Bloor
InBrief

Philip Howard – Research Director, Information Management
Daniel Howard – Senior Analyst, Information Management and DevOps

MUTABLE AWARD
2020

PLATINUM

neo4j

www.Neo4j.com

111 E 5th Avenue, San Mateo, CA 94401, USA
Tel: +1 855 636 4532
Email: info@neo4j.com

Neo4j

The company

Neo4j Inc (previously Neo Technology) was first conceived in 2000 and was formally founded in 2007 in Sweden, although it is now headquartered in the United States. Outside of these two countries the company also has offices in the UK and Germany with additional sales and service personnel across the EU, Middle East and Asia Pacific regions. The company's eponymous product is available in both Community and Enterprise Editions and is available both on-premises and via Google, Amazon and Microsoft Azure cloud platforms. Managed service options (Neo4j Aura) are also available, both in public and private clouds. The company also offers Neo4j Bloom as a visualisation engine and the Neo4j Graph Data Science Library.

> " Our Neo4j solution is literally thousands of times faster than the prior MySQL solution, with queries that require 10-100 times less code. At the same time, Neo4j allowed us to add functionality that was previously not possible. "
> *eBay Shutl*

CREATIVITY   SCALE

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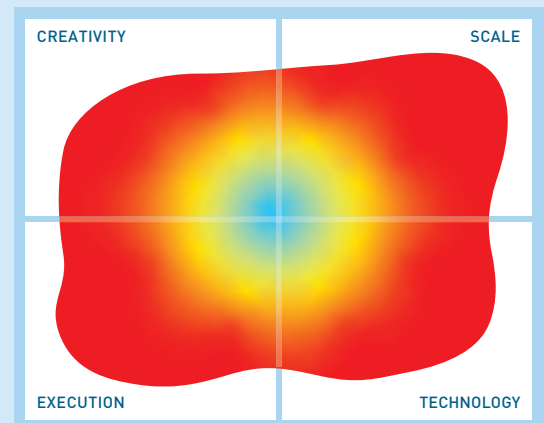
The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.
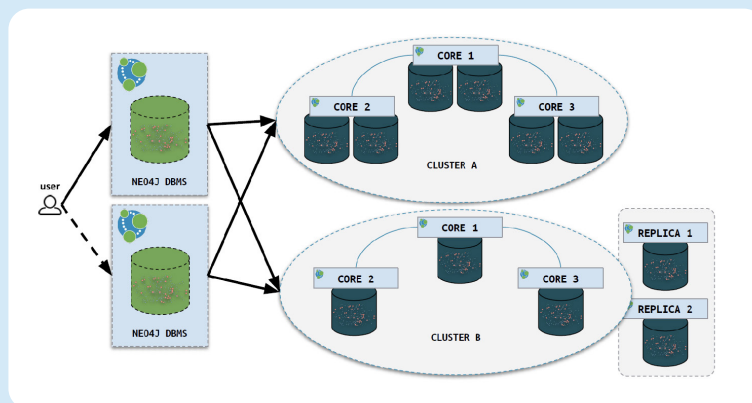

Figure 1 – Scaling out with Neo4j

The company has a significant partner base. Notable amongst these are Confluent (Kafka), Linkurious, Thales, Tom Sawyer, IBM, EY, GraphAware and NEORIS amongst many others.

What is it?

Neo4j is a property graph database with a native engine that is targeted at operational, hybrid operational/analytic (HTAP) and pure analytic use cases. It is ACID compliant and supports immediate consistency. Additional technologies and tooling are available to support the Neo4j environment. Since version 4.0 was released (4.1 is the current version) the product has supported scale-out as well as scale-up, as shown in *Figure 1*, which depicts the (geographically) distributed environment that Neo4j now supports. This is based on the introduction of support for sharding, which extends the horizontal multi-cluster scaling that was introduced in version 3.4. The replicas illustrated refer to read replicas, which have been available within the product for some time. Also included in the most recent release is support for much more granular security than was previously the case.

Most users (see below) employ Cypher or openCypher (the open source version), which is the declarative language developed by Neo4j. It is notable that SAP, Redis, Memgraph and others have adopted OpenCypher and it is also being used within several open source projects including Cypher for Apache Spark, and Cypher for Gremlin, as well as in research projects like InGraph for streaming queries. As with any declarative language this is best implemented along with a database optimiser and the company has devoted considerable resources to this, extending beyond an original rules-based optimiser so that it is now primarily cost-based, supporting optimisation for writes as well as reads.

Analytics	★★★★	Language	★★★★
Ease of use	★★★★	Operations	★★★★⯪
Features	★★★★	Performance	★★★★
Integration	★★★★	Scalability	★★★★⯪

> I'd like to comment on Neo4j's scalability and capability of looking at millions and millions of nodes. We have a "big data" problem — not only in structured data, but in unstructured data — and we are continually gathering more data. At NASA, my focus right now is on the unstructured data. And I need a product or an application that can go across and develop millions if not billions of nodes, connect that information and at fast speeds. Neo4j is that tool. **"**
>
> **NASA**

What does it do?

Unusually for a property graph, SPARQL is supported. So too is Gremlin (part of the Apache Tinkerpop project). Perhaps more significantly, the company has introduced a *"BI Connector"* which translates SQL queries into Cypher, with an initial focus on supporting Tableau so that you can use this instead of, or alongside, Neo4j Bloom, where the latter provides a visualisation and communication interface for non-technical users so that they can explore, edit and search graphs, and create storyboards. This is illustrated in *Figure 2*.
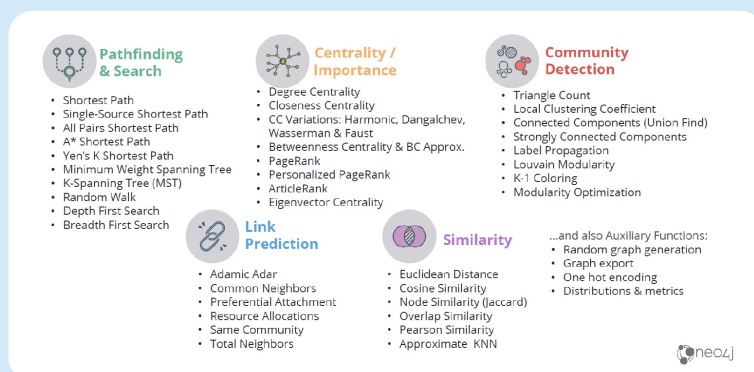

Figure 2 – Neo4j Bloom

Figure 3 – Graph algorithms provided by Neo4j

Also, importantly, the company is a driving force behind GQL (graph query language), which is intended to be a common standard for graph databases. under the ageas of the ANSI standards committee. This initiative is supported by a range of technology vendors including Talend, SAP, Tableau and others.

A significant recent release is the Neo4j Graph Data Science Library, which works in conjunction with Neo4j Bloom to support advanced analytics and machine learning. More than 50 graph algorithms (see *Figure 3*) are supported and have been optimised for robust scale and parallelised for performance. This last point is important because there are some vendors offering parallelised graph algorithms (consider MADlib for example) running against relational databases. And the problem there is that there are only a limited number of such algorithms can be parallelised in a relational environment, whereas Neo4j is able to offer a much comprehensive set of capabilities. Finally, in the context of analytic and query support, it is also worth noting that Neo4j supports 3D geospatial capabilities.

Why should you care?

Neo4j is the clear market leader in the graph space. It has the most users, it uses and drives a widely adopted query language. In many respects, it has consistently been a lot more innovative than its competitors. This is in part because of the maturity of the product and partly because its success has meant that it has the resources to introduce such developments more quickly. Its competitors have historically argued that the product did not scale well but the multi-clustering and sharding that are now available should knock that argument on its head. Some vendors that specialise in analytics will claim that they can outperform Neo4j and this may be valid, but Neo4j does not have this limited focus: it is, in effect, the Oracle or SQL Server of the graph database world. It is not the equivalent of Teradata. In other words, it is a general-purpose graph database, and it is no coincidence that it is the leading product in this space.

The Bottom Line

Whenever we talk to a vendor in the graph database space it is Neo4j they compare themselves to. Even if they do something different and address a different market, Neo4j is the benchmark – the company claims more than 400 enterprise customers globally. In pretty much every instance Neo4j should be on your shortlist.

FOR FURTHER INFORMATION AND RESEARCH CLICK HERE