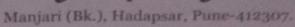
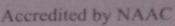


Pune District Education Association's College Of Engineering







Accredited by WANE	
DSBOAL	Assignment No: 11
*	Aim: - write a code in Jova for a single simple word count the number of occurrence of
	map reduce frame Hook on local stand alone setup
*	Theody: - map reduce word count is a frework Which splits the chunk of Jata, slots the
	map reduce tasks. A file-system stores the oip & iip of jobs. Re-execution of failed tas of the frame Hook and monitoring them is
	the task of the fromework.
*	map function- create and process the impost pata take in data converts it into a set of other
	data where the breakdown of individuals element into these tuples is done.
1	No API contract requiring a cortain
	reduce function:
7 7	mappers output is posted into the reduct

processes the data into something usable

overy single mapper is passed into the

reduced function.

The new old values are saved in HDFS.

A concept called streaming is used in a Writing a code for Word count.

In python using mapreduce, let's look at

In python using mapReduce, let's look at the reducer code and How to execute that using a streaming a Jar Alle.

* HOLH MapReduce WOOKS?

The mapReduce algorithm contains two important task, namely map and Reduce.

The map task takes a set of data & converts it into another set of data, where individual elements are broken down into tuples (key - value pairs).

mapreduce task takes the output from
the map as an input and combines those
data typies into a smaller set of typies.
The reduce task is always performed
after the map job.

D Input phase -

Here we have a record reader that translates each record that in on input file and sends that passed data to the mapper in the form of key.



Pune District Education Association's

College Of Engineering



Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

map is a user defined function, which takes a series of key value pairs and processes each of them to generate zero or more key value-pairs.

they key-value pairs generated by
the mapper are known as intermediate keys

* combines:- वहजन हिताय, बहजन सखाय।

A combiner is a type of local reducer that groups similar data from the map phase into identifiable sets.

shuffle and sout The reducer task starts with the shuffle and sout step.

Pune District Education Association

* REDUCES-

The Reduces takes the grouped keyvalue paired data as input and owns a
reduces function on each of them.

* conclusion: - Hence whe studied about

Q.D Explain Apache Hadoop:
- Apache Hadoop project develops open-

source slw for reliable, scalable & distributed computing:

-Apache Hadoop slld library is a frame Hook.

- It allows distributed processing of large data sets across clusters of computers using simple programming models.

* modules in Apache Hadoop:
O mapReduce
- This is a System for parallel processing of large sets.

This is a frame work for job scheduling and clusters resource management.

3) HDFS - (Hadoop Distributed file system).

-HDFS is a unique design that provides

storage for extermely large files

With streaming data.

W Hadoop common:
- These are common utilities that

Support the other Hadoop modules.

Q.2) Explain mapReduce.

mapReduce is a programming model for processing large datasets using parallel, distributed and clustered compute nodes-