People in Motion

Implementing an Edge-Fog-Cloud Architecture for Stream Data Management

Lilian Hernandez, Hung Cao, And Monica Wachowicz
University of New Brunswick

CISCO

UNB
EST. 1785
UNIVERSITY OF NEW BRUNSWICK

Outline

- Overview
- Data Life Cycle Tasks and Processes
- System Architecture
- System Implementation
- Preliminary Results
- Future Work

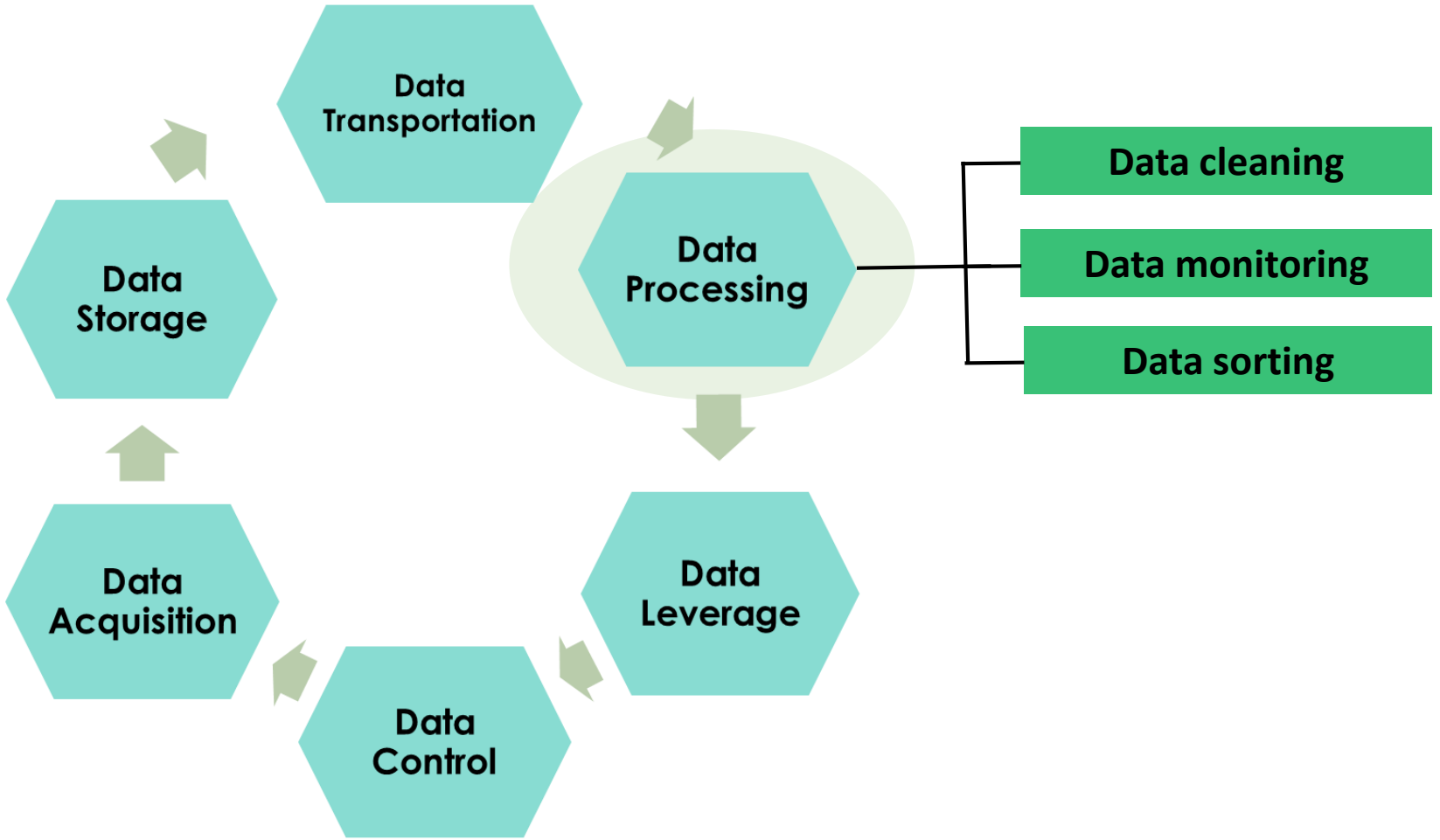
Overview

Data life cycle tasks integrate pre-processing ranging from sorting, cleaning, and monitoring to support more complex processes such as querying, aggregation, and analytics.

- How to handle the complexity of the data life cycle for increasing data rates.

- How to automate workflow tasks performed on IoMT data streams.

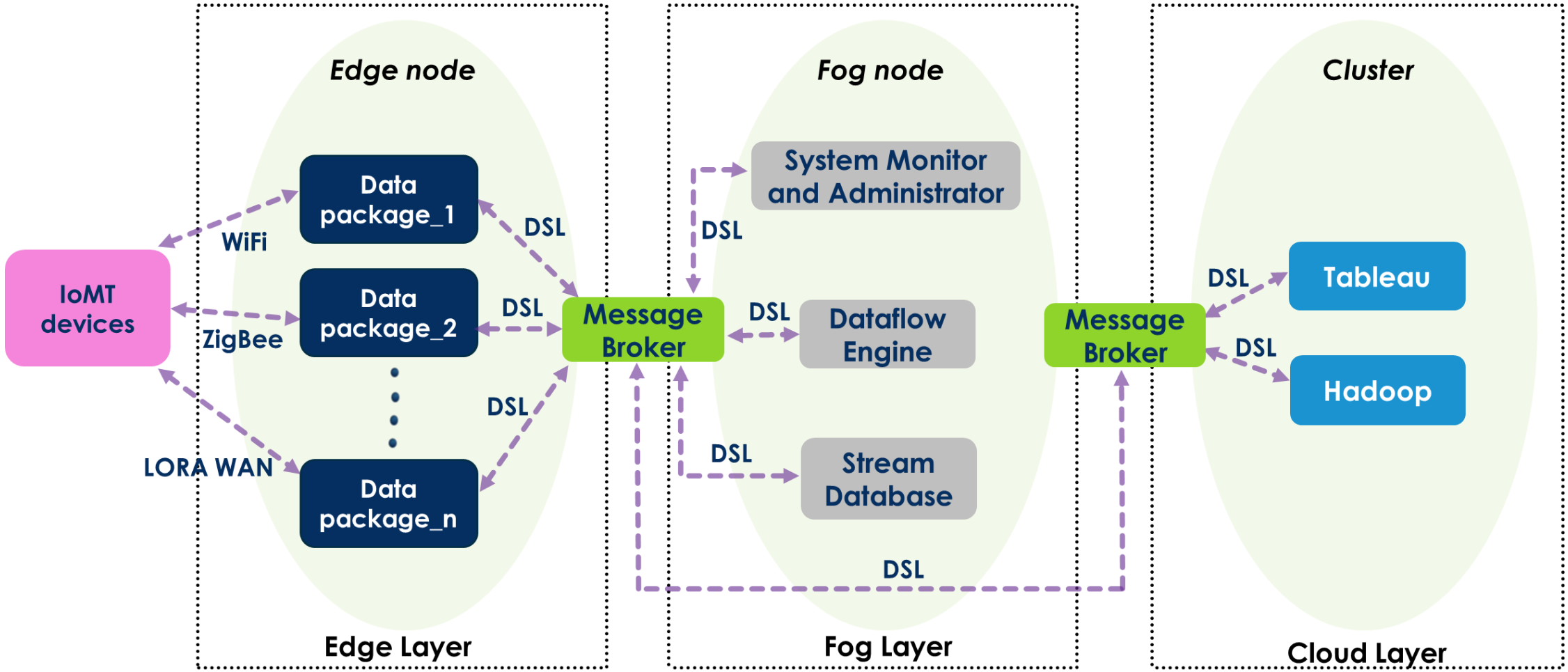
Data Life Cycle Tasks



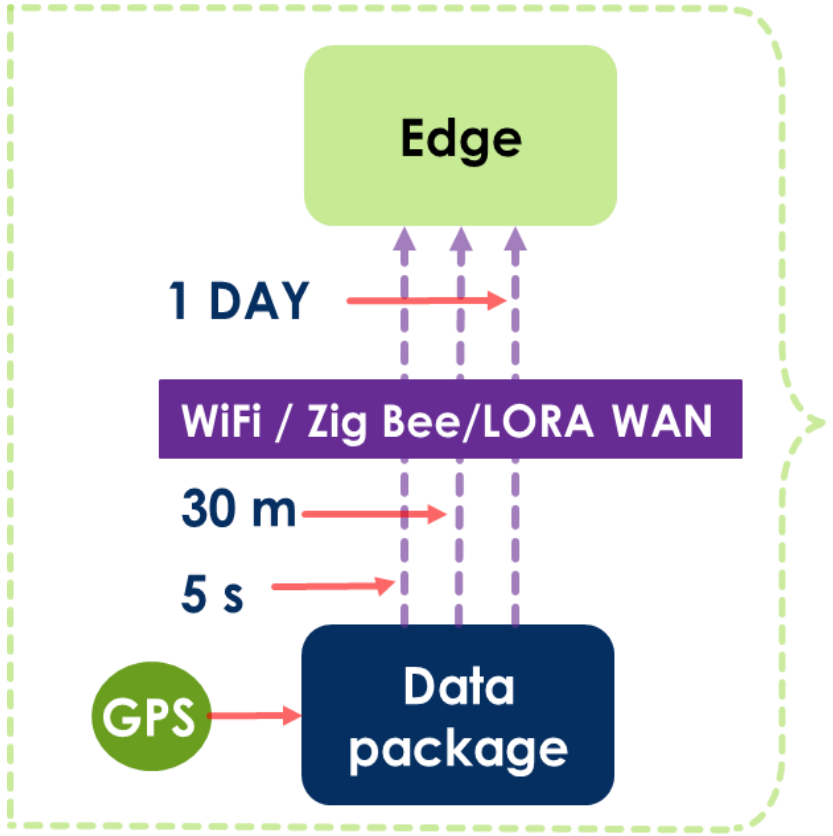
Commercial Stream Data Management Platform

Data life cycle tasks	**EFF (Cisco)**	Segment (Segment)	IBM Watson (IBM)	Axon Predict (Greenwave Systems and Intel)
Local Notification	**Yes**	Yes	Yes	Yes
Processing	**Yes**	No	Yes	Yes
Acquisition	**DSL**	Java API	Java API	Yes
Storage	**Yes**	Yes	Yes	Yes
Leverage	**Yes**	Yes	Yes	Yes
Data Control	**Yes**	Yes	Yes	Yes

System Architecture: 3 Tier



Data Streams at the Edge



The data streams consist of a sequence of out-of-order tuples:
T1=(S_1, x_1, y_1, t_1)

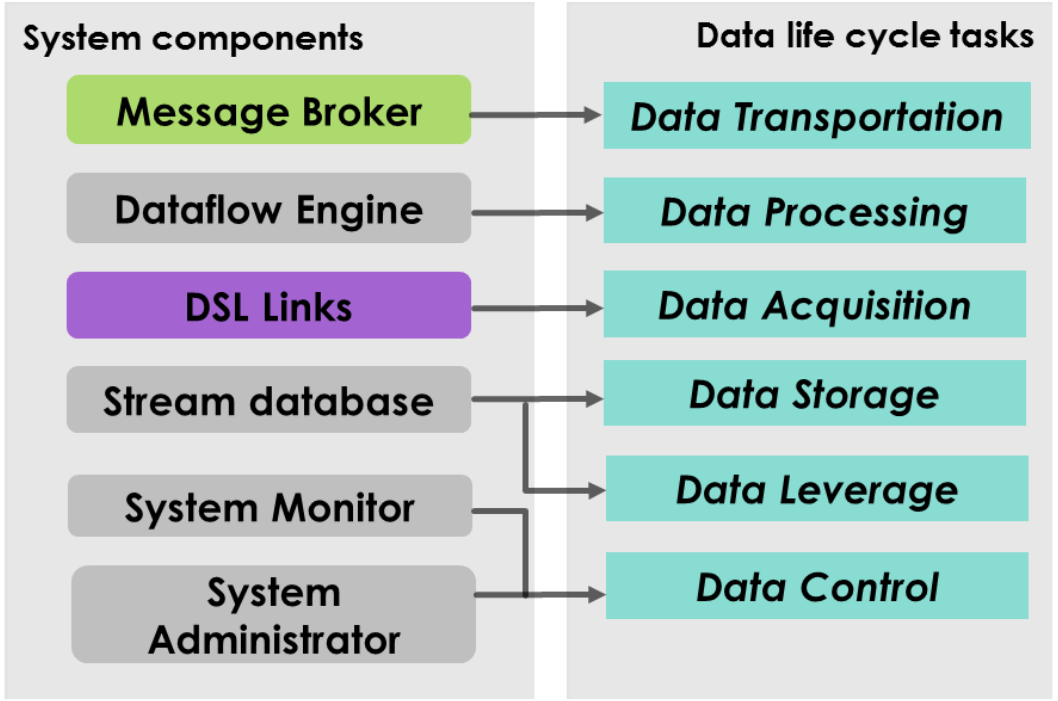
where:

S_1*:* is a set of attributes containing information about each IoMT device.
x_1, y_1, t_1 : is the geographical location of a device at the timestamp *t*.

Data Streams at the Fog

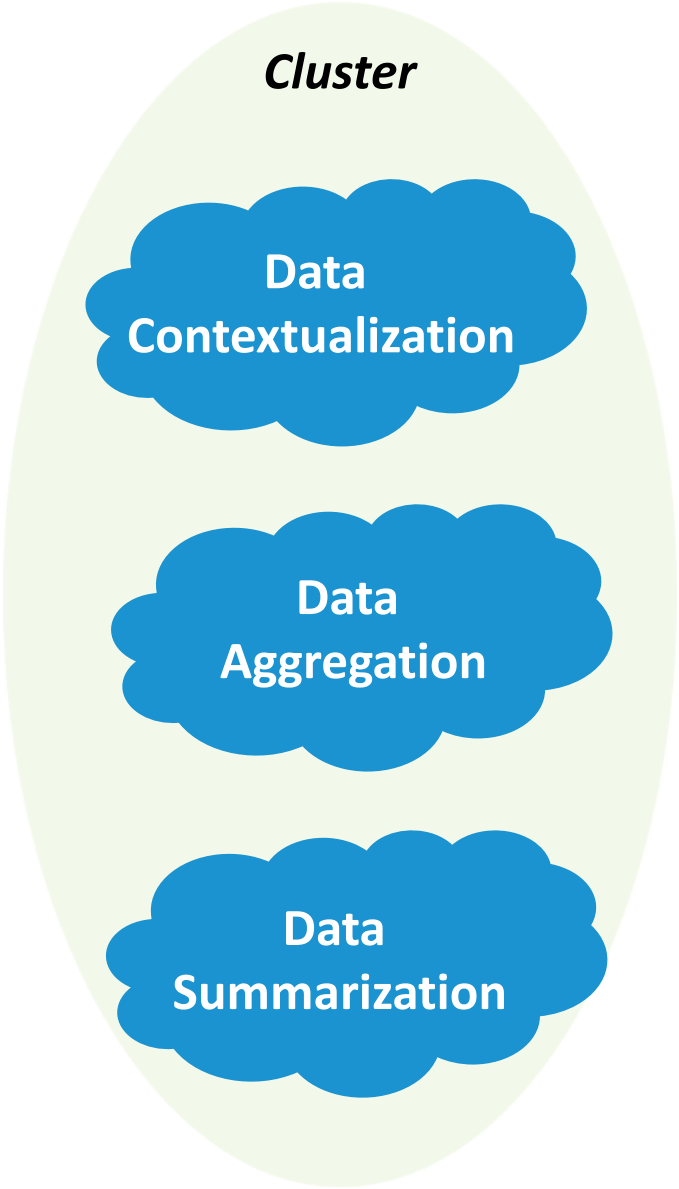
Proposed Streaming Model	
Phases	***Objectives***
Data flow	DSL links
Execution	Task sequence and data dependency
Control flow	Explicit/Implicit

Low Level Computing Processes	
Pre-processing tasks	***Objectives***
Data sorting	Order the data
Data cleaning	- Delete redundant tuples - Delete tuples with missing attributes - Delete tuples with misspelt attributes
Data monitoring	Task completion and data dependency

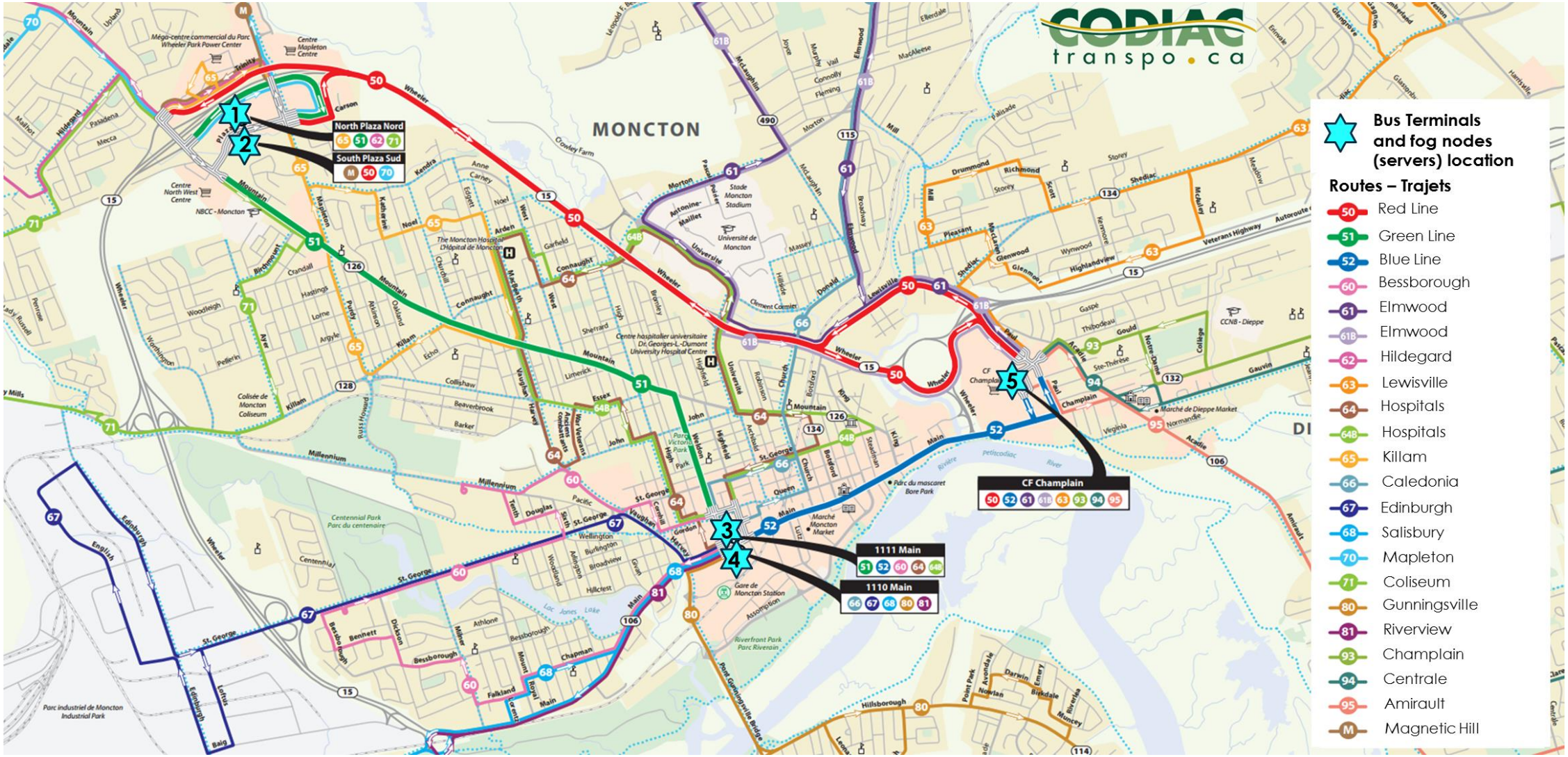


Data Streams at the Cloud

As a result of decentralizing computing processes to the edge and fog, the cloud may focus its functionalities to perform complex processes:



Transit Network Experiment



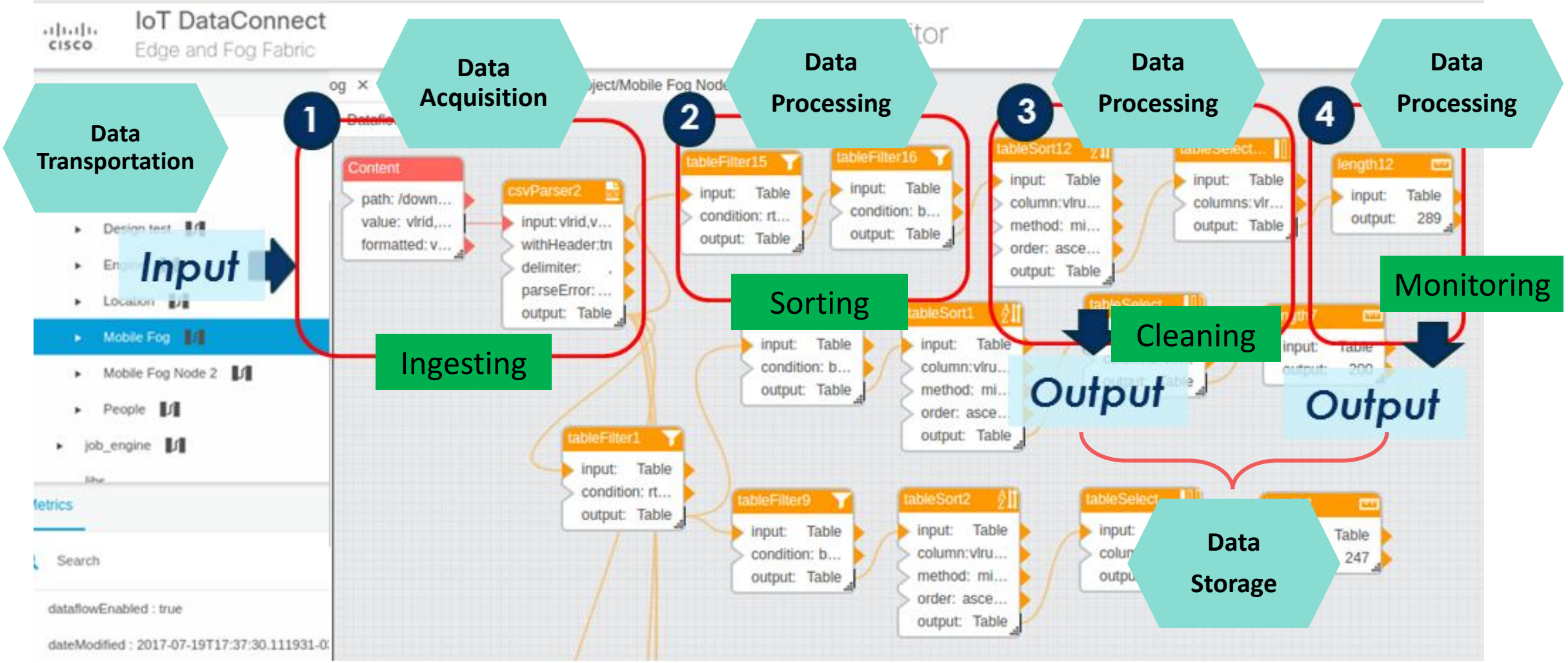
At the Edge

65 million tuples from June 1st 2016 to May 25th 2017

Attributes	Description
1. vlr_id	The ID of the data point in the vehicle location reports table.
2. route_id_vlr	The route ID in the vehicle location reports table.
3. route_name	The route name.
4. route_id_rta	The route ID in the route transit authority table.
5. route_nickname	The abbreviation of the route.
6. trip_id_br	The trip ID in the bid route table.
7. transit_authority_service_time_id	Transit authority service time ID.
8. trip_id_tta	Transit authority trip ID.
9. trip_start	Start time of the trip.
10. trip_finish	Finish time of the trip.
11. vehicle_id_vab	Vehicle ID.
12. vehicle_id_vlr	Vehicle ID in the vehicle locations reports table.
13. vehicle_id_vlr_ta	Descriptive name of the bus.
14. bdescription	Bus description.
15. lat	Latitude.
16. lng	Longitude.
17. timestamp	Timestamp of the data point.

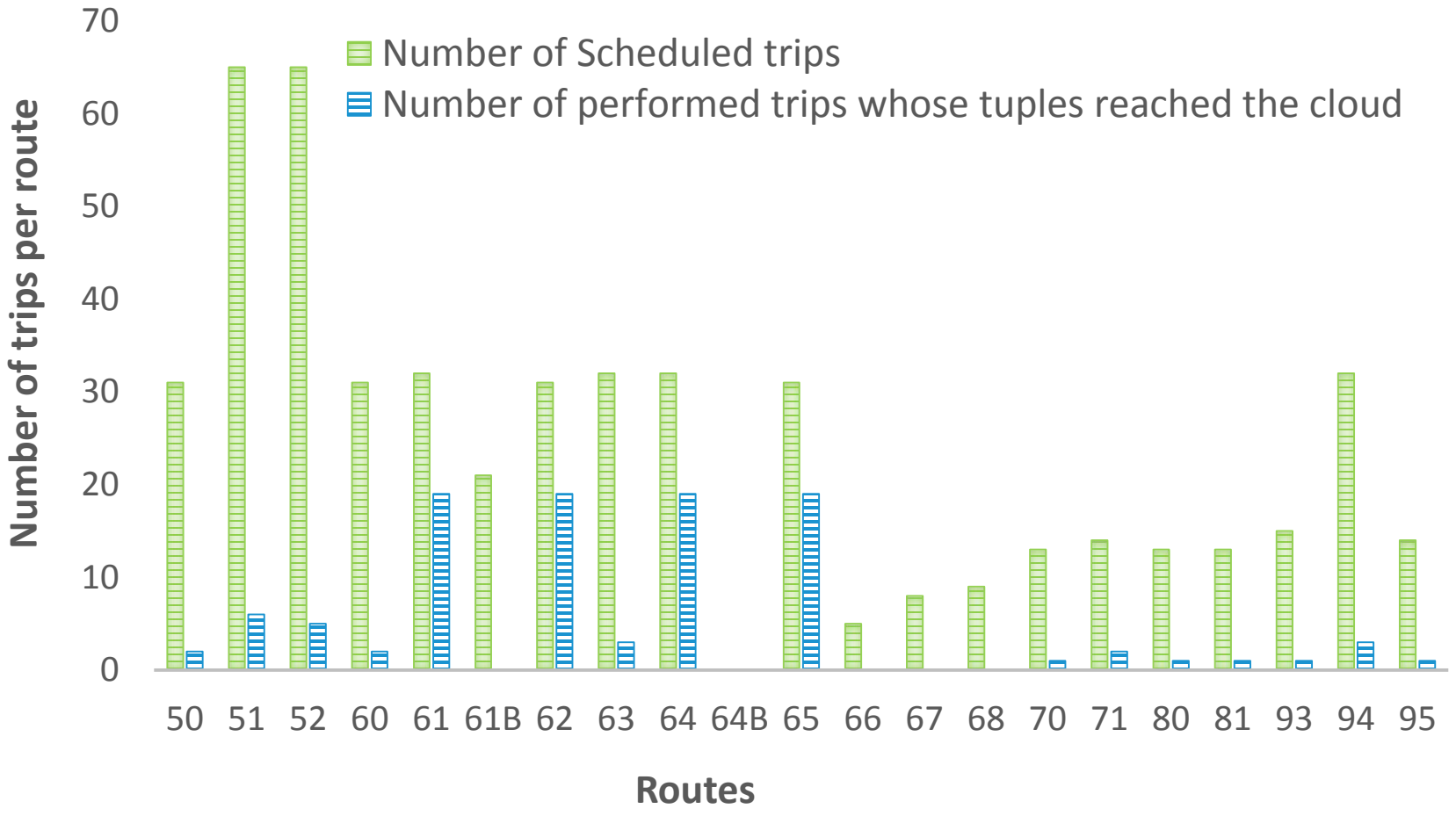
At the Fog

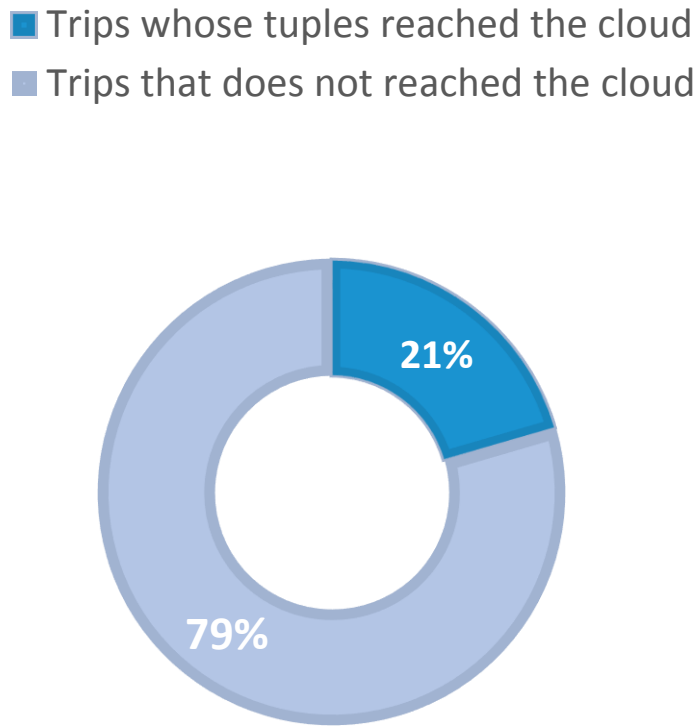
38.5 million tuples out of 65 million were deleted



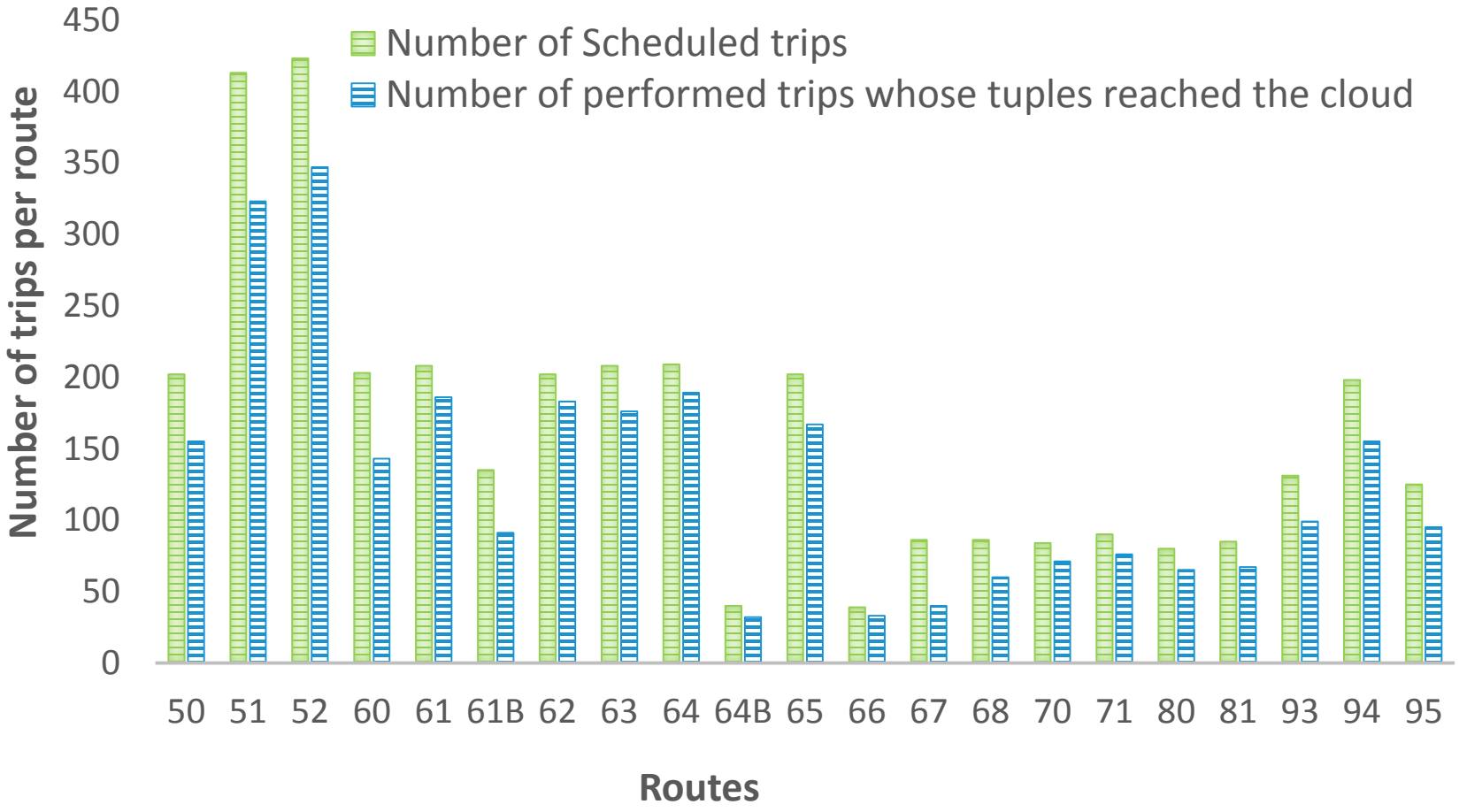
At the Cloud: One Day Performance

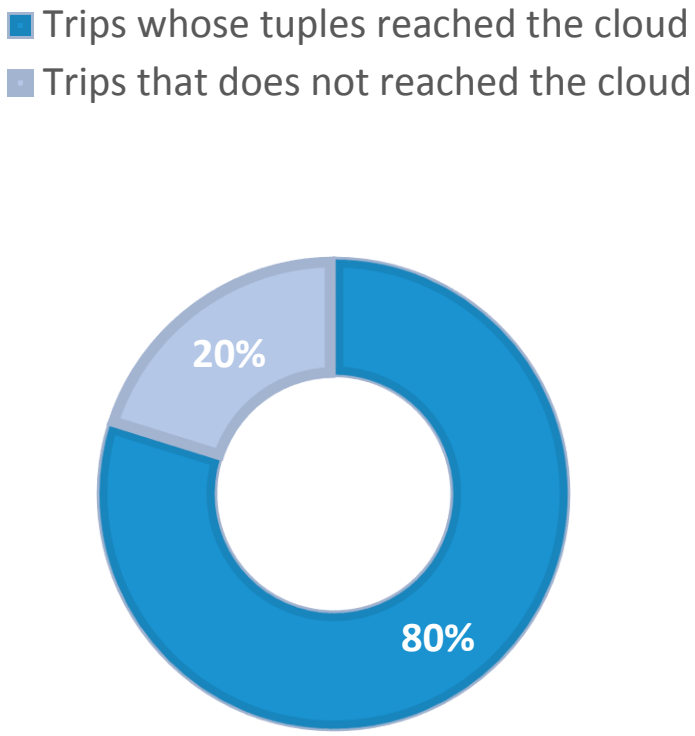
Only 26.5 millions tuples out of 65 millions were transported to the cloud





At the Cloud: One Week Performance





Conclusions and Future Work

- To make data available at real-time over an end to end architecture.
- To perform the implementation at the edge level.
- To explore a temporary storage at the edge.
- To collect and monitor other parameters, i.e. weather conditions, and passenger ridership.
- To create a process mining model through the use of Petri nets.
- To migrate our development from EFF Cisco to Kinetic Cisco platform.

Thanks for your attention!

Lilian Hernandez
lhernand@unb.ca
University of New Brunswick
People in Motion Lab