Intro…

Data - 3T

Jvm

Jar

Winscp –

7zip

Putty

Exam – command for java programs, for Hadoop, for hdfs and ofcourse MCQs

Unix commands

Hive – sql command to process big data

Chapter 1:

Java, python, c++, ruby can be used for map reduce

1. Amazon – aws, installation is easier, no need to buy hardware
2. 2. Google vs Microsoft

Hadoop File – map r – processing

--Hdfs – storing

Slide 17

o/p of map ( key n value in this case) is i/p of reduce

Gzip: g is for GNU; GNU is an op sys, unix like but

Context allows the mapper/reducer application to interact with the Hadoop system environment and to get the environment information. It contains configuration data for the jobs as well as interfaces which allow it to emit output.

Through the context object, users can access data like:

1. Job information (job configuration, mapper/reducer classes names, job name);
2. Status of the currently executed tasks;
3. Current key, value, input split.

Integer = 32 bits; long integer = 64 bits

Substring (number1, number2) includes the starting character, but not the ending character

+ve sign start from 0 so 88 to 92

88 89 90 91 92

0 1 2 3 4

So start from 89 to 92

For -ve sign, starts from 1

So 88 to 91 so 4

VALUES(102,25,65,458,125) where 102 is 10.2

Value = 102

S3 = Hadoop file system HDFS

|

|

EMR – Map Reduce – processing component