Blue Sun

Sireesh Gururaja & Chae Jubb

Overview

- Problem
- Solution
- Architecture
- Future Work
- Lessons Learned

Picture This!

- Apache Solr
  - Search engine (Apache Lucene)
  - Java
  - Scalable
  - Fault tolerant
  - GET requests
  - JSON
  - other buzzwords that won’t fit on this slide

But...

- Only works on normalized data

Picture This!

- Gremlin
  - More Java
  - Graph Engine
  - Database
  - Tinkerpop3

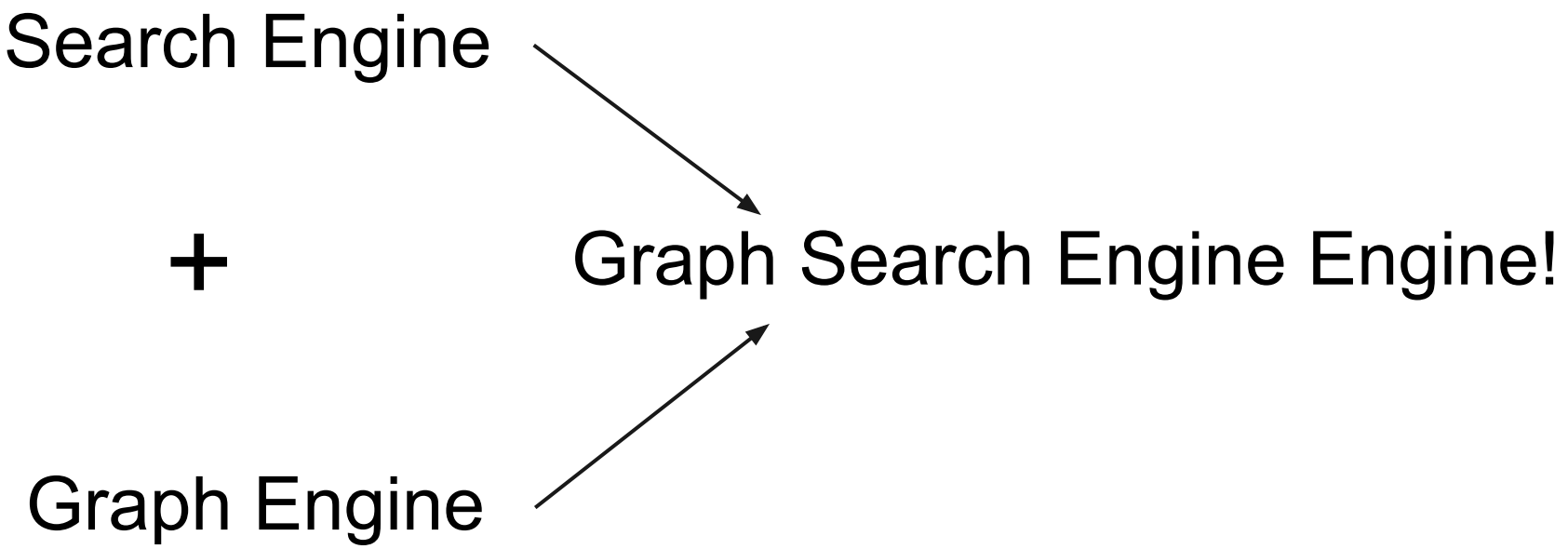
Problem

- Search Engine can only deal with normalized data
- We have
  - Search Engine
  - Graph Engine
- Graph simulates denormalized data operations

Graph Database

- Existing documents in database are vertices of graph
- Relational database operations like joins are simulated by graph vertices
- Data need no longer be normalized

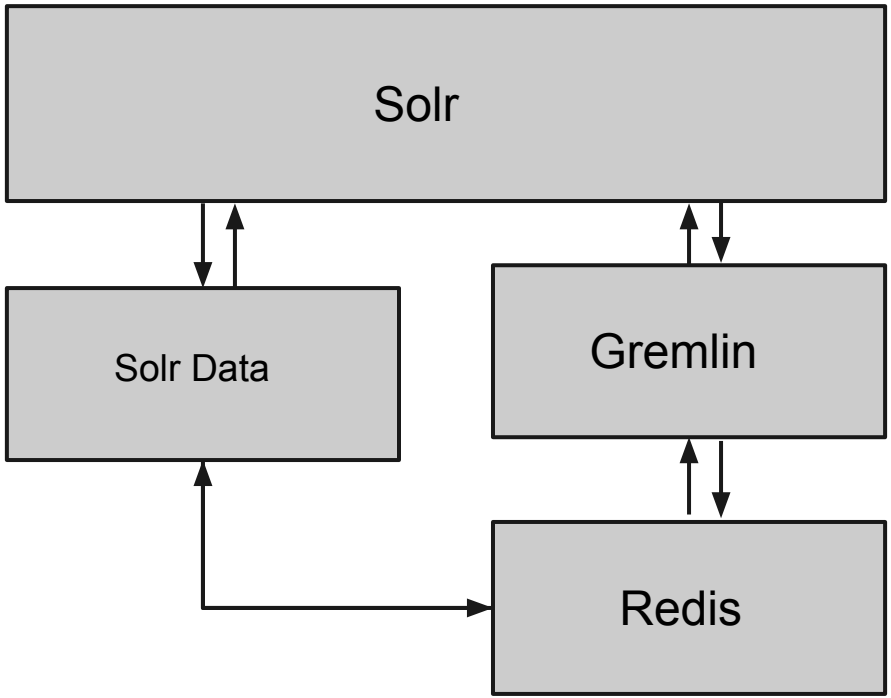
Solution



Architectural Components

- Solr
- Tinkerpop
  - Gremlin
- (Redis)
  - Underlying data store
  - Agnostic to this

Architecture Diagram



Process

- Basic Solr query returns Solr data
- Graph query returns list of nodes, mapped to documents

- Final result: intersection of Solr data and graph data

Modules

- Ingestion - Illinois/CMU
- Visualization - CMU
- Post Filtering - Columbia
- Redis Graph Implementation - MIT/CMU

Future Work

- Merge into Solr
  - Benefits
    - Graph search capability
    - It’s a “hot” area
  - Obstacles
    - Multiple jars
  - Module!
- More complex graph syntax

How did it go?

Major Obstacles

- Project Description
  - Didn’t match expectations
  - Outside of our specialties
- Poor Communication
  - Us
  - Mentor

Were we to do it again...

- Be more strict with ourselves
  - Press harder to set weekly meetings
  - Intermediate milestones
- Know when to cut losses
  - Adjust expectations

Questions?