

* **Testing the Project on Cloud with AWS:**

1. After testing the sample dataset on Cloudera VM, the project solution is set using AWS.
2. All the dataset files uploaded in .csv format to S3 after creation of a bucket by the name noaa-weather-dataset.
3. Data from years 1763-2019 was uploaded along with the source code jar file.
4. Snapshot of the project on AWS :

A screenshot of a computer screen

Description automatically generated

1. Dataset used in almost 17GB in size with more than 20M instances.

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

1. Clusters were created on EC2 and initiated after making the necessary changes.

A screenshot of a computer screen

Description automatically generated

* **Mapper-Reducer Code:**

For the mapper, the max temperature class is static, this method takes the input as text data type. Leaving the first five tokens, the 6th token is taken as the temp\_max and the 7th as temp\_min.

Now temp\_max value is set to be >35.0 and the temp\_min is set to be <10.0 and are now passed to the reducer step.

If the temp values for the day are >35.0 output as Hot Day and if <10.0 output as a Cold Day.

For the Reducer method, it takes the input as key and the pairs would be the list of values from the Mapper.

Now **Aggregation** is applied, and it produce the next result.

For the main method, it is used for setting up all the configuration properties. This will be acting as the driver for our Map Reduce code.

* **Below is the Complete Source Code used:**

A screenshot of text

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

* **Eclipse IDE:**

Now the project is created in the Eclipse IDE to analyze the sample dataset.

The jar file is then exported after having no issues with the project files.

A screenshot of a social media post

Description automatically generated

The next step was to send the sample dataset onto HDFS.

**Hdfs dfs -put Downloads/Noaa\_Weather\_data.txt /**

**A screenshot of a cell phone

Description automatically generated**

Run the Jar file for output.

**Hadoop jar temperature.jar /Noaa\_weather\_data.txt /output\_hotandcold**

**A screenshot of a social media post

Description automatically generated**

Check the Output directory in the HDFS.

A screenshot of a social media post

Description automatically generatedA screenshot of a social media post

Description automatically generated

**Results analysis:**

Depending on the 2015 sample dataset only two days above 35.0 recorded.

**A screenshot of a social media post

Description automatically generated**

