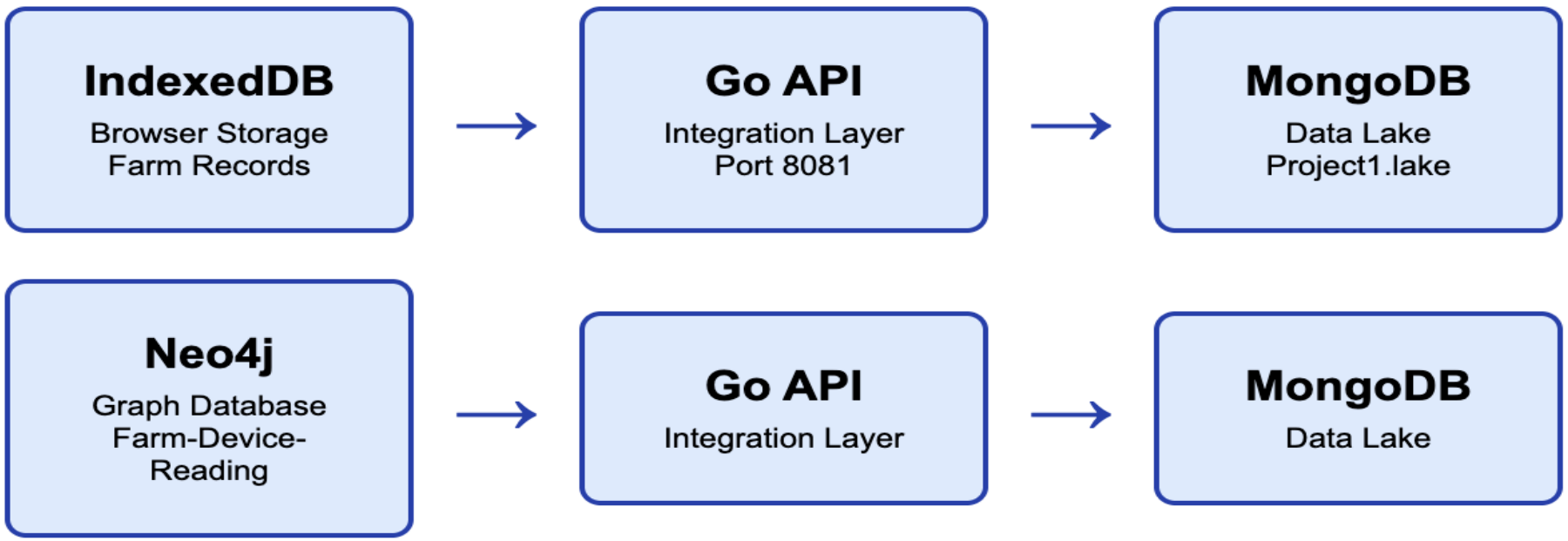
Multi-Source Data Lake Integration



📋 Metadata Standardization

sourceDB: Origin database | **ingestedAt**: UTC timestamp | **tags**: Keywords array

Lessons Learned: Extending the Data Pipeline

- **API-Driven Integration**
  - Centralized Go API simplified data ingestion from multiple sources
- **Metadata is Critical**
  - Standardized fields enable data lineage tracking
- **Schema Flexibility**
  - MongoDB accommodated diverse data structures
- **Data Transformation**
  - Converting graph to documents required careful planning
- **Multi-Database Management**
  - Proper resource handling for multiple connections
- **Cross-Platform Integration**
  - Browser and server databases unified via REST

What Worked Well vs. Challenges Faced

✓ What Worked Well

- **Go API Integration Hub**
  Single endpoint for both sources
- **Connection Management**
  Multiple clusters handled well
- **Metadata Automation**
  Automatic tagging ensured consistency
- **RESTful Design**
  Clear endpoints, easy testing
- **Browser Interface**
  Simplified IndexedDB operations

⚠ Challenges Faced

- **Data Model Transformation**
  Graph to document conversion complex
- **Cross-System Debugging**
  Multiple error sources
- **Timing Dependencies**
  Required proper startup sequence
- **Network Configuration**
  CORS and multi-cluster setup
- **IndexedDB Limitations**
  Manual data seeding required