

SQL Server 2019 Preview Brings Hadoop, Spark

and AI into DBMS



A tilted white banner containing the text "SQL SERVER 2019" in a bold, sans-serif font. To the left of the text is a stylized wireframe globe icon. Below the text, the URL "ttmind.com" is visible in a smaller font.

A SQL Server 2019 Preview Brings Hadoop and Spark into the Fold, Positioning the Database as an All-Purpose Data Portal. Azure Cosmos DB Also Gets an Update.

Microsoft continued this week to move its SQL Server database platform beyond relational data and deeper into a variety of data types.

The company released a SQL Server 2019 preview that supports Apache Spark and the Hadoop Distributed File System (HDFS), along with various machine learning packages that could combine to make SQL Server a test bed for many shops' first forays into big data analytics.

A slew of open source Hadoop components come in the new version of the venerable database management system (DBMS). Besides HDFS and the Spark data processing engine, Apache's Knox authentication gateway, Ranger se-

curity framework and Livy job scheduler will be available in Linux containers running on Kubernetes clusters as part of SQL Server 2019.

In recent years, Apache Spark has become a go-to tool for many organizations leading the big data charge. It enables high-volume analytics; large-scale extract, transform and load conversions; machine learning; and other workloads.

Overcoming fear of big data

But installing Spark along with HDFS in clusters has required skills beyond those common in most IT shops. The news on the inclusions of Spark and HDFS in SQL Server 2019 came at the Microsoft Ignite 2018 conference in Orlando, Fla.

“Basically, Microsoft is going to bundle SQL Server along with the installation processes needed to have an HDFS and Spark cluster running, as well,” said Warner Chaves, a principal consultant at technical services provider Pythian, based in Ottawa. “That is important, because many customers have felt too intimidated to do their own big data infrastructure.”

Doug Henschen, a principal analyst at Constellation Research, agreed this offering may appeal primarily to green-field deployments for organizations that don’t have other big data infrastructure.

But, he continued, it helps SQL Server shops to plan their data platform strategy of the future.

“Data scientists, data analysts and even progressive database management types have been increasingly blending data from structured SQL databases and big data sources, such as HDFS,” Henschen said. “They amass unstructured data, such as log files, social streams, JSON from mobile apps, clickstreams and other variable sources.”

Microsoft SQL Server 2019 is intended to unify these worlds into a single DBMS platform built to run on Kubernetes, he said, either on premises or in the cloud on IaaS.

Taken together with Azure Data Studio, which became generally available Sept. 24, this can combine to provide a single developer interface supporting both SQL data access to structured data contained in the SQL Server store and notebook-style access to all the data and the Spark engine running on the same platform, Henschen said.

Initially, the big data clusters capability with Spark and HDFS is available only to users that sign up for Microsoft’s SQL Server 2019 early adopter program. Otherwise, the preview release can be downloaded for use on Windows, Linux and Docker without the big data features.

Tomorrow, the Cosmos DB

Among tools and APIs discussed at Ignite 2018 were plug-ins to the aforementioned Azure Data Studio, which allows developers to work with SQL Server 2019, and PolyBase connectors to Oracle, Teradata, MongoDB, PostgreSQL and other external databases for analyzing data stored in them from SQL Server.

Azure Cosmos DB, Microsoft’s all-purpose multimodel cloud database, has also been updated with multimaster read and write capabilities across all cloud regions.

With Cosmos DB, developers can build global cloud systems that “can auto-replicate elastically at scale,” said Scott Guthrie, executive vice president of the Microsoft cloud and enterprise group, as part of the product rollout during the conference.

Guthrie said developers can have ready entry into the database via common APIs to MongoDB, MariaDB and other data stores that are now generally available. At the conference, Guthrie described a new addition: a Cassandra API plug-in for Azure Cosmos DB.

Today, SQL Server; tomorrow, Cosmos DB. Microsoft’s Scott Guthrie described available APIs for data developers at Ignite 2018.

Shell’s game

Much of the spotlight at Microsoft Ignite 2018 was on data as an enabler of AI and business transformation.

Onstage at the event was Yuri Sebrechts, executive vice president for technology and CTO at oil giant Shell, which, he said, employs Microsoft’s Azure cloud and data tools to assist IoT efforts. He commended Microsoft’s embrace of data-related open source technologies.

“We like a platform that allows us to bring in the latest developments in open source options,” he said, adding that Shell looked to exploit open source not just on the cloud, but also on the edge, which includes, in this case, Shell’s 44,000 retail outlets.

Organizations need to be ready to shift their general notions about data and analytics, Sebrechts indicated.

“Putting the data centric to what you do is different from the past when opinions mattered more than data,” he said.

All and all, with its diverse software debuts, Microsoft Ignite 2018 provided plenty of impetus for such a shift in thinking about data.