EL-V BE IT
EXP 7

Aim: Design and develop a distributed application to find the coolest/hottest year from the available weather data. Use weather data from the Internet and process it using MapReduce.

Steps:
First install hadoop (if not installed yet) by,
https://sl6it.wordpress.com/2015/12/04/1-study-and-configure-hadoop-for-big-data/

Download **dataset.zip** file (attached with this post)
It contains NCDC weather data from year 1901 to year 1920.
Copy and extract dataset.zip in your **home** folder

Open terminal
whoami
It will display your user name, we will use it later.

Open eclipse->new java project->project name **exp7**->new class->**MaxTemperatureMapper**

Add following code in that class

```
package exp7;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class MaxTemperatureMapper extends Mapper<LongWritable, Text, Text, IntWritable>
  {
        private static final int MISSING= 9999;
        @Override
public void map(LongWritable key,Text value, Context context) throws IOException,InterruptedException
          {
                    String line = value.toString();
                    String year = line.substring(15, 19);
                    int airTemperature;
                    if (line.charAt(87)=='+')
                      {
                              airTemperature = Integer.parseInt(line.substring(88, 92));
                      }
                    else
                      {
                              airTemperature = Integer.parseInt(line.substring(87, 92));
                      }
                    String quality = line.substring(92, 93);
                    if (airTemperature != MISSING && quality.matches("[01459]"))
                      {
                         context.write(new Text(year), new IntWritable(airTemperature));
                      }
          }
  }
```

Save the file

It will display some errors, so we are going to import two jar files in our project.

Copy hadoop-mapreduce-client-core-2.7.1.jar from ~/hadoop/share/hadoop/mapreduce directory
In eclipse-> right click on exp7 project- >paste
Right click on pasted hadoop-mapreduce-client-core-2.7.1.jar-> Buid path-> add to buid path

#Copy hadoop-common-2.7.1.jar from ~/hadoop/share/hadoop/common directory
In eclipse-> right click on exp7 project- >paste
Right click on pasted hadoop-common-2.7.1.jar-> Buid path-> add to buid path

Right click on project exp7->new class-> **MaxTemperatureReducer**
Add following code in that class

```
package exp7;

import      java.io.IOException;
import      org.apache.hadoop.io.IntWritable;
import      org.apache.hadoop.io.Text;
import      org.apache.hadoop.mapreduce.Reducer;
public class MaxTemperatureReducer
 extends Reducer<Text,IntWritable,  Text, IntWritable>
  {
      @Override
      public void reduce(Text key, Iterable<IntWritable> values, Context
context) throws IOException,  InterruptedException
    {
      int maxValue = Integer.MIN_VALUE;
      for (IntWritable value  : values)
              {
             maxValue   = Math.max(maxValue, value.get());
         }
      context.write(key,      new   IntWritable(maxValue));
    }
}
```

Save the file

Right click on project exp7->new class-> **MaxTemperature**
Add following code in that class (**replace `your_user_name` by your own username**)
hdfs port number here is 1234, replace it with your port no (if different).

```
package exp7;
import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.util.ArrayList;
import java.util.List;
import java.util.Scanner;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileStatus;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
public class MaxTemperature
  {
    public static void main(String[] args) throws Exception
       {
          if (args.length != 2)
           {
      System.err.println("Usage:MaxTemperature  <input path> <output path>");
      System.exit(-1);
           }
      @SuppressWarnings("deprecation")
      Job job     = new Job();
      job.setJarByClass(MaxTemperature.class);
      job.setJobName("Max temperature");
      FileInputFormat.addInputPath(job,   new   Path(args[0]));
      FileOutputFormat.setOutputPath(job, new   Path(args[1]));
      job.setMapperClass(MaxTemperatureMapper.class);
      job.setReducerClass(MaxTemperatureReducer.class);
      job.setOutputKeyClass(Text.class);
      job.setOutputValueClass(IntWritable.class);
      job.waitForCompletion(true);

       Configuration conf = new Configuration();
     conf.set("fs.defaultFS", "hdfs://localhost:1234/user/your_user_name/");
     FileSystem fs = FileSystem.get(conf);
     FileStatus[] status = fs.listStatus(new Path(args[1]));

     //copy hdfs output file to local folder
       for(int i=0;i<status.length;i++){
         System.out.println(status[i].getPath());
fs.copyToLocalFile(false, status[i].getPath(), new Path("/home/your_user_name/"+args[1]));
     }
     System.out.println("\nYear\tTemperature\n");
       //display contents of local file
BufferedReader br = new BufferedReader(new
FileReader("/home/your_user_name/"+args[1]));
    String line = null;
    while ((line = br.readLine()) != null) {
      System.out.println(line);
    }
    br.close();
      Scanner s = new Scanner(new File("/home/your_user_name/"+args[1]));
      List<Integer> temps = new ArrayList<Integer>();
      List<String> years = new ArrayList<String>();

      while (s.hasNext())
      {     years.add(s.next());
            temps.add(Integer.parseInt(s.next()));
      }
   int max_temp=0,min_temp=999,i=0,j=0;
   String hottest_year="", coolest_year="";
   for (int temp: temps)
      {if(temp>max_temp)
            {     max_temp=temp;
                  hottest_year=years.get(i);
          }
            i++;
      }

      float max_temp1=max_temp;
System.out.println("Hottest Year:"+hottest_year);
```

```
System.out.println("\tTemperature:"+max_temp1/10+" Degree Celcius");

      for (int temp: temps)
      {
            if(temp<min_temp)
            {
                  min_temp=temp;
                  coolest_year=years.get(j);
          }
            j++;
      }
    float min_temp1=min_temp;
System.out.println("Coolest Year:"+coolest_year);
System.out.println("\tTemperature:"+min_temp1/10+" Degree Celcius");
      s.close();
   }
}
```

Save the file

In eclipse->Right click on project exp7-> export->java->jar file->next-> select the export destination -> /home/**your_user_name**/exp7.jar -> next -> next -> select main class ->browse -> MaxTemperature -> finish

exp7.jar file will be created in your home folder

Open terminal

Now Start NameNode daemon and DataNode daemon:

~/hadoop/sbin/start-dfs.sh

Make the HDFS directories required to execute MapReduce jobs (if not already done)

~/hadoop/bin/hdfs dfs -mkdir /user

~/hadoop/bin/hdfs dfs -mkdir /user/**your_user_name**

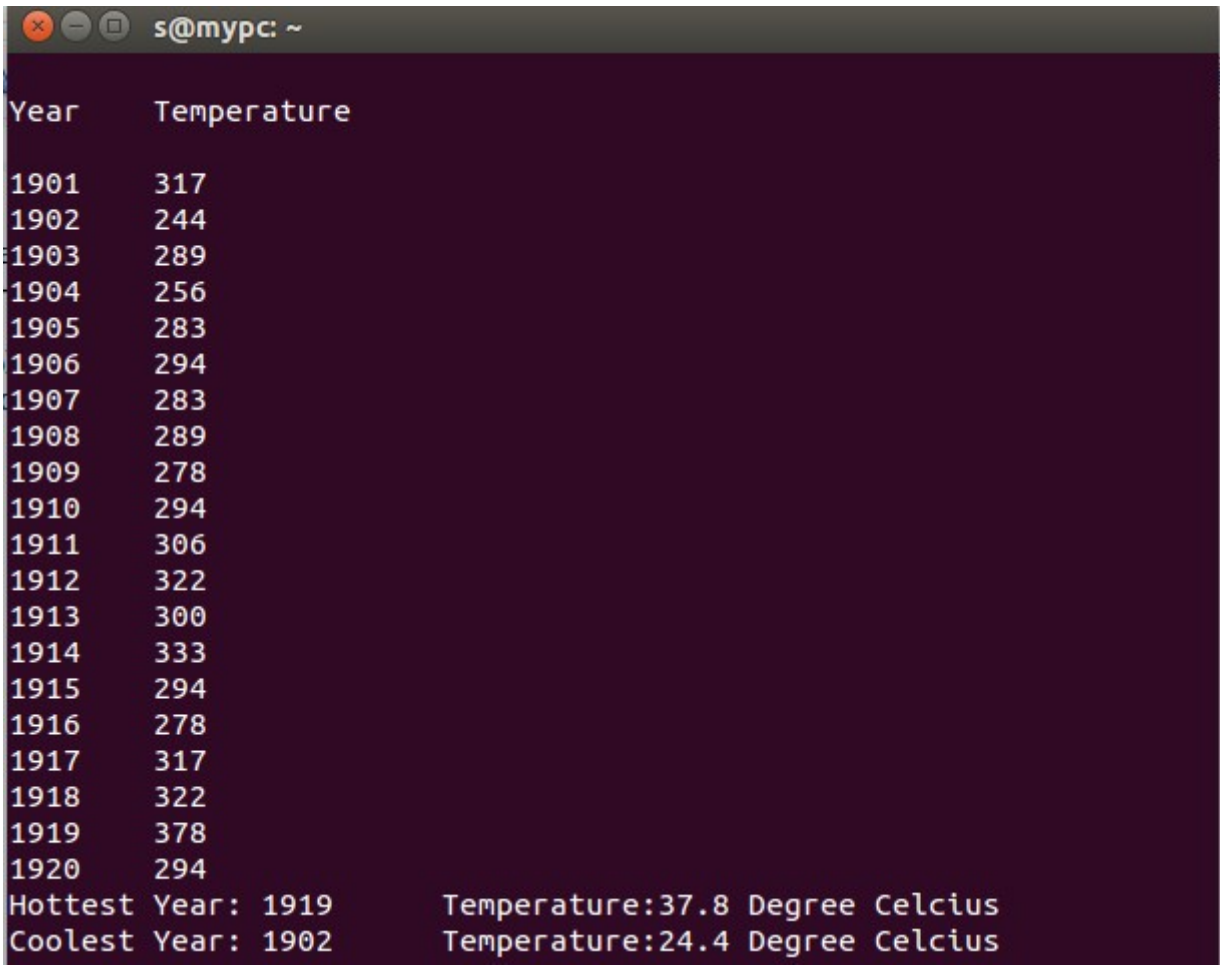
Put NCDC weather dataset in hdfs

~/hadoop/bin/hdfs dfs -put ~/dataset input_dataset

Perform MapReduce job

~/hadoop/bin/hadoop jar ~/exp7.jar input_dataset output_dataset

Output



Stop haddop

~/hadoop/sbin/stop-dfs.sh

jps

Reference : Hadoop the definitive guide, O'Reilly Publications, by Tom White

---

Prepared By,
Prof. S. T. Kolhe
(Department of I.T – S.R.E.S C.O.E Kopargaon)