DataEng S24: Project Assignment 4

Final Presentation Video

Due date: Jun 7, 2024 at 10pm

Congratulations on getting your two data pipelines flowing, integrated and capable of visualizing the TriMet GPS sensor data, and future proofing them with data maintenance strategies. The only job remaining is for you to summarize your project with a final presentation video.

Keep your pipelines flowing at least until the day when you submit your video. Any demonstration done in the video should utilize the latest data stored in your database.

Chase Verbout

Trae

Will

John

Link to submission video on Google Drive :https://drive.google.com/file/d/1Re4miYdU0decp9MFQbrTmuC6GTm7sy21/view

Instructions for Video

Create a 10 minute (maximum) video describing:

1. The architecture of your system
2. Any relevant details of each stage of your pipeline
3. Describe the daily amount of data flowing through your pipeline
4. Describe the total amount of data stored in your database
5. Demonstrate 2 (or more) visualizations of your choice. For each visualization:
    a. Describe what it is and why you chose it
    b. Show the query/queries used to obtain the data for it
    c. Show the visualization
    d. Interact with the visualization using pan and zoom appropriately.

Absolutely must not be longer than 10 minutes. No minimum time requirement.

Introductions

One of the first few slides in the video must show the team's name, the name of all team members, the name of your university, college, department and course, and the names of your professor and TA.

Each teammate must present for at least part of the video. Clearly say and list (on the slide) the name of each presenting team member as they present. The best way to do this is to have each team member introduce themselves as they begin talking and to list that team member's name on each slide that they present. It is fine for some team members to do more of the presentation than others as long as each team member does at least part of the presentation.

System Description

As part of your introduction, describe what your system is and what it does. Don't assume that viewers already know what the data is or what the system does. Briefly explain what TriMet is, don't assume that your viewers know what TriMet is or what "breadcrumbs" are. Your video might be viewed by people who do not know the details of what was assigned to you.

In the slides we gave a brief introduction to Trimet and what breadcrumbs are.

Extra Credits

List and describe any extra credit items that you accomplished in your project.

These are some of the extra credit feats we were able to accomplish as part of our project.

1. Using Folium
2. We set up our backups early with the bucket
3. Krabby Patty illustrations/example with Pub/Sub (Amazing analogy)

Challenges you Overcame

Describe any difficulties or challenges that you encountered during the project and describe how you overcame these challenges. Also, include any general comments about the project or lessons learned, especially lessons that might aid a future Data Engineer working on a similar project.

We did not talk about challenges we overcame in the video. One of the first challenges we had was our data was taking astronomically large amounts of time to be processed. Once we did the COPY_FROM in class assignment and learned that was a great way to process the data, we implemented it and that fixed our issue. Another difficulty we could not quite resolve were abnormalities in the data. Some of the trip speeds were unrealistically high. We attributed this to stops being skipped or entries being skipped messing with the calculation of the speed.

Architecture

After your introduction, describe the architecture of your system. Specifically, display a component diagram that includes each of the items below. No need to talk through each component in detail, but feel free(We will) to describe any part of the site that you want to. The goal here is to convince viewers that you developed a sophisticated system with many coordinated parts.

- The source of the data
- Any GCP VMs used
- The publisher(s)
- Google cloud Pub/Sub system
- The receiver(s)
- The database server
- Any other relevant details, like interesting validations or transformations needed for the breadcrumb or stop data.

This is where our amazing Pub/Sub example is

Data Description

Show a slide that summarizes the amounts of data handled by your system. Include measurements/estimates of average daily data (both weekday and weekend) as well as total amounts of data. Data can be measured in terms of data records or bytes or even better both record count and byte volume. The idea here is to convince viewers that you built something much more realistic than a small toy example.

Visualizations

Next, show at least two separate visualizations in detail. Try to choose visualizations/cases that are not only visually appealing but also provide unexpected or unusual insights about the TriMet transit system. Describe the purpose/goals of each visualization. Show the SQL query/queries used to extract the data from your database, and then display the visualization. Interact with and explore the visualization to show any interesting or essential details. *(Trae Was here)*

Conclusion

Finish the presentation with a “Thank You” slide. Feel free to thank anybody who helped you, but at the very least be sure to thank TriMet for sharing their valuable real-world data for your project. Use the TriMet logo on this slide.

10 minutes does not give you much time, so plan and edit your presentation accordingly.

Video Judging

Your video will be judged by our celebrity guest judges at the incredibly prestigious **Third Annual Data Engineering Awards Ceremony** on Wednesday, June 12, 2024 at 12:30pm in Lincoln Hall #247.This is our regularly scheduled exam time, we will hold the award ceremony in place of an exam.

The judges will be looking at content more than slickness of presentation. Fun prizes will be awarded based on categories like best team name, best tooling, and best overall presentation. These come with the honor of bragging to all your friends and family about how you had the best pipeline ever! (and then explaining to them what a pipeline is)

This video does not need to be super professional or polished. The judges will be looking at content more than slickness of presentation. Use your choice of video capture technology. For example, you can use QuickTime screen recording on a Mac. On any system, you can hold a Zoom meeting with your teammates, share your screen, and record it.

Submission

Upload the final video to either Youtube or GDrive with public permissions enabled so that the judges can view it. Be sure to include a URL reference to it in your submission form (where the form requests “... URL of your assignment submission document”). Also, be sure to list your accomplished extra credit items in the form as well so that it doesn’t get missed (where the form requests “... Any suggestions or advice” because we don’t have a designated area for this yet).

Link to submission video on Google Drive :https://drive.google.com/file/d/1Re4miYdU0decp9MFQbrTmuC6GTm7sy21/view