Evan Lavertu
CS455
Dr. Pallickara
Assignment 3

Assignment 3

1. **Which state has the most monitoring sites across the United States? (Note: a site is identified by the combination of the state code, county code, and site number.)**

   California had the most monitoring sites across the United States on both datasets. For this question I created 2 mapper functions and 2 reducer functions. The first mapper mapped unique site id's consisting of their state code, county code, and site number. The first reducer just flipped the key and val and broke after the first occurence. This output was sent to a temp directory. The second mapper read from the temp directory and outputted 1 each time it found an instance of a unique site identifier. The second reducer counted all the occurrences for each state and outputted them to the final output.

2. **Does the East Coast or West Coast have higher mean levels of SO2? (Note: there are a total of 4 and 16 states in the West Coast and East Coast, respectfully.)**

   The East Coast had a higher mean of SO2 levels when the program was run on the gasses dataset. For this question I used 1 map function and 1 reduce function. The first map function used a enum of coastal states tied to their state code. It then outputs whether the state was East Coast or West Coast and the measurement of SO2 for that entry. The reduce function loops through all the measurements and outputs the mean for both East Coast and West Coast.

3. **What time of day (GMT) has the highest SO2 levels between 2000 - 2019? Capture the mean SO2 levels for each hour (GMT) over all 20 years to justify your answer.**

   When the program was run on the gases dataset, the time of day that had the highest mean level of SO2 was 16:00. For this question I used 1 map function and 1 reduce function. The map function finds SO2 measurements for dates within the bounds each hour. Then outputs the time and measurement. The reduce function loops through all the measurements and outputs the time and mean.

4. **Has there been a change in SO2 levels over the last 40 years? Capture the mean SO2 levels for each year to justify your answer.**

   Running the application on the gases dataset showed that over the past 40 years, the SO2 levels have been steadily declining. For this question I used 1 map function and 1 reduce function. The map function finds SO2 measurements within the past 40 years

and outputs the year and measurement. The reduce function counts up all the measurements and outputs the year and mean.

5. **What are the top 10 hottest states for the summer months (June, July, August)? Capture the mean temperature levels for the summer months (GMT) to justify your answer.**

   Running the application on the meteorological dataset, showed that Arizona, Puerto Rico, Virgin Islands, Texas, Nevada, Mississippi, Florida, Louisiana, Arkansas, and Oklahoma had the hottest summers. For this question I used 1 map function and 1 reduce function. The map function finds measurements during the months of June, July, and August for each state. It then outputs the state and measurement. The reducer loops through all the measurements and outputs the state and mean.

6. **What is the mean SO2 levels for the hottest states found in Question 5?**
   Using the gases dataset. The mean SO2 levels for the hottest states are Arizona with 67.72, Arkansas with 63.34, Florida with 73.21, Louisiana with 68.21, Mississippi with 68.50, Nevada with 62.70, Oklahoma with 60.81, Puerto Rico with 81.70, Texas with 68.62, and Virgin Islands with 79.82. For this question I used 1 map function and 1 reduce function. The map function looked for SO2 measurements for the 10 hottest states. It then outputs the state and measurement. The reducer counted all the measurements and output the state and mean.