

GenAI-Powered SQL Refactoring Assistant

SQLite → HiveQL Migration

Milestone 1

Presented by Group 3: Ajay, Madhavan R Mohan, Sanjay Rajesh Manwani, Senthilkumar N, Veral Sharma

WHY THIS MATTERS?

The Problem We're Solving

Businesses migrating from **SQLite** → **Hive** face:

Manual Refactoring

Manual refactoring of 1000s of SQL queries

Syntax + Semantic Differences

Date functions (strftime → from_unixtime)

Typing Issues

Dynamic → static typing inconsistencies

Quote Handling Issues

Incompatible quote handling between dialects

- ✖ Current solution: Manual = Error-prone + Slow



Our Solution

INPUT

SQLite query

OUTPUT

Functionally equivalent HiveQL query

Guarantees

- Same result set guaranteed
- RAG for syntax compliance

Architecture

Scalable for future dialects (MYSQL → SparkSQL, etc.)



Technical Approach

RAG + LLM Pipeline



RAG

Syntax/docs retrieval

Few-shot

Accuracy

Chain-of-Thought

Reasoning

To test across: **GPT, Claude, Gemini, Deepseek**



KEY PAPERS ANALYSED

Literature Review



SQL-GEN

NL2SQL limitations



CrackSQL

Rule+LLM (no RAG → hallucinations)



MALLET

RAG-based (our base)



RISE

Chain-of-Thought (no RAG)



PARROT

Evaluation benchmark



Our edge: RAG + CoT + Few-shot



TWO-LAYER EVALUATION

Evaluation Strategy

Layer 1: Execution

- Docker: SQLite + Hive
- Compilation success
- Data equivalence

Layer 2: CodeBLEU

- AST similarity
- Structural quality

Dataset

PARROT (SQLite→Hive pairs)



INNOVATION OPPORTUNITIES

Research Gaps We're Addressing

Current research ignores:

- Semantic mismatches
- Function transformations
- Typing inconsistencies
- Distributed execution needs

Our Fix: Schema-aware RAG

Dynamic metadata + function docs

 **TEAM MILESTONE 1**

Team Contributions

**Ajay**

Problem Definition & Objectives

**Senthil**

Literature Review

**Sanjay**

Baselines & Evaluation

**Veral**

Gaps & Opportunities

**Madhavan**

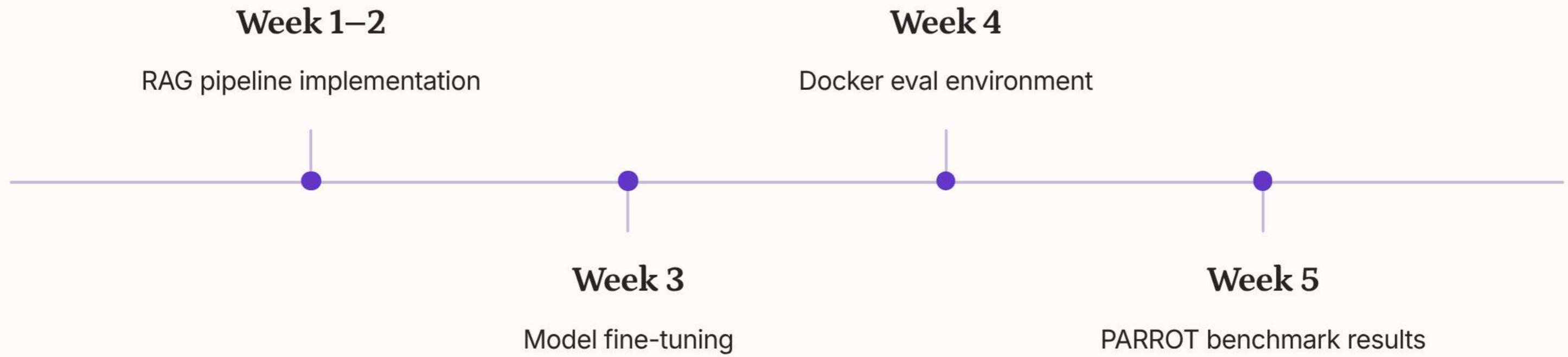
Consolidation + Final Report

**Milestone 1: COMPLETE**



MILESTONE 2 ROADMAP

Next Steps (Milestone 2)



- Target: Beat PARROT baseline scores**

THANK YOU!!!

Presented by Group 3: Ajay, Madhavan R Mohan, Sanjay Rajesh Manwani, Senthilkumar N, Veral Sharma