

Semantic-Aware Orchestration and Energy-Optimized Data Management across the Edge–Cloud Continuum: The GLACIATION Approach

Aidan O’Mahony¹, Pournima Sonawane¹ and Shraddha Gupta¹

¹Applied Research, Dell Technologies, Cork, Ireland

Abstract

The rapid growth of data-intensive applications across the edge-cloud continuum presents a dual challenge: immense energy consumption and complex data governance. This paper presents the GLACIATION project, which tackles these issues through a platform for energy-efficient and privacy-preserving data operations. We showcase an integrated approach that combines a Distributed Knowledge Graph (DKG) with AI-driven orchestration to automate and optimize data management. Key innovations include the *Green Index*, a real-time metric for sustainable energy-aware workload shifting; a bio-inspired scheduling engine using Ant Colony Optimization; and a lightweight, zero-shot Large Language Model (LLM) interface that enables semantic exploration of the DKG via natural language. The effectiveness of this framework is demonstrated through its validation in real-world industrial and public sector pilot deployments.

Keywords

Distributed Knowledge Graphs, Data Placement, Workload Orchestration, AI Scheduling, SPARQL with LLMs, Edge-Cloud Systems

1. Introduction

The proliferation of IoT and data-intensive applications is driving a shift towards the edge-cloud computing continuum, a paradigm that introduces significant challenges related to data governance, latency, and energy consumption. As data generation at the edge explodes, the need for intelligent, automated, and efficient data operations becomes critical. Traditional cloud-centric models are often ill-equipped to handle the complex trade-offs between processing data locally to reduce latency and moving it to centralized clouds for powerful analytics, all while adhering to strict privacy regulations and minimizing the environmental footprint.

This paper outlines the key innovations of the GLACIATION platform. We begin by describing the project’s architecture and the core semantic metadata model (GLC-MRM) that enables interoperability. We then detail the platform’s AI-driven optimization capabilities, including the use of Ant Colony Optimization for data discovery, a bi-level scheduling approach for managing trade-offs, and a lightweight Large Language Model (LLM) interface for democratizing data access. Finally, we present the validation of our approach through real-world industrial pilots, with a focus on the *Green Index* for sustainable, energy-aware orchestration in the energy sector.

2. The GLACIATION Project

GLACIATION (Green responsibLe privACy preservIng dATA operaTIONS) is a Horizon Europe project designed to address the significant energy consumption and carbon emissions resulting from the rapid growth of big data analytics across the edge-to-cloud continuum. The project’s central ambition is to create a platform for energy-efficient, privacy-preserving data operations. The core of this initiative is the development of a novel Distributed Knowledge Graph (DKG) that spans the entire edge-core-cloud

SEMANTICS’25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria

✉ aidan.omahony@dell.com (A. O’Mahony); pournima.sonawane@dell.com (P. Sonawane); shraddha.gupta@dell.com (S. Gupta)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

architecture. By leveraging AI-enforced minimal data movement and optimizing the physical location of analytics and data storage, GLACIATION aims to achieve substantial reductions in power consumption.

The technical approach revolves around a modular, microservices-based platform that supports the DKG. The conceptual architecture, illustrated in Figure 1, is comprised of key components including a Metadata service to manage data annotations, a Trade-off service to balance latency and resource use, and a Prediction service for forecasting data popularity and workload needs. A vital element is the project’s metadata framework, which provides the tools to embed privacy and trust requirements directly into data operations. The efficacy and generality of the GLACIATION platform will be validated through four demanding, real-world use cases in the public-service, manufacturing, enterprise, and energy sectors, led by partners MEF/Sogei, Dell, SAP, and IPTO, respectively¹.

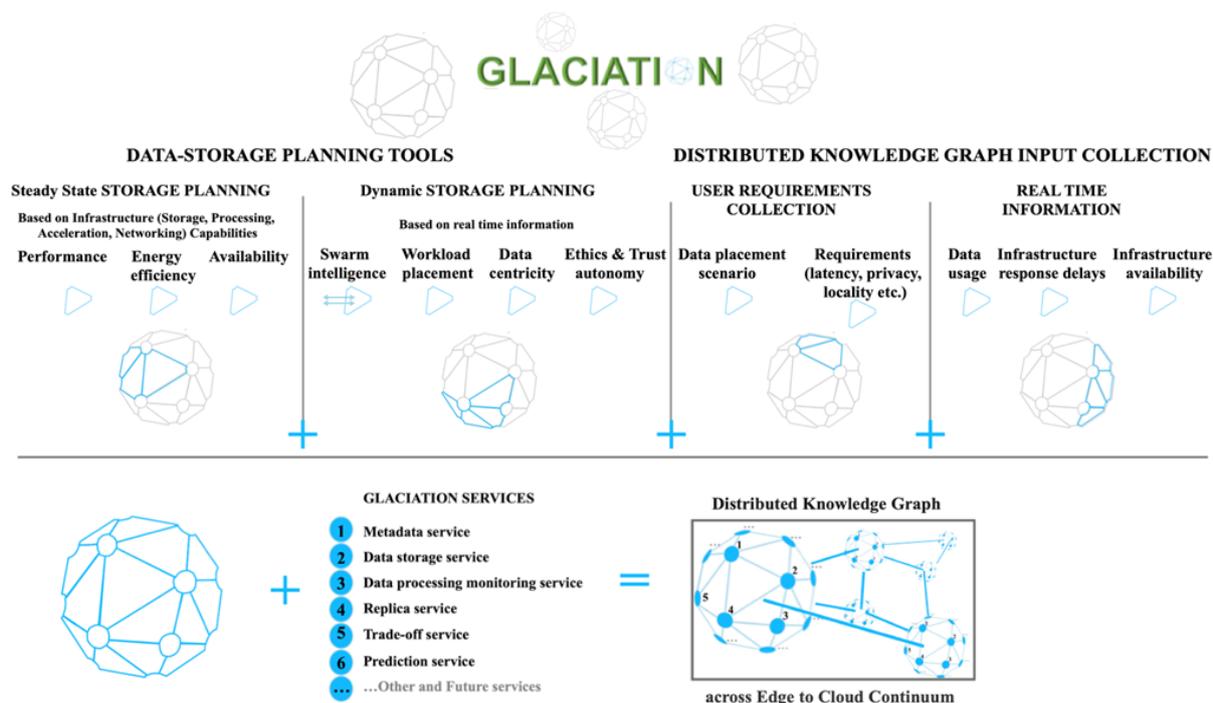


Figure 1: The GLACIATION Conceptual Architecture, illustrating the Distributed Knowledge Graph (DKG) and its supporting microservices operating across the edge-to-cloud continuum.

3. Semantic Metadata Model

GLACIATION introduces the *GLACIATION Metadata Reference Model* (GLC-MRM) [1], a flexible and extensible ontology designed to enable interoperable data and metadata representation within a Distributed Knowledge Graph (DKG). Founded on Linked Data principles, the GLC-MRM uses standards like RDF, RDFS, and OWL to formalize a generic conceptualization of a task scheduling and resource monitoring environment. The core of the model consists of primary classes such as Task, Resource, Constraint, and Measurement, which describe the assignment of tasks to resources under specific hard or soft constraints, while monitoring their real-time and predicted performance.

The GLC-MRM is designed for extensibility through a use-case-driven specialization methodology, where the core ontology is adapted with modular vocabularies to describe specific assets, services, and contexts. This includes detailed specializations for Kubernetes resources (nodes, pods, workloads), IoT devices, and crucial energy metadata such as the Green Index. To ensure broad interoperability, the model is formally mapped to established external ontologies, including the DCAT vocabulary for describing datasets and the ETSI SAREF ontology for modeling devices and energy-related measurements.

¹The source code and use cases of the GLACIATION platform are available at <https://github.com/glaciation-heu>

This semantic framework, supported by technologies like SHACL for data validation and JSON-LD for lightweight data exchange, enables both declarative SPARQL queries and automated policy enforcement, forming the cornerstone of the platform’s orchestration capabilities.

4. Semantic-Based Optimization

The GLACIATION optimization strategy is centered on a Distributed Knowledge Graph (DKG), which serves as the foundation for a novel metadata fabric codenamed IceStream. The approach taken is inspired by the Agri-Gaia project [2]. The DKG’s primary objective is to furnish a real-time, comprehensive, and global perspective on data distributed across the network, encompassing cluster information, resource usage, and application status. This is achieved by representing all entities—such as nodes, services, data, and their associated properties—within a graph-based data model built on W3C standards like the Resource Description Framework (RDF) and Web Ontology Language (OWL). This approach creates an expressive, machine-readable semantic layer that describes the entire state of the edge-cloud continuum, enabling AI-driven decisions to optimize resource allocation and minimize energy consumption.

This rich semantic representation facilitates powerful optimization mechanisms through standardized protocols. The architecture supports declarative, decision-making queries via the SPARQL query language, allowing the orchestration engine to retrieve precise metadata for various microservices. For instance, the Trade-off Service queries the DKG to acquire metadata on latency, resource usage, energy, privacy, and security to inform workload placement decisions. Furthermore, the system leverages the Shapes Constraint Language (SHACL) to validate RDF data against a set of predefined conditions and policies, ensuring data quality, consistency, and compliance with domain-specific constraints. This combination of semantic querying and constraint validation provides a robust framework for implementing and enforcing contextual policies, enabling dynamic, automated, and intelligent optimization across the platform.

5. AI-Driven Scheduling and Trade-Off

GLACIATION employs a bio-inspired approach for distributed data discovery and movement, utilizing a variant of Ant Colony Optimization (ACO) [3]. In this model, search queries generate forward ants that traverse the Distributed Knowledge Graph (DKG) following pheromone trails. These trails are dynamically laid by backward ants, which retrace the path of successful queries, reinforcing routes that lead to desired RDF triples. Beyond discovery, the established pheromone gradients guide a data movement strategy, relocating frequently accessed data closer to its point of demand. This process improves search efficiency and hit rates by optimizing the physical location of data within the edge-fog-cloud architecture, thereby reducing network traffic and query latency.

The scheduling and orchestration logic is framed as a Bi-level Multi-Objective Optimization Problem (BMO-TSLB), addressed by an Improved Multi-Objective Ant Colony Optimization (IMOACO) algorithm [4]. At the upper level, the system optimizes for makespan, cost, and energy consumption of tasks. This is dependent on the lower-level optimization, which focuses on load balancing by minimizing response time and maximizing resource utilization across available fog nodes. A dedicated *Trade-Off Service* orchestrates this process, reasoning over the often-conflicting objectives of performance, power efficiency, and policy compliance. By integrating the outputs of the ACO-based data discovery with workload predictions, the service selects optimal workload placements that satisfy the multi-faceted constraints of the system.

6. Interfacing with LLMs

To democratize access to the Distributed Knowledge Graph (DKG), we introduce a framework for question answering that leverages Large Language Models (LLMs) in a zero-shot setting [5]. This

approach enables non-technical domain experts to query the DKG using natural language, eliminating the need for familiarity with SPARQL or the underlying KG schema. The framework operates by taking a user’s question and automatically extracting relevant context, such as pertinent class types and predicates from the KG schema. This contextual information, along with the natural language question, is then passed to an LLM, which generates a corresponding SPARQL query. This initial query can be further refined through an iterative enhancement loop, where the LLM itself parses, validates, and improves the generated query, thereby increasing the likelihood of a successful execution without requiring pre-existing question-query training pairs or model fine-tuning.

Our preliminary evaluations demonstrate the viability and effectiveness of this approach. For the task of generating SPARQL queries from natural language questions, using models such as Llama2-7B, Llama3-8B, and Mistral-7B, we observed that increasing the number of examples (n-shot prompting, $n \geq 1$) generally leads to significant improvements in BERT scores (Precision, Recall, and F1) for pretrained LLMs compared to a 0-shot setting. Notably, a fine-tuned Mistral-7B model (Mistral-7B FT), even when fine-tuned on a different dataset, achieved the best overall F1 score of 0.9671 with 5-shot prompting. This fine-tuned model also significantly improved performance in a 0-shot setting compared to its non-fine-tuned counterpart, indicating its adaptability to new datasets for the SPARQL2Q task. The query enhancer component within our framework has been particularly effective in significantly improving the quality of generated queries. For example, the average Acc@10 score (percentage of correct answers out of 10 runs) increased from 0.22 (without enhancer) to 0.57 (with enhancer) across a set of 30 questions. For specific complex questions, the Acc@10 improved from zero to 0.9 and 1.0 respectively with the enhancer.

7. Real-World Evaluation

The GLACIATION framework is validated across three diverse, real-world pilot deployments, demonstrating its practical applicability in industrial and public sector settings. These use cases serve to test and measure the effectiveness of the platform’s core capabilities, including energy-aware workload distribution, privacy-preserving data orchestration, and semantic policy enforcement.

The pilots showcase the integrated functionality of the GLACIATION approach. At the Independent Power Transmission Operator (IPTO) in Greece, the system utilizes the *Green Index* derived from real-time SCADA data to perform energy-aware workload distribution for grid anomaly detection tasks across three interconnected data centers [6]. Once computed, its values, along with their temporal and spatial context, are represented in the DKG as RDF triples, following the GLC-MRM ontology.

In a smart manufacturing scenario in Ireland, the platform manages privacy-preserving orchestration using data from robotic and IoT sensors, enforcing data sovereignty and access control policies to govern data movement and computation. A third pilot with MEF/SOGEL in Italy focuses on optimizing data movement and energy consumption in a public administration context. Across these deployments, the Open Policy Agent (OPA) framework is utilized to implement semantic policy control, with real-time dashboards providing visibility into performance and compliance [7].

7.1. Example Pilot - Green-Aware Orchestration at IPTO

To facilitate green-aware workload placement, GLACIATION partner IPTO introduces the *Green Index*, a supply-side metric that quantifies the availability of sustainable energy at potential execution sites. The index is designed to be comparable across different locations and times, enabling the orchestrator to dynamically steer computational tasks toward sites and times with higher green energy availability, thereby minimizing environmental cost. It is derived directly from real-time Supervisory Control and Data Acquisition (SCADA) telemetry data provided by the Transmission System Operator (TSO), which monitors the power flow at grid substations.

The calculation of the Green Index begins with the net power balance (P_{net}) at a substation, indicating whether the site is a net exporter or importer of energy. This local value is then combined with the national ratio of renewable energy sources (r_{RES}) to produce a RES-adjusted power balance, R . To

ensure comparability, this value is normalized using a signed percentile rank over a one-week historical window and scaled to a final range of 0 to 1. The resulting index is strongly correlated with national CO₂ emissions, where a higher index value (approaching 1) signifies greater availability of green energy and signals optimal conditions for executing workloads.

8. Conclusion

This paper presented the GLACIATION framework, a platform for energy-optimized and semantically-aware data management. We demonstrated how its integrated technologies—a Distributed Knowledge Graph (DKG) based on the GLC-MRM ontology, an AI-driven scheduling engine using Ant Colony Optimization, and the Green Index—create an intelligent orchestration platform.

The successful validation highlights the potential of semantic technologies to solve complex optimization problems in distributed environments. The DKG, paired with a lightweight LLM interface, is a significant step toward making sophisticated data infrastructures more autonomous, efficient, and accessible. Future work will focus on scaling the AI models, optimizing data movement strategies for heterogeneous networks, and improving human-in-the-loop support for domain experts.

Acknowledgments

This work is funded by the European Union’s Horizon Europe research and innovation programme under grant agreement No 101070141 (GLACIATION).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] GLACIATION Consortium, GLACIATION Metadata Reference Model (GLC-MRM) v1.1.0, 2024. URL: <https://glaciation-project.eu/MetadataReferenceModel/1.1.0/>, accessed: 2025-07-03.
- [2] T. Wamhof, A. Bernardi, D. Martini, M. Leinberger, A. Sinha, H. Tapken, A. Schliebitz, H. Graf, Metadata management and asset exchange in the agricultural data ecosystem of the project agri-gaia, *Datenbank-Spektrum* 23 (2023) 107–115.
- [3] H. Hamann, et al., Ant-search algorithm for distributed knowledge graphs, in: *Swarm Intelligence: 14th International Conference, ANTS 2024, Konstanz, Germany, October 9–11, 2024, Proceedings*, volume 14987, Springer, 2024, p. 243.
- [4] N. Kouka, V. Piuri, P. Samarati, Tasks scheduling with load balancing in fog computing: a bi-level multi-objective optimization approach, in: *Proceedings of the Genetic and Evolutionary Computation Conference, 2024*, pp. 538–546.
- [5] G. Piao, M. Mountantonakis, P. Papadacos, P. Sonawane, A. OMahony, *Toward exploring knowledge graphs with llms* (2022).
- [6] O. Vantzou, D. Skipis, E. Chassioti, I. Moraitis, Green index: A measure of locally available green energy, in: *Proceedings of the 25th International Conference on Environment and Electrical Engineering (EEEIC), 2025*. Presented at the 25th International Conference on Environment and Electrical Engineering (EEEIC 2025).
- [7] S. Paraboschi, M. Abbadini, J. Böhler, F. Capano, S. De Capitani di Vimercati, D. Facchinetti, S. Foresti, G. Livraga, G. Oldani, M. Rossi, P. Samarati, Deliverable D4.1 – Policies and Techniques for Data Protection in Modern Distributed Environments, Technical Report 101070141, GLACIATION Consortium, 2023. EU Horizon Europe Project GLACIATION.