?**

The main configuration parameters in “MapReduce” framework are:

* Input locations of Jobs in the distributed file system
* Output location of Jobs in the distributed file system
* The input format of data
* The output format of data
* The class which contains the map function
* The class which contains the reduce function
* JAR file which contains the mapper, reducer and the driver classes