

REPORT REPRINT

On the Yellowbrick road: data-warehousing vendor emerges with funding and flash-based EDW

JAMES CURTIS

30 AUG 2018

After years in the making, Yellowbrick has emerged from stealth with an on-premises, flash-based data-warehousing system aimed squarely at large enterprises looking for flexibility in carrying out operational analytics.

THIS REPORT, LICENSED TO YELLOWBRICK DATA, DEVELOPED AND AS PROVIDED BY 451 RESEARCH, LLC, WAS PUBLISHED AS PART OF OUR SYNDICATED MARKET INSIGHT SUBSCRIPTION SERVICE. IT SHALL BE OWNED IN ITS ENTIRETY BY 451 RESEARCH, LLC. THIS REPORT IS SOLELY INTENDED FOR USE BY THE RECIPIENT AND MAY NOT BE REPRODUCED OR RE-POSTED, IN WHOLE OR IN PART, BY THE RECIPIENT WITHOUT EXPRESS PERMISSION FROM 451 RESEARCH.



©2018 451 Research, LLC | WWW.451RESEARCH.COM

After four years in development, Yellowbrick emerged from stealth by rolling out its maiden data-warehousing product. Although the company's offering is an on-premises-based enterprise data warehouse, it does integrate with cloud services, such that it could be categorized as a hybrid system. The company employs a flash-based architecture, delivering what it is calling the ability to carry out 'native flash queries.' Off to a quick start, Yellowbrick announced that it has secured \$44m in a series A funding round.

THE 451 TAKE

At a time when many vendors are looking to the cloud, Yellowbrick comes to market with a deliberate on-premises system, albeit with integration to the cloud, positioned as a hybrid system. The company is further differentiated by its flash-based architecture leveraging SSDs. This approach differs from many other vendors that have gone the route of in-memory processing. While leveraging flash is not entirely new, Yellowbrick's architecture has the option to bypass RAM altogether, although it can be used for demanding joins or similar workloads. Our belief is that Yellowbrick can likely spur some performance/pricing discussions in the market, given the dense nature of flash memory over RAM. However, with an initial on-premises product, the company will appeal less to a cloud-looking customer and more to those with eyes for on-premises deployments.

CONTEXT

Founded in 2014, Yellowbrick took four years to develop its data-warehousing product. The startup is based in Palo Alto, California, where its executive team consists of a collection of former Fusion-io employees that also have experience at several noted technology firms such as Google, IBM, Informix, Intel, Microsoft, Netezza and Snowflake Computing.

Yellowbrick just emerged from stealth in July, but already boasts customers in the 'high single digits,' according to management. Some notable customers include Teoco, Symphony AI, Melco and Overstock.com. The company also secured \$44m in a series A funding round that includes participation from DFJ, GV, Menlo Ventures, Samsung Ventures and Third Point Ventures.

PRODUCTS

To understand Yellowbrick's data-warehousing product is to understand the company's reasoning for coming to market. It is estimated that data warehouses first emerged in the early 1980s, so the market is multiple-decades old, and boasts a number of established players. Therefore, a new entrant is not necessarily unheard of, but it is less common.

Regardless of market maturity, Yellowbrick believes that there still exist many shortcomings here, specifically pointing out the lack of high availability, inability to run ad hoc queries, poor scalability, inability to run mixed real-time workloads, and challenges handling high concurrent numbers of users. As such, the Yellowbrick Data Warehouse product is targeted specifically at these challenges.

The company's data-warehouse initial offering is an on-premises system but has the ability to integrate with cloud services, such that Yellowbrick refers to it as a hybrid system. However, management notes that a cloud-specific product is in the works, although no specific timetable was provided. While coming to market with an on-premises system may seem a bit out of the ordinary, given the growth of cloud computing today, Yellowbrick managements notes that it is targeting the largest of large enterprises – where cloud computing could be cost-prohibitive, or where the cloud is not an option, based on regulatory compliance, security or other factors.

The company also points, however, to its small-footprint architecture as providing an attractive on-premises option for its clientele. Yellowbrick is based on an all-flash architecture, but it's how the data moves through the system that constitutes its differentiation, what it calls its 'native flash queries' capability. In a traditional disk-based system, data moves from the storage media (spinning disk or SSD) to the RAM, where data is then moved to CPU cache for processing. Depending on storage speed, throughput and CPU speed, performance can vary.

With the Yellowbrick Data Warehouse, data is stored in flash memory SSDs, and comes off this memory as discrete data packets. It goes directly to the CPU cache where it can then be processed in parallel by the CPU cores. Effectively, Yellowbrick bypasses the RAM; however, larger joins, aggregations, or hashes that need to be carried out can be moved into RAM, and then to CPUs for processing. There are a few benefits to this type of architecture, claims Yellowbrick, such as the fact that SSDs are significantly denser, so the overall system footprint is greatly reduced, leading to reduced costs.

From a core database perspective, Yellowbrick is natively built, in that the company has developed its own full stack to include the operating system, storage engine, memory manager, cluster, and query planner, all components that make up the data management system. However, from an interoperability standpoint, Yellowbrick is compatible with PostgreSQL such that it can leverage the tools and resources that constitute the broader PostgreSQL ecosystem.

COMPETITION

Given Yellowbrick's offering of an on-premises system, we expect its primary competition to be the incumbent data-warehousing vendors. These incumbent vendors all started out providing on-premises systems that vary between leveraging specific, optimized hardware to the use of commodity hardware. Many of the known vendors include Oracle with its Exadata system that is part of the company's Engineered Systems lineup, Microsoft's SQL Server Data Warehouse, and IBM's Db2 lineup that includes the IBM Integrated Analytics System, combining Db2 and Netezza technology.

Still, there are others that include SAP's Business Warehouse that leverages the SAP HANA platform, Teradata's Analytics Platform and Database offerings provided in appliance form on the company's IntelliFlex and IntelliBase hardware, Pivotal's Greenplum that is available as open source software, Actian's Vector that is also available as Vector in Hadoop, Micro Focus' Vertica, Exasol and MemSQL.

While all of the previous vendors noted they offer their wares as cloud deployments, some vendors are pushing cloud-only data warehouses, although these offerings may not be direct competitors with Yellowbrick. Vendors include Amazon Redshift on AWS, Microsoft SQL Data Warehouse on Azure, Google BigQuery on Google Cloud Platform, Snowflake Computing, and IBM's Db2 Warehouse on Cloud (formerly dashDB).

Also worth noting are the distributed data processing framework vendors (previously referred to as Hadoop distributors) that leverage open source tools as well as some proprietary components. Vendors include Cloudera with its Analytic DB offering, Hortonworks with its Enterprise Data Warehouse product, and MapR with its Data Warehouse Optimization and Analytics offering.

SWOT ANALYSIS

STRENGTHS

Yellowbrick provides a differentiated flash-based architecture that allows data to bypass the RAM and feed directly into the CPU cache, providing a compelling and potentially cost-reducing scenario.

WEAKNESSES

The company's initial data-warehouse product will appeal primarily to an on-premises customer because the offering is a physical system setup, with a cloud offering still under development for later release.

OPPORTUNITIES

There are still many customers to be had that are looking to refresh their legacy on-premises data-warehouse systems, and that might be looking to reduce overall costs while at the same time needing to maintain performance SLAs.

THREATS

While there continues to be a list of longtime data-warehouse vendors, there are emerging competitors ranging from cloud-based vendors to Hadoop-based, to even some NewSQL vendors that are interested in peeling away certain data-warehousing workloads.