

Dell Technologies Solutions for Microsoft SQL

SQL Server 2019 Big Data Clusters: A Big Data Solution Using Dell Technologies Infrastructure.



Benefits of Microsoft SQL Server Big Data Clusters on a Dell Technologies Foundation

- Enable intelligence over all your data
- Remove the limitations created from data silos by combining both structured and unstructured data across the entire data estate
- Deploy scalable clusters using Apache Spark, HDFS containers with Kubernetes, and SQL Server
- Deliver secure, isolated virtualized containers quickly and easily
- Create persistent storage for stateful applications
- Simplify provisioning, management and orchestration of container storage via Container Storage Interface (CSI) plugins
- Linear scalability for consistent performance and minimal latency

Keeping up with Big Data

Big Data is comprised of structured and unstructured data being generated by social networks and the internet of things (IoT), personal electronics and apps, marketing questionnaires, product purchases, and so much more. The expansion of IoT, connected devices and people is generating volumes of data that is exceeding the storage capacity of traditional database systems. By some estimates, we generate as much as 2.5 quintillion bytes of data each day¹.

With Big Data comes big challenges, including:

- The inability to quickly and easily access all types of data, locked within storage silos, in order to perform business analysis and reporting.
- Time, cost and resource consuming extract, transform, load (ETL) processes which also inhibits near real-time analysis.
- IoT data sets are unstructured and without business context, often in formats that are not suitable for storing in relational database tables or for querying using relational query semantics.

You require a solution that bridges business/operational data stored in relational databases with IoT data to generate true, real-time, Business Intelligence (BI) that leads to self-learning intelligent applications and business processes.

Microsoft SQL Server 2019 – Unified Data Platform

Microsoft SQL Server 2019 is a database management system that is continuing to extend its capabilities as a platform for data unification. SQL Server 2019 is a hub for data, with the ability to query over data in Oracle®, Teradata and Hadoop® in a secure and highly performant way — all without moving or copying it. The key capabilities of SQL Server 2019 for unifying virtual business data and operational data stored in relational databases with IoT for true real-time BI and embedded Artificial Intelligence (AI) and Machine Learning (ML) include:

Linux and Containerization

Microsoft first introduced SQL Server containers on Linux with the release SQL Server 2017. SQL Server 2019 expanded these capabilities and makes it even easier to adopt SQL Server in containers by enabling: new high availability (HA) scenarios;

Linux-based container images on Microsoft Container Registry; Red Hat-Certified Container Images; and the SQL Server operator for Kubernetes, which makes it easy to deploy Always-on Availability Groups.

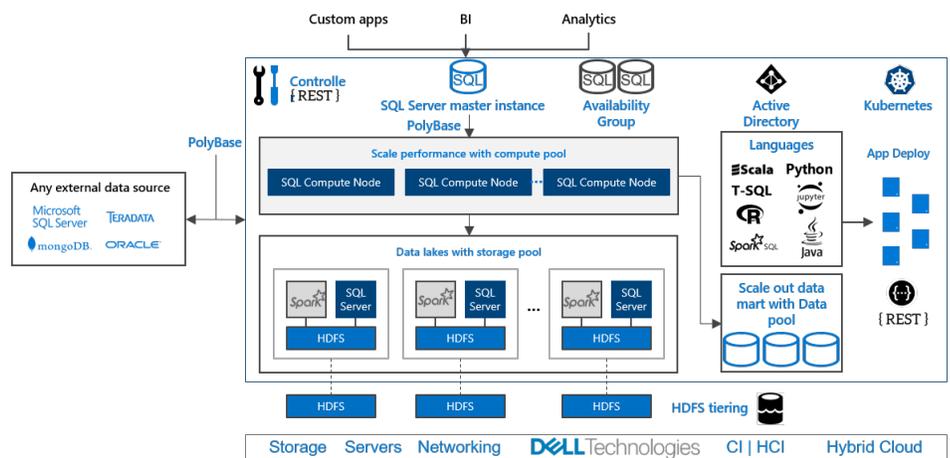
Companies can configure a SQL Server instance on Kubernetes with persistent storage for high availability. Also, Kubernetes orchestrates instances of SQL Server in container images that participate in a SQL Server Always On Availability Group providing improved health monitoring, faster recovery, rolling upgrades, offload backup, and read scale out.

Data Virtualization via PolyBase

PolyBase is a technology that accesses and combines both non-relational and relational data, residing in different locations, all from within SQL Server. It enables applications and users to query a variety of datastores including those supporting open database connectivity, NoSQL, relational databases and big data stores in Hadoop Distributed File System (HDFS)-compatible distributions and file systems.

Big Data Clusters

The addition of support for Linux, containerization and data virtualization via PolyBase, laid the foundation for SQL Server Big Data Clusters (BDC). SQL Server BDC allows you to deploy scalable clusters of SQL Server, Spark, and HDFS containers running on Kubernetes. These components run side by side to enabling you to read, write, and process big data from Transact-SQL or Spark. SQL Server BDC allows you to easily combine and analyze your high-value relational data with high-volume big data.



SQL Server 2019 Big Data Clusters combine SQL Server, PolyBase, HDFS, Spark, and Machine Learning in one solution and is supported on a Dell EMC infrastructure.

Machine Learning

SQL Server 2019 integrates additional services such as machine learning. Machine learning is the capability to use data in developing models. Incorporating predictive analytics into applications enables businesses to predict outcomes against new data. The closer the data is to the application the faster the analytics can be run, so you can make better decisions—quickly. For example, using machine learning, a predictive model can be created to determine the right time to perform maintenance on a car engine or train tracks to avoid failures. It can also be used to predict future business revenue based upon consumer buying habits, historical trends or even to determine proper staffing so the company can ensure most efficient coverage of staff to capture maximum profit. Machine Learning Services enables developers to use Java, R, and Python code in SQL Server and provides a pathway for data scientists to use Spark and HDFS tools to develop a multitude of models.

The right foundation for Microsoft SQL Server 2019 BDC

Scale-out to manage Big Data

SQL Server and similar databases were designed primarily for online transaction processing (OLTP) and are scale-up systems. In a scale-up system, performance benefits come from adding more compute and memory resources in the host server or migrating to a larger server. In contrast, a scale-out database system is designed to use multiple networked servers with storage to distribute data and data processing across a cluster. Scale-out systems are designed to manage Big Data challenges that do not work well through the traditional scale-up systems approach.

With SQL Server 2019, Microsoft offers an option for hosting scale-out database services. SQL Server 2019 BDC is a scale-out Big Data solution that combines SQL Server, Spark, and HDFS across a cluster of servers. In Big Data Clusters,

the compute pool processes the query across the SQL Server instances. This parallelization enables faster reading of large datasets, thus saving time in returning the results.

When you have a unified data platform with scale-out data services, you need a scale-out storage solution designed to support the platform. Dell Technologies offers a range of all-flash array and software-defined storage solutions designed to scale-up and out based on your workload needs. Performance, such as IOPS and transactions per second, scale linearly as the environment scales.

Storage designed for performance and parallelism

New storage technologies, like non-volatile memory express (NVMe) deliver new levels of performance and parallelism, paving the way for Big Data Clusters. NVMe drives were designed to overcome the bottlenecking that occurs when fast flash-based storage collides with legacy data transport technologies. Dell Technologies storage infrastructure offerings, such as Dell EMC PowerStore and PowerMax, use NVMe to take full advantage of the tremendous speed and low latency of this next generation media, with greater device bandwidth and queue depth. In addition, because NVMe maximizes the power of flash drives, it opens the door to the next media disruption with storage class memory (SCM). SCM provides performance and endurance beyond those of flash and approaching the speed of DRAM.

CSI plugins: Integration for better orchestration, automation and management of container storage

Container engines such as Docker and orchestration frameworks such as Kubernetes, provide a standardized way to package applications — including the code, runtime and libraries — and to run them in a consistent manner across the entire software development life cycle. The CSI is a standard for exposing arbitrary block and file storage systems to containerized workloads on Container Orchestration Systems (COs) like Kubernetes.

Dell Technologies storage solutions, including PowerStore and PowerFlex, provide unique CSI Plugins which allow customers to deliver persistent storage for container-based applications for both development and production scale. The combination of Docker, the Kubernetes orchestration system and Dell EMC storage CSI plugins enables simplified provisioning of containers and persistent storage.

Virtualize containers for greater security

Data center architects who have standardized on VMware virtualization can benefit from hosting containers on virtual machines. You can virtualize containers on VMware vSphere virtual machines (VMs) to securely isolate a Big Data Cluster instance on a Dell Technologies shared infrastructure. Employing VMware vSphere VM security can prevent accidental access to resources by anyone outside the respective project teams.

Virtualizing containers provides additional isolation and consolidation benefits. For example, a bare-metal implementation with one host operating system would force multiple projects to use the same stack: operating system, Docker, Kubernetes, and Dell EMC CSI plug-in. However, virtualizing containers on a Dell EMC hyperconverged infrastructure enables you to use multiple host operating systems on one server.

Dell Technologies' modern infrastructure solutions for SQL Server BDC

Dell Technologies designs and validates solutions for Microsoft SQL Server to help organizations jump-start their application development transformation with a modern infrastructure. This includes realizing the advantages of using Microsoft SQL Server 2019 Big Data Cluster hosted on a modern Dell Technologies infrastructure as a scalable data management and analytics platform.

To learn more about these solutions, read our white papers:

- [Microsoft SQL Server 2019 Big Data Clusters on VxRail](#)
- [Microsoft SQL Server 2019 Big Data Clusters on Dell EMC PowerFlex](#)
- [Dell EMC PowerStore: Microsoft SQL Server 2019 Big Data Clusters](#)

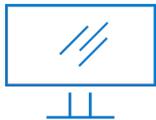
Key Component	Capabilities
<p data-bbox="245 348 347 380">Storage</p>  <p data-bbox="228 506 367 531">PowerStore</p>  <p data-bbox="191 831 402 856">PowerFlex Family</p>	<p data-bbox="513 260 1495 506">Dell EMC PowerStore is designed to support any workload by delivering multi-protocol Block, File, and now vVOL storage in a performance optimized appliance that supports end-to-end NVMe and can scale up and out when demands increase. Administrators can also access the hypervisor to deploy applications directly on the PowerStore appliance, using the same VMware tools and methods they use with external hosts. This game-changing capability, known as AppsON, is ideal for data-intensive workloads in core or edge locations where infrastructure simplicity and density is required,</p> <p data-bbox="513 522 1463 611">Comprehensive integration with open management frameworks, containerization platforms, DevOps platforms and virtualization enable PowerStore to seamlessly support the demands of Microsoft SQL Server Big Data Clusters.</p> <hr/> <p data-bbox="513 646 1487 827">The PowerFlex family creates a server-based SAN (2-layer or single layer (HCI) architecture) by combining storage virtualization software, known as PowerFlex. PowerFlex is a scale-out block storage service designed to deliver flexibility, elasticity, and simplicity with predictable high performance and resiliency at scale. Local storage resources are combined to create a virtual pool of block storage with varying performance tiers.</p> <p data-bbox="513 844 1492 995">The PowerFlex family provides enterprise-grade data protection, multi-tenant capabilities, and add-on enterprise features such as QoS, thin provisioning, and snapshots. PowerFlex offers true block storage as a service, making it an excellent complement to Kubernetes for stateful applications, such as databases, continuous integration, logging and monitoring platforms.</p>
<p data-bbox="191 1136 402 1167">Hyperconverged</p>  <p data-bbox="256 1272 337 1297">VxRail</p>	<p data-bbox="513 1031 1495 1182">Dell EMC VxRail hyperconverged infrastructure offers the performance, capacity, and graphics capabilities that are needed to meet the infrastructure requirements of a small or mid-size enterprise. VxRail provides a simple, cost-effective solution that solves virtualization infrastructure challenges and supports a wide range of applications and workloads.</p> <p data-bbox="513 1199 1503 1409">Developers and data scientists can easily support a SQL Server Big Data Cluster ecosystem with VxRail, VMware vSphere, Docker, and Kubernetes. VxRail allows customers to start small and grow by scaling up capacity and performance. The vSphere CSI driver with VxRail enables comprehensive automation, orchestration, and coordination of server and storage systems for implementing a software-defined infrastructure. When combined with its ability to seamlessly provision persistent storage, VxRail systems able to handle big data workloads like SQL Server 2019.</p>

Key Component	Capabilities
<p data-bbox="256 302 342 327" style="text-align: center;">Server</p>  <p data-bbox="185 478 412 506" style="text-align: center;">PowerEdge Family</p>	<p data-bbox="516 260 1479 348">Dell EMC PowerEdge servers provide a scalable business architecture, intelligent automation and integrated security for your high value data management and analytics workloads.</p> <p data-bbox="516 369 1490 548">Dell Technologies offers one of the broadest selections of servers enabling customers to configure their compute to match business requirements. The vast configuration choices in using PowerEdge server means you can optimize per core licensing for containers, Docker's Enterprise Edition as an example. The key to getting the greatest return on your containerized environment is consolidation that maximizes the efficiency of CPU utilization.</p>

An agile IT foundation for the new era of data management

In the data-driven age, information is as valuable as currency. Harnessing more data from more sources to provide actionable insights quickly and cost-effectively is the key to business innovation. It is clear that running traditional data warehouses and databases on isolated infrastructure is an untenable IT model in the era of petabyte data stores and real-time analytics.

Microsoft SQL Server BDC enables the union of large data initiatives and easy access to existing data sources in your datasphere but to truly exploit the benefits of this unified data platform, you need an equally powerful foundation. Dell Technologies provides foundational solutions that allow you to accelerate business innovation and achieve competitive advantage. From ingestion and prep of SQL relational data into data lakes, to artificial intelligence and machine learning models, this streamlining of architecture lets you take advantage of next-generation analytics and go behind your data to draw valuable insights.



[Learn More](#) about
Dell EMC Ready
Solutions for Microsoft



[Contact](#) a Dell Technologies Expert

¹. Source 1: Forbes, "How much data do we create every day? The mind blowing stats everyone should read", May 2018