On Demand RDF Databases in the Cloud

Presentation for the Ontology Forum

March 3, 2016
Company Brief

What we offer? Semantic graph database and Content management solutions

How we do it? Managing huge knowledge graphs and using them for text analysis and tagging

Best for: Content publishing and Information discovery

Clients: BBC, FT, IET, OUP, Wiley, AstraZeneca, Getty
Customers
What is S4?

- Capabilities for *text analytics, content enrichment* and *smart data management*
  - Text analytics for news, life sciences and social media
  - RDF graph database as-a-service
  - Access to large open knowledge graphs
- Available on-demand, anytime, anywhere
  - Simple RESTful services
- Simple *pay-per-use pricing*
  - No upfront commitments
Vision

Semantic Search
- Semantic, exploratory search
- Metadata driven content

Text Mining & Interlinking
- Interlink people, locations, organisations, topics
- Discover implicit relations
- Reuse open knowledge graphs

Graph Database
- Flexible RDF graph data model
- Ontology based metadata layer
S4 Big Picture

RDF Data

Knowledge Graphs

Locations

Social Media

Text Analytics

Organisations

Persons

Biomedical entities

Today's webinar focus

Deputy Defence Minister Kostas Tsichos told the BBC northern Europe must do more to rescue and shelter migrants.
Benefits

- Enables quick prototyping
  - Instantly available, no provisioning & operations required
  - Focus on building applications, don’t worry about infrastructure
- Free tier!
- Easy to start, shorter learning curve
  - Various add-ons, SDKs and demo code
- Based on enterprise semantic technology by Ontotext
Available Knowledge Graphs

• SPARQL query endpoint to the FactForge semantic data warehouse
  – 500 million entities / 5 billion triples

• Key LOD datasets integrated
  – DBpedia, Freebase/WikiData, GeoNames, WordNet
  – Dublin Core, SKOS, PROTON ontologies and vocabularies

• Learn more at: http://factforge.net/
GraphDB Features

- High performance RDF database, 10s of billions of triples
- Integration with text analysis pipelines
- Standards compliant Full SPARQL 1.1 support
- Various reasoning profiles, including custom rules
- Efficient data integration (“sameAs” optimisations) and deletion of statements & their inferences
- RDF Rank, full-text search, 3rd party plugins
- Connectors to Solr, ElasticSearch, NoSQL DBs
- GraphDB Workbench
Target Users

- Ideal for customers who are...
  - still evaluating and testing RDF technology
  - in the early phase of adoption / PoC

- Enterprise grade RDF database in the Cloud
  - No need for upfront payments for licenses & hardware
  - Pay only for what you use, when you use it
  - Instantly operational within minutes
  - No need for complex planning - use as many DB instances for as long as needed
  - Timely upgrades to the latest version

- *Self-managed* and *fully managed* options
S4 Options

• Available from AWS Marketplace, “1-Click” purchasing

• Variety of hardware configurations
  – 2 to 8 CPU cores / 8 to 61 GB RAM
  – IOPS performance & encryption (EBS)

• Manage large data volumes

• Pay-per-hour pricing

• Users take care of operations
  – Backups, restores
S4 Advantages

- Low-cost graph DBaaS available 24/7
- Ideal for small & moderate data & query volumes – database options: 1M, 10M, 50M, 250M & 1B triples
- Instantly deploy new databases when needed
- Zero administration – automated operations, maintenance & upgrades
- Users pay only for the actual database utilisation
- Standard OpenRDF REST AP
### S4 Sizing Options

<table>
<thead>
<tr>
<th>Database type</th>
<th>Max triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>micro</td>
<td>1 million</td>
</tr>
<tr>
<td>XS</td>
<td>10 million</td>
</tr>
<tr>
<td>S</td>
<td>50 million</td>
</tr>
<tr>
<td>M</td>
<td>250 million</td>
</tr>
<tr>
<td>L</td>
<td>1 billion</td>
</tr>
</tbody>
</table>
## S4 API

<table>
<thead>
<tr>
<th>resource</th>
<th>operations</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>/repositories</td>
<td>GET</td>
<td>Get info on DB repos</td>
</tr>
<tr>
<td>/repositories/&lt;REPOSITORY&gt;</td>
<td>GET, POST, PUT, DELETE</td>
<td>Create*, delete, query a repository</td>
</tr>
<tr>
<td>/repositories/&lt;REPOSITORY&gt;/size</td>
<td>GET</td>
<td>Gets the number of triples in a repository</td>
</tr>
<tr>
<td>/repositories/&lt;REPOSITORY&gt;/statements</td>
<td>GET, POST, PUT, DELETE</td>
<td>Add, read, update, delete statements</td>
</tr>
<tr>
<td>repositories/&lt;REPOSITORY&gt;/rdf-graphs/&lt;GRAPH&gt;</td>
<td>GET, POST, PUT, DELETE</td>
<td>Same as above</td>
</tr>
<tr>
<td>/settings</td>
<td>GET, PUT</td>
<td>Configure the DBaaS*</td>
</tr>
</tbody>
</table>
Try and Learn More

- Web: http://s4.ontotext.com/
- Documentation: http://docs.s4.ontotext.com/
- Support: s4-info@ontotext.com
- Tweet: @Ontotext_S4
- Call: (929) 239-0659
- In person: Peio Popov  
  <peio.popov@ontotext.com>