Developing an Analytics Everywhere framework for the Internet of Things in Smart City Applications

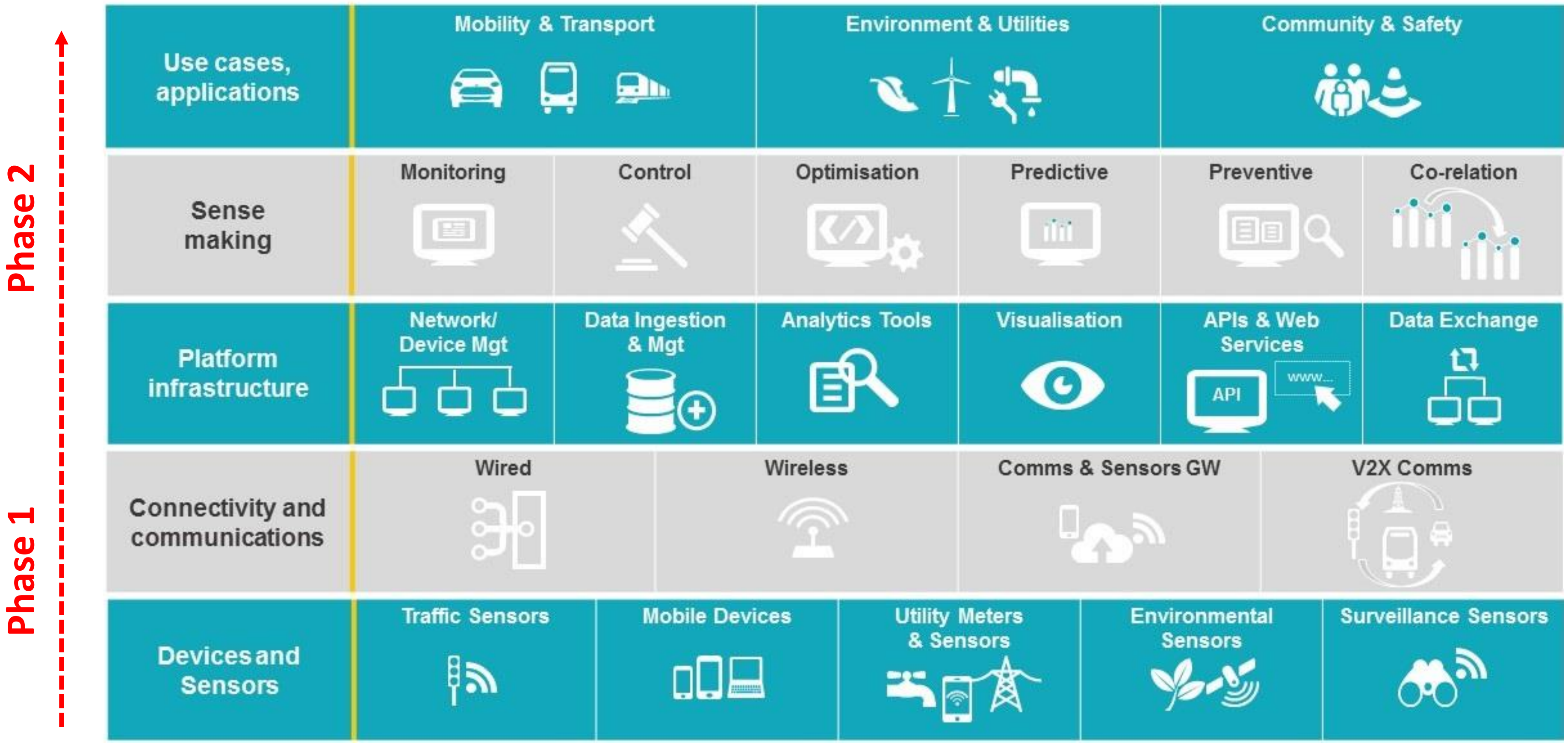
Hung Cao

Outline

- Background/Research Motivation
- Research Questions & Objectives
- Methodology
- Implementation
- Applications & Results
- Scientific Contributions
- Recommendations for Future Work

Current Research in IoT

Figure: https://www.beca.com/



Al-Fuqaha, Guizani, Mohammadi, Aledhari, & Ayyash (2015); Li, Da Xu, & Zhao (2015); Ngu, Gutierrez, Metsis, Nepal, & Sheng (2017); Gazis (2017)

Current Research in IoT Computing

- Focus on processing & analyzing IoT data streams using one computing resource:
  - Cloud Computing
  - Fog Computing
  - Edge Computing

Current Research in IoT Geomatics

- IoT-GIS is still in its infancy.
- No GIS can handle IoT stream data
- IoT-GIS platforms are needed for contextualizing IoT data streams that can significantly improve the results of analytical workflows.

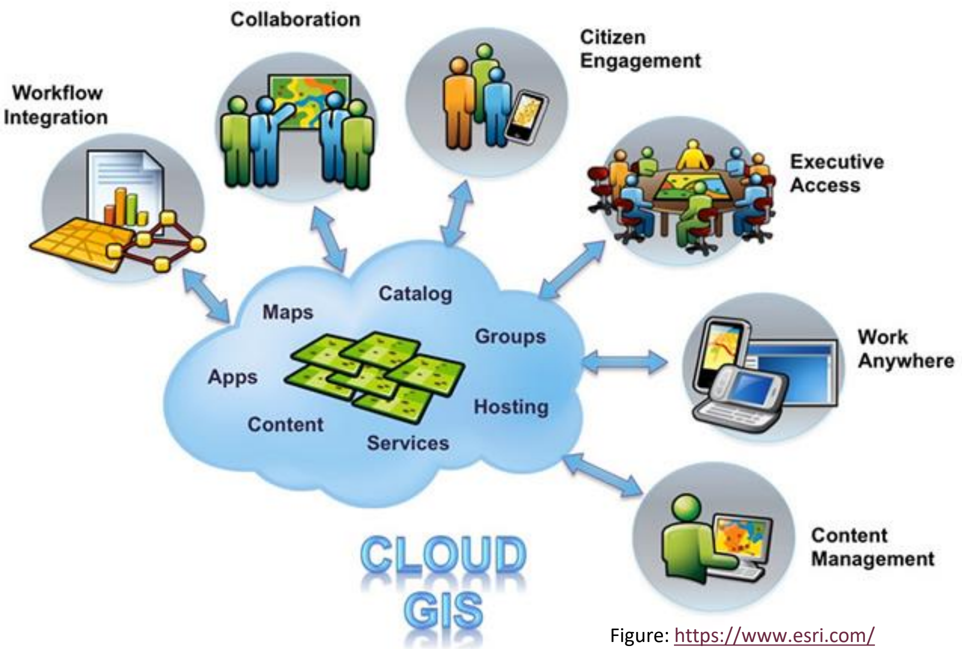


Figure: https://www.esri.com/

Research Challenges

Figure: https://ecf.com/

- Mobility and co-location of IoT devices in smart cities will require
  - avoiding bottlenecks
  - reducing data latency

- Vast amount of IoT data streams will require
  - streaming automated analytical tasks
  - manage, process, and retrieve high velocity, variety, and volumes of data.

Overall Research Goal

- Develop an *"Analytics Everywhere"* framework to explore the edge-fog-cloud continuum for performing automated analytical tasks capable of:
  - providing higher-level intelligence from continuous IoT data streams
  - generating long-term predictions from accumulated IoT data streams.

RQ1: How can **automated analytical tasks** be developed for supporting **analytical capabilities** such as **streaming descriptive/diagnostics/predictive algorithms/methods**?

- Identify the algorithms/methods that can be used for supporting automated streaming analytical tasks (analytical capability).
- Identify the potential IoT applications in smart cities while taking into account the availability of IoT data streams.

RQ2: How can **continuous** and **accumulated streams** be combined while taking into account different IoT **data life-cycles**?

- Develop data life-cycles for executing automated analytical tasks and coping with continuous IoT data streams.
- Develop data life-cycles for executing automated analytical tasks and coping with accumulated IoT data streams.
- Implement the data life-cycles for the real-world experiments.

RQ3: How can **IoT** and **GIS** be integrated into the **edge-fog-cloud continuum** without compromising **resource capabilities**?

- Identify the off-the-shelf tools that can be used to implement the proposed framework (computational resources).
- Build the architecture for this new framework based on a continuum of edge-fog-cloud nodes.

RQ4: What are the **benefits** and **limitations** of the proposed **"Analytics Everywhere" framework** in smart cities?

- Implement the "Analytics Everywhere" framework in real-world experiments to evaluate the proposed approach. Two experiments were selected according to data availability: smart transit and smart parking.
- Validate the proposed "Analytics Everywhere" framework.

Methodology

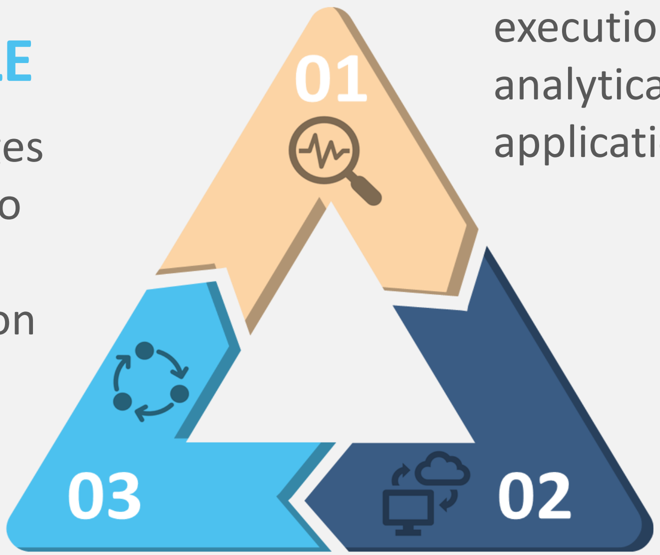
Analytics Everywhere Framework

ANALYTICAL CAPABILITY

Consists of algorithms for the execution of a network of analytical tasks for IoT applications

DATA LIFECYCLE

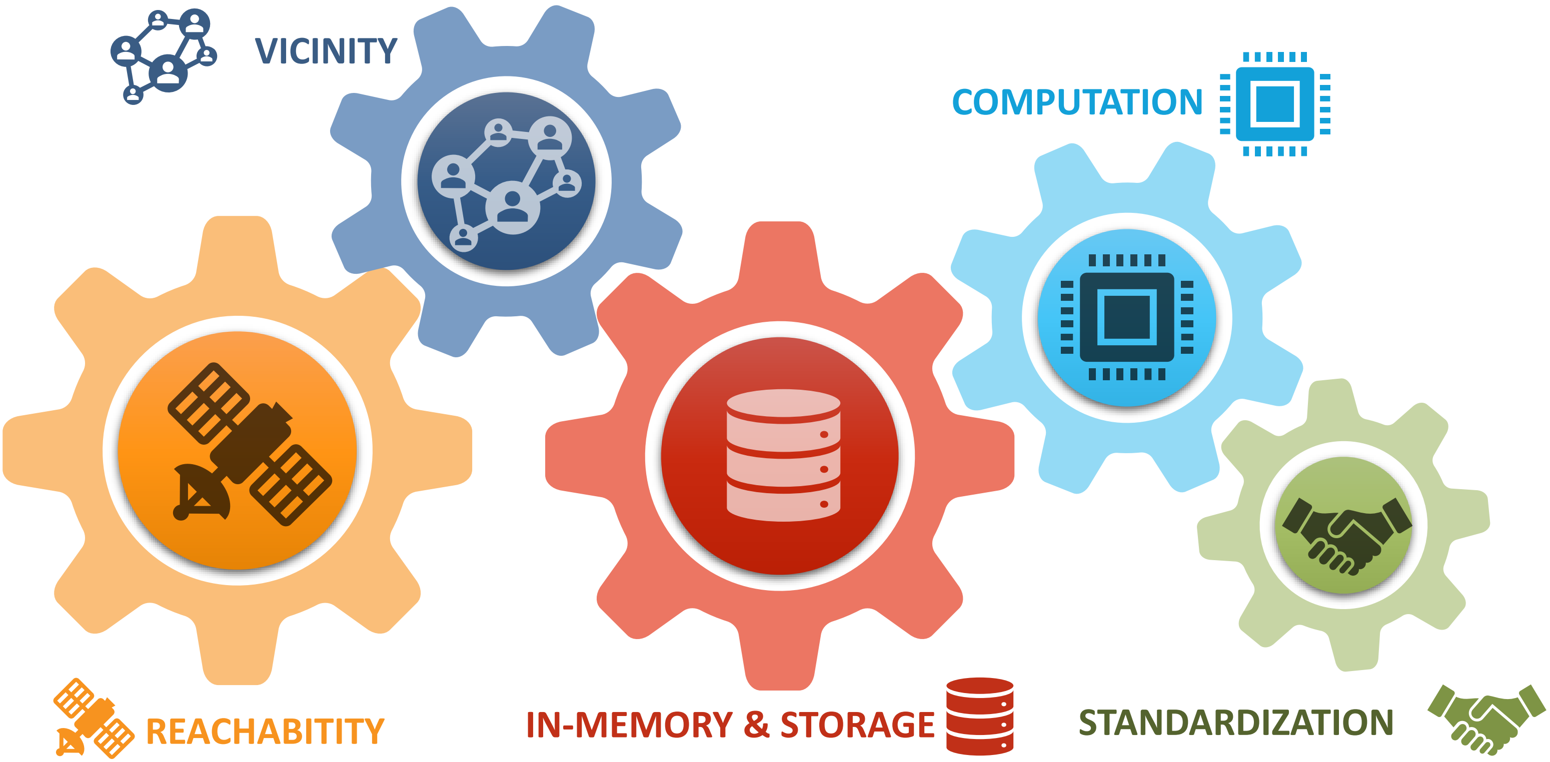
Manages the changes that data streams go through during the automated execution of a network of analytical tasks



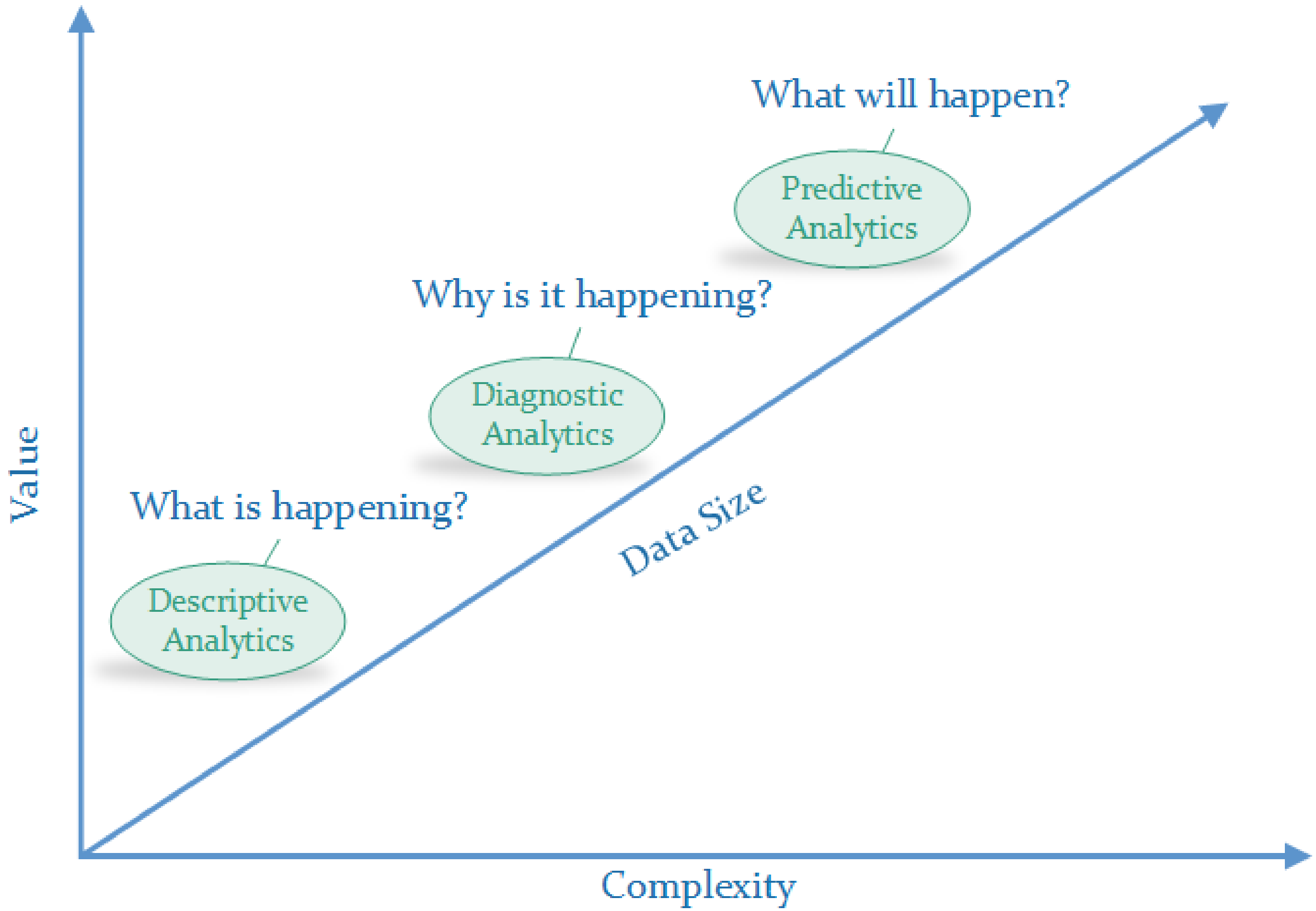
RESOURCE CAPABILITY

Consists of distributed compute nodes (i.e. cloud, fog, and edge nodes) that provide I/O, storage, computation and processing power for the execution of a network of analytical tasks

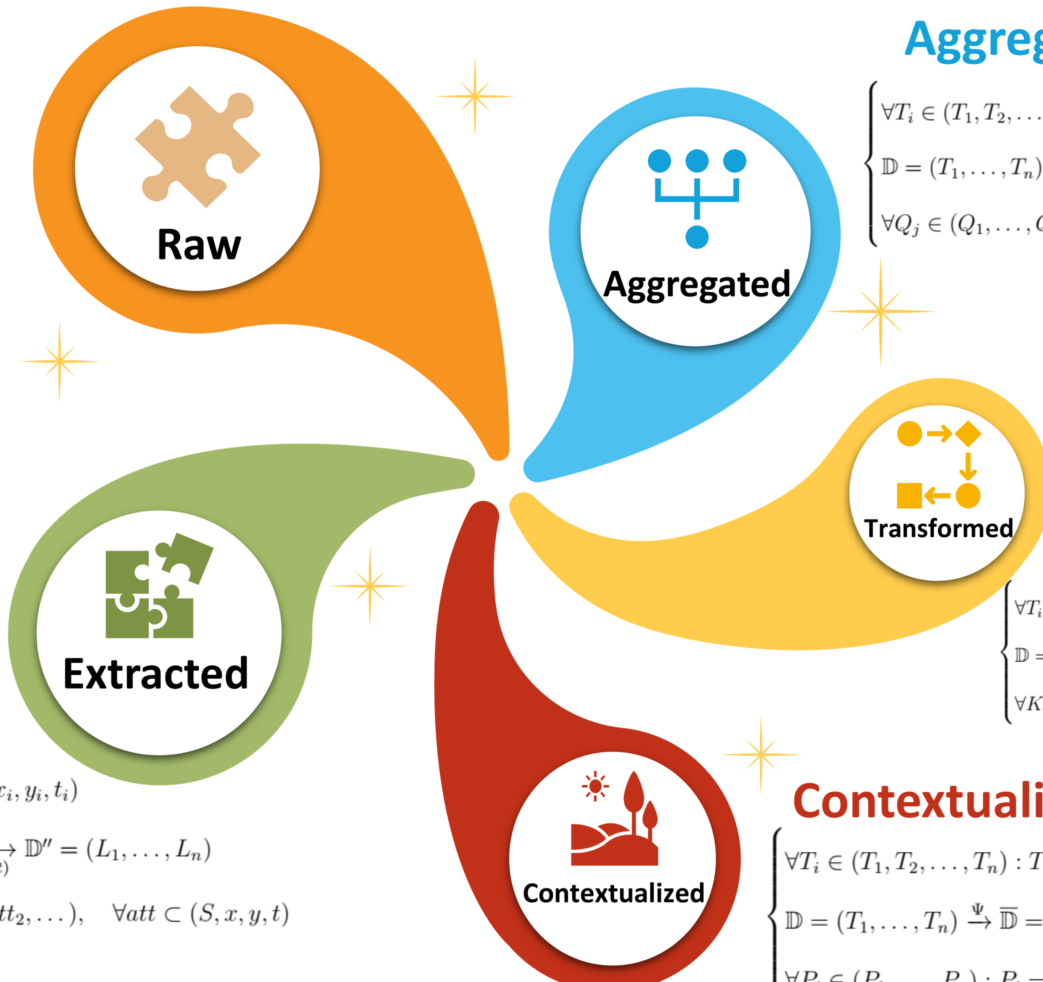
Resource Capability



Analytical Capability



Data lifecycle



Raw Data

$$T_i = (S_i, x_i, y_i, t_i)$$

Aggregated Data

$$\begin{cases} \forall T_i \in (T_1, T_2, \dots, T_n) : T_i = (S_i, x_i, y_i, t_i) \\ \mathbb{D} = (T_1, \dots, T_n) \xrightarrow[\text{on attribute } S]{\Phi} \widehat{\mathbb{D}} = (Q_1, \dots, Q_m) \\ \forall Q_j \in (Q_1, \dots, Q_m) : Q_j = (Agg\_value_1, Agg\_value_2, \dots) \end{cases}$$

Transformed Data

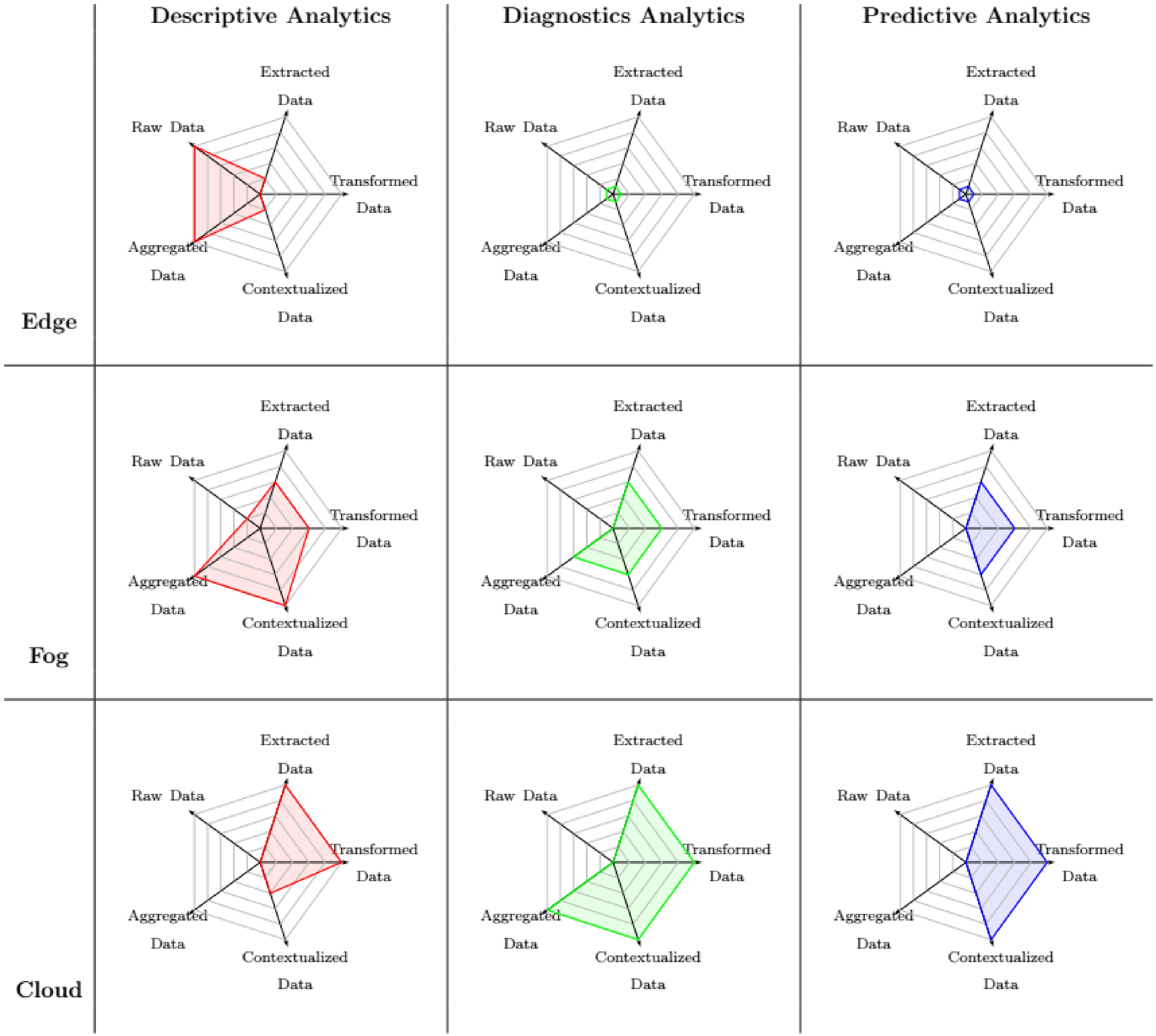
$$\begin{cases} \forall T_i \in (T_1, T_2, \dots, T_n) : T_i = (S_i, x_i, y_i, t_i) \\ \mathbb{D} = (T_1, \dots, T_n) \xrightarrow{\Upsilon} \ddot{\mathbb{D}} = (K_1, \dots, K_n) \\ \forall K_i \in (K_1, \dots, K_n) : K_i = (Trans\_value_1, Trans\_value_2, \dots) \end{cases}$$

Contextualized Data

$$\begin{cases} \forall T_i \in (T_1, T_2, \dots, T_n) : T_i = (S_i, x_i, y_i, t_i) \\ \mathbb{D} = (T_1, \dots, T_n) \xrightarrow{\Psi} \overline{\mathbb{D}} = (P_1, \dots, P_n) \\ \forall P_i \in (P_1, \dots, P_n) : P_i = (S_i, x_i, y_i, t_i, Context_1, Context_2, \dots) \end{cases}$$

Extracted Data

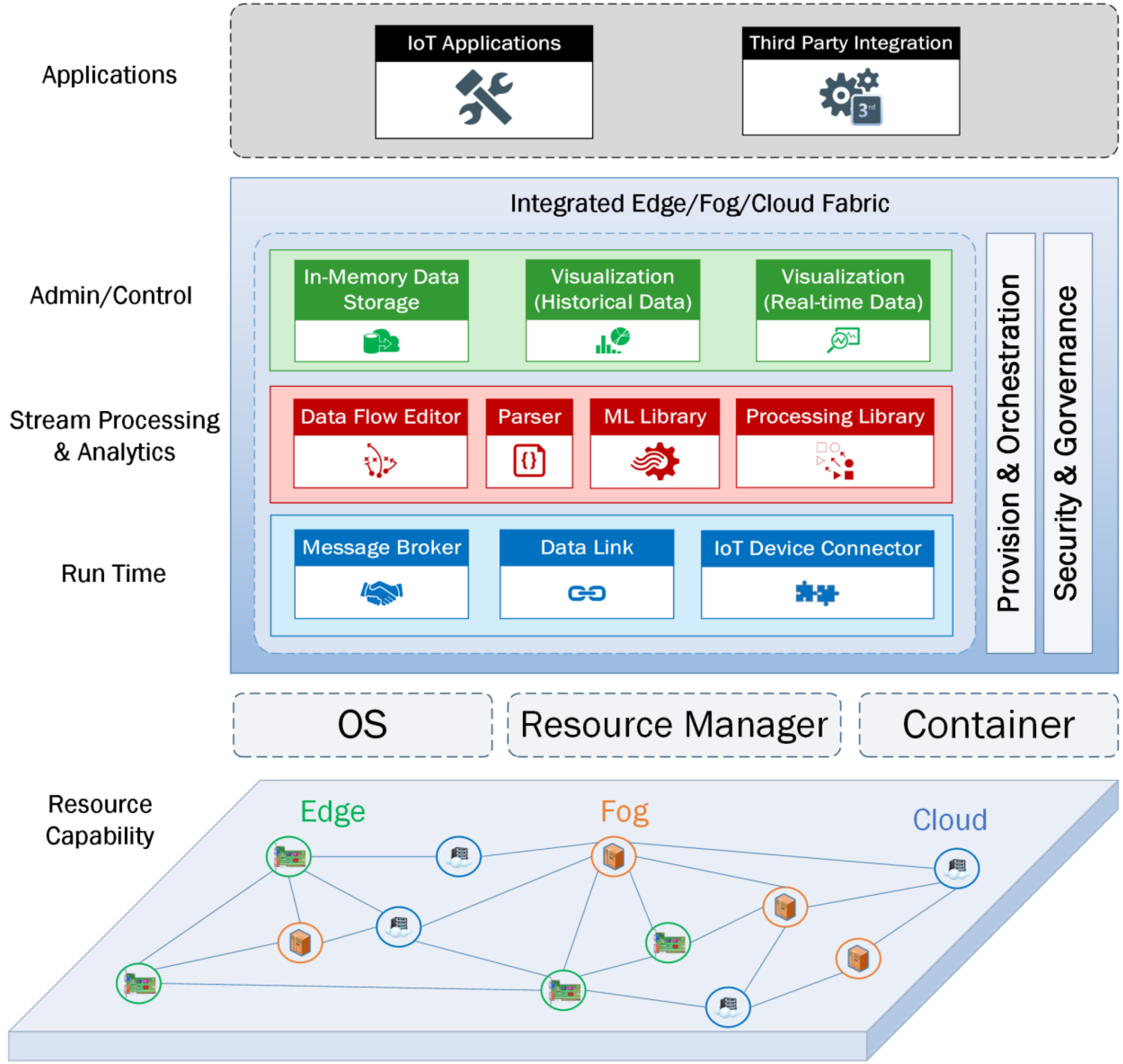
$$\begin{cases} \forall T_i \in (T_1, T_2, \dots, T_n) : T_i = (S_i, x_i, y_i, t_i) \\ \mathbb{D} = (T_1, \dots, T_n) \xrightarrow[\text{on attributes } (S|x|y|t)]{\Omega} \mathbb{D}'' = (L_1, \dots, L_n) \\ \forall L_i \in (L_1, \dots, L_n) : L_i = (att_1, att_2, \dots), \quad \forall att \subset (S, x, y, t) \end{cases}$$



Descriptive Analytics
Diagnostics Analytics
Predictive Analytics
Edge
Fog
Cloud
Extracted Data
Raw Data
Transformed Data
Aggregated Data
Contextualized Data

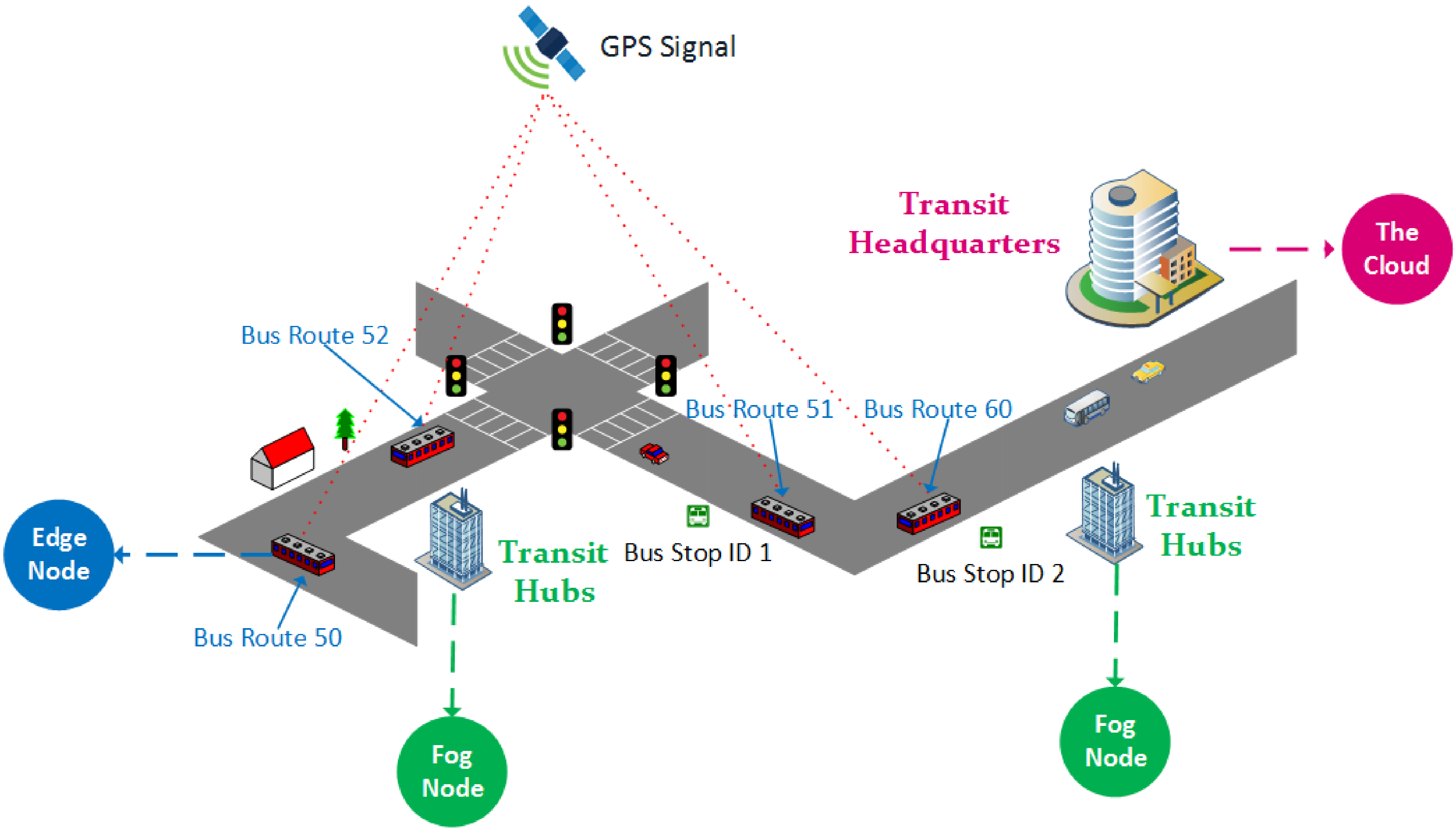
Implementation

Architecture

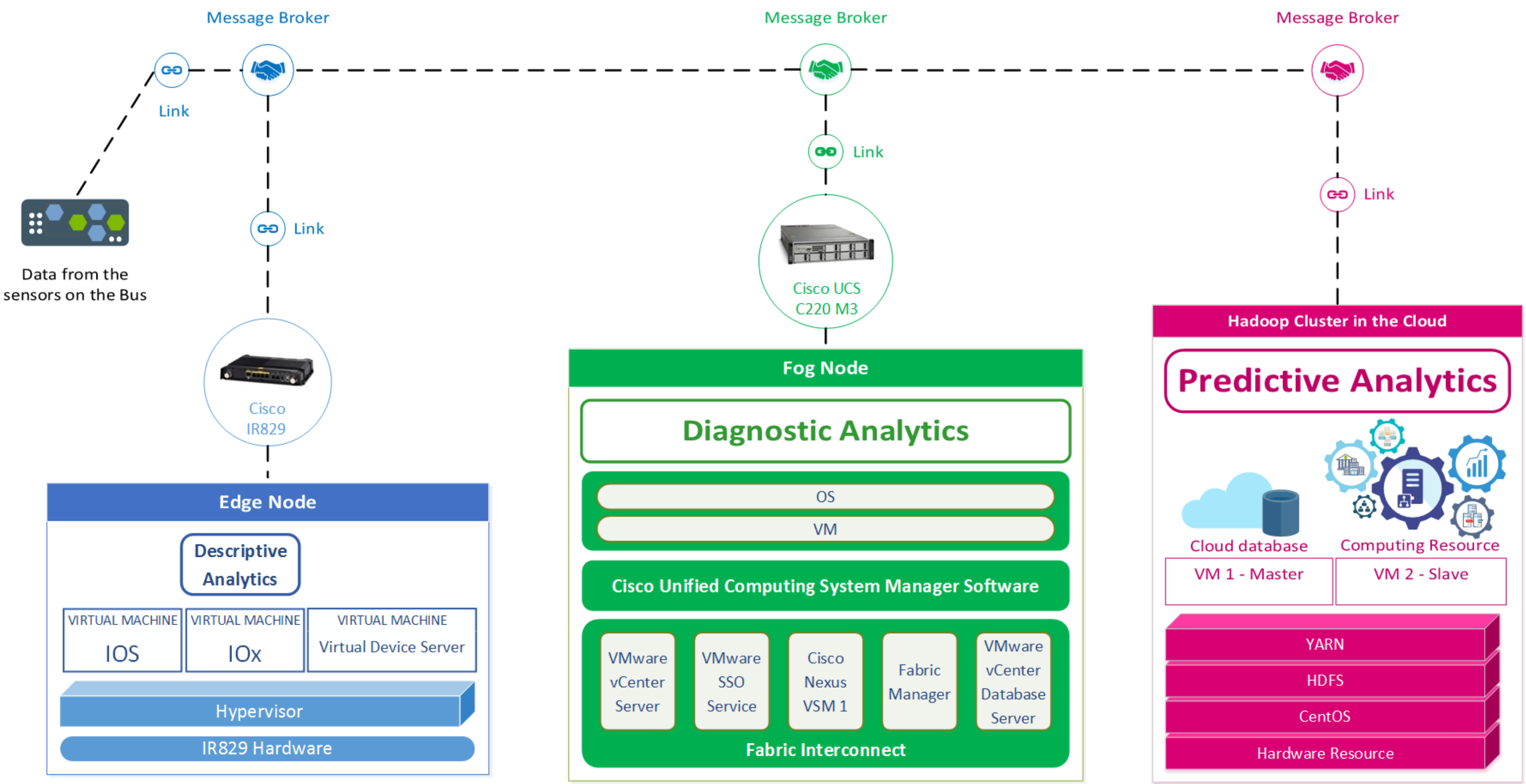


IoT Use Cases	Analytical Capability	Techniques	Applications	Target Group of Users
Smart Transit	Descriptive	Statistics	Schedule adherence	Transit Operators
	Diagnostic	Affinity Propagation Clustering	Abnormalities detection	Bus Drivers
	Predictive	Random Forest	Trip behaviors prediction	Passengers
Smart Parking	Descriptive	Statistics	Parking usage	Parking Commission
	Diagnostic	Agglomerate Hierarchical Clustering	Events/Incide nts diagnostic	Policy makers
	Predictive	Adaptive Random Forest	Empty spot prediction	Drivers

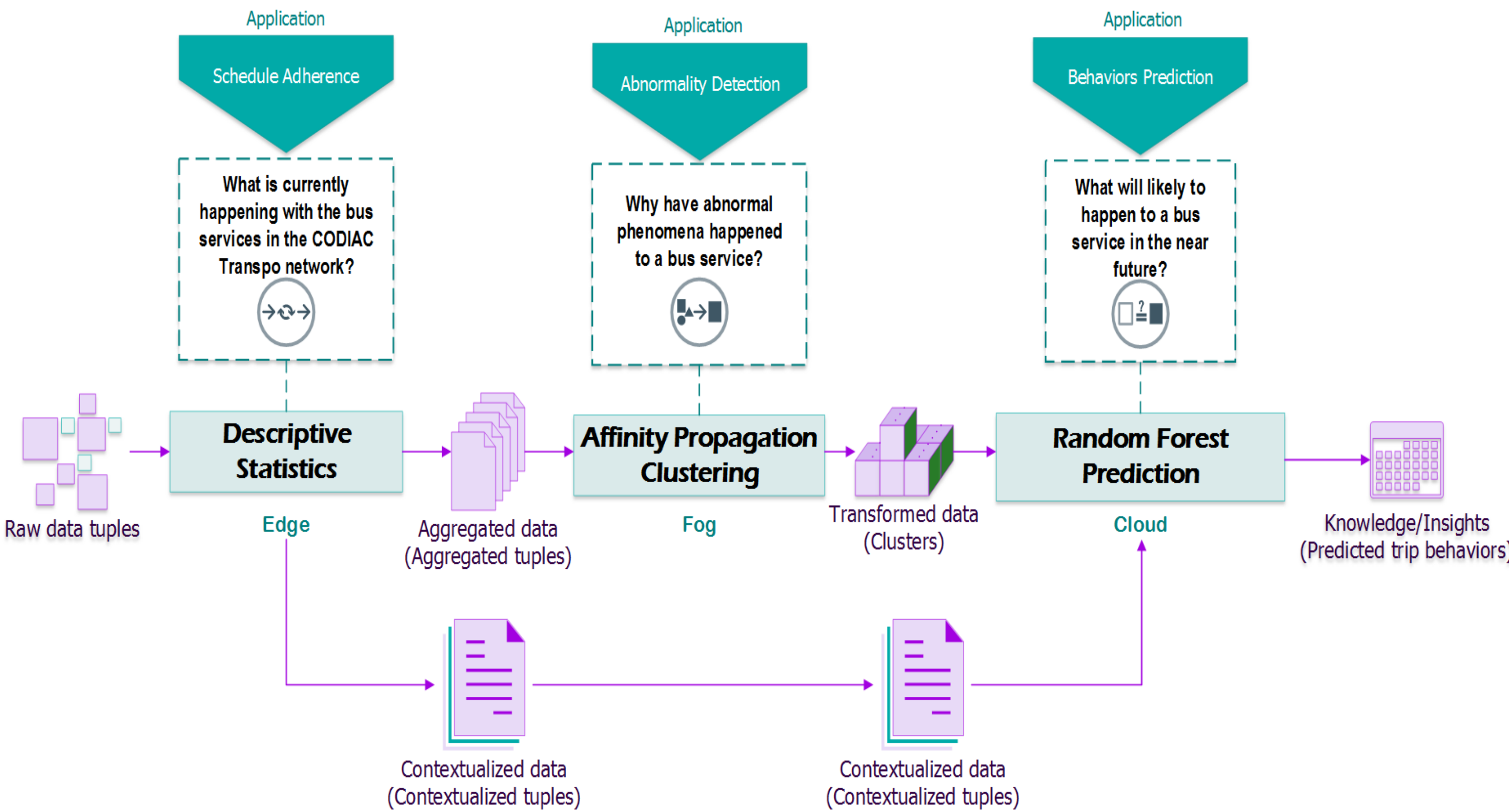
Smart Transit



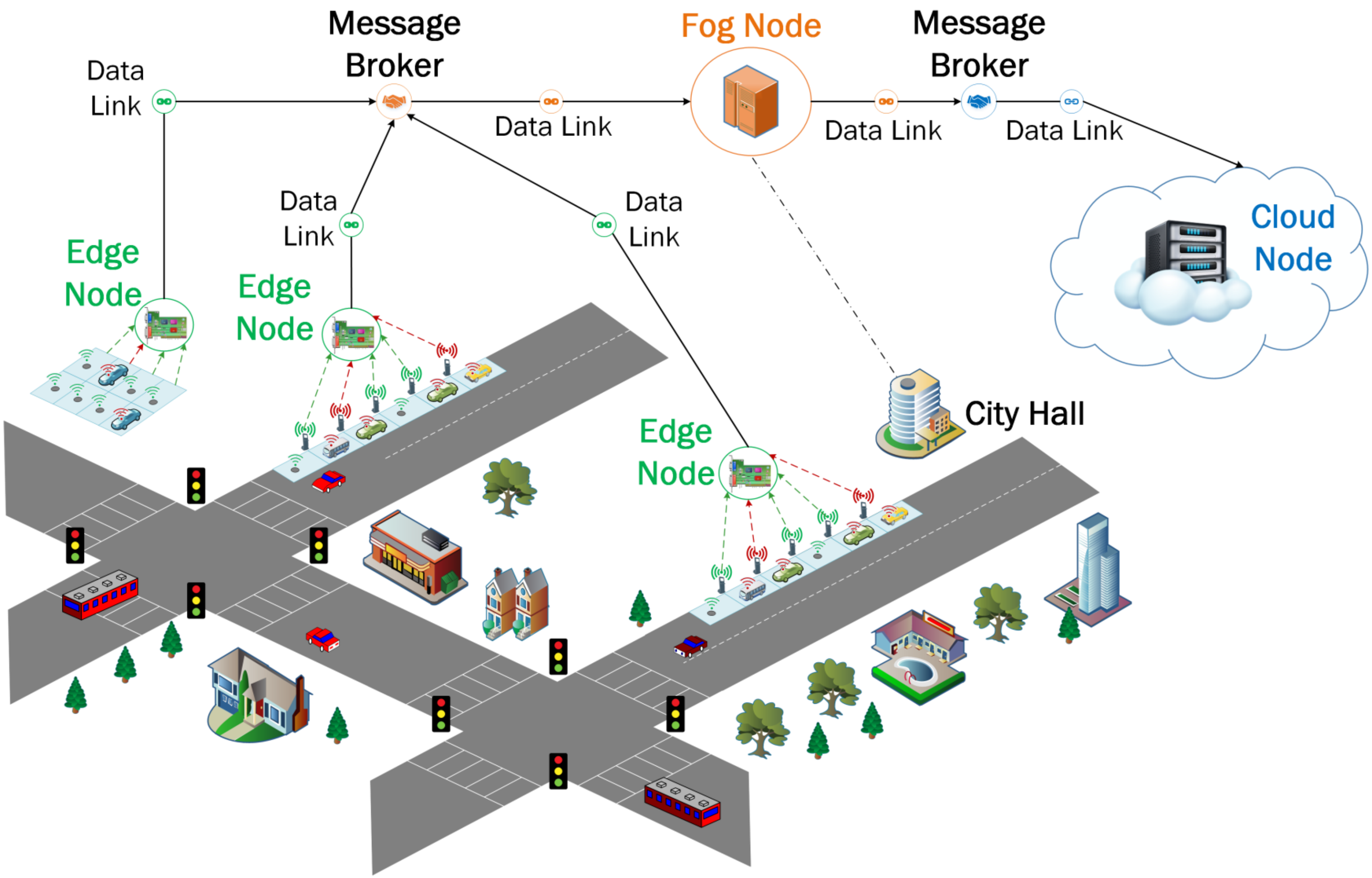
Smart Transit



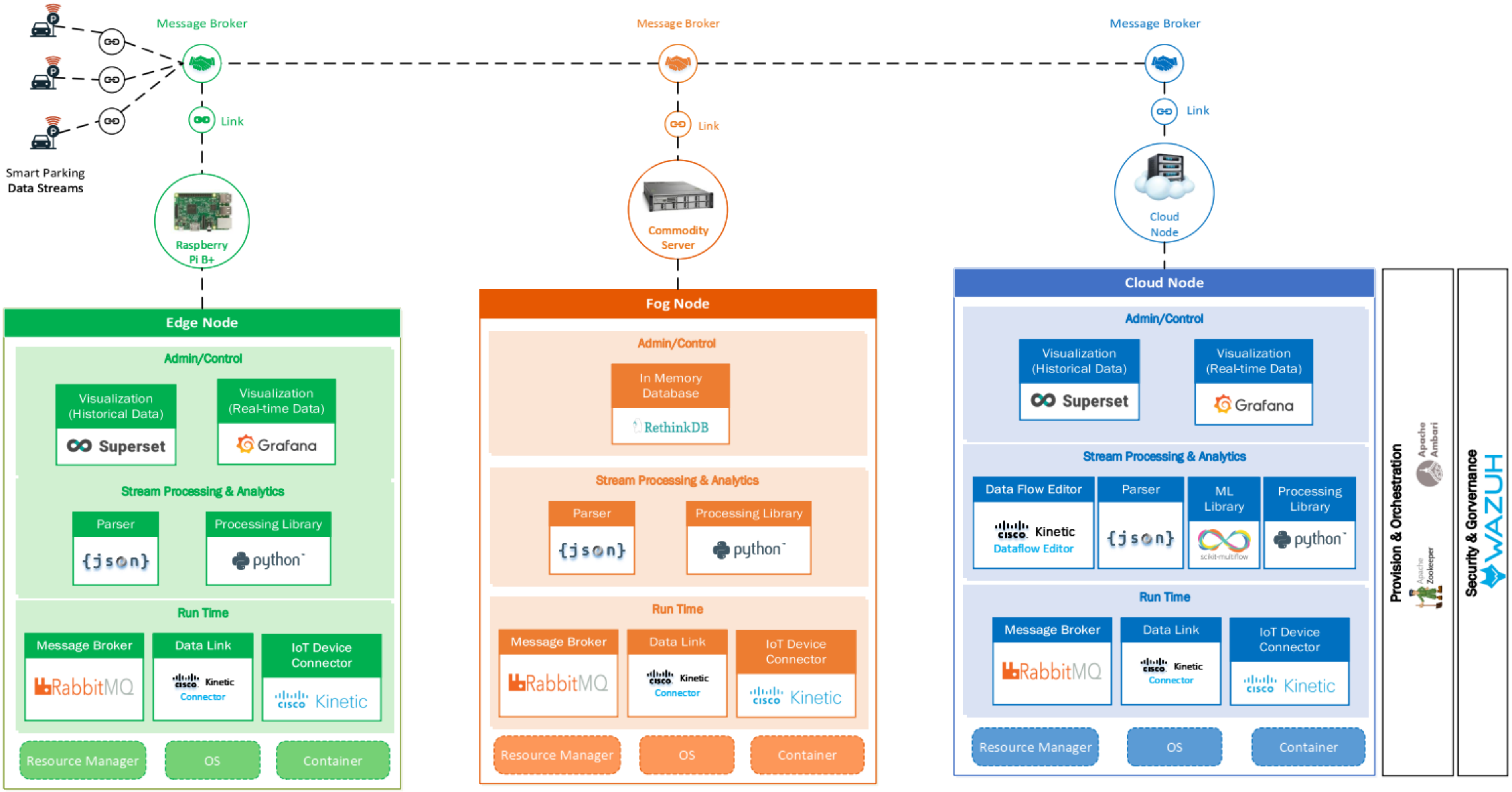
Smart Transit



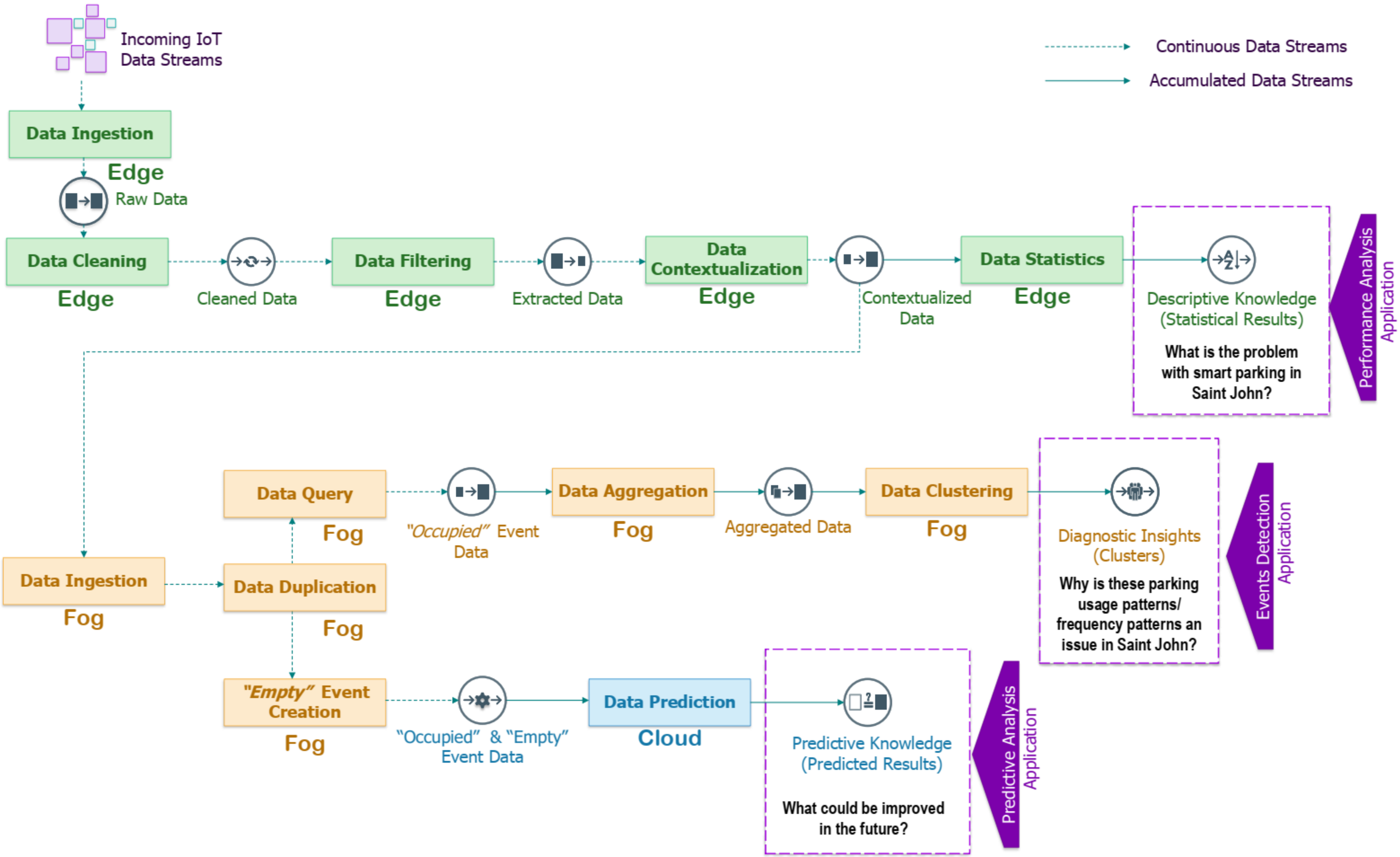
Smart Parking



Smart Parking



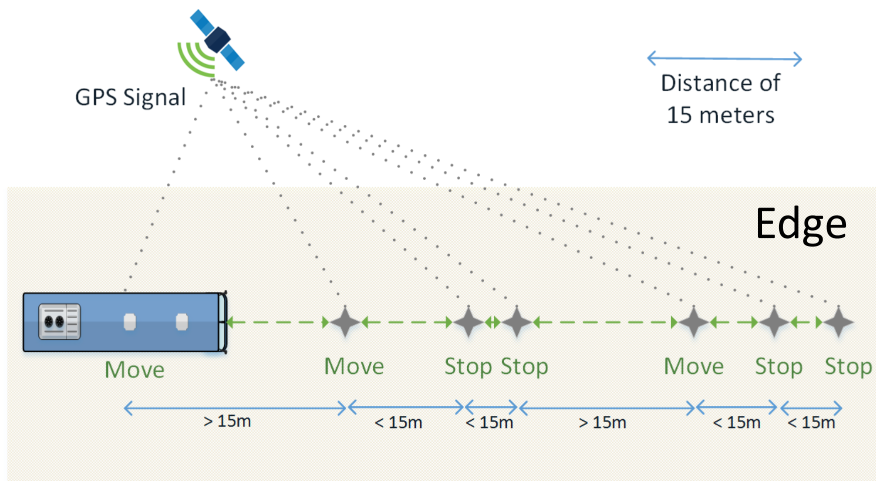
Smart Parking

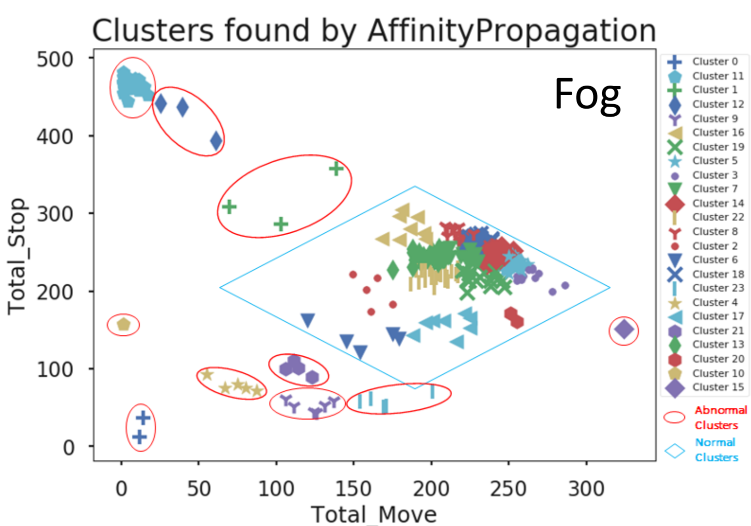


Results

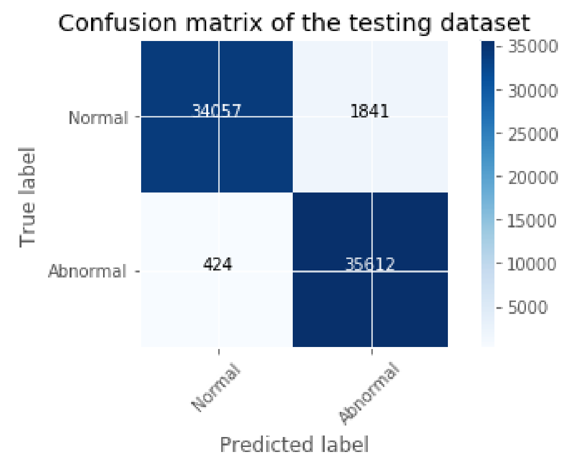
Smart Transit

Analytics at the Edge-Fog-Cloud continuum



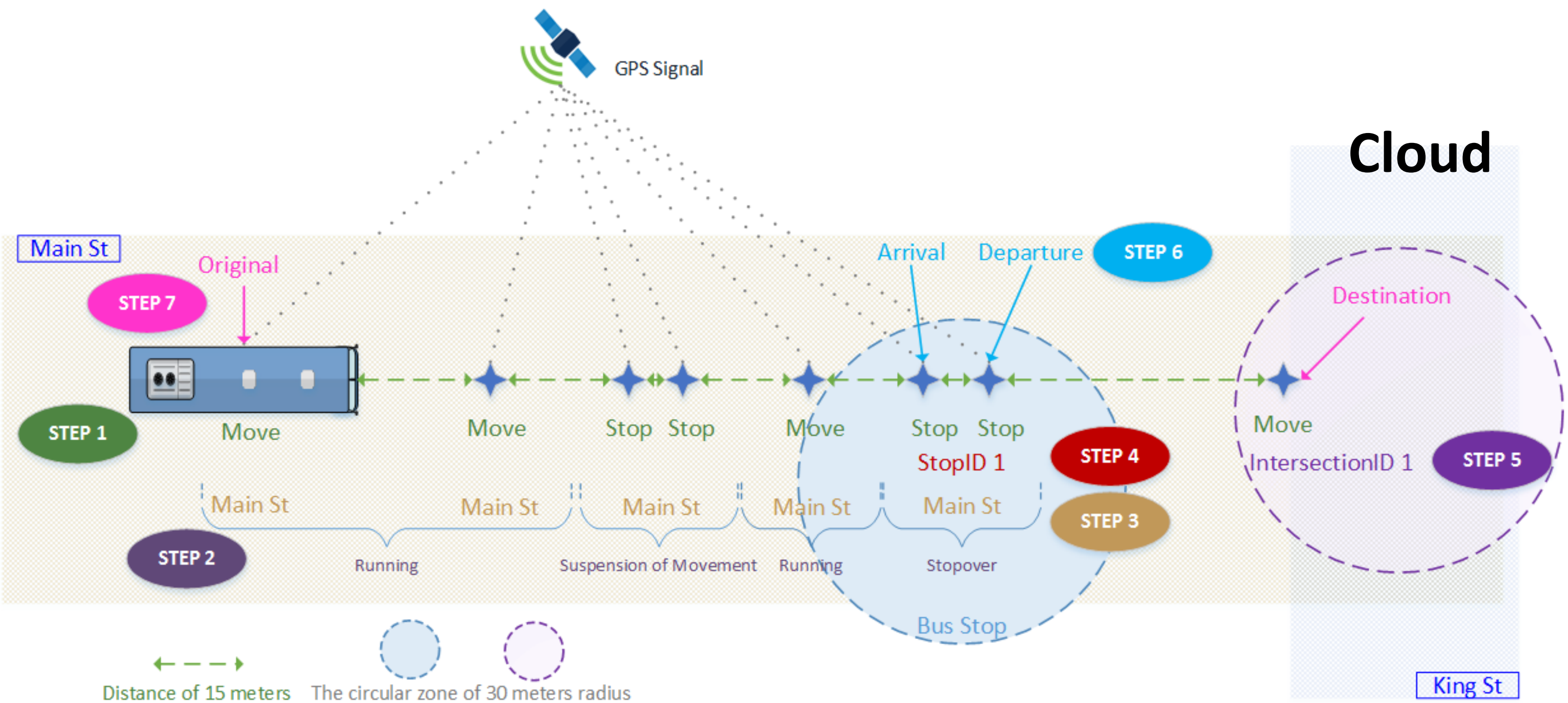


Cloud

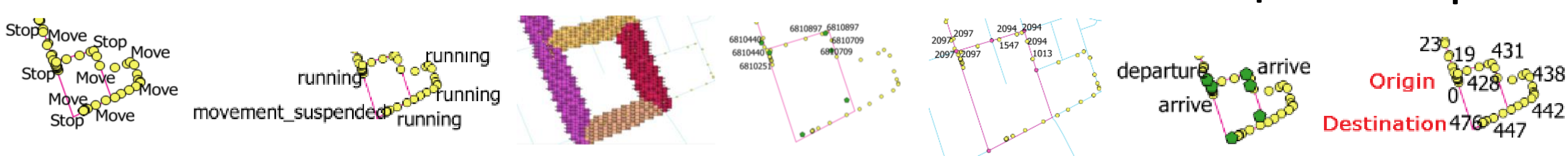


Smart Transit

Integrating GIS into our Edge-Fog-Cloud continuum

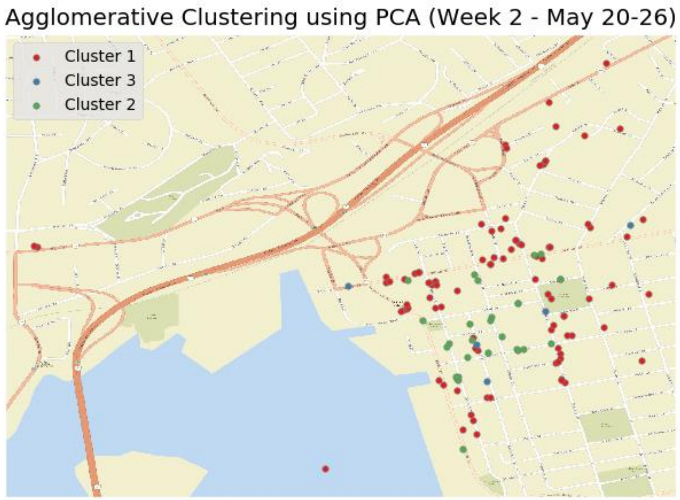


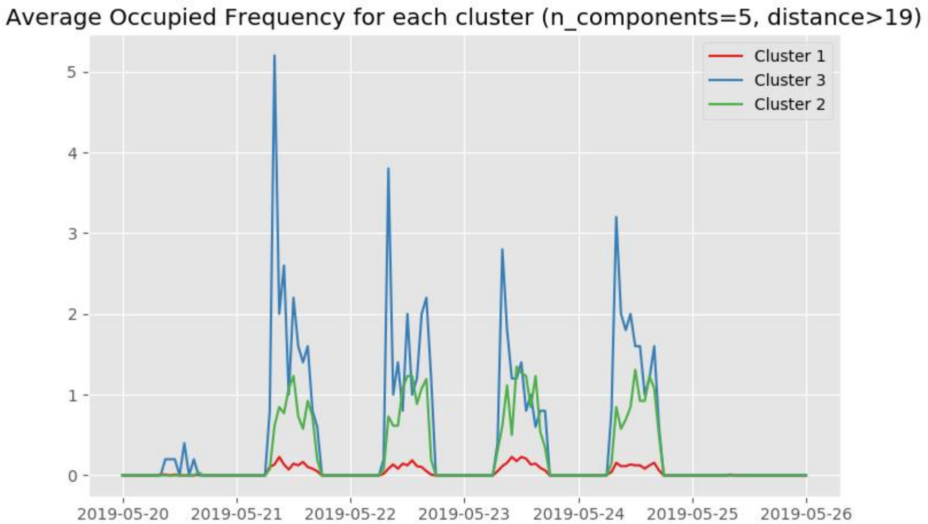
Step 1 **Step 2** **Step 3** **Step 4** **Step 5** **Step 6** **Step 7**

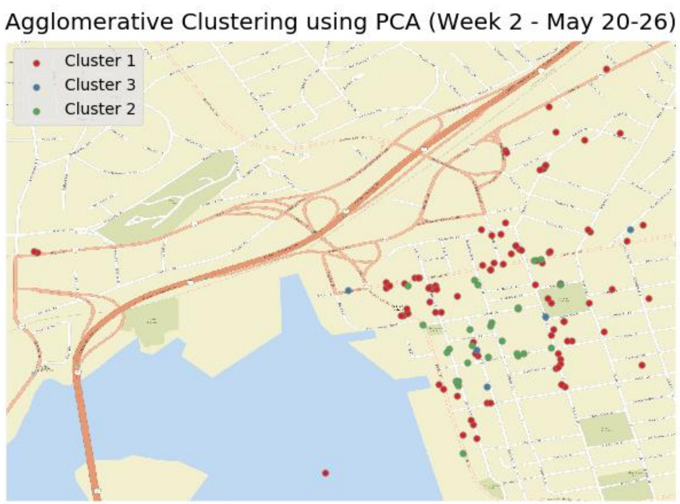


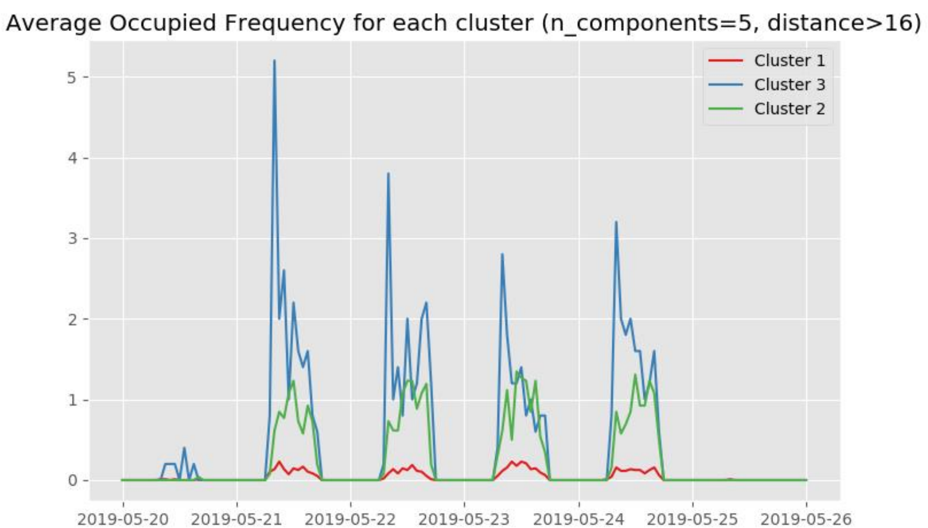
Smart Parking

Why are the discovered usage patterns an issue in Saint John?









https://youtu.be/iS6-WxYHkv8

Scientific Contributions

IEEE Access
Received April 17, 2019, accepted May 20, 2019, date of publication May 28, 2019, date of current version June 12, 2019.
Digital Object Identifier 10.1109/ACCESS.2019.2919514

Analytics Everywhere: Generating Insights From the Internet of Things

HUNG CAO[1], MONICA WACHOWICZ[1], CHIARA RENSO[2], AND EMANUELE CARLINI[2]
[1]People in Motion Lab, University of New Brunswick, Fredericton, NB E3B 5A3, Canada
[2]HPC Lab, ISTI-CNR, 56127 Pisa, Italy
Corresponding author: Hung Cao (hcao3@unb.ca)
This work was supported in part by the NSERC/Cisco Industrial Research Chair under Grant IRCPJ 488403-14.

sensors
MDPI

Article
An Edge-Fog-Cloud Architecture of Streaming Analytics for Internet of Things Applications

Hung Cao * and Monica Wachowicz
People in Motion Lab, University of New Brunswick, Fredericton, NB E3B 5A3, Canada
* Correspondence: hcao3@unb.ca

Contents lists available at ScienceDirect
Computers, Environment and Urban Systems
journal homepage: www.elsevier.com/locate/ceus
ELSEVIER

The design of an IoT-GIS platform for performing automated analytical tasks

Hung Cao*, Monica Wachowicz
People in Motion Lab, University of New Brunswick, Canada

1. Proposed an *"Analytics Everywhere"* framework which demonstrates that a single computational resource is not sufficient to support IoT applications.
2. Proposed a new architecture based on an integrated fabric of edge-fog-cloud nodes for executing a network of analytical tasks.
3. Automated analytical tasks have been implemented into a IoT-GIS platform which is capable of handling accumulated streams.
4. Implemented the proposed framework using two use cases in smart transit and smart parking.

Recommendations for Future Work

- The Analytics Everywhere framework
  - Streaming analytical tasks and computational resources could be **dynamically** mapped together in the next development step.
  - Extending GIS functionalities to the edge and fog computing environment
- The proposed architecture could be extended by considering:
  - Security
  - Latency
  - Fault tolerance
  - Privacy requirements of IoT applications
  - Concept-drift

Recommendations for Future Work

- The next research question is:
  - *How can a streaming analytical task in the network of tasks reuse and exploit the gained knowledge from the other tasks to resolve its own problems?*
- A potential approach that could answer this research question is the transfer learning methods.

People in Motion Lab

www.people-in-motion-lab.org

About me: www.hungcao.me