REPORT REPRINT

Yellowbrick builds its case for hybrid- and multicloud data warehousing

FEBRUARY 04 2021

By Matt Aslett

The high-performance data-warehousing specialist is seeing increasing customer demand for cloud-based adoption. That, along with a greater emphasis on performance from the cloud providers themselves, is helping to boost its case for hybrid- and multicloud data warehousing.

451 Research

S&P Global
Market Intelligence

Introduction

Yellowbrick emerged in 2018 with its approach to high-performance data warehousing, delivering an all-flash on-premises data-warehousing appliance, as well as availability as a service on the public cloud, initially particularly for high availability and disaster recovery (HA/DR). While the bulk of its early adoption deployments were on-premises, driven by the need for near-real-time analytics in the financial services, insurance and telecom sectors, Yellowbrick is seeing the majority of its new opportunities being driven by the cloud as customers look to take advantage of the availability of high-performance cloud instances and embrace strategic hybrid-IT architecture.

451 TAKE

While Yellowbrick's focus on flash-based performance naturally led to a preponderance of early on-premises engagements, it has always addressed cloud availability. Cloud (both private and public) and edge deployments are clearly going to be a bigger part of its future, especially given the growing availability of high-performance cloud instances. As these become more widely available, the performance arguments in favor of deploying Yellowbrick in the cloud, rather than on-premises, will increase. The vendor can remain agnostic, however, and is well-placed to benefit from the case in favor of a hybrid IT and multicloud data platform layer.

Details

Founded in 2014 by former executives of flash memory provider Fusion-io, as well as a variety of database firms, Yellowbrick emerged four years later with the Yellowbrick Data Warehouse, a massively parallel, PostgreSQL-compatible data-warehousing appliance designed to take advantage of the performance advantages of flash-based solid-state disk that also took advantage of replication to cloud-based variants, particularly (although not exclusively) for high availability and disaster recovery.

The company was initially successful in industries with a mission-critical need for near-real-time analytics and has established an installed base in the financial services, insurance and telecom sectors. While Yellowbrick doesn't disclose precise customer numbers, named adopters include LexisNexis, BMW Group Financial Services, Symphony RetailAI, and TEOCO.

Given the focus on performance and the appliance architecture, the majority of those initial deployments were naturally on-premises, but Yellowbrick has noted growing demand for its cloud services among new opportunities, particularly as the cloud suppliers themselves have placed heightened focus on the performance side of the price/performance equation, increasingly delivering high-performance instances based on NVMe SSDs.

This trend plays into the company's hands, enabling it to deliver support for stand-alone cloud data warehouses, as well as hybrid IT architecture that spans both on-premises and the cloud. Yellowbrick can be deployed as an endpoint in a customer's virtual private cloud on Amazon Web Services and Microsoft Azure, taking advantage of AWS PrivateLink and Azure Private Link (with plans to do the same on Google Cloud Platform when an equivalent is generally available), and delivers asynchronous data replication across instances with support for read-only secondary instances that support geo-distributed analytics, as well as HA/DR.

Secondary instances do not need to have the same topology as the master, providing support for global orchestration of data while maintaining regional data sovereignty requirements, and also making Yellowbrick potentially suitable for deployment for edge analytics use cases (such as data aggregation and filtering). Other key features that Yellowbrick believes differentiates it from its competition – which includes big-name data-warehousing providers Oracle, IBM, Microsoft and Teradata, as well as newer market entrants such as Snowflake and Cloudera, plus cloud offerings from AWS, Google Cloud and Microsoft Azure – include real-time streaming ingestion and parallel bulk loading.

Specifically, Yellowbrick offers connectors for streaming or batch bulk inserts using Apache Spark and Apache Kafka, with real-time loading at up to three million rows per second, and parallel bulk loading at up to 10TB per hour. Both of these capabilities could be used to accelerate migration projects, whether on-premises or in the cloud.

Toward the end of 2020, the company also boosted its advanced workload management capabilities with the launch of Yellowbrick Data Release 5, which included self-healing clusters for fault tolerance and more granular query visibility. Release 5 also included a new security model with more fine-grained security controls, as well as additional support for SQL user-defined functions.

Concurrent with the Release 5 launch, Yellowbrick also added a new Standard service plan, priced at $10,000 per month aimed at smaller-scale use cases with up to 10TB of data, including development and test environments and departmental data marts. The company's Enterprise service plan is also charged on a subscription basis, both on-premises and in the cloud.

Yellowbrick now has approximately 130 employees, with offices in Palo Alto, Salt Lake City, London and Hong Kong. The company has raised $173m in funding. The latest $81m series C round was announced in June 2019 and led by DFJ Growth with participation from IVP and BMW i Ventures, as well as Next47, Third Point Ventures, Menlo Ventures, Threshold Ventures, and GV.