

## REPORT REPRINT

# At MarkLogic World, it's all about making the Data Hub 'smarter' and easier to consume

**JUNE 04 2019**

**By James Curtis, Paige Bartley**

At its annual user conference, MarkLogic made several announcements, from enhanced security to a specific application for the pharmaceutical sector, confirming its commitment to be 'all in' with its data hub strategy as a means to combat the data-siloed environment.

---

THIS REPORT, LICENSED TO MARKLOGIC, 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.



### Introduction

At the company's latest user conference held in Washington DC, MarkLogic made several product announcements, all of which had some tie-in to the company's data hub offering that it has been focusing on over the past few years. But the cloud was also a strong theme of the conference and is expected to play a key part of the company's future. Regarding MarkLogic Data Hub, which can be deployed on-premises or in the cloud, the company will enhance its security certifications and add a machine learning engine, and it has rolled out a pharmaceutical-specific version of the data hub. MarkLogic Data Hub Service, announced in October 2018 for AWS, will now be available on Microsoft Azure.

### 451 TAKE

Even though MarkLogic announced official data hub support at its conference in 2018, the company has been peddling a data consolidating positioning for some time. The message coming out of this year's event, however, was that the company's data hub offering is becoming clearer, if not maturing to some extent, based on the company's most recent announcements. The announced Pharma Research Hub, for example, is an abstraction layer that sits above the MarkLogic Data Hub and is geared to the life sciences vertical, making it easier for researchers to find relevant data. Further, the company has embedded machine learning capabilities that are expected to lay the foundation for modern data management capabilities at scale for the future and affect a number of other functional areas. For instance, none of the immediate announcements at the conference were specifically focused on governance functionality, but we expect machine learning to have major roadmap implications. Also, with the MarkLogic Data Hub already having embedded data mastering and curation capabilities, further machine learning functionality will very soon make these functions more accurate and 'smarter.' With the announcements, the company is clearly making the data hub strategy its central focus, which appears to be resonating with many enterprises, but MarkLogic can be deployed in other use scenarios, so the hope is that the former doesn't overshadow the latter.

### Context

As we have noted in previous research regarding MarkLogic, the company has been driving a data silo consolidation strategy for many years. From its early days as an XML database, MarkLogic has evolved to a multi-model NoSQL database, enabling not only XML data types but adding JSON (document) and semantic (graph) capabilities in 2013.

Previously, MarkLogic has stressed its operational and transactional capabilities, but the company's specific focus on a data hub as a primary deployment pattern for its database really started to take shape at 2018's MarkLogic World conference. At the event, the company announced a specific Data Hub product framework that could be downloaded as open source from GitHub. The GitHub offering was geared to automate and simplify the setting up and deploying of MarkLogic as an operational data hub, where a variety of data types, including XML, JSON, time series, text, geospatial and columnar data, for instance, can be ingested into the system. Once ingested and indexed, the data hub can function as a type of serving layer, whether that be specific applications, analytics or other downstream systems.

### Products and announcements

Since the official release of the Data Hub offering in May 2018, the company also rolled out its MarkLogic Data Hub Service in October 2018, which was initially available on AWS but is also now available on Microsoft Azure, as announced at the conference. The Data Hub Service is functionally similar to the on-premises deployment setup, but as a cloud service, MarkLogic handles certain aspects of the service such as hardware provisioning, managing back-ups, scaling, and updates and patching, along with providing monitoring and support.

While the Data Hub Service is fairly new, MarkLogic sees the service playing a critical role for the company going forward, particularly since the service can be spun up quickly without the need to work with enterprises on procuring hardware, for instance. But besides a quicker path to adoption, the company points to its pricing model as a key aspect of the service and favorable for enterprises because it offers a consistent subscription price while also being able to handle unplanned demand. The pricing approach is similar to how the telecoms manage cell phone rollover minutes. Essentially, enterprises can 'build up' credits that are accrued and can be applied when demand rises. The speed at which credits are used and accrued are based on parameters set during the initial set up of the service, with the primary objective of providing predictable pricing for enterprises.

Because the service is fairly new, MarkLogic did not disclose the number of current customers but the company did report that it has over 30 customers running the Data Hub in production and over 50% of its customers are running MarkLogic in some type of cloud environment.

With a clear focus on its data hub strategy, the company announced a new cloud-based, vertical-based data hub service targeting the pharmaceutical industry, called the MarkLogic Pharma Research Hub. The new offering is built atop the MarkLogic Data Hub and is geared to pharmaceutical researchers and others in similar roles and consists of built-in, ready-to-use data services specific to the pharmaceutical industry. Researchers often struggle to find relevant research, wondering whether the research exists or if incorrect search terms were applied in finding the data. Regardless, the new Pharma Research Hub incorporates machine learning capabilities, including other technologies based on semantics and fuzzy matching, for instance, to aid in the organizing, navigating, collaborating and retrieving of relevant data that can be carried out with a 'single pane of glass' interface.

### Governance and data management roadmap

The new Data Hub Service, with its higher levels of abstraction and natural pairing with the announced machine learning for the Data Hub, offers the potential for disruption of the traditional ETL tooling market. While it is not uncommon for customers today to still use ETL products to get data into the MarkLogic environment, the current abilities of the Data Hub Service will streamline the process of ingesting and curating data even further. The newly announced partnership with MuleSoft, in particular, will further help populate the underlying database. Because data can be ingested as-is and then handled via the Data Flow Orchestration and Smart Mastering capabilities, with curation done directly in Data Hub, much of the dependencies on the 'transform' step of ETL are potentially eliminated.

Data management capabilities are further bolstered by the scope of the MarkLogic World 2019 conference announcements, although largely indirectly through their future roadmap potential. There may have not been a single pivotal 'home run' announcement in terms of focused data governance functionality at the conference – MarkLogic has always placed a heavy emphasis on governance regardless, perhaps unique in the non-relational space – but the embedding of machine learning-driven automation into the database ecosystem and Data Hub will infuse existing security and governance controls with the ability to detect abstract, ill-defined relationships at scale: something that will increasingly be needed as the definition of 'personal data' becomes more esoteric with the proliferation of global data privacy and data protection regulations.

451 Research foresees a future in which non-relational data management is neither a free-for-all 'wild West' data access scenario nor a combative, locked-down and security-centric paradigm that blocks users from data access by default. In theory, there is a happy medium, and this is what MarkLogic has been ostensibly working toward for the last several years. Given its demonstrated track record with the US government agencies, perhaps a government aphorism is appropriate: 'trust, but verify.' MarkLogic's conference this year underscored that it is not interested in security for security's sake, but rather in the facilitation of secure collaboration with data. Simply blocking access to data is not, and should not, be the goal. Any attempt to prevent end users from seamlessly sharing work products and ideas is frowned upon because the security of the system should ultimately be firm but invisible to the end user.

### Competition

With its focus on a data hub strategy, MarkLogic is likely to see a number of competitors from a number of areas, primarily due to the fact that there is no agreed-upon industry definition for a 'data hub.' But MarkLogic has been driving a data silo consolidation story for some time, so a move to position as a data hub is a natural progression.

That being the case, a data hub is sometimes likened to a data lake, which puts the company up against the distributed data-processing framework vendors (Hadoop) such as Cloudera and Hortonworks, which it acquired in 2018. But MapR and Qubole are also vendors in the data lake business, as is Cazena. Data lake management providers include the likes of Zaloni, offering some analogous governance capabilities. But so-called data lakes with the promise of loading disparate data sources to object storage while providing analytical access, for instance, can be built with cloud services. On AWS, some services might include AWS EMR (Hadoop), Kinesis (streaming), DynamoDB (NoSQL), Neptune (graph), Elastic Search (search) and on Azure, services can include HDInsights (Hadoop), Data Factory (data integration) and Cosmos DB (multi-model NoSQL). SAP, too, has its own Data Hub branded offering, although it uses federated push-down processing without resting on a single database, and is focused more on the overall orchestration and pipelining of data.

Regarding other specific NoSQL competitors, MarkLogic is considered one of the older and more established vendors; it got its start in 2001, while the general NoSQL space didn't become a recognized market until 2009. Regardless, MarkLogic will likely see other NoSQL multi-model vendors such as MongoDB, DataStax, Redis Labs, Couchbase, MapR-DB, ArangoDB and OrientDB as competitors. InterSystems Cache is another, and while not technically categorized as a NoSQL database, it provides NoSQL-like document database capabilities.

It's also noteworthy to point out that MarkLogic is realistically likely to meet the large relational database players, specifically Oracle, MySQL, IBM and Microsoft, in the sales cycle. MarkLogic does displace systems from these vendors, but a common scenario is working alongside them and then over time taking on more workloads. Displacement may not always be the result; a complementary strategy is also common, which can play into the company's data silo-consolidation strategy. Regarding its strategy against these larger relational database vendors, MarkLogic often cites its data flexibility – the ability to ingest a variety of data types and then serve up that data upon ingest.

SWOT Analysis

**STRENGTHS**

MarkLogic provides a stable and highly secure NoSQL database offering that has an established track record of driving mission-critical applications for large organizations. Governance capabilities come first and foremost for the company, so it's little wonder that that MarkLogic still often competes against more-established relational players in the sales cycle, rather than their non-relational technological peers.

**WEAKNESSES**

In the non-relational data management market, having a strong developer following is a valuable brand asset. MarkLogic struggles with this aspect because it tends to focus marketing resources on top-down efforts targeting high-level decision-makers. Despite the widely available free version of the company's offering, it faces an uphill battle against some of its non-relational peers in terms of developer recognition, adoption and usage.

**OPPORTUNITIES**

The newly announced Pharma Research Hub, while targeted for life sciences, has piqued the interest of other industries looking for like-minded offerings and suggests that the MarkLogic Data Hub can be applied both horizontally as well as for industry-specific verticals.

**THREATS**

Major cloud providers are the major threat here; however, it is possible to architect MarkLogic's capabilities from individual cloud providers' offerings and components. Orchestration and ongoing maintenance are, of course, the perpetual struggle should the enterprise take this route.