

Business Driven Insight via a Schema-Centric Data Fabric

Dougal Watt¹, Brad Bebee², and Michael Schmidt²

¹ Meaningful Technology, Auckland, NZ

² Amazon Web Services, Seattle, WA 98101, USA

Abstract. Research shows that Analytics and Business Intelligence (BI) projects have high failure rates of up to 80 percent of projects, and suffer from low uptake of big data and AI tooling. Our internal research indicates two core problems that customers face when generating business outcomes from analytics. Firstly, business customers are often totally overwhelmed in knowing where to start. They report that even the simplest modern BI tools provide little or no help in analyzing their data and do not reflect the needs of their business. Secondly, when they do generate insights from their data they often have no trust in these, as they cannot understand provenance information about where the data came from, what was done to it by systems and individuals in their organization. We present a new approach to solving these analytics challenges using a new semantic ‘data fabric’ platform for business computing, based on the Amazon Neptune graph database service from Amazon Web Services (AWS) and other AWS cloud technologies. This architecture is purpose built around a semantic schema and custom tools that model and manage business data in a form understandable by business users. As a schema-centric platform, all insight is grounded from the definitions and relationships in actual business data, and provenance data captured throughout the platform and across all data flows.

Keywords: Business Model, Analytics, Insight, Business Intelligence, Cloud, Neptune, RDF, SPARQL, Architecture, Schema

1 Issues and Opportunities in Analytics

Recent research has highlighted problems with organizations deploying analytics and business intelligence tools and processes to gain business insights, and use these to effect business performance improvements. Gartner [1] estimates that until 2022, around 80 percent of analytics insights will not deliver business outcomes, while a survey by NewVantage [2] showed that 77 percent of businesses report challenges in the adoption of big data and AI, with many companies failing to use deployed tools.

1.1 Solving the Analytics Challenge

Existing BI and AI tools typically require moderate to significant domain knowledge to operate. For business users in the SME segment, these issues are often a considerable barrier as they typically do not have staff with the necessary skills and experience, and limited funds to afford consultants with the requisite skills. Similarly, traditional tools require significant effort in data preparation for data loads for machine learning, or data extraction, transformation and loading into data warehouses. The complex

steps required frequently result in loss of context for business users, such that the link between business information and insight becomes lost and insights remain unused.

To solve these issues, we created a novel approach to analytics using a ‘schema-centric data fabric’ architecture. A schema directly models business structures and terminology, and is used to drive the user experience of selecting, integrating, preparing, and shipping data to external analytics tools integrated with the platform. This removes the need for specialist, time consuming data preparation and tooling integration. All interactions with data are conformed to this schema, so any externally generated insights are immediately understandable by business users in their own language.

Our architecture addresses five key use cases: 1. manage a business model; 2. integrate data from cloud applications and data stores; 3. store/conform data into a semantic database according to the schema; 4. generate business insight from this integrated data; and 5. simplify tooling and operations via an integrated suite of interfaces, custom tools, cloud deployment architectures, and pre-built cloud services.

In our presentation we will outline our use of RDF/OWL to model business models, SHACL as a constraint and query language for integrating cloud apps, a full integrated suite of Amazon Web Services including the Neptune Database to store and manage conformed data in the form of a semantic RDF graph, and a set of custom components and user interfaces to generate insight. We will also showcase user interfaces used to manage these artifacts, with reference to specific customer use cases.

We highlight key aspects of Amazon Neptune that are relevant to our use case, including (i) high availability through data replication and automated failover, (ii) its transaction semantics, which facilitates parallel updates from independent sources with transactional guarantees, and (iii) Neptune’s support for lightweight in-database analytics. The flexibility to scale the graph database horizontally with read replicas on demand and using them for analytical workloads and large data exports allows us to scale compute resources and to separate resource intensive queries from the continuous OLTP workload, thus avoiding interference of analytics with regular operations.

Early work extending this architecture in novel directions includes the ability of Neptune to integrate with other data analytics and ML frameworks in the AWS ecosystem such as the Amazon Forecast service, a fully managed ML service to deliver forecasts that we will feed back into our insight tools. We will also showcase extensions to automate the operation of traditional data warehouses, using our insight tooling and Neptune’s Change Data Capture feature to subscribe to changes in graph data.

We are currently applying our approach with several businesses in New Zealand, with results showing rapid creation of analytics dashboards, simplified whole-of-company performance reporting, and feeding existing data warehouses with harmonized graph data. Use cases and lessons learnt from these customers will be discussed.

References

1. Gartner, https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/, last accessed 2019/06/25.
2. NewVantage, <https://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings.pdf>, last accessed 2019/06/25.