**GenAI Engineer: Technical Assessment - Retrieval-Augmented Generation (RAG) System for Salesforce**

## 1. Introduction

This assessment evaluates your ability to design and implement a Retrieval-Augmented Generation (RAG) system. The objective is to build a solution that provides Question & Answer (Q&A) and summarization capabilities based on Salesforce's quarterly earnings presentation transcripts. You will be required to present your solution in a walkthrough session.

**Note:** This assessment is a composite of requirements from various Altimetrik clients and uses Salesforce as a representative example.

## 2. Assessment Details

- **Estimated Completion Time:** 4 hours (flexible scheduling)
- **Data Set:** Publicly available Salesforce quarterly earnings presentation transcripts. Download them here: Salesforce Earnings Call Transcripts

## 3. Project Goal

Develop a fully functional RAG system that integrates a Large Language Model (LLM) and a Vector Database to enable Q&A and summarization of Salesforce's earnings call transcripts through a Conversational User Interface.

## 4. Key Requirements

- **RAG System:**
  - Integrate an LLM to enhance output with retrieved documents from the Vector Database.
  - Select a technology stack that includes an LLM, a Vector Database, and a Conversational UI.
- **System Capabilities:**
  - **Question Answering (Q&A):** Provide concise, accurate answers to user questions.
  - **Summarization:** Generate summaries of key points, specific topics, or trends.
- **Example Questions:**
  - "When was the most recent earnings call?"
  - "What are the risks that Salesforce has faced, and how have they changed over time?"
  - "Can you summarize Salesforce's strategy at the beginning of 2023?"
  - "How many earnings call documents do you have indexed?"
  - "How many pages are in the most recent earnings call?"
- **Technical Specifications:**
  - **Vector Database:** Choose a vector database (e.g., Pinecone).

    - **LLM:** Select an appropriate LLM (e.g., OpenAI models).
    - **Conversational UI:** Develop a simple interface.
- **Data Handling:** The system must effectively handle both structured (tables, summaries) and unstructured (raw text) data.

**5. Business Case: Salesforce Earnings Analysis**

Salesforce needs an AI system to help analysts quickly understand key information from quarterly earnings transcripts. This system should:

- Summarize key points from earnings presentations.
- Provide direct answers to questions about company performance, strategies, and risks.

**6. Additional Considerations**

- **Scalability:** Describe how the system can be scaled for larger datasets.
- **Design Decisions:** Be prepared to discuss your architectural choices and integration methods.
- **Deployment (Optional):** Explain how you would deploy the system (e.g., Docker, cloud platforms).

**7. Evaluation Criteria**

- **Architecture & Design:** Clarity and effectiveness of the system architecture.
- **Functionality:** Accuracy of Q&A and summarization.
- **Completeness:** Fulfillment of core requirements within the time frame.
- **Presentation:** Ability to explain and demonstrate the solution clearly.
- **Code Quality:** Clean, well-documented code.

**8. Deliverables**

1. **Code Repository:** Provide a link to a repository (e.g., GitHub) with all code, instructions, and scripts.
2. **Documentation:** Include a README file with setup and usage instructions.
3. **Architecture Diagram:** (Optional) A visual representation of your system design.

**9. Walkthrough Session**

You will be invited to a session to present your solution, discuss your design, and demonstrate its functionality.