Enterprise Multi-Database Natural Language Query Engine

Version 1.0

Prepared for

myOnsite Healthcare, LLC.

Document Control

Rev. No.	Description of Change	Effective Date
1.0	Initial Release	11st Aug 2025

Authored By

Name	Role	Signature	Date
Het	Team Lead		11st Aug 2025

Reviewed and approved By

Name	Role	Signature	Date

Enterprise Multi-Database Natural Language Query Engine

Advanced Text-to-SQL System with Dynamic Schema Adaptation

Project Overview

Build a production-grade, multi-database natural language query engine that can understand complex business questions, generate optimized SQL across different database systems, handle real-time schema evolution, implement advanced security controls, and provide enterprise-level governance and observability.

Time Allocation: 5 hours
Complexity Level: Principal Engineer Challenge
Focus Areas: Advanced NL-to-SQL, multi-database orchestration, real-time schema adaptation, enterprise security

---

System Architecture Requirements

You're building an enterprise query engine that must:

- Support **multiple database systems** (PostgreSQL, MySQL, SQLite, MongoDB) with dialect-specific optimizations
- Handle **complex natural language queries** including temporal reasoning, multi-table joins, and business logic inference

- Implement **real-time schema introspection** with automatic adaptation to structural changes
- Provide **advanced security controls** including SQL injection prevention, role-based access, and query governance
- Support **concurrent query execution** with intelligent caching and performance optimization
- Include **comprehensive audit logging** and explainable AI for query reasoning
- Handle **ambiguous queries** with clarification mechanisms and confidence scoring

---

Database Environment

Multi-Database Setup

You'll work with **4 interconnected databases**:

1. **PostgreSQL** (Primary OLTP)
    - **Tables:** customers, orders, products, inventory, suppliers, categories
    - **Complex features:** Foreign keys, indexes, triggers, stored procedures, views
    - **Data volume:** ~100K customer records, ~500K order records
    - **Advanced structures:** JSON columns, arrays, custom types, partitioned tables
2. **MySQL** (Analytics/Reporting)
    - **Tables:** sales_analytics, customer_segments, product_performance, regional_data
    - **Features:** Window functions, CTEs, materialized views
    - **Data volume:** ~1M aggregated records with time series data