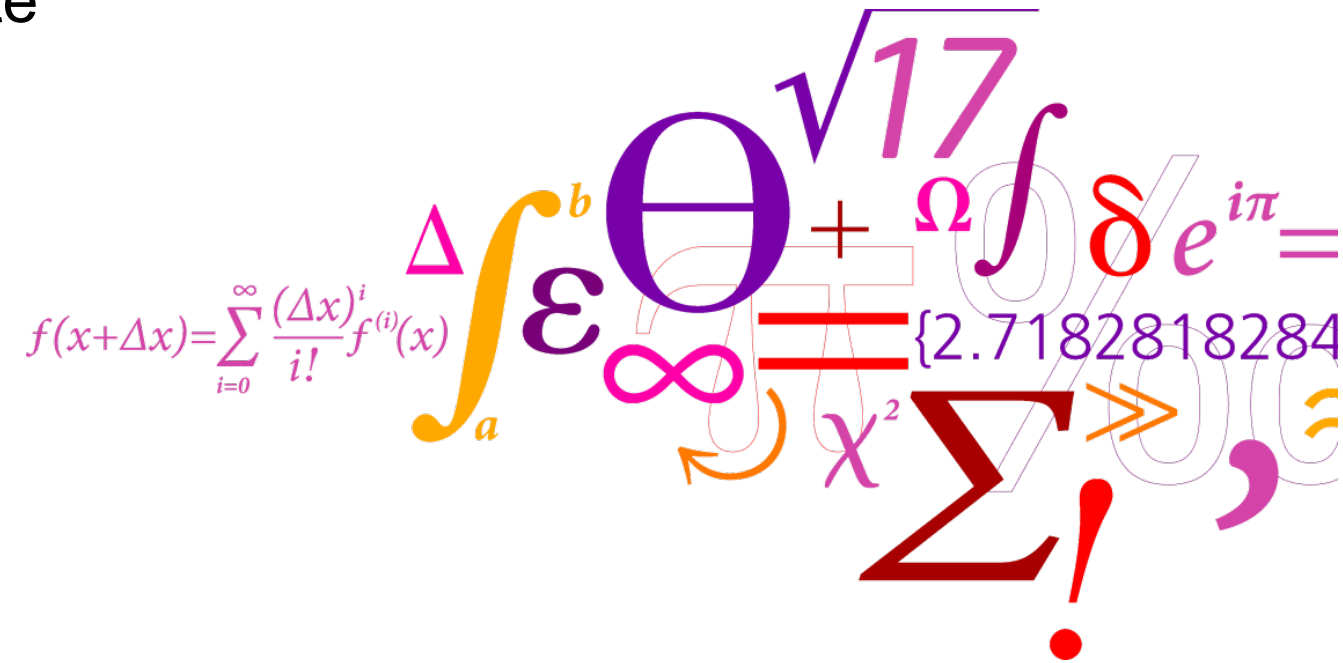
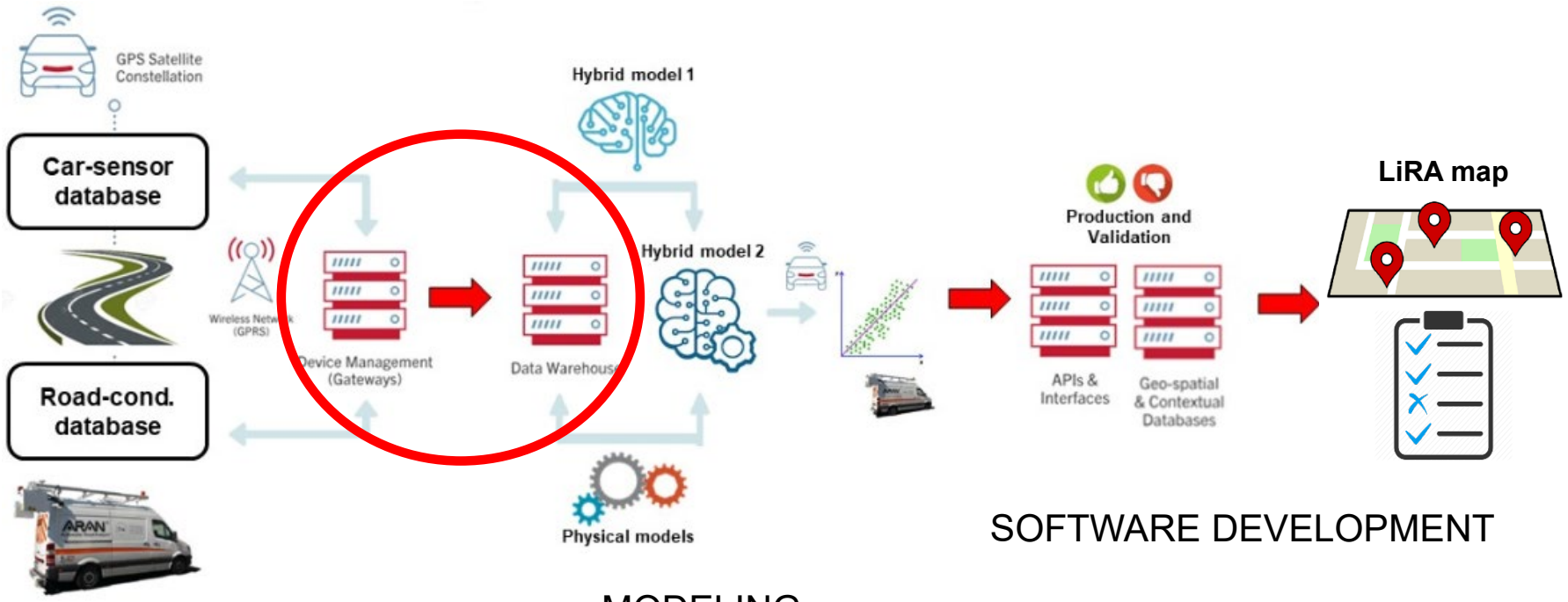
Shahrzad M. Pour

Postdoctoral Researcher at DTU Compute

email: shmp@dtu.dk



Overview


GPS Satellite Constellation
Car-sensor database
Road-cond. database
Wireless Network (GPRS)
Device Management (Gateways)
Data Warehouse
Hybrid model 1
Hybrid model 2
Physical models
Production and Validation
APIs & Interfaces
Geo-spatial & Contextual Databases
LiRA map
MODELING
SOFTWARE DEVELOPMENT

Overview

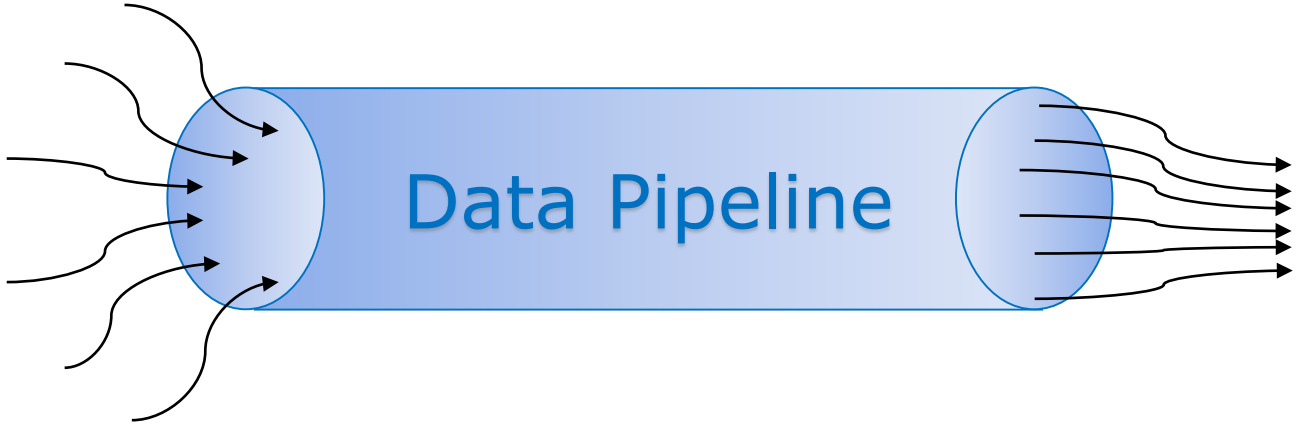
DIFFERENT SOURCES OF DATA

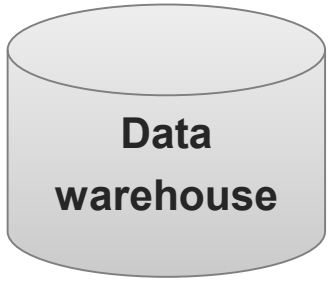
SINGLE LOCATION

AutoPi sensors

ARAN devices

Historical Condition data





Outline

- **Introduction**
  IoT & ROAD Transportation &
  – The data infrastructure Requirements in LiRA

- **LiRA Data Pipeline**
  – Components
  – Proposed pipeline Architecture
  – A bit of Implementation

- **LiRA Data Warehouse**
  – Designing the Data warehouse
  – IoT Design methodology
  – A bit of Implementation & size of data

- **Pros & Cons**

Introduction to IoT - Internet of Things

- A Network of **Physical** objects – “Things”

- **IoT is an Eco-System**
  - A network of Things
  - Technologies
  - Goal

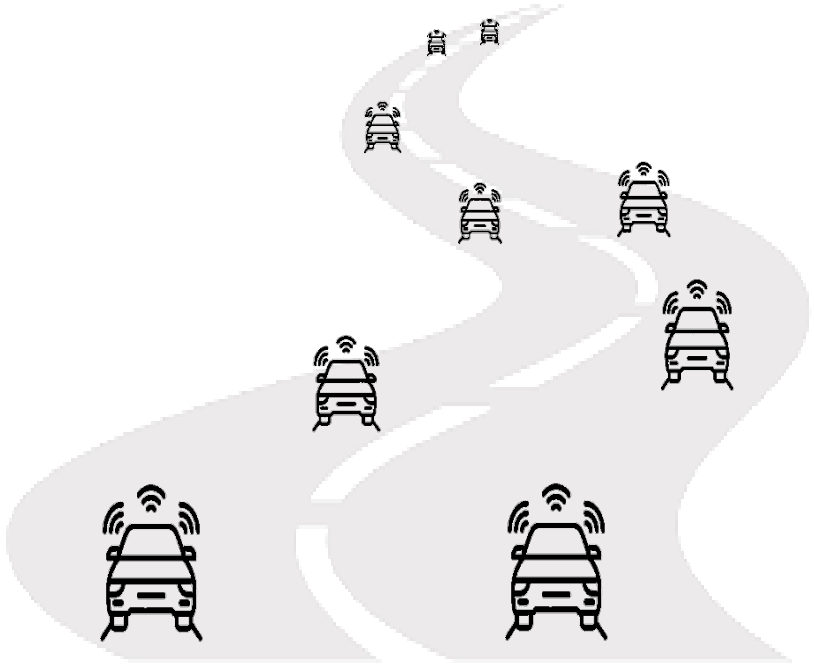
- **Machine learning**

Rouse, Margaret (2019). "internet of things (IoT)". IOT Agenda. Retrieved 14 August 2019

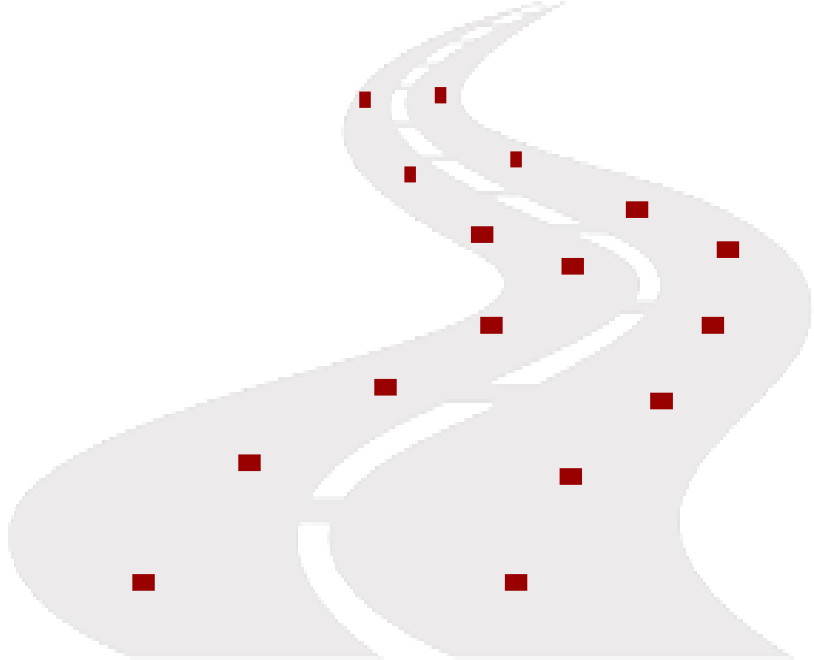
Internet of Things In ROAD Transportation

THINGS:

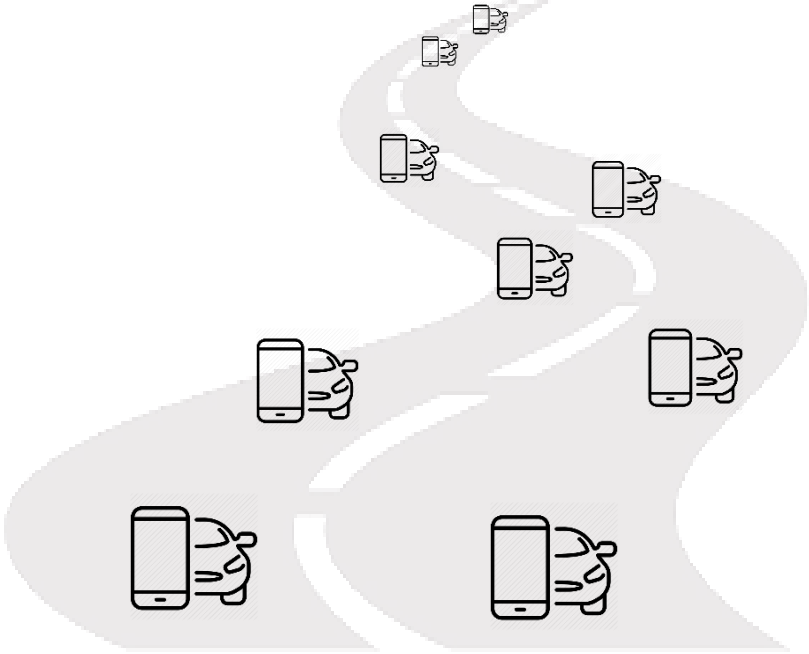
The vehicle



The infrastructure



The driver/user



The SCOPE of LiRA

Outline

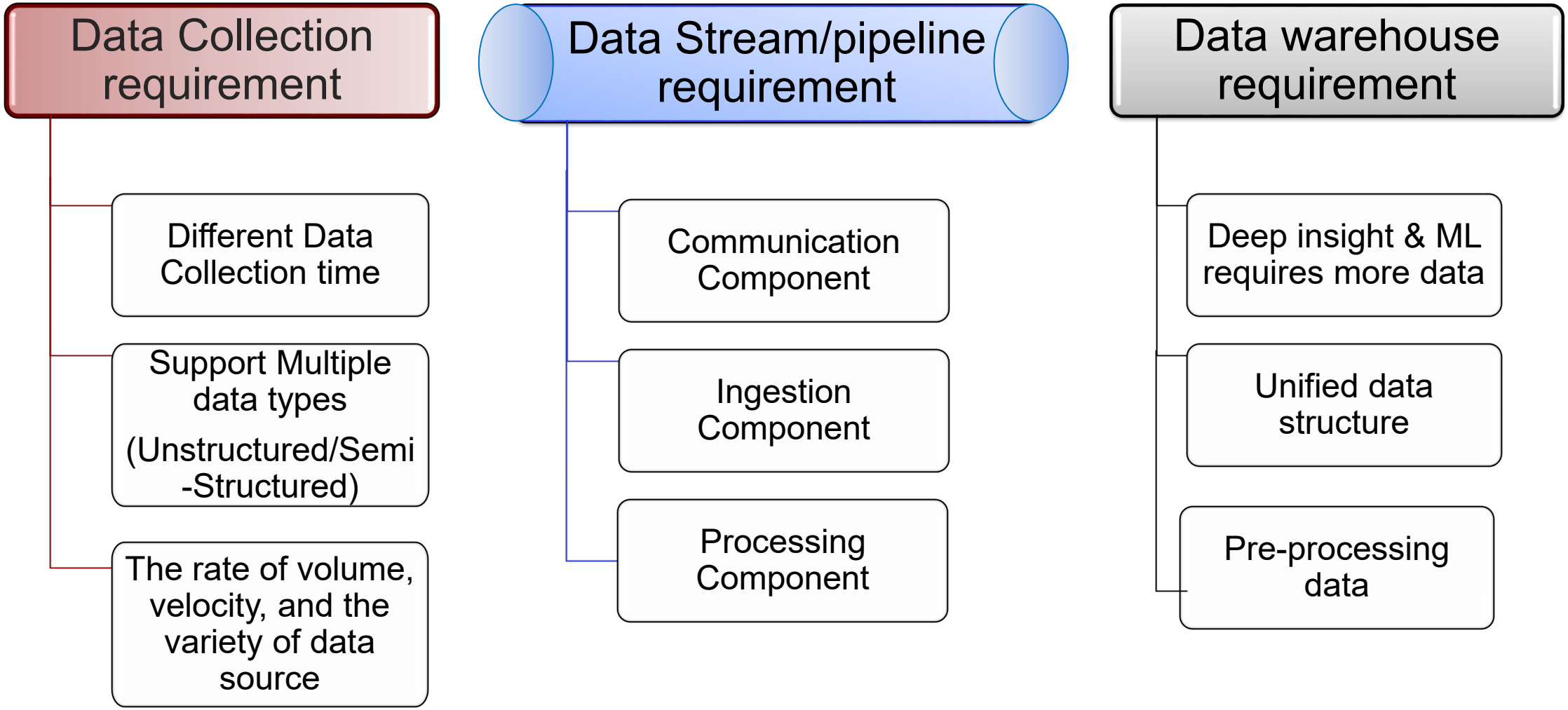
- **Introduction**
  IoT & ROAD Transportation &
  – The data infrastructure Requirements in LiRA

- **LiRA Data Pipeline**
  – Components
  – Proposed pipeline Architecture
  – A bit of Implementation

- **LiRA Data Warehouse**
  – Designing the Data warehouse
  – IoT Design methodology
  – A bit of Implementation & size of data

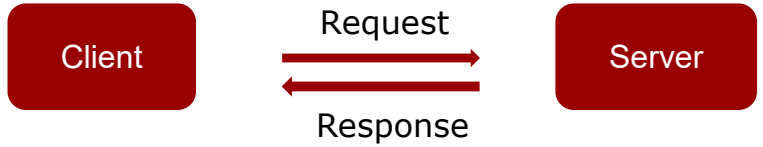
- **Pros & Cons**

The Architectural data infrastructure Requirement



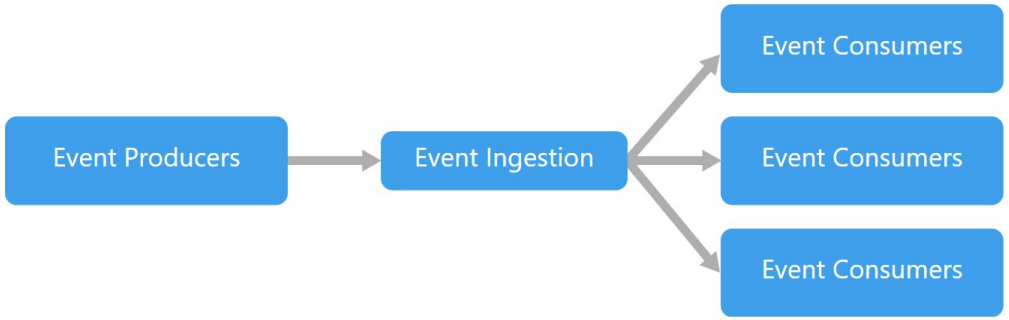
Data pipeline Components

- **Communication Component:**
  - **Response-Request** Communication model



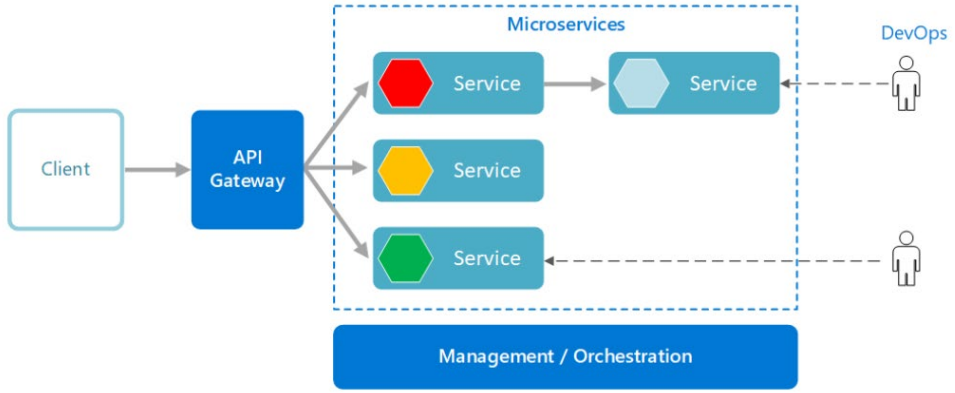
- **Ingestion Component:**
  – Hybridization of **Event-Driven Architecture** & **Micro-service Architecture**

Event-Driven Architecture



https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/event-driven

Micro-Service Architecture



https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/microservices

Data pipeline Components

- **Data Processing Component:**
  - **Data filtering/Cleaning**
    - Remove extrapolated measurements
    - Sorting Measurements based on Timestamp
  - **Data Interpolation**
    - Linear
    - Quadratic
    - Cubic
  - **Map-matching**
    - Hidden Markov Model ((Paul Newson and John Krumm, 2009)
    - Map Matching Service from OSRM library
      - » from Open Source Routing Machine(OSRM)
      - » It matches/snaps given GPS points to the road network

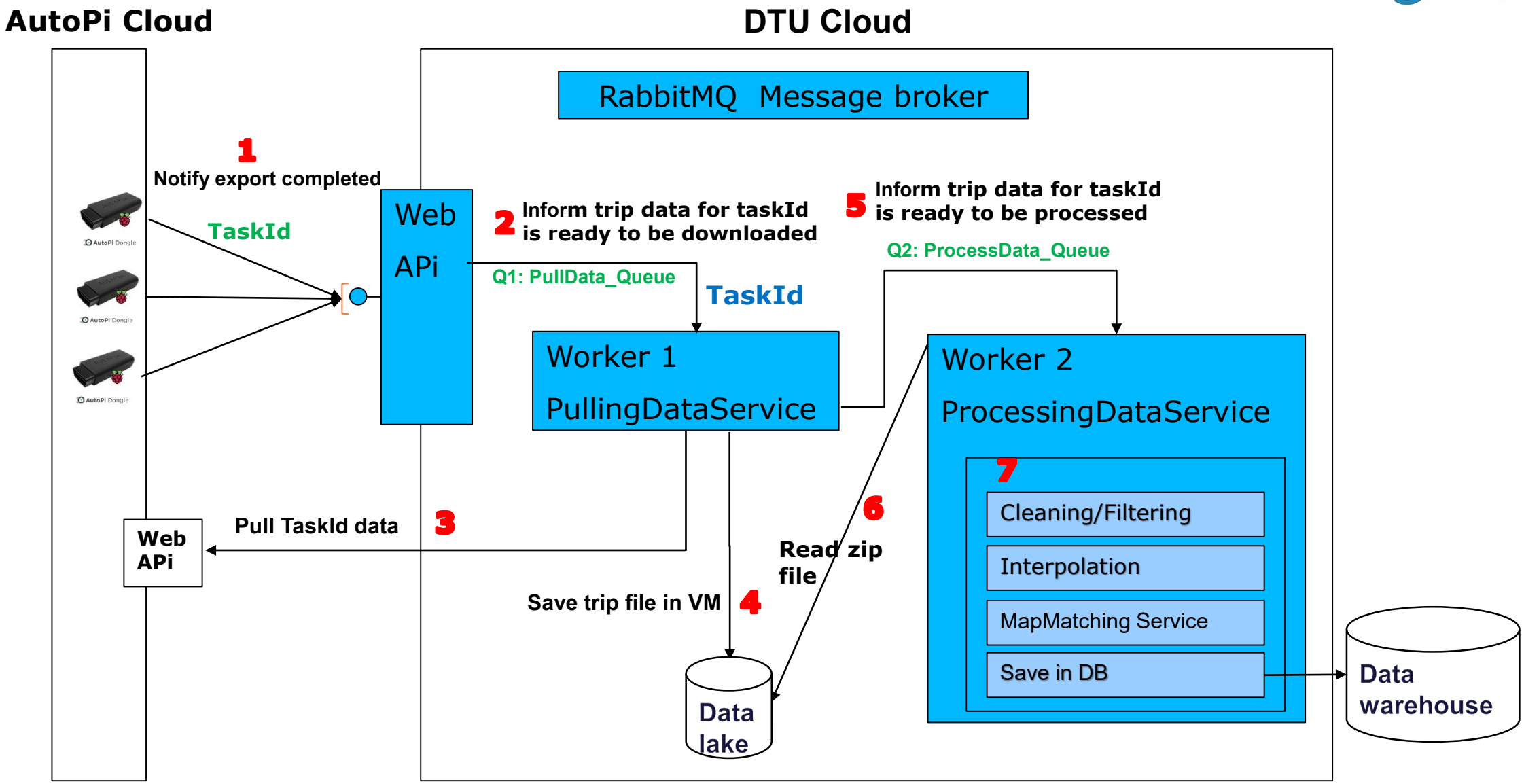
Raw data

Interpolation

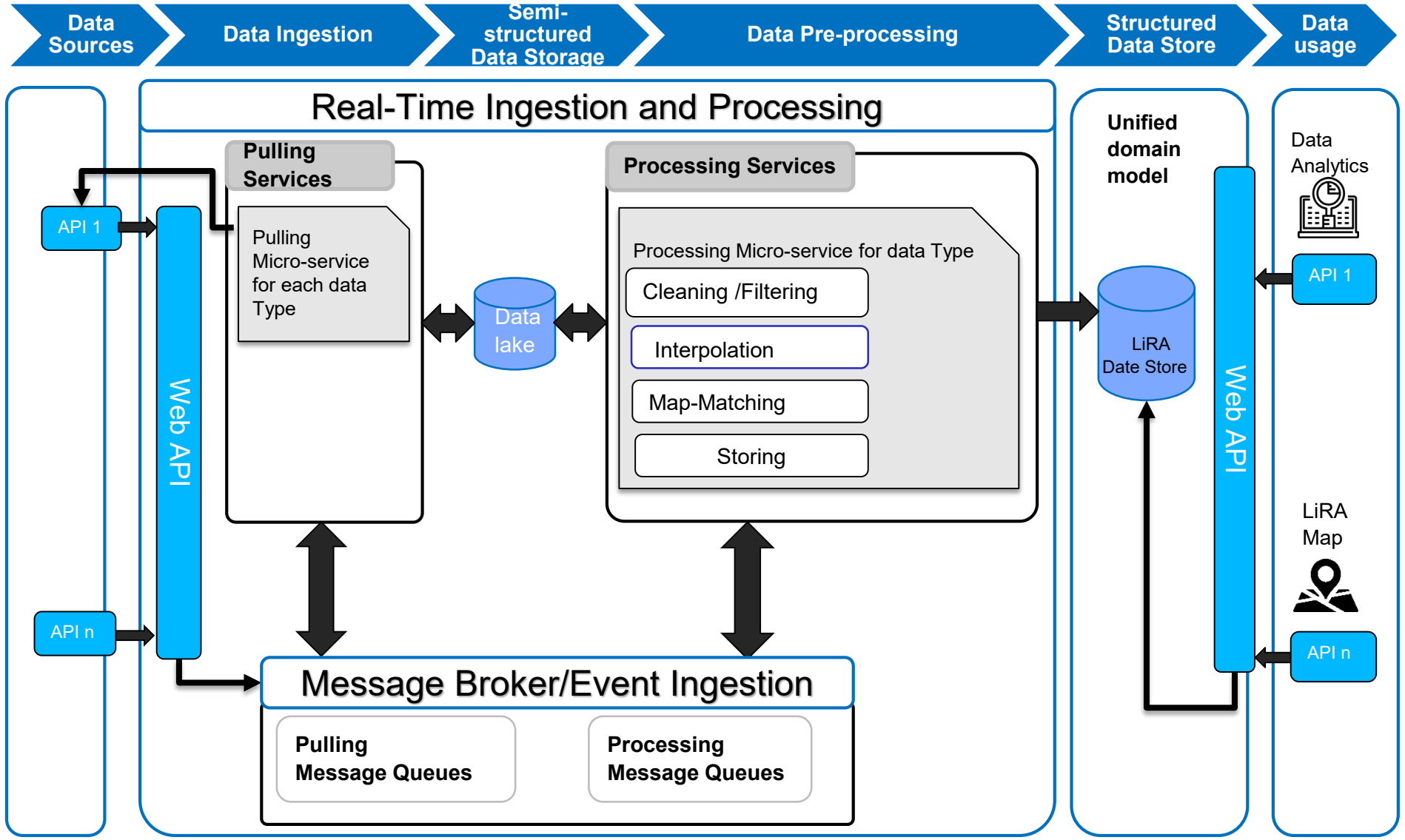
Interpolated data

Map matched data

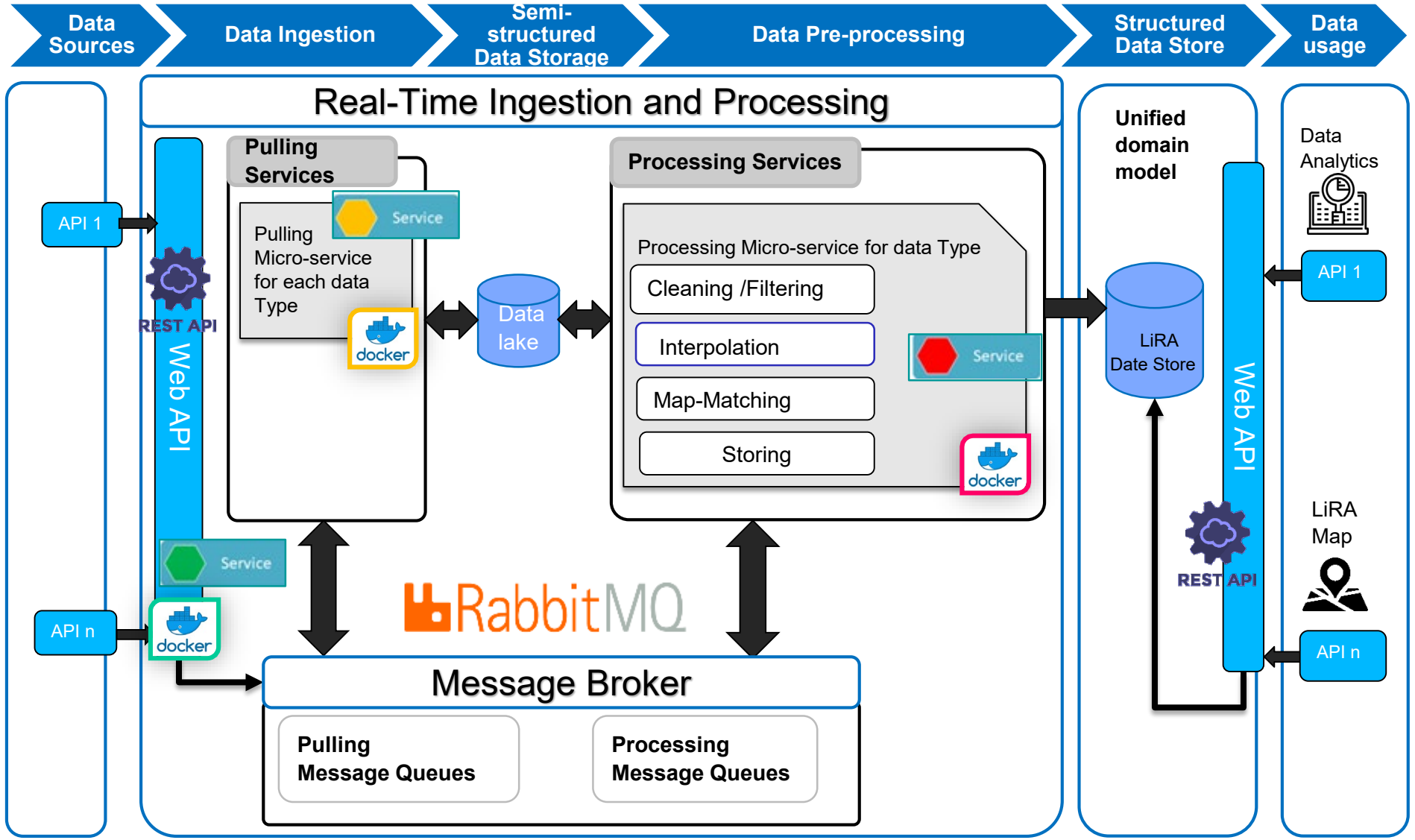
An example of ingesting one data source



Data Pipeline Architecture



A bit of Implementation



Outline

- **Introduction**
  IoT & ROAD Transportation &
  – The data infrastructure Requirements in LiRA

- **LiRA Data Pipeline**
  – Components
  – Proposed pipeline Architecture
  – A bit of Implementation

- **LiRA Data Warehouse**
  – Designing the Data warehouse
  – IoT Design methodology
  – A bit of Implementation & size of data

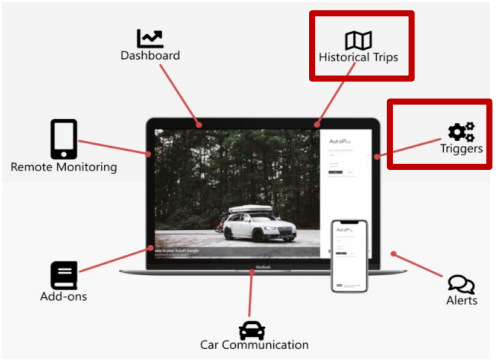
- **Pros & Cons**

Designing the Data warehouse

- **Data format & Behavior**

Auto Pi Device

Behavior



Distinct trip

IoT Design methodology [1]

Device: AutoPi

- SourceType
  - SourceType Id

Physical entity: Vehicle

- Device/vehicle
  - Device Id

- Trip
  - Start and end position
  - Start and end time
  - Device Id
  - List of events
  - Tag (Personal/Business)

- Measurement types
  - track position (GPS)
  - accelerometer
  - device events
  - obd
  - rpi

- Measurement
  - Object Id
  - Timestamp
  - Type
  - Attributes (flat/structured)

Virtual Entity: Trip

- MapReference
  - Object Id
  - Timestamp
  - Type
  - Attributes (flat/structured)

1. Internet of Things: A Hands-On Approach (Introduced by Bahga & Madisetti, 2015)

A bit of Implementation & size of data

- **PostgreSQL** Database Management System(Open-Source & relational database)
- **Creating database using** Code-First Approach and by **Entity Framework**
- **600+** number of trips have been imported to the database
- Disk Space ~ 100 GB

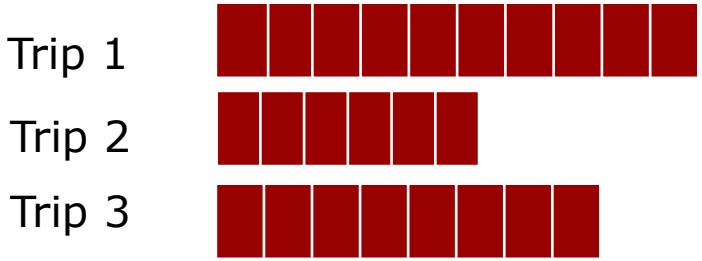
- Range of data of trips

AutoPi Type	Shortest Trip	Number of measurements	Longest Trips	Number of measurements	No of Measurement Types
Gen 2	6 min	4726	1 h & 2 min	108,852	66
Gen 3	2 min	5898	4 h & 15 min	15,682,197	66

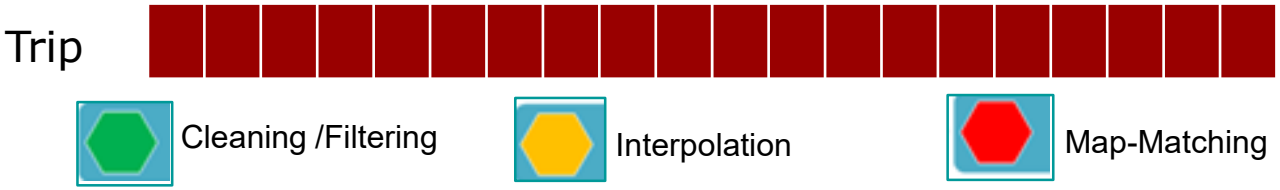
Pros & Cons

Data Pipeline

- Pros
  - Unified & Generic architecture
  - Scalabe in multiple Trips level
    - Messaging mechanism
    - Micro-service architecture



- Cons
  - Not scalable within one single Trip



Data warehouse

- Pros
  - Generic structure
  - So far supports two different sources of data
  - Easy to use structurally (feedback from ML team)
- Cons
  - No interface is provided to ML & CIVIL yet
  - No proper parameter tuning has been done yet (size of batches)

TQ

?