Transit Data Pipeline: Real-Time GPS Tracking for TriMet

GoTriMet

Saeah Go

Introductions

Portland State University, Maseeh College of Engineering and Computer Science
Data Engineering – Spring 2025

Professor: Bruce Irvin
TA: Alex Harris
Kira Klingenberg

System Description

- What's TriMet?
  - TriMet is the regional transit organization for the metropolitan area of Portland, Oregon.TriMet generates almost 2 million GPS readings per day. This data is extremely valuable for understanding traffic patterns, delays, and route efficiency.
- What are breadcrumbs?
  - These location points are called 'breadcrumb data' because they act like a digital trail, showing where the bus has been throughout the day
- What did I build?
  - Built data pipelines to ingest, validate, transform, store, and visualize TriMet transit data

Architecture Overview

Source:
TriMet APIs
- Breadcrumb data (JSON)
- Stop event data (HTML)

GCP VM:
dataeng-vm

publisher.py
publisher_stop.py

Google Cloud Pub/Sub
- bus-data-sub (for breadcrumb data)
- stop-event-data-sub (for stop event data)

GCP VM:
dataeng-vm

subscriber.py
subscriber_stop.py

Validations & Transformations

Breadcrumb data:
- Remove rows with null latitude and longitude
- Filter out speeds > 80 mph

Stop event data:
- Extract trip_id from table name (h2) when loading data
- Only extract first row of the table

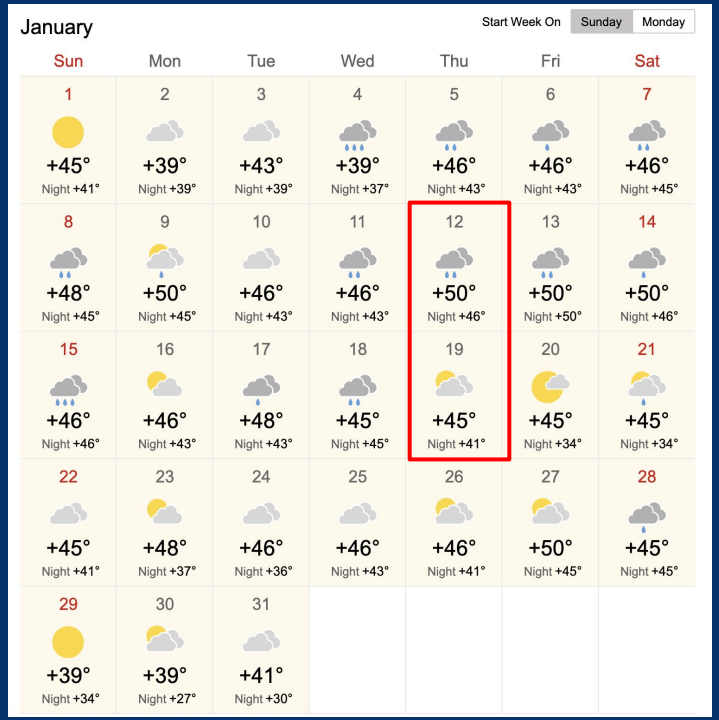
load_to_postgres.py

PostgreSQL Database
- breadcrumb table
- trip table

Data Description

- 🚌 21M+ Breadcrumb Records
- 📅 ~577K per Weekday / ~322K per Weekend
- 💾 2.3 GB Total Ingested
- 🔄 Daily Ingest: ~66 MB
- 🧩 2 Tables: breadcrumb, trip

Visualization #1: Effect of Rain on Bus Speeds



January

Start Week On: Sunday | Monday

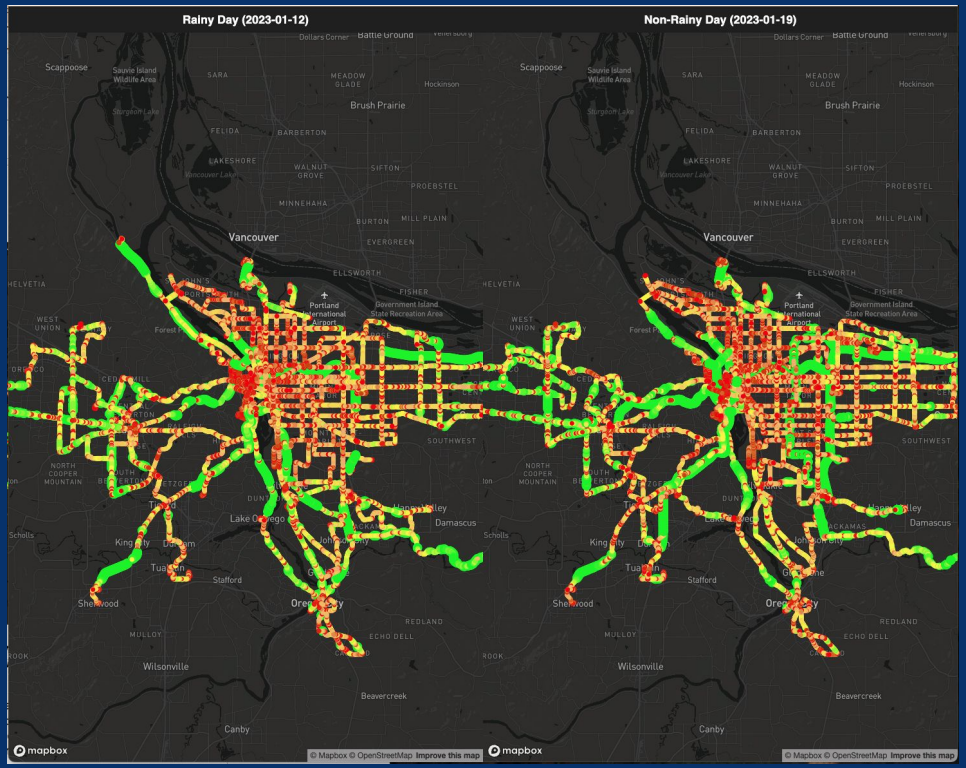
Sun	Mon	Tue	Wed	Thu	Fri	Sat
1 +45° Night +41°	2 +39° Night +39°	3 +43° Night +39°	4 +39° Night +37°	5 +46° Night +43°	6 +46° Night +43°	7 +46° Night +45°
8 +48° Night +45°	9 +50° Night +45°	10 +46° Night +43°	11 +46° Night +43°	12 +50° Night +46°	13 +50° Night +50°	14 +50° Night +46°
15 +46° Night +46°	16 +46° Night +43°	17 +48° Night +43°	18 +45° Night +45°	19 +45° Night +41°	20 +45° Night +34°	21 +45° Night +34°
22 +45° Night +41°	23 +48° Night +37°	24 +46° Night +36°	25 +46° Night +43°	26 +46° Night +41°	27 +50° Night +45°	28 +45° Night +45°
29 +39° Night +34°	30 +39° Night +27°	31 +41° Night +30°				

https://world-weather.info/forecast/usa/portland_2/january-2023/

Jan 12, 2023 Weather Info

Jan 19, 2023 Weather Info

Visualization #1: Effect of Rain on Bus Speeds



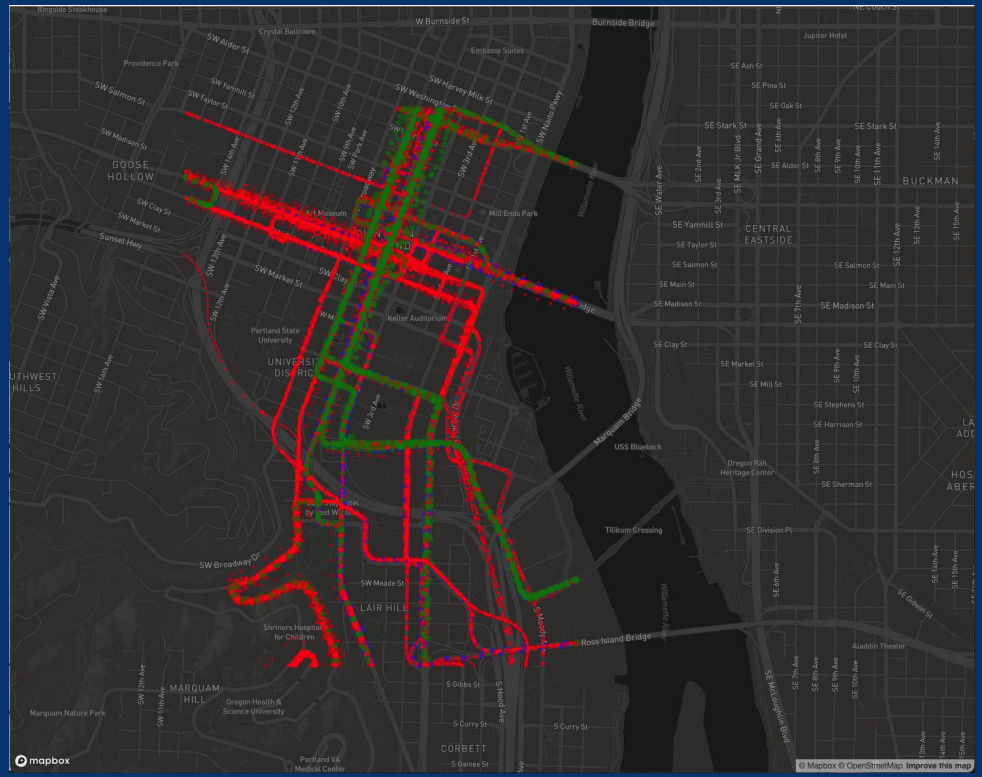
Side-by-side visualization that compares bus movement speeds across Portland during midday hours (10:00 AM – 1:00 PM) on two contrasting weather days: a rainy day (January 12, 2023) and a non-rainy day (January 19, 2023).

- Purpose: Compare traffic flow on a rainy vs. dry day
- Query for the Visualization:

```
SELECT latitude, longitude, speed,
  CASE
    WHEN DATE(tstamp) = '2023-01-12' THEN 'rainy'
    WHEN DATE(tstamp) = '2023-01-19' THEN
'non-rainy'
  END AS weather
FROM breadcrumb
WHERE DATE(tstamp) IN ('2023-01-12', '2023-01-19')
  AND tstamp::time BETWEEN '10:00' AND '13:00';
```

- Insights: Despite rain, overall trend is similar (no drastic change). But still dry day shows faster speeds than rainy day for some routes.

Visualization #2: Weekday vs. Saturday vs. Sunday



Bus activity during the morning rush hour (7:00 AM – 9:00 AM) within a focused area of Portland (latitude 45.50–45.52, longitude -122.69 to -122.67) by Service Key (Weekday, Saturday, and Sunday) from January 15, 2023 to January 21, 2023.

- Purpose: Compare traffic flow on weekday vs. Saturday vs. Sunday
- Query for the Visualization:

```
SELECT
    b.tstamp, b.latitude, b.longitude,
    b.speed, b.trip_id, t.service_key
FROM breadcrumb b
JOIN trip t ON b.trip_id = t.trip_id
WHERE b.tstamp::date BETWEEN '2023-01-15' AND
'2023-01-21'
   AND b.tstamp::time BETWEEN '07:00' AND '09:00'
    AND b.latitude BETWEEN 45.50 AND 45.52
    AND b.longitude BETWEEN -122.69 AND -122.67
    AND t.service_key IN ('W', 'S', 'U')
ORDER BY
    b.tstamp;
```

- Insights: Weekday (red) has more bus traffics compared to Saturday (green) and Sunday (blue)

Challenges & Learning

Challenges

- VM disk filled up → rebuilt and resized; hit credit limit
- First time using virtual machines — unfamiliarity with setup and management
- Working with systemd for automation was tricky
- Required JSON to DataFrame conversion before processing

Newly Learned

- Managed cloud VMs, storage, and automation tools
- Built robust data pipeline using OOP and environment variables
- Parsed messy HTML and validated streaming JSON data
- Gained practical experience with Pub/Sub and PostgreSQL

Thank you!

Big Thanks to

Professor Bruce

TA Alex

and Kira

Special thanks to