

BIG DATA LAB 3
NAME: JESWIN MS
USN: 1MS21CI024

Objective: How to write a mapreduce odd even program and execute on hadoop

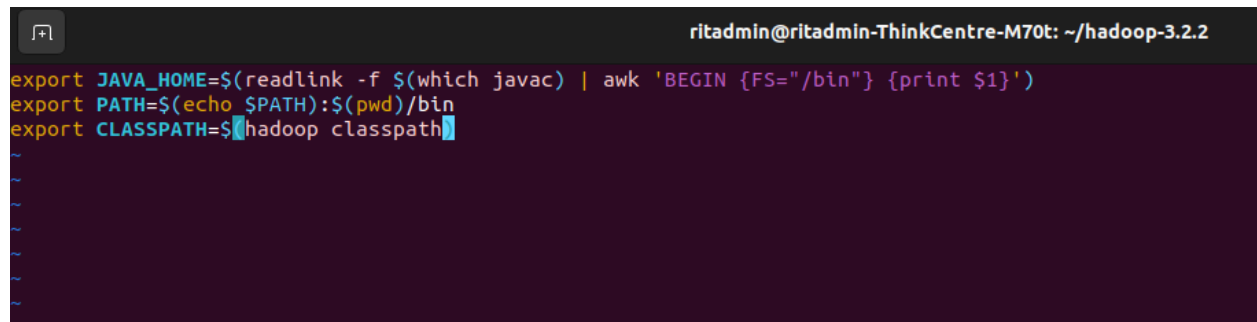
Setup:

- We download the `hadoop-3.2.2.tar.gz` file and extract the file contents in a folder in home directory
- We create a `bash.sh` file inside the folder using `vim bash.sh`. If vim is not installed, use `sudo apt install vim`
- Next we paste this content below inside the `bash.sh` file.

```
export JAVA_HOME=$(readlink -f $(which javac) | awk 'BEGIN
{FS="/bin"} {print $1}')
```

```
export PATH=$(echo $PATH):$(pwd)/bin
```

```
export CLASSPATH=$(hadoop classpath)
```

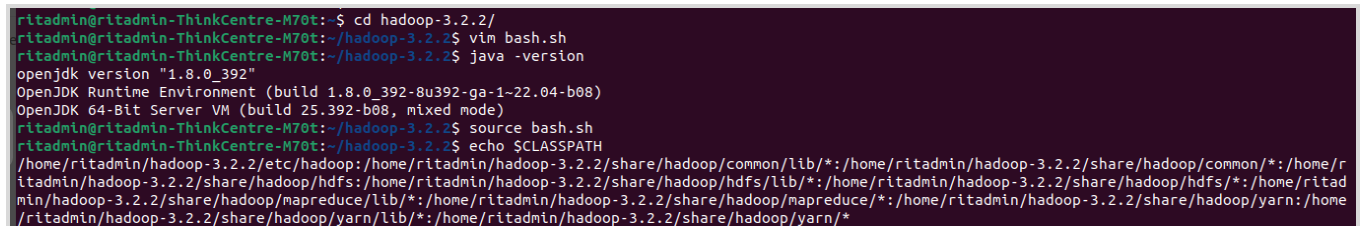
A terminal window with a dark purple background. The prompt is 'ritadmin@ritadmin-ThinkCentre-M70t: ~/hadoop-3.2.2'. The user has entered three lines of code: 'export JAVA_HOME=\$(readlink -f \$(which javac) | awk 'BEGIN {FS="/bin"} {print \$1}''', 'export PATH=\$(echo \$PATH):\$(pwd)/bin', and 'export CLASSPATH=\$(hadoop classpath)'. The cursor is at the end of the third line.

```
ritadmin@ritadmin-ThinkCentre-M70t: ~/hadoop-3.2.2
export JAVA_HOME=$(readlink -f $(which javac) | awk 'BEGIN {FS="/bin"} {print $1}')
```

```
export PATH=$(echo $PATH):$(pwd)/bin
```

```
export CLASSPATH=$(hadoop classpath)
```

- Run the command `source bash.sh`

A terminal window with a dark purple background. The prompt is 'ritadmin@ritadmin-ThinkCentre-M70t: ~'. The user has entered several commands: 'cd hadoop-3.2.2/', 'vim bash.sh', 'java -version', 'source bash.sh', and 'echo \$CLASSPATH'. The output shows the Java version and the full classpath for Hadoop.

```
ritadmin@ritadmin-ThinkCentre-M70t:~$ cd hadoop-3.2.2/
```

```
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ vim bash.sh
```

```
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ java -version
```

```
openjdk version "1.8.0_392"
```

```
OpenJDK Runtime Environment (build 1.8.0_392-8u392-ga-1-22.04-b08)
```

```
OpenJDK 64-Bit Server VM (build 25.392-b08, mixed mode)
```

```
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ source bash.sh
```

```
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ echo $CLASSPATH
```

```
/home/ritadmin/hadoop-3.2.2/etc/hadoop:/home/ritadmin/hadoop-3.2.2/share/hadoop/common/lib/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/common/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/hdfs:/home/ritadmin/hadoop-3.2.2/share/hadoop/hdfs/lib/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/hdfs/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/mapreduce/lib/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/mapreduce/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/yarn:/home/ritadmin/hadoop-3.2.2/share/hadoop/yarn/lib/*:/home/ritadmin/hadoop-3.2.2/share/hadoop/yarn/*
```

- Now the environment is set up in order to run the map reduce program.

Instructions for running odd even program:

driver.java

```
package oddeven;

import java.io.*;
import java.util.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.fs.Path;

public class driver
{
    public static void main(String args[]) throws IOException
    {
        JobConf conf=new JobConf(driver.class);
        conf.setMapperClass mapper.class);
        conf.setReducerClass(reducer.class);
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf,new Path(args[1]));
        JobClient.runJob(conf);
    }
}
```

mapper.java

```
package oddeven;

import java.io.*;
import java.util.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.io.*;

public class mapper extends MapReduceBase implements
Mapper<LongWritable , Text , Text , IntWritable>

{

public void map(LongWritable key,Text
value,OutputCollector<Text,IntWritable> output,Reporter r)
throws IOException

{

    String[] line=value.toString().split(" ");

    for(String num:line){

        int number=Integer.parseInt(num);

        if(number%2==0) {

            output.collect(new Text("even"),new
IntWritable(number));

        }

        else{

            output.collect(new Text("odd"),new
IntWritable(number));

        }

    }

}
```

```
}
```

reducer.java

```
package oddeven;

import java.io.*;

import java.util.*;

import org.apache.hadoop.mapred.*;

import org.apache.hadoop.io.*;

public class reducer extends MapReduceBase implements
Reducer<Text,IntWritable,Text,IntWritable>

{

public void reduce(Text key,Iterator<IntWritable>
value,OutputCollector<Text,IntWritable> output ,Reporter r)
throws IOException

{

    int sum=0,count=0;

    while(value.hasNext()){

        sum+=value.next().get();

        count++;

    }

    output.collect(new Text("Sum of "+key+" Numbers"),new
IntWritable(sum));

    output.collect(new Text(key+" Number count"),new
IntWritable(count));

}

}
```

- Create a Directory named `oddeven` inside the hadoop directory which was initially setup and move to it.(mkdir and cd)

- Create 3 java files(driver,mapper and reducer) as shown above and save them (vim <file-name>.java and :wq to save the file
- Run `javac -d . *.java` inorder to compile all the programs present inside the wordcount directory and the respective classes will be saved inside a folder named wordcount in the same directory.
- Run `echo Main-Class: oddeven.driver > Manifest.txt`
- Run `jar cfm oddeven.jar Manifest.txt oddeven/*.class`
- Create an input text file (vi oeinput.txt) specifying the text on which the wordcount operation will be performed.
- `hadoop jar oddeven.jar oeinput.txt output` Inorder to run the program on the input.txt file and store the results on output.
- `cat output/*` to display the output of the mapreduce program.

```

ritadmin@ritadmin-ThinkCentre-M70t:~$ cd hadoop-3.2.2
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ vim bash.sh
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ source bash.sh
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ mkdir oddeven
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2$ cd oddeven
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ vim driver.java
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ vim mapper.java
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ vim reducer.java
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ javac -d . *.java
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ echo Main-Class: oddeven.driver > Manifest.txt
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ jar cfm oddeven.jar Manifest.txt oddeven/*.class
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ ls
driver.java  Manifest.txt  mapper.java  oddeven      oddeven.jar  reducer.java
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ vi oeinput.txt
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ cat oeinput.txt
1 2 3 4 5 6 7 8 9
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ hadoop jar oddeven.jar oeinput.txt output

```

```

Bytes written=96
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ cat output/*
Sum of even Numbers      20
even Number count       4
Sum of odd Numbers      25
odd Number count        5
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ cat oeinput.txt
1 2 3 4 5 6 7 8 9
ritadmin@ritadmin-ThinkCentre-M70t:~/hadoop-3.2.2/oddeven$ █

```