Session 10

Graph Databases

Big Data Analytics Technology, MSc in Data Science,
Coventry University UK

Miyuru Dayarathna

Presentation Outline

- Introduction
- Native Graph Storage
- Graph Query Languages
- Performance and Scalability
- Conclusion

Property Graph Data Model

- It contains nodes and relationships
- Nodes contain properties (key-value pairs)
- Relationships are named and directed, and always have a start and end node
- Relationships can also contain properties

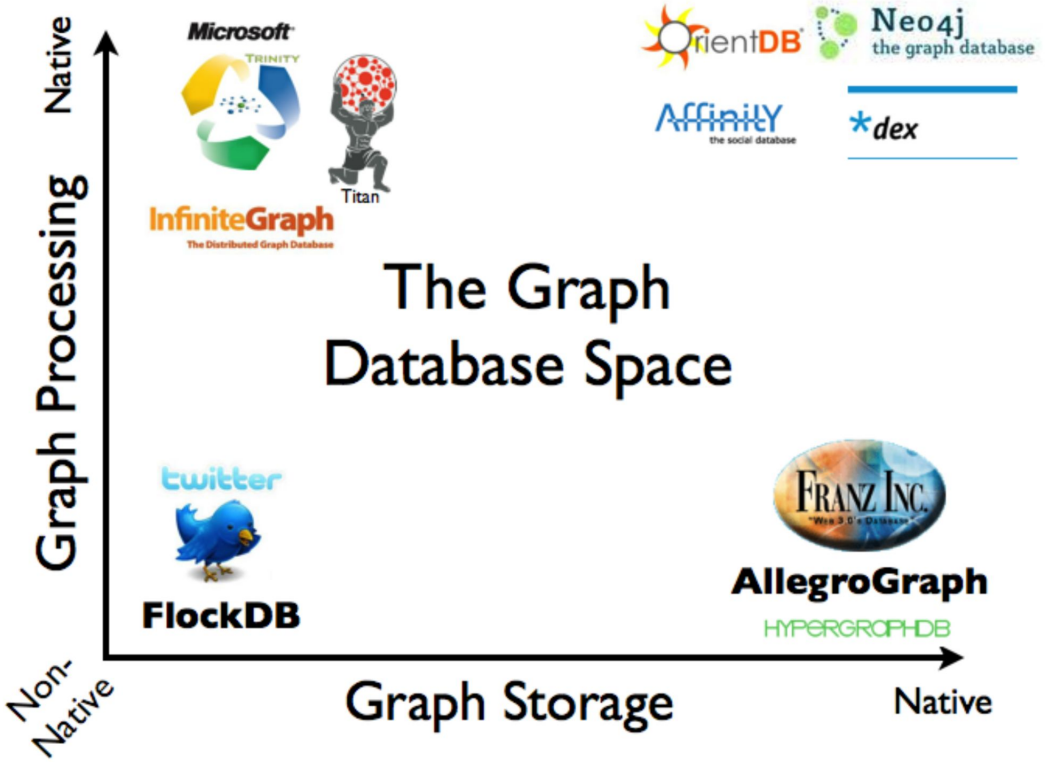
Graph Databases

- A graph database management system (henceforth, a graph database) is an online database management system with Create, Read, Update, and Delete (CRUD) methods that expose a graph data model.
- Graph databases are generally built for use with transactional (OLTP) systems.
- Accordingly, they are normally optimized for transactional performance, and engineered with transactional integrity and operational availability in mind.

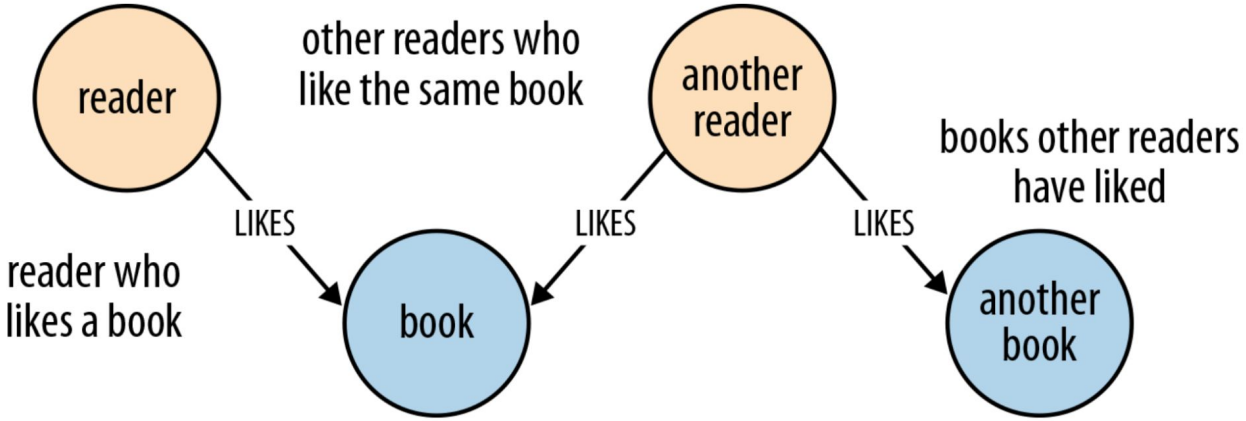
Two key properties of Graph Databases

- The underlying storage
  - Some graph databases use native graph storage that is optimized and designed for storing and managing graphs. Not all graph database technologies use native graph
  - storage, however. Some serialize the graph data into a relational database, an object oriented database, or some other general-purpose data store.
- The processing engine
  - the significant performance advantages of index-free adjacency, and therefore use the term native graph processing to describe graph databases that leverage index-free adjacency

Overview of Graph Database Space



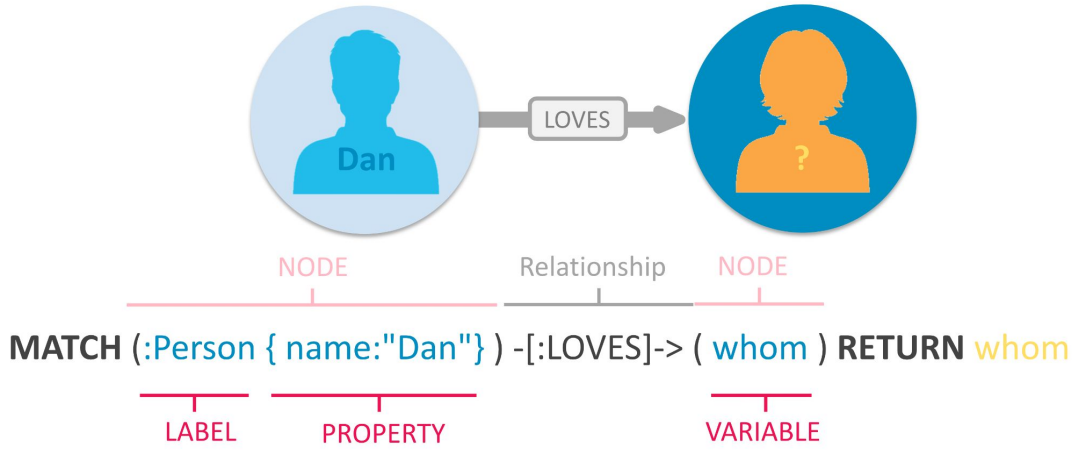
Graph Query Languages




```
START reader=node:users(name={readerName})
      book=node:books(isbn={bookISBN})
MATCH reader-[:LIKES]->book<-[:LIKES]-other_readers-[:LIKES]->books
RETURN books.title
```

Cypher Query Language

- Cypher is Neo4j's graph query language that lets you retrieve data from the graph.
- It is like SQL for Graphs

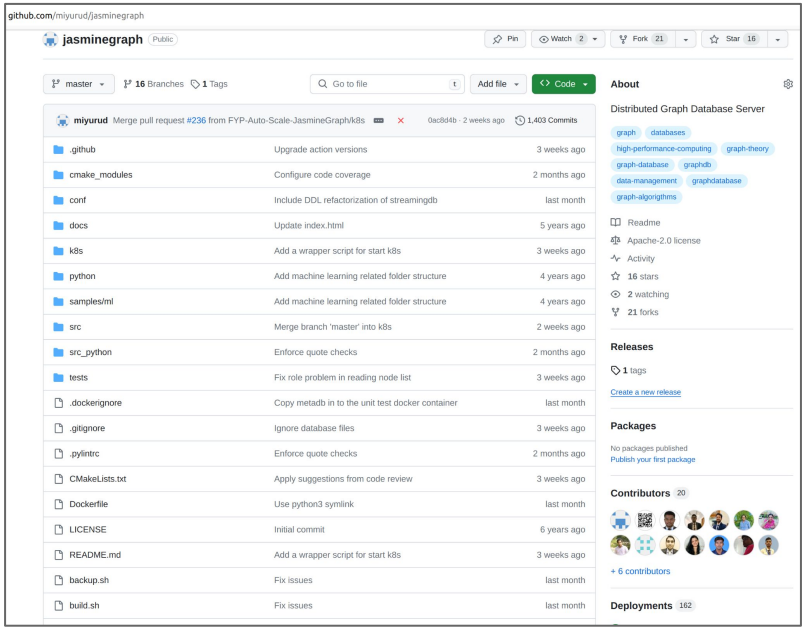


Distributed Graph Database Servers

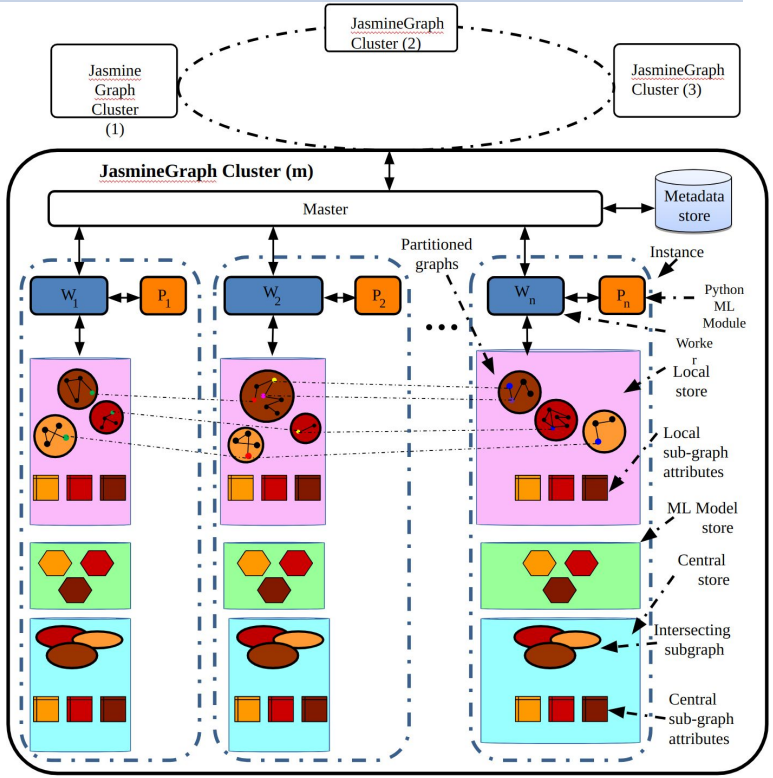
- Requirement
- Implementations
  - JasmineGraph
  - TigerGraph
  - Amazon Neptune
  - Neo4j Distributed

JasmineGraph

- An open source distributed graph database server
  https://github.com/miyurud/jasminegraph



JasmineGraph System Architecture



Graph Triangle Counting - Practical Session

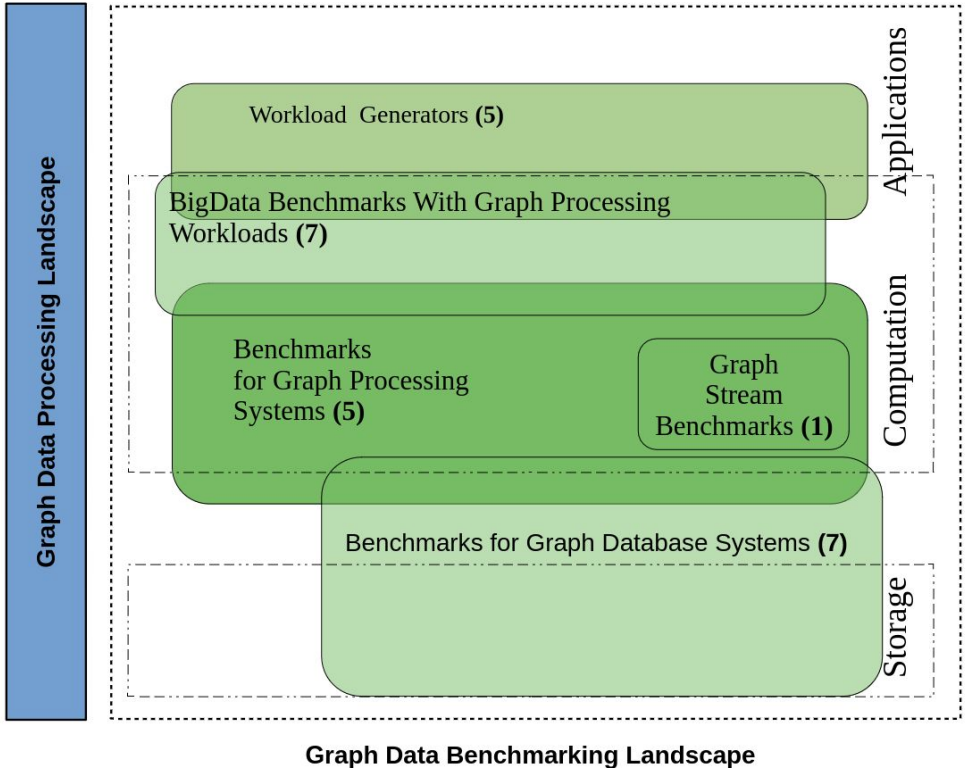
- Download the following graph and upload to JasmineGraph
- Count the number of triangles using *trian* command of JasmineGraph and verify the results obtained with the reported triangle count in the below link

  https://snap.stanford.edu/data/email-Enron.html

Graph Processing System Performance Measurement

- Graph 500
- LDBC (Linked Data Benchmark Council)
- Graphalytics

Summary view of graph benchmarking landscape



https://arxiv.org/pdf/2005.12873.pdf

Thank you!