Developing an analytics everywhere framework for the Internet of Things

Ph.D. Research Proposal by Hung Cao

Research Motivation

The Internet of Things

1. **IoT devices** require **pushing** the data streams from the devices to near-by fog nodes and remote clouds.
2. The IoT data streams have **fast incoming data rates.**
3. Analytics needs to **generate new insights in a timely way** before the data streams become outdated.
4. IoT applications require **seamlessly computation of analytical tasks and distributed storage** over a vast geographical area.

Current Research in IoT

1. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). **Internet of things: A survey on enabling technologies, protocols, and applications.** *IEEE Communications Surveys & Tutorials*, *17*(4), 2347-2376.
2. Li, S., Da Xu, L., & Zhao, S. (2015). **The internet of things: a survey.** *Information Systems Frontiers*, *17*(2), 243-259.
3. Ngu, A. H., Gutierrez, M., Metsis, V., Nepal, S., & Sheng, Q. Z. (2017). **IoT middleware: A survey on issues and enabling technologies.** *IEEE Internet of Things Journal*, *4*(1), 1-20.
4. Gazis, V. (2017). **A Survey of Standards for Machine-to-Machine and the Internet of Things.** *IEEE Communications Surveys & Tutorials*, *19*(1), 482-511.
5. Singh, K. J., & Kapoor, D. S. (2017). **Create Your Own Internet of Things: A survey of IoT platforms.** *IEEE Consumer Electronics Magazine*, *6*(2), 57-68.
6. Verma, S., Kawamoto, Y., Fadlullah, Z., Nishiyama, H., & Kato, N. (2017). **A Survey on Network Methodologies for Real-Time Analytics of Massive IoT Data and Open Research Issues.** *IEEE Communications Surveys & Tutorials*.

First Phase

Core research challenges were related to **connectivity**, **physical infrastructure**, **sensors**, and **hardware configurations**.

Second Phase

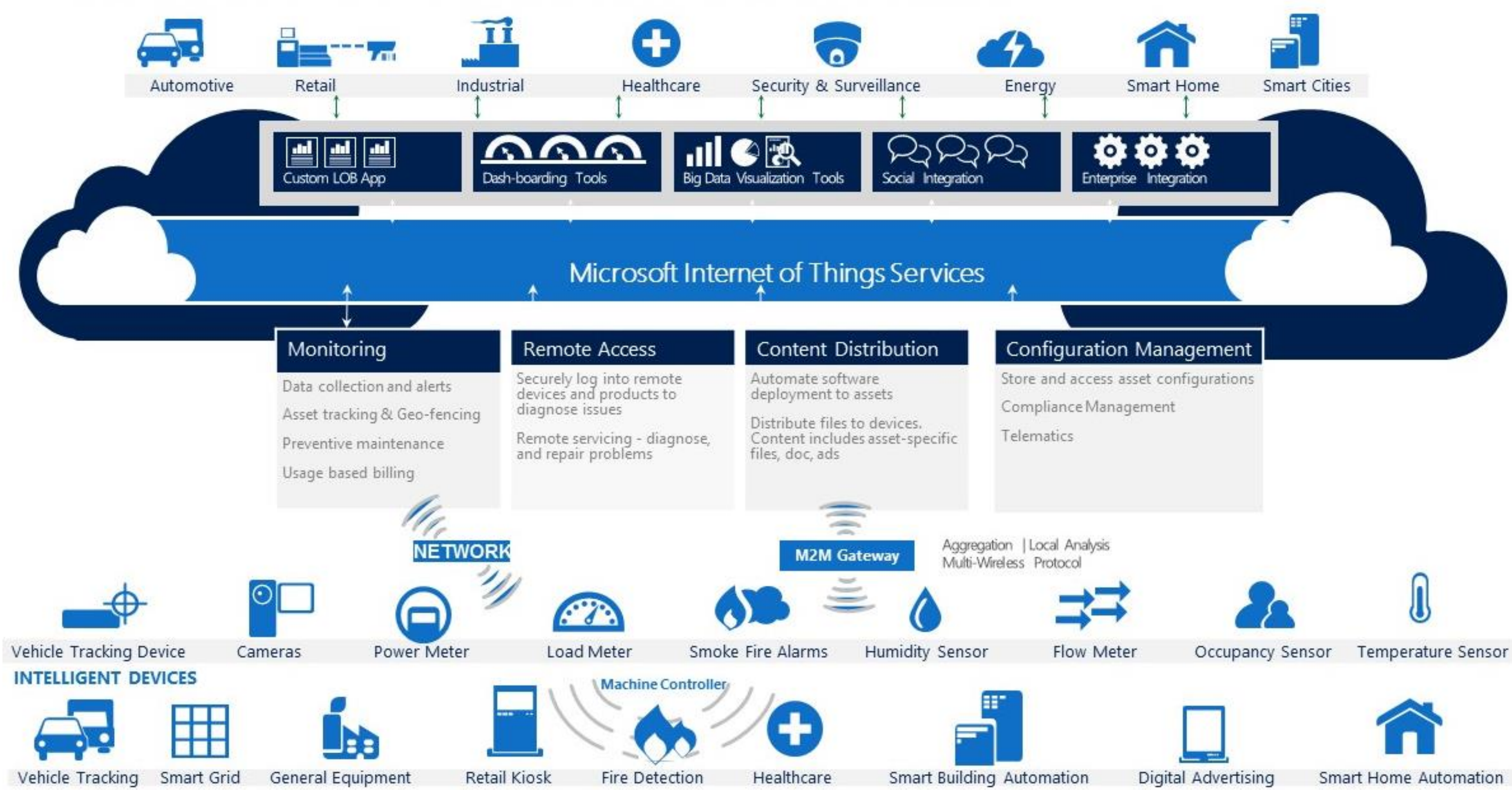
Core research and challenges are shifting to **software**, **analytics**, and **platform configuration**.

Research Status Quo

Many studies related to processing & analyzing IoT data streams tend to stick to one computing resource:

- Cloud Computing
- Edge Computing
- Fog Computing

Cloud Computing



Source: http://select.advantech.com/build-your-own-iot_windows-10/

Cloud Computing

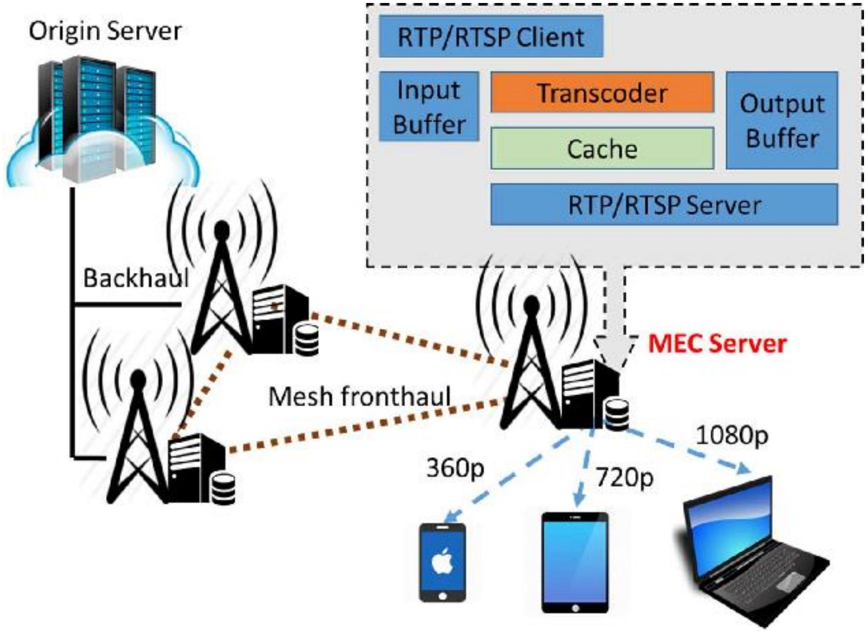
Advantages

- on-demand processing service
- high availability
- rapid elasticity
- virtually unlimited resources
- low costs for deployment
- complex data analytical tasks

Disadvantages

- not adequate to support the short response time
- generate bottlenecks
- batch processing with historical datasets rather than for the IoT data streams

Edge Computing



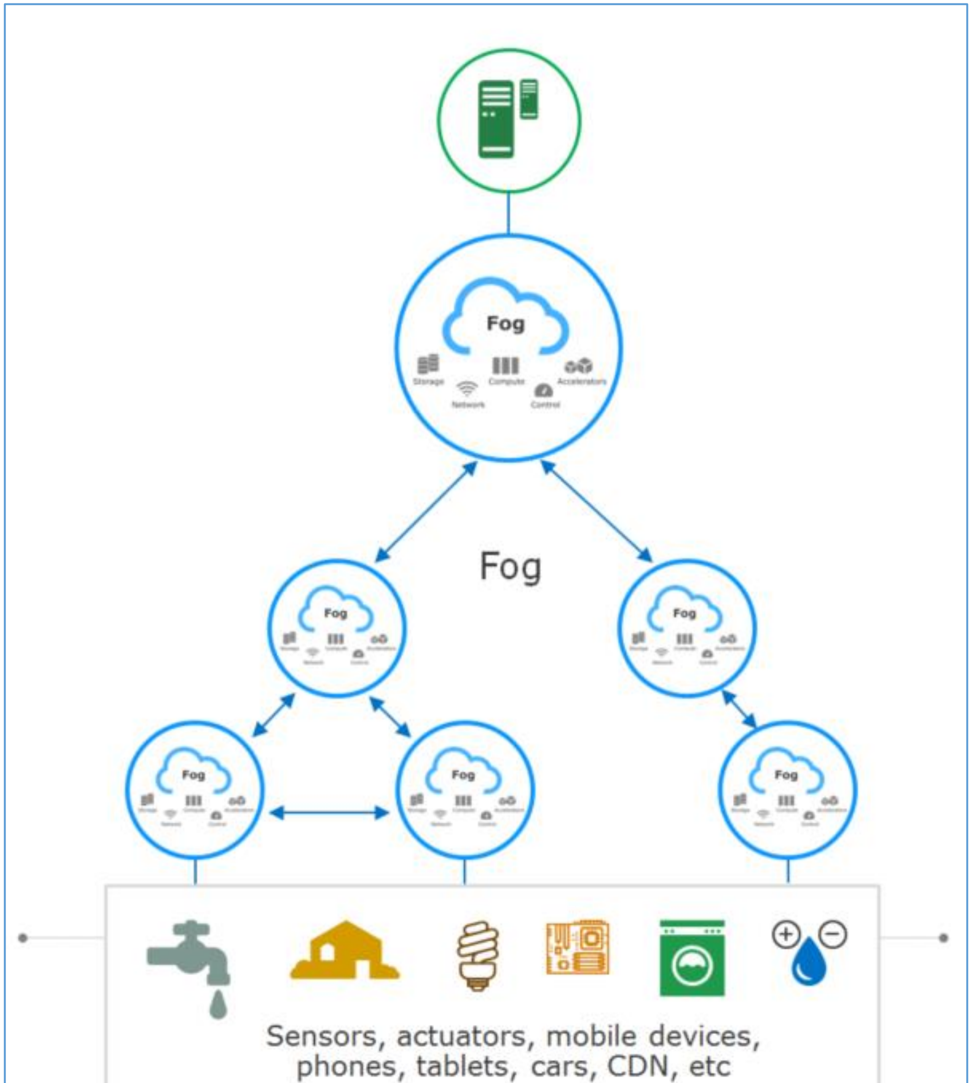
Mobile Edge Computing

Tran, T. X., Hajisami, A., Pandey, P., & Pompili, D. (2017). Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges. *IEEE Communications Magazine*, *55*(4), 54-61.

Edge Computing

Advantages	Disadvantages
• vicinity of IoT devices • decreasing network congestion • alleviating the saturation of network bandwidth • accelerating analytical tasks	• lightweight with low processing and storage capabilities • resource contention • increase processing latency • few attempts were found in the literature using edge computing for supporting data analytics

Fog Computing



OpenFog Consortium. (2017). *OpenFog Reference Architecture for fog computing*. Tech. Rep., February

Fog Computing

Advantages

- Fog computing as an **intermediate resource** that can seamlessly integrate edge and cloud resources.
- Fog computing can **eliminate** resource **contention** at the edge by supporting several analytical tasks at the fog nodes and **coordinating** the use of **geographically distributed IoT devices** more efficiently than in the cloud

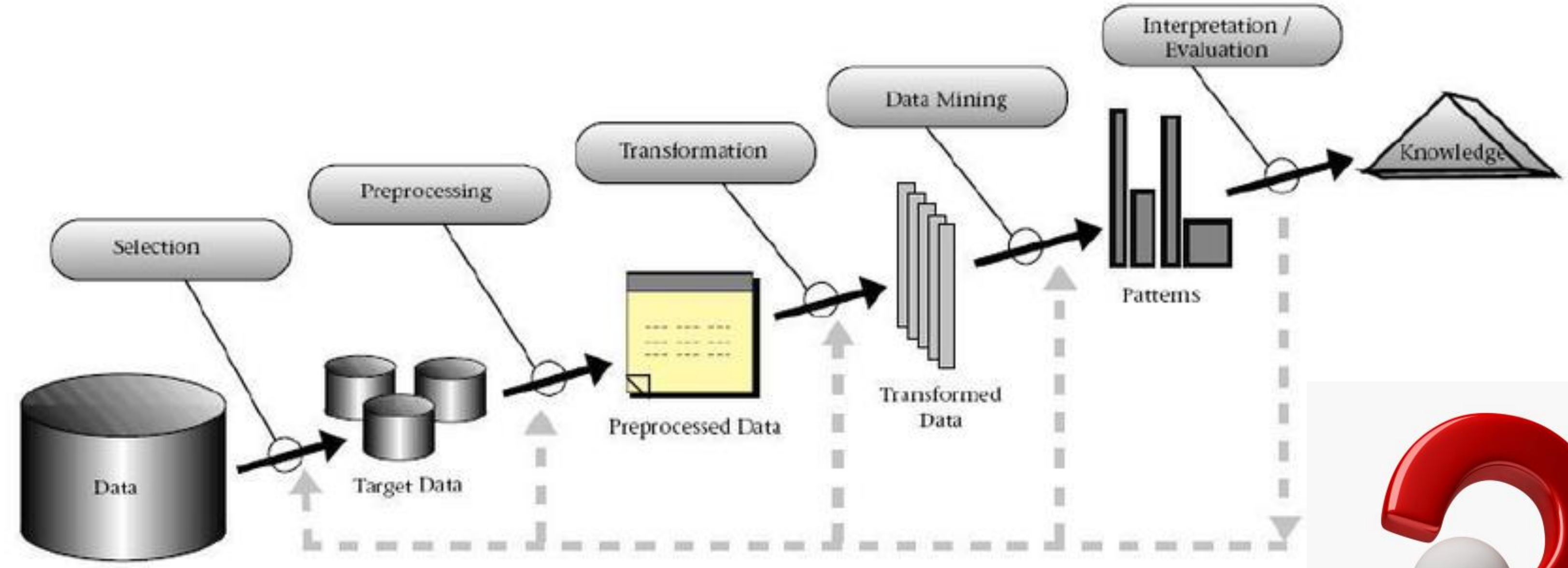
Disadvantages

- Resource capability is at the medium level.
- Not suitable to store large IoT data streams.
- Waste of resources if used to analyze small datasets.

Research Challenges

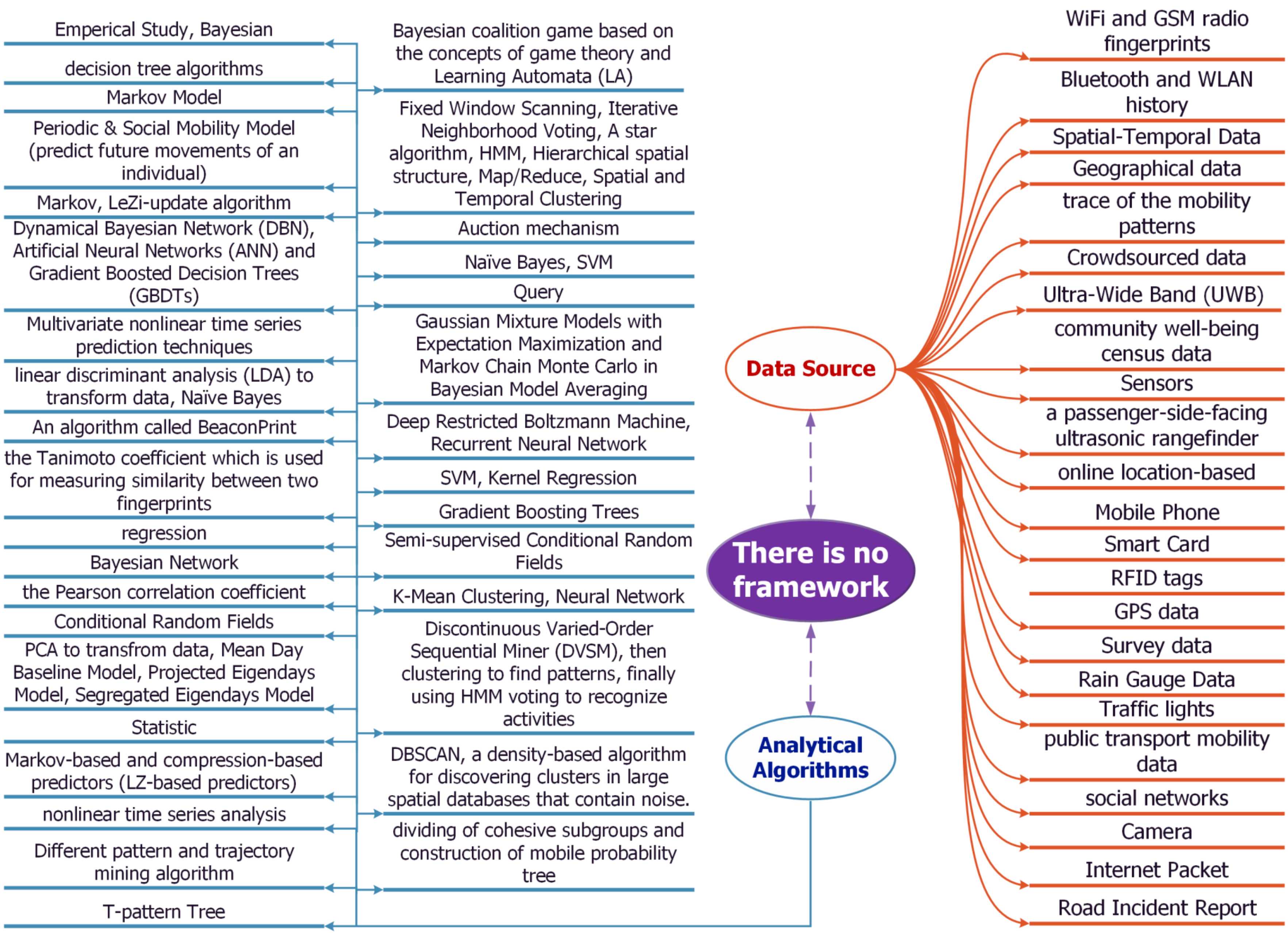
- Analytical tasks/algorithms for processing IoT data streams are not necessarily the same tasks/algorithms previously developed for analytical workflows in data mining or KDD process.

KDD Workflow (tasks)

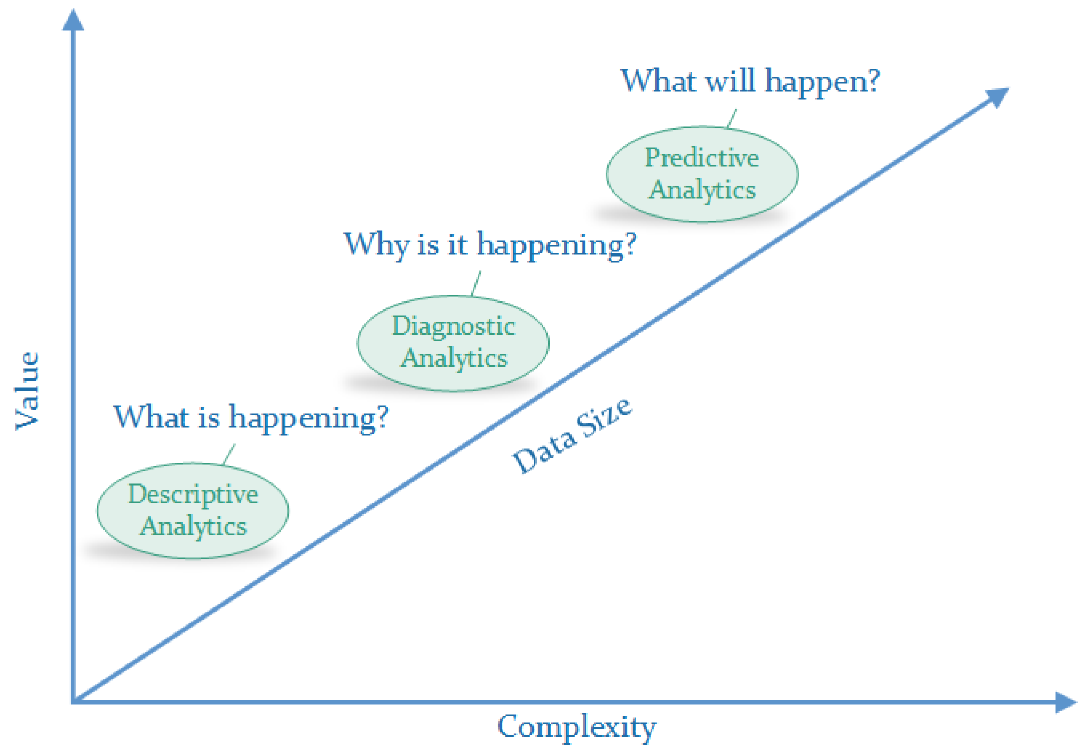


An overview of the steps that compose the knowledge discovery in databases (Fayyad et al. 1996)

Analytical Algorithms

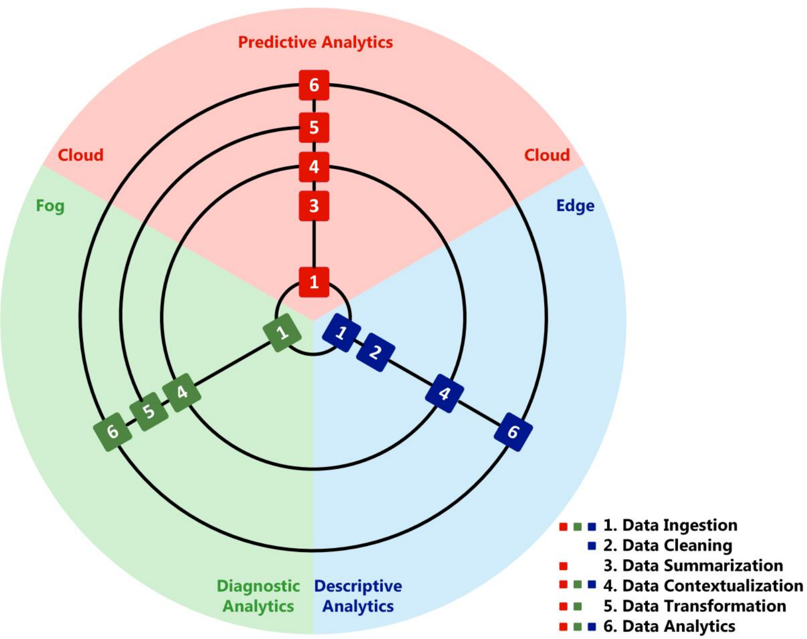


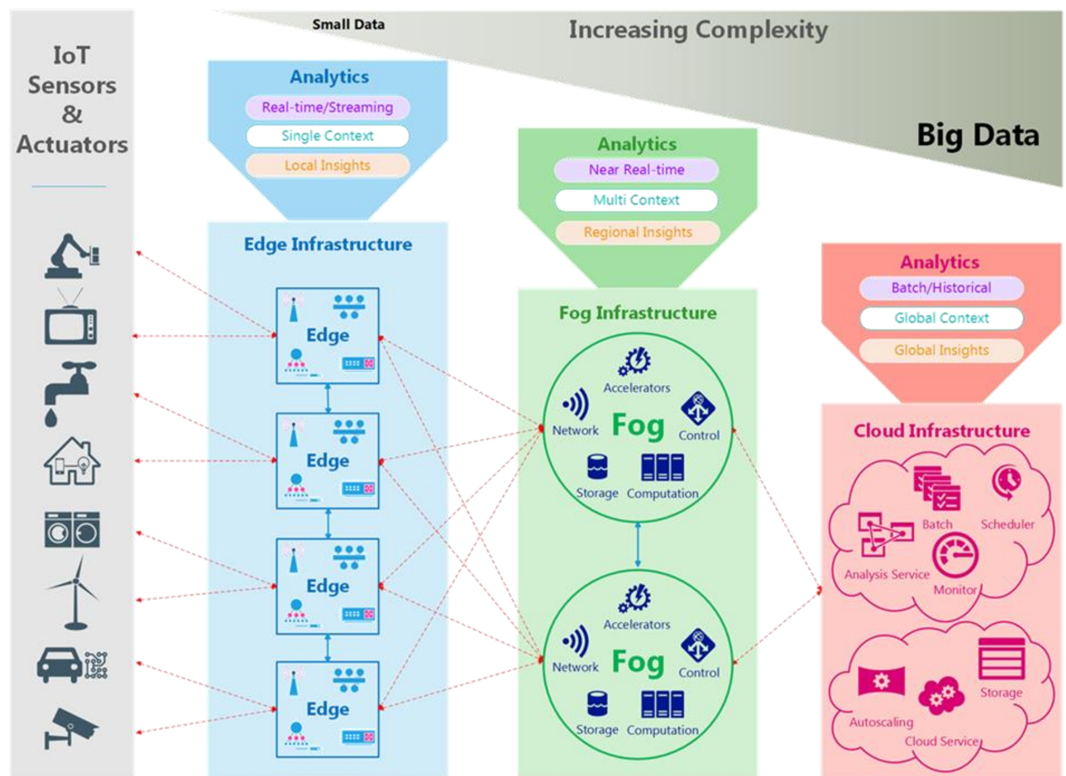
Analytical Capability



Main Scientific Contribution

Analytics Everywhere Framework





Research Questions

RQ1: How can we combine analytical algorithms without compromising the accuracy of the overall analytical workflow?

RQ2: What is the feasibility of distributing the analytics considering that edge and fog computing have less processing power than the cloud?

RQ3: What are the benefits and limitations of the proposed analytics everywhere framework?

Research Objectives

RQ1: How can we combine analytical algorithms without compromising the accuracy of the overall analytical workflow?

- Develop and implement an analytical workflow and their corresponding sequence of tasks.
- Select and combine different algorithms to execute the same analytical task running at different resources.
- Select one algorithm to execute the same analytical task running at different resources.

Research Objectives

RQ2: What is the feasibility of distributing the analytics considering that edge and fog computing have less processing power than the cloud?

- Develop a set of metrics to provide feedback about the optimal performance of the framework
- Develop a set of metrics for evaluating the real-time, near real-time, and batch processing time serving different IoT use cases.

Research Objectives

RQ3: What are the benefits and limitations of the proposed analytics everywhere framework?

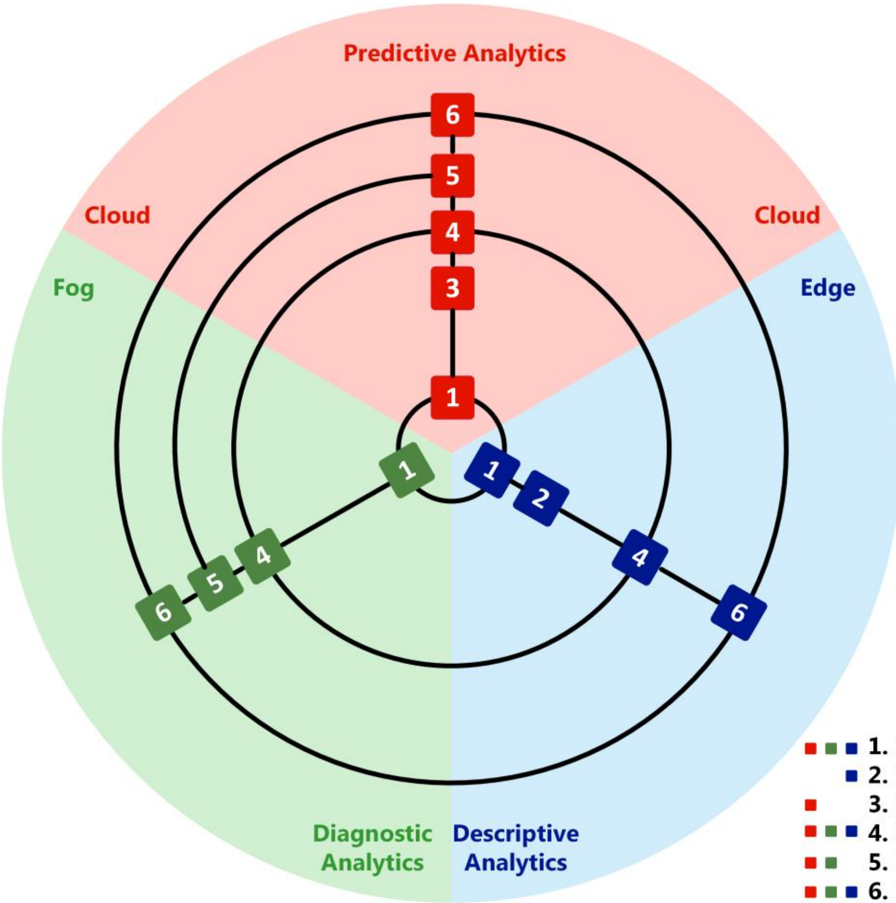
- Identify the balance point of the framework aiming to reduce the unpredictable network latency, saving band-width, offloading the resource, and handling the mobility and geo-distribution of IoT devices and mobile edge nodes.
- Study the advantages and disadvantages of combining a particular analytics (e.g. descriptive analytics) with other analytics (e.g. diagnostic and predictive analytics) by analyzing the new insights at the local, regional and global scales.

Methodology

- **Streaming analytical workflow**: consists of a set of automated tasks designed to process the data streams from the IoT devices.
- **Analytical algorithms**: are used to execute the automated tasks. They include data mining, machine learning and statistical techniques.
- **Distributed resource architecture**: is designed to provide seamlessly access to computing resources available at the edge and fog nodes, as well as the cloud clusters.

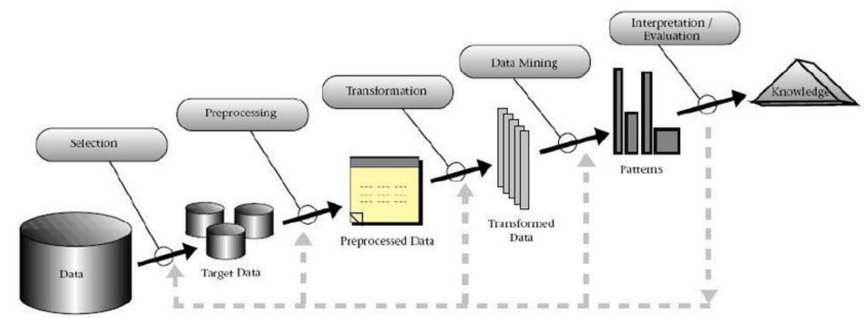
Streaming Analytical Workflow

Streaming Analytical Workflow



1. Data Ingestion
2. Data Cleaning
3. Data Summarization
4. Data Contextualization
5. Data Transformation
6. Data Analytics

KDD Workflow

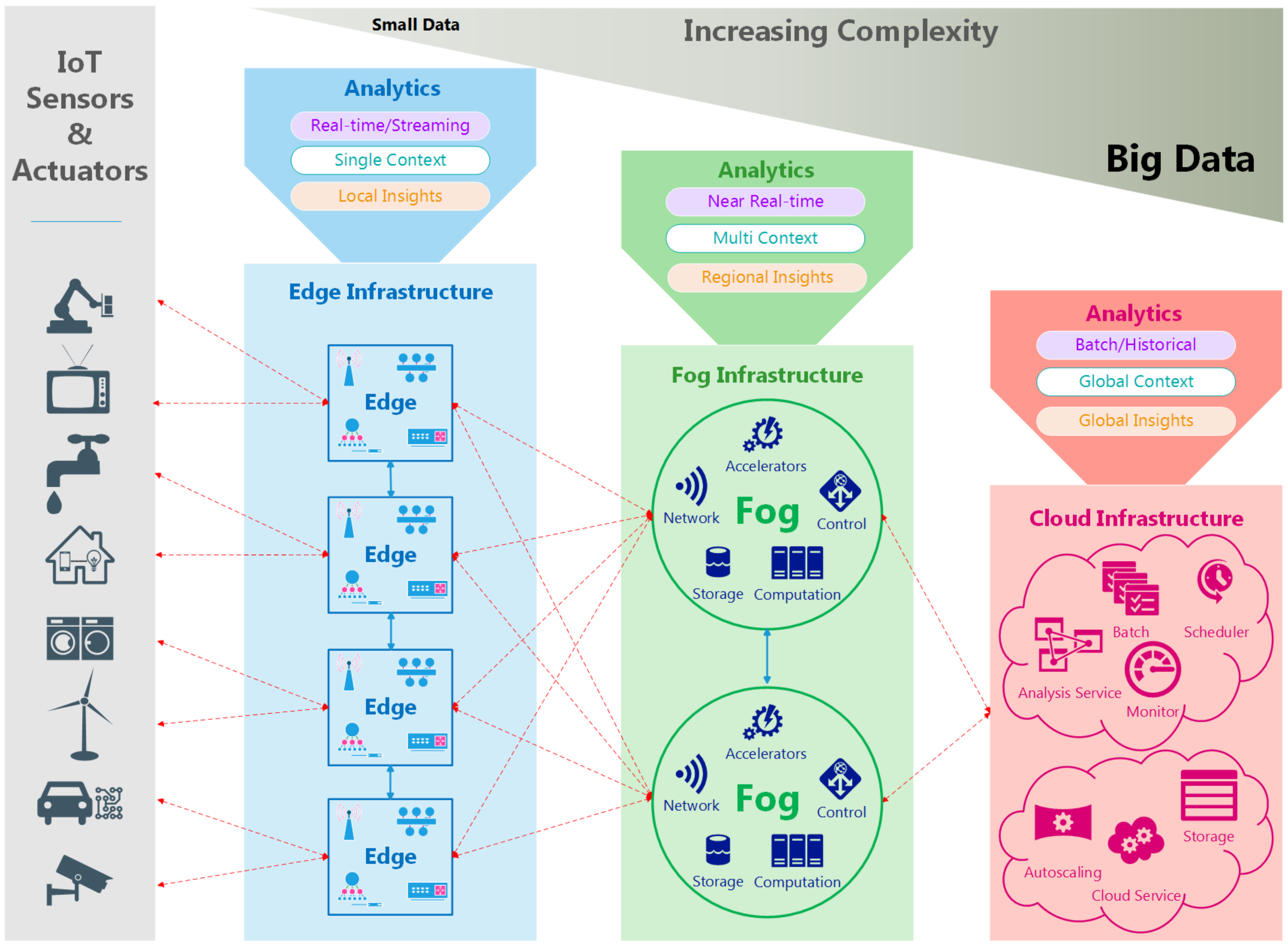


An overview of the steps that compose the knowledge discovery in databases (Fayyad et al. 1996)

Analytical Algorithms

- Statistic
- SVM
- DBSCAN
- HMM
- Random Forest
- Neural Network

Distributed resource architecture

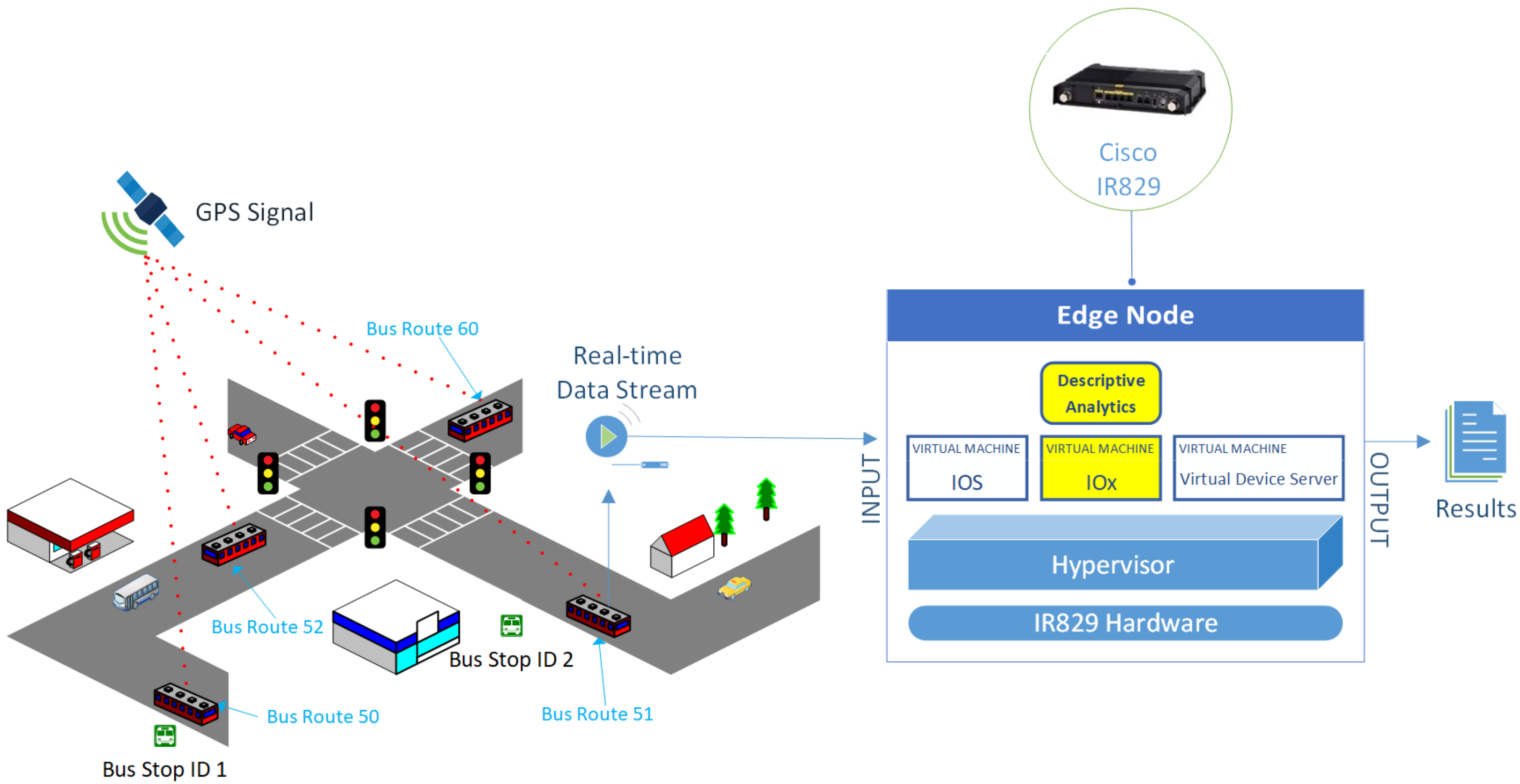


IoT Use Cases

IoT Case Studies	Analytics	Techniques	Applications
Intelligent Fleet Management	Descriptive	- Statistic	- Schedule adherence - Abnormalities detection.
	Diagnostic	- DBSCAN - SVM	- Aggressive driving - Fuel Inefficiency driving
	Predictive	- Random Forest - HMM - Neural Network	- Predicted arrivals at each station - Origin-Destination predictions - Predicted load for each trip
Smart Parking	Descriptive	- Statistic	- Parking density usage - Real-time parking spot reservation
	Diagnostic	- Clustering	- Decrease parking search - Dynamic pricing strategy
	Predictive	- Random Forest - Neural Network	- Predicted available parking - Predicted parking demand

Preliminary Results

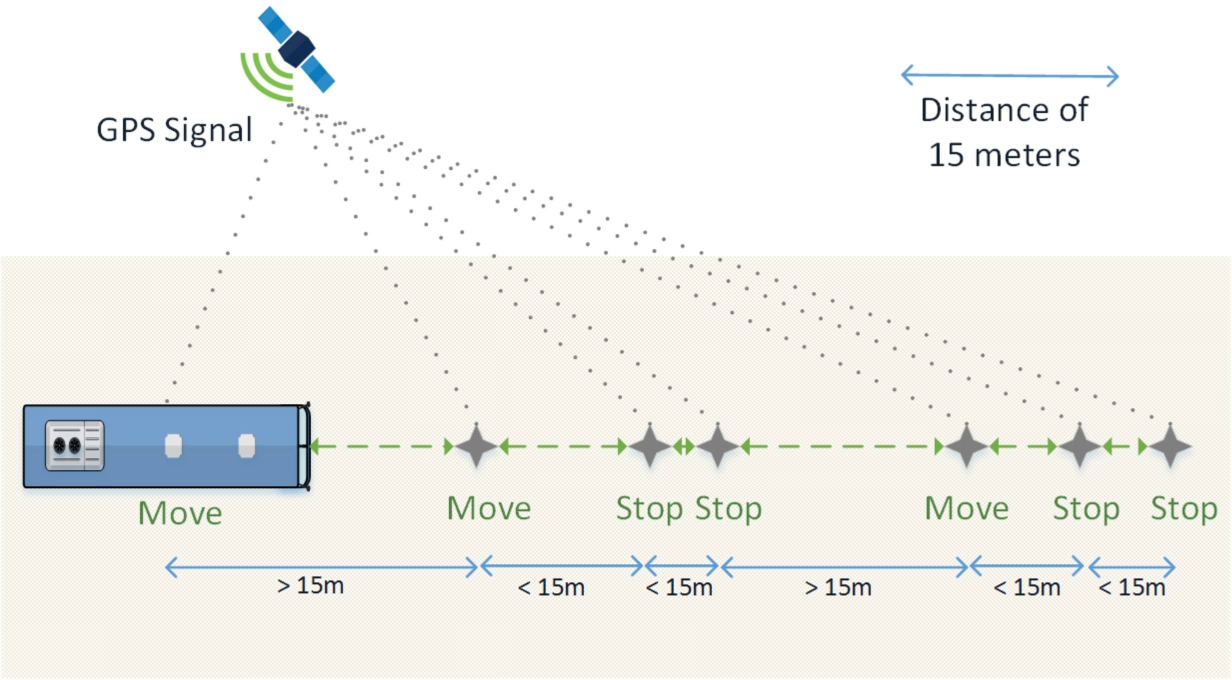
Intelligent Fleet Management at the Edge



Cao, H., Wachowicz, M., & Cha, S. (2017, December). **Developing an edge computing platform for real-time descriptive analytics.** In *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 4546-4554). IEEE.

Preliminary Results

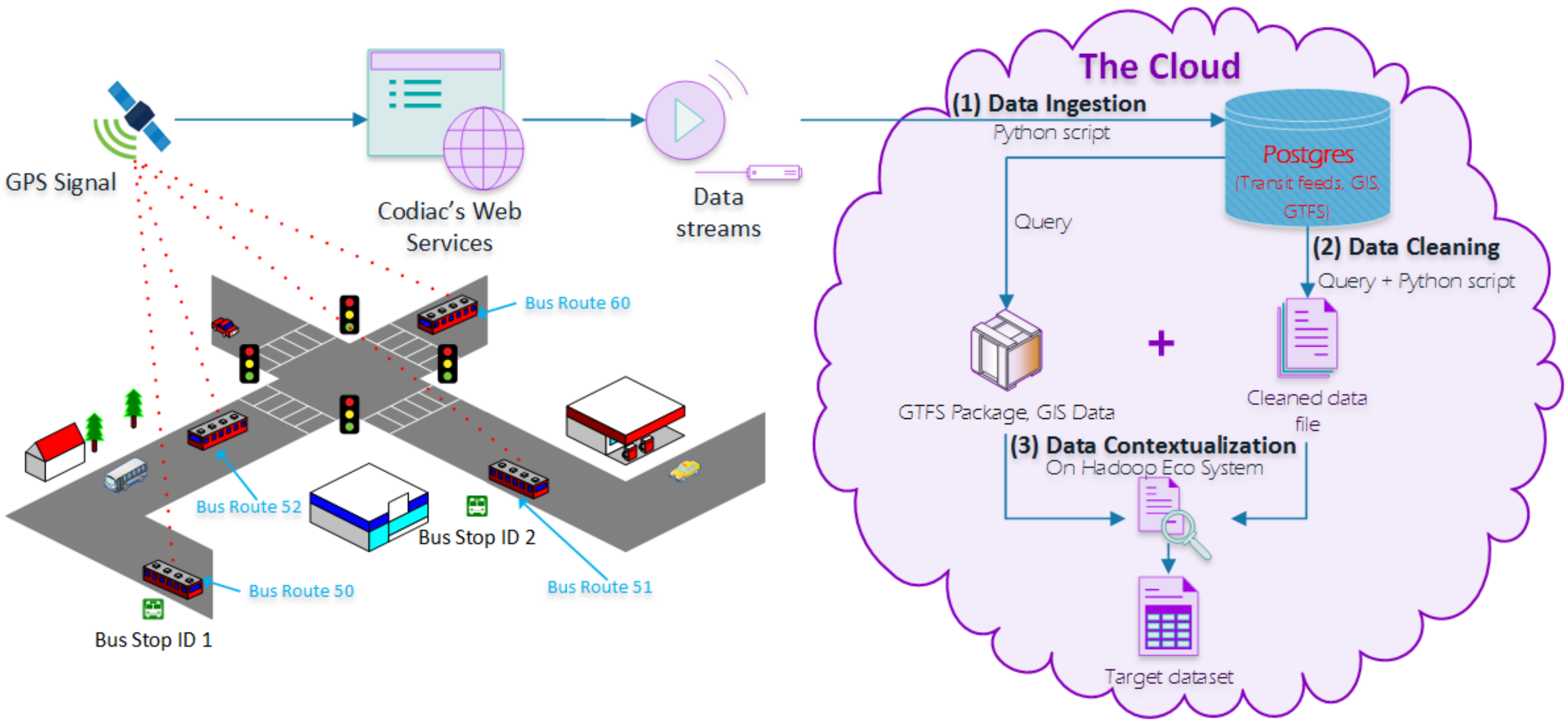
Descriptive Analytics at the Edge



Cao, H., Wachowicz, M., & Cha, S. (2017, December). **Developing an edge computing platform for real-time descriptive analytics.** In *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 4546-4554). IEEE.

IoT Use Cases

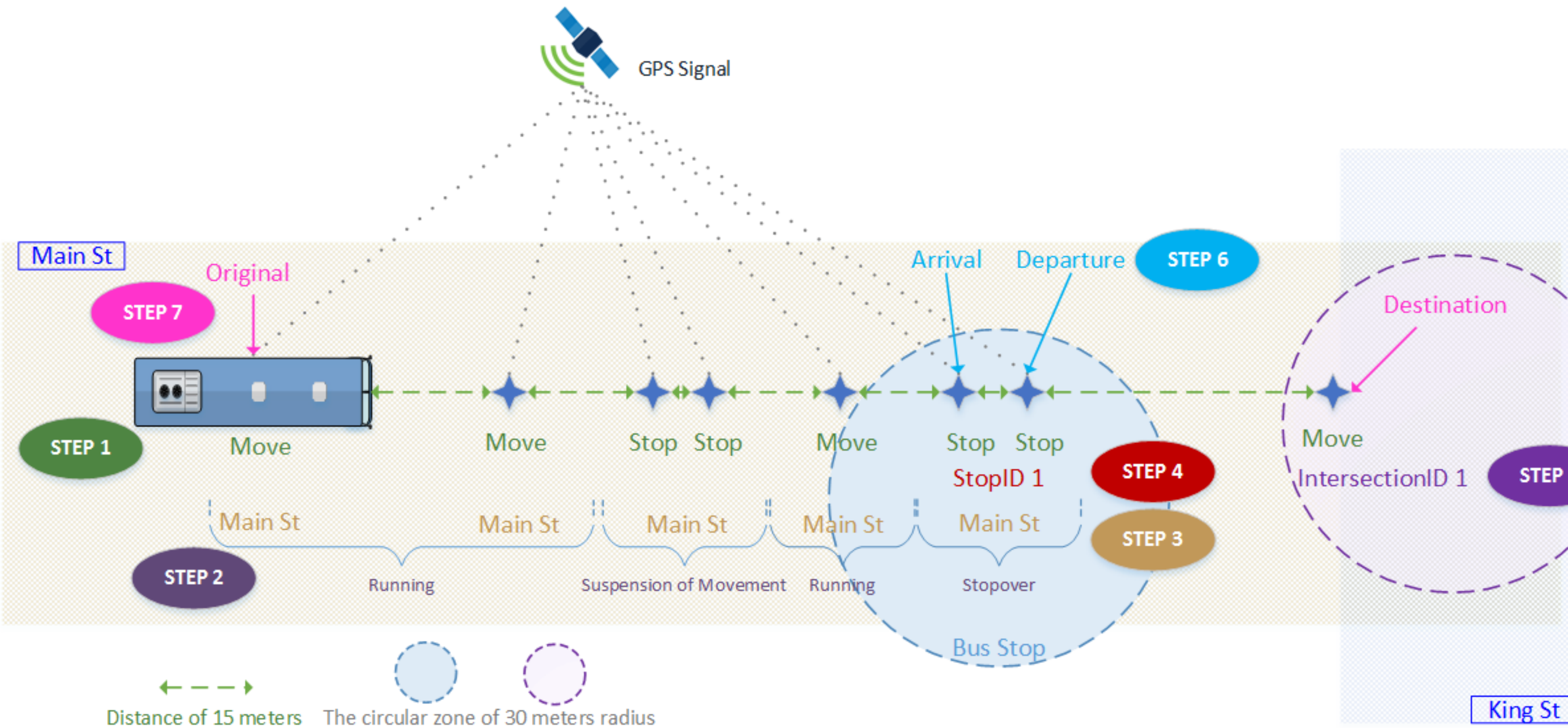
Intelligent Fleet Management in the Cloud



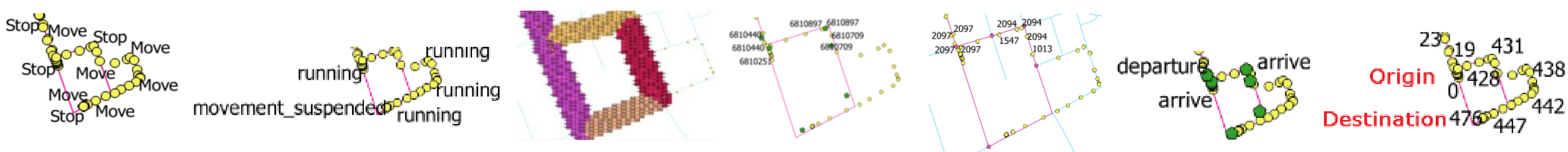
Cao, H., & Wachowicz, M. (2018). **The design of an automated analytical workflow for processing mobility contexts.** (Under submission)

Preliminary Results

Predictive Analytics in the Cloud



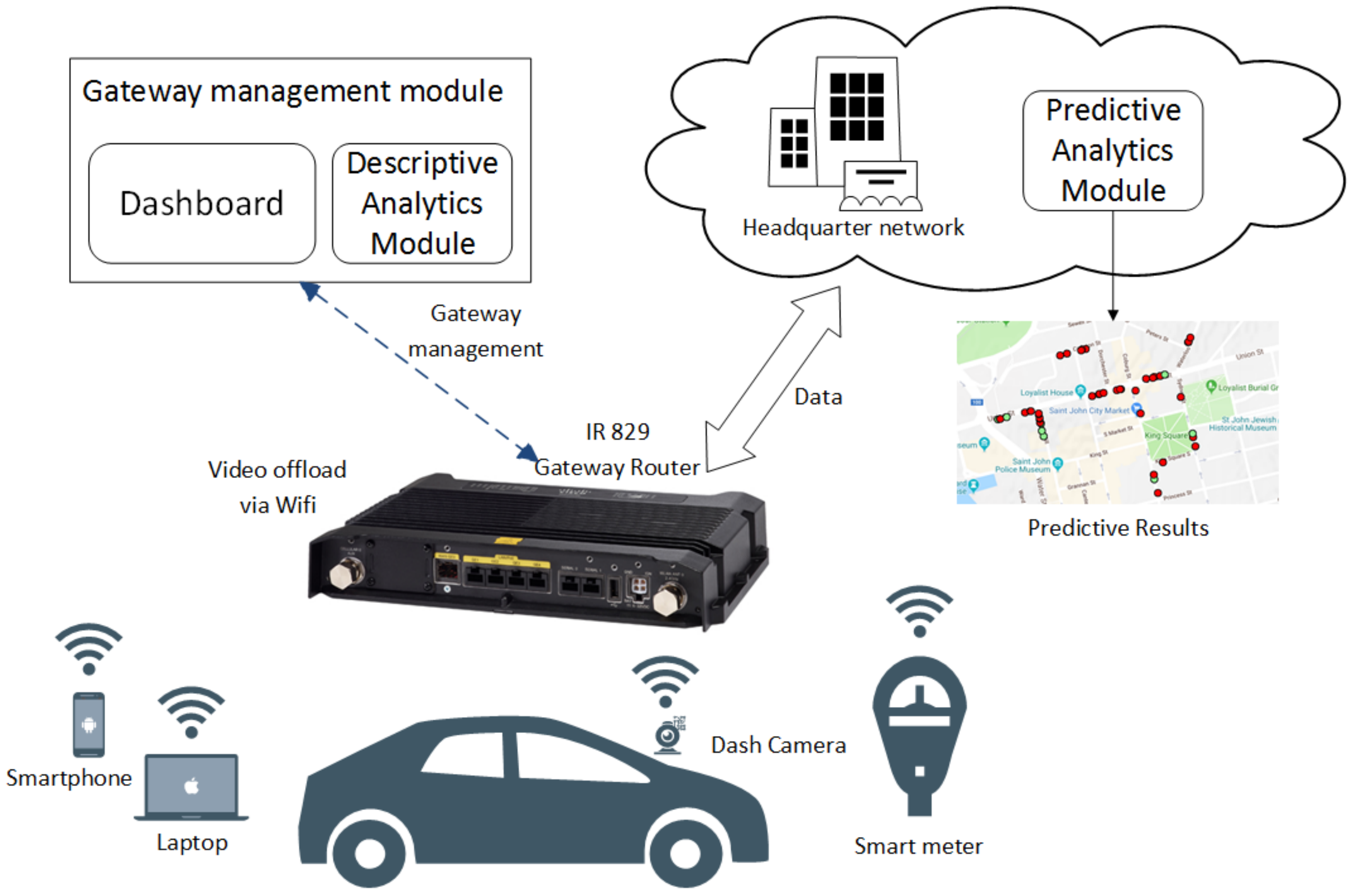
Step 1 **Step 2** **Step 3** **Step 4** **Step 5** **Step 6** **Step 7**



Cao, H., Wachowicz, M., Renso, C., & Carlini, E. (2018) **An edge-fog-cloud platform for anticipatory learning process designed for Internet of Mobile Things.** (Under submission)

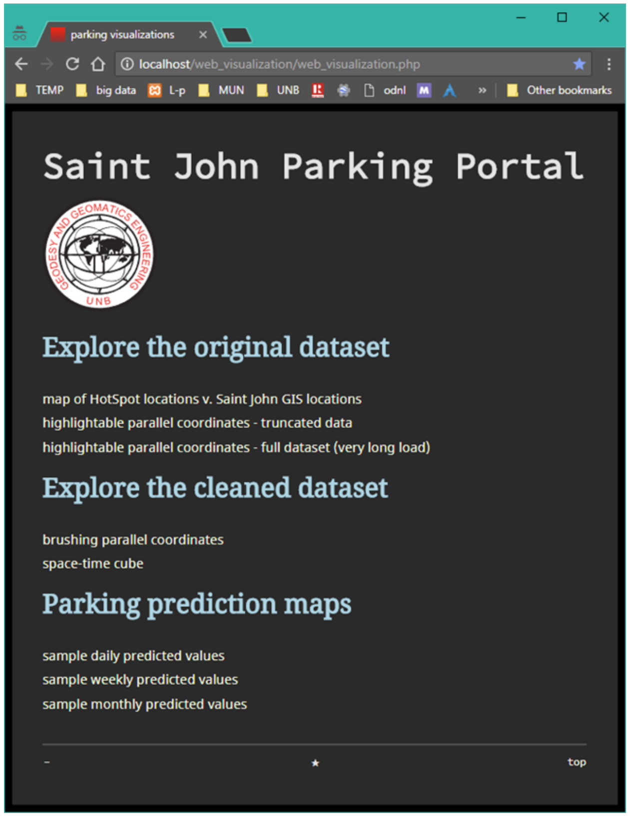
IoT Use Cases

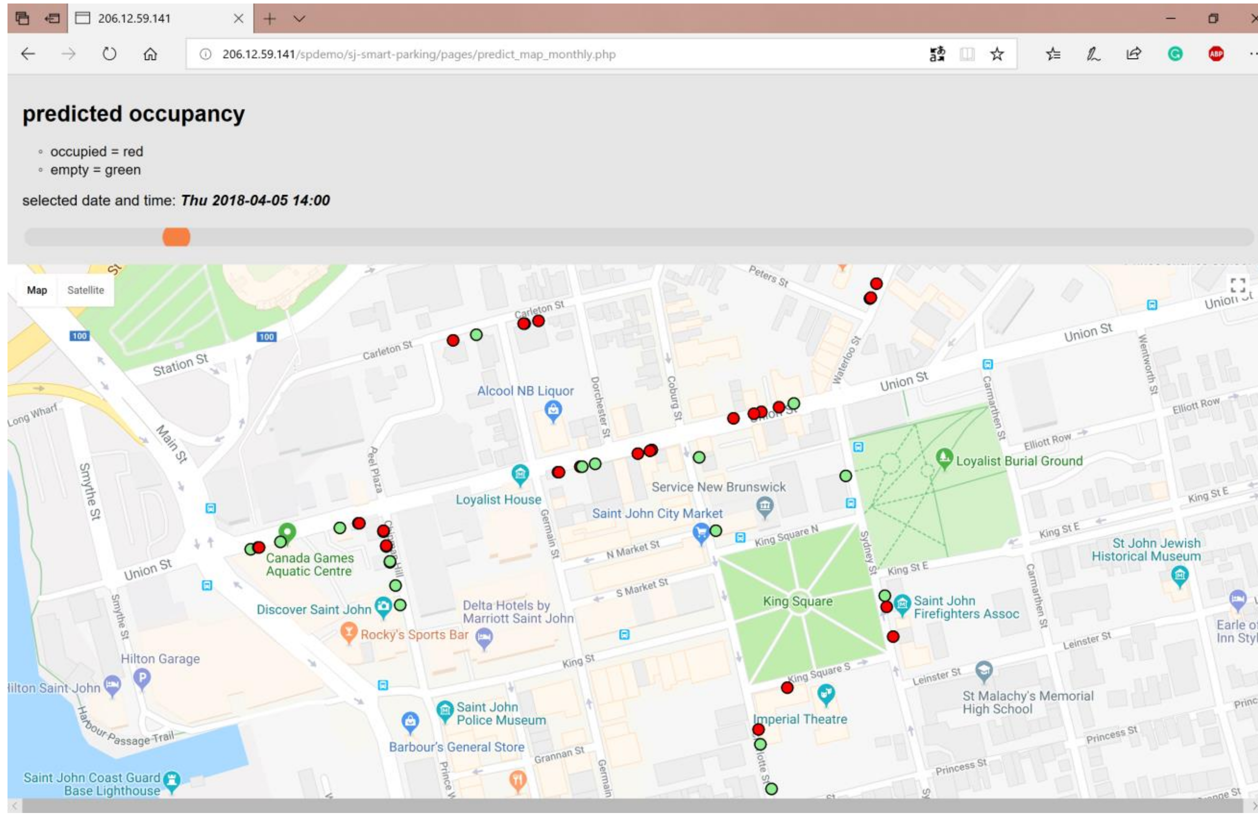
Smart Parking



IoT Use Cases

Smart Parking





Demo: http://206.12.59.141/spdemo/sj-smart-parking/web_visualization.php

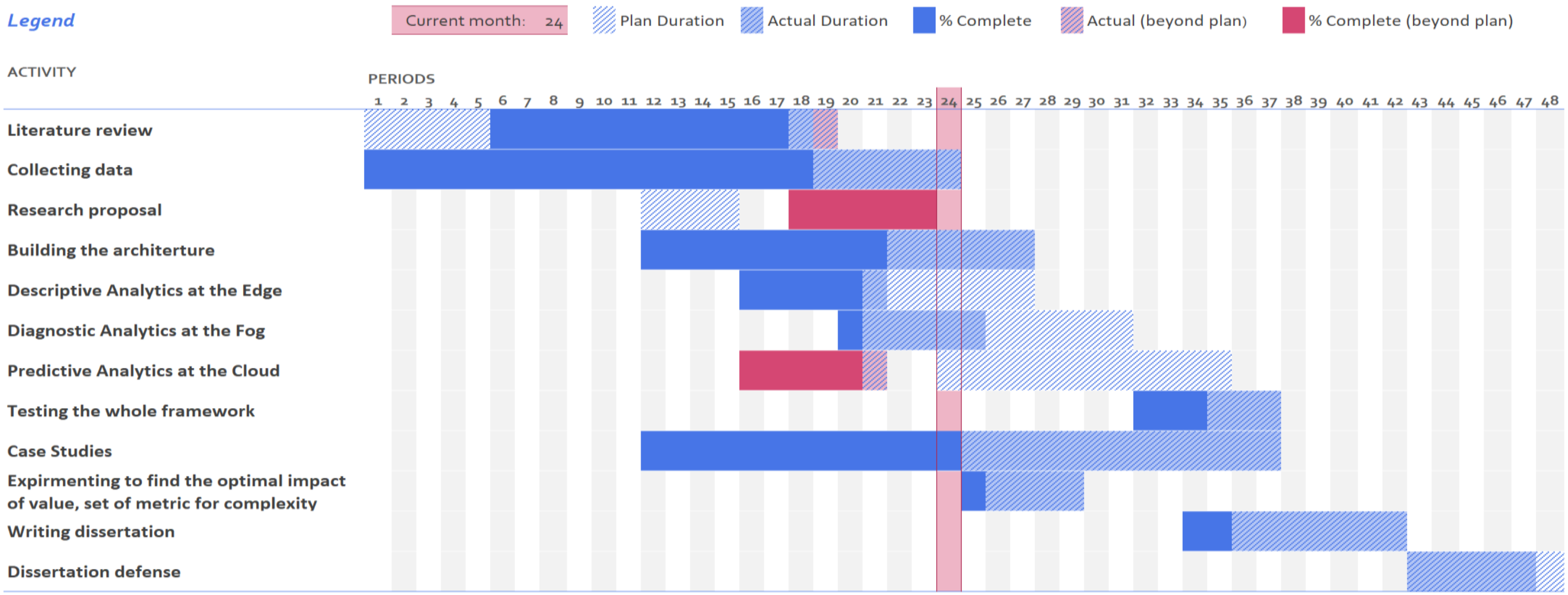
Expected Results and Significance

- Design of a flexible end-to-end architecture in conjunction with a streaming analytical workflow.
- Develop a unique “analytics everywhere” framework for different IoT Use Cases.
- Scientific evidence to confirm the research hypothesis that combining edge-fog-cloud resources will bring more effective analytics for IoT systems.

Resources Required

- Datasets:
  - Codiac Transit Data Streams
  - Hotspots smart parking data
  - Rimots telecommunication data
  - City of Ottawas transit data
  - Open datasets
- Hardware:
  - Edge node: IR 809/829, IE 4000, ISR 4000
  - Fog node: Cisco UCS C220, Cisco UCS C240
  - Cloud: WestCloud, EastCloud provided by Compute Canada
- Software:
  - Cisco Kinetic
  - Database: PostgreSQL, MongoDB, RethinkDB, Neo4j, BigchainDB, Hbase, Cassandra
  - Centralized cluster data management platform: Hortonwork Data Platform

Time Schedule



Activity	Plan Start	Plan Duration	Actual Start	Actual Duration	Complete
Literature review	1	18	6	14	90%
Collecting data	1	24	1	24	75%
Research proposal	12	4	18	6	100%
Building the architerture	12	16	12	16	66%
Descriptive Analytics at the Edge	16	12	16	6	85%
Diagnostic Analytics at the Fog	20	12	20	6	10%
Predictive Analytics at the Cloud	24	12	16	6	90%
Testing the whole framework	32	6	32	6	50%
Case Studies	12	26	12	26	50%
Experimenting to find the optimal impact of value, set of metric for complexity	25	5	25	5	10%
Writing dissertation	34	9	34	9	30%
Dissertation defense	43	6	43	5	0%

I hope that my ***cool*** proposal of an

“analytics everywhere”

framework

will help you ***soothe*** the ***heat*** of this summer.

Thank you – Questions & Answers