

Project name:
Cardano Analytics Platform (CAP)

Project URL:
<https://milestones.projectcatalyst.io/projects/1300034>

Project ID:
1300034

Team:
Marcio Moreno, PhD
(marciofmoreno@gmail.com)

Rafael Brandão, PhD
(rafael@mobr.ai)

Date project started:
January 20th, 2025

Date project completed:
December 10th, 2025

1. Introduction

The Cardano Analytics Platform (CAP) is designed to simplify access to Cardano blockchain data through natural language queries powered by large language models (LLMs). Funded under Project Catalyst F13 (Project #1300034), CAP addresses the challenge of accessing and structuring on-chain data in the Cardano ecosystem by building an innovative analytics platform integrated with large language models (LLMs). This initiative aimed to support users of varying technical expertise to interact with blockchain data through natural language queries, thereby streamlining decision-making processes and enhancing overall ecosystem engagement. The development process followed a structured milestone-based approach, ensuring systematic progress from foundational research to prototype deployment. All specified outputs have been made fully open-source, allowing for ongoing community contributions and adaptations. This report outlines the work completed, and the tangible achievements realized throughout the project.

2. Challenge KPIs

The Cardano Analytics Platform (CAP) project tackled several key challenges in blockchain analytics and LLM integration, with success measured through the development of a functional prototype that achieved factual accuracy for the designed use cases, supporting natural language access, efficient data handling, and overall feasibility validation via testing and open-source release. These challenges were addressed systematically over 11 months, resulting in a platform that mitigates limitations in LLMs and knowledge graphs, streamlines on-chain data access, and handles domain-specific data issues.

Challenge KPI	How It Was Addressed
Factual consistency and domain-specific knowledge in LLMs, leading to hallucinations	Integrated LLM with a formal ontology and knowledge graph to ensure responses are grounded in accurate Cardano blockchain data and relationships, reducing inconsistencies through structured reasoning.
Requirement for specialized query languages in knowledge graphs, lacking natural language flexibility	Combined knowledge graphs with LLMs to enable natural language inputs and outputs, allowing users to interact via everyday queries translated into SPARQL while maintaining formal querying precision.
Accessing and structuring on-chain data in the Cardano ecosystem	Developed data extraction and ETL mechanisms, and a knowledge graph using Virtuoso triplestore, automating ingestion from cardano-node and PostgreSQL database to transform raw data into RDF triples aligned with the ontology.

3. Project KPIs

CAP met its key performance indicators by delivering a fully functional LLM-powered prototype for blockchain analytics, including a detailed ontology, knowledge graph, ETL pipelines, natural language query engine, user interface for visualizations, and comprehensive testing that validated feasibility with success across defined use

cases, ultimately releasing the open-source platform on GitHub under GPLv3 to promote community collaboration and reduce technical barriers in the Cardano ecosystem.

Project KPI	Achieved Outcomes
Ontology Development	Created a comprehensive ontology extending general blockchain concepts with Cardano-specific elements like EUTXO, native tokens, Ouroboros protocol, and governance frameworks, documented with Protégé tool and available online.
Knowledge Graph Bootstrapping	Established a functional knowledge graph using Virtuoso triplestore, integrated with API for querying, and demonstrated accurate RDF-triple representations from initial Cardano data extraction.
ETL Pipeline Implementation	Refined extractors and transformers for entities like blocks, transactions, and governance actions, optimized with batch processing and error-handling to ingest real-time and historical data into the knowledge base.
LLM Query Engine Development	Built an engine translating natural language queries to SPARQL, supported by API endpoints and utilities for result conversion to key-value pairs and Vega visualizations, enabling complex queries like large ADA transactions.
User Interface and Dashboard Creation	Designed an interface for composing custom dashboards with widgets linked to queries, incorporating tools for visualizations such as bar, pie, and line charts.
Testing and Validation	Conducted unit, integration, and end-to-end tests covering ETL, SPARQL, and NL pipelines, and public open-source release.

4. Key Achievements and Impact

The Cardano Analytics Platform (CAP) project achieved several key milestones over its 11-month duration, culminating in a fully functional open-source prototype that integrates large language models (LLMs) with knowledge graphs to enable natural language queries, insightful analytics, and dashboard visualizations of Cardano on-chain data. Notable accomplishments include the development of a comprehensive ontology modeling Cardano-specific elements like the Extended UTXO model, native tokens, and governance frameworks; the implementation of ETL pipelines for efficient data extraction, transformation, and loading into an RDF triplestore; and the creation of an LLM-based query engine that translates user prompts into SPARQL queries, delivering accurate responses with utilities for data conversion and Vega-based charts. Released publicly on GitHub under GPLv3, CAP reduces technical barriers for non-experts, broadens ecosystem participation by enabling users to access on-chain insights without specialized knowledge, while supporting community collaboration, transparency, and innovation within the Cardano network.

5. Key Learnings

The project key learnings relies on the integration of large language models (LLMs) with knowledge graphs for blockchain analytics, highlighting the effectiveness of combining ontological reasoning with natural language processing to overcome LLM hallucinations and ensure factual accuracy through structured data querying. A

primary insight was the critical role of a well-defined ontology, enabling semantic clarity and formal reasoning while addressing challenges in domain-specific knowledge representation. The development process underscored the importance of ETL pipelines for handling high-volume, and real-time blockchain data to compose a knowledge graph capable of contextualizing CAP's LLM, which are essential for maintaining data integrity in analytics platforms. The development process underscored the importance of robust ETL pipelines for handling high-volume, real-time blockchain data, including custom utilities to manage issues such as large integer preservation and hexadecimal decoding, which are essential for maintaining data integrity in analytics platforms. However, the ETL implementation in Python using cardano-db-sync proved inefficient, as the relational database structure demanded numerous joins and repeated consultations to derive the most appropriate RDF representations, leading to performance issues. This prompted us to initiate efforts designing a C++ with Ogmios alternative, for more direct and performant data extraction. CAP's modular architecture proved vital for scalability, maintainability, and secure deployment, with API and service modules facilitating seamless user interactions and performance monitoring. Extensive testing across unit, integration, and end-to-end scenarios revealed the need for comprehensive validation, emphasizing empirical evidence in proving solution feasibility. Finally, the open-source release under GPLv3 demonstrated transparency and innovation by reducing technical barriers and enabling broader participation in the Cardano ecosystem.

6. Importance

This project is important because it eases access to Cardano's on-chain data, bridging the gap between sophisticated blockchain analytics and everyday users by mitigating LLM hallucinations with knowledge graphs, thereby streamlining decision-making, enhancing transparency in transactions, governance and staking, and supporting greater ecosystem engagement. The Cardano community should be excited about CAP's accomplishments as it enables non-experts to explore real-time insights without technical skills, promoting innovation, collaboration, and informed participation that can drive the network's growth and inclusivity.

7. Next Steps

Following the completion of the Cardano Analytics Platform (CAP) prototype, the next steps focus on transitioning from proof-of-concept to a production-ready solution through productization efforts, including enhanced deployment options, scalability improvements, and broader integration with Cardano's evolving ecosystem. The team envisions incorporating off-chain analytics to complement on-chain data, providing richer insights for users, while fine-tuning the LLM model on specialized Cardano query datasets to boost accuracy and reduce hallucinations. Community collaboration will be pivotal, with the open-source GitHub repository under GPLv3 inviting contributions to expand the ontology, refine ETL pipelines for real-time efficiency, and develop new use cases beyond the initial ten. Additionally, ongoing testing and validation will ensure robustness, with potential expansions in multilingual support, advanced visualizations, and API enhancements to further lower barriers for developers, researchers, and enthusiasts, ultimately driving greater transparency, engagement, and innovation in blockchain analytics.

8. Final Remarks

The Cardano Analytics Platform represents a significant advancement in facilitating access to blockchain data analysis. By combining a domain ontology with natural language processing capabilities, it bridges the gap between technical blockchain data and meaningful insights accessible to a wide range of stakeholders. The platform's architecture introduces several key innovations for the Cardano community. The use of semantic technologies enables rich context-aware queries that understand the relationships across different aspects of the Cardano ecosystem. The LLM-powered query system enhances accessibility, allowing users without technical

expertise in programming languages or blockchain internals to perform complex analytics. In addition, the modular design supports continuous improvements to both the ontological model and analytical capabilities as the Cardano ecosystem evolves. This project has fulfilled its objectives, delivering an LLM-powered analytics prototype that can transform how users interact with Cardano on-chain data. Through development across all milestones, the prototype stands as an open-source tool that enhances accessibility and usability. The achievements underscore the project's success in addressing key challenges in blockchain analytics, paving the way for future enhancements and community-driven expansions. This report marks the end of the Fund 13 phase, with the source code now available for widespread adoption, contribution and further development to achieve productization.

9. References

Close-out Video:

https://www.youtube.com/watch?v=nRsa_qiGhN0

Deployed Prototype:

<https://cap.mobr.ai>

GitHub Repository:

<https://github.com/mobr-ai/cap>

Next step productization (Fund 15 proposal):

<https://app.projectcatalyst.io/proposal/019aa15b-0cf6-7699-8207-542200aef2ba?version=019ac167-a02f-7a60-8d74-d0de5ef5dba7>

Technical Report:

https://github.com/mobr-ai/cap/blob/abb8a0fa51f12e1e067f32631fb447da855f2b7f/documentation/pdf/technical_report.pdf

Project Execution Report:

https://github.com/mobr-ai/cap/blob/b70b19387dcf929b5022c2f09a9f436f51017364/documentation/pdf/project_execution_report.pdf

Company Website:

<https://mobr.ai>