



Semantic Web Success Story

Practical Integration of Semantic Web Technology

Chris Chaulk, Software Architect
EMC Corporation

Who is this guy ?

- Software Architect at EMC
 - 12 years, Storage Management Software
 - Lead architect for EMC's ProSphere product
 - Technical direction for distributed engineering team
 - U.S./India, 220+ engineers

The Goal (2008)

New product...better than the old one.

Architectural drivers

- Standards-based
- Virtual appliance-based deployment
- De-centralized and distributed
- High scalability
- Dynamic, flexible information model
- Semantics in model, not application code
- RESTful architectural style

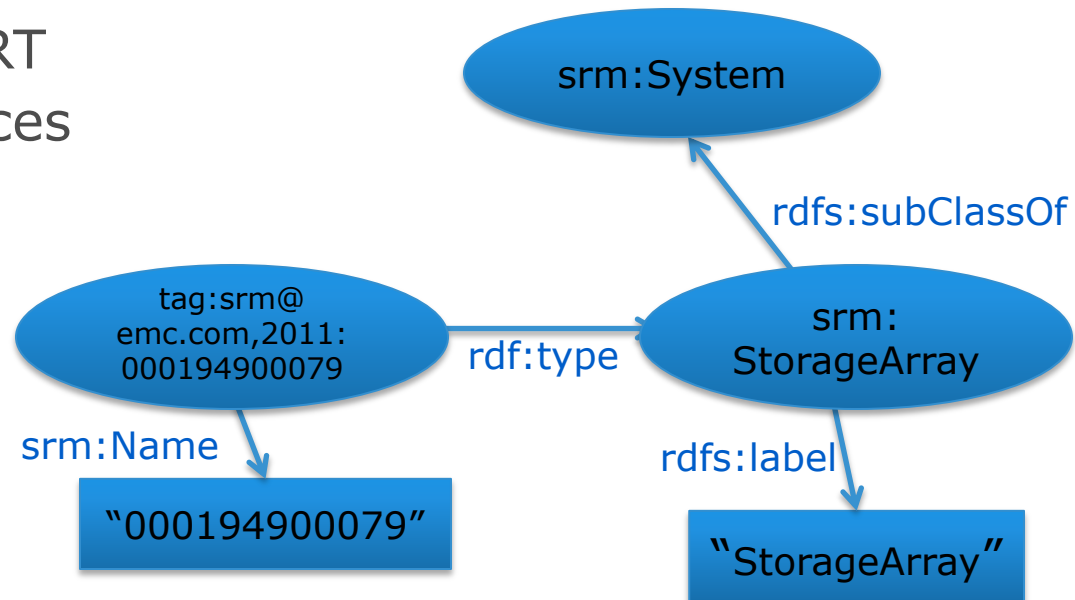
Information Architecture



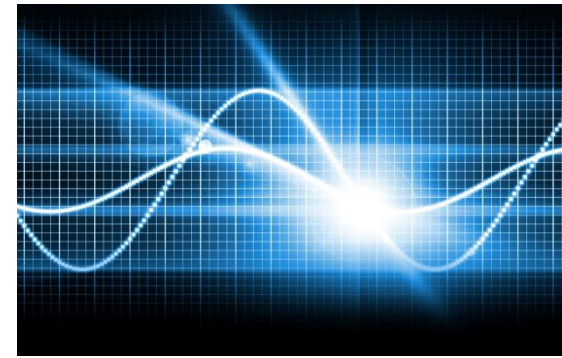
- Key requirements
 - Visualize network relationships
 - host-to-storage
 - Support new platforms rapidly
 - There is no **single** model
 - *There are multiple models, and there will always be*
 - Fundamentally a data integration project
- Decision
 - Use open standards
 - RDF and OWL chosen for information model
 - Model entire domain
 - Use model consistently

Information Architecture

- “Use RDF in the middle; Convert to/from RDF at the edges” – *thanks to David Booth*
- Source models transformed to RDF
 - Transforms:
 - CONSTRUCT/INSERT
 - RDFS/OWL inferences
 - Procedural



The bigdata[®] experience



- Development and training
 - Rapid development of new features
 - On-site training and consulting
- Support
 - Extremely fast response times (minutes)
 - Very collaborative
 - Dealing directly with developers, not support team
- Technology
 - Solid and scalable
 - Very few defects found
 - Feature-rich

bigdata[®] at EMC

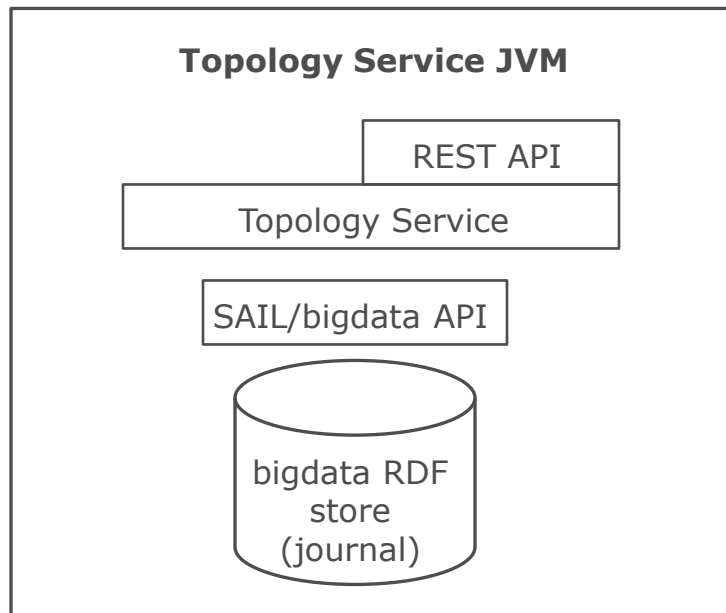
- Embedded RDF store
 - Persistence for multiple information types
 - Other SQL/NoSQL technologies utilized (RDBMS)
- Encapsulated RDF store
 - Maven project, Spring configured
 - Prior to bigdata-generated Maven project
- Funded implementation, consultation
 - ChangeLog, Statement Identifier in-lining

Information Architecture

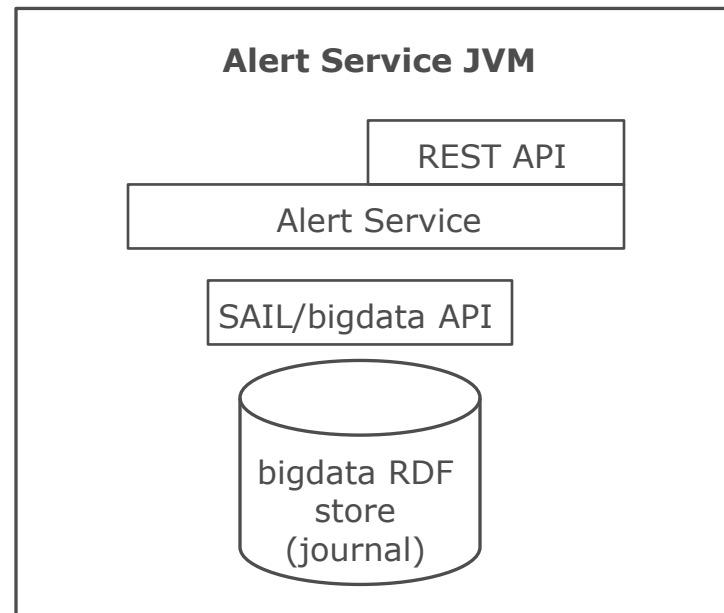


- REST driven by RDF model
 - Attributes and relationships modeled in RDF as Properties Class
 - URI = `https://{ServerName}/topology/instances/StorageArray/{arrayID}`
 - `{arrayID}` can be any Identifier RDF property
- Tag URIs used for most individuals
- Ontologies centralized in Maven project

bigdata[®] usage: EMC ProSphere

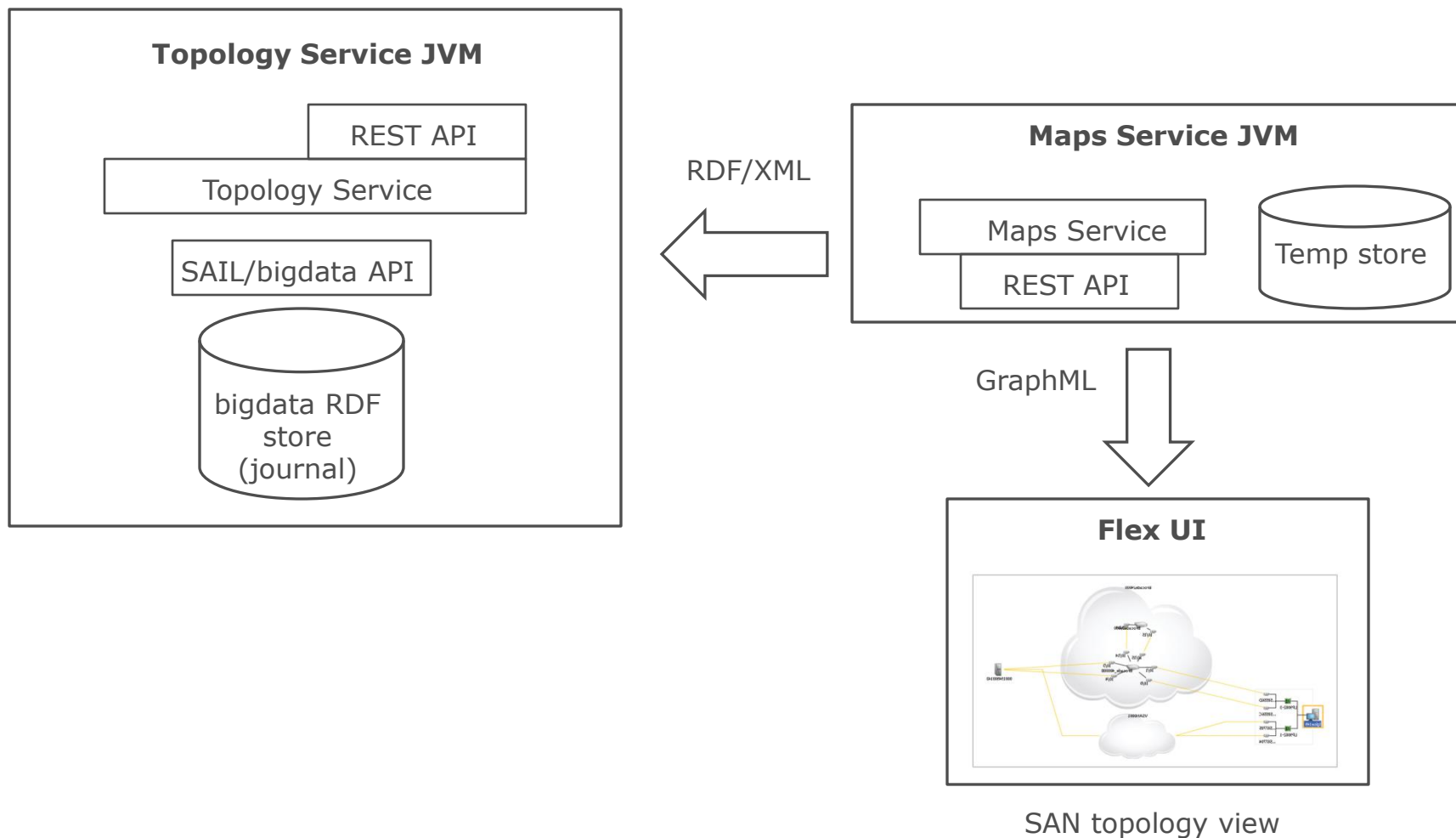


↑
SAN topology and configuration



↑
SAN topology and configuration events (asynchronous)

bigdata[®] usage: EMC ProSphere



Semantic Web in the Enterprise: *What makes it harder ?*



- Education and Expertise
 - Can't ask an engineer to "Go learn the Semantic Web"
 - Not every developer needs to know SPARQL
- Real-world examples
 - What's the "Killer App" for Semantic Web applications ?
- Integration/Extensions
 - E.g. SPARQL and need for full-text search
 - bigdata full-text search implementation
 - OpenShara ElasticSearch integration

Semantic Web in the Enterprise: *What makes it harder ?*

- Traditional development parallels
 - Comfort-level with SQL, Objects
 - E.g. I want a findObject(objectID) interface
- REST frameworks
 - Following Linked Data principles and patterns
- Awareness
 - Add RDF databases to the NoSQL discussion!
 - Linked Data as an integration architecture
 - E.g. Linked Data Integration Framework (LDIF)



What's next?

- Leverage new bigdata[®] features
 - Additional SPARQL 1.1 features
 - MemStore
 - History table
 - Solution set cache
 - bigdata[®] B-Trees
 - Proven, compare to LevelDB
 - Graph analytics
 - Clustered database & replication
- bigdata[®] integrations and RDF evangelism
 - ElasticSearch, Spring-Data like API



EMC²®