In [85]:

```
from google.colab import files
uploaded = files.upload()
```

Choose File **No file selected**

**Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.**

```
Saving 10k_view_data.csv to 10k_view_data.csv
```

In [86]:

```
import pandas as pd
```

In [87]:

```
import io
df = pd.read_csv(io.BytesIO(uploaded['10k_view_data.csv']))
```

In [88]:

```
even_stations = df.loc[df["station_num"] % 2 == 0]
prime_even = even_stations.loc[(even_stations["event_time"] >= 200000) & (even_stations[
"event_time"] < 230000)]
```

In [89]:

```
with open('prime_even.txt', 'w') as file:
    file.write(prime_even.to_csv(index=False, header=False))
```

In [90]:

```
!pip install mrjob
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publi
c/simple/
Requirement already satisfied: mrjob in /usr/local/lib/python3.9/dist-packages (0.7.4)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.9/dist-packages (fr
om mrjob) (6.0)
```

In [91]:

```
%%file lab_02.py

from mrjob.job import MRJob
from mrjob.step import MRStep

class lab_02(MRJob):

  def steps(self):
        return [
            MRStep(mapper=self.mapper,
                   reducer=self.reducer),
            MRStep(reducer=self.reducer_find_max_program)
        ]

  def mapper(self, key, row):
    yield (row.split(","))[5], 1

  def reducer(self, program, counts):
    yield None, (sum(counts), program)

  def reducer_find_max_program(self, _, count_program_pairs):
    #count_program_pairs is a list of tuples
    yield max(count_program_pairs)
```

```
if __name__ == '__main__':
  lab_02.run()
```

```
Writing lab_02.py
```

In [92]:

```
!| python lab_02.py < prime_even.txt > output.txt
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/lab_02.root.20230416.222532.881979
Running step 1 of 2...
reading from STDIN
Running step 2 of 2...
job output is in /tmp/lab_02.root.20230416.222532.881979/output
Streaming final output from /tmp/lab_02.root.20230416.222532.881979/output...
Removing temp directory /tmp/lab_02.root.20230416.222532.881979...
```

In [93]:

```
with open('output.txt') as file:
  line = (file.readlines())[0]

max_prog = line.split()[1]
prog_name = (max_prog.split('"'))[1]
prog_df = df.loc[df["prog_code"] == prog_name]
print(prog_df.shape[0], max_prog)
```

```
257 "SP003189730000"
```