Data Engineering Project Assignment -1

Team Name - Data Foundry

Team Members:
Kiranmayi Thanikonda
Jayita Banerjee
Navya Sri Ambati

Step-1:

Construct a table showing each day for which your pipeline successfully, automatically processed one complete days' worth of sensor readings. It might be a good idea to have logging of some kind for the project so you can gather not only this data, but errors that might occur as you develop your pipeline throughout the term.

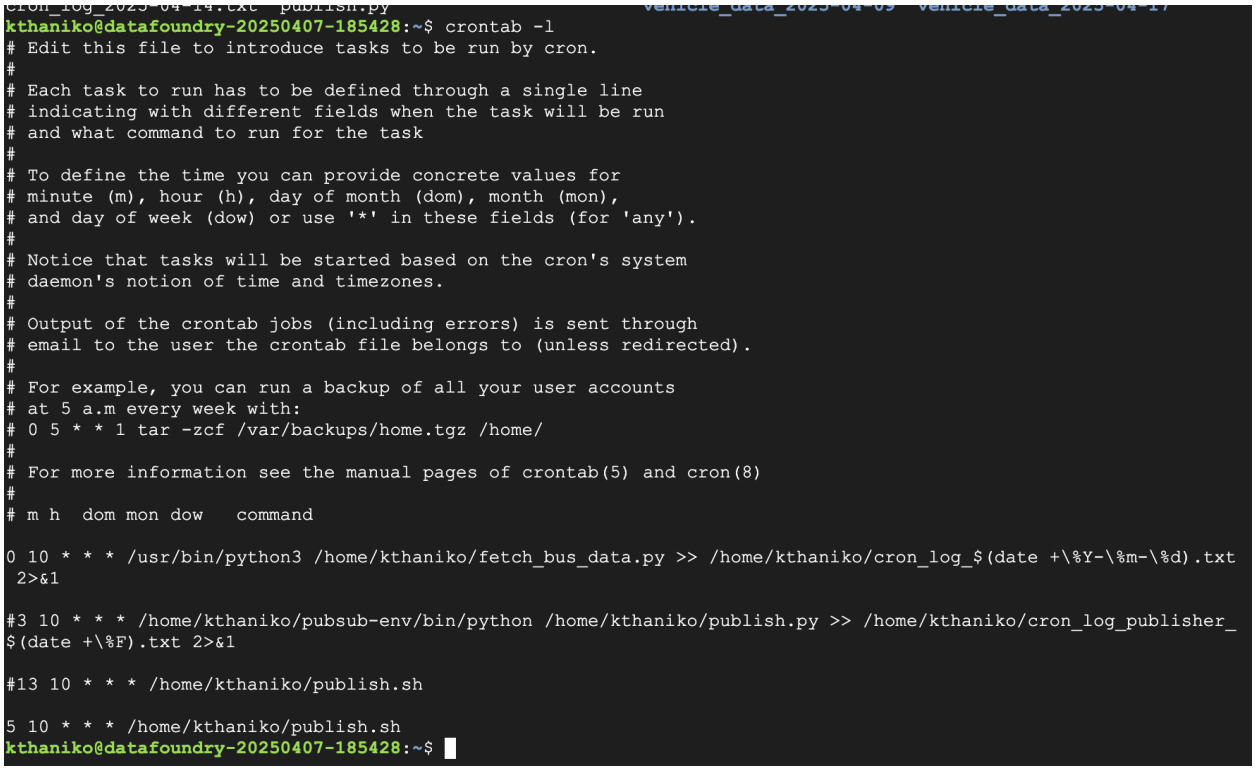
Date	Day of Week	Approximate Time of day for your data access	# Sensor Readings	Total Data Saved (KBs)	# Pub/Sub messages published and received
04/07/2025	Monday	12:00 AM	663247	181788.25 KB	663247
04/08/2025	Tuesday	10:00 AM	709902	194718.97 KB	709902
04/09/2025	Wednesday	10:00 AM	729793	200164.08 KB	729793
04/10/2025	Thursday	10:00 AM	688427	188748.67 KB	688427
04/11/2025	Friday	10:00 AM	672597	184510.95 KB	672597
04/12/2025	Saturday	10:00 AM	No Data	No Data	No Data
04/13/2025	Sunday	10:00 AM	No Data	No Data	No Data
04/14/2025	Monday	10:00 AM	744055	173536.76 KB	744055
04/15/2025	Tuesday	10:00 AM	677762	162173.72 KB	677762
04/16/2025	Wednesday	10:00 AM	603623	144467.21 KB	603623
04/17/2025	Thursday	10:00 AM	664401	159022.37 KB	664401

04/18/2025	Friday	10:00 AM	685748	164139.78 KB	685748
04/19/2025	Saturday	10:00 AM			
04/20/2025	Sunday	10:00 AM			

Step-2:

Additionally, include screenshots for the parts C, H and I

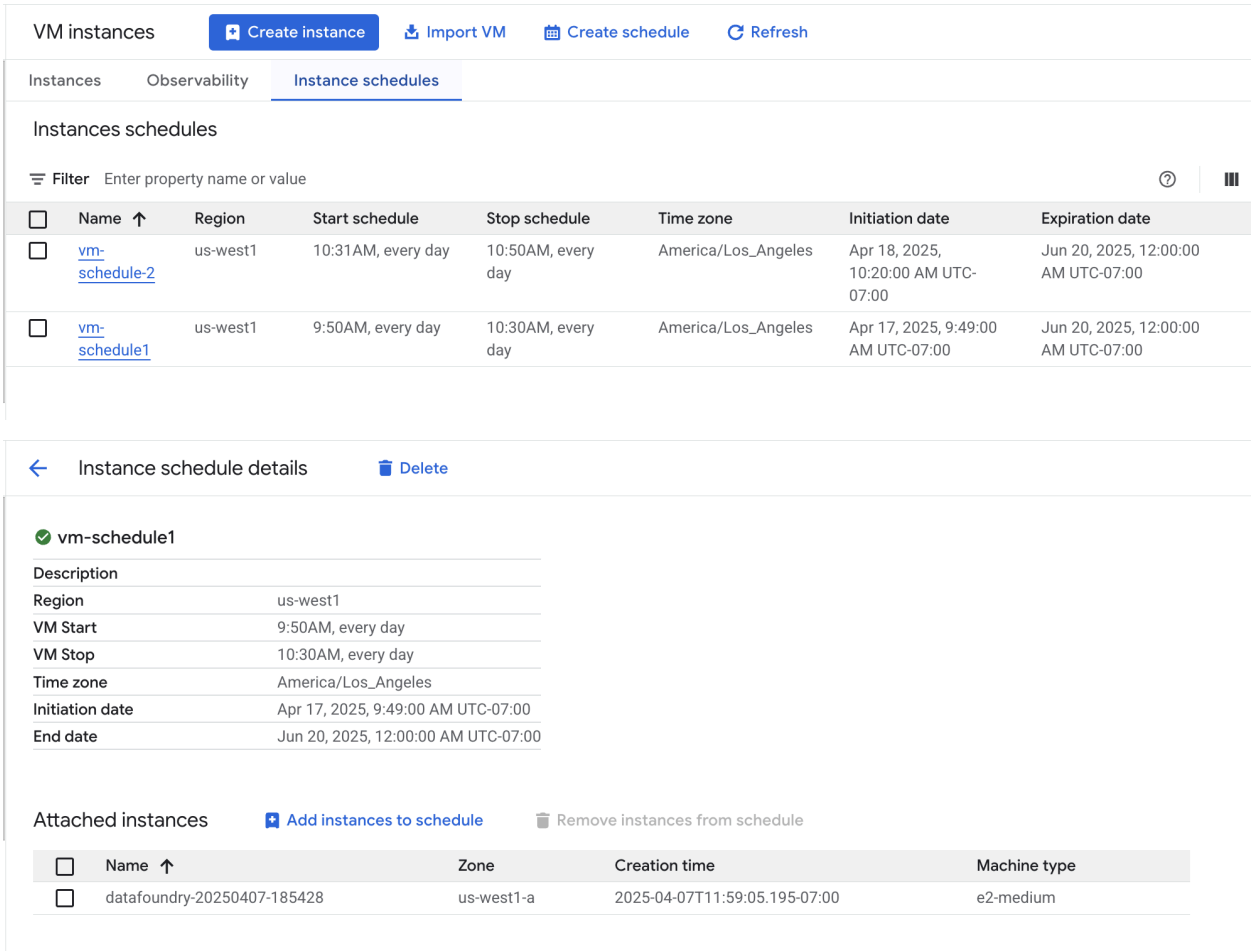
1. Output of crontab -l: Your scheduled cron jobs.

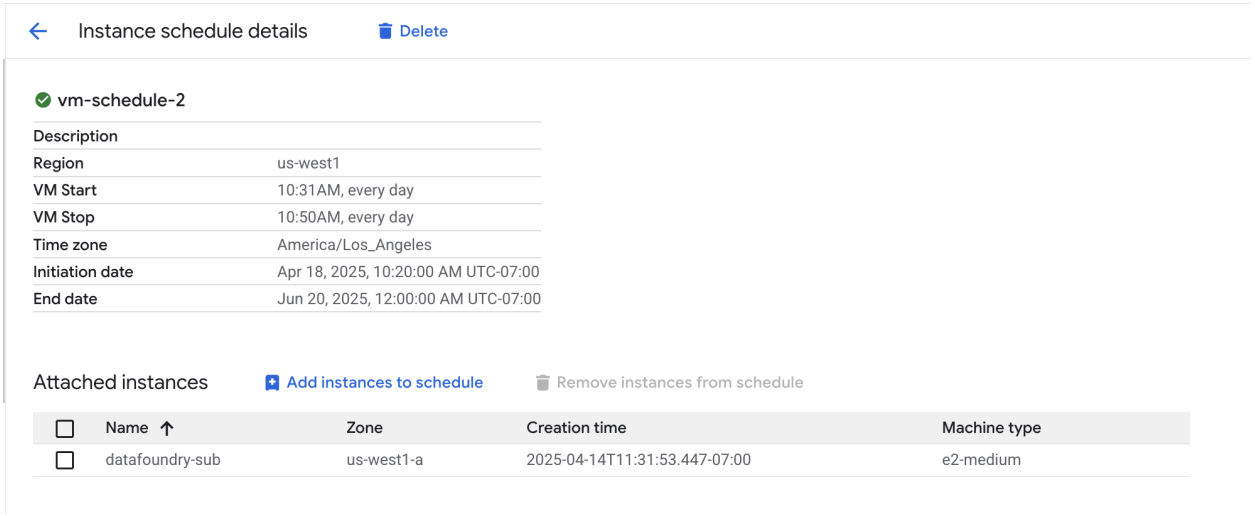


2. systemctl status: This will show the status of your receiver program.

```
● datafoundry-sub
    State: running
    Units: 258 loaded (incl. loaded aliases)
     Jobs: 0 queued
   Failed: 0 units
    Since: Thu 2025-04-17 12:22:29 PDT; 42s ago
  systemd: 252.36-1~deb12u1
   CGroup: /
           ├─init.scope
           │ └─1 /sbin/init
           ├─system.slice
           │ ├─cron.service
           │ │ └─891 /usr/sbin/cron -f
           │ ├─dbus.service
           │ │ └─380 /usr/bin/dbus-daemon --system --address=systemd: --nofork --nopidfile --systemd-activation>
           │ ├─exim4.service
           │ │ └─808 /usr/sbin/exim4 -bd -q30m
```

3. VM instance schedule: This will display the schedule settings for your GCP VM instance.





GitHub Link:

https://github.com/kthanikonda/DataEngineering/tree/main/Project/Data%20Gathering%20and%20Transport

Permissions: