

The Landscape of Cloud-based Semantic and Property Graph Databases

AUTHOR

Harsh Thakker, PhD.

in

Issued Date

November 2022

 +49 241 94313 0

 office@osthus.com

 www.osthus.com



Table of Content

Recap of Whitepaper 1 + Objective	03
Vendors in the Semantic (RDF) and Property Graph Landscape	04
Vendors Grouped by Type of Graph Data Model	06
Vendors in the Cloud-based Semantics (RDF) and Property Graphs Landscape	08
Who is using Semantic and Property Graphs and what for?	10
The use of Graph Databases in Life Sciences	12
Conclusion	13

Recap of Whitepaper #1 + Objective

In the previous white paper - [“Let’s talk about Graph Data models and why they are relevant for your Data and IT Strategy”](#) - we briefly touched upon the concept of graph data models such as RDF and Property Graphs (PG), web semantics, ontology, and a Knowledge graph using a running Elon Musk data example. We also reflected upon key differences between the RDF and PG data models.

Furthermore, we described several reasons or use cases where adopting a graph data model is advantageous compared to a traditional relational database approach.

Now we continue our journey and summarize the current state of the art in cloud-based semantics and PG database platforms offered by the commercial and open-source providers. We first present a categorization of key players (providers) by the data model and platform they offer. Then, we list the real-world use cases and selected case studies, along with the consumers/corporations that have adopted these platforms.

Please note, that the information documented in the scope of this article has been accumulated from the publicly available use case reports, case studies, technical papers, and market surveys published by the respective vendors, and third-party sources. Given the rapid pace of research and innovation in the graph and semantics landscape (both on premise and cloud-based platforms), the current survey document may not reflect the most recent advancements in the domain. It serves as a focal point for gathering a broad understanding of the use of graph and semantics data and IT platforms, which have been or are currently being adopted by several big and small corporations.

So let's dive in.

Vendors in the Semantics (RDF) and Property graph Landscape

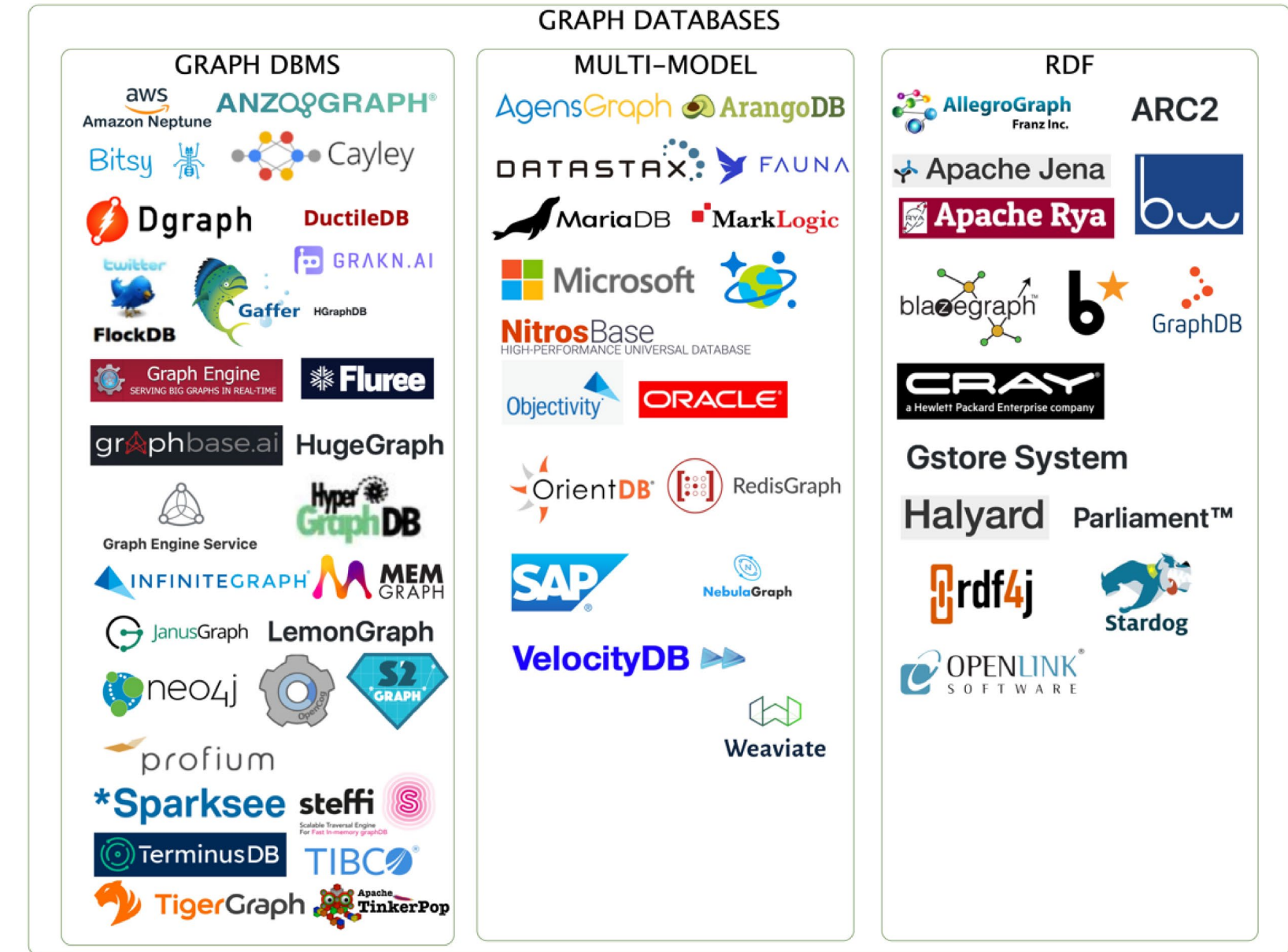


With the maturation of the graph ecosystem in the last two decades, there has been an explosion of vendors (and users) of graph databases and platforms - be it standalone, on premise, cloud-based Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS) providers.

The extent of graph technology's impact is evident from the fact that all the big players such as Amazon, IBM, Microsoft, Oracle, etc. have thrown their hats in the ring for global domination of the graph market. This is driven by the spectrum of use cases and applications that can be powered using graphs. We listed a few in [part 1 of this white paper series](#). To summarize, graphs are currently being used in not only conventional use cases such as - data migration, data cataloging, data lineage, data governance, data lake acceleration, data virtualization, supply chain, etc.; but also in modern complex analytical use cases such as - fraud detection, knowledge graphs, questions answering, real-time recommendation systems/engines, Identity & Access Management, semantic (graph-based) search, bioinformatics & Life Sciences, network & IT operations, GDPR - the list goes on!

Due to such a high applicability and dynamic growth of the graph technology stack, it is extremely difficult to stay up to date on the recent developments and offerings in this landscape. As a result, for the past couple of years, [Graphaware](#) and [Matt Truck](#) (amongst others) have been publishing a snapshot of the annual state of the Graph, Big Data, and AI-Machine Learning landscape as an effort to consolidate all the different vendors, third-party providers, use cases, and applications currently being used. Below we borrow a slice from Graphaware's most recent well-consolidated [landscape snapshot](#) in order to highlight the various vendors in the broader graph technology ecosystem.

And yes: we'll elaborate on this in the next section.



Vendors Grouped by Type of Graph Data Model

The graph technology landscape is full of vendors offering solutions that support a wide variety of features and graph data models. One such categorization can be referenced from [Steve Sarsfield's blog post on this topic](#) (see image on the next page). Depending on the type of data model(s) your application will require, one can select a single model (such as PG or RDF or only) solution or a multi-model (RDF, PG, JSON, Key-Value, etc.) solution.

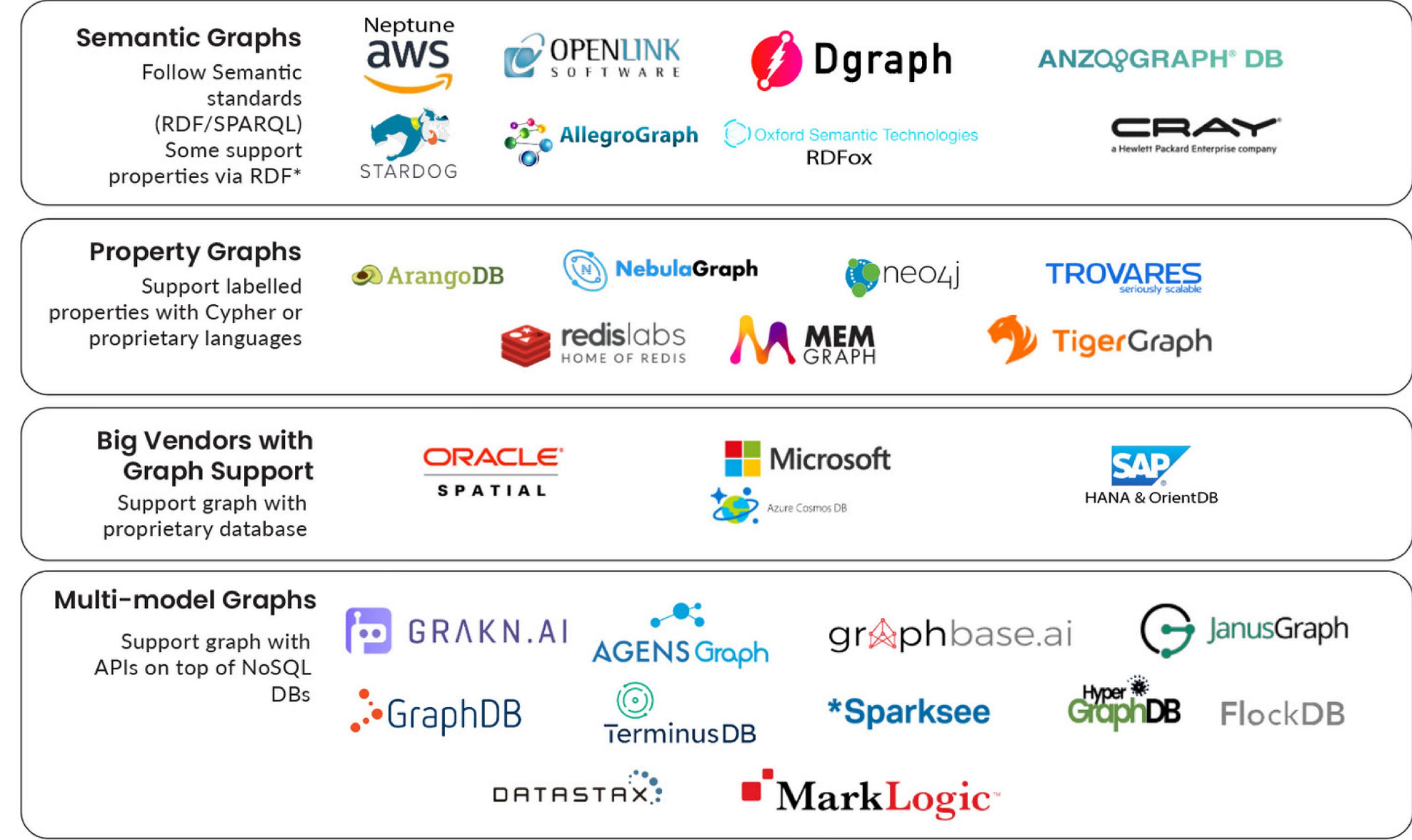
It is advisable to consider the choice between proprietary graph databases and query languages they support beforehand in order to avoid possible vendor lock-in scenarios (e.g. choosing between Neo4j which supports Cypher and Datastax which supports TinkerPop Gremlin).

However, since the past couple of years, the risk of a significant vendor lock-in scenario has become less severe due to the implementation and adoption of multi-model and multi-query language solutions by all key players. A multi-model data and query solution allows the use to store and query data using multiple data formats and query languages.

Thereby avoiding a vendor lock-in scenario where integration of other data ETL and processing tools becomes feasible without relying on a particular vendor's custom software and data interfaces/plugins.

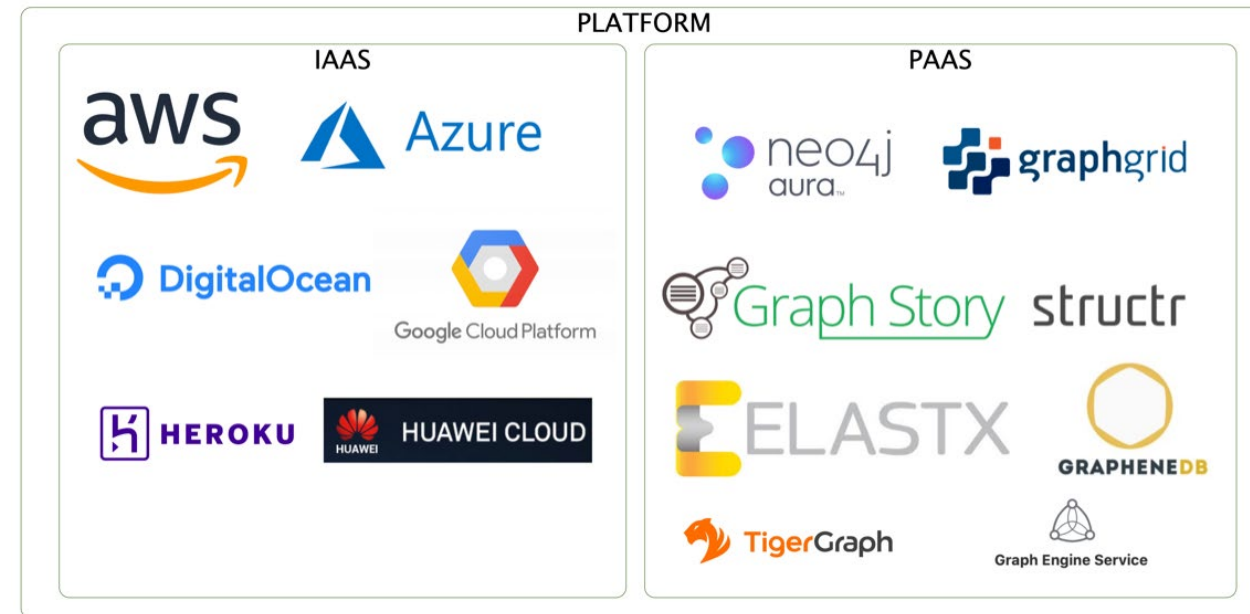
For example AWS Neptune, Neo4j, Microsoft Azure CosmosDB, Oracle Graph Server + Graph Studio on Autonomous Database, and Google via the Neo4j AuraDB PaaS on Google Cloud, to name a few.

Furthermore, given the recent ongoing standardization effort in the PG space (new property graph query language - [GQL standard](#)), the interoperability and lock-in issues should vanish as the adoption rate increases among the vendors.



Vendors in the Cloud-based Semantics (RDF) and Property Graphs Landscape

We borrow the following neat categorization (see figure below) of graph technology vendors in the cloud space from Graphaware's Janos Szendy-Varga. There are various alternatives for storing your graph data in the cloud. If you go with infrastructure-as-a-service (IaaS), you may pick one of the prominent providers and host a Virtual Private Server (VPS) with your graph database on it, which could be any cloud provider of your choice. However, a few key players enable operating a graph database as a cluster or add-on. Furthermore, some vendors provide the graph database as an off-the-shelf solution in the cloud, i.e., utilize the graph database as Platform-as-a-Service (PaaS).



On the next page, we have compiled a list of prominent cloud providers, their solutions along with the data models they support to the best of our knowledge. Please note that the following list reflects the status in time when the article was compiled. We propose the interested readers follow the links and check out the solutions that may seem of interest.

VENDOR	DATABASE / SOLUTION(S)	RDF (YES / NO)	PG (YES / NO)	COMMENTS
Amazon	AWS Neptune	Yes	Yes	It is an either/or choice between RDF and PG; as both models are not interoperable.
	Neo4j AuraDB for AWS	No	Yes	-
Google	Neo4j AuraDB for GCP	No	Yes	-
IBM	Db2 Graph	Yes	Yes	Supports querying via Apache TinkerPop Gremlin graph traversal language. RDF support from IBM® Db2® Version 11.1 Enterprise Server Edition (Db2 Enterprise Server Edition) databases.
Microsoft	Azure CosmosDB	No	Yes	Also supports document, wide column, and key-value store data models
Neo4j	Neo4j Graph Database	*Yes	Yes	*RDF support via neosemantics
Oracle	Graph Server and Graph Studio on Autonomous Database	Yes	Yes	Offers PG data access via its proprietary query language PGQL



A complete list, relative ranking, and system specification of vendor solutions in the entirety of the database technology landscape is available on the [db-engines website](#) - [Graph database systems](#) and [RDF database systems](#).

Who is using Semantic and Property Graphs and what for?

In this section, we present a compiled list of notable consumers of prominent graph technology vendors and their reported use cases. We have documented this information from the publicly available use case reports, case studies, technical papers, and market surveys published by the respective vendors, customers, and third-party sources. We link the corresponding source(s) for each reported use case in case you are interested in further reading.

DATA MODEL	DATABASE / SOLUTION(S)	CONSUMERS	APPLICATIONS	USE CASES / CASE STUDIES
RDF, PG	AWS Neptune	Audible, Siemens, Zerobase, ADP, Zeta, Cox Automotive, Games24x7, Careem, etc. (source)	Enterprise Knowledge Graph, Identify Graph, Fraud Detection, Machine Learning, Data Migration. Data Analytics, Data Lake, Edge, Sustainability, Business Applications, Data Virtualization, Data Lineage, Recommendation System, Digital Twins, etc.	Audible , Siemens , Cox Automotive , ADP , Games24x7 , Careem , Zeta , Recommendation Systems , Geospatial Systems , RhizomerEye , Data Lineage , Digital Twins , and many more
PG	Microsoft Azure CosmosDB	MARS, Mercedes-Benz, Chipotle, ASOS, Symantec, etc. (source) It also has a list of third-party system integrator and tooling partners (see list)	Enterprise Knowledge Graph, Machine Learning, Data Science, Data Migration. Data Analytics, Data Lake, Data Virtualization, Data Lineage, Recommendation System, Digital Twins, Distributed Cloud Data Platform, etc.	Mercedes-Benz , Chipotle , ASOS , Symantec , and 21 more blog posts on Azure-driven architectures can be found here

DATA MODEL	DATABASE / SOLUTION(S)	CONSUMERS	APPLICATIONS	USE CASES / CASE STUDIES
RDF, PG	Neo4j Graph Database, AuraDB and AuraDS for GCP and AWS	NASA, Zenapse, Orita, Levi Strauss & Co., Adeo, BOSA, APIAX, Boston Scientific, Department for Education (UK), Dun & Bradstreet, Lucinity, Maritz, PwC, The Orchard, Aker BP ASA, Tourism Media, Worldline, Amity, Qualicorp, and several more (sources: Neo4j , Graph Database , AuraDB , AuraDS)	Enterprise Knowledge Graph, Business Intelligence, Master Data Management, Fraud Detection, Identity & Access Management, Logistics, Social Network, Real-time Recommendation Engine, Graph-based Search, Network and IT Operations, Machine Learning, Data Transformation, Bioinformatics, Life Sciences, and several more	A detailed list of well-maintained documentation of around 65 customer case studies for various Neo4j solutions can be referred from here
RDF	Openlink Virtuoso	Fluor, Bae Systems, Verison, American Express, Bayer, Deutsche Bank, Wells Fargo, NASA, Novartis, Sanofi, Abbot, UBS, Janssen, Nestle and many more (source)	Enterprise Knowledge Graph, Linked Data, Data Integration, Data Virtualization, Master Data Management, Identity & Access Management, Drug Discovery, Question Answering, Semantic Search, Supply Chain, Network Analysis, Middleware, Data Visualization, Big Data, NoSQL, etc.	Openlink Virtuoso has several documented use cases broadly from the following links: Knowledge Graph , GDPR , Openlink Blog
RDF, PG	Stardog	NIH, NASA, Springer Nature, eBay, Bosch, Boehringer Ingelheim, FINRA, IDB, National Cancer Institute, US Department of Defense, Schaeffler, Scheider Electric, etc. (source)	Enterprise Knowledge Graph, Data Catalogue, Data Integration, Data Virtualization, Master Data Management, Data Lake Acceleration, Question Answering, Drug Discovery, Semantic Search, Supply Chain, Data Fabric, Operational Risk, Analytics Modernization, etc.	NASA , NIH , Springer-Nature , Boehringer Ingelheim , Drug Discovery , Schneider Electric , QIMR Berghofer , Casalini Libri

The use of graph databases in Life Science

As OSTHUS provides consulting and system integration to the leading players of the Life Science industry, we'd like to give some specific context for this particular audience: Graph databases allow addressing a diverse set of problems and use cases that are typical for your domain. The graph data model is key for capturing dynamic relations between objects that are required in a typical Life Science use case.



For instance, use cases such as

- [Life Science/pharma global data management](#), (in collab with EDMC),
- [supporting semantic interoperability using ontologies](#),
- [reference and master data management](#),
- [medical supply chain analysis](#),
- [building an enterprise-wide FAIR data foundation](#),
- [R&D lab data lineage and management](#),
- [drug discovery using Knowledge Graphs](#),
- [drug development](#),
- [Knowledge management via global data integration and access](#),
- [flexible research data modeling and access](#),
- [data analytics and mining over complex biological data](#) (clinical recommendation engine)

are only a few examples of organizations solving complex business problems harnessing the power of semantic and Property Graphs, the majority of which are examples of previous OSTHUS projects.

We urge interested readers to check out the references from vendors such Neo4j, Stardog, AWS Neptune, and others (cf. table above) for an in-depth analysis of the relevant use cases and case studies. We are observing an increasing adoption of graph databases in the Life Science and Pharma industry over the past years and anticipate the trend continuing over the next years owing to the demand for handling highly complex, connected, and diverse (biological/chemical) data.

In Conclusion

We hope that this article serves as a solid reference point for its readers who are interested in a quick and somewhat comprehensive overview of the graph technology landscape. Overall, this two-part white paper series on graphs had the following purpose in mind for its diverse audience:

Part I: Using Graph Data models for your Data & IT Strategy

To introduce the concept of graph data models such as RDF and Property Graphs (PG), web semantics, ontology, and a Knowledge graph (using a relatable real-world example, i.e. Elon Musk). To touch upon key differences between the RDF and PG data models and describe reasons and suitable use cases where adopting a graph data model is advantageous compared to a traditional relational database approach.

Part II (aka this very research paper ;-)


To present a comprehensive summary of the current state of the art in cloud-based semantics and PG database platforms offered by the commercial and open-source vendors. To serve as a reference point for documenting several real-world use cases, case studies, and consumers/corporations for which the vendors are leveraging the power of graphs via their platforms.

If you are interested in learning more about the applications of cloud-based semantics and Property Graphs in the Life Sciences or your other data-driven business use cases, please do reach out to us via LinkedIn, Twitter or Email. Furthermore, stay tuned to check out our upcoming blog-posts, research & white papers on this topic in the future.



Get In Touch

 +49 241 94314 0
 www.osthus.com
 office@osthus.com

 Eisenbahnweg 9-11,
52068 Aachen

