Data Ontology Cache System

Company Background
- Process-driven investment manager with billions in assets under management
- Offices in New York, Houston, and London
- Invent trading algorithms that utilize petabytes of financial data
- Employ state-of-the-art technologies such as event stream processing and grid computing

Project Motivation
- Many different teams produce artifacts stored on disparate, unconnected systems
- Currently difficult to ask questions about related data that span multiple systems
- Unable to sanity check and normalize data for storage on specific volumes

System Features Delivered
- Distributed ontology cores
  - Security built into the ontology
  - Accepts data pushes from collectors
  - Responds to SPARQL queries from clients
  - Data validation against ontology schema
- Web application
  - Relationship discovery with graph visualizer
  - Full-text search engine
  - Groovy reporting module
- System status monitor
- UNIX-style command line tools
- Java API for query clients and collectors

Data Scenario

System Architecture

Dual-Model Process
- First Quarter: Evolutionary Prototype
  - Technology spikes and feasibility studies
  - Minimize risk
- Second Quarter: Time-Boxed Iterations
  - Implementation of features
  - Maximize value