



JAYOTI VIDYAPEETH WOMEN'S UNIVERSITY, JAIPUR
Government of Rajasthan established
Through ACT No. 17 of 2008 as per UGC ACT 1956
NAAC Accredited University

Faculty of Education and methodology

Department of Computer Science and Engineering

Faculty Name- Jv'n Narendra Kumar Chahar (Assistant Professor)

Program- B.Tech 6thSemester

Course Name – Web Intelligence, HADOOP and Big Data Analysis

Session no.: 10

Session Name- Hadoop - MapReduce

Academic Day starts with –

- Greeting with saying '**Namaste**' by joining Hands together following by 2-3 Minutes Happy session, Celebrating birthday of any student of respective class and **National Anthem**.

Lecture starts with- quotations' answer writing

- Review of previous Session- **Hadoop – Command Reference**

Topic to be discussed today- Today We will discuss about – **MapReduce**

- Lesson deliverance (ICT, Diagrams & Live Example)-
 - Diagrams

Introduction & Brief Discussion about the Topic – **An overview**

HADOOP -MAPREDUCE

MapReduce is a framework using which we can write applications to process huge amounts of data, in parallel, on large clusters of commodity hardware in a reliable manner.

What is MapReduce?

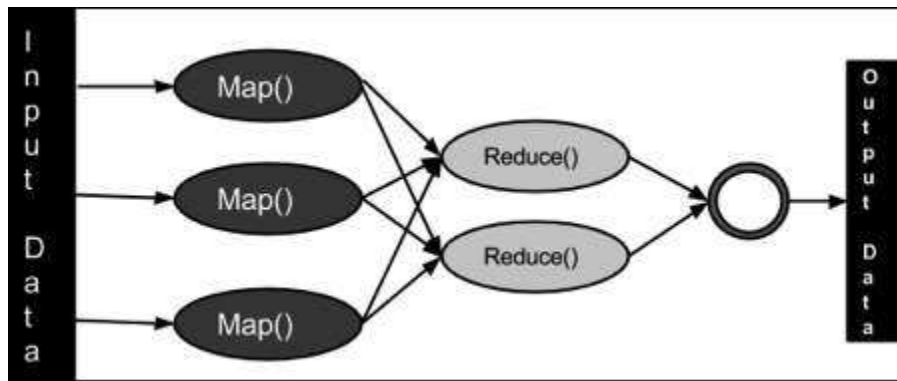
MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into *mappers* and *reducers* is sometimes nontrivial. But, once we write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change. This simple scalability is what has attracted many programmers to use the MapReduce model.

The Algorithm

- Generally, MapReduce paradigm is based on sending the computer to where the data resides!
- MapReduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.
 - **Map stage** – The map or mapper's job is to process the input data. Generally, the input data is in the form of file or directory and is stored in the Hadoop file system (HDFS). The input file is passed to the mapper function line by line. The mapper processes the data and creates several small chunks of data.

- **Reduce stage** – This stage is the combination of the **Shuffle** stage and the **Reduce** stage. The Reducer’s job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.
- During a MapReduce job, Hadoop sends the Map and Reduce tasks to the appropriate servers in the cluster.
- The framework manages all the details of data-passing such as issuing tasks, verifying task completion, and copying data around the cluster between the nodes.
- Most of the computing takes place on nodes with data on local disks that reduces the network traffic.
- After completion of the given tasks, the cluster collects and reduces the data to form an appropriate result, and sends it back to the Hadoop server.



References-

1. **Book:** Hadoop: The Definitive Guide by Tom White, 3rd Edition, O’reilly Hadoop in Action by Chuck Lam, MANNING Publications
2. **Online:** <https://www.tutorialspoint.com/>
3. **Online:** <http://www.oracle.com>

QUESTIONS: -

Q1. What is MapReduce in Hadoop?

Q2. What does the algorithm define for MapReduce?

Next, we will discuss about Inputs and Outputs (Java Perspective).

- Academic Day ends with-
National song 'Vande Mataram'