

Welcome to
Big Data & Hadoop
Session

Session 4 - Adv. Map Reduce



+91-9538998962

sandeep@knowbigdata.com

WELCOME - KNOWBIGDATA

- Interact - Ask Questions
- Lifetime access of content
- Class Recording
- Cluster Access
- 24x7 support
- Real Life Project
- Quizzes & Certification Test
- 10 x (3hr class)
- Socio-Pro Visibility
- Mock Interviews



ABOUT ME

2014	KnowBigData	Founded
2014	Amazon	Built High Throughput Systems for Amazon.com site using in-house NoSql.
2012		
2012	InMobi	Built Recommender after churning 200 TB
2011		
	tBits Global	Founded tBits Global Built an enterprise grade Document Management System
2006		
	D.E.Shaw	Built the big data systems before the term was coined
2002		
2002	IIT Roorkee	Finished B.Tech somehow.



Hadoop

Know BIG DATA

Sandeep Giri

COURSE CONTENT

I	Understanding BigData, Hadoop Architecture
II	Environment Overview, MapReduce Basics
III	Adv MapReduce & Testing
IV	Analytics using Pig
V	Analytics using Hive
VI	NoSQL, HBASE
VII	Oozie, Mahout,
VIII	Zookeeper, Apache Storm
IX	Apache Flume, Apache Spark
X	YARN, Big Data Sets & Project Assignment



TODAY'S CLASS

- **Map Reduce Adv. Concepts**
 - **Input Splits**
 - **Data Locality Optimization)**
 - **Records**
 - **Multiple Reducers**
 - **Combiners**
 - **Partitioner**
 - **Network Topology - Rack Awareness**
-



MAP / REDUCE

Discussion - Problem I

Code M/R for all the problems in your favourite language.



MAP / REDUCE

Discussion - Problem 2

Based on the content from a very large text archive, formulate the next words recommendation.

For each word, prepare a top 5 recommendations of the word that would go next.

happy birthday, newyear, marriage
how are, do, did

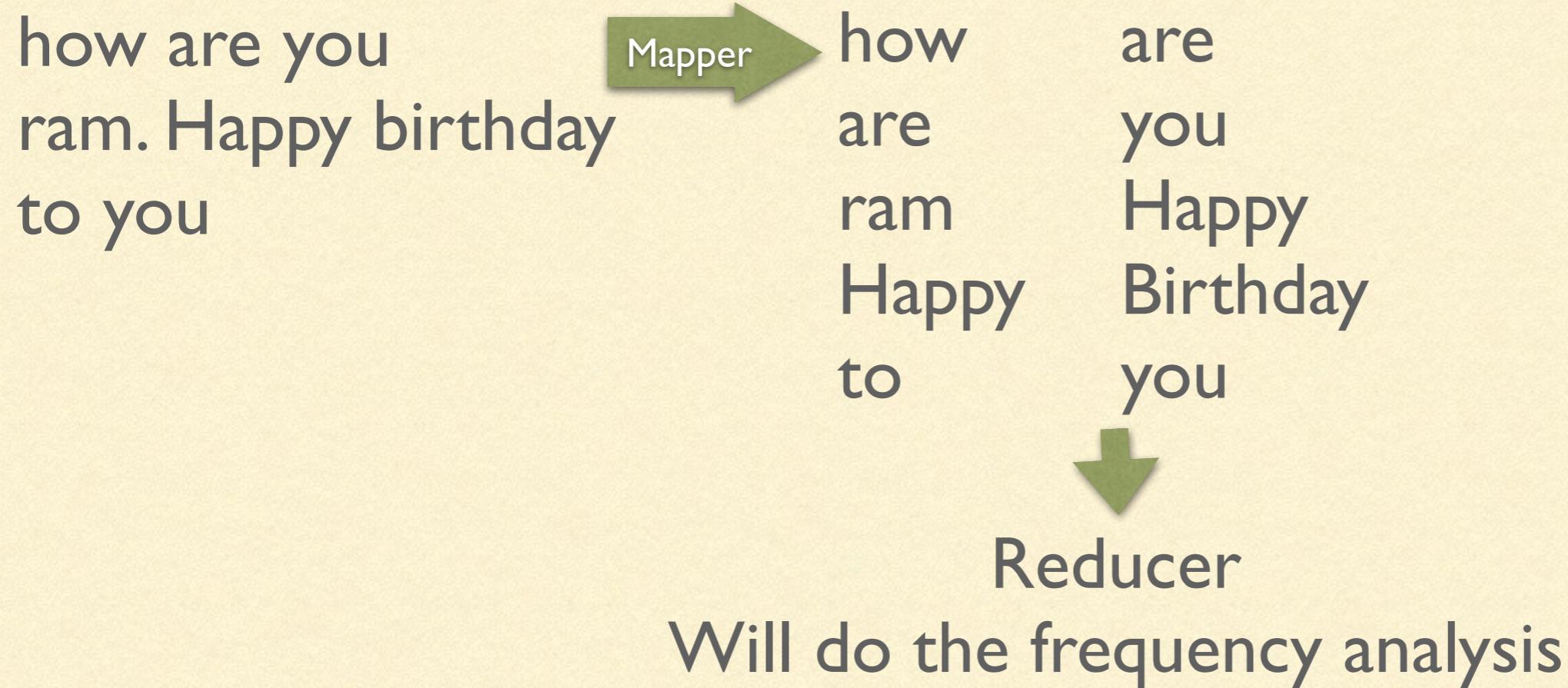
...

Download the content from wikipedia using this:

<http://www.evanjones.ca/software/wikipedia2text.html>



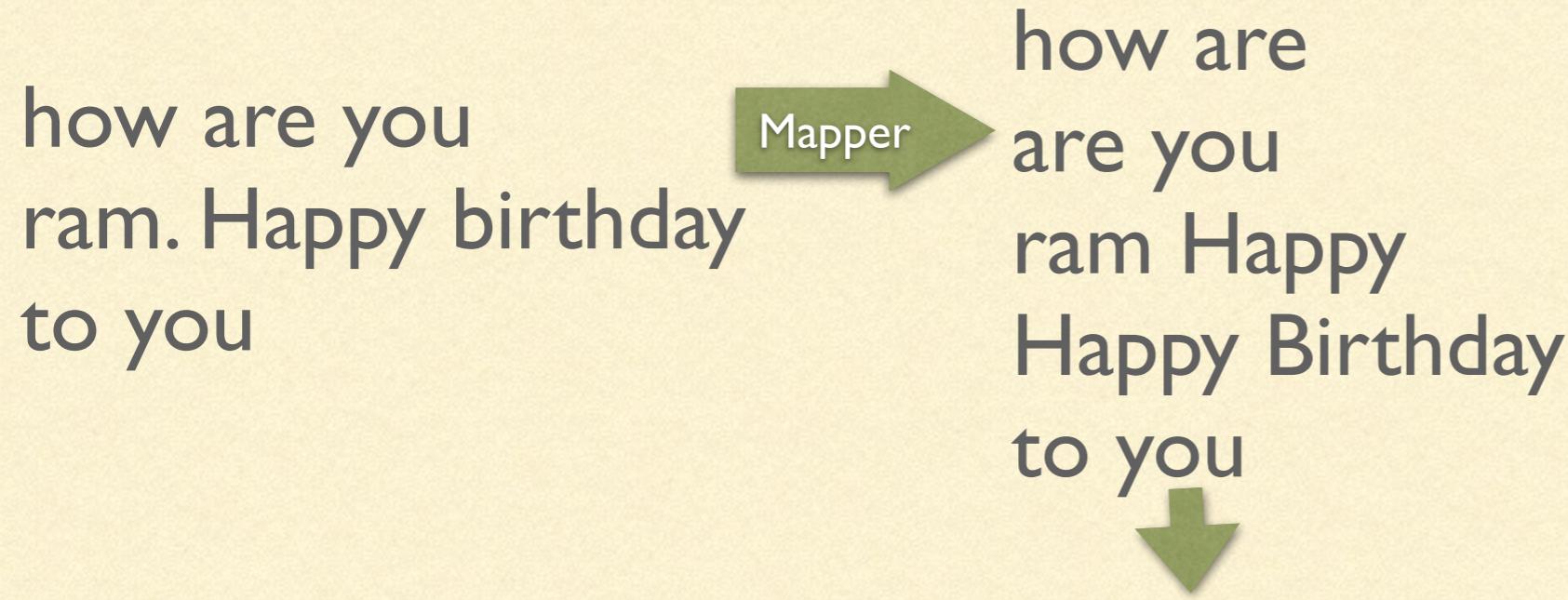
MAP / REDUCE



Problems?



MAP / REDUCE



Reducer

Will do the frequency analysis

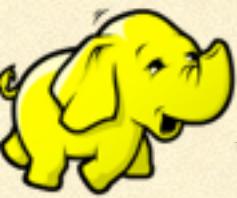
Problems:

1. Last word of line should pair with first word of next line
2. The last word of a sentence should not be paired
3. Computation of frequency can further be mapred'ed



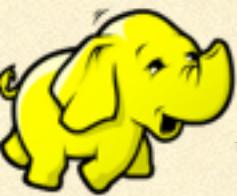
MAP / REDUCE

The data generated by the mapper is given to reducer and then it is sorted / shuffled [Yes/No]?



MAP / REDUCE

The data generated by the mapper is given to reducer and then it is sorted / shuffled [Yes/No]?

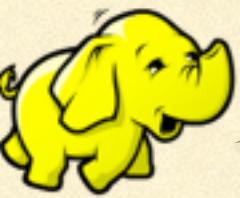


No. The output of mapper is first shuffled/sorted and then reducers are called.



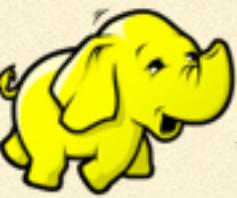
MAP / REDUCE

The mapper can only generate a single key value pair for an input value [Yes/No]?



MAP / REDUCE

The mapper can only generate a single key value pair for an input value [Yes/No]?

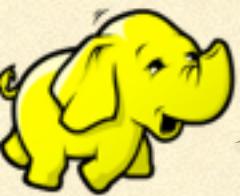


No. Mapper can generate as many key-value pair as it wants for an input.



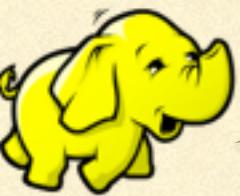
MAP / REDUCE

By default there is only one reducer in case of streaming job [Yes/No]?



MAP / REDUCE

By default there is only one reducer in case of streaming job [Yes/No]?



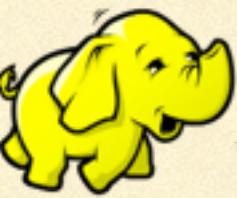
Yes. By default there is a single reducer job but it can be split by specifying cmd option : mapred.reduce.tasks
In case of Java MapReduce, the number of reducers are decided dynamically.



MAP / REDUCE

What is the role of job tracker?

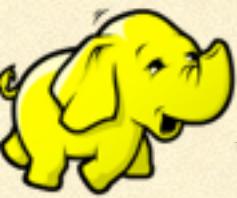
- A: Executing the Map/Reduce Logic
- B: Delegate the Map/Reduce Logic to task tracker.



MAP / REDUCE

What is the role of job tracker?

- A: Executing the Map/Reduce Logic
- B: Delegate the Map/Reduce Logic to task tracker.

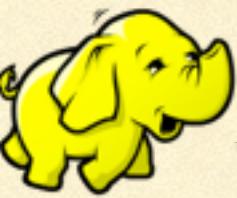


B.



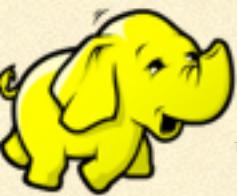
MAP / REDUCE

Q: The Map logic is executed preferably on the nodes that have the required data [Yes/No]?



MAP / REDUCE

Q: The Map logic is executed preferably on the nodes that have the required data [Yes/No]?

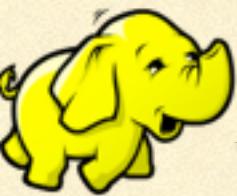


Yes.



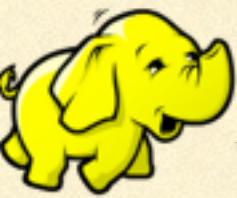
MAP / REDUCE

The whole of the chunk available on a node is fed to the Mapper [Yes/No]?



MAP / REDUCE

The whole of the chunk available on a node is fed to the Mapper [Yes/No]?

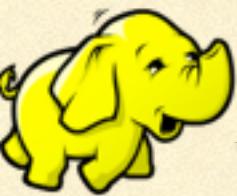


No. Input is split using InputSplits and part of which is transferred to other mappers

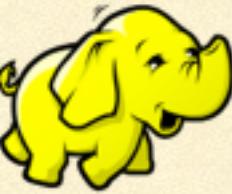


MAP / REDUCE

Where does Hadoop Store the result of reducer?
In HDFS or Local File System?



MAP / REDUCE



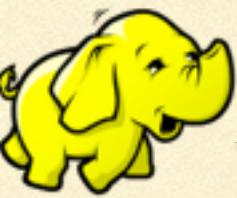
Where does Hadoop Store the result of reducer?
In HDFS or File System?

In HDFS.



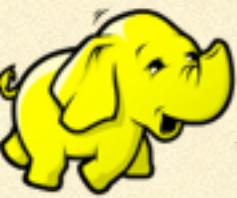
MAP / REDUCE

Where does Hadoop Store the intermediate data such as
output of Map Tasks?
In HDFS or File System?



MAP / REDUCE

Where does Hadoop Store the intermediate data such as
output of Map Tasks?
In HDFS or File System?



In filesystem.



MAP / REDUCE - JAVA

Why in Java?

- Primary Support
- Can modify behaviour to a very large extent

Writing Map-Reduce in Java

1. Install Eclipse
2. Create a Java project
3. Add Libs: hadoop-mapreduce-client-core.jar, hadoop-common.jar
4. Change JDK from 7 to 6.0
5. Change the Java Compiler settings to 1.6



MAP / REDUCE - JAVA

6. Create a Mapper

Byte
Offset

Input
Line

```
public class StubMapper extends Mapper<Object, Text, Text, LongWritable> {  
  
    @Override  
    public void map(Object key, Text value, Context context)  
        throws IOException, InterruptedException {  
  
        String[] words = value.toString().split("[ \t]+");  
        for(String word:words)  
        {  
            context.write(new Text(word), new LongWritable(1));  
        }  
    }  
}
```



MAP / REDUCE - JAVA

7. Create a Reducer

```
public class StubReducer extends Reducer<Text, LongWritable, Text, LongWritable> {  
  
    @Override  
    public void reduce(Text key, Iterable<LongWritable> values, Context context)  
        throws IOException, InterruptedException {  
  
        long sum = 0;  
        for(LongWritable iw:values)  
        {  
            sum += iw.get();  
        }  
        context.write(key, new LongWritable(sum));  
    }  
}
```



MAP / REDUCE - JAVA

8. Create a Driver

```
public class StubDriver {  
  
    public static void main(String[] args) throws Exception {  
        Job job = new Job(conf, "wordcount");  
        job.setJarByClass(StubDriver.class);  
        job.setMapperClass(StubMapper.class);  
        job.setReducerClass(StubReducer.class);  
        job.setOutputKeyClass(Text.class);  
        job.setOutputValueClass(LongWritable.class);  
        job.setMapOutputKeyClass(Text.class);  
        job.setMapOutputValueClass(LongWritable.class);  
        FileInputFormat.addInputPath(job, new Path(args[0]));  
        FileOutputFormat.setOutputPath(job, new Path(args[1]));  
        boolean result = job.waitForCompletion(true);  
        System.exit(result ? 0 : 1);  
    }  
}
```



MAP / MAP / REDUCE - JAVA

Writing Map-Reduce in Java (Continued)

9. Export jar

10.scp jar to the hadoop server

11.Run it using the following command:

`hadoop jar sandeep/training2.jar StubDriver <arguments>`

e.g: `hadoop jar sandeep/training2.jar StubDriver /users/root/wordcount/input /users/root/wordcount/output16/`

12. In case there is a need use -use-lib

13. Testing: Add all the jars provided

Using external Jars:

`$ export LIBJARS=/path/jar1,/path/jar2`

`$ hadoop jar my-example.jar com.example.MyTool -libjars ${LIBJARS}`



MAP / REDUCE - JAVA

14. Create Test Case

```
public class StubTest {  
  
    @Before  
    public void setUp() {  
        mapDriver = new MapDriver<Object, Text, Text, LongWritable>();  
        mapDriver.setMapper(new StubMapper());  
        reduceDriver = new ReduceDriver<Text, LongWritable, Text, LongWritable>();  
        reduceDriver.setReducer(new StubReducer());  
        mapReduceDriver = new MapReduceDriver<Object, Text, Text, LongWritable, Text, LongWritable>();  
        mapReduceDriver.setMapper(mapper);  
        mapReduceDriver.setReducer(reducer);  
    }  
  
    @Test  
    public void testXYZ() {  
        ....  
    }  
}
```



MAP / REDUCE - JAVA

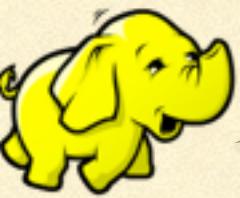
15. Create a test case

```
@Test  
public void testMapReduce() throws IOException {  
  
    mapReduceDriver.addInput(new Pair<Object, Text>  
        ("1", new Text("sandeep giri is here")));  
    mapReduceDriver.addInput(new Pair<Object, Text>  
        ("2", new Text("teach the map and reduce class is fun.")));  
    List<Pair<Text, LongWritable>> output = mapReduceDriver.run();  
    for (Pair<Text, LongWritable> p : output) {  
        System.out.print(p.getFirst() + “-“ + p.getSecond());  
        //assert here  
        ....  
    }  
}
```



MAP / REDUCE

BREAK for 10 mins



MAP / REDUCE - ADVANCE CONCEPTS

Input Splits

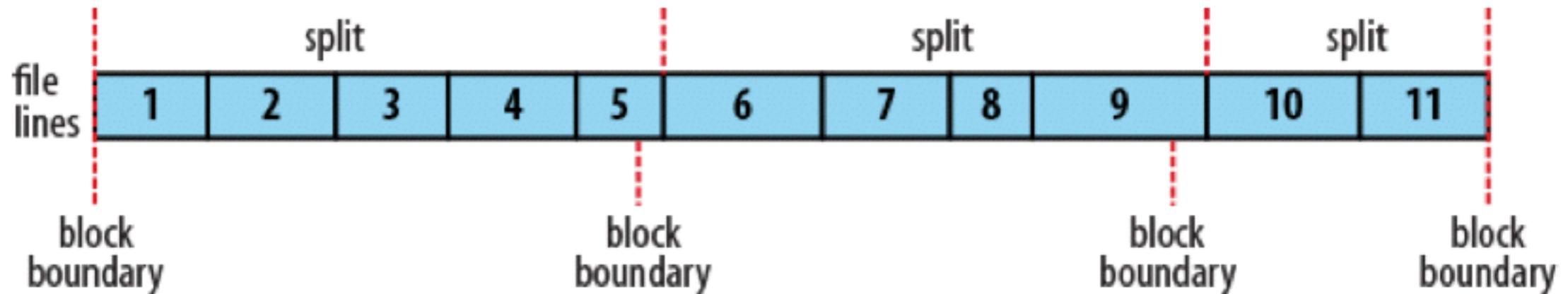
- Hadoop divides input into fixed-sized pieces called Splits
- One mapper instance per input split
- Map runs on each *record* of split
- Small Splits => too many => High Overhead
- Good Split Size tends to be equal to HDFS Block (64MB)
- Hadoop tries to run maps where data is located
 - => Data Locality Optimisation
- If all nodes hosting replica of data busy
 - Maps execute at the same rack different node
 - Otherwise off-rack is used

```
public abstract class InputSplit {  
    public abstract long getLength()  
    public abstract String[] getLocations()  
}
```



MAP / REDUCE - INPUT SPLITS (CONT.)

The Relationship Between Input Splits and HDFS Blocks



Line1 - 25MB
Line2 - 25MB
Line3 - 25MB

Ch1
Line1 - 25MB
Line2 - 25MB
Line3 - 14MB

ch2
Line3 - 11MB

MAP / REDUCE - INPUT SPLITS (CONT.)

- Has length and locations
- Largest gets processed first
- InputFormat creates splits
 - Default one is TextInput Format
 - Extend it to custom splits/records

```
public abstract class InputSplit {  
    public abstract long getLength()  
    public abstract String[] getLocations()  
}
```

```
public abstract class InputFormat {  
    List<InputSplit> getSplits (JobContext);  
    RecordReader createRecordReader  
        (InputSplit,TaskAttemptContext);  
}
```



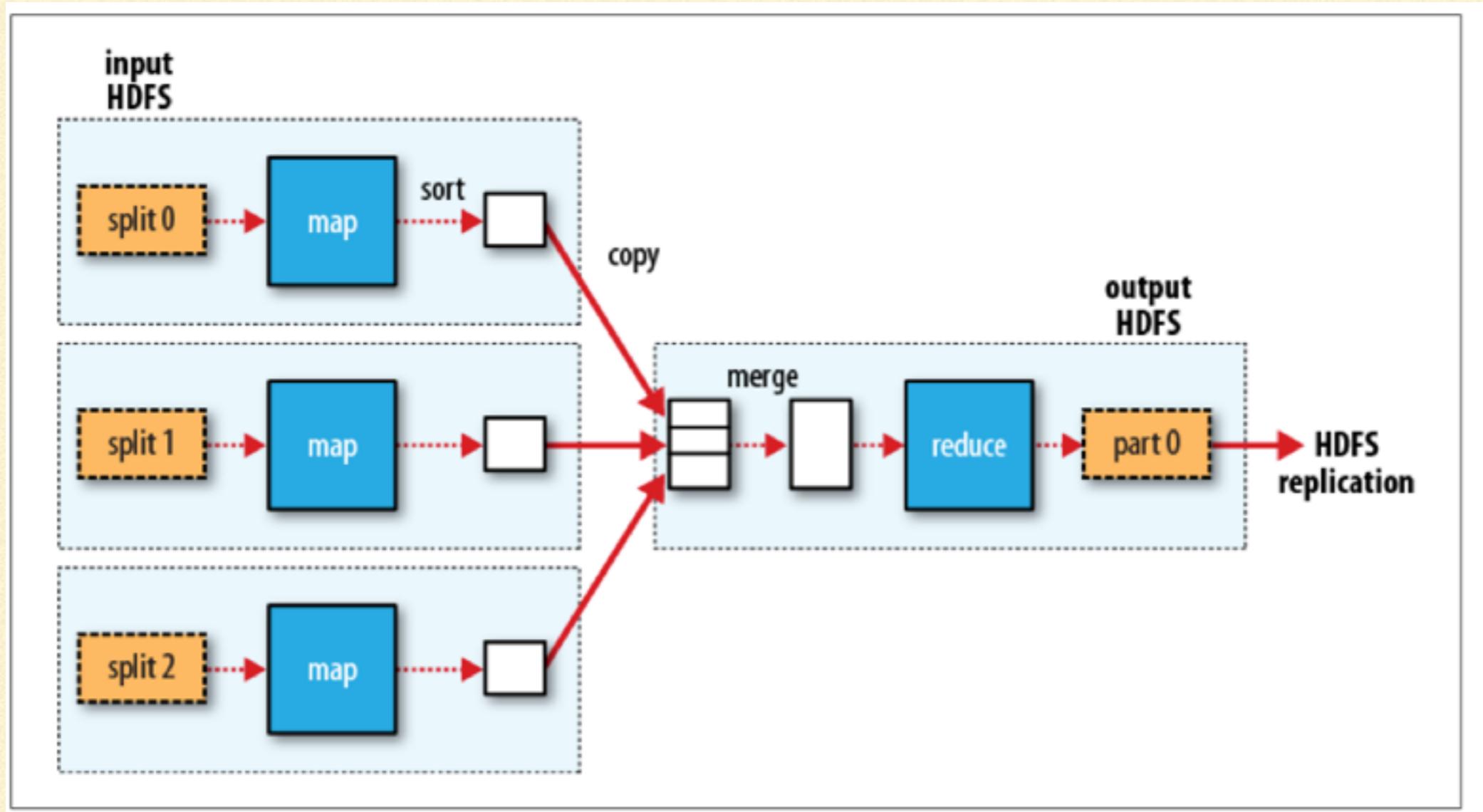
MAP / REDUCE - RECORDS

Records

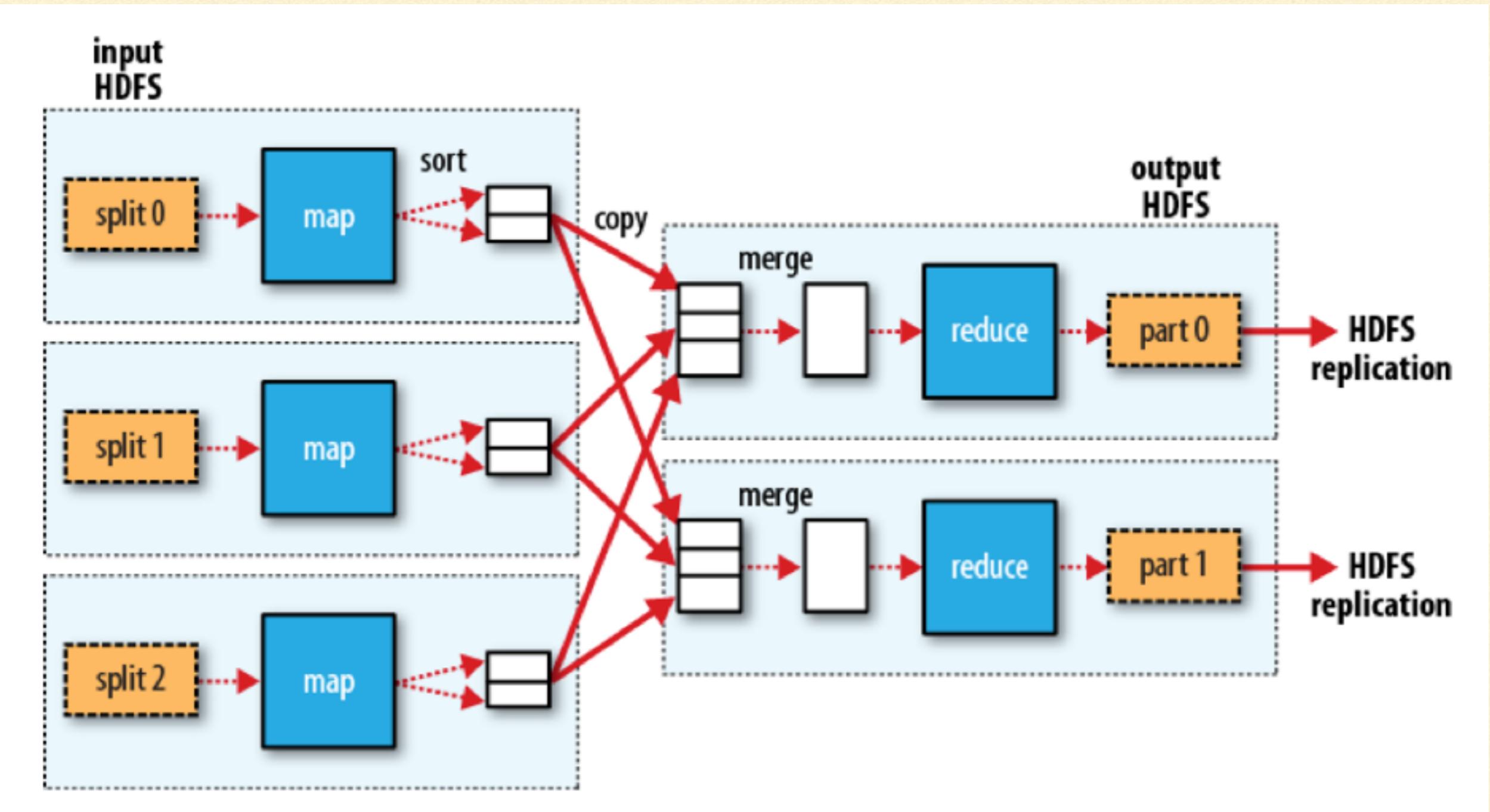
- Each Input Split is further divided into records
- A map function is applied on each record
- A record is a logical entity
- Analogy: RDMBS:: Row
- InputFormat creates Records



MAP / REDUCE - DATA FLOW WITH SINGLE REDUCER



MAP / REDUCE - MULTIPLE REDUCERS



MAP / REDUCE - HOW MANY REDUCERS?

- By Default One
- Too many reducers effort of shuffling is high
- Too few reducers, computation takes time
- Tune it to have total number of slots



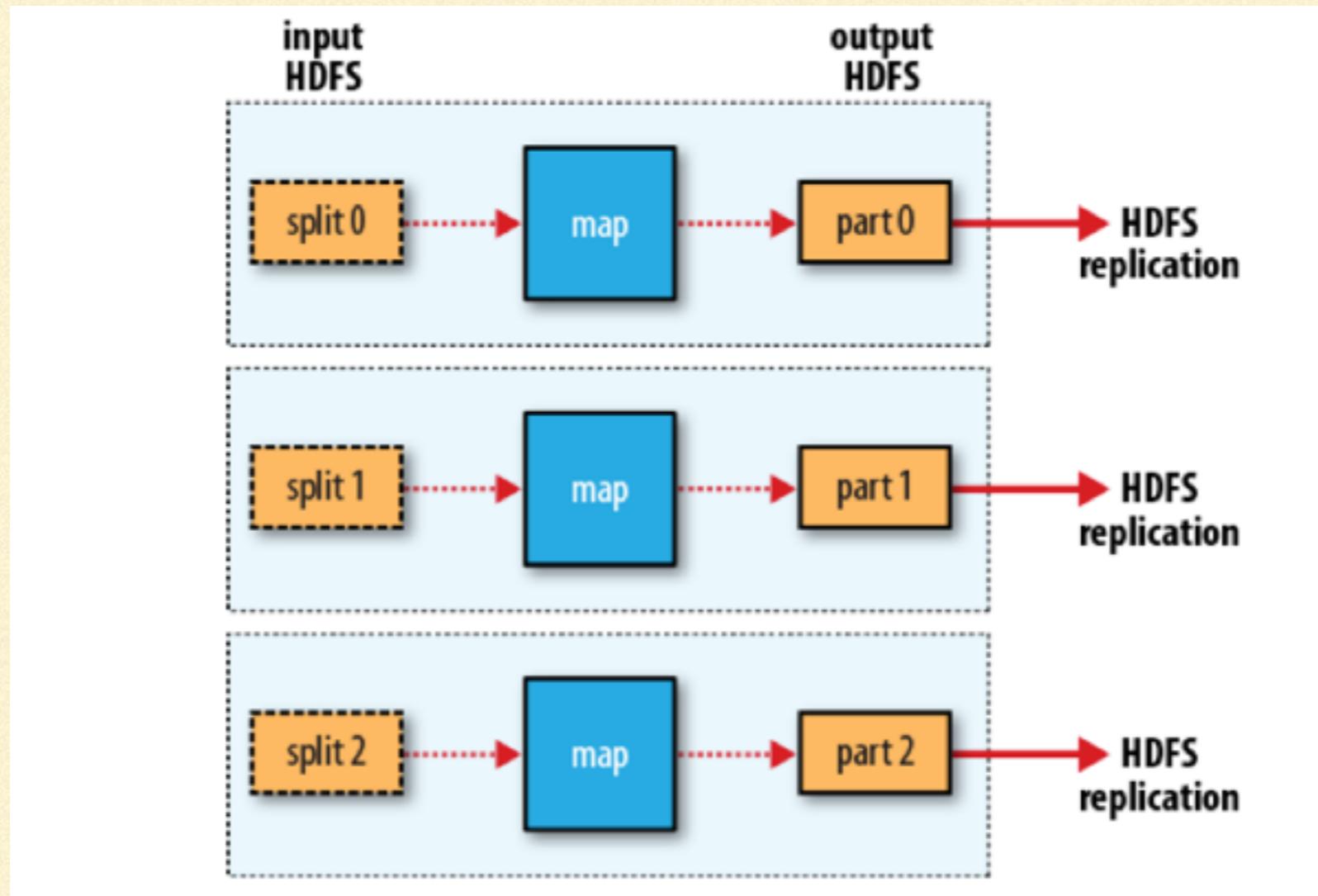
MAP / REDUCE - PARTITIONER

- Defines the key for partitioning
- Decides which data goes to which reducers

```
public static class AgePartitioner extends Partitioner<Text, Text> {  
    public int getPartition(Text ageInt, Text value, int numReduceTasks) {  
        //this is done to avoid performing mod with 0  
        if(numReduceTasks == 0)  
            return 0;  
        //if the age is <20, assign partition 0  
        if(ageInt <=20)  
            return 0;  
        //else if the age is between 20 and 50, assign partition 1  
        if(ageInt >20 && ageInt <=50)  
            return 1 % numReduceTasks; // = 1000 % 2 <= 2  
        //otherwise assign partition 2  
        else  
            return 2 % numReduceTasks;  
    }  
}
```



MAP / REDUCE - COMBINER FUNCTIONS



- Runs on the same node after Map has finished
- Processes the output of Map
- Helps in minimise the data transfer
- Does not replace reducer



MAP / REDUCE - COMBINER FUNCTIONS

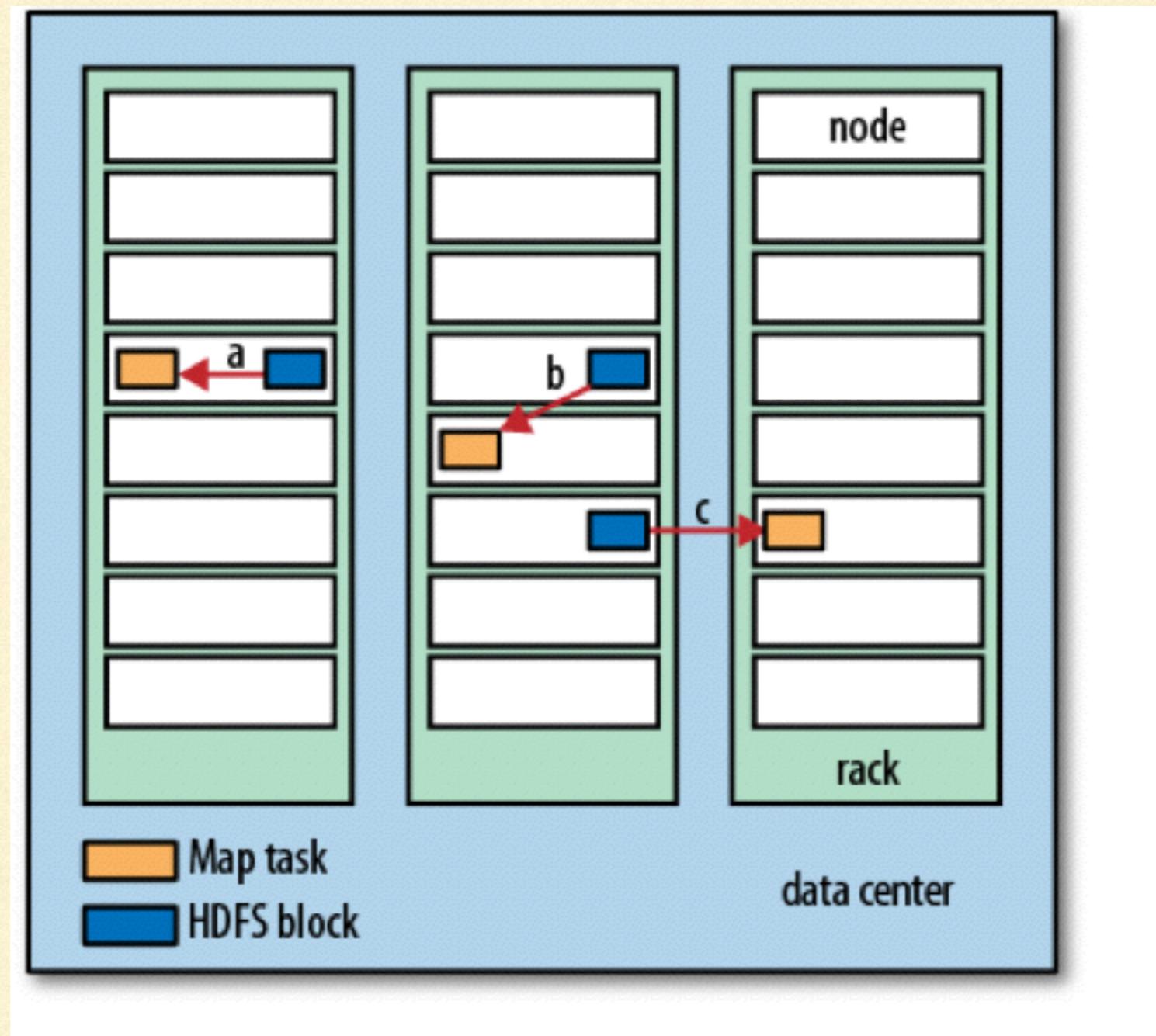
- Defined Using Reducer Class
- No matter in what way it is applied, output should be same
- Examples: Sum, Min, Max
- $\max(0, 20, 10, 25) = \max(\max(0, 20), \max(10, 25)) = \max(20, 25) = 25$
 - $= \max(\max(0, 10), \max(20, 25)) = \max(10, 25) = 25$
- Not: average or mean
 - $= \text{avg}(\text{avg}(0, 20), \text{avg}(10, 25)) = \text{avg}(10, 17.5) = 13.75$
 - $= \text{avg}(\text{avg}(0, 10, 20), \text{avg}(25)) = \text{avg}(10, 25) = 17.5$
- is function $f(a, b, c, \dots) = \{\text{return } \sqrt{a^2 + b^2 + c^2};\}$

```
job.setCombinerClass(MaxTemperatureReducer.class);
```



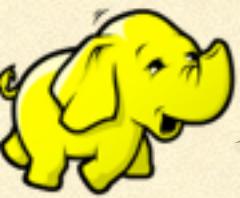
MAP / REDUCE - ADVANCE CONCEPTS

Data Locality Optimisation



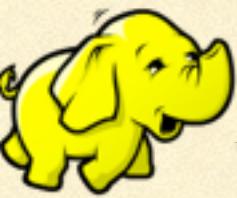
MAP / REDUCE

Can reducer take a benefit of Data Locality Optimization?



MAP / REDUCE

Can Reduce task take a benefit of Data Locality Optimization?

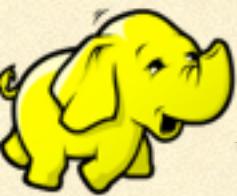


No. Input to reducer is from all maps generally.



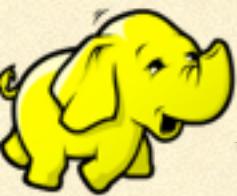
MAP / REDUCE

If you have to store a file with replication factor of 3, on which all nodes will hadoop store this file?



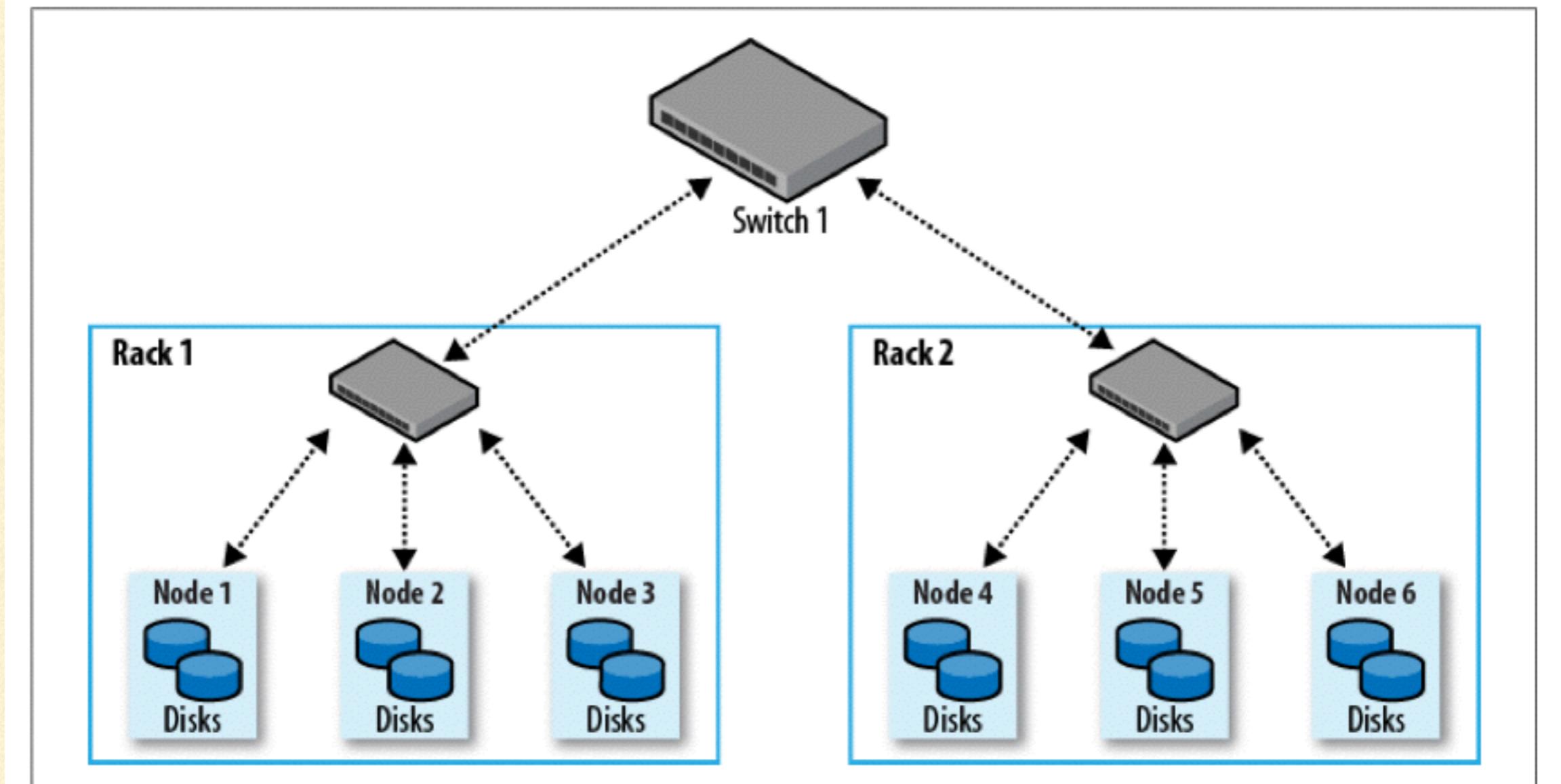
MAP / REDUCE

If you have to store a file with replication factor of 3, how will HDFS store it?



The first one will be on the local node, other two will be on off-rack.

MAP / REDUCE - NETWORK TOPOLOGY



MAP / REDUCE - RACK AWARENESS

Rack Awareness

- Understands the networks as a tree
- Similar to file path /a/b/c
- DNSToSwitchMapping provides translation
 - from machine name to network location
- Config Script Location Key: *topology.script.file.name*
- Distance is computed as:
 - Sum of distance from Common Ancestor

192.168.1.1 => /192/168/1/1

192.168.1.2 => /192/168/1/2



MAP / REDUCE - RACK AWARENESS

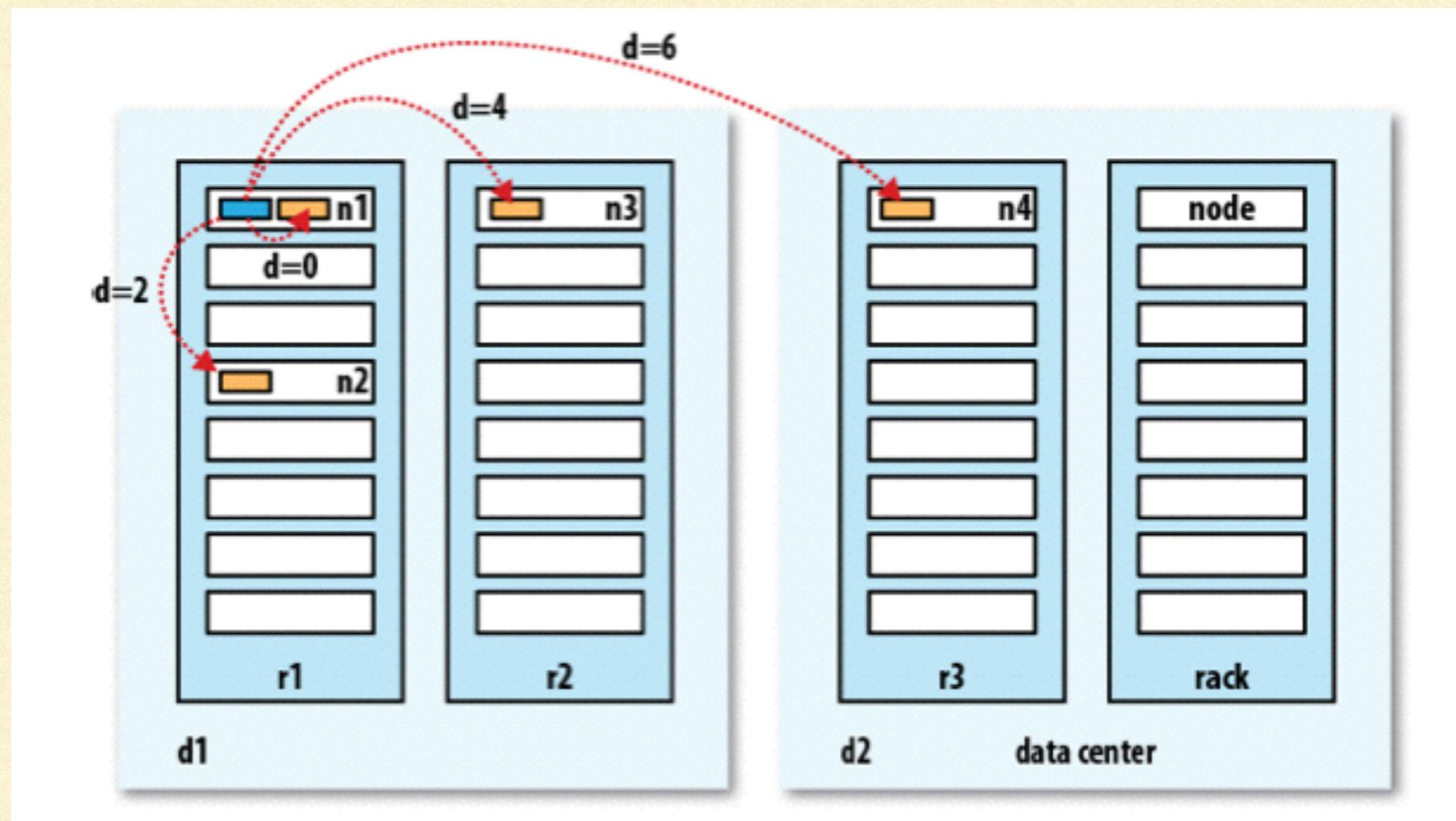
Bandwidth

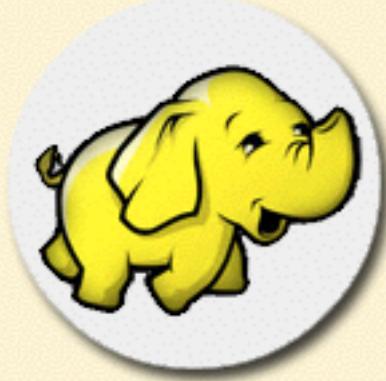
$\text{distance}(/d1/r1/n1, /d1/r1/n1) = 0$ (processes on the same node)

$\text{distance}(/d1/r1/n1, /d1/r1/n2) = 2$ (different nodes on the same rack)

$\text{distance}(/d1/r1/n1, /d1/r2/n3) = 4$ (nodes on different racks in the same data centre)

$\text{distance}(/d1/r1/n1, /d2/r3/n4) = 6$ (nodes in different data centres)





Big Data & Hadoop

Thank you.



+91-9538998962

sandeep@knowbigdata.com