Market Insight Report Reprint

Yellowbrick adopts Kubernetes, adds control plane for data warehousing in the distributed cloud era

April 26 2021

by Matt Aslett

With the recent addition of its Yellowbrick Manager and support for Kubernetes, Yellowbrick is positioning itself to address the growing requirements for data warehousing that is distributed across on-premises datacenters and multiple clouds.

451 Research

S&P Global
Market Intelligence

Introduction

We recently noted that data warehousing software and appliance provider Yellowbrick had adapted its offering to reflect the requirements for hybrid- and multicloud data warehousing by providing a consistent experience on-premises and in multiple clouds. Recently the company went one step further, delivering a preview of its Yellowbrick Manager functionality, which is designed to provide a unified control plane for managing data warehouse deployments distributed across multiple clouds and datacenters. The company has also adopted cloud-native architecture via Kubernetes to support that distributed cloud approach and has also delivered the latest version of its optimized hardware instances for private cloud. Known as Andromeda, the second generation of Yellowbrick's all-flash optimized architecture is said to deliver 3x performance improvements.

THE 451 TAKE

At the beginning of this year we outlined the case for adopting an infrastructure-agnostic data platform layer that can run on multiple clouds as well as on-premises datacenters. The ability to use the same database in multiple locations is already addressed by most database providers today but managing data across those multiple locations is easier said than done. Yellowbrick will be aiming to position the functionality being delivered by Yellowbrick Manager as a differentiator, alongside its Andromeda architecture and Kubernetes support, to help enterprises as they look to evolve their investments in hybrid IT and manage data consistently across distributed clouds that span private cloud datacenters, public cloud services and edge devices.

Details

Despite much consolidation over the years, the data warehousing sector remains crowded, with established incumbents jostling with emerging startups for attention. Providing differentiation is key to gaining traction. Yellowbrick, which was founded in 2014 and emerged from stealth four years later, has a couple of factors it can point to: a high performance an all-flash data-warehousing appliance for on-premises, and availability as a service on the public cloud with asynchronous data replication to support hybrid IT and multicloud requirements. The company recently announced updates to Yellowbrick Data Warehouse that deliver enhancements related to both these factors, including the general availability of its Andromeda second-generation optimized hardware architecture, support for Kubernetes and a preview of Yellowbrick Manager for managing data across a distributed cloud and datacenter architecture.

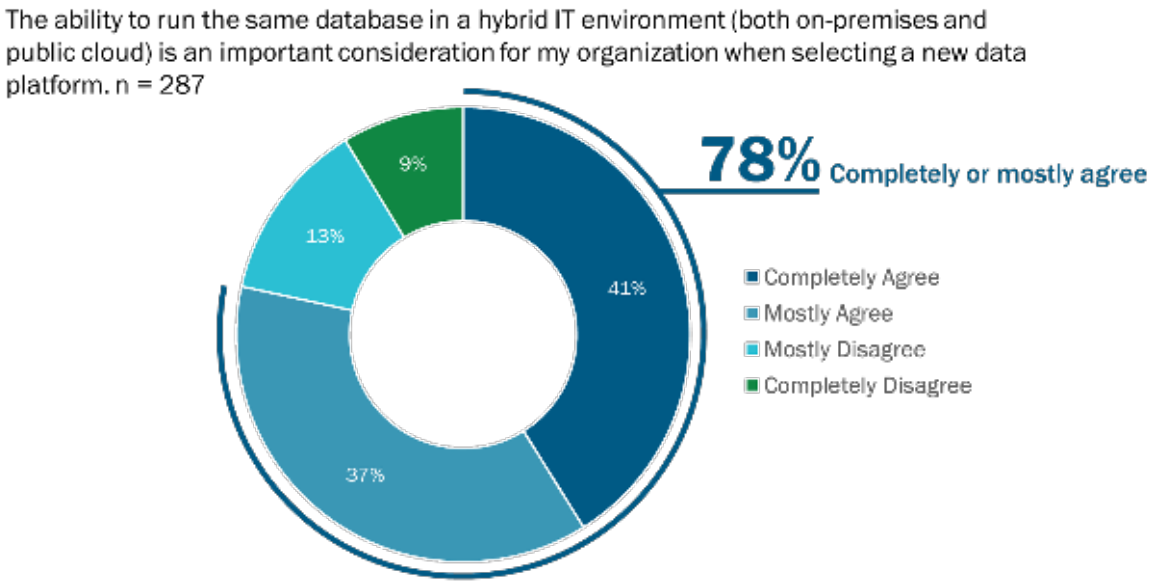
Adopting Kubernetes as the core component of its cloud-native architecture better enables Yellowbrick to address portability across multiple environments – including running on Kubernetes in a customer's own datacenter, as well as public cloud Kubernetes offerings, and also on edge architecture. Kubernetes is also being built into the Andromeda optimized instances later this year. In the interim, Andromeda incorporates new 'Kalidah' scan accelerators for data scan rate improvements and also delivers network performance improvements, and improved linear scalability (up to 6PB across 40 nodes). The Andromeda instances are primarily positioned for on-premises private clouds but can also be managed as a service via AWS PrivateLink and Azure Private Link (with plans to do the same on Google Cloud Platform when an equivalent is generally available).

With its 5.2 release, Yellowbrick has also delivered the general availability of improved support for data lake architecture integration – specifically enabling data to be loaded into Yellowbrick Data Warehouse from tables in cloud object stores, such as Amazon S3, Azure Blob Storage and Google Cloud Storage, as well as MinIO object stores and the Hadoop Distributed File System. The loading of data from these object stores can be performed via SQL integration tools, or orchestrated via Yellowbrick Manager, which is now in preview and due to be generally available before the end of the year. Yellowbrick Manager is a significant addition to Yellowbrick Data Warehouse in that it is designed to provide a unified control plane to enable users to manage data distributed across multiple locations – including private cloud, public cloud or edge devices.

Initially aimed at data engineers and developers, with additional management console functionality to come aimed at database administrators, Yellowbrick Manager provides functionality to enable users to view the configuration and connection details of multiple Yellowbrick Data warehouse instances, as well as load and move data, and explore data and create queries with a built-in SQL editor. Yellowbrick Manager is designed to provide a unified management experience that complements the company's existing data loading and replication functionality, including the previously mentioned connectors for streaming or batch bulk inserts using Apache Spark and Apache Kafka, as well as asynchronous data replication across instances with support for read-only secondary instances.

While many operational and analytic database vendors offer their products for deployment on-premises as well as consumption via public clouds, 451 Research believes that the ability to manage data across these environments is becoming increasingly critical. As the figure below illustrates, more than three quarters (78%) of respondents to our Voice of the Enterprise: Data & Analytics, Data Platforms, 2021 agreed that the ability to run the same database in a hybrid IT environment is an important consideration for their organization when selecting a new data platform.

Importance of Database Support For Hybrid IT



Source: 451 Research, Voice of the Enterprise: Data & Analytics, Data Platforms, 2021

CONTACTS

The Americas
+1 877 863 1306
market.intelligence@spglobal.com

Europe, Middle East & Africa
+44 20 7176 1234
market.intelligence@spglobal.com

Asia-Pacific
+852 2533 3565
market.intelligence@spglobal.com

www.spglobal.com/marketintelligence