



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.: 11

Session Name- Inputs and Outputs (Java Perspective)

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 – MapReduce**

Topic to be discussed today- Today We will discuss about – **Inputs and Outputs (Java Perspective)**

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

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

Inputs and Outputs (Java Perspective)

The MapReduce framework operates on <key, value> pairs, that is, the framework views the input to the job as a set of <key, value> pairs and produces a set of <key, value> pairs as the output of the job, conceivably of different types.

The key and the value classes should be in serialized manner by the framework and hence, need to implement the Writable interface. Additionally, the key classes have to implement the Writable-Comparable interface to facilitate sorting by the framework. Input and Output types of a **MapReduce job** – (Input) <k1, v1> → map → <k2, v2> → reduce → <k3, v3>(Output).

	Input	Output
Map	<k1, v1>	list (<k2, v2>)
Reduce	<k2, list(v2)>	list (<k3, v3>)

Terminology

- **PayLoad** – Applications implement the Map and the Reduce functions, and form the core of the job.
- **Mapper** – Mapper maps the input key/value pairs to a set of intermediate key/value pair.
- **NamedNode** – Node that manages the Hadoop Distributed File System (HDFS).
- **DataNode** – Node where data is presented in advance before any processing takes place.
- **MasterNode** – Node where JobTracker runs and which accepts job requests from clients.
- **SlaveNode** – Node where Map and Reduce program runs.
- **JobTracker** – Schedules jobs and tracks the assign jobs to Task tracker.
- **Task Tracker** – Tracks the task and reports status to JobTracker.
- **Job** – A program is an execution of a Mapper and Reducer across a dataset.
- **Task** – An execution of a Mapper or a Reducer on a slice of data.

- **Task Attempt** – A particular instance of an attempt to execute a task on a SlaveNode.

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 are inputs and outputs in perspective of Java?

Q2. Explain various keywords/terminology in Hadoop.

Next, we will discuss about Compilation and Execution of Process Units Program.

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