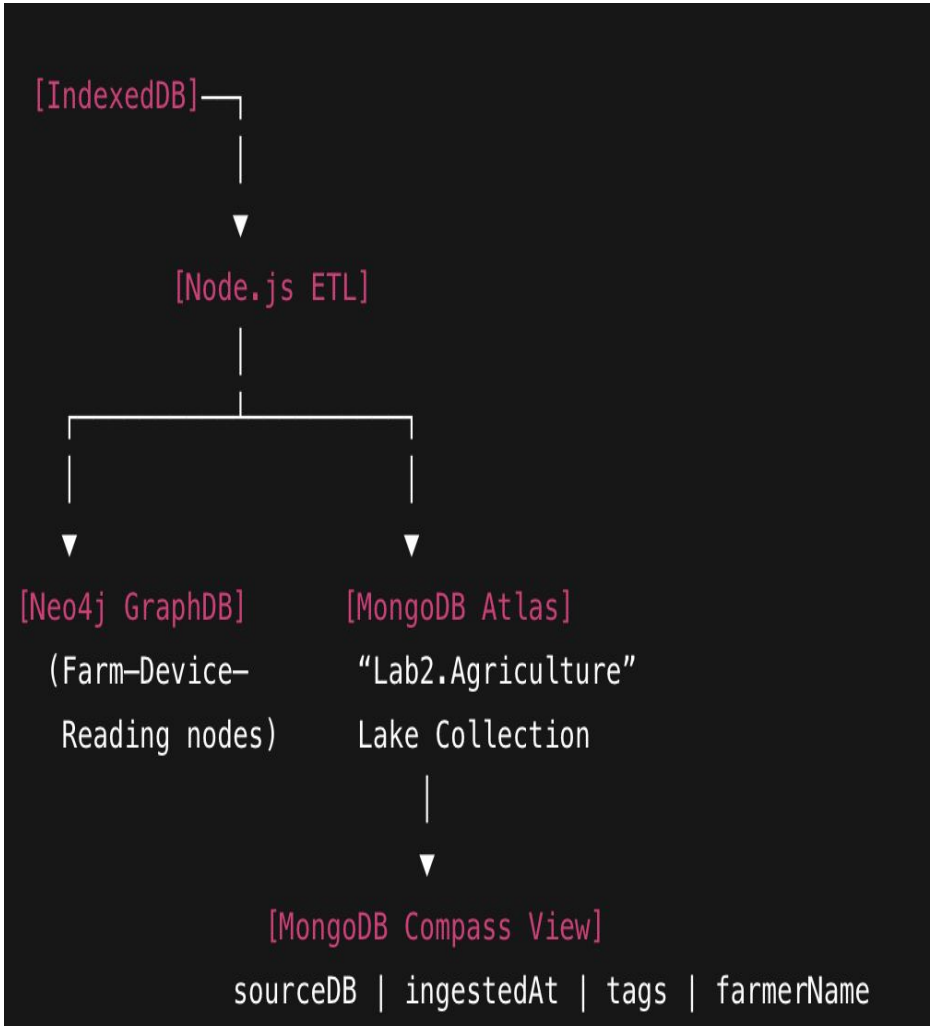
Integrated Data Pipeline: IndexedDB + Neo4j → MongoDB Lake



- IndexedDB → JSON export of offline sensor data
- Neo4j → Graph traversal via Cypher query
- Node.js ETL merges both → MongoDB Atlas 'lake'
- Metadata auto-added:
  - sourceDB (IndexedDB / Neo4j)
  - ingestedAt (current UTC)
  - tags (keywords)
  - farmerName = 'Utkarsh Gawande'

Lessons Learned from Extending the Pipeline

- Connected heterogeneous data sources: IndexedDB + Neo4j + MongoDB Atlas.
- Learned Cypher queries to extract Farm–Device–Reading relationships.
- Integrated multiple drivers (neo4j-driver, mongodb) in Node.js.
- Added metadata enrichment and schema harmonization.
- Debugged Compass verification with sourceDB filters.
- Built a complete mini data-lake workflow for smart agriculture and IoT.

Challenges & Reflections

✅ What Worked Well:
- Successful Neo4j–MongoDB integration using one ETL script.
- Correct Cypher graph data retrieval.
- Visible metadata fields in Compass (sourceDB, ingestedAt, farmerName).

⚠️ Challenges:
- Initial module-installation errors (missing 'mongodb').
- Confusion between local vs Atlas database paths (Lab2 vs Project).
- Needed consistent naming for Compass visibility.

🌾 Future Improvements:
- Automate logging and timestamp dashboards.
- Add error-handling for failed inserts.
- Extend graph for multi-sensor and time-series analytics.