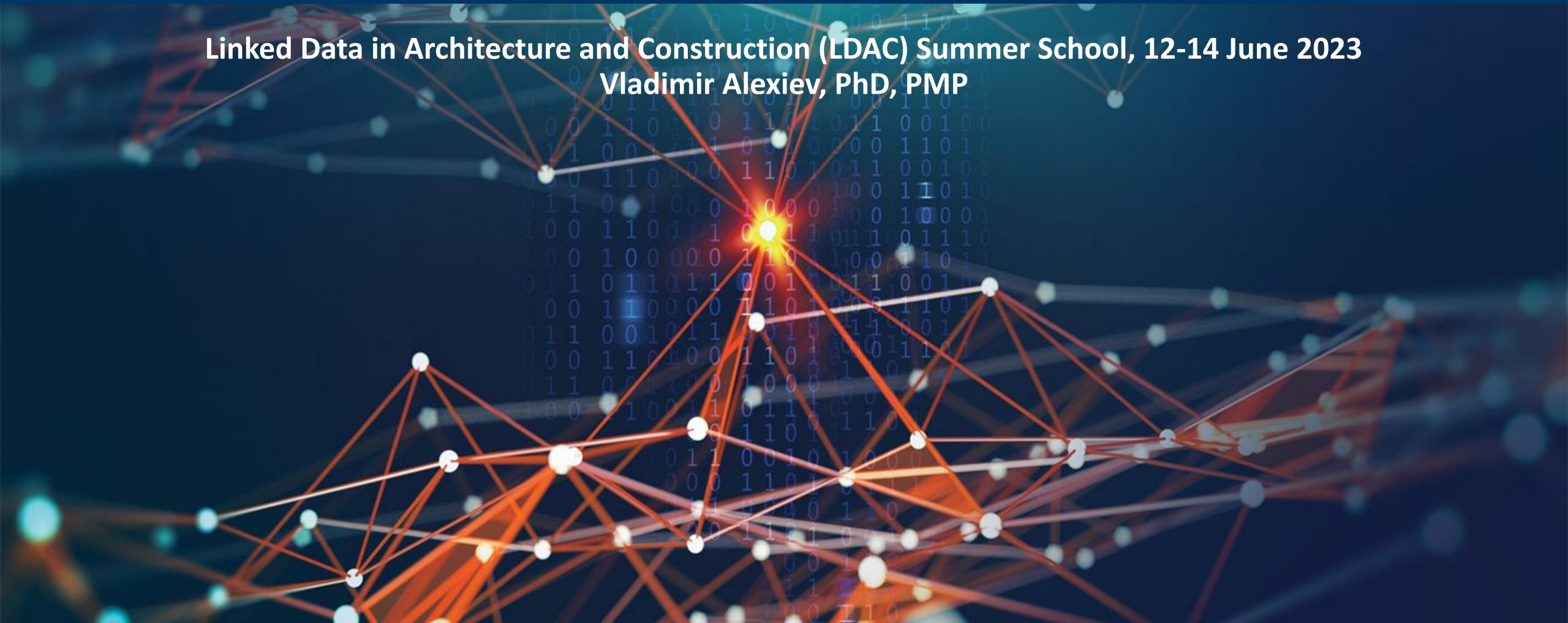


Ontotext Products Overview

GraphDB (database), Semantic Objects (GraphQL), OntoRefine (ETL)

Linked Data in Architecture and Construction (LDAC) Summer School, 12-14 June 2023
Vladimir Alexiev, PhD, PMP



- **GraphDB**

- **Essentials**

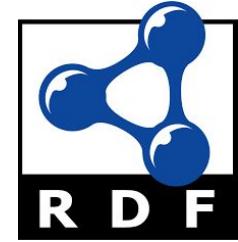
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

- **Semantic Objects**

- **Ontotext Refine**

GraphDB Essentials

- **Scalable and Dependable RDF 1.1 engine**
 - Predictable performance across wide range of workloads
- **Platform Independent (100% Java)**
- **W3C Standards Compliant**
 - Comprehensive support for SPARQL 1.1, OWL 2, RDF* and SHACL
- **Reasoning and Consistency Checking**
- **High-Availability Cluster and Enterprise-Grade Security**
- **Extensible Plug-in Architecture**
- **Excellent Support**



Architecture

GraphDB Workbench

- User friendly interface for database administration

The screenshot shows the GraphDB Workbench interface. On the left, there is a sidebar with various icons for managing datasets and configurations. The main area is titled "SPARQL Query & Update". It contains a code editor with a SPARQL query:

```
1 SELECT ?yr ?name ?document
2 WHERE {
3   ?class rdfs:subClassOf foaf:Document .
4   ?document rdf:type ?class .
5   ?document dc:terms:issued ?yr .
6   ?document dc:creator ?author .
7   ?author foaf:name ?name
8   OPTIONAL {
9     ?class2 rdfs:subClassOf foaf:Document .
10    ?document2 rdf:type ?class2 .
11    ?document2 dc:terms:issued ?yr2 .
12    ?document2 dc:creator ?author2
13    FILTER (?author=?author2 && ?yr2<?yr)
14  } FILTER (!bound(?author2))
15 }
```

Below the code editor is a "Run" button. To the right, there is a results table with columns "s" and "o". The table displays the following data:

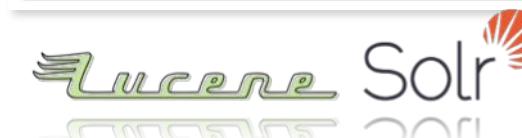
| s | o |
|----|--|
| 1 | vin:adjacentRegion |
| 2 | vin:locatedIn |
| 3 | http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine |
| 4 | http://www.w3.org/TR/2003/CR-owl-guide-20030818/wine |
| 5 | vin:Wine |
| 6 | vin:Winery |
| 7 | vin:Region |
| 8 | vin:Vintage |
| 9 | vin:WineGrape |
| 10 | vin:WhiteWine |

GraphDB Engine

- REST API for database access
- Plugin / Connectors

The screenshot shows the GraphDB Engine REST API documentation. It is organized into sections:

- repositories : Repository management**
 - sparql : SPARQL**
 - DELETE /repositories/{repositoryID}/statements** Deletes statements from the repository.
 - GET /repositories/{repositoryID}/statements** Fetches statements from the repository.
 - POST /repositories/{repositoryID}/statements** Performs updates on the data in the repository
 - PUT /repositories/{repositoryID}/statements** Updates data in the repository, replacing any existing data with the supplied data
- contexts : Contexts management**
- namespaces : Namespaces management**



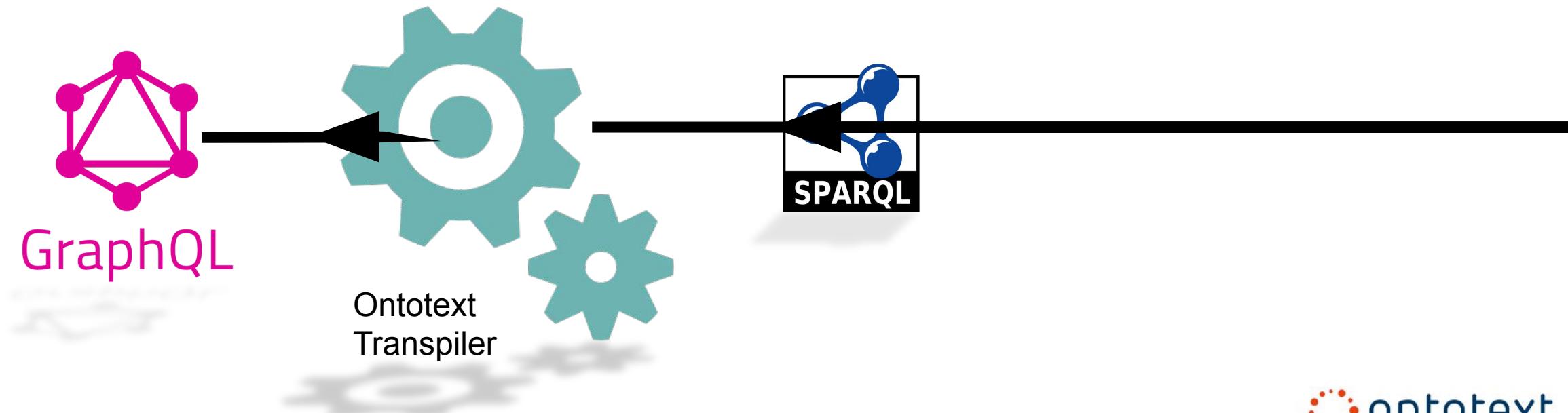
GraphDB Access and Interfaces

- **SPARQL Protocol** (end-point)
- **GraphQL**
- **REST APIs**
- **rdf4j open-source API**
 - Embedded and remote
 - Ontotext is major contributor to rdf4j
- **JDBC Driver** for SQL access to KGs
- **JS Driver** optimized for node.js



GraphQL Access via Semantic Objects

- Knowledge Graph access and updates via GraphQL
- Data validation via RDF Shapes
- Semantic Business Objects definitions done by business analysts
 - GraphQL Schema and shapes generated from Semantic Objects



GraphDB Editions

FREE

Delivers a fully functional database optimized for desktop use and small commercial prototypes:

- No constraints on data scale
- Limited to 1 concurrent Query thread
- Limited to 1 Inference Thread
- Lucene connector for FTS indexing

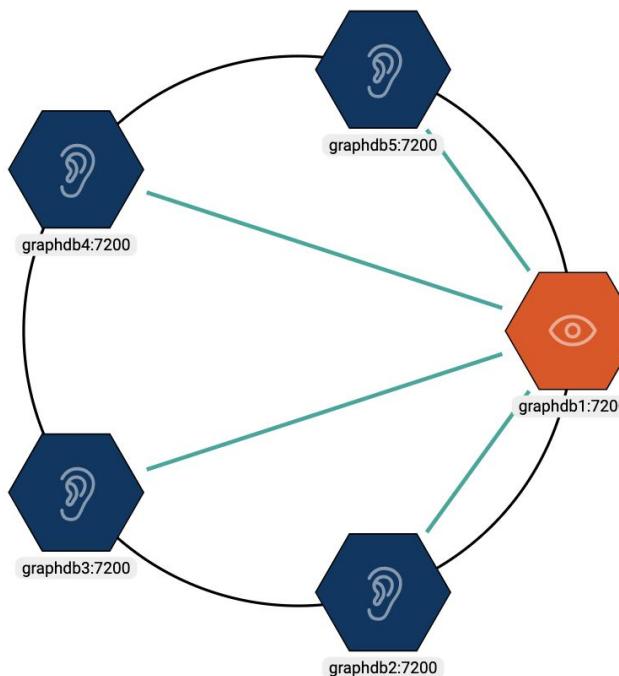
ENTERPRISE

Offers cluster support for enterprise resilience and high-availability:

- No single point of failure
- Multi-data center support
- Unlimited scalability of the read operations
- Elasticsearch and SOLR connectors

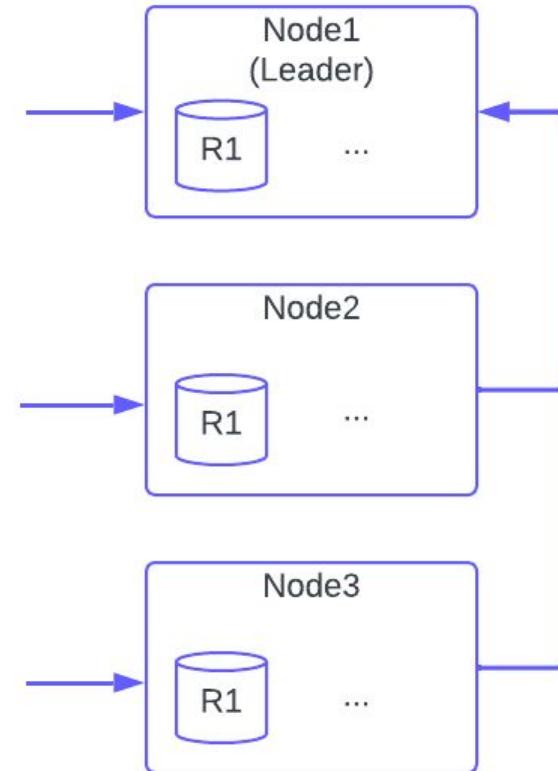
High Availability Cluster Architecture

- Coordinating all read and write operations
- Ensuring that all worker nodes are synchronized
- Propagating updates (inserts and deletes) across all workers and checking updates for inconsistencies
- Load balancing read requests between all available worker nodes



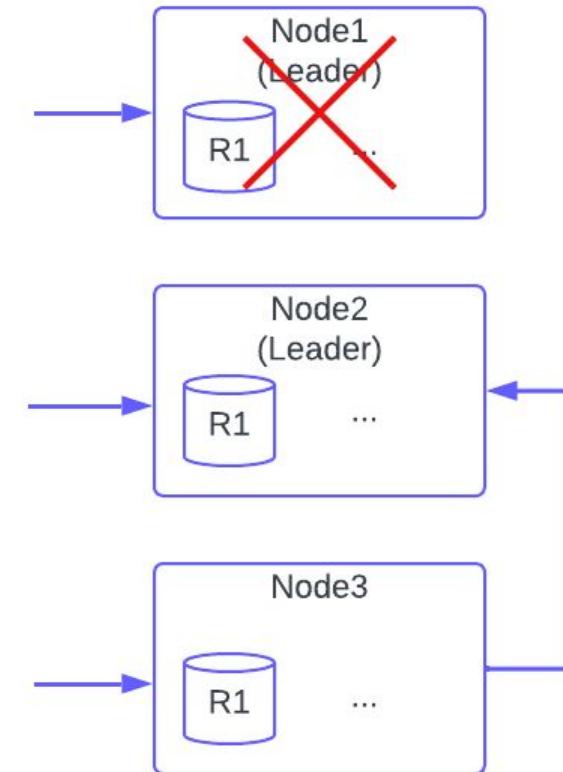
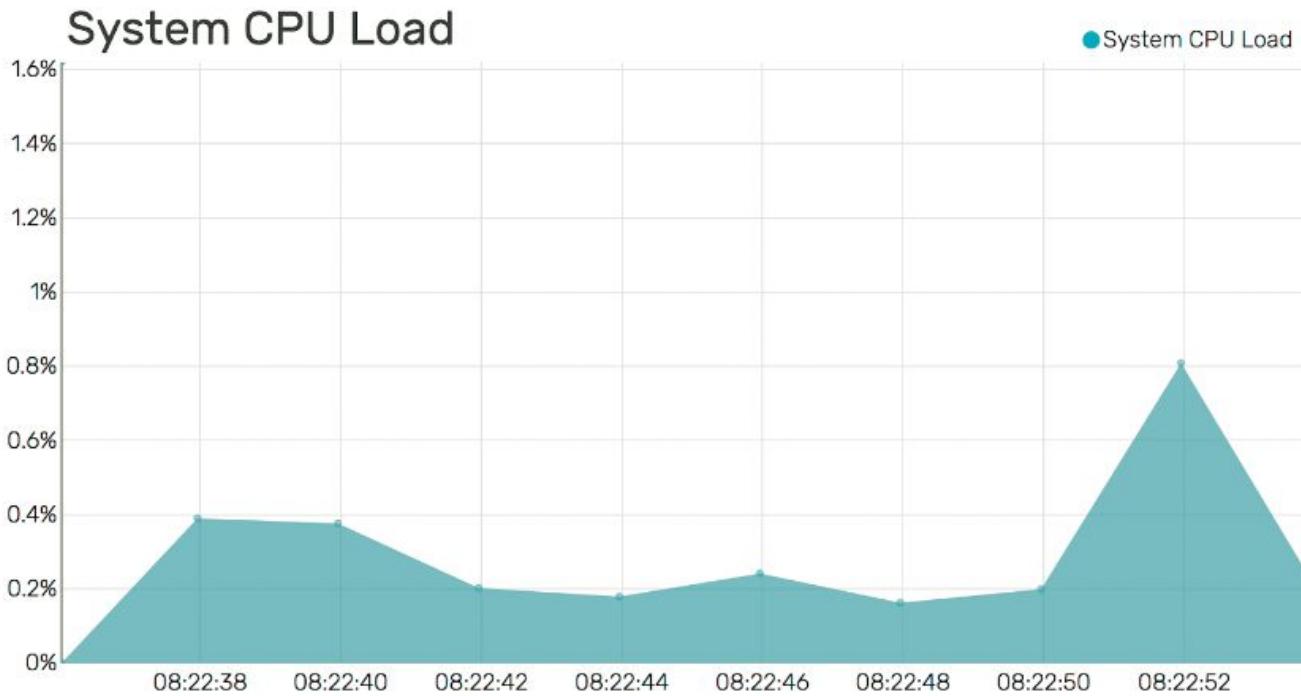
High Availability Cluster Architecture (ctd)

- **Improved resilience**
 - Failover, dynamic configuration
- **Improved query bandwidth**
 - Larger cluster means more queries per unit time
- **Deployable across multiple data centers**
- **Elastic scaling in cloud environments**
- **Integration with search engines**
- **Integration with MongoDB**



Cluster Management and Monitoring

- Automatic leader election to support recovery from any failure
- A smart client supporting multiple endpoints

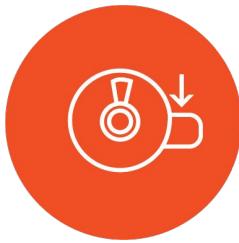


Enterprise-Grade Security and Access Control

- LDAP integration
- Kerberos
- OAuth support
- Role-Based Access Control (RBAC)
- Encryption in transit



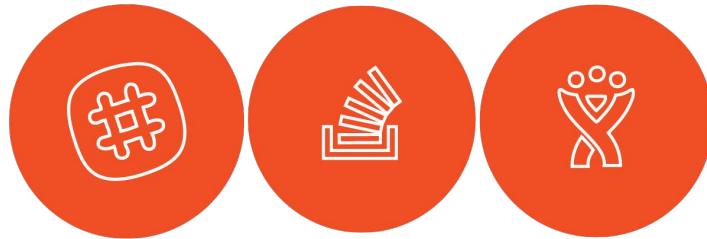
Easy Setup



- 100% Java based
- Platform Independent
- Native Installation Packages
- Open Maven
- Puppet Support
- Dockerized
- Kubernetes
- Cloud agnostic: AWS, Azure, Google Cloud, and on-premise



GraphDB Support



- Dedicated Support Team
- Community Support
 - StackOverFlow monitoring
- 24x7 Service Desk
 - Jira Issue-Tracking System
- Fully Managed Service and Custom SLAs (optional)
- Customized Runbooks
- Easy Slack Communication

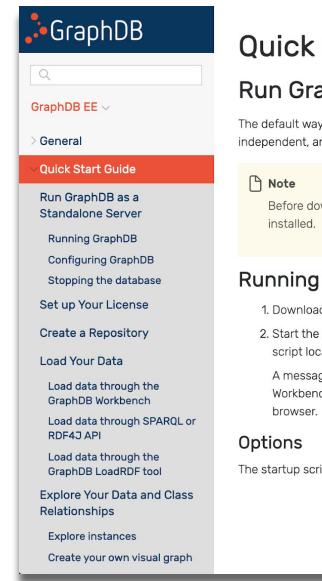
The collage includes:

- A screenshot of a Stack Overflow search results page for the query "-Denable-debug-rules=true not giving out statistics".
- A screenshot of a Stack Overflow question about Graph DB local data storage.
- A screenshot of the FactForge Jira issue tracking system.
- A screenshot of a Slack workspace showing a channel for #blogs and a post from a user named adam.
- A screenshot of a LinkedIn news feed.
- A screenshot of a Facebook news feed.

GraphDB Documentation & Training



- Always up-to-date
- Dedicated Doc team
- Startup Guide
- How-To Examples
- Video Tutorials
- Webinars and Live Trainings
- Online and on-site training courses



Quick Start Guide

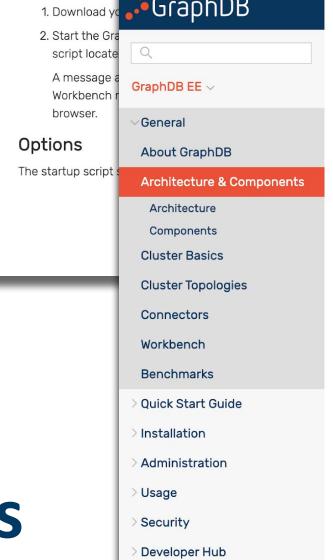
Run GraphDB as a Standalone Server

The default way of running GraphDB is as a standalone server. The server is platform-independent, and includes all recommended JVM parameters for immediate use.



Before downloading and running GraphDB, please make sure to have JDK or JRE installed.

Running GraphDB



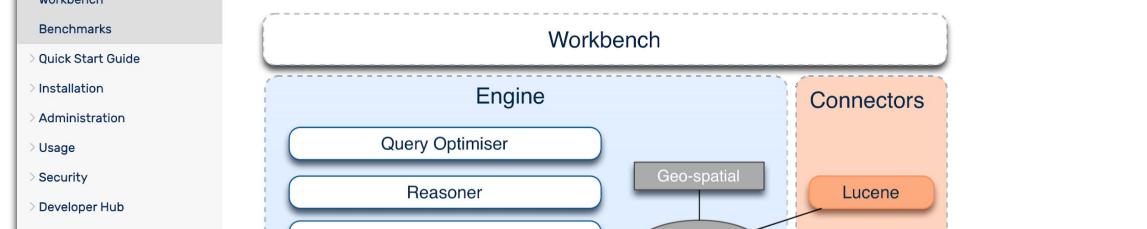
Architecture & Components

Architecture

GraphDB is packaged as a Storage And Inference Layer (SAIL) for [RDF4J](#) and makes extensive use of the features and infrastructure of RDF4J, especially the [RDF](#) model, RDF parsers, and query engines.

Inference is performed by the [Reasoner \(TRREE Engine\)](#), where the explicit and inferred statements are stored in highly optimized data structures that are kept in-memory for query evaluation and further inference. The inferred closure is updated through inference at the end of each transaction that modifies the repository.

GraphDB implements [The Sail API](#) interface so that it can be integrated with the rest of the RDF4J framework, e.g., the query engines and the web UI. A user application can be designed to use GraphDB directly through the RDF4J SAIL API or via the higher-level functional interfaces. When a GraphDB repository is exposed using the RDF4J HTTP Server, users can manage the repository through the embedded [Workbench](#), the RDF4J Workbench, or other tools integrated with RDF4J.





• GraphDB

- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

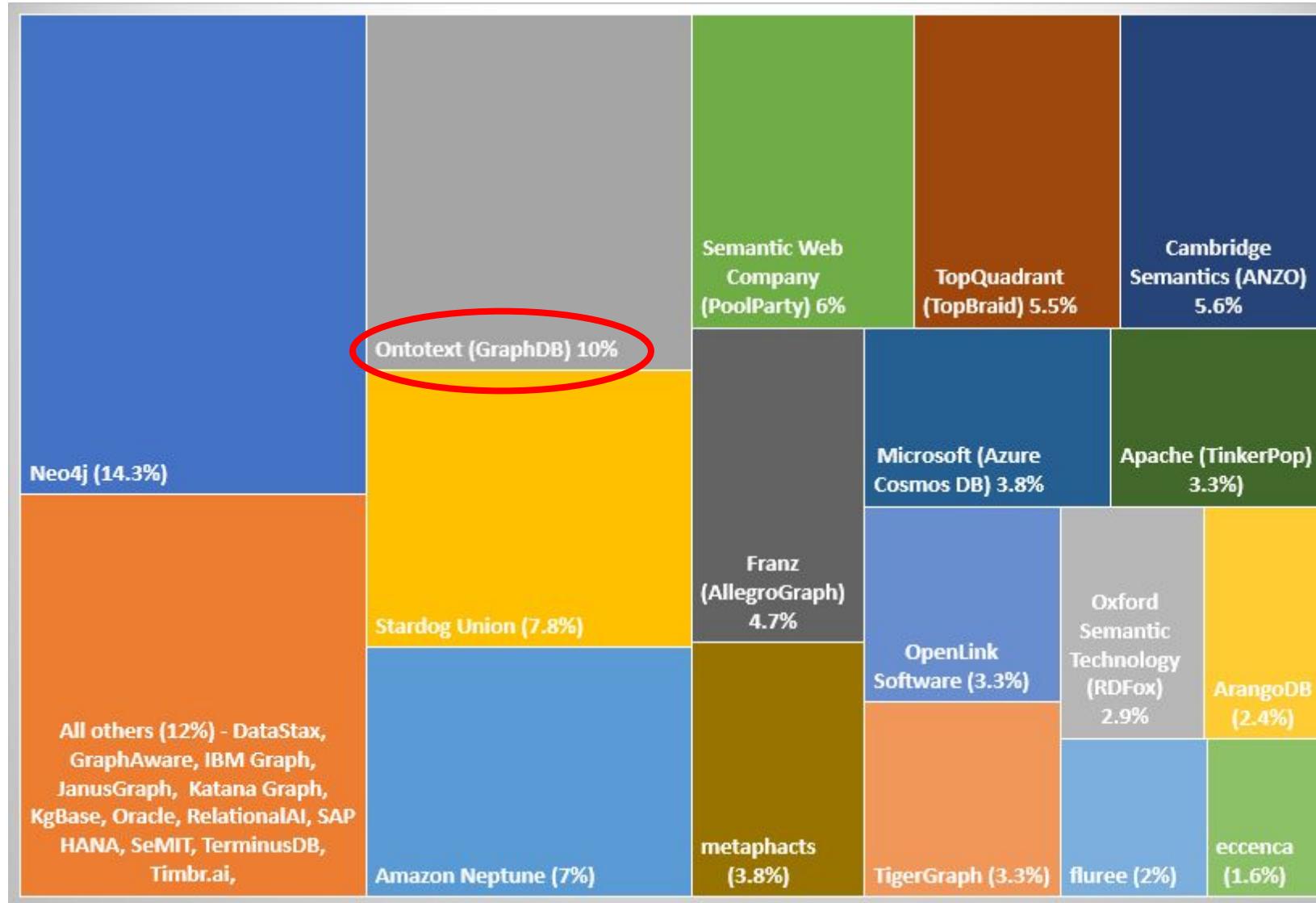
• Semantic Objects

• Ontotext Refine

- **Dependable performance** across a wide variety of workloads
 - ✓ The only engine passing both LDBC Semantic Publishing and Social Network Benchmarks!
 - Most of the competitors claim efficiency, but never publish audited benchmark results
 - Managing 1B facts in 16GB of RAM worth more than loading a trillion triples in a supercomputer
- **Ultimate extensibility, accessibility and deployment options**
 - ✓ Open-source workbench and engine plug-ins, FTS connectors, GraphQL, SQL/R2RML, JDBC, Kafka
 - ✓ Platform agnostic, Dockerized, portable 100% Java implementation
- **Standard compliance without compromises**
 - ✓ From OWL and SHACL to RDF-star and graph-path search in SPARQL
- **Regular release cycle, Stable quality & Clear Roadmap**
 - ✓ Robust dev. process allows for 5+ releases per year, rear bugs and no performance degradation

- **Scalable reasoning** across the full lifecycle of the data
 - ✓ Efficient reasoning: the only engine offering efficient inference *across all CRUD operations*
 - Stardog's backward-chaining fails when querying big data; ORACLE requires recompute upon update
 - ✓ SHACL data validation which does work well on sizeable transactions
- **Variety of analytics & search capabilities**
 - ✓ *Text analysis: built-in pipelines for entity linking against large knowledge graphs*
 - Corpus management environment, out-of-the-box vocabulary tagging and integration of 3rd-party NLP libraries
 - ✓ Semantic similarity search based on word and graph-embedding; autocomplete via PageRank
- **The richest set of seamlessly integrated partner tools**
 - ✓ Knowledge management, taxonomy & vocabulary management, ontology editors
 - ✓ Data management: data catalogues, ETL, data linking, graph & instance data editors
 - ✓ Search, exploration, visualization, chat bots

2nd Most Popular Knowledge Graph Tech Vendor



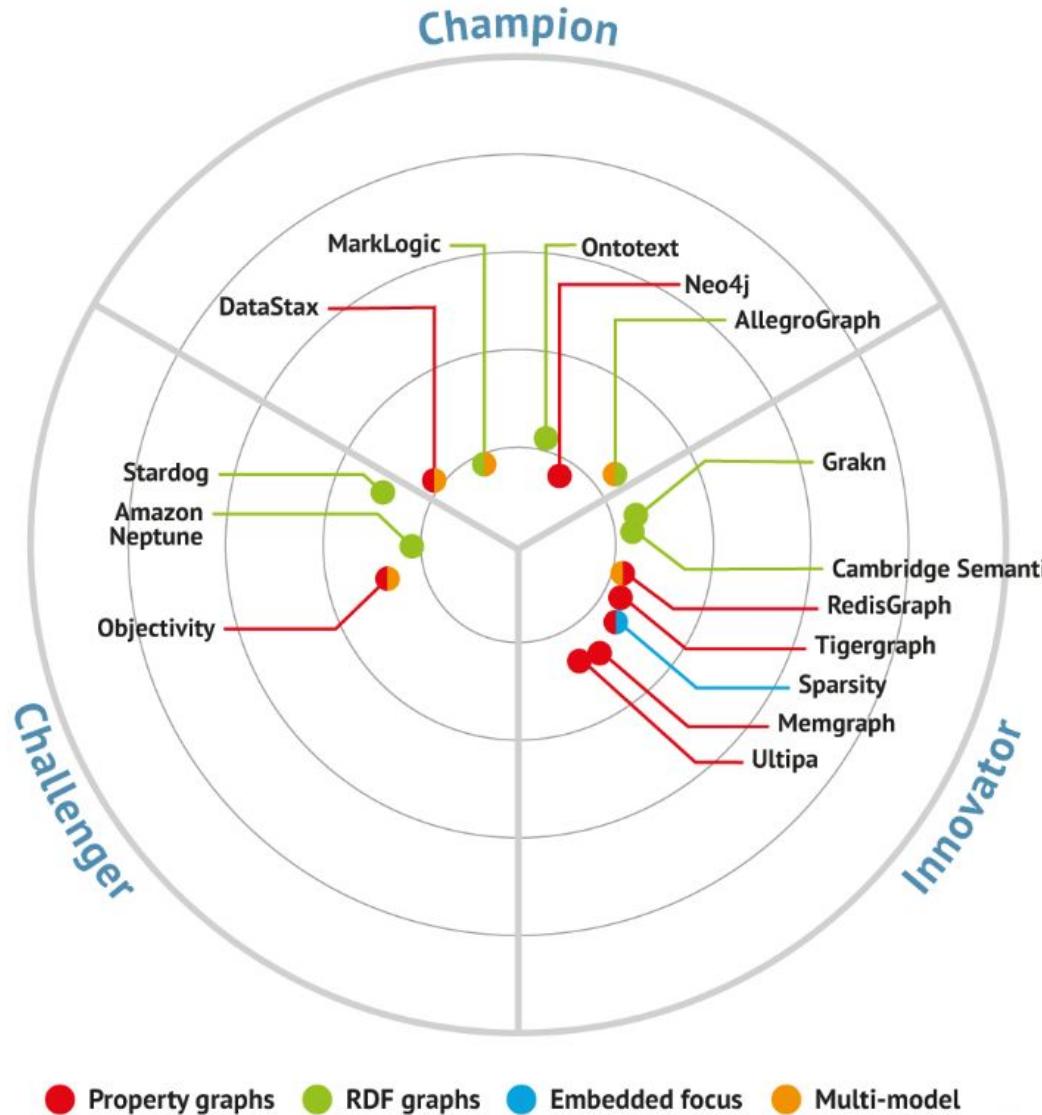
Source:
**Knowledge Graph Industry
Benchmarking Survey,**

*June 2022 • Data and Analysis
on Industry Maturity*

EKGf, KGC, Content Strategies

*Conducted among
representatives of 150
organizations at KGC'22*

Question: Which of the
knowledge graph vendors are
you testing or using?



We would argue that the **market leaders in this space continue to be Neo4j and OntoText (GraphDB), which are graph and RDF database providers respectively.** However, www.db-engines.com suggests that **MarkLogic is the leader in the RDF space.** This is a question of definition: **GraphDB is a pure-play RDF database with multi-model capabilities while MarkLogic is a multi-model database with an underlying XML engine that offers RDF capabilities.** In any case, they are both leading vendors in this space, along with **Amazon Neptune**

Bloor Research
Graph Database Market Update 2023

<https://www.bloorresearch.com/technology/graph-databases/>

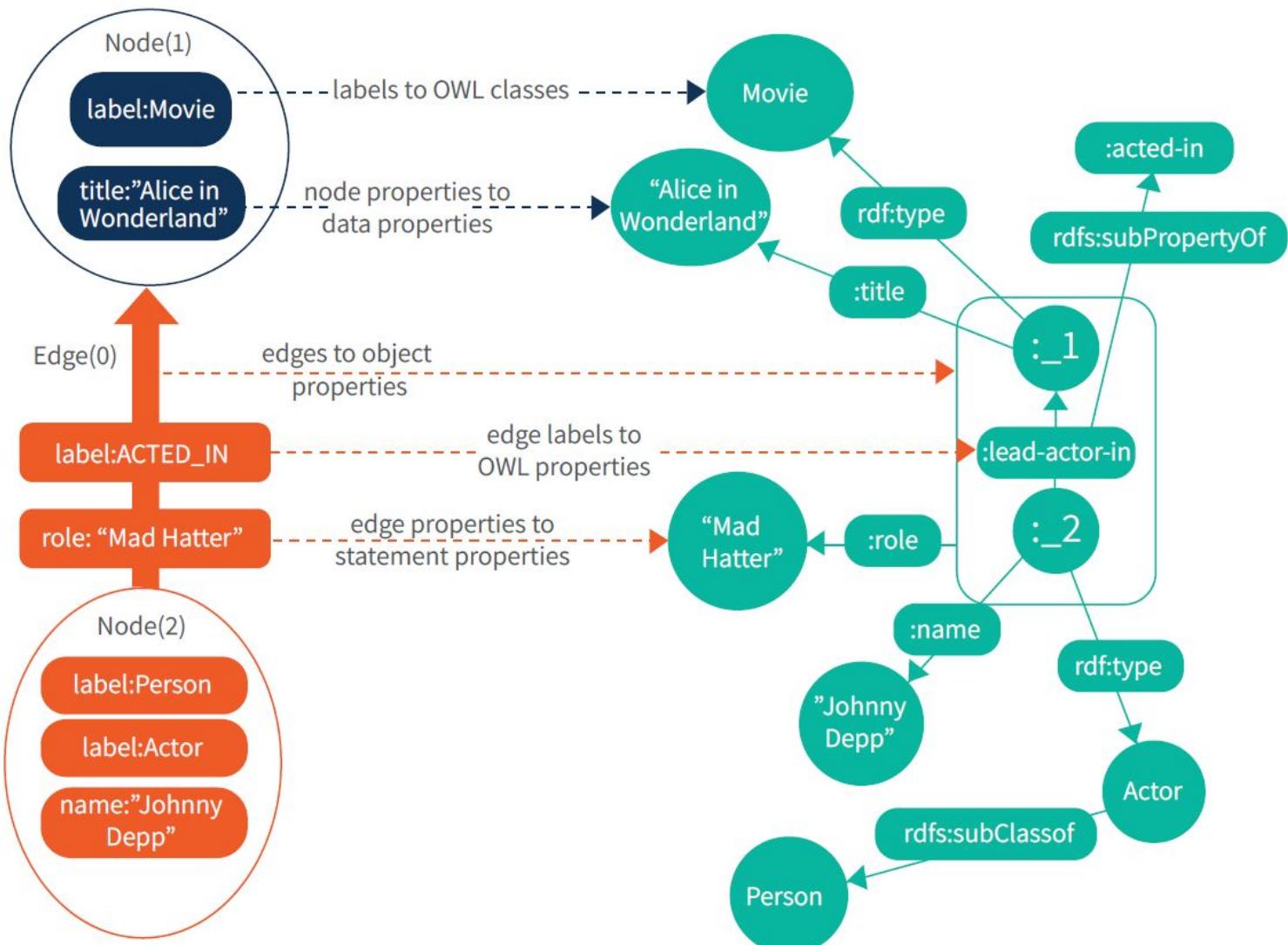
One of the Most Popular Graph Engines

| Rank | | | DBMS | Database Model | Score | | |
|----------|----------|----------|---|---|----------|----------|----------|
| Mar 2023 | Feb 2023 | Mar 2022 | | | Mar 2023 | Feb 2023 | Mar 2022 |
| 1. | 1. | 1. | Neo4j  | Graph | 53.51 | -1.92 | -6.16 |
| 2. | 2. | 2. | Microsoft Azure Cosmos DB  | Multi-model  | 36.10 | -0.40 | -4.79 |
| 3. | 3. | 4. | Virtuoso  | Multi-model  | 6.39 | +0.29 | +0.82 |
| 4. | 4. | 3. | ArangoDB  | Multi-model  | 5.04 | -0.26 | -0.57 |
| 5. | 5. | 5. | OrientDB | Multi-model  | 4.30 | -0.24 | -0.63 |
| 6. | ↑ 7. | ↑ 7. | Amazon Neptune | Multi-model  | 2.60 | -0.13 | -0.09 |
| 7. | ↓ 6. | ↑ 8. | JanusGraph | Graph | 2.56 | -0.24 | +0.09 |
| 8. | 8. | ↓ 6. | GraphDB  | Multi-model  | 2.32 | -0.12 | -0.52 |
| 9. | 9. | 9. | TigerGraph | Graph | 1.96 | -0.21 | -0.22 |
| 10. | ↑ 12. | ↑ 15. | NebulaGraph  | Graph | 1.83 | +0.04 | +0.70 |
| 11. | ↑ 13. | ↑ 20. | Memgraph  | Graph | 1.79 | +0.10 | +1.41 |
| 12. | ↓ 11. | ↓ 11. | Dgraph | Graph | 1.78 | -0.08 | +0.09 |
| 13. | ↓ 10. | ↓ 12. | Fauna | Multi-model  | 1.77 | -0.12 | +0.42 |
| 14. | 14. | ↓ 10. | Stardog  | Multi-model  | 1.77 | +0.14 | -0.13 |
| 15. | 15. | ↓ 13. | Giraph | Graph | 1.55 | -0.04 | +0.23 |

There are over 3 000 active installations of GraphDB running around the globe.

Source: db-engines.com [ranking of graph databases](#) based on popularity in social media, job portals, news, etc.

Note: This is not ranking by features or revenues – information on the later is not available for most of the vendors



- **LPG are easy to transform in RDF**
- **Advantages of RDF's finer grained model:**
 - ✓ Schema and data can be queried together
 - ✓ Property values are nodes and can be described
 - ✓ Full control on the structure of the metadata

GraphDB in the AECO Industry

- Johnson Control's BMS system: uses GraphDB in every installation
- Schneider Electric's BMS system: uses GraphDB in every installation
- Sequens, leader in social housing in the Ile de France region (310 municipalities, 105k properties, 230k tenants)
- SemmTech Laces, leading semantic asset management software (OTL/STL)
- ACCORD H2020 research project on Automated Compliance Checking
- Looking forward to collaborations with AECO companies!
- SSoLDAC 2023:
 - We hope that all students will use some GraphDB tools to develop their case
 - Post questions to WhatsApp SSoLDAC 2023 group, staff from Ontotext will answer



• GraphDB

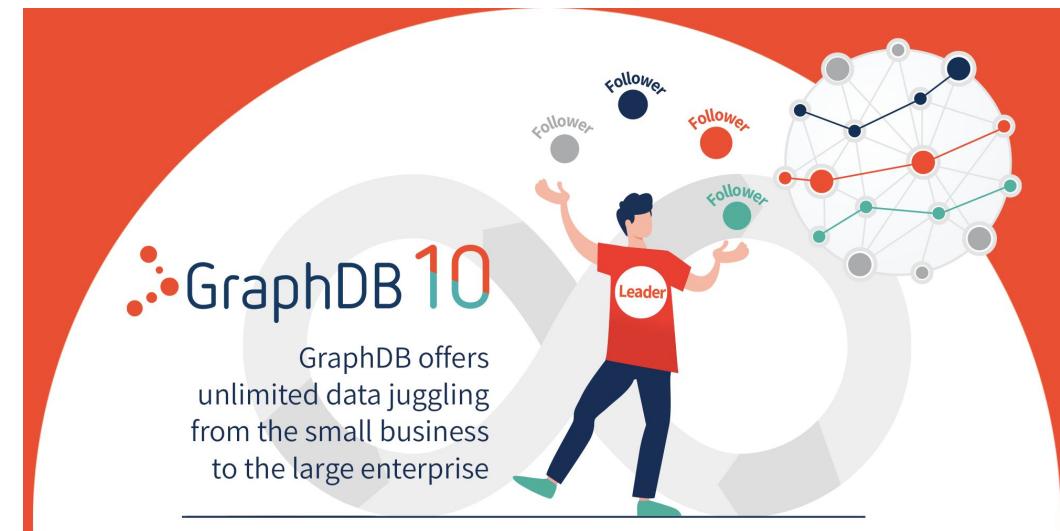
- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

• Ontotext Refine

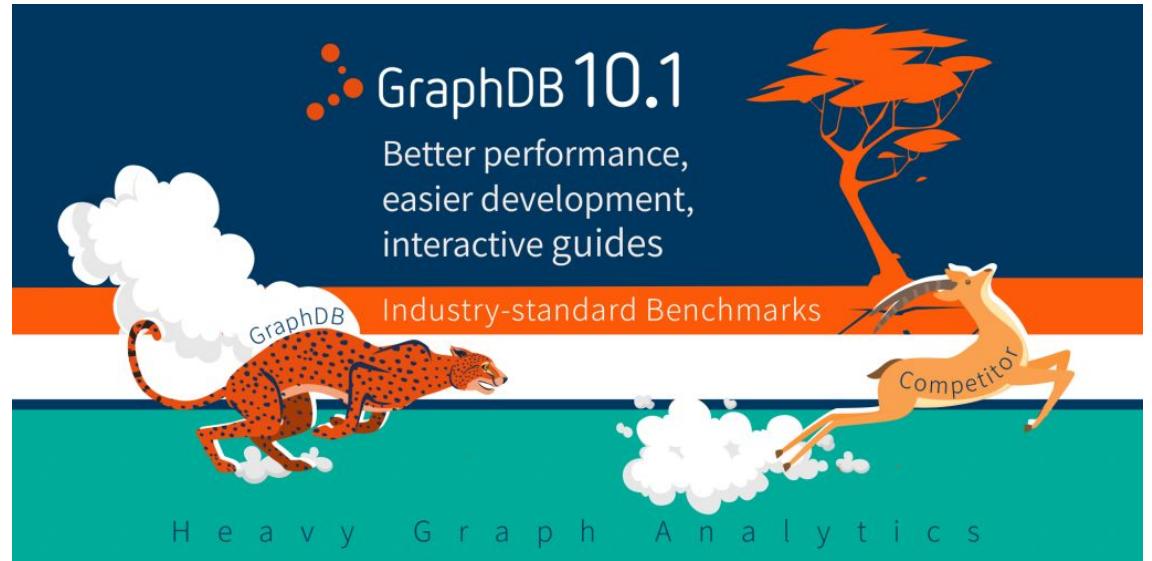
GraphDB 10.0

- New high-availability cluster
- Single distribution and repository type
- Licensing and parallelism
- Connector filtering redesign
- Upgraded to RDF4J
- Ontotext Refine made a standalone product (outside of GraphDB)



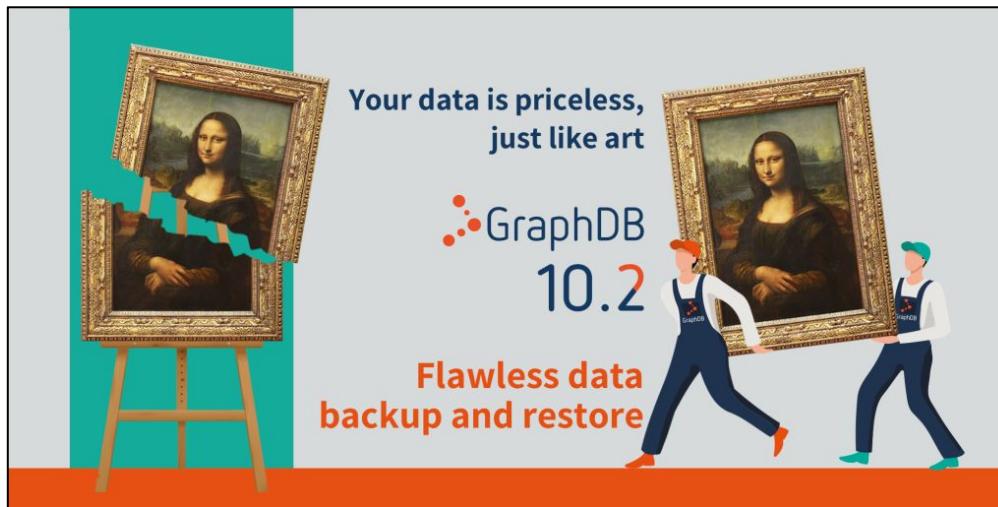
GraphDB 10.1

- Significant performance improvements and lower memory usage
- Easier start with the product
- New SPARQL function and better compatibility with Jena
- Simple full-text search
- Interactive user guides



GraphDB 10.2

- Improved cluster backup and support for cloud backups
- Lower memory requirements & improved transparency memory mode
- Better monitoring and support for Prometheus
- Flexible authentication options with X.509 certificate

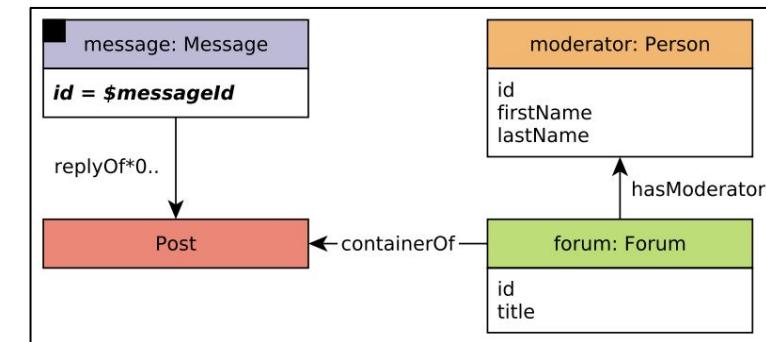


GraphDB 10.2 audited by LDBC

- GraphDB: the only audited engine passing multiple benchmarks

- LDBC Social Network Benchmark

- Social network graph with users browsing forums and messages
- **Insert operations** changing the graph connectivity
- Complex graph path search and analytical queries
- The most advanced graph analytics benchmark

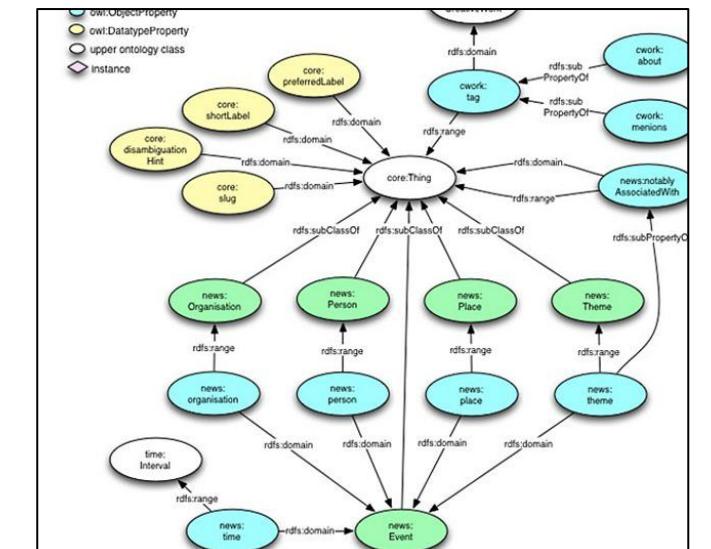


- LDBC Semantic Publishing Benchmark

- Based on BBC's Dynamic Semantic Publishing editorial workflow
- **Updates**, adding new metadata or updating the reference knowledge
- **Aggregation queries** retrieve content according to various criteria
- The only proper benchmark that involves reasoning and updates

Note: Benchmark results and significance elaborated in the following blog post

<https://www.ontotext.com/company/news/graphdb-passes-both-ldbc-benchmarks/>



GraphDB Roadmap

- Over 100,000 unique installations in the last year
- Over 7,620 active GraphDB installations in Jan 2023

| Strategic themes | Key features |
|--|---|
| Ease of use and dev experience | Language specific libraries, JSON-LD 1.1, UX upgrade |
| Improve the performance of the database | Better engine parallelization, Configurable indexes |
| Integrate ML & analytic algorithms | Integrate ChatGPT, Explain query results, |
| Integration with data management platforms | Increase the number of supported SQL sources |
| Easier consumption of knowledge graphs | Tighter integration with GraphQL |
| Deliver cloud offering over AWS and Azure | Publish GDB on AWS/Azure marketplace |
| Security and database reliability | Eliminate security vulnerabilities |
| Integrate with Large Language Models (ChatGPT) | 10.3 has functions to ask ChatGPT, and explain queries and results. Future: GraphDB to provide data to ChatGPT, natural language query to GraphQL and maybe SPARQL |



• GraphDB

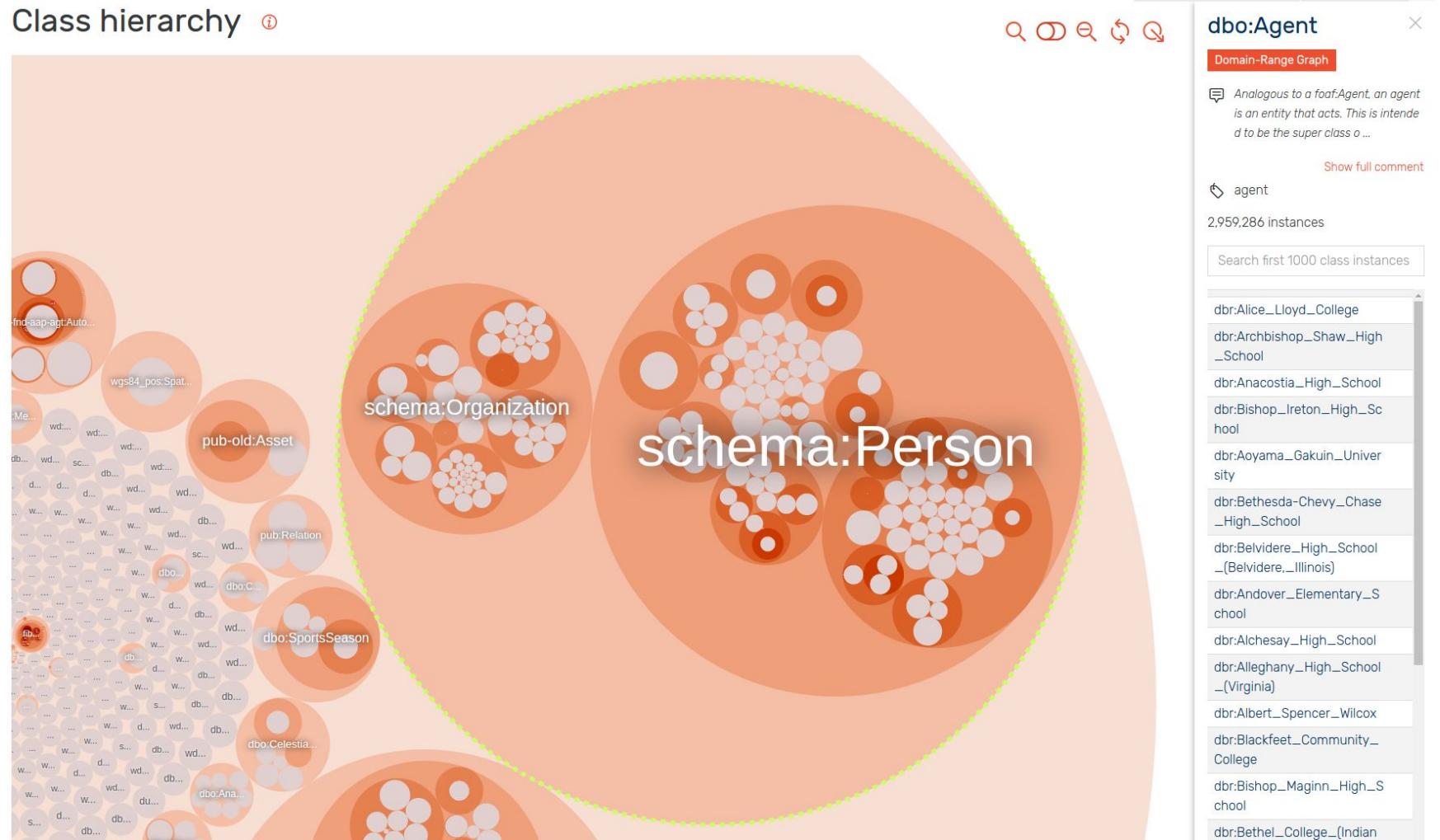
- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

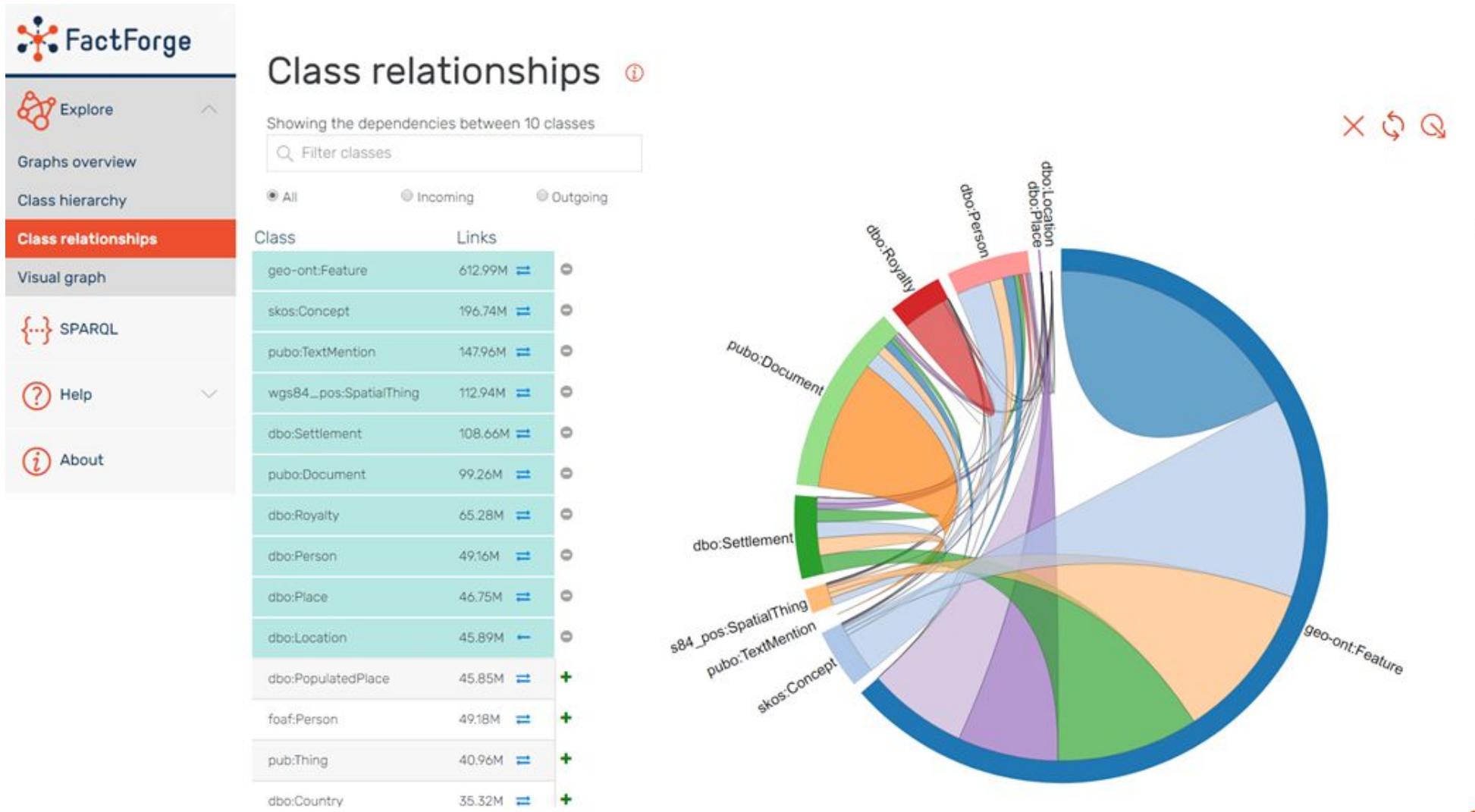
• Ontotext Refine

Class Hierarchy Exploration

- Explore ontologies of 1000+ classes
- Get sample instances

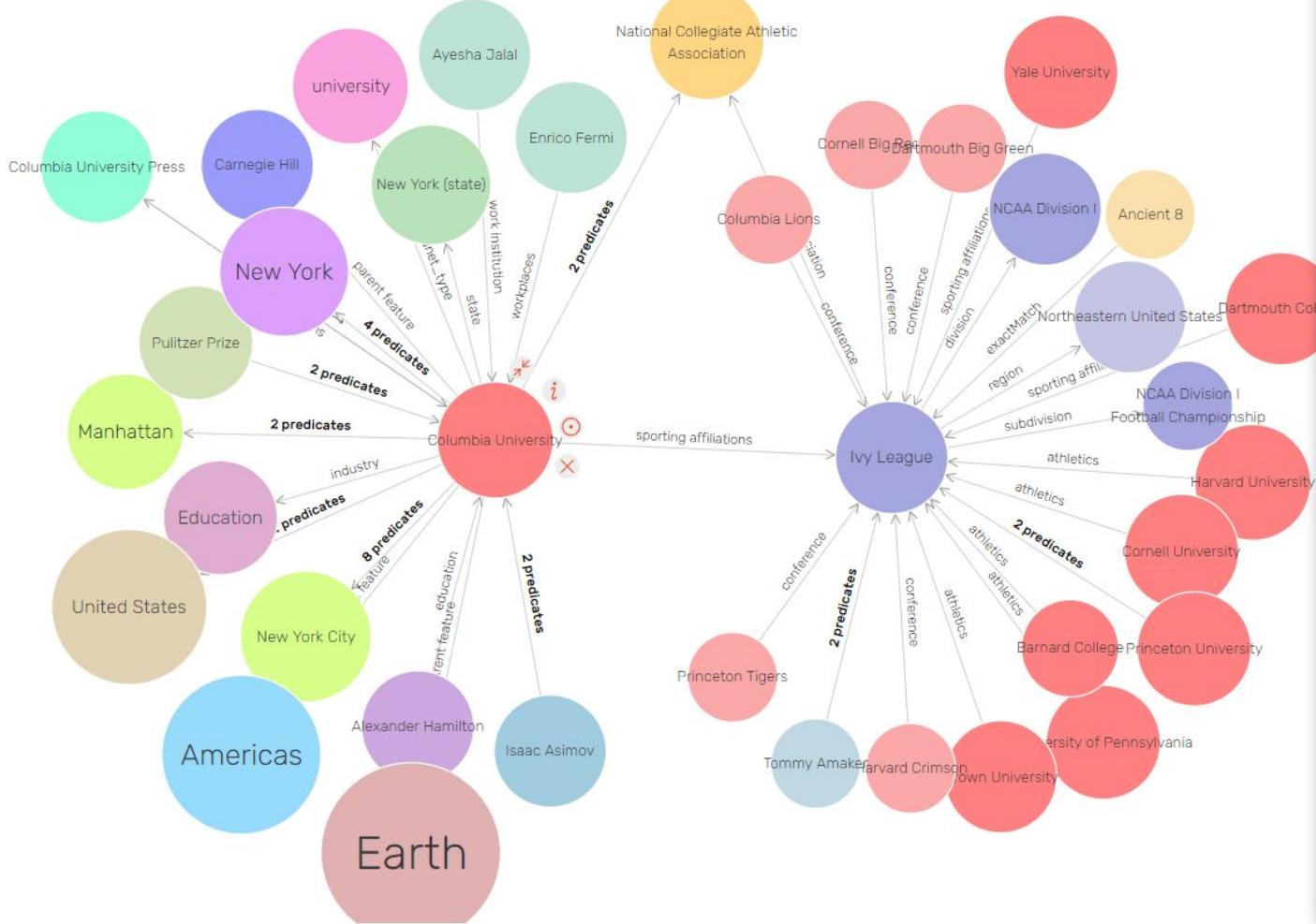


Explore Relationships Between Classes



Explore Nodes with Visual Graph

Visual graph [?](#)



Columbia University

Columbia University (officially Columbia University in the City of New York) is a private Ivy League ... [Show full comment](#)

[Columbia University](#)^{en}

Types:

- dbo:University
- gn:Feature
- geo-pos:SpatialThing
- skos:Concept
- wd:Q3918

RDF rank: 0.27

KING'S COLLEGE IN 1750

Search instance properties

dbo:abstract
Columbia University (officially Columbia

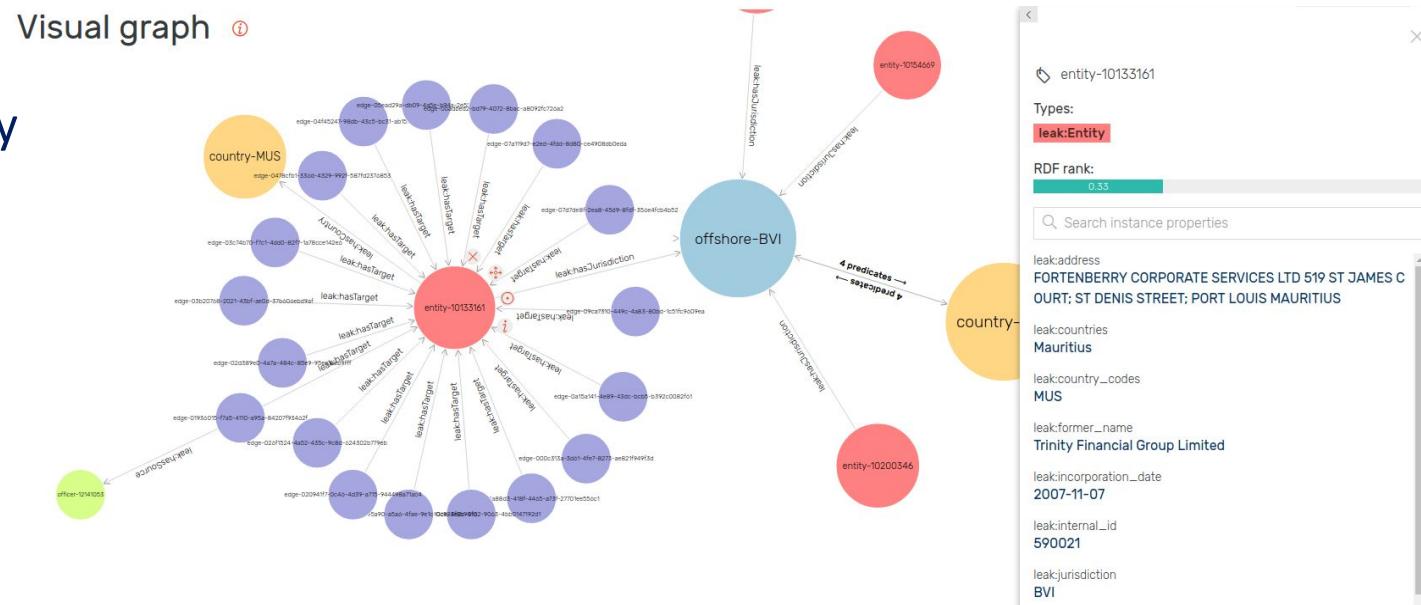
Explore Nodes with Visual Graph

- **Customize the visualization**

- Filter out abstract resources
- Include/exclude specific properties
- Generate new relationships on the fly
- Show/hide inferred statements

- **Better views via node importance ranking**

- Use GraphDB's RDFRank to get the importance/centrality rank of a node
- Use importance to chose the top-20 related nodes to be shown (configurable)
- Use importance ranks to size the nodes





• GraphDB

- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

• Ontotext Refine

GraphDB Benchmarking



- LDBC: TPC-like benchmarks for graph databases
- Members include: Ontotext, neo4j, CWI, UPM, ORACLE, IBM, *Sparsity
- LDBC Semantic Publishing Benchmark
 - Based on BBC's Dynamic Semantic Publishing editorial workflow
 - Updates, adding new metadata or updating the reference knowledge
 - Aggregation queries retrieve content according to various criteria
 - The only “proper” benchmark that involves reasoning and updates

LDBC SPB Results of GraphDB

- CPU: 1 x E5-1650
- RAM: 20G heap
- Dataset: LDBC SPB 256
- DB: GraphDB SE 8.0, RDF Statements:
 - 254,948,985 (explicit), 480,405,141 (total)
OWL-Horst-optimized rule set
- Creative works: 8,821,535

| Clients reading / writing | Reads/s | Writes/s |
|------------------------------|---------|----------|
| 0 / 1 | 0.0000 | 11.4067 |
| 0 / 2 | 0.0000 | 14.3033 |
| 0 / 4 | 0.0000 | 14.6700 |
| 0 / 8 | 0.0000 | 15.1067 |
| 1 / 0 | 17.8258 | 0.0000 |
| 4 / 0 | 43.0833 | 0.0000 |
| 8 / 0 | 70.3767 | 0.0000 |
| 16 / 0 | 83.2633 | 0.0000 |
| 8 / 2 | 52.5667 | 9.2867 |
| 8 / 4 | 54.0233 | 9.6167 |
| 8 / 8 | 54.9067 | 9.5733 |
| 10 / 2 | 59.9467 | 8.5333 |
| 10 / 4 | 62.2867 | 8.4767 |
| 10 / 8 | 61.7167 | 8.6067 |
| 16 / 2 | 68.8100 | 5.0600 |
| 16 / 4 | 70.3900 | 5.1067 |
| 16 / 8 | 70.2300 | 4.9967 |
| 16 / 16 | 70.9467 | 5.0567 |

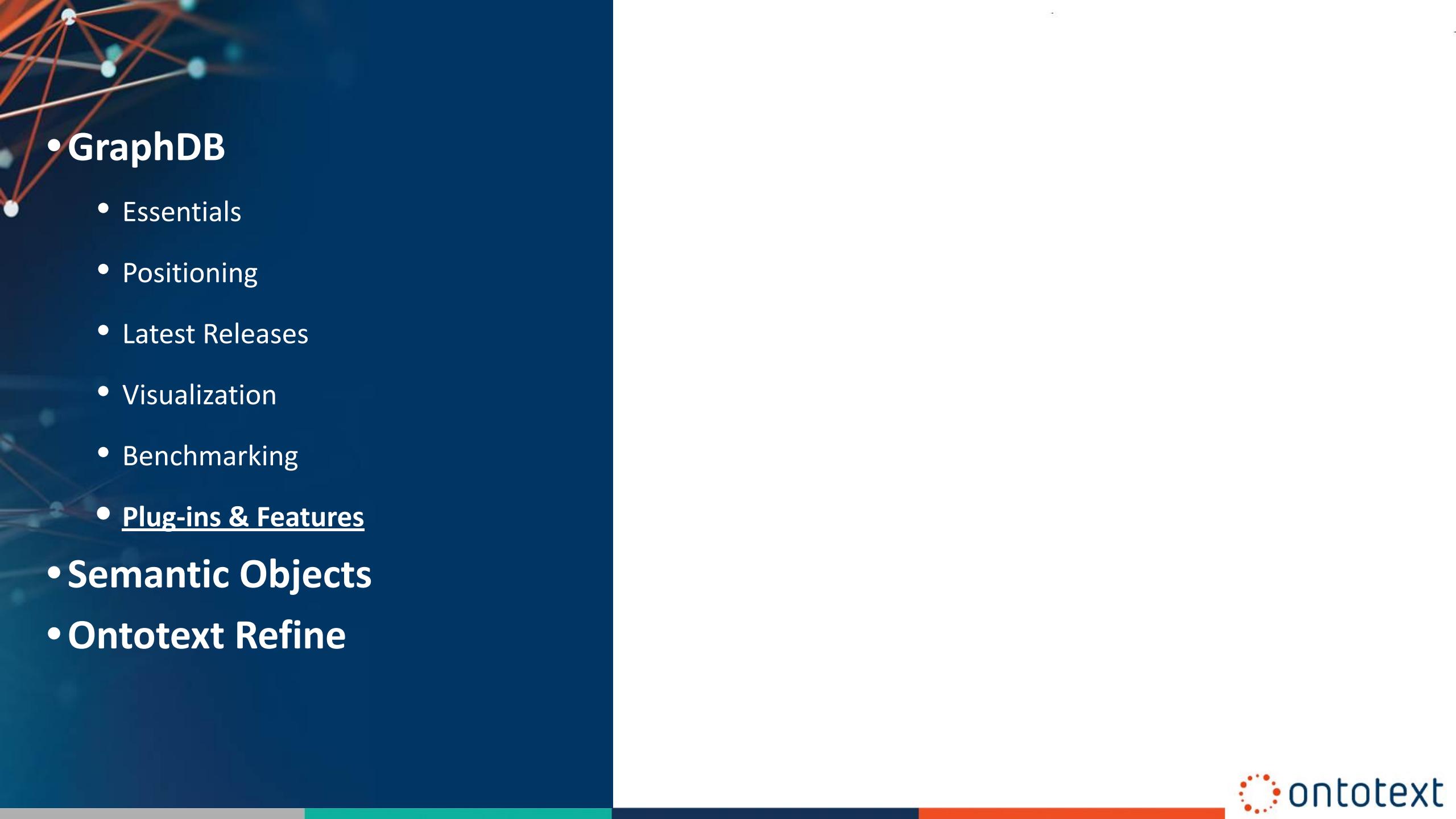
Data Loading: Semantic Publishing Benchmark 256M

| Editions | Ruleset | AWS instance | Cores | Loading time (min) |
|-----------|---------------------|--------------|-------|--------------------|
| 9.4 Free | RDFS-Plus-Optimized | i3.xlarge | 1* | 400 |
| 9.4 SE/EE | RDFS-Plus-Optimized | i3.xlarge | 2 | 316 |
| 9.4 SE/EE | RDFS-Plus-Optimized | i3.xlarge | 4 | 302 |
| 9.4 SE/EE | RDFS-Plus-Optimized | i3.2xlarge | 8 | 256 |
| 9.4 SE/EE | RDFS-Plus-Optimized | i3.4xlarge | 16 | 251 |
| 9.4 SE/EE | OWL2-RL | i3.large | 2 | 1625 |
| 9.4 SE/EE | OWL2-RL | i3.xlarge | 4 | 837 |
| 9.4 SE/EE | OWL2-RL | i3.4xlarge | 16 | 620 |

- Size: 237M explicit statements; Total: 385M (RDFS-Plus-optimized) or 752M (OWL2-RL)
- GraphDB Free uses single thread for data loading and reasoning

Production Load: SPB-256M

| AWS instance | Price / hour | Disk | Concurrent read agents | Read query mixes / sec | Concurrent write agents | Updates / sec |
|--------------|--------------|----------------|------------------------|------------------------|-------------------------|---------------|
| c4.4xlarge | \$0.796 | EBS (5K IOPS) | 0 | | 4 | 8.22 |
| i3.4xlarge | \$1.248 | local NVMe SSD | 0 | | 4 | 27.03 |
| c5d.4xlarge | \$0.768 | local NVMe SSD | 0 | | 4 | 30.39 |
| c4.4xlarge | \$0.796 | EBS (5K IOPS) | 16 | 52.27 | 0 | |
| i3.4xlarge | \$1.248 | local NVMe SSD | 16 | 60.81 | 0 | |
| c5d.4xlarge | \$0.768 | local NVMe SSD | 16 | 95.12 | 0 | |
| c4.4xlarge | \$0.796 | EBS (5K IOPS) | 12 | 28.41 | 4 | 2.56 |
| i3.4xlarge | \$1.248 | local NVMe SSD | 12 | 46.10 | 4 | 12.53 |
| c5d.4xlarge | \$0.768 | local NVMe SSD | 12 | 76.98 | 4 | 16.40 |



• GraphDB

- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

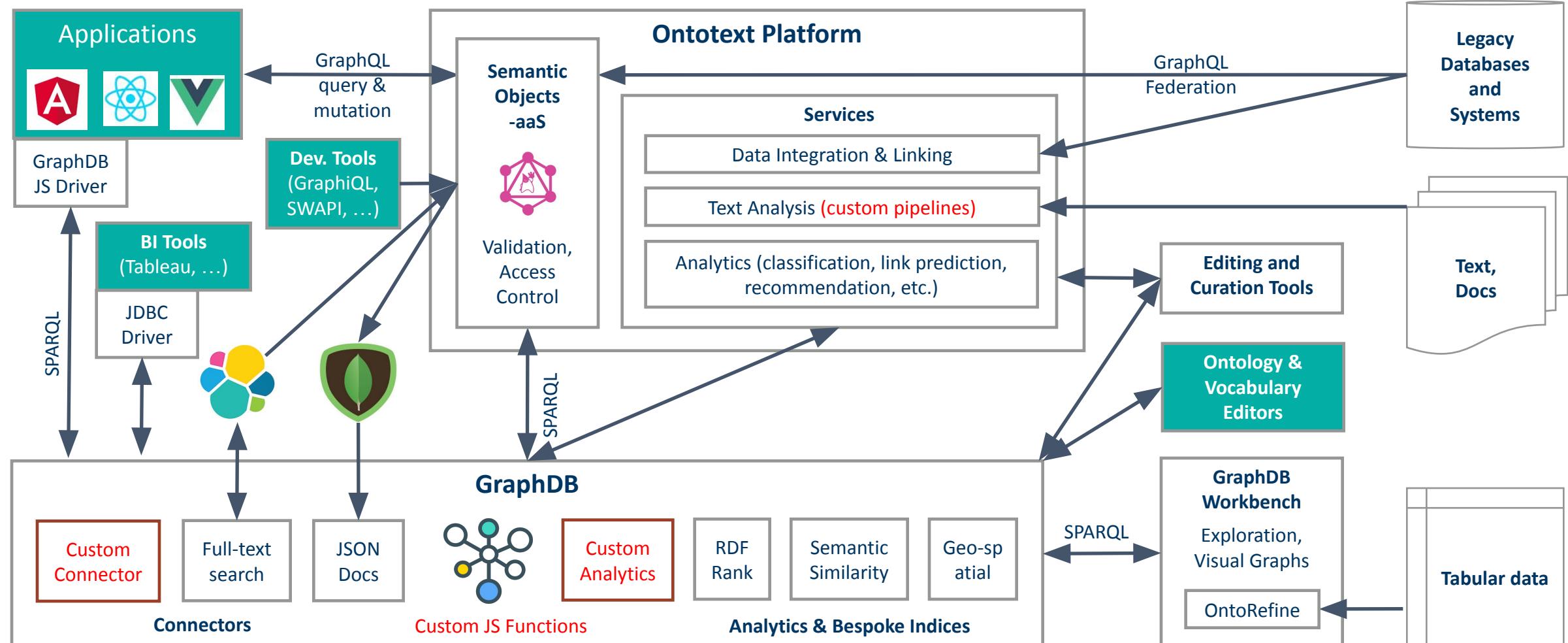
• Ontotext Refine

Extensible Plug-in Architecture

- **Special-purpose indices, connectors, analytics**
 - Accessing the internal data structures plug-ins can implement specific tasks very efficiently
 - Use a standard SPARQL syntax – no need of any special purpose non-standard languages
- **Sample Plug-ins**
 - **RDFRank**: node importance/centrality calculated via PageRank
 - **Geospatial**: supports functions like near-by via a special purpose index; GeoSPARQL
 - **FTS connectors**: search the KG via Lucene, Solr or Elasticsearch
 - **MongoDB connector**: access large documents and metadata from SPARQL
 - **Semantic Similarity** based on word and graph-embedding
 - **Change Tracking (History), Provenance and Proof** (why something was inferred)

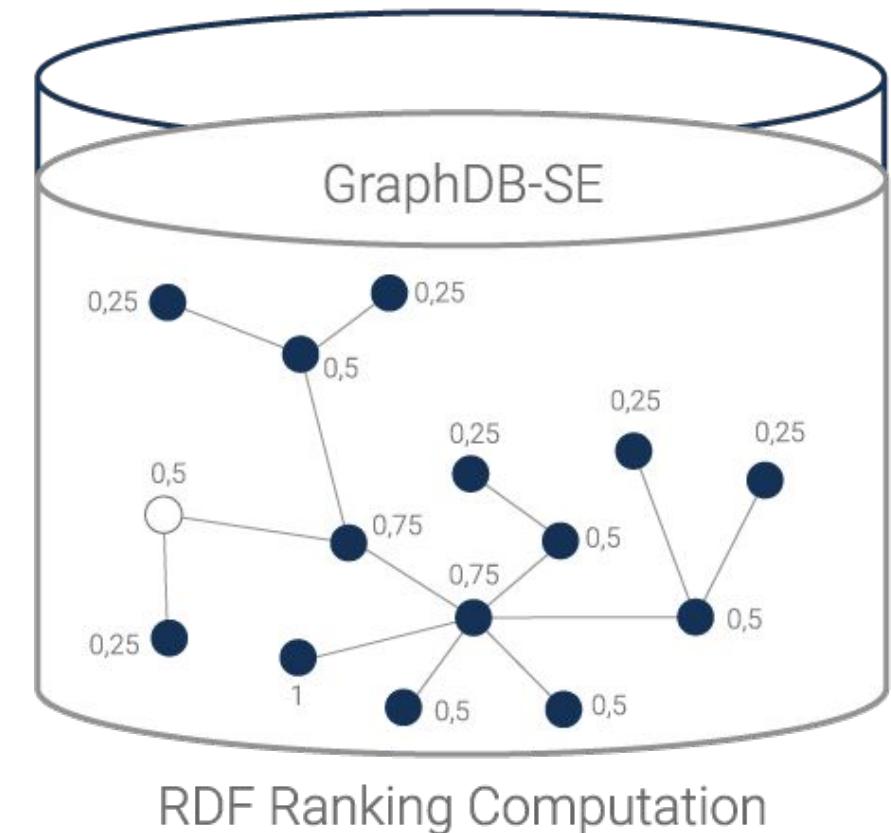
<https://graphdb.ontotext.com/documentation/standard/plug-ins.html>

Access, Customization and Integration Points



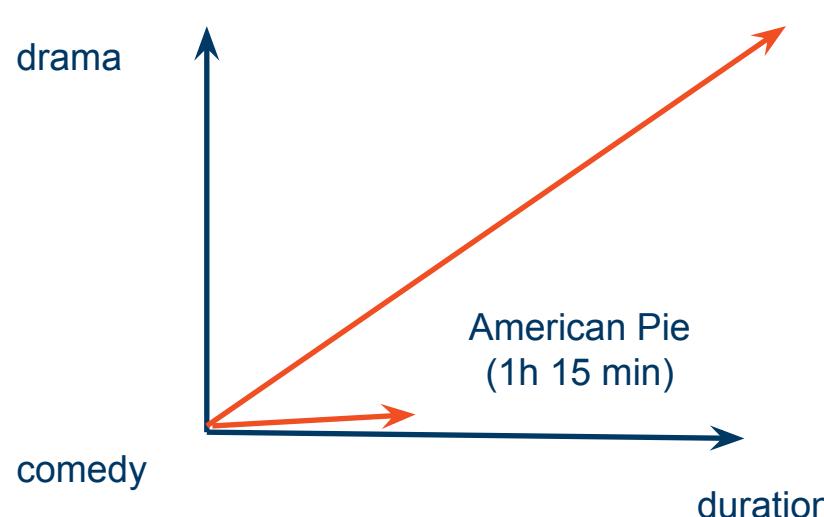
RDFRank: The Importance of a Node in a Graph

- **GraphDB's RDFRank plug-in computes PageRank**
 - Calculates “importance” based on nodes’ interconnectedness
 - Node ranks accessible via the rank:hasRDFRank predicate
 - Incremental RDF Rank calculation upon update is useful for dynamic data
- **RDFRank is used in GraphDB Workbench**
 - For ordering big lists of auto-suggest options in the SPARQL Editor and Search
 - To determine the node size of GraphDB’s Visual Graph
 - As feature for similarity ranking

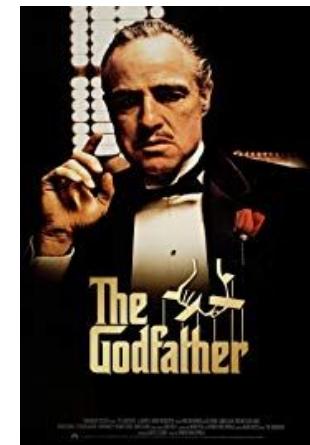


What is Knowledge Graph Embedding?

- Predict similar graph nodes or properties
- Require no input training data
- Mathematical representation of graph nodes as vectors



vs.



GraphDB Semantic Similarity Plug-in

- **Statistics similarity on KG using Semantic Vectors**
 - Creates statistical semantic models from your RDF data
 - Uses Lucene for scalable indexing and search for similar terms and documents
 - Reduced dimension vector space, e.g., 200 or 2000
- **Can judge similar nodes based on similar edges**
 - Not only an exact match of predicate-object pairs of the basic VSM
 - Example: <locatedIn, Manhattan> and <hasOfficeIn, New York City>
- **Both text-based and graph/predication-based embeddings**
 - Most interesting: combining text embeddings and graph embeddings models

Reasoning in GraphDB

- **Fast forward-chaining materialization**
 - Allows for efficient query evaluation on big datasets
- **Incremental for both inserts and deletes**
 - Inferred closure is updated transparently upon commit of transaction
- **Sample rules:**

ENTAILMENT

$p \text{ <} \text{rdf:type} \text{ >} \text{ <} \text{owl:FunctionalProperty} \text{ >}$

$x \ p \ y$

$x \ p \ z$

$y \text{ <} \text{owl:sameAs} \text{ >} z$

CONSISTENCY

$x \text{ owl:sameAs } y$

$x \text{ owl:differentFrom } y$

OWL 2 Reasoning

- Built-in rule-sets for: RDFS, RDFS+, OWL-Horst, OWL2-RL, OWL2-QL
- Optimized handling of owl:sameAs identifier mappings
- Custom rule-sets easily defined
 - Ruleset optimizer/profiler
- Configurations with multiple rule-sets
 - E.g., one with consistency checking to be used for internal data and another one with „open-world“ semantics for LOD and other external datasets
- Proof plug-in provides inference explanation



• GraphDB

- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

• Ontotext Refine

Semantic Objects: Easier Access of Knowledge Graphs

- **SPARQL Complexity**

- Skilled developers are required. SPARQL and RDF are perceived to be complex, difficult, unstable, and a niche technology stack. The average developer, customer, or enterprise just does not have the time, budget, or developers to make use of its power early in a product build

- **Developer community**

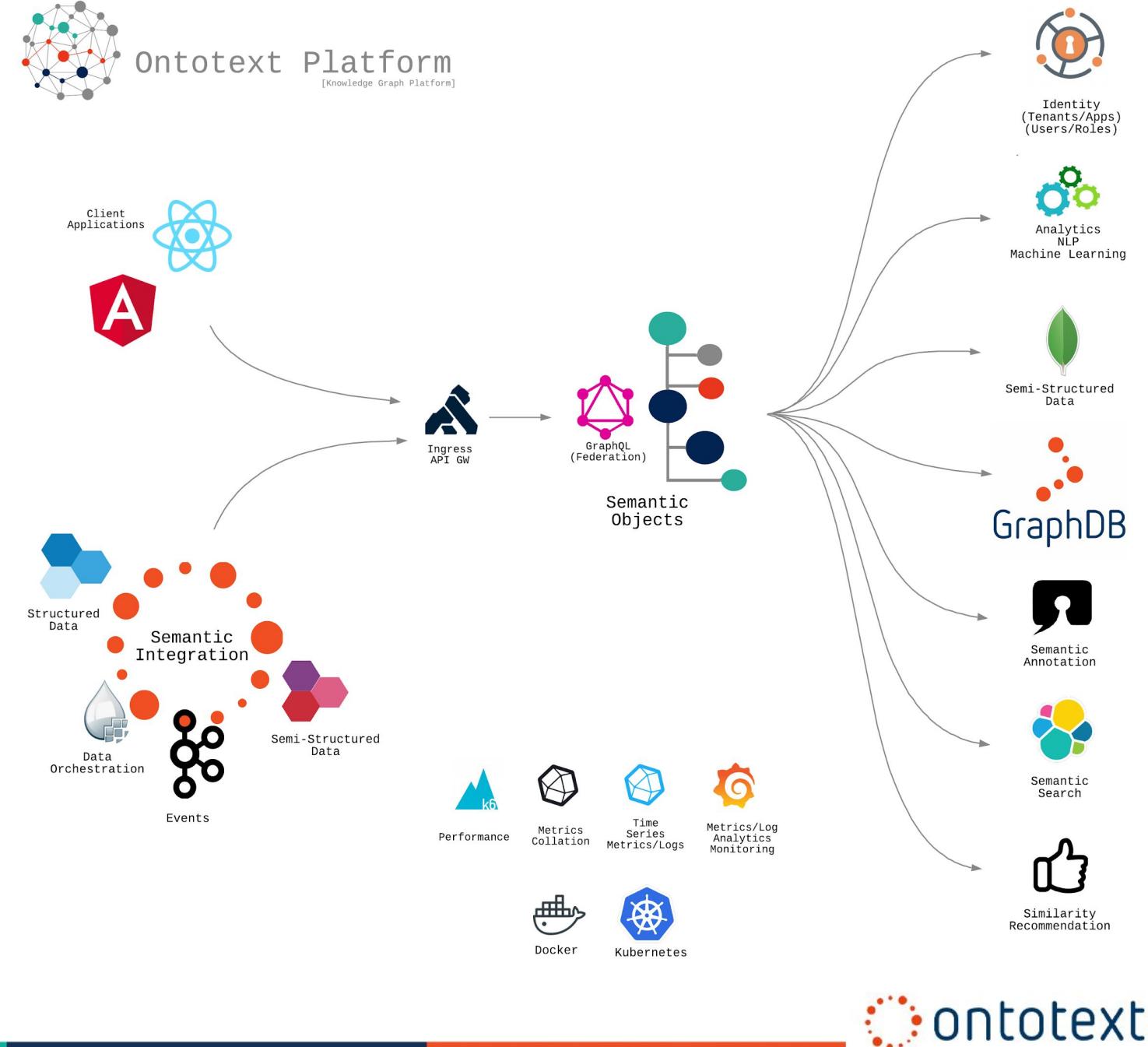
- Developers want something simple and easy that works most of the time. A groundswell of opposition has developed avoiding semantic web stacks due to complexity

- **Integration**

- New APIs are settling and moving towards GraphQL and JSON. Simple, declarative, and powerful enough for most use cases, GraphQL has a large developer community with many tools, frameworks, and huge momentum

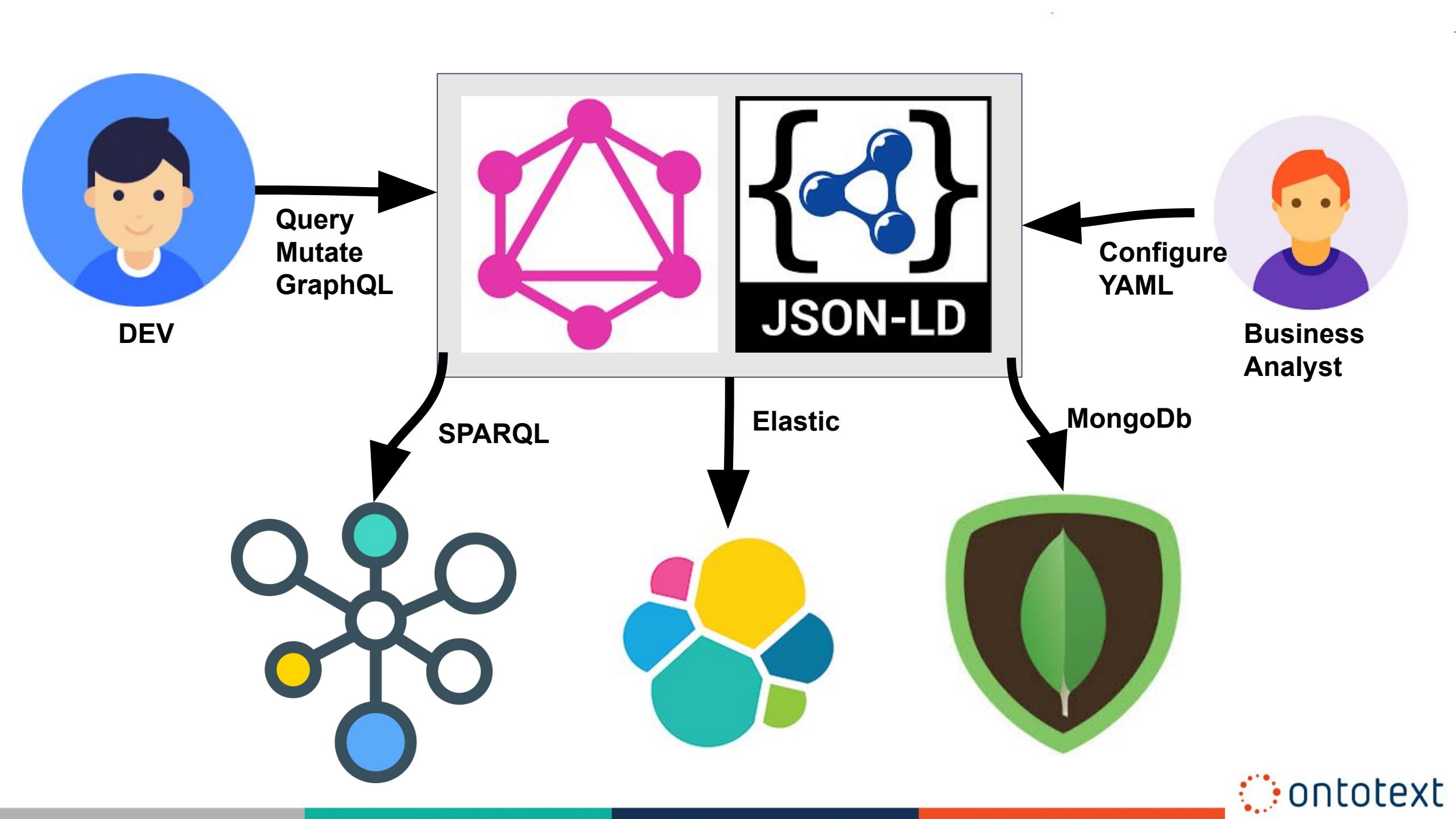
What is Ontotext Semantic Objects?

- Service for querying and mutating knowledge graphs
- Automatic GraphQL API generation
- Translation from GraphQL to SPARQL



Key Design Concepts

- **Fast knowledge graph development**
- **No server-side programming, no Object-RDF Mapping layers
(Backend as a Service)**
- **Automatic GraphQL API generation**
- **Optimized translation from GraphQL to SPARQL (Transpiling)**
- **Security over business objects**



Why Do I Need Another Modelling Language?

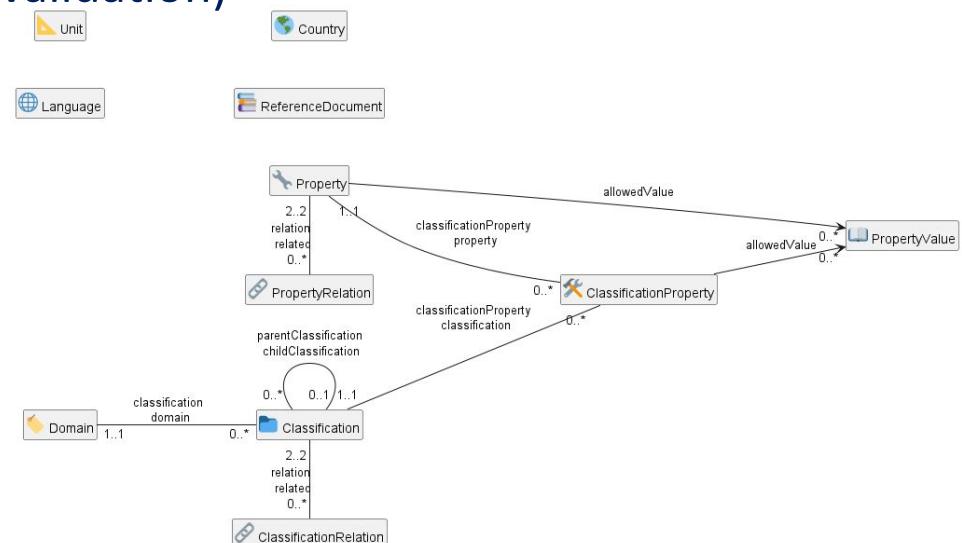
- **Semantic Object Modeling Language (SOML): simple YAML-based schema description:**
 - Makes a “bounded context” view over a Knowledge Graph
 - Can be generated from RDFS/OWL/Schema.org ontology
 - Generates GraphQL schema, query language (where, limit, offset, orderBy), API
 - Generates SHACL shapes (simple for referencing, full for validation)

• Why yet another modelling language?

- Ontologies often lack essential details like attachment of properties to objects, cardinalities, etc

- Can I see a schema graphically?

- ### ○ SOML visualization



GraphQL Overview

- **GraphQL is good for APIs with nested/interconnected data**
- **GraphQL has declarative schema (strong types) guiding its users**
- **Tooling helps developers (queries practically write themselves!)**
 - GraphiQL, GraphQL Playground, Typescript checking...
- **The client is king: they get exactly the data they want**
 - Hard for the server but easy for the client
- **Efficient translation (transpiling) of GraphQL queries to SPARQL**
 - Handle details like cardinalities (optional, union), avoiding Cartesian Product, field aliases...
 - Optimize based on database statistics
 - GraphQL has top-down execution strategy of nested queries;
SPARQL/SQL has bottom-up execution strategy of nested queries;
So we implemented Lateral Join (Coordinated Subqueries) in SPARQL

GraphQL access

- **Semantic Objects**

- Write SOML schema describing objects, props, cardinalities, RDF binding
- Or generate from ontology ([owl2somi](#)) then tune: bind props to classes, specify cardinalities, virtual inverses...
- Generates GraphQL schema and queries (**where language**)
- Lang tag handling
- SPARQL Templates for specialized queries
- Generates GraphQL mutations and SHACL shapes for validation
- Dynamic transpiling of GraphQL queries to SPARQL (next slide),
- Conversion of results to hierarchical JSON
- Example of Polyglot Modeling ([yaml-ls#19](#))

- **Semantic Search**

- Adds Elastic indexing and search

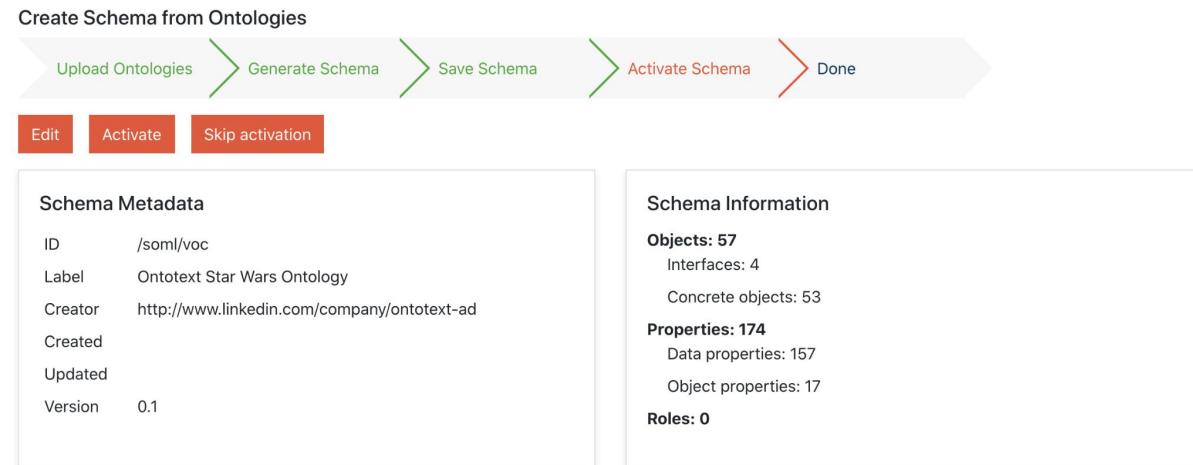
- **Powerful access control**

- **Federation**

- [Apollo Federation](#), [GraphQL Mesh](#), nullability

- **Workbench (interactive IDE)**

- Generate/edit SOML schemas
- Explore GraphQL schema
- Make queries (GraphiQL)



GraphiQL

query Tatooine { planet(orderBy: {name: ASC} where: {name: {IRE:"tatooi"}}) { id name type climate resident(orderBy: {height: ASC, name: DESC}) { name starship{ id } } } }

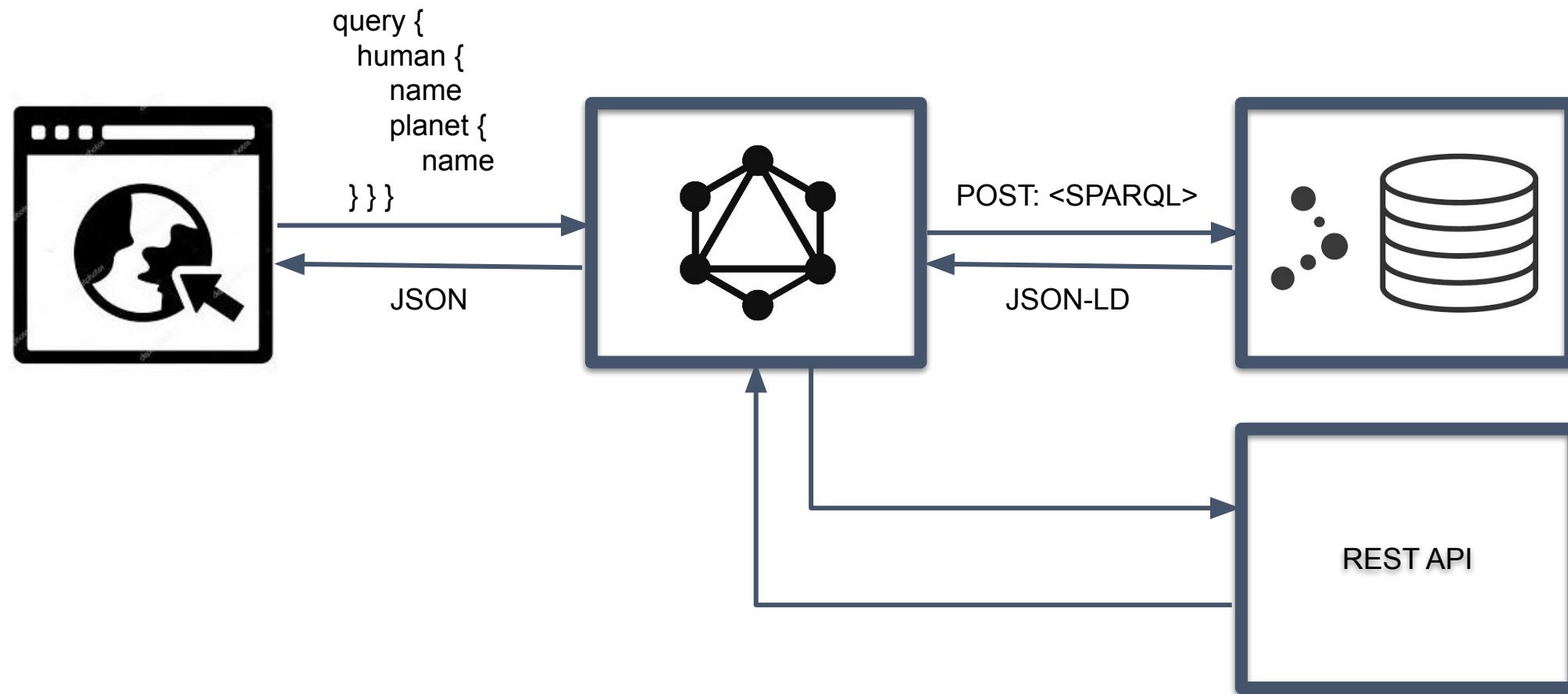
QUERY VARIABLES

Schema

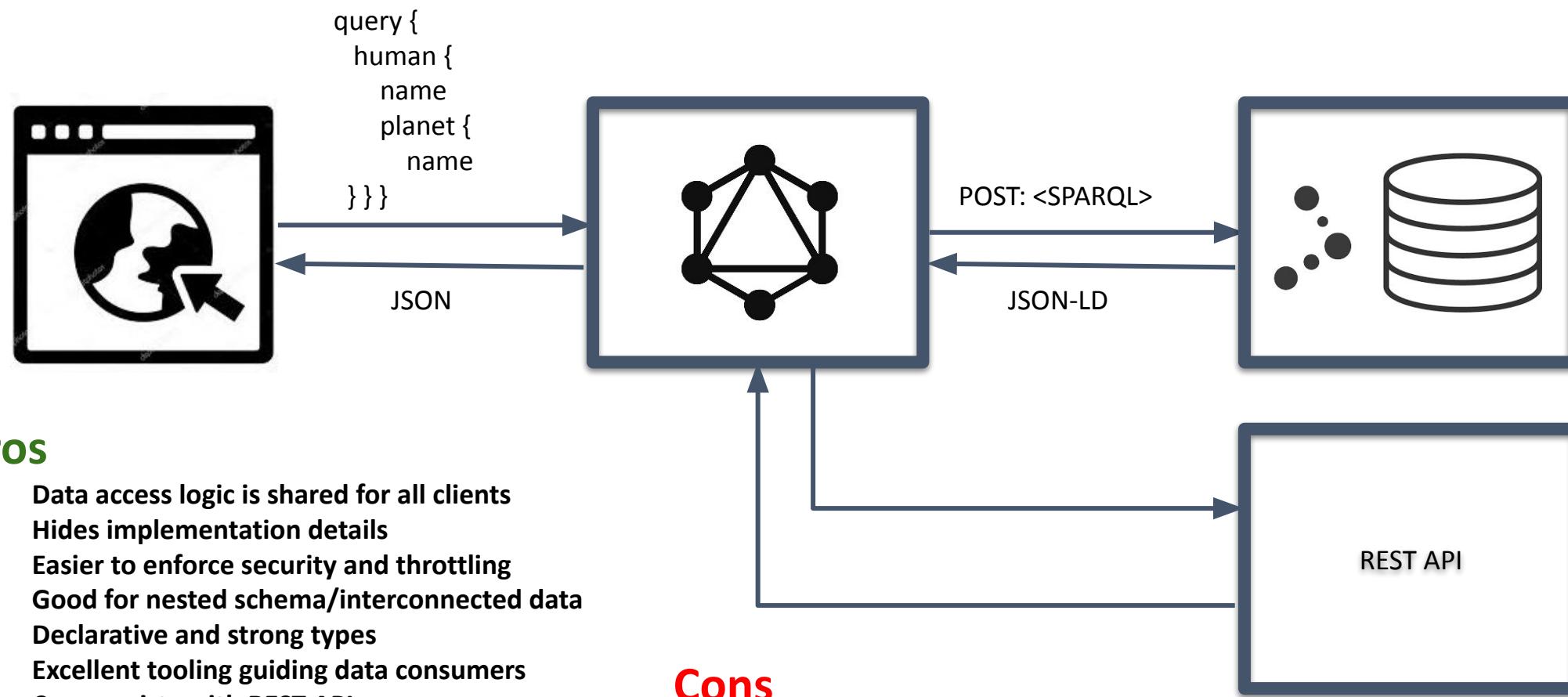
planet(
ID: [ID!]!
orderBy: Planet_OrderBy
where: Planet_Where_Multi
limit: PositiveInteger
offset: PositiveInteger
lang: String
): [Planet]!
Query for Planet

querrian(

GraphQL API



GraphQL API



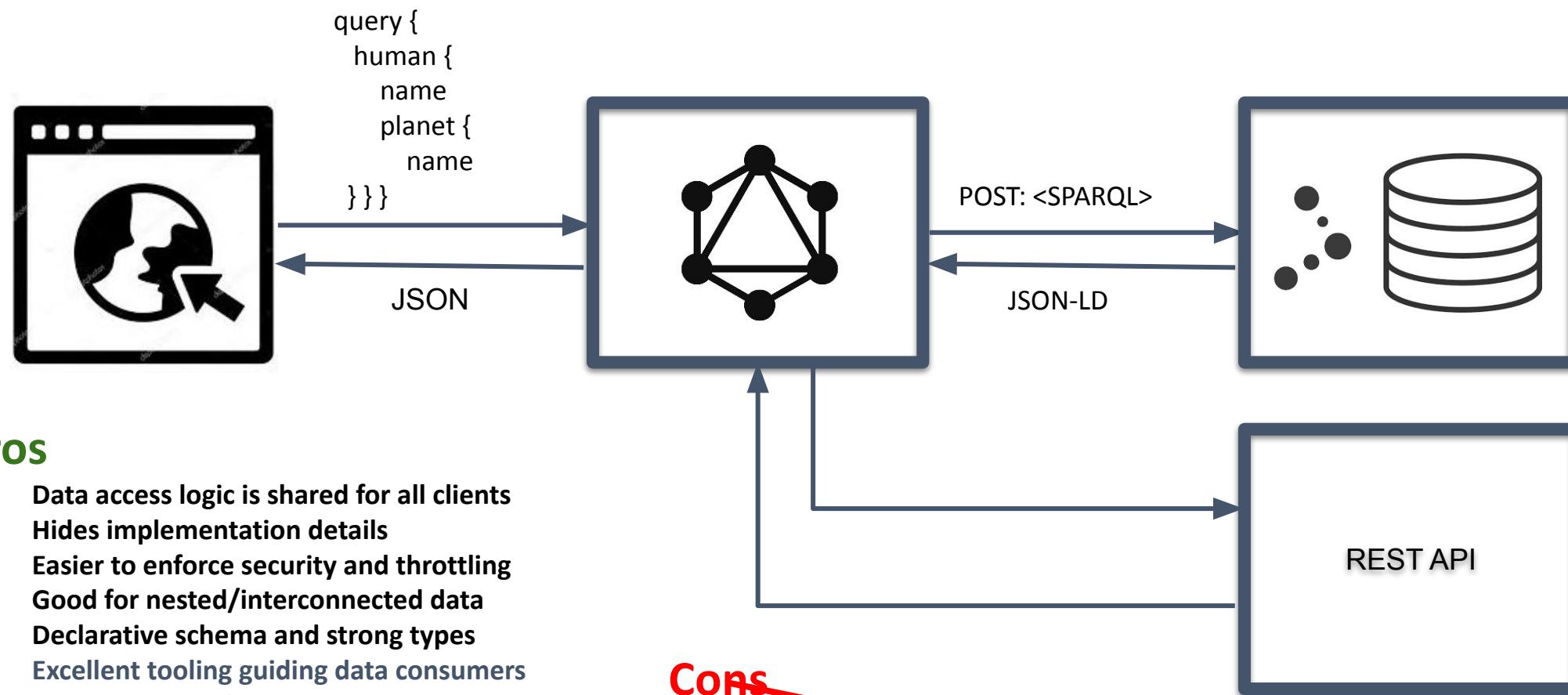
Pros

- Data access logic is shared for all clients
- Hides implementation details
- Easier to enforce security and throttling
- Good for nested schema/interconnected data
- Declarative and strong types
- Excellent tooling guiding data consumers
- Can coexists with REST APIs

Cons

- Harder to implement server logic
- Need tooling to generate GraphQL schema
- Need to implement GraphQL querying language (where with nested filters, limit, offset, orderBy...)

GraphQL API



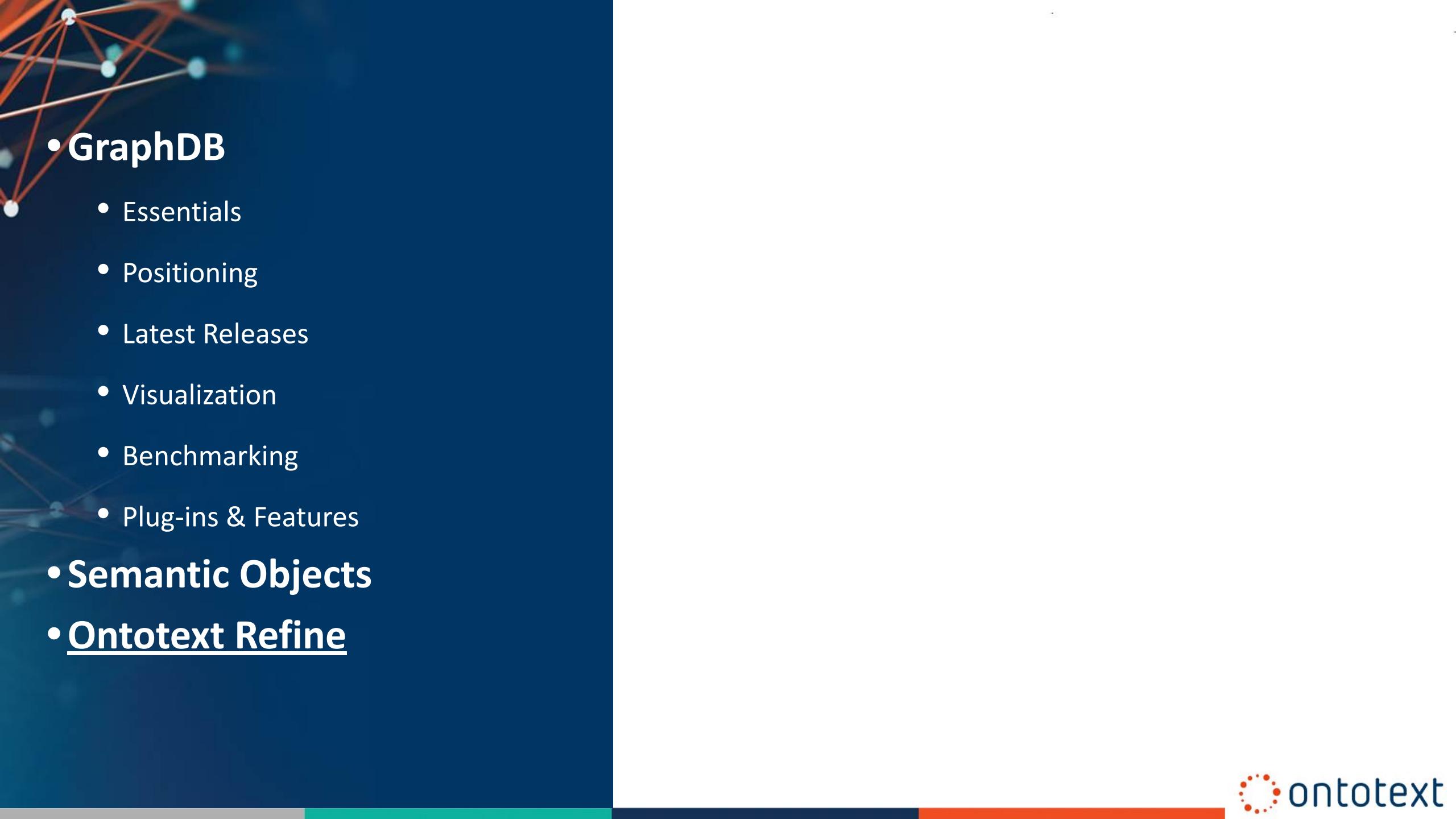
Pros

- Data access logic is shared for all clients
- Hides implementation details
- Easier to enforce security and throttling
- Good for nested/interconnected data
- Declarative schema and strong types
- Excellent tooling guiding data consumers
- Can coexists with REST APIs

Cons

- Harder to implement server logic
- Need tooling to generate GraphQL schema
- Need to implement GraphQL querying language (where with nested filters, limit, offset, orderBy...)

Semantic Objects

The background of the slide features a complex, abstract network graph composed of numerous small, glowing blue and white circular nodes connected by thin, translucent red and orange lines, creating a sense of data connectivity and complexity.

• GraphDB

- Essentials
- Positioning
- Latest Releases
- Visualization
- Benchmarking
- Plug-ins & Features

• Semantic Objects

• Ontotext Refine

Ontotext Refine: Data Transformation and Cleaning

- Easily import and clean your tabular data
- Transform your data
- Get real-time view as RDF with virtual SPARQL endpoint
- Import *or update* transformed RDF directly to a GraphDB repository using SPARQL Federation

name: TrumpWorld Data Public Person Org csv 

Open... Export ▾ Help

Facet / Filter Undo / Redo 0 Refresh

Reset All Remove All

Connection change
352 choices name count Cluster

President 492
Director 120
Member 75
Former director 63
Listed as asset/income on Public Financial Disclosure Report 63
Trustee 46
Board member 37
Chairman 28
Founder 20

1637 rows

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 50 next > last »

All Organization Person Connection Source(s)

| | | | | |
|----|---------------------------------|-----------------|--|---|
| 1. | 1116 SOUTH MAIN STREET, LLC | WILBUR ROSS | Managing member | https://www.commerce.senate.gov/public--public-questionnaire-redacted.pdf |
| 2. | 1186 BROADWAY, LLC | STEVEN MNUCHIN | Listed as asset/income on Public Financial Disclosure Report | https://extapps2.oge.gov/201/Presiden.ns |
| 3. | 2503(C) MINORITY TRUST FBO MPM | STEVEN MNUCHIN | Trustee | https://extapps2.oge.gov/201/Presiden.ns |
| 4. | 3126 CORPORATION | DONALD J. TRUMP | President | https://www.documentcloud.org/docume |
| 5. | 4 SHADOW TREE LANE LLC | DONALD J. TRUMP | President | https://www.documentcloud.org/docume |
| 6. | 4 SHADOW TREE LANE MEMBER CORP. | DONALD J. TRUMP | President | https://www.documentcloud.org/docume |
| 7. | 40 WALL | DONALD J. TRUMP | President | https://www.documentcloud.org/docume |

OPEN
Refine 

Ontotext Refine (continued)

- GraphDB Workbench offers interactive interface to:
 - Parse data in CSV, XLS, JSON, XML
 - Explore and correct the data
 - Clean/transform data with GREL expressions, Jython (Python) functions, SPIN functions
 - Reconcile input data values against existing knowledge graphs and datasets
 - Generate RDF data and store it in a local repository or remote endpoint
 - Automate the process and repeat in batch mode

name: TrumpWorld Data Public Person Org csv 

Open... Export ▾ Help

Facet / Filter Undo / Redo 0 Refresh

Reset All Remove All

Connection change
352 choices name count Cluster

All Organization Person Connection Source(s)

1637 rows

SPARQL endpoint Permalink SPARQL

| 1. | 1116 SOUTH MAIN STREET, LLC | WILBUR ROSS | Managing member | https://www.commerce.senate.gov/public--public-questionnaire-redacted.pdf | |
|----|---------------------------------|-----------------|--|---|--|
| 2. | 1186 BROADWAY, LLC | STEVEN MNUCHIN | Listed as asset/income on Public Financial Disclosure Report | https://extapps2.oge.gov/201/Presiden.ns | |
| 3. | 2503(C) MINORITY TRUST FBO MPM | STEVEN MNUCHIN | Trustee | https://extapps2.oge.gov/201/Presiden.ns | |
| 4. | 3126 CORPORATION | DONALD J. TRUMP | President | https://www.documentcloud.org/docume | |
| 5. | 4 SHADOW TREE LANE LLC | DONALD J. TRUMP | President | https://www.documentcloud.org/docume | |
| 6. | 4 SHADOW TREE LANE MEMBER CORP. | DONALD J. TRUMP | President | https://www.documentcloud.org/docume | |
| 7. | 40 WALL | DONALD J. | President | https://www.documentcloud.org/docume | |

OPEN
Refine 

- Based on OpenRefine

Data Reconciliation: Transform Strings to Things

- GraphDB integrates OpenRefine's UI for reconciliation
- Map string values to object identifiers via external reconciliation services (e.g. Wikidata or GeoNames)

OntoRefine  name: Fortune 500 2017 Fortune 500 csv 

Open... Export ▾ RDF Help

> 500 rows Permalink

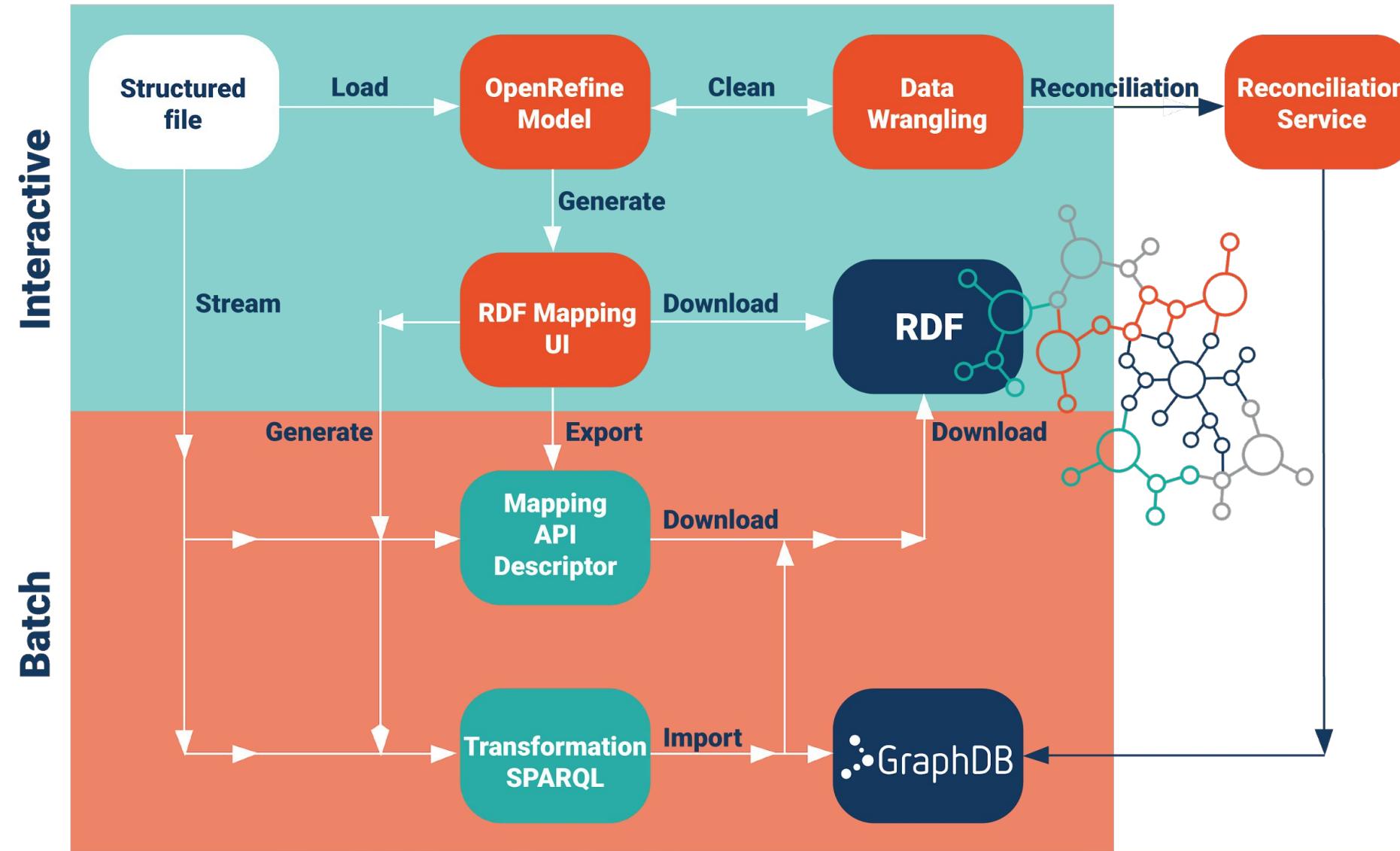
Show as: **rows** records Show: 5 10 25 **50** rows « first < previous 1 - 50 next > last »

| All | Rank | Title | Website | Employees | Sector | Industry | Hqlocation | Hqaddr |
|-----|------|-------------------------|---|-----------|------------|--|-----------------|------------------------|
| ☆ | 1. | 1 Walmart | http://www.walmart.com | 2300000 | Retailing | General Merchandisers | Bentonville, AR | 702 S.W. Eighth St. |
| | | ✓ ✓ Bharti Walmart [14] | | | | | | |
| | | ✓ ✓ Walmart Stores [14] | | | | | | |
| | | ✓ ✓ WalMart Canada [11] | | | | | | |
| | | ✓ ✓ Create new item | | | | | | |
| ☆ | 2. | 2 Berkshire Hathaway | http://www.berkshirehathaway.com | 367700 | Financials | Insurance: Property and Casualty (Stock) | Omaha, NE | 3555 Farnam St. |
| | | Choose new match | | | | | | |
| ☆ | 3. | 3 Apple Industries | http://www.apple.com | 116000 | Technology | Computers, Office Equipment | Cupertino, CA | 1 Infinite Loop |
| | | Choose new match | | | | | | |
| ☆ | 4. | 4 Exxon Mobil | http://www.exxonmobil.com | 72700 | Energy | Petroleum Refining | Irving, TX | 5959 Las Colinas Blvd. |
| | | Choose new match | | | | | | |

Visual RDF Mapping and Streaming Transformation

- **Visual interface in OntoRefine to define transformation of tabular data to RDF and mapping to existing schemas and ontologies**
 - Guidance in choosing the right predicates and classes
 - Defining the datatype to RDF mappings
 - Implementing arbitrary complex transformation using OpenRefine's GREL language
- **RDF mapping API with streaming support to transform tabular data**
 - Automation of ETL activities for building or updating knowledge graphs
 - Supports “data providers” like an OpenRefine project (see above) or posted CSV data stream
 - The streaming API guarantees no limitations on the size of the data

Visual RDF Mapping and Streaming Transformation



Data Virtualization: from Tables to Graphs and Back

- **Access data in relational databases as virtual graphs**
 - Ontop open-source data virtualization engine is integrated
 - Support for PostgreSQL, MySQL, Microsoft SQL Server, Oracle, IBM DB2, H2, Dremio
 - Virtual SPARQL endpoint configured by R2RML or OBDA descriptor
 - Virtualized data is accessible via virtual repositories to allow for strict access control
- **Access graph data in GraphDB via SQL, using GraphDB's JDBC driver**
 - Full SQL support via Apache Calcite
 - Graphs managed in GraphDB can be accessed from BI tools (e.g. Power BI and Tableau)
 - User-friendly interface to manage the SQL views (defined as SPARQL queries)

Give It a Try!

Play with GraphDB at [FactForge.net](#)

Explore a public read-only access to a GraphDB repository loaded with 2 billion facts, incl. DBpedia, Geonames and metadata for 1 million news. An easy way to experiment with GraphDB functionalities such as SPARQL editor, visualization, FTS connectors, geo-spatial and ranking.

Download and use the single-click installation of GraphDB Free

Delve into all GraphDB functionalities, including OntoRefine, which allows a WYSIWYG transformation of tabular data and reconciliation.

Get an Evaluation License for GraphDB Enterprise Editions



Data Pieces



Smooth Data Integration



THANK YOU!

www.ontotext.com

sales@ontotext.com

Worldwide: +359 2 974 61 60

N. America: +1 (929) 239 - 0659