The Semantic Web RDF standard from W3C is making increasing inroads into government, the enterprise, bioinformatics, telecommunications, and other demanding arenas. Its simplicity and flexibility makes it suitable to represent any structured or unstructured knowledge (be it call detail records from telecom providers or intelligence data from Homeland Security). However, for real-world applications the number of RDF triples can easily grow into millions or even billions, making it difficult to process efficiently with traditional means.

Systems that must load, manipulate, and query such enormous triple data stores require the best possible performance. Faced with this challenge, Franz has used its 20+ years of experience with persistent dynamic object technology to engineer a massively scalable persistent triple store - AllegroGraph.

AllegroGraph efficiently performs the three most important tasks for a triple store: to load, store, and query data.

- Loading of triples, through its highly optimized RDF/XML and N-Triples parsers, is best-of-breed, particularly on large files. With just standard x86 64-bit hardware, it can load gigabytes of RDF data in minutes. The following table displays the results of loading and indexing a variety of RDF files and Ontologies.

<table>
<thead>
<tr>
<th>AllegroGraph Load Test</th>
<th>#Triples</th>
<th>Size *</th>
<th>Time</th>
<th>Load Rate **</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCyc</td>
<td>255K</td>
<td>49MB</td>
<td>17 Seconds</td>
<td>15.0 K TPS</td>
</tr>
<tr>
<td>FMA Ontology Model of Human Anatomy</td>
<td>550K</td>
<td>84.5MB</td>
<td>37 Seconds</td>
<td>15.0 K TPS</td>
</tr>
<tr>
<td>LUBM, Lehigh U. Benchmark - 1,000 Files</td>
<td>6.88M</td>
<td>1.3GB</td>
<td>8.24 Minutes</td>
<td>13.9 K TPS</td>
</tr>
<tr>
<td>Uniprot, Universal Protein Knowledgebase</td>
<td>234M</td>
<td>28.6GB</td>
<td>3.3 Hours</td>
<td>19.36 K TPS</td>
</tr>
<tr>
<td>Reified LUBM + Movie &amp; Actor Database</td>
<td>1B</td>
<td>N/A***</td>
<td>14.03 Hours</td>
<td>19.80 K TPS</td>
</tr>
</tbody>
</table>

* Size = size of triple file, ** TPS = Triples per second, *** This RDF data is generated at runtime.

- Storage is persistent, including between application launches in on-disk binary trees. There is no additional serialization or deserialization overhead.

- Querying is both flexible and performant. Multiple indices support fast access through a simple triple-level API, Allegro Prolog, or SPARQL (the emerging W3C standard RDF query language). When querying for a particular subject with ten triples, AllegroGraph can retrieve about 40,000 triples per second, from disk.
High-performance Storage and Querying

AllegroGraph is designed with microsecond retrieval times as a basic target. On a 2GHz AMD64 quad-processor machine with 16GB of RAM, AllegroGraph hits a peak load speed of 20,000 triples per second, including thorough indexing and on-disk storage. Simple triple queries have achieved speeds with a worst-case time of around 30 microseconds, speeding up to as little as 300 nanoseconds for a cache hit.

Powerful and Expressive Reasoning and Querying

AllegroGraph provides the broadest array of mechanisms to query and access knowledge in RDF triples:

- Low-level APIs allow fast, ‘close-to-the-metal’ access to triples by subject, predicate, and object.
- RDF Prolog provides concise, powerful, industry-standard, domain-specific reasoning to stored knowledge for building relations on top of RDF data.
- SPARQL, the W3C standard RDF query language, gives native object, RDF, and XML responses to queries. Query over sockets, HTTP, Lisp or a Java / SAIL API.
- RDFS reasoning with owl:SameAs and owl:InverseOf predicates NEW

Furthermore, the Semantic Web reasoning system, RacerPro, has been integrated with AllegroGraph, exposing RDF data in AllegroGraph to Racer’s highly optimized Description Logic (DL) reasoner. It is most suitable for ontology-driven applications. RacerPro’s interfaces also include DIG over HTTP and support for rules (SWRL).

Robustness

AllegroGraph is robust at every level. Transactions provide safety when importing data and adding triples, and fully-journalled modification ensures the consistency and integrity of data in a failure scenario. Modifications can be continually streamed to backup files - even other databases, over the network to insure redundancy.

Other Features

AllegroGraph offers facilities beyond the standard RDF requirements. Triples have additional optional fields, usable for source tracking, named graphs or access control. This feature allows implementation of permissions, trust, and provenance layers easily. To support developing ontology-driven applications, simple inferencing (subclass, subproperty, identity, and inverse relations) is included out-of-the-box.

Platforms

- AllegroGraph runs on all popular 64-bit architectures (8 platforms), and is 100% cross-platform compatible.
- The AllegroGraph Java/HTTP version provides an HTTP interface for loading and retrieving RDF triples, and a Java API for developing large-scale semantic applications in Java.