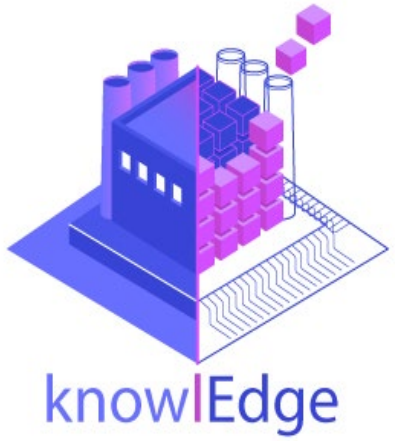


Towards AI powered manufacturing services, processes, and products in an edge-to-cloud-knowlEdge continuum for humans

Data collection through edge-fog-cloud continuum Final Review Meeting, Online,

2024-04-24

Edoardo Pristeri – LINKS Foundation

{name}.{surname}@linksfoundation.com

Motivation and Challenges

Motivation

- In the realm of **Industrial Internet of Things** (IIoT), modern applications should strategically harness a **distributed infrastructure** that seamlessly integrates **edge**, **fog**, and **cloud** nodes.

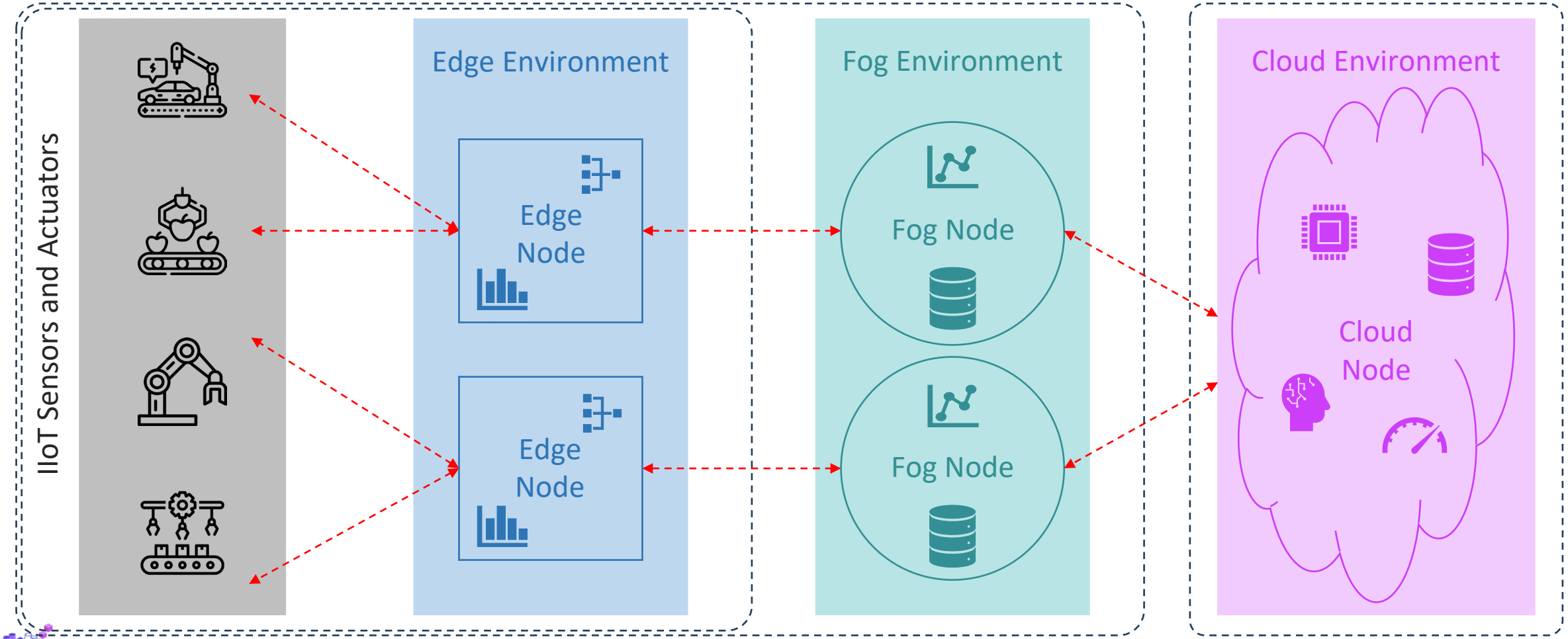
knowledge

Motivation

- **Edge nodes**:
  - These are the devices that are **closest** to the **data sources**, such as sensors, actuators and controllers. Edge nodes can perform **data acquisition** and **preprocessing**, as well as real-time **communication** and **control** of other devices. Edge nodes are suitable for applications that require **low latency**, high reliability, and **limited bandwidth**.
- **Fog nodes**:
  - These are the **intermediate devices** that are located between the edge and the cloud, such as **small servers**. Fog nodes can **aggregate** and **filter** data, as well as perform more **complex analysis** and storage of data. Fog nodes are suitable for applications that tolerate moderate latency and require scalability and security.
- **Cloud nodes**:
  - These are the **centralized devices** that are in remote **data centres**. Cloud nodes can store, process, and analyse **large amounts** of **data**. Cloud nodes are suitable for applications that require high performance, availability, and accessibility, such as **big data** analytics, **deep learning**, **decision support**, and **visualization**.

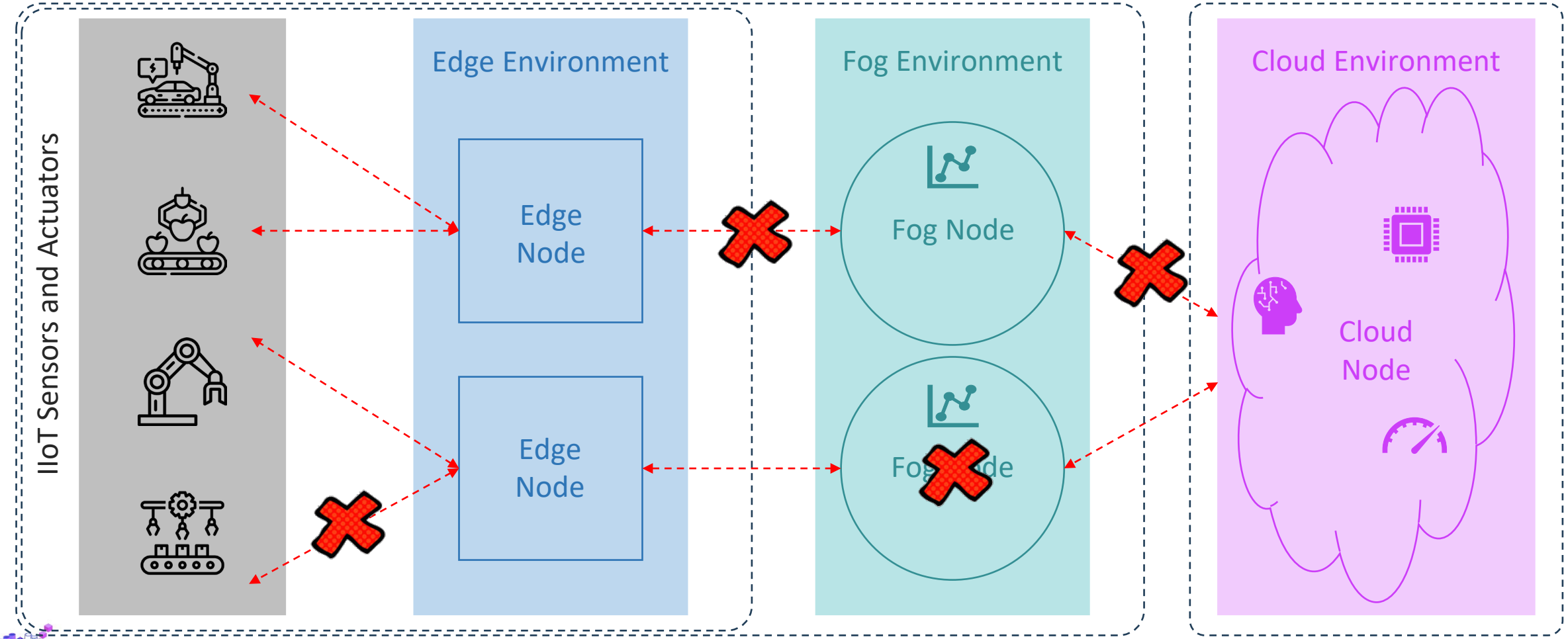
knowlEdge

Motivation



knowlEdge

Motivation



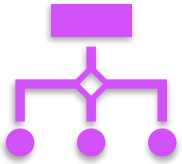
knowledge

Challenges

- **Computing Resources Diversity**
  - Inability to use all data produced in the environment due to different computing capacity at different levels of the compute continuum
- **Data Diversity and lack of Integration**
  - The presence of varied data types and collection methods poses challenges in integrating information into the lifecycle of processes, ensuring reliable fusion of heterogeneous data from multiple sources, and addressing interoperability issues.
- **Information Isolation**
  - The isolated exchange of information among different processes and the inaccessibility of data pipelines obstruct a comprehensive analysis of process data.

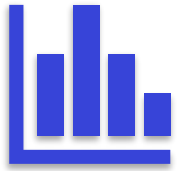
knowledge

Challenges



Data Integration

The **collection**, **integration**, and creation of **interoperable** data models.



Data Quality

The **refinement** and **assurance** of **data quality**

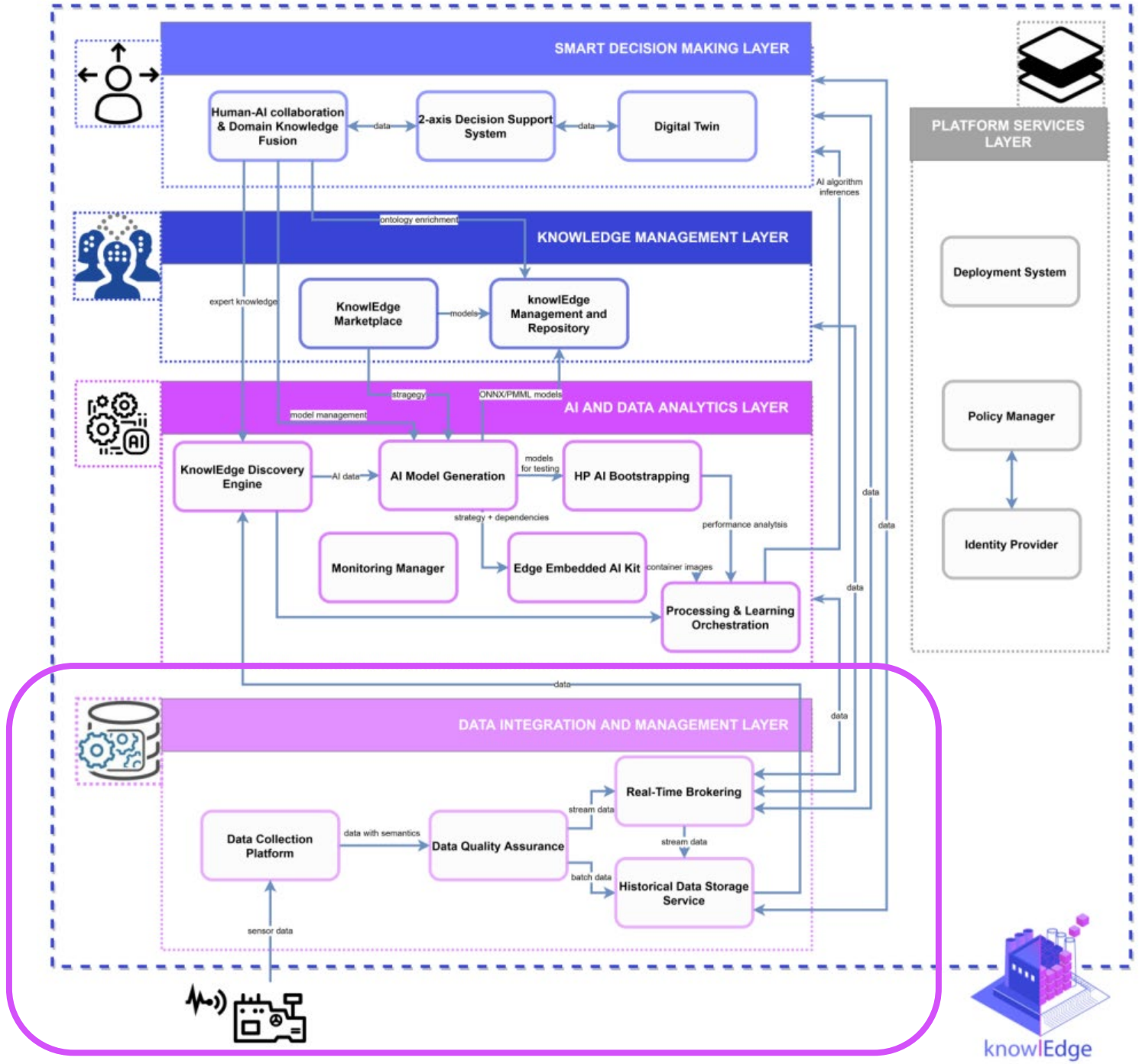
Data Governance

The **retrieval**, **access**, and **governance** of **real-time** and **historical** data across the **computing continuum**.

knowledge

Approach

knowlEdge Data Integration and Management Layer



knowlEdge

knowlEdge

Data Collection Platform

The goal of the Data Collection Platform is to Reads pilot's sensors and databases and expose unified data to knowledge components.

- **Southbound Interfaces** for connecting with custom external protocols
- **Data Transformation and Unification** adopting the knowlEdge Data Model
- **MQTT Northbound Interface** enabling connection with Real Time Broker

Data Quality Assurance

The Data Quality Assurance Architecture (**DQAA**) is a **micro-service-based** component composed of multiple services.

- **Connector Service** for *reading* data from the Real-Time Broker
- **Validator Service** for applying *data quality* rules on incoming data
- **Anonymization Service** for *anonymizing* only valid data

knowlEdge

Real Time Broker

The Real Time Broker distributes the data collected from the shopfloors and early processed by the data quality assurance and anonymization modules towards the various components of the upper layers of the knowlEdge system.

- **RabbitMQ Based** offering relayable queues and multi-protocol support
- **MQTT Plugin** enabling pub-sub capabilities and TLS support
- **Shovel Plugin** to enable worker nodes to securely distribute messages

knowlEdge

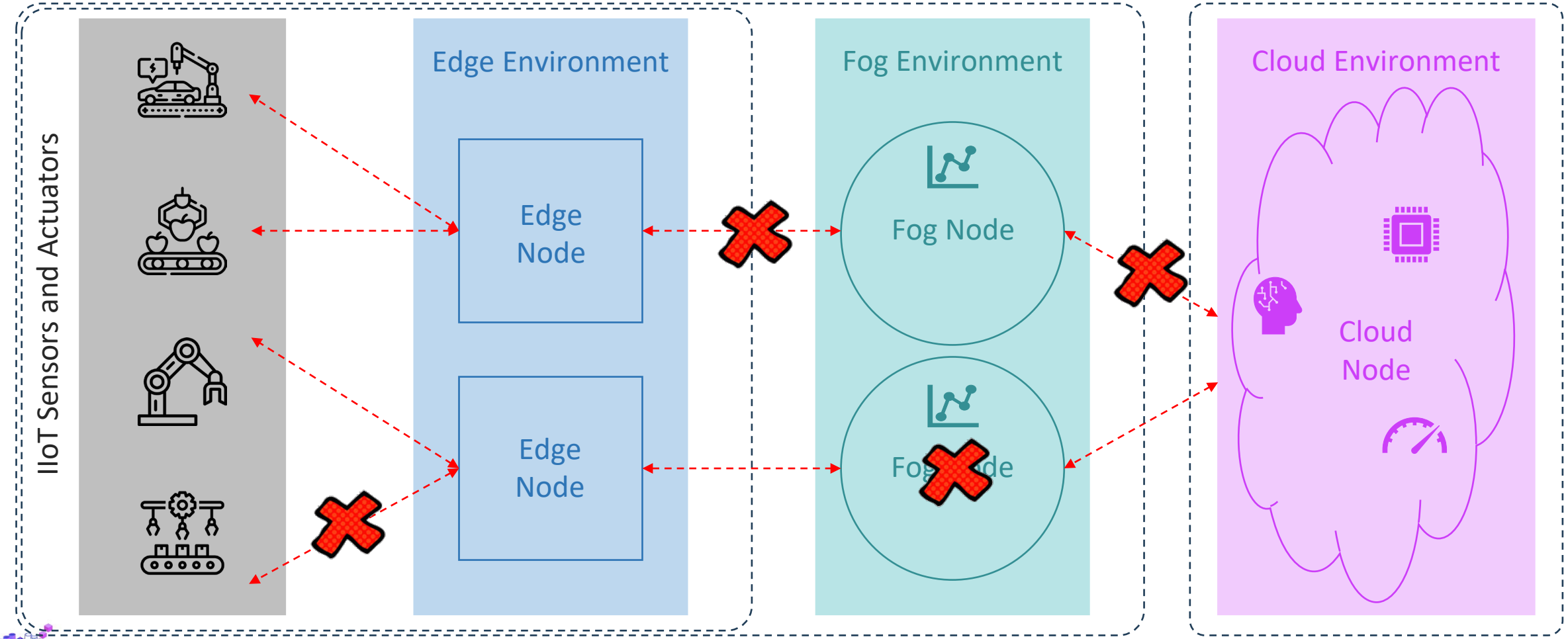
Historical Data Storage

The Historical Data Storage provide historical data across the factory premises and the cloud for the knowlEdge platform components. The service also capable to provide, data compression, data transformation and aggregation

- **Timeseries Storage** for storing raw, validated and anonymized datasets
- **Storage of Data Sets for AI Models** included in the Knowledge Repository
- **Storage of Intermediate Simulation Results** from AI components
- **Data Replication** across computing continuum, when needed

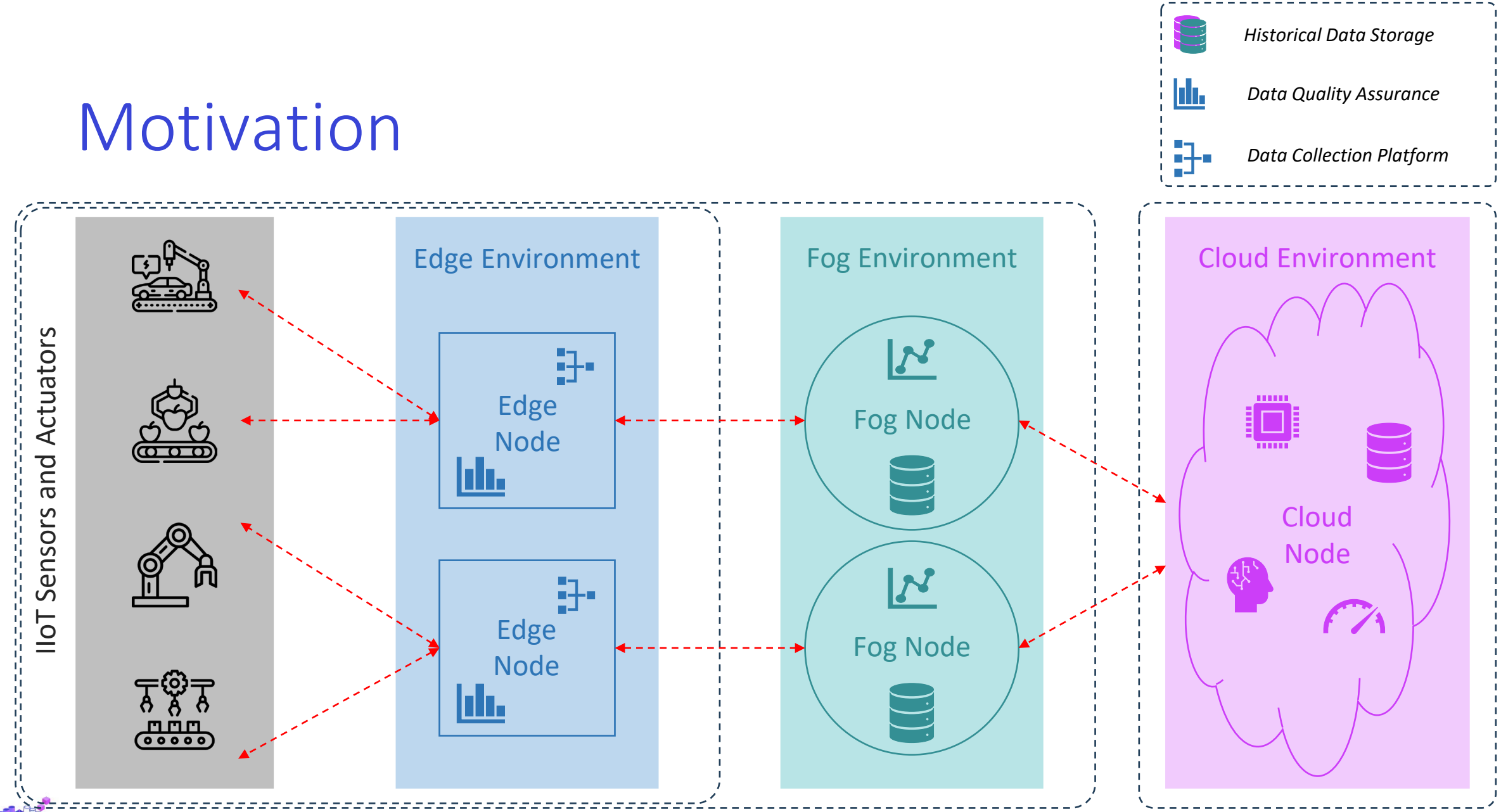
knowlEdge

Motivation



knowlEdge

Motivation



knowlEdge

Results

Solved Challenges: Data Integration

Enabled the **data collection** from the shop-floor in the **pilot** environments.

Implemented a support adaptation layer for the **interoperability** among all the **subsystems** involved.

Selected of a **data model** that considers all the **properties** and **features** of all the diverse communication **endpoints**.

Transformed shop-floor **heterogeneous** data for providing a proper **input** to the **high-level services**.

knowlEdge

Solved Challenges: Data Integration

knowlEdge

Solved Challenges: Data Quality

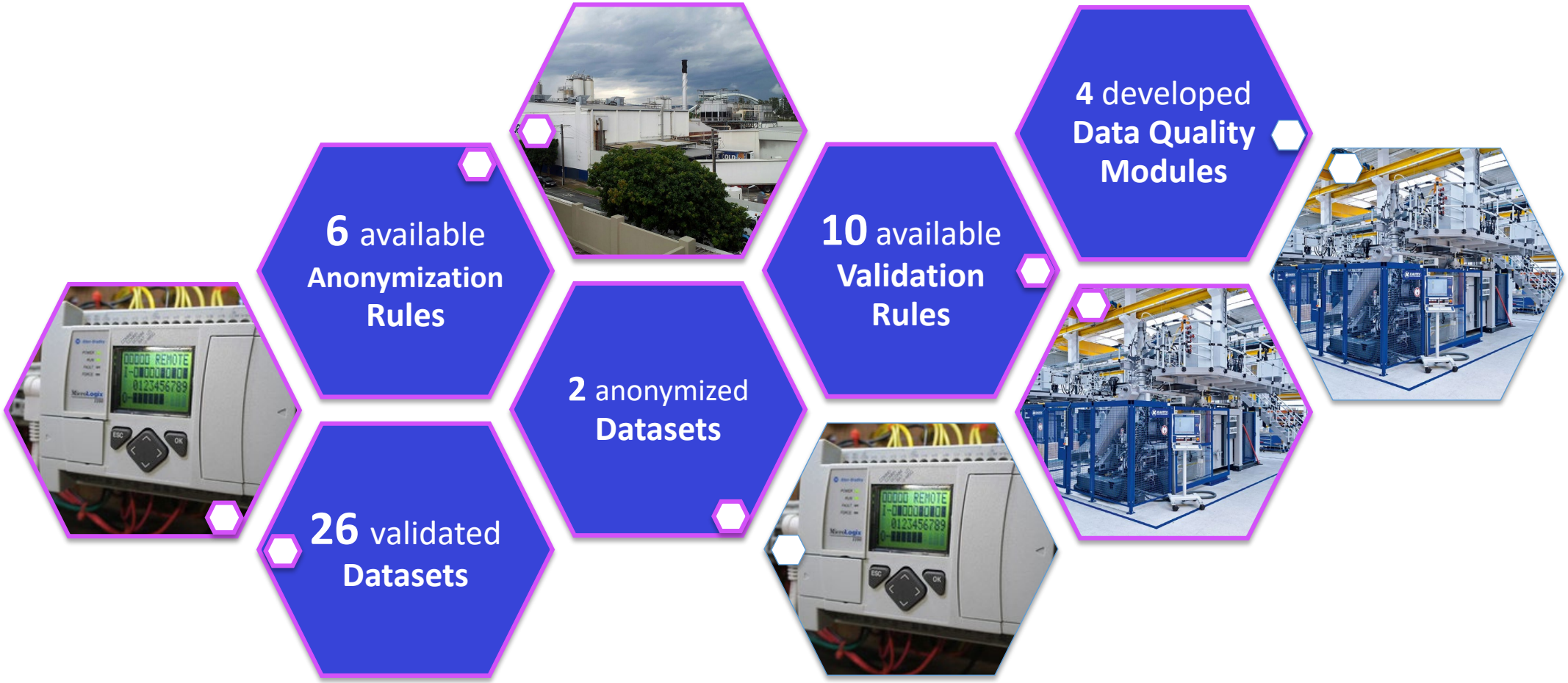
Developed a modular **edge pipeline** to **acquire** and **process** the shopfloor **data**

Provided modules and **algorithms** aimed to ensure **data quality**

Ensured privacy and **confidentiality** of the collected data by **masking** the **sensitive** data at the **edge** in **real-time** and **historical** datasets

knowlEdge

Solved Challenges: Data Quality



knowlEdge

Solved Challenges: Data Governance

Historical data managed at the factory premises for visualization and data mining purposes.

Effective queuing mechanism for the live data providing failure free data propagation across the nodes at the edge, fog and the cloud.

Ensure low footprint on processing, storage requirements and latency of the developed artefacts

Data orchestration and data access methods considering privacy and confidentiality needs of the fog computing paradigm

knowlEdge

Solved Challenges: Data Governance

knowlEdge

Lessons Learnt and Open Challenges

- REST Protocol Limitations:
  - Struggles with large data quantities.
    - Implemented data splitting and retrieval in smaller segments, although requiring frequent queries.
- MQTT Advantages:
  - Eliminates the need for synchronization and enhances data handling efficiency.
    - Usually limited to sensors data transmission in industrial context, REST much more present
- Data Quality Validation Rules:
  - Not always clear from the beginning
    - An iterative process has to be implemented to tune the configuration of the DQ component

knowlEdge

Final Remarks and Discussion

More Details

- knowlEdge Project Website
  - https://www.knowledge-project.eu/
- knowlEdge Deliverables
  - D3.2 – Final site-wide data collection and integration toolkit
  - D3.4 – Final Data Management and Data Quality modules
  - D3.6 – Final Site-wide Data Storage and Governance suite
- knowlEdge Scientific Publications
  - E. Alberti et al. - AI Lifecycle Zero-Touch Orchestration within the Edge-to-Cloud Continuum for Industry 5.0

knowlEdge

Discussion

Questions?