

Data Engineering Milestone - I

Date	Day of Week	Approximate Time of day for your data access	# Sensor Readings	Total Data Saved (KBs)	# Pub/Sub messages published and received
04/14/2024	Sunday	11:00 am		81076	
04/15/2024	Monday	11:00 am		84748	
04/16/2024	Tuesday	11:00 am		56636	
04/17/2024	Wednesday	11:00 am		66304	
04/18/2024	Thursday	11:00 am		92620	
04/19/2024	Friday	11:00 am		87032	
04/20/2024	Saturday	11:00 am	25.5k	92732	25.5k
04/21/2024	Sunday	11:00 am	29k	99964	29k

A. Create, Configure and Run Your Virtual Machine - Done

B. Initial Python Data Gatherer - Done

Data Gathering script pushed by Jathin

C. Run the Data Gatherer Daily - Done

Crontab setup by Siddhanth

```
siddhan@dataeng:~$ crontab -l
# Edit this file to introduce tasks to be run by cron.
#
# Each task to run has to be defined through a single line
# indicating with different fields when the task will be run
# and what command to run for the task
#
# To define the time you can provide concrete values for
# minute (m), hour (h), day of month (dom), month (mon),
# and day of week (dow) or use '*' in these fields (for 'any').
#
# Notice that tasks will be started based on the cron's system
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow   command
30 20 * * * ./home/siddhan/dataeng-project/runscript.sh
siddhan@dataeng:~$
```

D. Configure Google Cloud Pub/Sub - Done

Pub/Sub Config and pipeline setup by Vipul

E. Parse JSON into Individual Breadcrumb Records - Done

Pub/Sub Config and pipeline setup by Vipul

F. Send Breadcrumb Records to Pub/Sub Topic - Done

Pub/Sub Config and pipeline setup by Vipul

G. Pub/Sub Receiver - Done

Pub/Sub receiver script setup by Vipul

H. Configure Linux to Run Pub/Sub Receiver Continually - Done

Systemctl and service file configurations by Raghuram

```
CLOUD SHELL
Terminal (dataeng-gowda-vipulp) x (dataeng-gowda-vipulp) x + v
Open Editor

# For more information see the manual pages of crontab(5) and cron(8)
# m h dom mon dow command
30 18 * * * /home/vipulp/dataeng-project/runscript.sh
(env) vipulp@vm-dataeng:~/dataeng-project/trimet-sub-a$
(env) vipulp@vm-dataeng:~/dataeng-project/trimet-sub-a$
(env) vipulp@vm-dataeng:~/dataeng-project/trimet-sub-a$
(env) vipulp@vm-dataeng:~/dataeng-project/trimet-sub-a$ sudo systemctl status trimet-receiver.service
* trimet-receiver.service - Trimet service
   Loaded: loaded (/etc/systemd/system/trimet-receiver.service; enabled; vendor preset: enabled)
   Active: active (running) since Sun 2024-04-21 18:53:57 PDT; 23min ago
     Main PID: 107841 (python3)
        Tasks: 26 (limit: 4680)
       Memory: 43.0M
      CGroup: /system.slice/trimet-receiver.service
             └─107841 /home/vipulp/dataeng-project/env/bin/python3 /home/vipulp/dataeng-project/pub-sub/receiver.py

Apr 21 18:55:39 vm-dataeng python3[107841]: Received Message {
Apr 21 18:55:39 vm-dataeng python3[107841]:   data: b'{"EVENT_NO_TRIP": 224239311, "EVENT_NO_STOP": 2242...'}
Apr 21 18:55:39 vm-dataeng python3[107841]:   ordering_key: ''
Apr 21 18:55:39 vm-dataeng python3[107841]:   attributes: {}
Apr 21 18:55:39 vm-dataeng python3[107841]: }
Apr 21 18:55:39 vm-dataeng python3[107841]: Received Message {
Apr 21 18:55:39 vm-dataeng python3[107841]:   data: b'{"EVENT_NO_TRIP": 224239311, "EVENT_NO_STOP": 2242...'}
Apr 21 18:55:39 vm-dataeng python3[107841]:   ordering_key: ''
Apr 21 18:55:39 vm-dataeng python3[107841]:   attributes: {}
Apr 21 18:55:39 vm-dataeng python3[107841]: }
Apr 21 18:55:39 vm-dataeng python3[107841]: }
(env) vipulp@vm-dataeng:~/dataeng-project/trimet-sub-a$
```

I. Schedule your VM to start and stop automatically - Done

VM instance scheduling setup done by siddhanth and Jathin

Google Cloud dataeng-Gowda-vipulp Search (/) for resources, docs, products, and more

Instance schedule details DELETE

trimet-services

DescriptionTrimet Instance Schedule

Regionus-west1

VM Start8:00PM, every day

VM Stop9:30PM, every day

Time zoneAmerica/Los_Angeles

Initiation date

End date

Attached instances ADD INSTANCES TO SCHEDULE REMOVE INSTANCES FROM SCHEDULE

<input type="checkbox"/>	Name ↑	Zone	Creation time	Machine type
<input type="checkbox"/>	vm-dataeng	us-west1-b	2024-04-13T23:06:34.705-07:00	e2-medium