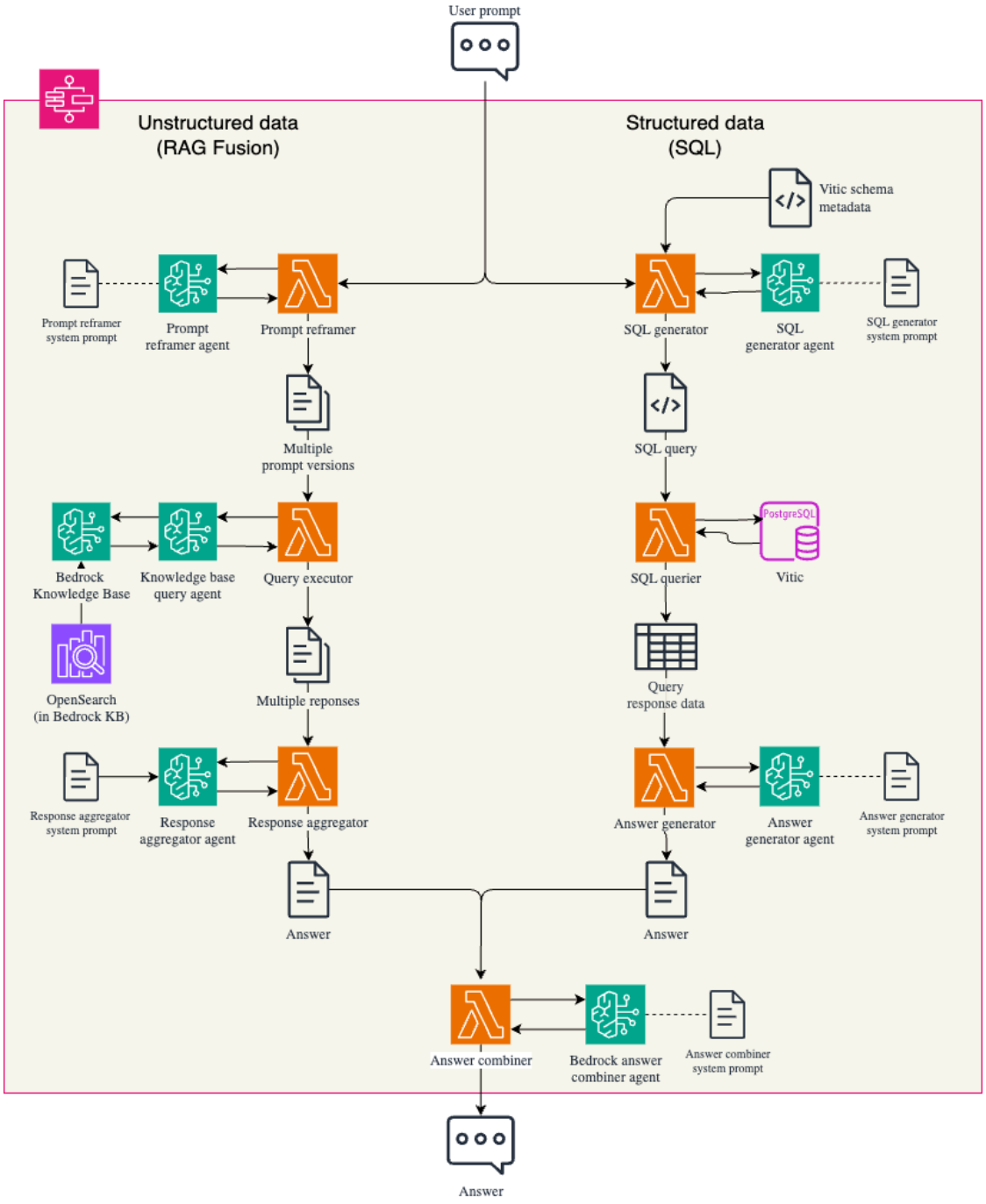
Initial concept for system architecture

Architecture diagram

Query Execution - RAG Fusion

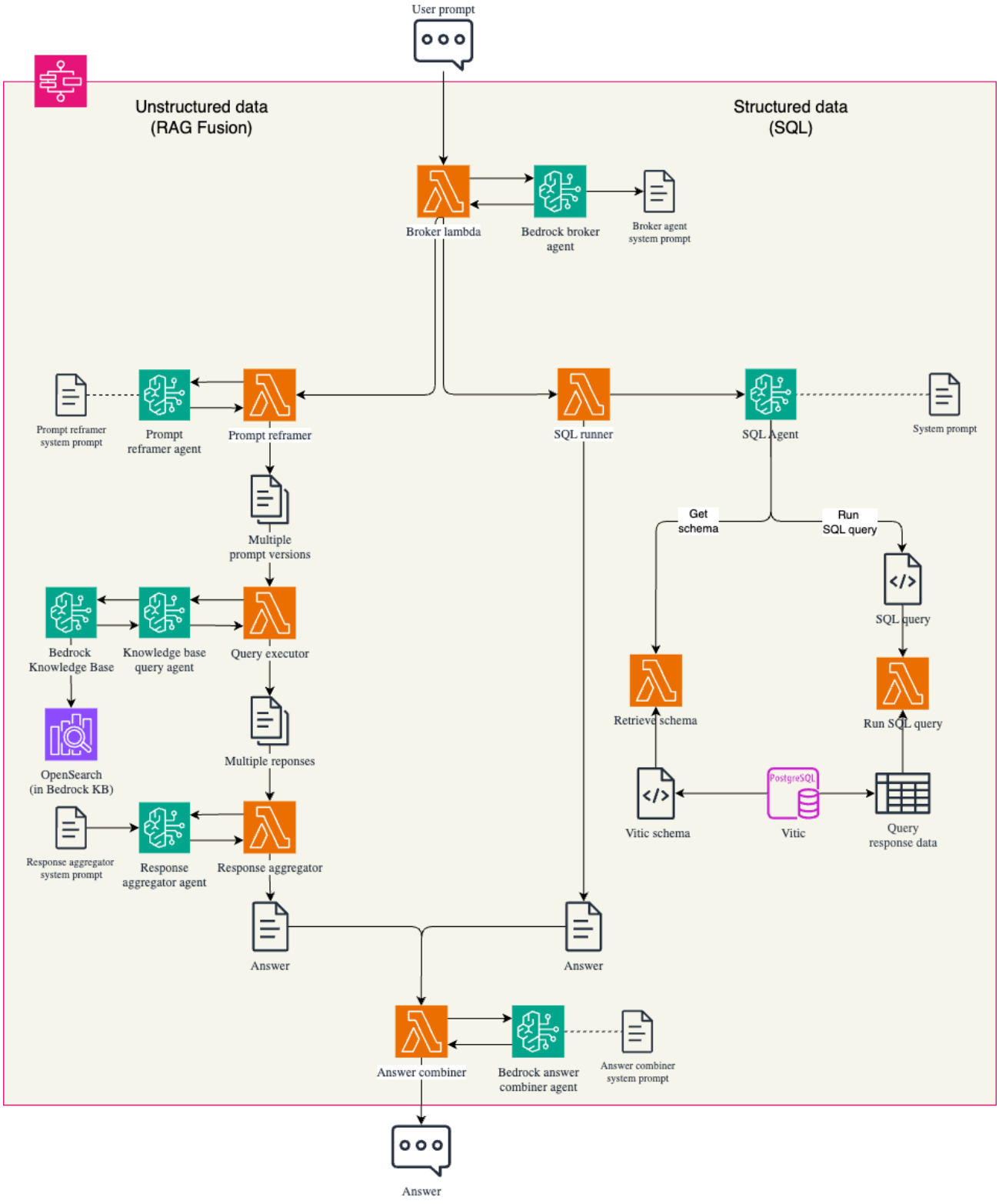


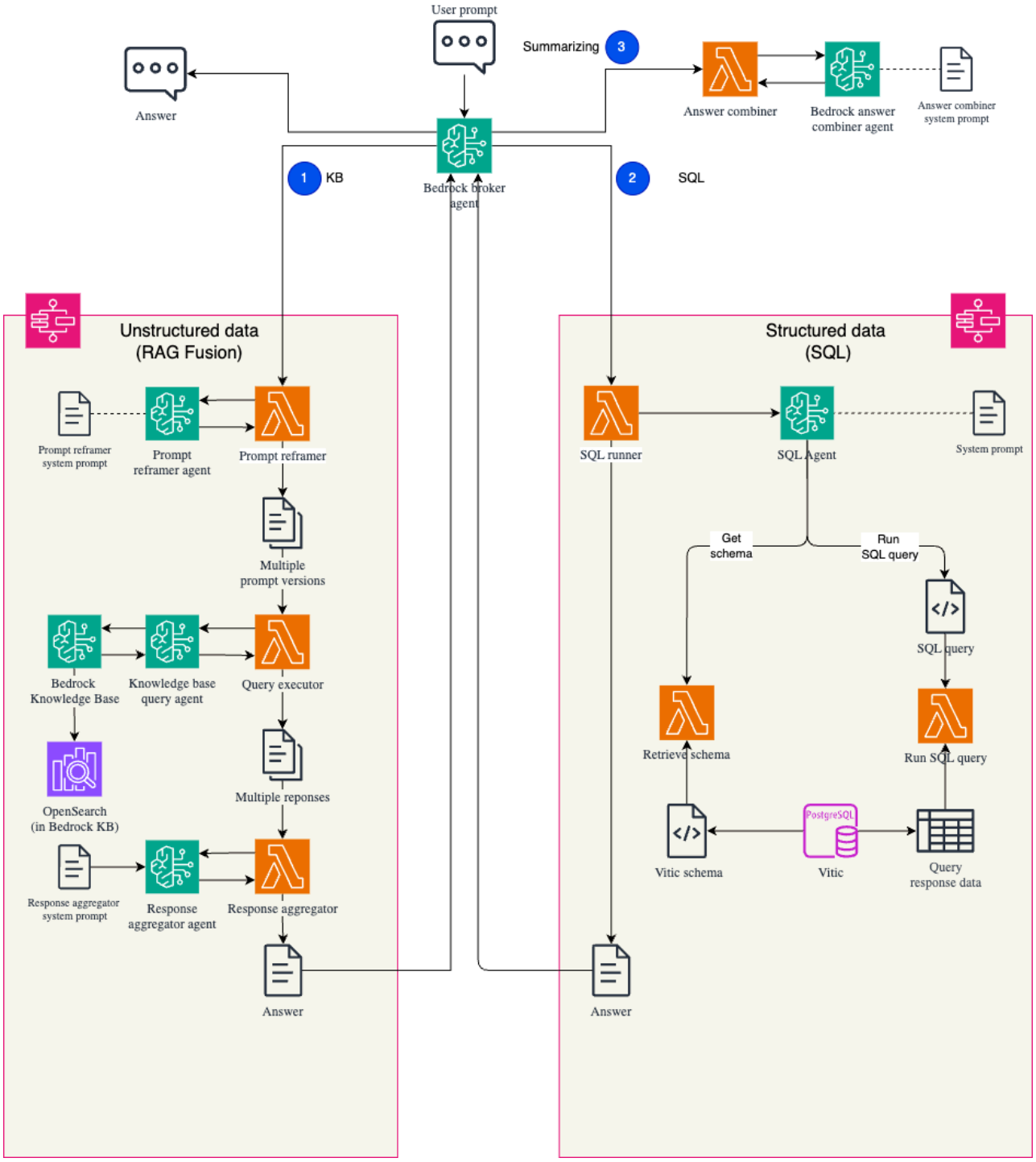
In the above diagram are not shown:

- logging (see implementation for details)
- analysis
- visualisation of results

Query execution - proposed evolution

The below diagram shows the proposed structure for the evolution of the RAG Fusion concept.


User prompt
Unstructured data
(RAG Fusion)
Structured data
(SQL)
Broker lambda
Bedrock broker
agent
Broker agent
system prompt
Prompt reframer
system prompt
Prompt
reframer agent
Prompt reframer
SQL runner
SQL Agent
System prompt
Multiple
prompt versions
Get
schema
Run
SQL query
Bedrock
Knowledge Base
Knowledge base
query agent
Query executor
SQL query
Retrieve schema
Run SQL query
OpenSearch
(in Bedrock KB)
Multiple reponses
Vitic schema
Vitic
Query
response data
Response aggregator
system prompt
Response
aggregator agent
Response aggregator
Answer
Answer
Answer combiner
Bedrock answer
combiner agent
Answer combiner
system prompt
Answer



Query workflows

The current query architecture is implemented using a combination of Step Functions, Lambdas, and Bedrock Agents invoked from Lambdas. Logging is performed using Lambdas and information is stored in JSON formats on a dedicated S3 bucket.

Knowledge Base workflow

This implements the basic functionality of Knowledge Base agent invoking workflow.

Structure and implementation are described in Knowledge Base RAG Fusion Agent.

Note: The component is glued to the Knowledge base RAG Fusion workflow and used to run the alternative queries against the Knowledge Base.

Knowledge Base RAG Fusion workflow

Structure and implementation are described in Knowledge Base RAG Fusion Agent.

SQL Query workflow

Structure and implementation are described in Vitic Agent V1.

Aggregated workflow

The initial query is analysed and distributed between the (RAG Fusion) Knowledge Base workflow and the SQL Query workflow. At the end, the result is summarised by a summarising agent.

Structure and implementation are described in Combined RAG Fusion V2.

Appendix

A1. Initial draft for the architecture diagram