Triple Stores

What is a triple store?

- A specialized database for RDF triples
- Supports a query language
    - SPARQL is the W3C recommendation
    - Other RDF query languages exist (e.g., RDQL)
    - Might or might not do inferencing
    - Most query languages don't handle inserts
- Triple stores might be in memory or provide a persistent backend
- Presistence provided by a relational DBMS (e.g., mySQL) or a custom DB for efficiency.

Example: HP's Jena Framework

- An OS Java system developed by HP
    - http://jena.sourceforge.net/
- Has internal reasoners and can work with DIG compliant reasoners or Pellet.
- Supports SPARQL
- Joseki is an add-on that provides a SPARQL endpoint via an HTTP interface

SPARQL Example

```
BASE <http://example.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ex: <properties/1.0#>
SELECT DISTINCT $person ?name $age
FROM <http://rdf.example.org/people.rdf>
WHERE { $person a foaf:Person ;
            foaf:name ?name.
        OPTIONAL { $person ex:age $age } .
        FILTER ! REGEX(?name, "Bob")
        }
LIMIT 3 ORDER BY ASC[?name]
```

Example: Kowari

- http://kowari.org/
- "An Open Source, massively scalable, transaction-safe, purpose-built database for the storage, retrieval and analysis of metadata"
  - Implemented in Java, not built on a conventional RDBMS
- Developed by Tucana, which also had a commercial version
  - Tucana closed down in Fall 2005, Northrop Grumman purchased the rights to the commercial system

Example: Sesame

- Sesame is an open source RDF framework with support for RDFS inferencing and querying
- http://www.openrdf.org/
- Implemented in Java
- Query languages: SeRQL, RQL, RDQL
- Triples can be stored in memory, on disk, or in a RDBMS

Issues

- Can we build efficient triple stores around conventional RDBMS technology?
- What are the performance issues?
  - Load time?
  - Interfencing?
- How well does is scale?