

ANZOGRAPH® DB

CASE STUDY: Building an ESG solution with NLP and a scalable graph analytics database

Why Parabole chose AnzoGraph® DB to power its cognitive analytics platform for financial services investments insights

Summary

Parabole offers a powerful cognitive analytics platform to bring together data from news sources, documents and other unstructured data from data streams and third-party databases in order to provide rich insights and analytics to guide socially responsible investing. Underneath the covers, Parabole chose AnzoGraph DB, a fast and highly scalable graph analytics database, as the analytical engine for this solution.

Company Overview

Founded in 2014, Parabole is an innovative leader of cognitive analytics software for the financial services industry. Parabole's vision is to "make unstructured data available for mainstream data analytics in an organization." By combining natural language processing and graph analytics, the Parabole Cognitive Analytics Platform is able to analyze unstructured text and combine the data therein with domain knowledge as materialized in knowledge graphs to build high-value AI/ML analytics applications.

The Challenge

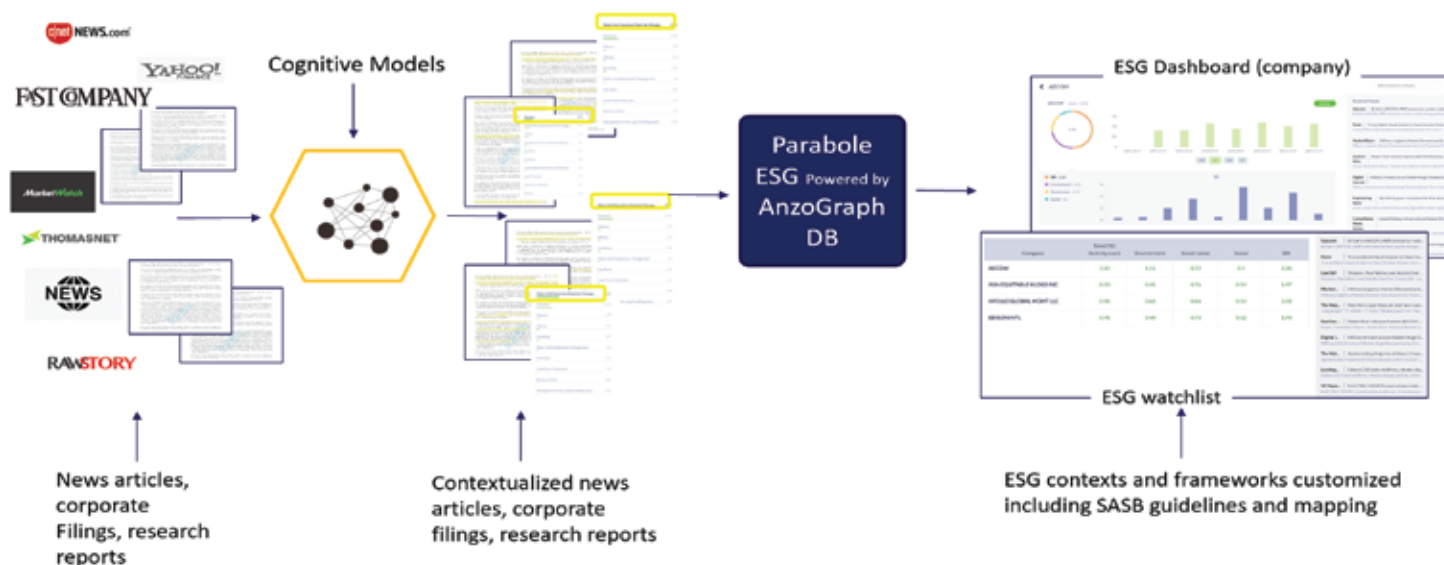
Known as ESG, short for Environmental, Social and Governance, investors today are more and more interested in the set of standards for a company's operations that ensure best practices and sustainability. While ESG often evaluates traditionally well-known issues such as greenhouse gas emissions, women's issues, and ethics, ESG also speaks to how a company is run overall. To keep tabs on ESG activities, investors have to consider a much wider set of data from diverse sources than typical investment analysis. Tools to aggregate, synthesize and provide relevant data from a growing number of diverse news, reports and data sources are lacking in the industry.

"Graph databases that can handle RDF and OWL are perfect for performing advanced analytics on Elastic Search and NLP output. We chose Anzograph DB for our ESG solution because we needed both a standards-based approach and unbounded scalability for analytics."

- Rajib Saha, CEO, Parabole

Parabole's alphaESG is an artificial intelligence and machine-learning solution, built to provide faster data harmonization and insights from diverse data for fund managers and investment analysts looking to analyze and invest in companies that are sustainable and responsible. Parabole recognized that their specialized knowledge in NLP and graph analytics could be used to create a cognitive analytics platform to support ESG analysts with data collection, curation, scoring and investment. By applying artificial intelligence and machine learning to corporate research, alphaESG can find relevant information to create a customizable corpus of data that can be used for company investment analysis using ESG criteria. The solution can tap into corporate filings, news feeds, PDF files, spreadsheets and other structured and unstructured sources to extract context and measure the signals of how well a company is performing ESG activities. The solution was named Parabole alphaESG.

Increasing use of non-traditional factors from alternative data makes automation and cognition critical



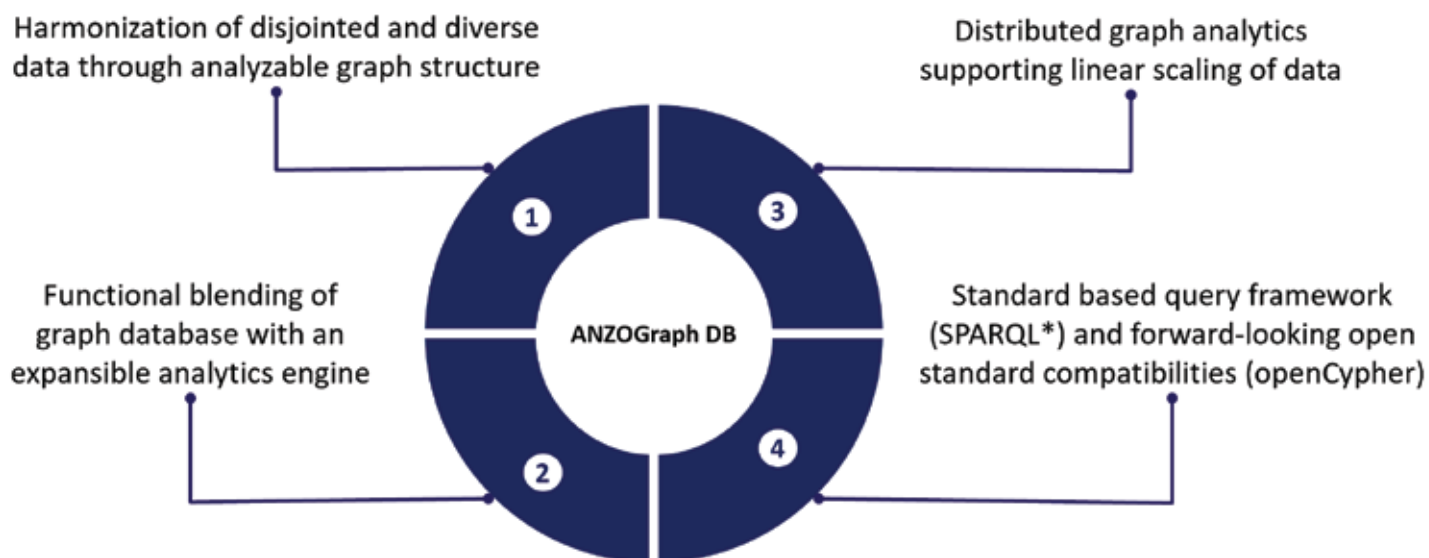
The Selection Criteria

The first challenge is that the data from many news sources need a flexible infrastructure that can adapt to address changing data needs. Most graph databases are well suited to handle unstructured data and diverse data. Data models/schemas can be easily evolved in graph databases to address new data sources, metadata and new types of data. For Parabole, using a graph database for harmonizing diverse data is critical for speed and efficiency of data handling.

Analytics was also crucial. Parabole alphaESG provides a variety of analytics to provide quick insights to investment analysts. These include business intelligence-style analytics such as aggregates, averages, Excel-style functions along with sorting, filtering and other such functions. With a graph, an analyst may also be interested in using graph algorithms, inferencing, data science functions and even build their own custom algorithms. AnzoGraph DB's ability to provide such analytics capabilities and to support standard SPARQL functions were an important consideration for Parabole to select AnzoGraph DB.

Scaling for big data was also an important factor. Parabole alphaESG needed to scale from running on one limited server for smaller portfolios to incrementally add and scale to multiple servers as data volumes and performance needs grew. AnzoGraph DB scalable infrastructure allows Parabole to scale its customers and Parabole's business as data and performance needs expand. Many databases claim to scale, but trials with other graph databases often hit a wall when asked to scale up.

Finally, since many NLP technologies output data in RDF format and come from various vendors, a standards-based database engine was important to ensure all would work together. As a part of standards, it was important that AnzoGraph DB support labeled property graph (SPARQL*/RDF*) and OpenCypher as it allows Parabole to provide value-add analytics functionalities to its customers.



The Solution

The company began exploring the options in the graph database market and eventually landed on the decision to use AnzoGraph DB from Cambridge Semantics as the engine behind alphaESG. Their decision rested on five main decision elements:

- **Graph Data Model:** The wide variety of data sources required in this use-case necessitates harmonization of data across a wide range of heterogeneous data sources. Use of graph database helped achieve harmonization of data and it also helped overcome the challenges of performing analytics outside of the rigid schemas of the underlying datasets.
- **Analytics:** Parable AlphaESG needed to provide a wide range of insight, making depth of analytical capability important. With AnzoGraph DB, Parable has the power to perform simple analytics and go beyond with customizable algorithms.
- **Scalability:** AnzoGraph DB easily allows adding or removing server nodes when they are needed. This means that servers can be spun up as necessary. Although many graph database solutions in the market support additional server nodes, these solutions add servers to replicate data for geographic or data insertion rather than analytics scaling. Parable has leveraged this scalability of Anzograph DB to store to start with as many as 40 million triples of the AlphaESG product for retrieval and analysis and expect to be able to scale the solutions to billions and even trillions of triples as data needs grow. Parable did not find other databases that provide this unlimited scaling.
- **Speed:** Anzograph DB empowers the AlphaESG product's capabilities with its high loading, retrieval and analysis speeds. With Anzograph DB we have shown that we can easily load over 300 triples per second and perform high-speed analysis, all to exceed our loading and processing needs.

- **Standards:** While there are many graph databases on the market, only a few follow W3C standards such as RDF and SPARQL. Since natural language processing (NLP) is crucial to this solution, RDF compatibility was necessary. Most NLP technologies use RDF for text and entity/relationship extraction. Using graph formats other than RDF & SPARQL leads to lossy and brittle systems as unstructured data is converted from RDF to other graph formats for analysis. AnzoGraph DB does comply with W3C standards for RDF and SPARQL and is active in developing upcoming standards like labeled property graphs.

Outcome

Since AnzoGraph DB is a standard SPARQL endpoint, it was relatively easy to leverage for the ESG solution. It took Parabole 10 weeks to fully integrate AnzoGraph DB into alphaESG and test the integrated solution. Parabole designed a simple deployment architecture where alphaESG can be deployed on all major clouds or behind the firewall.

By integrating AnzoGraph DB, the solution is now able to load data significantly faster than before. Since AnzoGraph DB follows SPARQL, RDF and OWL conventions, powerful inferencing features can be leveraged for developing the special ontologies needed in the alphaESG platform.

"The capability and response from the AnzoGraph DB team have been very positive" said Rajib Saha, president of Parabole. "It's good to see we have an analytics partner in Cambridge Semantics."

Conclusion

alphaESG is an analytical tool for researchers and fund managers to contextually analyze news feeds, corporate filings and research reports to uncover activity scores, risk signals and investment opportunities more quickly than your competition. The solution has seen much interest with fund managers and subject matter experts.

By leveraging a powerful graph database like AnzoGraph DB and combining that engine with machine-based knowledge mining and cognitive analytics, Parabole was able to enable a smarter analysis of alternative data and identification of ESG risks at scale.

More Information

Parabole alphaESG, see: <https://www.parabole.ai/environment-social-governance/>

AnzoGraph DB, see: <http://anzograph.com>