A **Federated** Data Management Model for **Scalable** Digital Twins

Julia Robles, **Cristian Martín,** Manuel Díaz
ERTIS Research Group and ITIS Software
University of Málaga, Spain
cristian@uma.es

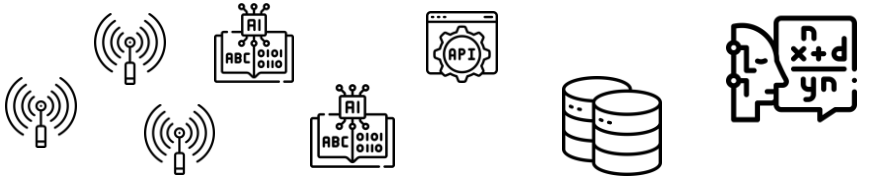
Context. The challenge DTs: scalability

Digital twin

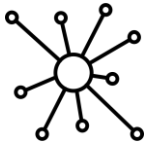
A formal digital representation of an asset that remains **synchronised** with its physical counterpart, capturing the attributes and behaviours relevant to its communication, storage and interpretation within **a specific context**

Two-dimensional scalability

Operational
Handling of large volumes of heterogeneous data with frequent updates



Structural
Ability to model thousands of entities and their relationships/interdependencies



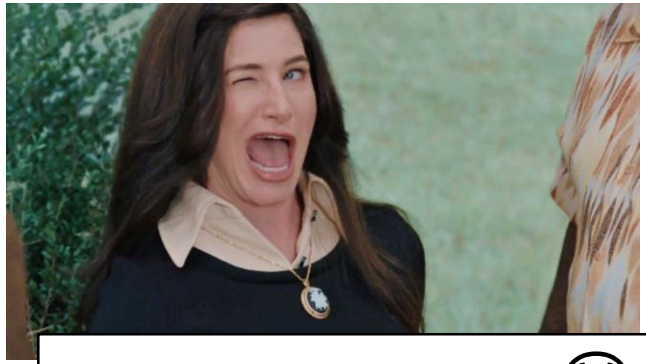

Gasoline
Diesel
Oil
Parafins
for apple

OpenTwins: An open-source framework for the development of next-gen compositional digital twins

opentwins Public

Innovative open-source platform that specializes in developing next-gen compositional digital twins

☆ 247 51



A start is welcome 😊

Robles, J., Martín, C., & Díaz, M. (2023). OpenTwins: An open-source framework for the development of next-gen compositional digital twins. *Computers in Industry*, *152*, 104007.

Rationale: Technical dichotomy in data management

Requirements for large-scale DT

Semantic expressiveness

Necessary for modelling the topology and relationships of a complex system

High frequency

Required for the mass collection of telemetry data

Typically, DTs manage their data centrally, but:

- Knowledge graphs (KG) suffer a **decline in performance under heavy write loads**
- Time-series database **(TSDB) architectures lack semantic and relational context**

Current architectures do not scale in terms of both semantic complexity and data volume

Proposal: Federated model and polyglotal persistence

Polyglotal persistence
Assignment of specific storage technologies based on the nature of the data to optimize performance

Semantic plane

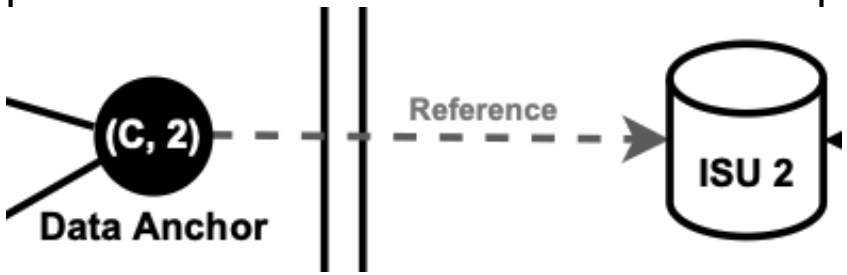
Structures and relationships (static/slow) managed in a KG

Data anchors

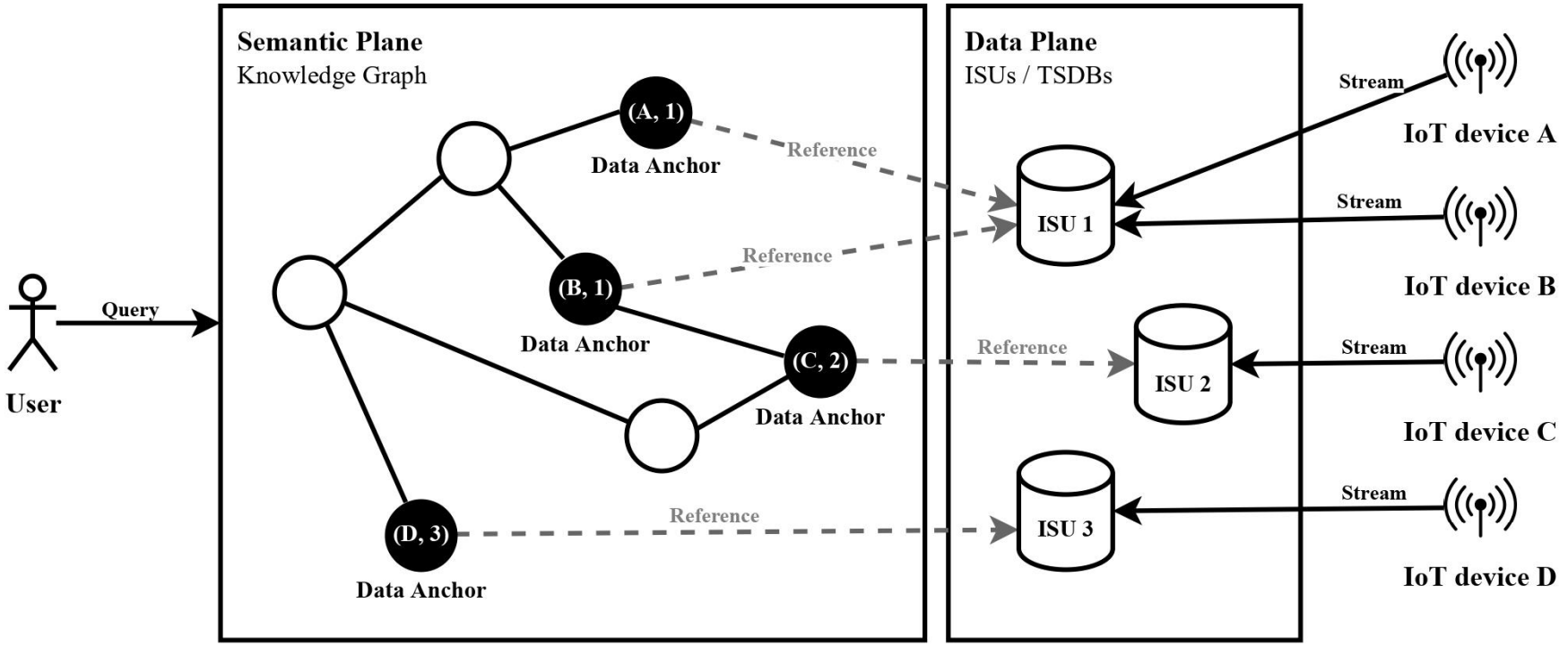
A logical abstraction in the KG that serves as a reference to an ISU in the Data Plane.

Data Plane

Volatile telemetry managed in TSDBs known as Independent Storage Units (ISUs).



Proposal: Data management **federated model**



Direct connection from the sensors to the ISUs, bypassing the KG

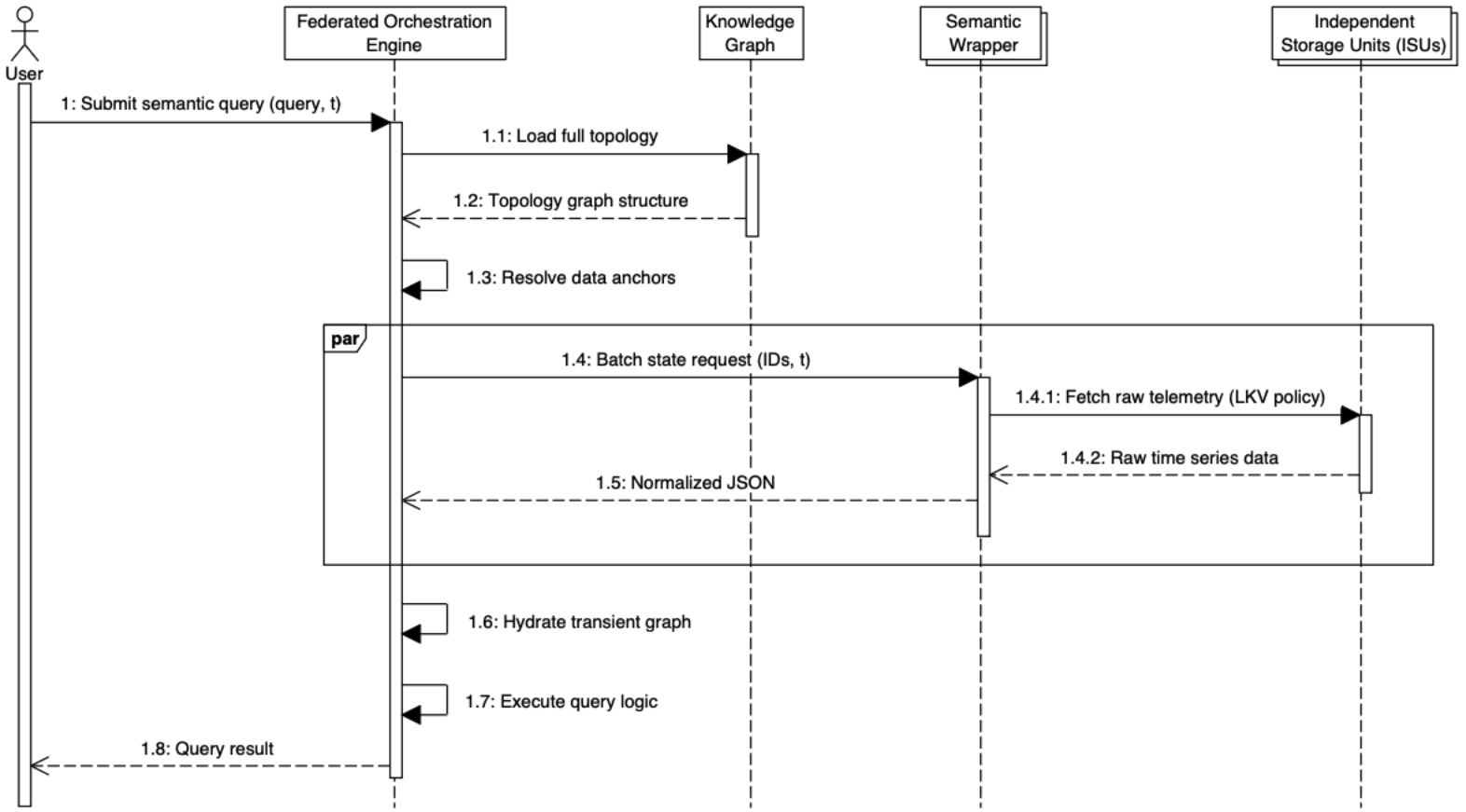
The KG serves as a benchmark index, not as a repository of values

State reconstruction mechanism and queries

Concept: The global state of the DT is neither static nor persistent but is **constructed dynamically at runtime**.

- A **lazy evaluation** strategy is used: data is retrieved and joined only when a query is executed.
- Use of a middleware component (**Federated Orchestration Engine**), which acts as an intermediary between the user and the planes, resolving anchors, requesting status data and dynamically generating the DT's global status.
- Use of adapters in each ISU (**Semantic Wrappers**), which standardize the output from the TSDBs to a canonical schema to abstract away heterogeneity and enable communication with the middleware.

Query orchestration mechanism



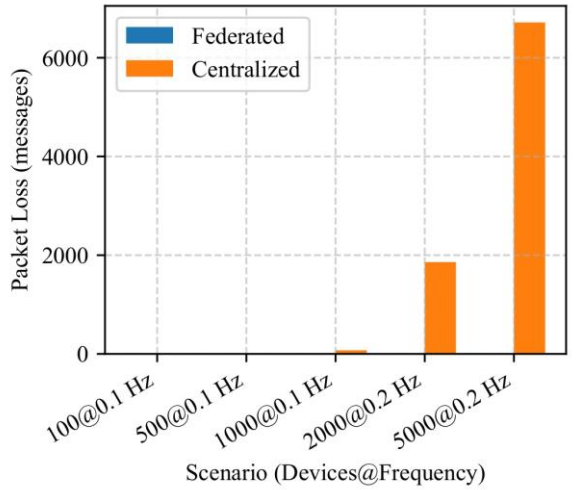
Experimental evaluation: Data ingestion

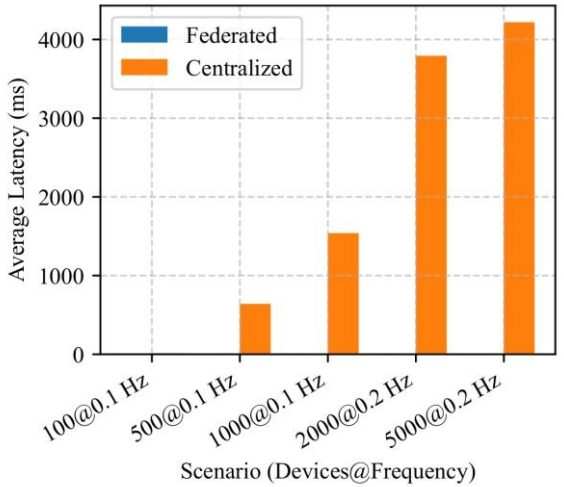
Scenarios: Up to 5,000 sensors transmitting at up to 0.2 Hz (1,000 messages per second).

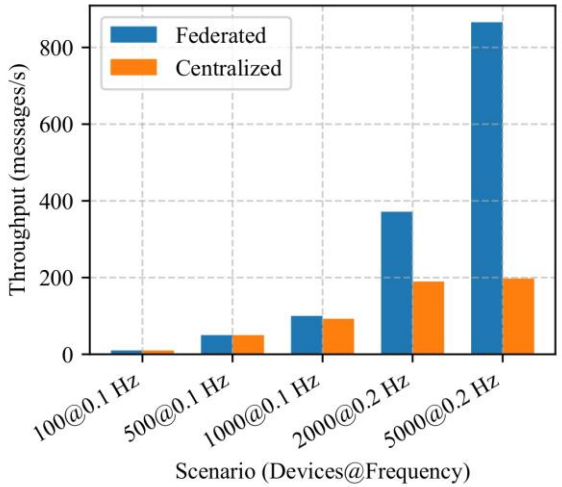
Results:

- **Federated:** Stable latency (<6ms). 0% loss of messages.
- **Centralized:** System failure at 200 messages per second. Latencies > 4000ms and high data loss.

Decoupling eliminates the serialization bottleneck in the KG







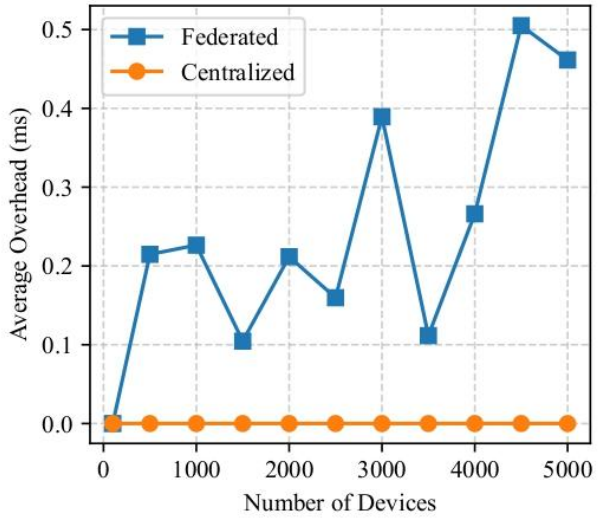
Experimental evaluation: Query

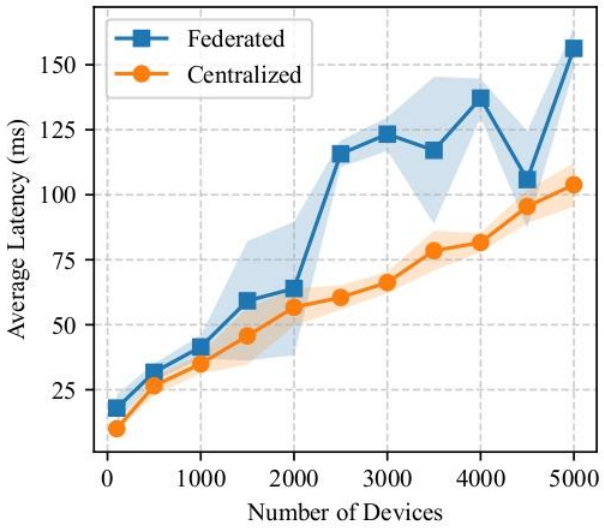
Scenarios: Incremental loading of 100 to 5,000 assets per query.

Results:

- **Total latency:** The federated model completes the reconstruction in 156ms for 5,000 assets (a ~1.5x increase compared to the centralised model).
- **Orchestration overhead:** Minimal algorithmic cost (< 0.5 ms). Latency is dominated by network transport (I/O).
- **Scalability:** Linear growth (O(N)) due to the batching strategy.

The cost of federation is a reasonable trade-off for achieving scalability in data ingestion





Conclusions

Pros	Cons
Horizontal scalability Allows multiple ISUs to be added without affecting the KG's performance.	**Distributed query latency** On-demand reconstruction introduces delays.
Lifecycle evolution: Changes to hardware or sensors only require updating "data anchors", not the entire semantic model.	**Approximate temporal alignment** Strict synchronisation is not guaranteed between heterogeneous sources (ISUs)
Dynamic snapshot consistency: Ability to generate a coherent view of the global state for any instant, facilitating temporal navigation.	
Resilience: The system maintains situational awareness through unavailability markers in the event of connectivity failures.	

Future and ongoing work

OpenTwins v2

Open-source platform for the development of composite DTs

This data management model is being implemented for version 2 of the platform

Future work:

Optimisation
Implementation of semantic cache layers to reduce latency in concurrent queries

Synchronisation
Development of advanced interpolation mechanisms for more precise temporal alignment

Validation
Deployment in complex industrial use cases with high-frequency data

Thank you. Any question?

A **Federated** Data Management Model for **Scalable** Digital Twins

Cristian Martín,
University of Malaga, Spain
(cristian@uma.es)

Malaga, Spain