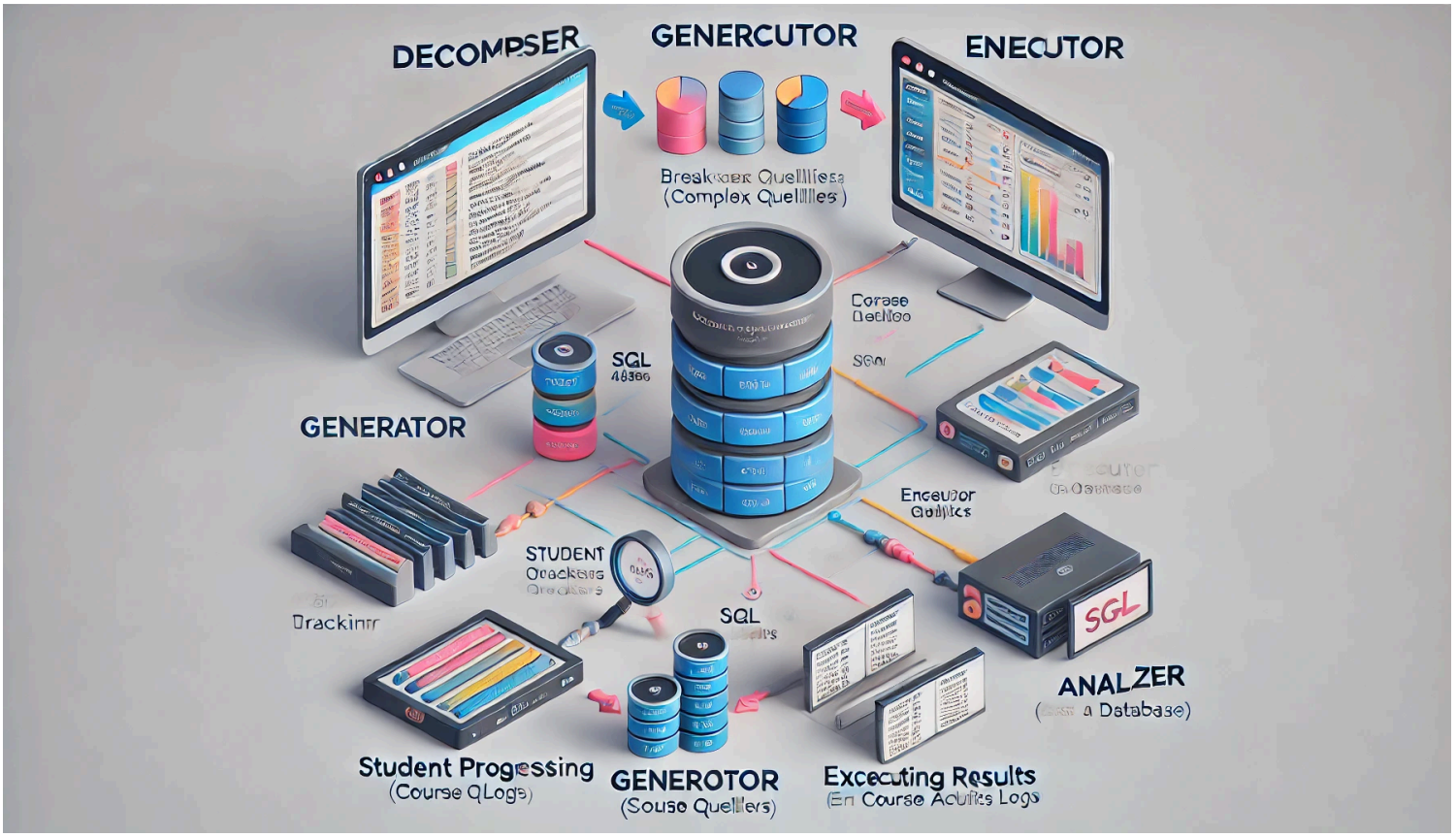


Proof of Concept Report

E-Learning Tracking System

1. Executive Summary

This Proof of Concept (POC) demonstrates an intelligent e-learning tracking system capable of processing natural language queries about student progress and course engagement. The system decomposes complex queries, generates SQL, executes them, and provides detailed insights and analysis.

2. Objectives

- Develop a system that understands and processes natural language queries about e-learning data.
- Implement a modular architecture separating concerns between query decomposition, SQL generation, execution, and analysis.
- Provide detailed insights and analysis of student progress and engagement.
- Create a user-friendly interface for interaction.
- Validate the feasibility of automated e-learning tracking.

3. Architecture

3.1 Core Components

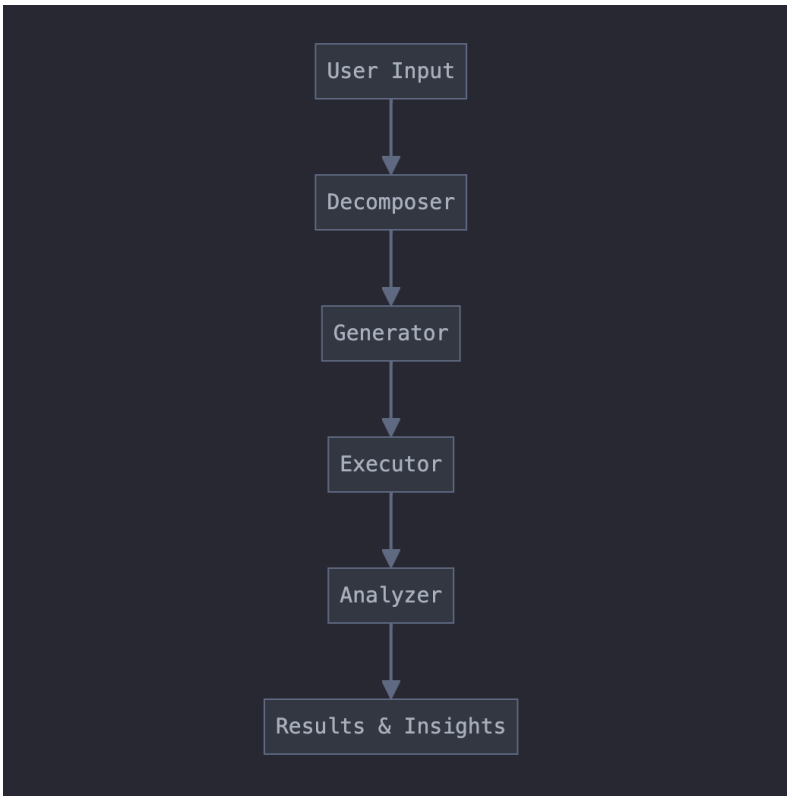
The system comprises four primary components:

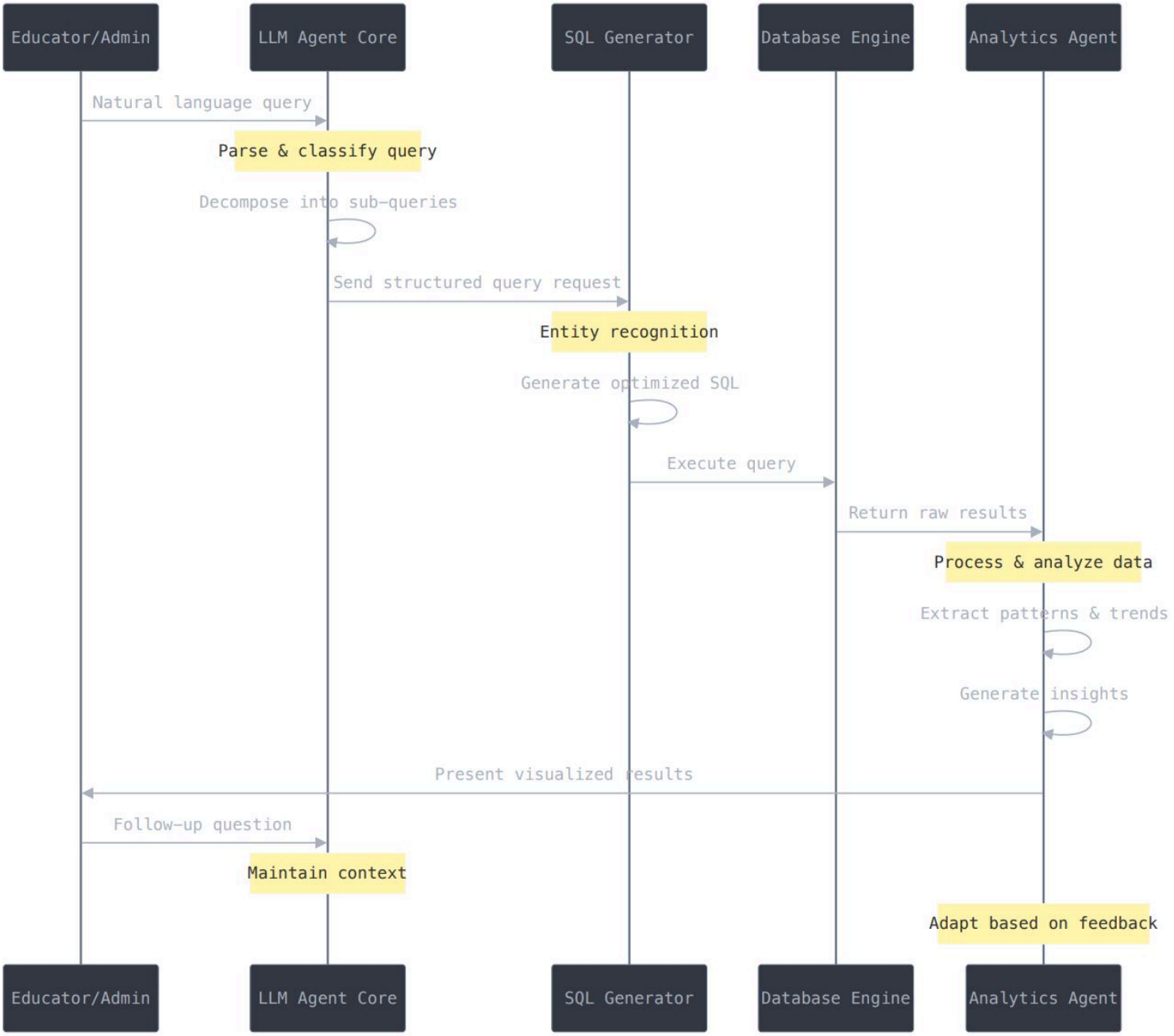
1. **Engine**
    a. **Decomposer (Sonnet)**
        - Breaks down complex queries into simpler sub-queries.
        - Select appropriate tables for sub-query execution.
        - Extracts relevant entities and matches them to the database schema.
    b. **Generator (Sonnet)**
        - Generates SQL queries from decomposed sub-queries.
        - Handles query optimisation and validation.
        - Manages query parameters and conditions.
    c. **Executor (In-House Execution)**
        - Executes SQL queries against the database.
        - Handles error conditions and query timeouts.
        - Manages database connections and transactions.
    d. **Optional: Analyzer (Haiku)**
        - Processes query results.
        - Generates insights and trends.
        - Provides detailed analysis of student engagement and performance.
2. **Metadata**
    a. Defines relationships and structure of the database.
    b. Guides entity extraction and SQL generation.
    c. Ensures consistency in query handling.
3. **Utils**
    a. **Search**
        - Implements search, entity extraction, and matching mechanisms.
        - Uses fuzzy matching techniques for improved accuracy.
    b. **Config**
        - Centralised configuration for managing system parameters.
        - Handles database settings, API configurations, and processing rules.
4. **Orchestrator**
    a. Builds the system into a graph-based structure.
    b. Converts the query processing pipeline into a structured chain of operations.
    c. Manages dependencies and execution flow.
        - **Query & Prompt Caching**
            1. Implements caching mechanisms for query results.
            2. Enables follow-up question capabilities by storing previous interactions.
            3. Reduces redundant query execution and enhances performance.

3.2 Data Flow

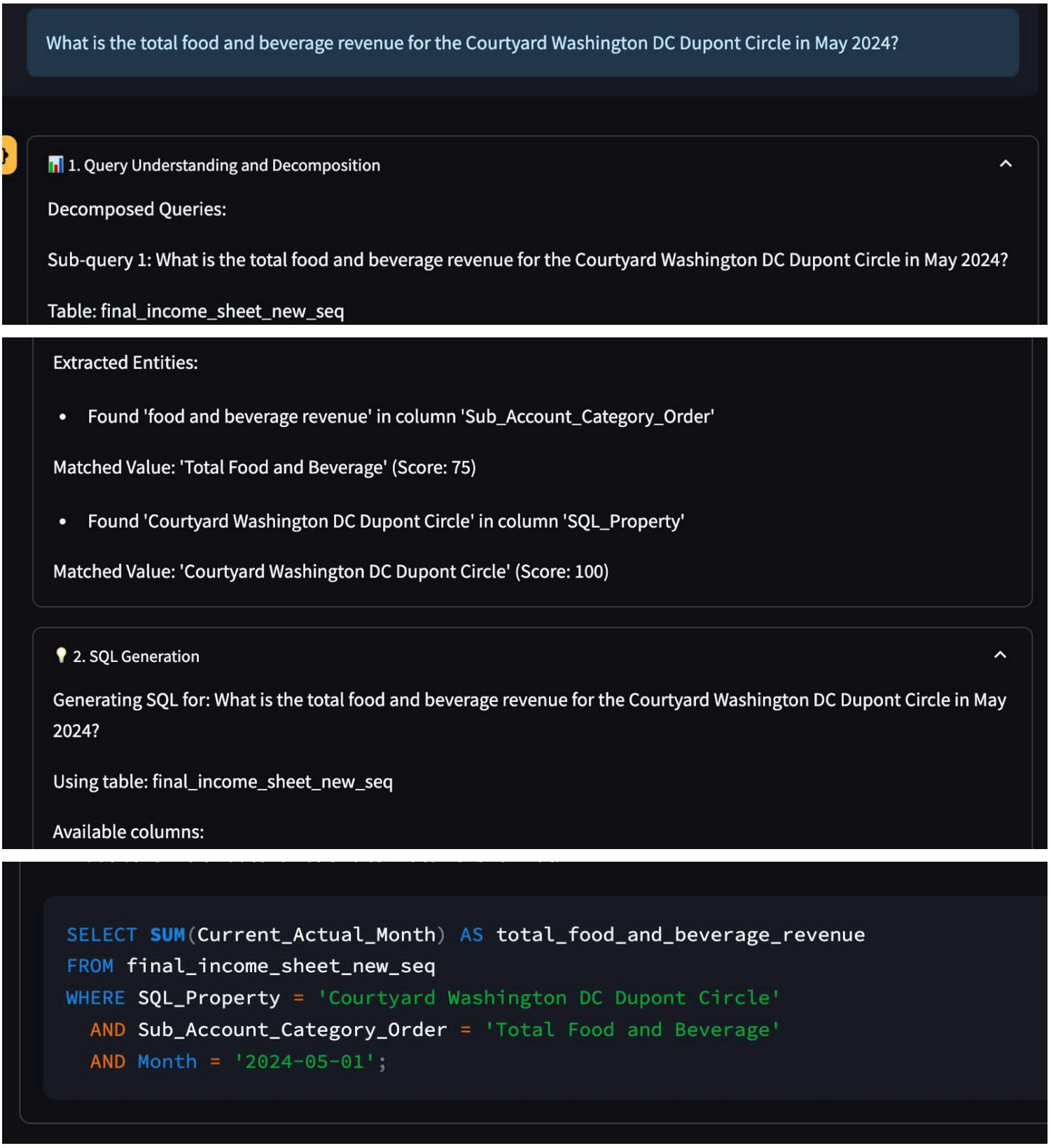
The system follows a structured query processing pipeline:

1. Query Understanding
2. Entity Extraction
3. SQL Generation
4. Query Execution
5. Results Analysis


User Input
Decomposer
Generator
Executor
Analyzer
Results & Insights


Educator/Admin
LLM Agent Core
SQL Generator
Database Engine
Analytics Agent
Natural language query
Parse & classify query
Decompose into sub-queries
Send structured query request
Entity recognition
Generate optimized SQL
Execute query
Return raw results
Process & analyze data
Extract patterns & trends
Generate insights
Present visualized results
Follow-up question
Maintain context
Adapt based on feedback
Educator/Admin
LLM Agent Core
SQL Generator
Database Engine
Analytics Agent

3.3 UI & User Interaction



⚡ 3. Query Execution

SQL validation passed

Execution successful: 1 rows returned

Results Preview:

total_food_and_beverage_revenue
23,116

📈 4. Results Analysis

Analysis Results:

Success: True

Sub-query count: 1

Total result count: 1

Key Findings:

📊 Summary: The total food and beverage revenue for the Courtyard Washington DC Dupont Circle in May 2024 is $23,116.

Key Findings:

📊 Summary: The total food and beverage revenue for the Courtyard Washington DC Dupont Circle in May 2024 is $23,116.

💡 Insights: ['The food and beverage revenue for this property and time period is a significant amount, indicating it is an important revenue stream for the hotel.', 'The data provides a specific, quantifiable figure for the food and beverage revenue, which can be used for financial planning and analysis.', 'The data is limited to a single month and property, so further analysis would be needed to understand the broader trends and performance.']

📈 Trends: ["This data point represents a single month's revenue, so it is difficult to identify any meaningful trends without additional data points over time.", 'To understand trends, it would be helpful to have food and beverage revenue data for multiple months or years for this property, as well as comparative data for other similar properties.']

🎯 Implications: ["The high food and beverage revenue suggests this is an important area of focus for the hotel's operations and profitability.", "The data could be used to inform pricing, menu planning, staffing, and other operational decisions related to the hotel's food and beverage offerings."]

🔗 Relationships: ['The data is linked to the specific property, Courtyard Washington DC Dupont Circle, and the time period of May 2024.', "The data is categorized under the 'Total Food and Beverage' sub-account, indicating it represents the combined revenue from all food and beverage sources at the hotel."]

4. Testing Strategy

4.1 Test Cases for Validating System Results

To ensure the accuracy and robustness of the system, a comprehensive suite of test cases will be created. These tests cover various aspects of data retrieval, entity analysis, and learning analytics.

Data Retrieval Queries

- General information or insights without specific constraints.
- Data retrieval for specific dates or date ranges.
- Queries involving multiple conditions for precise data retrieval.
- Comparison of data across different entities or periods.
- Analysing trends over time and handling ambiguous data interpretations.
- Summarising data through functions like SUM, AVG, COUNT, etc.

Entity and Scenario Analysis

- Identifying and extracting specific entities from the data.
- Exploring potential outcomes of hypothetical situations.
- Identifying distinct values within a dataset.
- Checking the validity and integrity of data or results.
- Testing the limits and boundaries of the system's capabilities.

5. Performance & Cost Considerations

The system will leverage **Claude LLM** for query processing and analysis. Based on past results:

- Each analysis can be completed within **5 seconds**.
- The estimated cost per analysis is **USD 0.05**. This ensures a balance between efficiency, cost-effectiveness, and real-time insights for e-learning tracking.
- As the database schema, dataset size, and method of interaction with the database have not been provided, actual response times may vary.
    - Factors affecting performance include:
        - Complexity of the database schema and relationships between entities.
        - Method of database query execution (local execution, server-based execution, etc.).
        - Number of distinct values and clear separation of entities within the dataset.

6. Timeline and Milestones

1. **Phase 1: Setup & Core Development** (3 weeks)
    - Environment setup, core components (engine, utils, metadata, testing_system, and UI) development, basic testing framework.
2. **Phase 2: Validation & Testing** (2 weeks)
    - Component integration, test suite development, performance optimisation.
3. **Phase 3: Refinement & Documentation** (2 weeks)
    - Bug fixes, documentation, final testing, and deployment.

This POC validates the feasibility of an intelligent e-learning tracking system. The modular architecture ensures scalability, while test cases ensure reliability. Performance and cost considerations will depend on database complexity, schema structure, and execution method.