Lab 5: Create AI Agents

Natural Language to Data

Concepts Introduced

AI Agents - LLM-powered systems that use tools to answer questions and take actions.

Key Concepts:

- **Genie** - Databricks' natural language to SQL agent for structured data
- **Knowledge Agent** - RAG-based agent that retrieves from document collections
- **Text-to-SQL** - LLM generates SQL from natural language questions
- **RAG (Retrieval Augmented Generation)** - Retrieve relevant documents, then generate answers

The Structured vs Unstructured Gap:

- Structured data: What customers *have* (accounts, positions, transactions)
- Unstructured data: What customers *want* (interests, goals, preferences)

Two Agent Types

Agent	Data Type	Capability
Genie	Structured (Delta tables)	SQL generation from natural language
Knowledge Agent	Unstructured (HTML docs)	Document retrieval and analysis

The Gap: Customers have stated interests in documents that aren't reflected in their portfolio data.

Genie Agent (Structured Data)

Queries the **14 Delta tables** we exported from Neo4j.

Sample Questions:

```
"Show me customers with investment accounts and their portfolio values"
"What are the top 10 customers by account balance?"
"Which customers have high risk profiles but conservative portfolios?"
```

Configuration:

- Define measures (Total Portfolio Value, Account Balance)
- Define filters (High Value Accounts, Recent Transactions)
- Add column descriptions and synonyms

Knowledge Agent (Unstructured Data)

Analyzes **HTML documents** in Unity Catalog Volumes.

Content Indexed:

- Customer profiles with investment preferences
- Bank and institutional profiles
- Company research and quarterly reports
- Market analysis and investment guides

Sample Questions:

```
"What investment interests does James Anderson have?"
"What renewable energy opportunities are in the research documents?"
```

Next: Combine both agents into a unified system