



# Modular GenAI via Model Context Protocol

TIAA Hackathon 2025

By: Mit Goswami

# The Problem We're Solving: The Data Divide

In today's enterprise, critical information is often trapped and siloed. We face:

## Fragmented Data Access

Data resides in disparate systems – structured databases like MySQL and unstructured knowledge bases like Confluence.

## Manual, Time-Consuming Reporting

Generating SQL queries and reports is a manual, error-prone process that drains valuable time and resources.

## Lack of Intelligent Automation

A significant gap exists in context-aware automation, limiting efficiency and real-time decision-making.

# Our Vision: Unified, Context-Aware AI

We envision a future where data access is seamless and intelligent.



## Modular MCP Servers

Decoupling data tasks into specialized, modular Model Context Protocol (MCP) servers.



## Intelligent Task Routing

Routing tasks intelligently using dynamic prompts, defined tools, and specific resources.



## Bridging Diverse Systems

Seamlessly connecting Large Language Models (LLMs), relational databases, vector search, and enterprise knowledge bases.

# MCP Architecture: Two Dedicated Servers

Our solution leverages a dual-server architecture for clear boundaries and optimized performance.

## MYSQL\_APP: Structured Data Master

Dedicated to handling structured SQL data queries and generating comprehensive reports.

- Manages relational database interactions.
- Specializes in report generation based on structured queries.
- Isolated tools, resources, and prompts for SQL operations.

## RAG\_APP: Unstructured Knowledge Hub

Focuses on Retrieval-Augmented Generation (RAG) from unstructured knowledge bases and vector operations.

- Retrieves information from Confluence
- Performs vector operations for efficient knowledge retrieval.
- Independent tools and prompts for RAG-specific tasks.

This design ensures each server operates with isolated tools, resources, and prompts, preventing overlap and enhancing efficiency.

# Robust Tech Stack for Seamless Integration

Our architecture is built on a foundation of cutting-edge technologies to deliver powerful and flexible solutions.



## LangChain

Orchestrates intelligent tool invocation and prompt management.



## OpenAI + Claude

Leveraging leading Large Language Models for advanced natural language understanding and generation.



## Qdrant

High-performance vector store integrated with HuggingFace embeddings for efficient semantic search.



## Confluence Cloud

Our primary Enterprise Wiki for unstructured knowledge retrieval and collaboration.



## MySQL

Robust Relational Database for structured data storage and querying.

# What Makes MCP Innovative?

Our approach introduces several key innovations that set us apart:



## FastMCP Modular Orchestration

Utilizes FastMCP for highly efficient and scalable LLM orchestration, enabling dynamic interactions between models and tools.



## Context-Aware Prompting

Intelligent prompts dynamically activate the correct tools and resources based on user intent and contextual understanding.



## Seamless SQL & RAG Automation

A unified system that elegantly combines structured SQL querying with unstructured RAG capabilities for comprehensive data insights.



## Decoupled, Independent Servers

MYSQL\_APP and RAG\_APP operate independently, ensuring resilience and modularity, yet seamlessly integrate for hybrid workflows.

# Example Workflow: Hybrid Data Retrieval

Let's walk through a typical scenario showcasing the power of MCP:

01

## 1. Natural Language Query

A user asks a question about MySQL data in natural language (e.g., "Show me Q4 sales figures for last year").

02

## 2. Schema Analysis & SQL Generation

Claude, via MYSQL\_APP, checks the MySQL schema (a defined resource) and intelligently generates the appropriate SQL query.

03

## 3. Data Retrieval & Report Creation

The SQL query is executed, and MYSQL\_APP transforms the results into a rich, formatted report within Confluence.

04

## 4. Knowledge Base Integration (Optional)

RAG\_APP automatically vectorizes the newly created Confluence report, indexing it in Qdrant for future semantic search.

05

## 5. Intelligent Answer Retrieval

When the user asks follow-up questions, RAG\_APP uses Qdrant to retrieve relevant information from both structured and unstructured sources, providing a comprehensive answer.

# Thank You!

We are incredibly excited about the potential of Modular GenAI via Model Context Protocol to revolutionize how TIAA interacts with its data.

## Bringing GenAI to Data

Our solution bridges the gap between advanced AI capabilities and complex enterprise data landscapes.



## Looking Forward to Your Feedback

Your insights are invaluable as we continue to refine and expand this powerful protocol. Q&A session starts now!