



www.cambridgesemantics.com

1 Beacon Street, 15th Floor

Boston, MA 02108

Tel: +1 617 245 0517

Email: info@cambridgesemantics.com

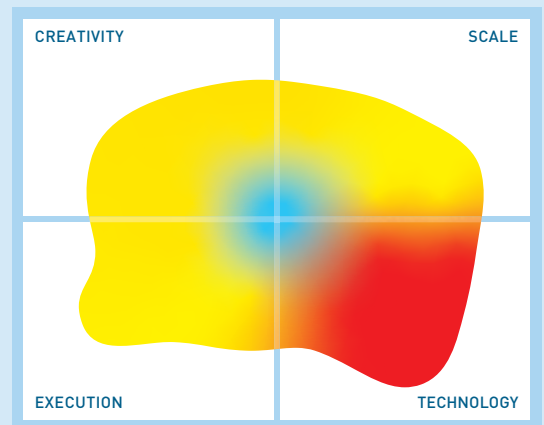
Cambridge Semantics AnzoGraph

The company

Cambridge Semantics is a private company based in Cambridge, Massachusetts. It was founded in 2007 to support a semantically oriented approach to big data and the enterprise information fabric. In January 2016 it acquired SPARQL

City, the company behind the development of the SPARQLverse graph database. Cambridge Semantics has various consulting, systems integrations and other technology companies as partners and its focus is in financial services, pharmaceuticals and life sciences, healthcare, Insurance, government, material sciences and retail. Customers include Merck, Johnson & Johnson, Bristol-Myers Squibb, Lilly, Credit Suisse, Roche, Novartis and The Mayo Clinic amongst others.

“Faster and more accurate regulatory compliance reporting Flexibility, consistency, and accuracy throughout the entire clinical trial lifecycle Slashed development lifecycles by 4x.”
Global Pharmaceutical company



The image in this Mutability Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

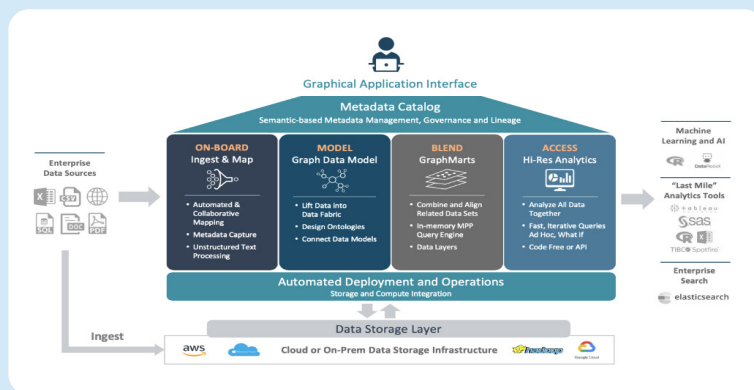


Figure 1 – The architecture of Anzo

The company offers two products: Anzo, which is a data fabric whose architecture is illustrated in **Figure 1**, and AnzoGraph. The latter is included in the former and, as a stand-alone capability is available in both a Free Edition (which has usage limitations) and an Enterprise Edition. It is available both in-cloud and on premises.

What is it?

AnzoGraph is a massively parallel graph database that runs on HDFS, NFS and other big data platforms.

It is ACID compliant, but the emphasis is more on analytics, including high performance interactive queries that extend from OLAP style functionality through to the provision of out-of-the-box graph algorithms (see **Figure 2**) and support for both Zeppelin and Jupyter notebooks. In the case of the former, the product ships with extended functionality including tutorials, sample code and so forth. In addition to analytics the company is also seeing significant growth in supporting data harmonisation (combining structured and unstructured data) through the construction of knowledge graphs.

While AnzoGraph is technically an RDF graph database it supports the proposed RDF* and SPARQL* W3C standards. This means that you can attach labels to vertices and edges in a similar manner to that of labelled property graphs. This capability has allowed Cambridge Semantics to add support (currently in preview) for OpenCypher. As RDF approaches tend to be favoured by information architects and property graphs are generally preferred by developers and business analysts this gives you the best of both worlds.

What does it do?

AnzoGraph’s key claim to fame is its analytic performance and scalability while running on low cost cloud or on-premises commodity servers.

| | |
|-------------|-------|
| Analytics | ★★★★★ |
| Ease of use | ★★★★ |
| Features | ★★★★ |
| Integration | ★★★★ |

| | |
|-------------|-------|
| Language | ★★★★ |
| Operations | ★★★★ |
| Performance | ★★★★★ |
| Scalability | ★★★★★ |

A leading biopharma firm found its AnzoGraph-powered analytics query performance was up to 250x faster than the graph database previously used.

To support its performance, the database uses various constructs. To begin with, all queries are compiled (into C++). Secondly, it uses forward chaining in its inference engine which, while consuming more memory than backward chaining, provides better performance. Thirdly, it supports both materialised and dynamic views. More fundamentally, it uses a shared nothing, massively parallel architecture.

Apart from its support for OpenCypher query processing is via an extended version of SPARQL. AnzoGraph supports many additional BI analytics functions such as conditional expressions, windowed aggregates, named queries, views and multi-graphs. Being a declarative language, there is an appropriate database optimiser for SPARQL, and it is noteworthy that the personnel at Cambridge Semantics have a significant history in this subject.

Two further significant capabilities are also in beta as of the time of writing. The first of these is support for data virtualisation, which will allow you to query external data sources without any requirement to move relevant data into AnzoGraph. And the second is support for Kafka. While there are existing facilities to directly import OLTP data into AnzoGraph (and to export results data from AnzoGraph into your OLTP system) using traditional ACID functions such as inserts, this is not fast enough when transaction volumes are high, hence the move to support Kafka, which will allow transactions to be posted in micro-batches.

Why should you care?

We do not believe that users want to agonise over whether they should choose an RDF graph or a property graph. Both have advantages and, in the longer term, we expect these markets to merge. However, that hasn't happened yet. Nevertheless, Cambridge Semantics is in the vanguard of vendors saying that you don't have to make that choice, that you can have both.

Another significant capability, once it is out of beta, is the geospatial support that AnzoGraph will be offering. This is significantly in advance of the (relatively few) other graph products that offer geospatial support and it will enable Cambridge Semantics to address IoT and other use cases that have not historically been well served by graph database suppliers.

And finally, from a purely technical perspective, a major reason for liking AnzoGraph is that there are relatively few vendors in the graph database market that have the performance and scalability to focus on analytics. Moreover, of those few competitors that are in this space they tend to be either not be semantically oriented and/or leverage proprietary languages and/or rely on (very expensive) hardware acceleration to get their performance.

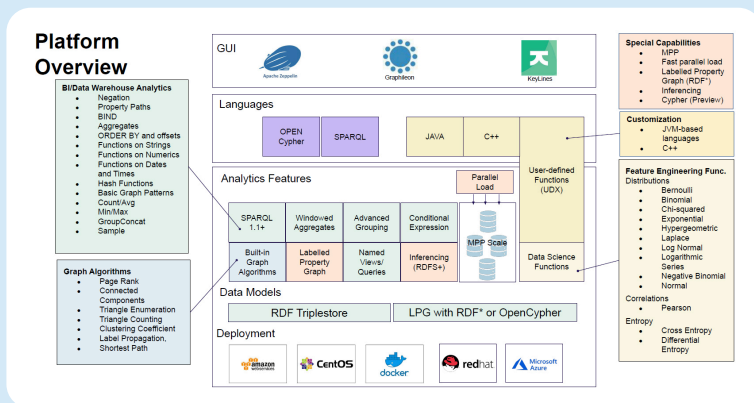


Figure 2 – Platform overview

As far as direct analytics support is concerned, various BI and analytic functions are provided out of the box, including graph algorithms such as PageRank and Shortest Path. As one would expect, inferencing is built-in, and there is support for integration to third party tools and vendors including R, SAS, Qlik, Tableau and Spotfire as well as specialist graph products such as Keylines. Notably, the company has announced a partnership with ESRI with all of that company's geospatial library of (approximately) 150 functions available inside the AnzoGraph database. This is a fully 3D (not just 2D) implementation and supports features such as shape files, polygon handling and so forth.

The Bottom Line

In our last report into the graph database market we reported that "AnzoGraph looks to be well positioned" to capitalise on the growth in demand for analytic graph processing. Our view today is that the product is fulfilling that promise and it is unquestionably a leader within its market segment.

[FOR FURTHER INFORMATION AND RESEARCH CLICK HERE](#)