THE DESIGN OF AN AUTOMATIC STREAMING ANALYTICAL WORKFLOW FOR PROCESSING MASSIVE TRANSIT FEEDS

Hung Cao and Monica Wachowicz
Outline

• Introduction
• Related work
• Streaming analytical workflow
• Stream processing architecture
• Conclusion and Future research
Introduction

What is a streaming analytical workflow?
  • Pre-build connector
  • Low-latency database
  • Streaming processing environment.

Why do we need them for?
  • Support the processing of continuous computation of data streams such as transit feeds and IoT streams.
Related work

• Most of the research work found in analyzing transit feeds (e.g. AVL and GPS) is based on manually batch processing using a cloud platform.

• Huang et al. (2014) proposes the use of descriptive analytics for the Beijing Transportation Department that is already reaching a data ingestion of 15,000 GPS records per second for 30,000 buses.
Our Streaming Analytical Workflow

[Diagram showing the workflow]
Data contextualization
Step 1: Stop/Move Detection
Step 2: Stop/Move Classification
Step 3: Street Name Annotation
Step 4: Bus Station Identification
Step 5: Street Intersection Identification

GPS location of the bus during Trip 51
Intersection
Step 6: Arrival/Departure Times Identification

25/05/2018
Step 7: Origin/Destination Trip Identification
Stream processing architecture
Conclusions

• Analytics performed over contextualized transit feeds could potentially revolutionize transit network services.

• The outcomes from the data cleaning task indicate that it is not worth it to send all the data streams to the cloud (from a total of 65.1 million tuples, 38.1 million tuples have been deleted).

• Other computing architectures such as mobile fog computing.
People in Motion Lab
www.people-in-motion-lab.org