Dell EMC Storage for Microsoft SQL Server 2019 Big Data Clusters

Gaining insights on data wherever it sits while ensuring optimal performance and agility with Dell EMC storage platforms.

What are Big Data Clusters?

Data Virtualization

By leveraging SQL Server PolyBase, SQL Server Big Data Clusters can query external data sources without moving or copying the data. SQL Server 2019 introduces new connectors to data sources.

Data lake

A SQL Server big data cluster includes a scalable HDFS storage pool. This can be used to store big data, potentially ingested from multiple external sources. Once the big data is stored in HDFS in the big data cluster, you can analyze and query the data and combine it with your relational data.

Scale-out data mart

SQL Server Big Data Clusters provide scale-out compute and storage to improve the performance of analyzing any data. Data from a variety of sources can be ingested and distributed across data pool nodes as a cache for further analysis.

Integrated AI and Machine Learning

SQL Server Big Data Clusters enable AI and machine learning tasks on the data stored in HDFS storage pools and the data pools. You can use Spark as well as built-in AI tools in SQL Server, using R, Python, Scala, or Java.

Source: https://docs.microsoft.com/en-us/sql/big-data-cluster/big-data-cluster-overview?view=sql-server-ver15

Linux Based Containers for Microsoft SQL Server 2019

Microsoft has continued to innovate on Microsoft SQL Server over the course of the past 25 years. However, 2019 seems to stand out among the previous releases with robust, powerful new PolyBase features that vastly extend on features introduced in Microsoft SQL Server 2017.

Support for Linux based containers and orchestrating those containers through Kubernetes allows for an environment that can scale to meet the needs of Big Data Clusters. Big Data Clusters is likely to be a widely adopted feature and the first feature of Microsoft SQL Server that companies will want to implement.

Gaining Insights through Data Virtualization

Big Data Clusters include new connectors, leveraging SQL Server PolyBase, to allow for the querying of external data services where the data sits. This will virtually eliminate costly ETL processes that hinder the ability for companies to make better decisions based on faster access to their greatest asset, their data.
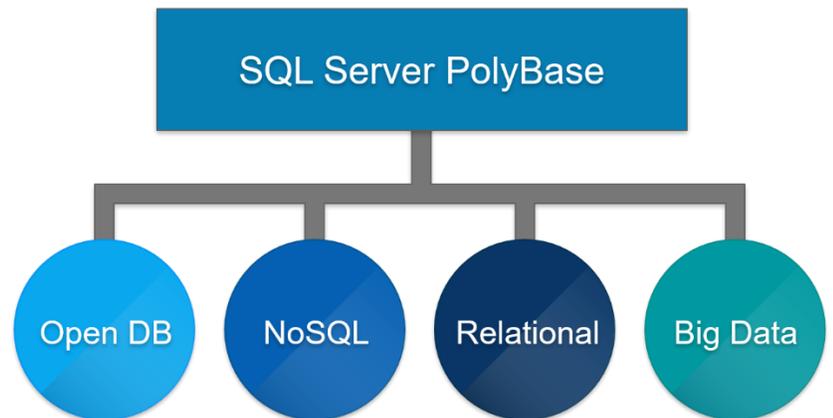
Moving to Microsoft SQL Server 2019 will not happen all at once, however, the implementation of Big Data Clusters will enable companies to gain greater insight on data that is likely sitting in multiple earlier versions of MSSQL. Consolidation of these data sources to the right storage platform(s) can elevate the performance and availability of earlier versions while maximizing the success of deploying Big Data Clusters and executing virtualized queries across block, file and object data throughout the datacenter.

Modernization is Not Optional

To improve competitiveness in a fast-paced world, organizations are embracing software as a prime differentiator. A software-driven business is more agile and efficient, innovates faster and responds dynamically to changing market and customer demands. Increasingly, gaining these compelling advantages requires adopting cloud-native applications. The benefits of containerized Microsoft SQL Server lie in the efficient provisioning, startup times, DevOps improvements, and being able to significantly reduce the overall storage capacity versus a VM.

Containers are ephemeral by nature, so the data that needs to be persistent must survive through the restart/re-scheduling of a container. When containers are rescheduled, they can die on one host, and might get scheduled on a different host. In such cases, the storage should also be shifted and made available on a new host for the container to start gracefully. This is functionality is delivered via the CSI plugin, developed to provide an open API to connect persistent storage to containers.
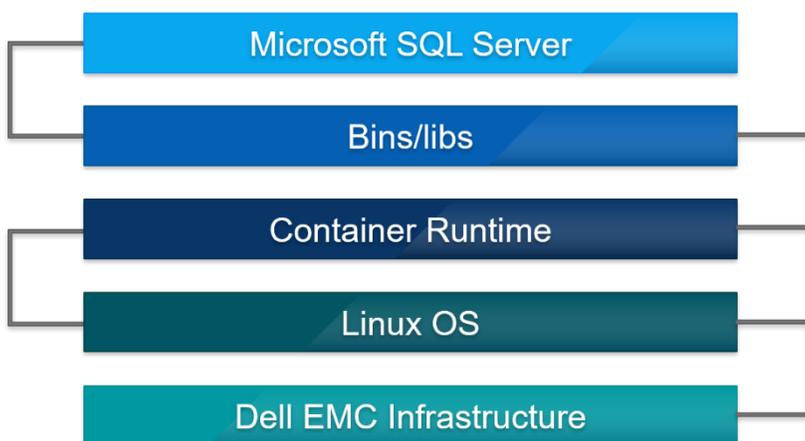
Dell EMC brings the power of persistent storage to a stateless platform. Dell EMC's storage and CSI plug-ins enable container orchestrators, like Kubernetes, to easily provision highly available and scalable container volumes for stateful containerized applications and Microsoft SQL Server databases.

As of the release of Microsoft SQL Server 2019, Dell EMC presently has support for CSI on Dell EMC Unity XT, PowerMax, VxFlex and XtremIO. CSI driver capabilities are constantly being improved upon so, be sure to leverage the latest version and follow along with updates on GitHub.com/Dell.

HDFS Tiering: Seamless Access to Isilon

Unification of data lakes is not an easy task, the rapid growth of data on-premises and in the cloud causes a wide variety of performance. HDFS Tiering allows for data stored on Isilon to be remotely mounted into the big data cluster for use by Spark and Microsoft SQL and leverage analytics and BI.

Isilon storage solutions are designed for enterprises that want to manage data, not storage. Simple to install, manage, and scale. Nodes can be added to an Isilon cluster as needed without downtime, manual data migration, or application logic reconfiguration, thereby saving IT management resources and avoiding operational interruptions. Unlike traditional Hadoop deployments, with Isilon you also gain the flexibility to scale storage resources independently from your compute resources. And our solutions stay simple no matter how much capacity is added, how much performance is required, or how business needs change in the future.

Protecting Big Data Clusters

Big Data solutions are relatively new and while they are many different combinations and deployments, most face a common issue. The amount of data stored in these environments is in the terabytes to petabytes, but backup and recovery is often an afterthought. The size of the data combined with multiple sources can make them a real challenge to backup and recover. Common data protection strategies include rebuilding and rehydrating the big data environment, or maintaining parallel copies, both of which can be costly and time-consuming. As these Big Data Solutions become increasingly critical to business, the availability requirements will inevitably increase making these strategies unacceptable.

The storage snapshot capabilities of Dell EMC storage allow these environments to be protected by storage snapshots in a matter of seconds, allowing for fast, efficient point-in-time copies, and recovery times in minutes, not days or months. For more information on Dell EMC Storage solutions for Microsoft SQL Server, continue reading over at DellEMC.com/Storage-for-SQL