**Lab Task 4**

**Code explanation:**

I read weather.json file in a pyspark dataframe by using shell.

After that, I ran query for filtering the records where I assumed feels\_like temperature for the day would be under 15 degree Celsius.

For this query I explode the daily array in the dataframe. Explode function will explode the daily array and give the “col” name. I utilized withColumn function to rename that “col” column to “daily” column [1].

Then, I used filter function of the dataframe for filtering the data.

Lastly, I saved the filtered dataframe content to the cold\_weather.json file.

The weather data is in weather.json file, filtered data in cold\_weather.json and python code in LabTask4.py file.

Output Screenshots are attached below.

**Screenshots:**

Text

Description automatically generated

Text

Description automatically generated



**References:**

[1] “Spark : How do I exploded data and add column name also in pyspark or scala spark?,” Stack Overflow. [Online]. Available: https://stackoverflow.com/questions/48748727/sparkhow-do-i-exploded-data-and-add-column-name-also-in-pyspark-or-scala-spar. [Accessed: 22- Jun-2022].