

Graph Representation for Subway Networks and Passenger Flow Analytics

RIFFENDY, Kennard
Supervised by Professor Brian Mak



Introduction

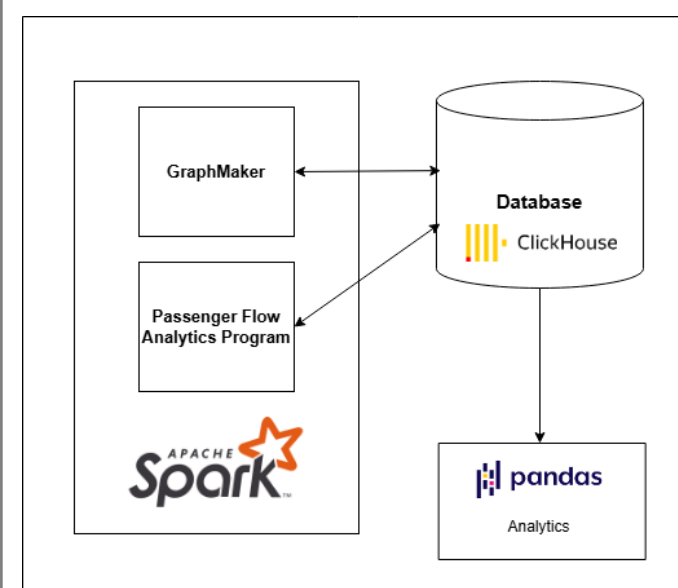
The logistical management of metro networks plays a pivotal yet unnoticed role in enhancing the passenger travel experience. Decisions as to when to increase train frequencies cannot always be made at a rational basis. This project seeks to expand upon the existing framework that leverages Graph Technology to analyze metro networks and generate passenger flow analytics metrics that vary greatly at different times of day, with the ultimate goal of making it deployable and beneficial for other metro networks.

Objectives

- Scale up the GraphMaker, an algorithm for generating metro network topologies as graphs to operate on simplified datasets
- Scale up the Passenger Flow Analytics, a system for processing network data and passenger information efficiently across simplified data sets throughout the business day.
- Develop a visual analytics program to visualize the metrics generated by the passenger flow analytics program.
- Architect an end-to-end pipeline: Automates the process of building all project prerequisites and initialize environments to streamline operations.

Methodology

Standalone System

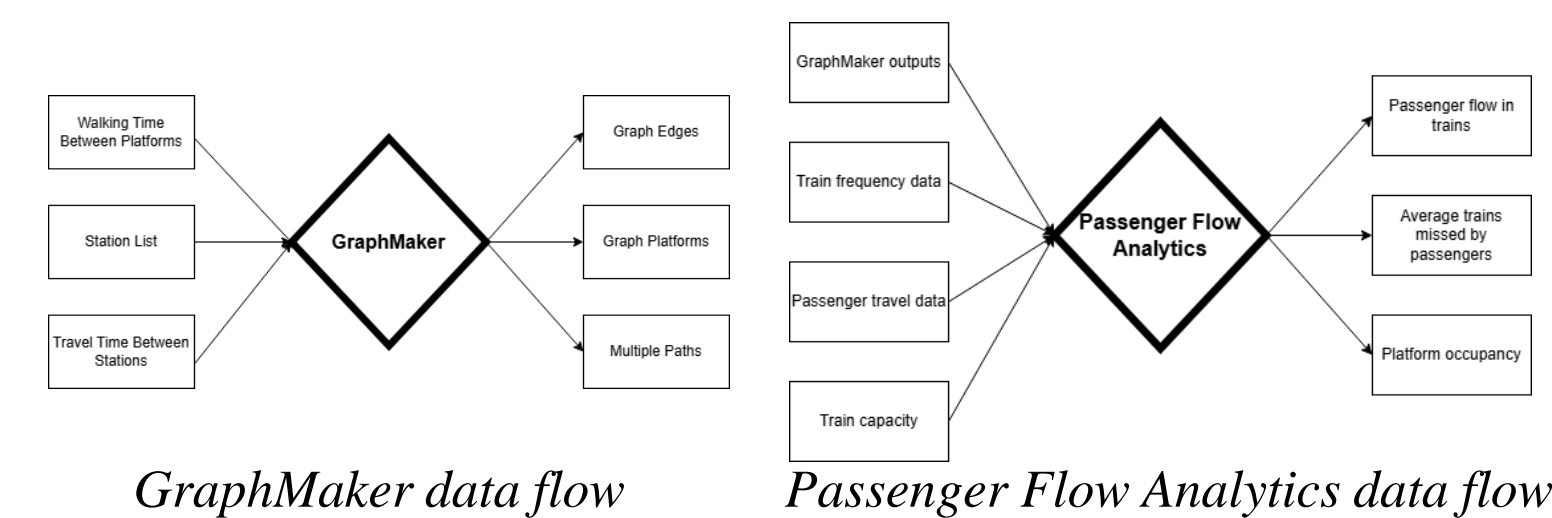


GraphMaker and Passenger Flow Analytics runs inside Apache Spark Cluster – distributed computing framework.

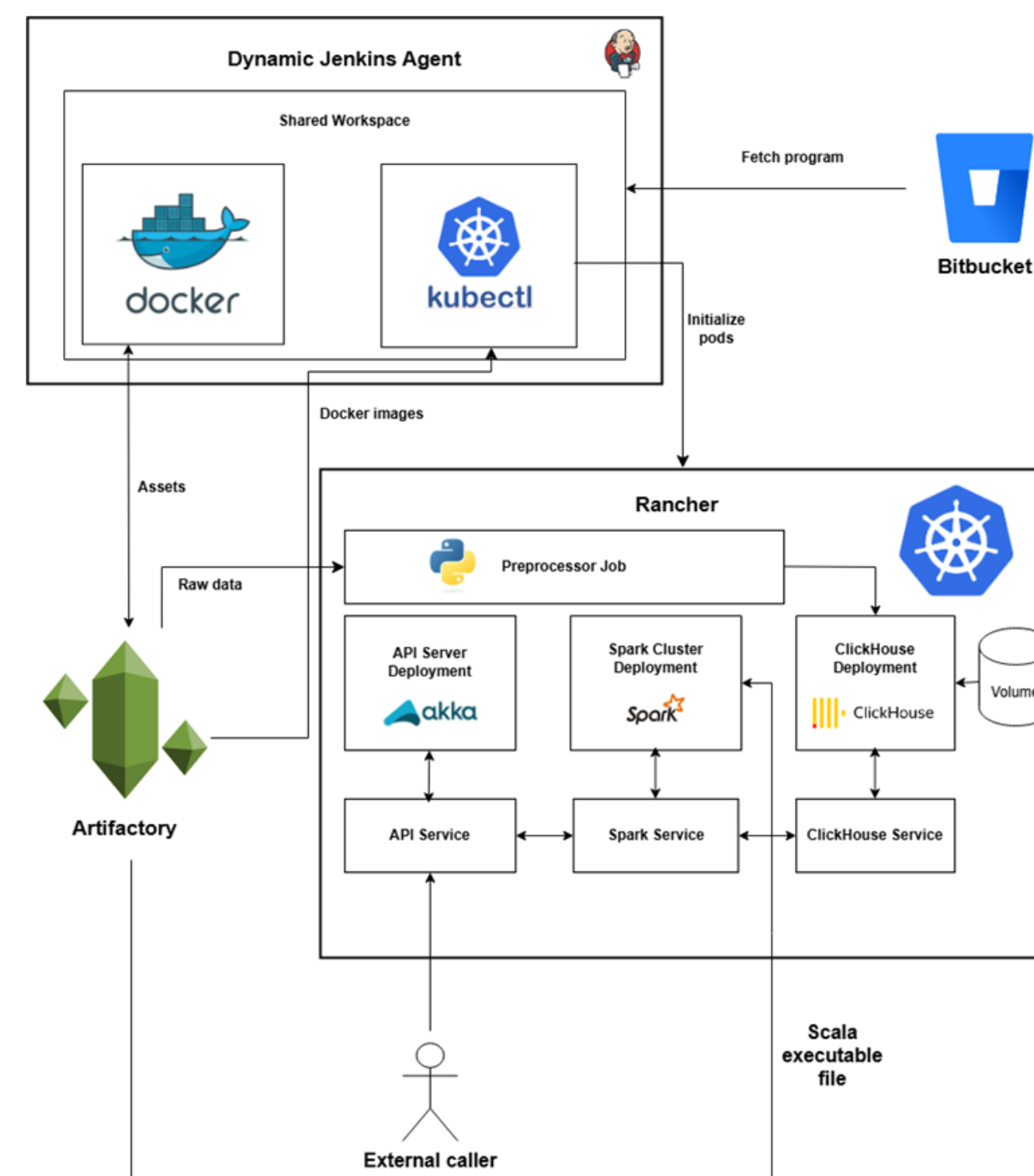
ClickHouse SQL is used as our database, which is a column-oriented database suitable for big data processing

Analytics program that takes metrics generated by Passenger Flow Analytics to visual graphical representations.

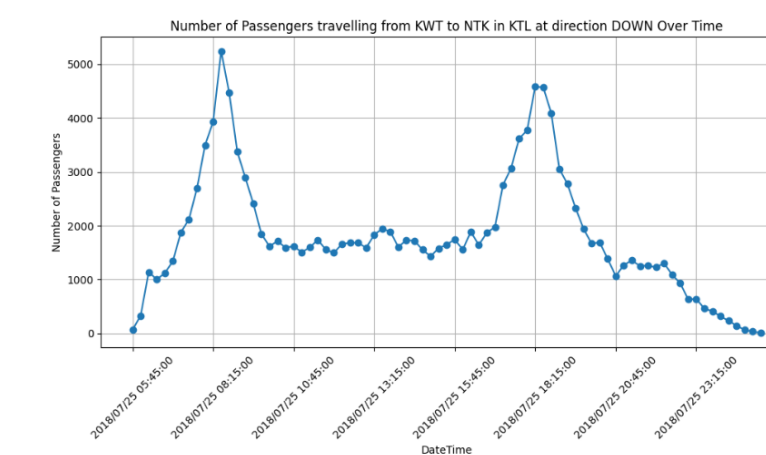
Data Flow



Automation Pipeline for hosting application in Deployment

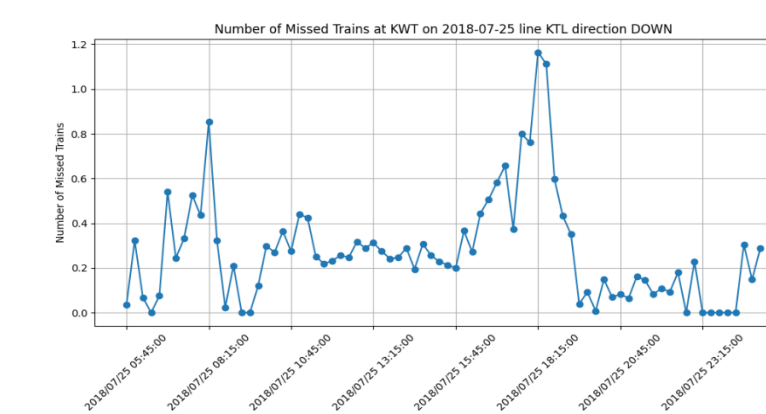


Generated metrics



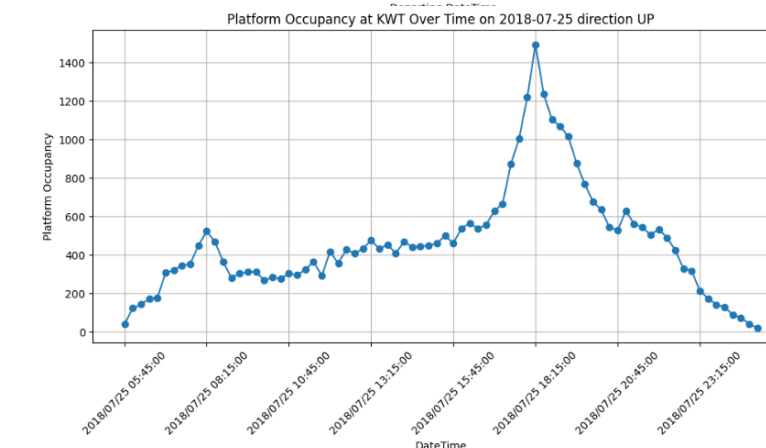
Passenger Flow

How many passengers are travelling from one station to its next station for every 15-minute interval?



Average Missed Trains

On average, how many trains does a passenger have to wait between the time of arrival at platform and boarding when travelling specific route every 15 minutes?



Platform Occupancy

How many passengers are waiting at a particular platform for every 15-minute interval?

Conclusion

Our framework for representing a metro network's topology as a graph data structure and conducting in-depth passenger flow analysis is scaled up to cater to more metro systems than the existing framework allows, as part as our commercial goals to sell to more customers.

On top of that, in order to elevate our prototype into a commercial standard, we developed an automation pipeline that abstracts the complex process of building prerequisites and initialize the required runtime environments for GraphMaker and Passenger Flow Analytics, allowing customer to simply run pipeline and invoke a simple API to execute the programs.