Final Development Plan
Automated Data Extraction and Automation

Project Goal
Build a complete agentic system that extracts structured data from unstructured text and stores it using both vector and graph databases.

Total Estimated Time
Approximately 3 hours.

Core Technologies
LangGraph, LLM (OpenAI/Gemini/Ollama), Weaviate, NebulaGraph, Python.

Phase 0 – Environment Setup
Create project structure, install dependencies, and prepare sample unstructured text.

Phase 1 – Entity Extraction
Use an LLM to convert raw text into structured JSON containing entities like person, organization, date, and amount.

Phase 2 – Agentic Workflow (LangGraph)
Implement a multi-node agent flow: Extract → Validate → Store → End.

Phase 3 – Vector Database (Weaviate)
Store documents and metadata for semantic search and similarity-based retrieval.

Phase 4 – Knowledge Graph (NebulaGraph)
Store extracted entities as nodes and relationships as edges, enabling structured reasoning.

Phase 5 – Integration & Demo
Demonstrate end-to-end pipeline and execute one semantic query and one graph query.

Expected Outcome
A working prototype that automates conversion of unstructured text into structured, queryable knowledge.

Evaluation Justification
The system combines agent orchestration, semantic understanding, and relational reasoning for scalable automation.