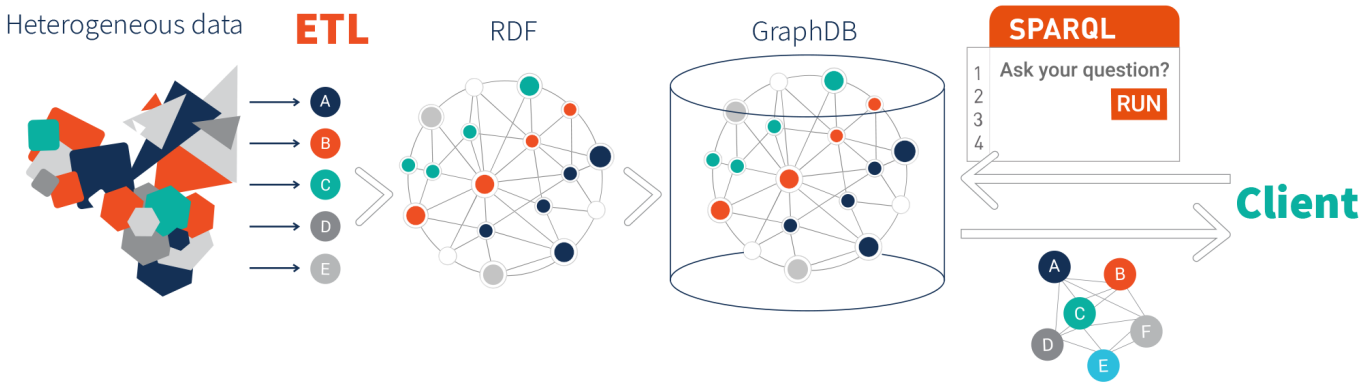
ontotext

The beauty and power of knowledge graphs is their abstraction away from the fiddly implementation details of our data. The data and information is organized in a way human-users understand it regardless of the physical location of the data, the format and other low-level technical details. This is because the RDF of the knowledge graph enables a schema-less, or a schema-agnostic, approach to facilitate the integration of multiple heterogeneous datasets.

Semantic technology defines how data and information is inter-related. These relationships give context and that context is what gives our data meaning that can be understood by humans AND machines. That's where the knowledge part of the graph comes from and it is a powerful way of providing a single view on disparate pieces of information.

## ETL is Still Your Best Bet

When possible it's better to pay the initial costs of the ETL process. In a previous blog post, we talked about how knowledge graphs generously repay that investment of time and effort taken in data preparation tasks. However, there are a number of reasons that it is impossible or impractical, such as the size of the dataset or the data exists in a critical legacy system where an ETL process would create more problems than it fixed. In these cases, it is better to take a data virtualization approach.



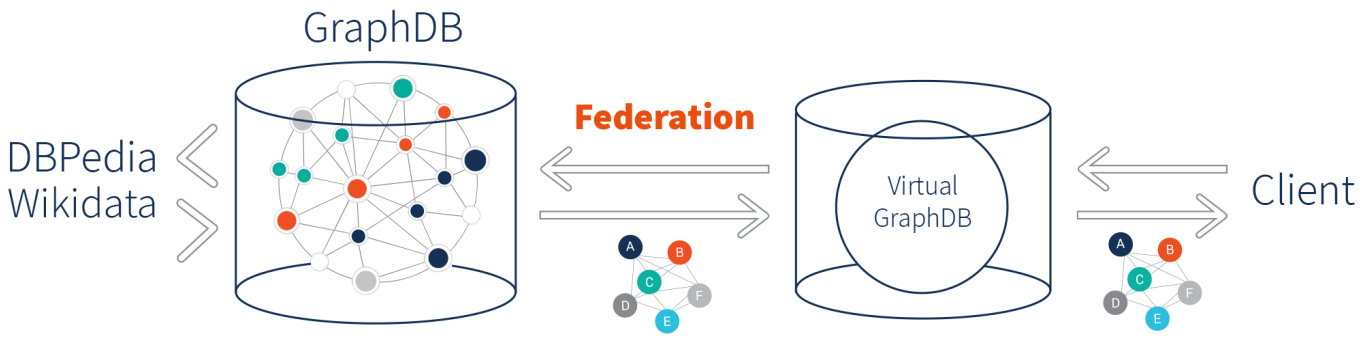

Ontotext GraphDB provides data virtualization functionality to realize the benefits of real-time access to external data, remove or reduce the need for middle-ware for query processing, reduce support and development time, increase governance and reduce data storage requirements.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

## Firstly, There's Federation.

RDF is the language of the semantic web. If you are working with Linked Data, it opens up a world of billions upon billions of factual statements about the world, which is probably why you chose to work with linked data in the first place. Nice work! And that means I don't have to tell you that DBPedia, a single data set among hundreds, has three billion triples alone. You are no longer limited by the data your organization holds. Queries about internal data can be seamlessly integrated with multiple external data sources.




For example, suppose you want to query well-known people and their birth places for a map application. It's possible to create a single query that gets the person's information from DBPedia, which would give you the birthplace and take those results to query another data source like Geonames to provide the geographic coordinates to be able to add them to a mapping application. Since both of these data sources are linked data, it's relatively straightforward to write a SPARQL query that retrieves the information.

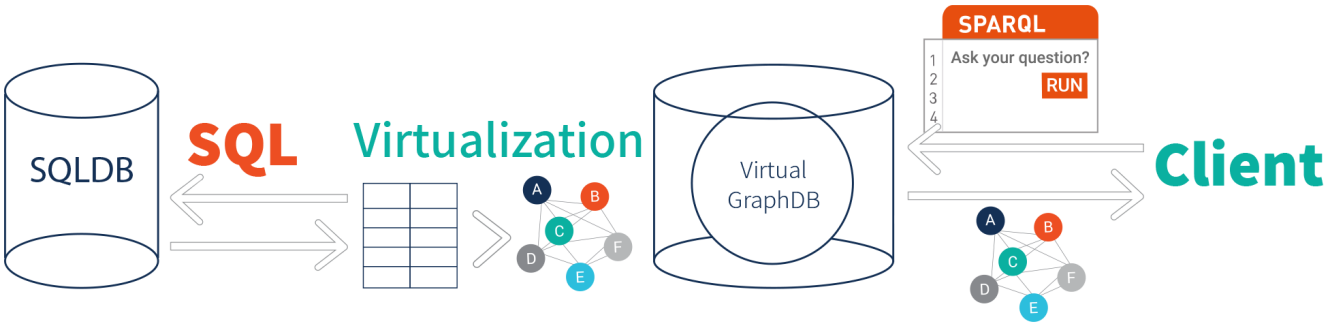
It doesn't even have to be another instance of GraphDB. It's part of the reason Ontotext insists on using open standards. With any equally W3C-compliant knowledge graph that supports a SPARQL endpoint, it is possible to retrieve the information you want and add it to your own knowledge graph to do with as you please. A single query could pull information from multiple external data sources to get the data you are after, which is why federation is an incredibly powerful tool to have.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

# The Business Intelligence Ecosystem Runs on SQL.

Ontotext is committed to lowering the costs of creating and consuming knowledge graphs. Not every app developer or DBA in an organization is going to have the time to work directly with the RDF data models. A previous version of GraphDB 9.4 added the JDBC driver to ensure those who need to think and work in SQL can access the power of the knowledge with SQL.

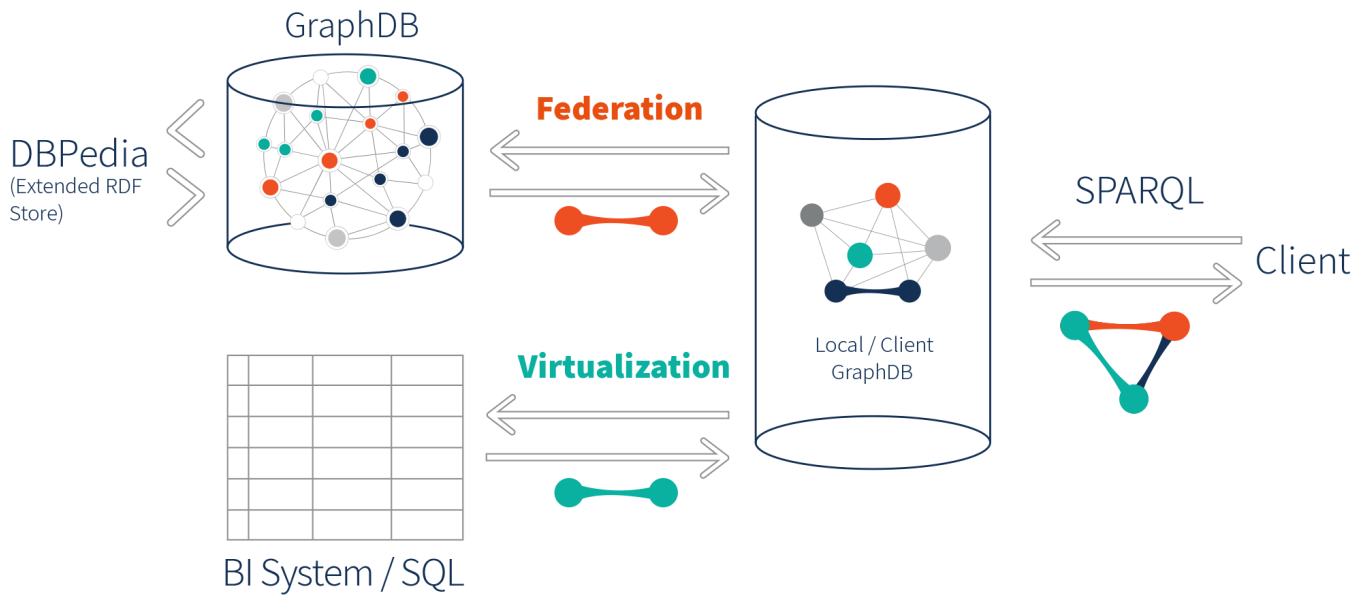



Knowing the importance and prominence of SQL for many applications, we have a webinar demonstrating how GraphDB does SQL-to-SPARQL transformation and query optimization and how Microsoft's Power BI and Tableau can be empowered by knowledge graphs. GraphDB provides a SQL interface to ensure those who prefer a SQL view of the world can have it.

# Virtualization vs ETL

The most recent GraphDB release has added virtualization functionality beyond simple federation. It is now possible to create a virtual graph by mapping the columns and rows of a table to entities in the graph. It becomes possible to retrieve information from external relational databases and have it play nice with our knowledge graph. We aren't bound by data that exists in our graph or even in RDF format. Of course it would be easier and certainly quicker to ETL a data source into a single graph and perform the query, but it is not always possible, because either the size of the dataset is too large, it gets updated too frequently or both.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com




For example, in the basement of your organization is a diesel-powered database that has a geological strata of decades old data and that is critical to the organization. You know and I that database is never going to be ETLed into the graph. Virtualization is your best bet, by creating a virtual graph and mapping that decade old format for client orders by saying, “When I query about ‘client order’, you need to go to this table and this column in that behemoth server belching black smoke that’s run by Quismodo and return the results”.

There will be an inevitable hit to query performance but there are a number of situations where slow is better than not at all. Such as the over-egged example above. It is important to understand the trade-offs and practicalities between ETL and virtualization. The important thing for Ontotext is to make sure GraphDB is capable of both and provide a combined approach to maximize flexibility. There is also a webinar on this topic, introducing the open-source Ontop.

## GraphDB Gives You Data Agility

Data virtualization and federation come with costs as well as benefits. There is no way we are ever going to master where the data you need exists and what format. The days of centralized control are over. It’s about finding the technology that gives your agility and

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

GraphDB’s added virtualization capabilities enables you to create queries that include external open sources and merge it seamlessly with your own knowledge graph. Virtualization of relational databases creates incredible opportunities for applications to provide users with a single coherent view on the complex and diverse reality of your data ecosystem.

Ontotext’s GraphDB
**Give it a try today!**
Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com