DELL EMC

Whitepaper

Winning in the New Era of Data Management

How Microsoft SQL Server 2019 will change the way you manage your data estate

Summary

In the data-driven age, information is as valuable as currency. Entire industries are being disrupted by new data sources, emerging technologies and tech-native startups. Harnessing more data from more sources to provide actionable insights more quickly and cost-effectively is the key to outpacing the competition.

As organizations evaluate IT strategies for harnessing data, it's 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.

The competitive advantage will go to organizations that can virtualize data so users can quickly and easily access data across all platforms and environments to make better decisions in real time — without the time and cost burden of traditional extract, transform, load (ETL) processes. Key stakeholder imperatives include:

- IT must embrace a new paradigm for managing data virtually, from the edge to the core, and from the data center to the cloud.
- Business and IT leaders will need to find a way forward that preserves legacy IT investments while setting a foundation for the future.

Traditional relational databases such as Microsoft® SQL Server® are evolving to act as a virtualized data hub, allowing users to access and manage unstructured and structured data from across the data estate using a single, unified interface.

This paper explores the ramifications of data gravity, the possibilities of SQL Server 2019 data virtualization and outlines considerations for building an agile IT foundation for the new era of data management.

October 2019

Table of contents

Data management in the data-driven era

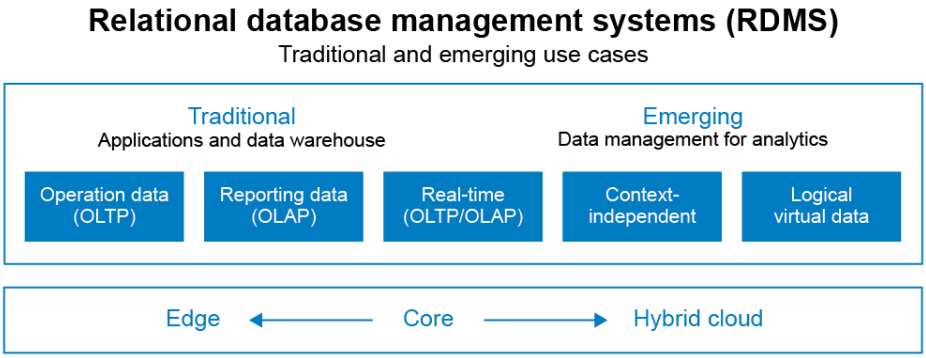
The evolving role of relational database management platforms

The data-driven age is here. Almost every industry is being disrupted by new data sources and emerging technologies that promise to harness them, like artificial intelligence (AI), machine learning (ML) and deep learning (DL). Tech-native startups are disrupting entire industries as well with the ability to bypass decades of legacy investments and take real-time data capabilities straight to consumers. With information being as valuable as currency, the advantage goes to organizations that can harness more data from more sources to provide actionable insights more quickly and cost-effectively than the competition.

Savvy IT leaders know that this new era requires a modern approach to data management. One that replaces monolithic, siloed databases with new models for using and sharing data across the organization.

In this new era of data management, SQL Server 2019 will be a game changer. But to understand the impact of Microsoft SQL Server 2019 and its implications on IT strategies, it's important to understand the context for how traditional relational databases have evolved for this new era.

Traditional relational databases have long been the backbone for operational data (online transaction processing [OLTP]) and traditional data warehouses (online analytics processing [OLAP]), supporting reporting and some analytics. However, as organizations look for ways to unlock the insights hidden in their data — across different platforms and environments — database management solutions have evolved to include additional analytics.



This critical change in database management strategy by large vendors has driven Gartner to define a new category, called "Data Management Solutions for Analytics" (DMSA), in which Microsoft has been recognized as a leader.[1] Below are Gartner-defined use cases for working with distributed data.[2]

- **Traditional data warehouse.** This use case involves managing structured historical data coming from multiple sources. Data is mainly loaded through bulk and batch loading. The traditional data warehouse can manage large volumes of data and is primarily used for standard reporting and dashboarding. To a lesser extent, it is also used for free-form ad hoc querying and mining, or operational queries.
- **Real-time data warehouse.** This use case adds a real-time component to analytics use cases, with a goal of reducing latency between when data is generated and when it can be analyzed. This use case primarily manages structured data that is loaded continuously via micro batching and/or streaming ingest in support of real-time decision support, embedded analytics in applications, real-time data warehousing and operational data stores.
- **Logical data warehouse.** This use case manages data variety and volume of data for both structured and other content data types, where the DMSA acts as a logical tier to a variety of data sources. Besides structured data coming from transactional applications, this use case includes other content data types such as machine data, text documents, images and videos.

[1] Microsoft, "Gartner names Microsoft a leader in the Magic Quadrant for Data Management Solutions for Analytics (DMSA)", January 2019.

[2] Gartner, "Magic Quadrant for Data Management Solutions for Analytics," January 2019.

- **Context-independent.** This use case allows exploration of new data values, data form variants and relationships. It supports search, graph and other capabilities to uncover new information models. This use case is primarily used for free-form queries to support forecasting, predictive modeling or other mining styles, as well as for queries supporting multiple data types and sources.

Designing and building the best solution for your data estate — from edge to core, and from on-premises to hybrid cloud — requires a combination of the right expertise and infrastructure. Not a one-size-fits-all approach like those of the past, but a consumable, flexible and scalable approach that builds on current investments while preparing you for a successful digital future.

In this paper we will explore the SQL Server journey in concert with Dell EMC's vision for managing your data capital as part of your overall data estate. Dell Technologies is uniquely positioned to deliver edge to core to hybrid cloud solutions to accelerate, protect and reuse the tooling that Microsoft SQL 2019 delivers.

Data has gravity

Data mass creates centers of gravity that pull in apps and services

It's no secret that data is big — and getting bigger. By some estimates, we are creating 2.5 quintillion bytes of data every day.[3] As cloud adoption soars, mobile devices proliferate and internet of things (IoT) expands, data is amassing in large quantities outside the confines of the traditional data center.

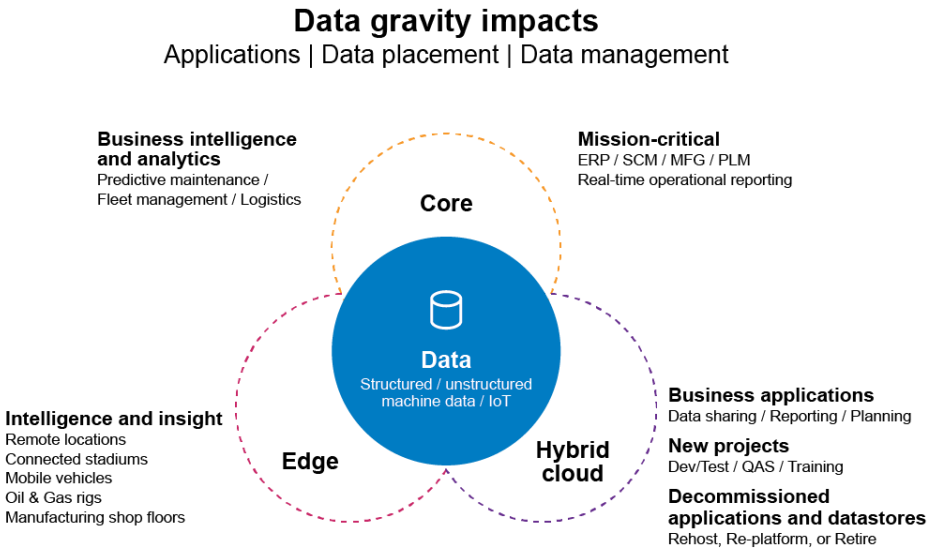
Data has become so big and so distributed that where the data is created and where it is used have become critical considerations for speeding workloads and decision making. The larger the amount of data, the harder it is to move. Therefore applications, services and even other data will be attracted to that data and orbit it as though it were a center of gravity. Just as gathering clouds of stardust once condensed into stars and planets that exert a gravitational force on surrounding objects, data is gaining mass and pulling applications and services toward it.

By putting applications and services in close proximity to the data, you can reduce latency and improve performance. Analytics run faster, so you can make better decisions more quickly. Response times are shorter, improving the customer experience. To put it succinctly, data needs to be ingested, aggregated, analyzed and acted upon where it resides.

For example, an autonomous vehicle needs to take in multiple points of data—from cameras, radar and sonar—and process them in real time. If an object moves into the car's path, there's no time to send the data to the cloud or back to the data center and wait for a response. The analytics have to come to the data at the edge. If the data is streaming to and from the cloud — say, for mobile services — then the apps and analytics need to live in the cloud so they can process the data in real time. By the same token, the data that fuels applications for mobile and remote users can safely stay in the data center, enhancing protection for privacy-sensitive personal and financial information.

As modern businesses adapt to the digital economy, IT needs to be prepared with a data management strategy, enabling the business to make use of data that lives in distributed hybrid environments from edge to core to cloud.

[3] Forbes, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," May 2018.



By adopting an IT strategy that respects the principles of data gravity, you can leverage your data no matter where it resides. That's where data virtualization comes into play, and it is set to change the world as we know it.

Data virtualization

Democratizing data enables comprehensive analytics

Data virtualization refers to abstracting data from different sources, locations and formats — without copying it or moving it — into a single layer that allows users to query it in real time from a single, unified interface. This also allows sensitive data to be controlled from a single location. By eliminating the need to create multiple copies of data, the cost of storing and managing it are likewise minimized.

Some of the key drivers of data virtualization are the challenges posed by traditional data warehouses used for business intelligence (BI). At the core, it's a requirement to copy and load data into the data warehouse reporting platform. Historically, ETL processing — or data pipelining — has been employed for data transformation. This process has introduced challenges that can't be ignored as businesses strive to become more responsive to customers, partners and market dynamics in real time. Some of the larger challenges include:

- **Data latency.** ETL pipelines introduce an inherent delay. A recent study found that more than 80% of data sets delivered by ETL pipelines are between two and seven days old by the time they reach an analytics system. A full 75% of businesses reported that delays in data processing had inhibited business opportunities.[3]
- **Big data and analytics.** With the expansion of IoT, connected devices and people are generating volumes of data that exceed the storage capacity of any traditional database system. By some estimates, 90% of all the data in the world was generated in the last two years at a rate of 2.5x1018 bytes of data per day.[4] This new type of data is often in formats that are not suitable for storing in relational database tables or for querying using relational query semantics.
- **Big data needs "context."** Many IoT data sets are unstructured and overwhelming in volume, velocity and variety. And without business context, have no value. IT needs to build the bridge for business/operational data stored in relational databases to come together with IoT for true real-time BI and response that leads to self-learning intelligent applications and business processes leveraging AI and ML.

[4] Microsoft technical white paper, Microsoft SQL Server 2019 Big Data Clusters, September 2018.

Unified data management

Real-time data access and visibility is increasingly critical for businesses to maintain a competitive edge. At the same time, new applications and new use cases — such as those enabled by AI and IoT — are surpassing the limits of traditional siloed data centers and processes like ETL. In fact, IT is now switching the “T” and the “L” to speed data movement, streamlining data into the target and allowing the target system to perform the translation. Transformation is now being driven by the data platform where the data sits.

With increasing adoption of data virtualization comes the need for data management platforms and infrastructure that can manage this virtual data layer with a unified view that bridges multiple applications, users, data stores and locations. SQL Server 2019 is a database management system that is continuing to extend its capabilities as a platform for data unification.

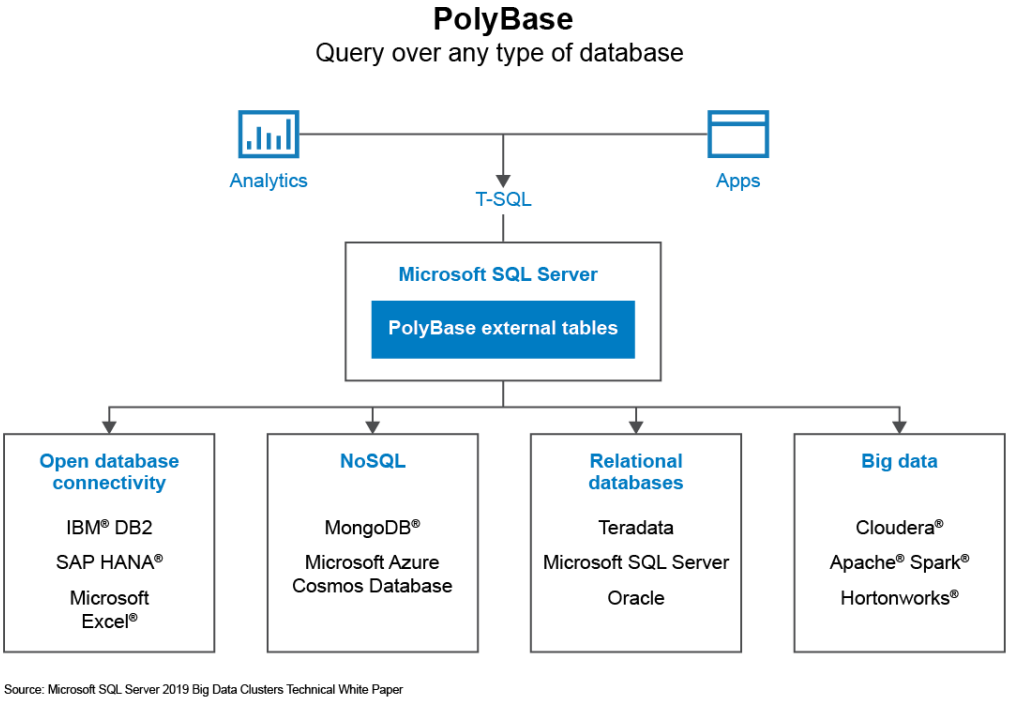
SQL Server 2019 is a hub for data, with the ability to query over data in Oracle®, Teradata and Hadoop® in a secure and highly performant way — all without moving or copying it. The key capabilities of SQL Server 2019 for unifying virtual business data and operational data stored in relational databases with IoT for true real-time BI and embedded AI and ML include:

Polybase: query over any data type

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.

PolyBase pushes some computations down to optimize performance. It pushes operations such as projections, predicates, aggregates, limits and homogeneous joins to the source system so it can take advantage of the query optimizer in each of the source systems. Only filtered results are returned, reducing the amount of data to transfer and thus increasing performance.

SQL Server 2019 extends the capabilities of PolyBase with new connectors to create external tables that link to a variety of data stores, including SQL Server, Azure® SQL Database, Azure SQL Data Warehouse, Oracle, Teradata, MongoDB, Azure Cosmos DB or any open database connectivity (ODBC)-compliant data source via a generic ODBC driver.



Source: Microsoft SQL Server 2019 Big Data Clusters Technical White Paper

When teams can easily consume data from multiple sources, regardless of location and type, you've effectively removed the boundaries that inhibit comprehensive data analysis and expedited decision making and value creation throughout your organization.

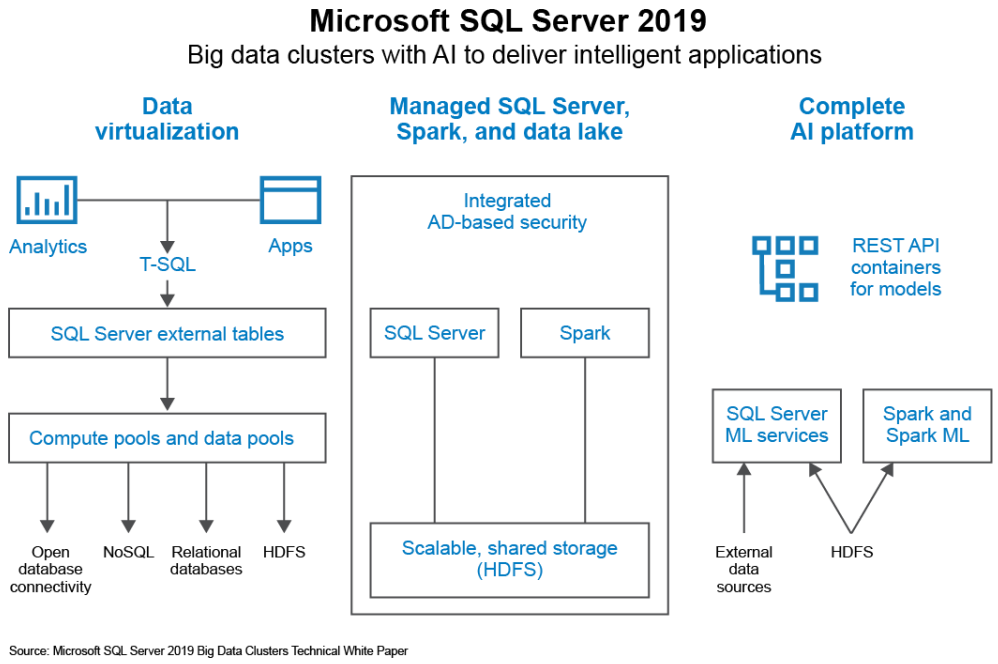
Additionally, in many instances, developers can keep T-SQL procedure result set attributes returned to the applications and analytics consistent, even though the data is being returned from a new source via PolyBase external tables.

Big data clusters: scale and diversify compute and storage

SQL Server 2019 big data clusters leverage PolyBase to improve the data virtualization experience by enabling faster, more secure integration between SQL Server and Apache HDFS and Spark systems.

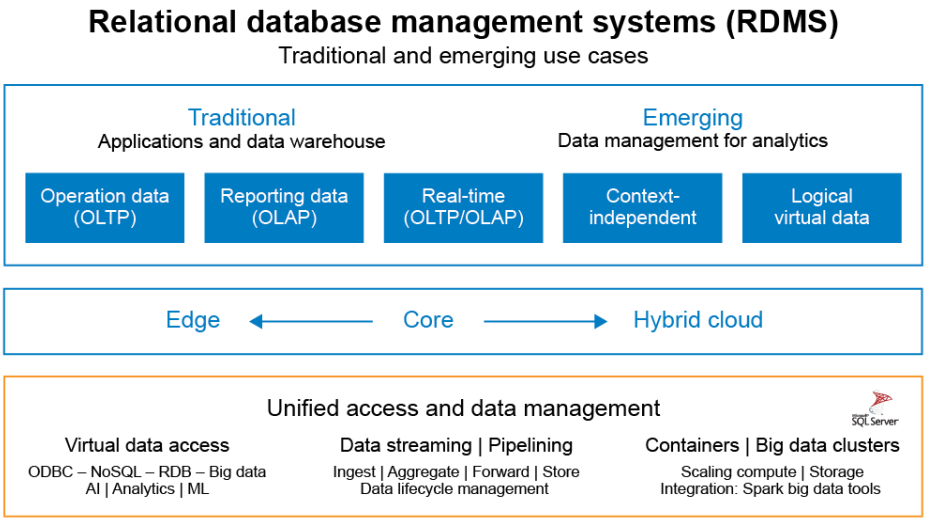
SQL Server big data clusters integrate SQL Server and industry-standard big data tools into a single package that's easy to deploy and manage. They provide three major functions:

- **Data virtualization** enables combining data from multiple sources without moving or replicating it. You can boost performance by scaling out compute and caching.
- **Managed SQL Server, Spark and data lake** allow you to store high volume data in a data lake and access it via SQL or Spark. This includes management services and integrated security to simplify management.
- **Complete AI platform.** Integrated data from many sources can be easily fed into model training; you can ingest and prepare data, then train, store and operationalize models in a single system.



Source: Microsoft SQL Server 2019 Big Data Clusters Technical White Paper

The support of Docker containers, Linux® and PolyBase in SQL Server set the stage for SQL Server 2019 big data clusters. Big data clusters fully embrace containers on Kubernetes to deploy applications — ensuring a predictable, fast and elastically-scalable deployment, regardless of where the Kubernetes cluster is deployed. As organizations look to utilize containers for stateful application use cases, the need for data persistency, protection and management will play a more integral role. They will need to consider storage solutions that deliver the data persistency and services required to support these platforms.



Data virtualization has powerful implications. Making data accessible to more than just the most highly skilled data scientists means it can now be included in multiple reports, dashboards and applications — multiplying its value.

It's clear that increasing data virtualization and embracing new capabilities for a unified data management platform require a new strategy for data services and infrastructure that complement SQL Server 2019. The next section lays out the Dell EMC strategy for data services and infrastructure that complement SQL Server 2019.

Emerging technologies

Data services and infrastructure

It may feel like you've been here before. The IT landscape has changed significantly, and now you need to pivot again to stay competitive. In the past, that meant ripping out and replacing relatively recent investments. That's not an ideal scenario to say the least.

This time it's different. SQL Server 2019 is designed to integrate the data estate, enabling you to evolve from living in a this "or" that world to developing a more strategic approach of this "and" that. You can leverage this unified layer for virtualized data across on-premises and cloud on traditional three-tier architectures leveraging all-flash arrays and software-defined storage and converged infrastructures and hyper-converged infrastructures … and so on.

Of course, you'll want to consider deploying some of the latest infrastructure technologies to unlock all the performance and agility capabilities within SQL Server 2019 and exploit working with virtual data. Emerging infrastructure technologies will help you:

- **Accelerate** and future-proof your environment while completely modernizing your SQL infrastructure. A revamped perspective on storage will leverage storage-class memory, persistent memory, such as NVDIMM, and other memory technologies to maximum effect.
- **Protect** your databases with industry-leading backups, replication, snapshots, resiliency and self-service deduplicated copies.
- **Reuse** snapshots for operational recovery, dev/test repurposing and continuous integration/continuous delivery (CI/CD) pipelines. For example, you could take a storage level copy of a database from production and scrub the data so it can become a source for a PolyBase external tables, without impacting the product database.

Aligning along these pillars will bring efficiency and consistency to a unified approach of data estate management. The combination of a strong, consistent, high-performance architecture supporting the database platform will make your IT team the modernization execution masters.

Let's take a moment to explore emerging infrastructure technologies (compute, storage, networks) and use cases for enabling data management with SQL Server 2019.

Compute level

Beginning with the compute level, real-time in-memory data processing leveraging DRAM is front and center for accelerating performance-dependent uses case tied to Microsoft SQL. Take, for example, real-time data analytics and context-independent data analysis, SQL Server is leveraging in-memory processing to:

- Accelerate real-time analytics and reporting by optimizing performance of OLTP and aggregation of OLAP data for BI including data ingestion, data load and transient data scenarios;
- Support the exploration of new data values, data form variants and relationships to perform free-form queries to support forecasting, predictive modeling or other mining styles, as well as for queries supporting multiple data types and sources; and
- Processing larger data sets of data for both structures and other content types such as machine data, text documents, images and videos.

Real-time in-memory data processing

Relational databases such as SQL Server have adopted in-memory processing for OLTP and OLAP to optimize performance of transaction processing, data ingestion, data load and transient data scenarios.

While in-memory processing is not new to SQL Server, database sizes are growing and increasingly being measured in multiple terabytes. The good news is that servers, such as two- and four-socket Dell EMC servers powered with Intel® processors can start small and grow to 6TB of RAM without disruption, delivering scalability as SQL Server deployments grow.

While the servers support larger memory-optimized databases for scale-up and scale-out system landscapes, data persistence traditionally is maintained at the storage layer to protect against data loss during planned and unplanned system downtime.

In 2019, Dell EMC released PowerEdge Servers with Intel Optane™ DC Persistent Memory. Intel Optane DC is an innovative memory technology that delivers a unique combination of affordable large capacity and support for data persistence. This is a new class of memory and storage that can accelerate insights from the massive amounts of data companies manage today.

Extending in-memory for larger data sets

Intel Optane DC persistent memory can be configured with traditional DRAM acting as a cache to transparently integrate into the memory subsystem, making it appear like DRAM with no changes required to the operating system (OS) or applications.

This provides a couple of benefits. First, you can double today's four-socket Dell EMC PowerEdge from 6TB DRAM systems to 12TB of usable system memory (3TB per socket).[5] This enables consolidation by reducing the number system nodes. Second, you can swap out higher-cost DRAM-only memory with lower-cost Intel Optane DC persistent memory modules.

[5] Intel, "The Challenge of Keeping Up With Data," April 2019.

Data persistency with non-volatile memory

Applications designed to run in-memory (DRAM) typically require the database tables and data to be available in-memory to execute the OLTP and OLAP processes. In the event of planned or unplanned system downtime, all data running in-memory is lost. Upon restart, the database tables and data need to be reloaded into memory before applications can restart.

From a storage performance standpoint, memory is fastest, since it is closer to the processor on the memory bus. Larger database tables stored as persistent storage on Optane DC persistent memory will be available for faster data load following an outage.

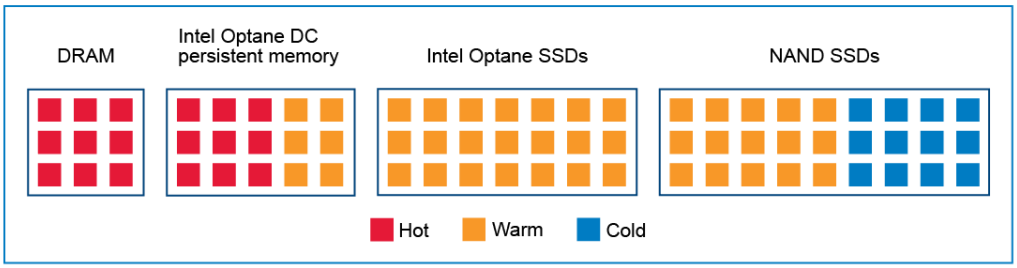
The benefit of persisting larger tables and data on Intel Optane DC persistent memory is faster data load and recovery time objectives (RTOs). For example, when testing Microsoft SQL 2019 in a preview environment and leveraging Optane DC persistent memory, Dell EMC testing showed that users can see up to a 2.7X improvement in SQL database performance.[6]

Data tiering and placement

Data usage changes over time. Initially, a set of data may be used frequently but that frequency decreases over time as the data ages. This is not a new revelation, however when you consider how many places your data is stored — edge to core to hybrid cloud — businesses should re-evaluate their data lifecycle management strategies. This includes storage strategies for tiering hot, warm and cold data and moving frozen data to the cloud, such as Microsoft Azure.

Data tiering encompasses working across various memory and storage classes, we will explore these uses case further in the next section.

Memory pools and data tiering



This figure provides for pooling and data tiering using DRAM, Intel Optane DC persistent memory and Intel Optane DC SSDs with hot data represented in red, warm data in orange, and cold data in blue.

Storage level

Data virtualization is empowering organizations with the ability to access data anywhere it resides, and storage plays a key role in helping businesses to unlock the value of their data — and maximize data capital. Attention must be paid to where the data resides, how it's managed and how it's protected. Storage has always been a driving force for the concepts of accelerate, protect and reuse — as previously discussed — and new advancements in storage technology continue to help organizations exploit working with virtual data.

To unlock the value of business data, attention must be paid to where data resides, how it's managed and how it's protected.

We are seeing companies shift from a "cloud first" approach to a "data first" mindset. This allows you to ensure that your most valuable asset is in the right place at the right time, and with the right service level agreement (SLA). Organizational needs and business models should be driving the decision on where data resides, while taking into consideration consolidation strategies that can help reduce administrative tasks and accelerate performance.

For instance, SQL Server supports both in-memory and disk-based placement for tables and data. However, where you choose to place the data is based on factors including the application profile, types of data and IT services required to support the users.

[6] StorageReview, "Dell EMC PowerEdge Servers Refreshed with New Intel CPUs and Optane DC Persistent Memory", April 2019.

While Microsoft does provide some guidance for table and data placement to facilitate IT planning, data placement typically requires a more thoughtful exercise balancing SLA requirements for performance and availability, including the data lifecycle management, versus the IT cost for delivering and managing the service and data.

NVMe, over fabrics and storage-class memory

It's important to take note of emerging storage technologies that are available as you plan your IT foundation for data management use cases and services with SQL Server 2019. For example, all-flash storage enables high performance, throughput and data replication for uses cases such as system copy/refresh for dev/test, and operational recovery along emerging practices such as continuous integration-delivery.

When it comes to performance, flash storage has already changed the game in delivering high performance with one millisecond or less latency for OLTP and OLAP workloads. For many customers, modernizing to an all-flash array means significantly more IO and bandwidth resulting in reduced management of systems, reduced calls to support desks and order of magnitude improvements in the time it takes to produce reports.

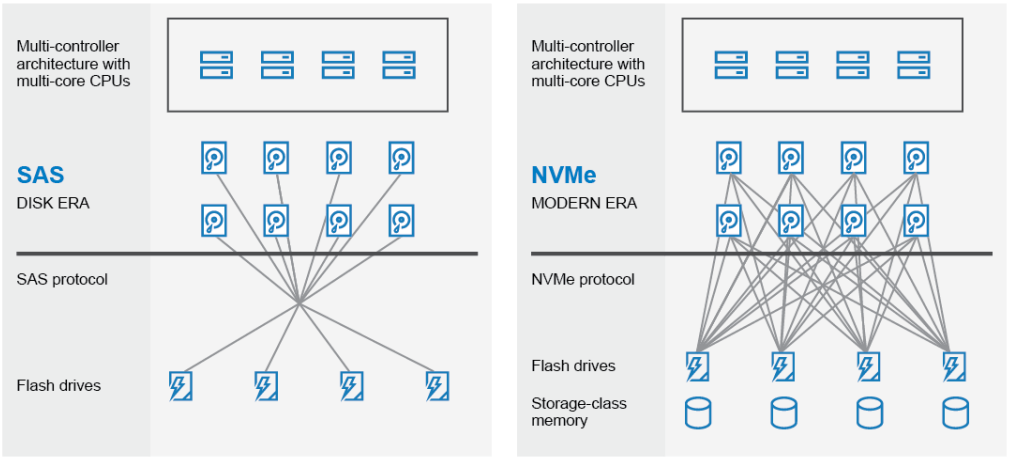
Now, with emerging BI, real-time analytics and big data uses cases with SQL Server 2019, flash is just the beginning. New technologies including non-volatile memory express (NVMe), NVMe over fabrics (NVMe-oF) and storage-class memory (SCM) are extending the value of flash SSDs, which are becoming foundational to storage platforms supporting these use cases.

Let's take moment to review how these technologies have evolved and how they are impacting performance.

NVMe has been around since 2012. NVMe was architected to take advantage of the parallelism of modern CPUs and SSDs. It provides faster data transport to the host controller system for example, 32GB/s for NVMe versus 12GB/s for SAS 3.0, providing almost three times the speed.

Maximizing the performance of a multi-controller architecture

Sub-millisecond response time to microsecond response time



NVMe breaks through the bottlenecks that occur when fast flash-based SSD storage collides with legacy data transport technologies. And, most importantly, opens the door to the next media disruption with SCM.

As an example, Dell EMC has storage built with a multi-controller architecture and end-to-end NVMe to maximize performance with modern media, including flash SSDs and SCM. You can invest in all-flash storage from Dell EMC today and non-disruptively add NVMe-OF and SCM drives tomorrow.[7]

NVMe-OF, published in 2016, takes the next step with a transport protocol that enables a host computer, such as a Dell EMC servers running Microsoft SQL Server, and a target solid-state storage device or system, such as Dell EMC storage over Fibre Channel (FC).

[7] CTT, "SATA vs SAS vs NVMe Performance Breakdown," October 2018.

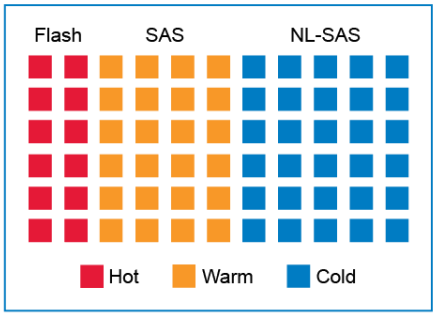
SCM, also known as persistent memory, may be the most disruptive storage technology innovation of the decade. While SCM is slightly slower than DRAM, its value is that it is persistent, meaning that, like traditional storage, its content is preserved during a power cycle.

The key point to take away is that, while running databases in-memory (DRAM) is faster than disk-based storage, storage platforms have been adopting emerging technologies to support the performance requirements for analytics and big data scenarios such as transaction processing, data ingestion, data load and transient data.

High performance for disk-based data sets

SQL Server provides flexible choices of tables and data placement, leveraging in-memory servers and flash-based storage options. This is valuable in that it allows IT the flexibility of allocating infrastructure resources to meet performance requirements with optimal total cost of ownership (TCO).

Storage pools and data tiering



Data placement and consolidation

An earlier section of this paper discussed data virtualization and the emerging capabilities of SQL Server 2019 as a unified data management platform for data residing in SQL Server and data located externally. But IT needs to consider the larger picture of data placement for the entire data estate.

From a storage perspective, IT should have a platform strategy that can support the inevitable landing spots for persistent data. This includes storing data on the right type of storage such as scale-up, scale-out, file, block, object, and long-term protection and retention strategies.

Another consideration is data consolidation. The vast majority of Microsoft SQL Server environments span many versions, which have disparate features and capabilities. Consolidation of these versions to a modern all-flash array allows for the offloading of responsibilities to the infrastructure and helps to provide a more consistent and elevated experience to all versions. The benefits that are experienced are largely based on the data services provided by the array and these are highly differentiated across the industry with respect to the effectiveness of the data services.

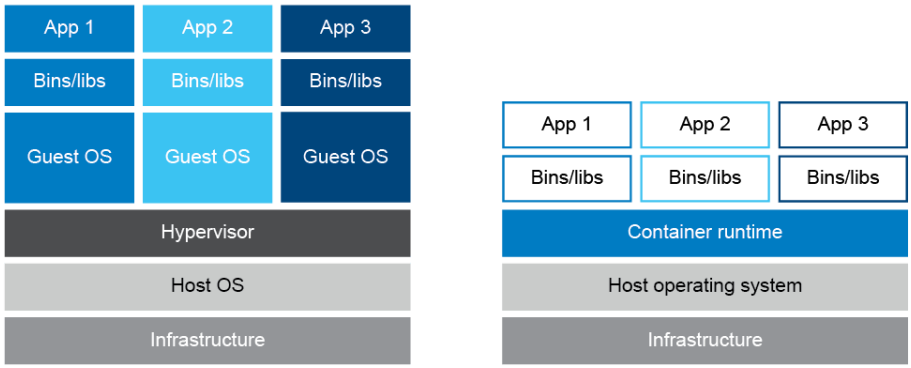
Across the Dell EMC product portfolio there are several key differentiators and variations on data services features.

- **Usable capacity:** A commonly overlooked aspect is the overhead, or lack thereof, that the system and the underlying RAID create. Dell EMC systems provide some of the highest usable capacity in the industry with some systems providing as much as 88% usable capacity.
- **Thin provisioning:** Commonplace today, this is one of the key features for maximizing the cost effectiveness of a shared storage environment. For a database administrator (DBA), this allows them to allocate the database files to the expected size in the future and not have to increase the size of it later. Space is only consumed as the data is written to the array.
- **Compression:** While this is done in different ways on the Dell EMC portfolio of storage arrays, each system typically will provide better than 2:1 compression on a database that isn't already leveraging row or page compression within Microsoft SQL Server.
- **Deduplication:** There are many reasons a DBA or application owner would need to create database copies. Test/Dev, patch testing, near real-time analytics, or just making a copy for protection these database copies will be 100% deduplicated on the array. Only the changes that are made to the copies or production will result in new data being written to the array.

- **Intelligent snapshots**: As efficient as deduplication is, an even more efficient and faster way to make a copy is to simply take a snapshot of a volume or set of volumes and then mount the snapshots to another SQL Server. Making database copies in this way ensures little to no impact to the production database and allows copies of even very large databases to be provided in under five minutes.

Storage data services for containers

Containers have reshaped the way companies think about developing, deploying and maintaining applications and software. Relational databases, including SQL Server, have embraced Docker containerization together with Kubernetes orchestration framework.



Virtual machines Containers

Containers were originally designed to be short-lived, or stateless, by nature, which makes them ideal for Test/Dev purposes. However, containers can also be valuable for production SQL Server database environments. Kubernetes and other container orchestraters are extending their reach to long-running processes, especially in the big data and analytics space. Gartner predicts that by 2022, more than 75% of global organizations will be running containerized applications in production.[8]

Application developers most often work outside of the server environments their programs need to run in. To minimize conflicts in library versions, dependencies and configuration settings, the production environment needs to be recreated multiple times for development, testing and pre-production integration.

The dilemma, when it comes to building or updating databases like SQL Server, is that the data needs to be persistent, and must survive through the restart, rescheduling or deletion of a container. When containers are rescheduled, the storage should also be shifted and made available on a new host for the container to start without incident.

To effectively address the challenges of stateless containers and the need for persistent storage and replication software, Dell EMC supports a Kubernetes Containerized Storage Interface (CSI) plug-in. This integration makes it possible for developers, DBAs and storage administrators to utilize interfaces they are most comfortable with — such as command line interface (CLI) and the Kubernetes dashboard — to provision and manage persistent storage tiering aligned with Dell EMC storage data services.

The adoption of containers and the benefit they specifically provide SQL Server, will continue to grow. Organizations that embrace containerized database workloads will be able to greatly simplify database development cycles, along with production deployments, upgrades and high-availability orchestration. One key tenant of a scalable containerized cluster workload is the SQL Server 2019 big data cluster capability.

[8] Gartner, "6 Best Practices for Creating a Container Platform Strategy." April 2019.

Network level

Today's applications and databases call for new thinking when it comes to the network architecture. The underlying network performance and utilization can adversely impact database acceleration and scalability. For example, application throughput can be reduced if you are hitting network bandwidth limits.

To support a data estate characterized by data gravity and virtualized data, IT organizations will need to ensure network platforms and topologies are ready for the increasing volume, variety and velocity of data movement from the edge, to the core, to the cloud and back.

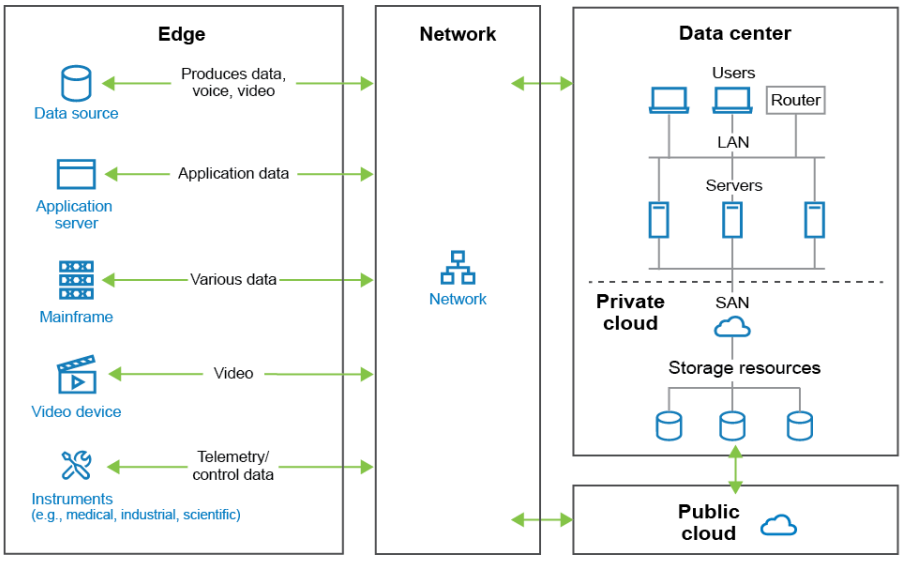
Open networking technologies

In preparing to support a data estate characterized by data gravity and virtualized data, IT organizations will need to ensure network platforms and topologies are ready for the increasing volume, variety and velocity of data movement from the edge, to the core, to the cloud and back again.

Open networking technologies, such as software-defined networking (SDN), enable organizations deploy a network that is more programmable, flexible and automated, which in turn:

- Removes network bottlenecks
- Makes adopting a cloud service delivery model easier
- Speeds application deployments
- Improves virtual machine (VM) recovery times to minimize application downtime
- Enables IT staff to focus more on strategic-level IT advancement
- Delivers higher levels of scalability, allowing the network to meet traffic demands of workloads as they grow and fluctuate over time

According to Enterprise Strategy Group (ESG), "organizations preparing for Internet of Things (IoT) initiatives and connecting to the cloud will have to explore a software-defined approach to networking. SDN helps meet high-bandwidth needs and accommodates the evolving traffic patterns of a modern digital business".[9]



[9] ESG, "The Network's Foundational Role in IT Transformation," May 2018.

There are also several areas to consider, including:

Data center network

With the emergence of new use cases encompassing data pipelining of IoT, machine and big data, it's imperative to ensure that network connections have enough bandwidth and performance to handle peak loads and the increasing amount of east-west traffic patterns.

Storage network

Typically, to optimize performance and support business continuity and disaster recovery, large organizations have deployed FC or IP-based storage area networks (SANs) to leverage shared enterprise storage.

As flash storage technology continues to advance, the SQL Server environment will take advantage of end-to-end low-latency fabrics leveraging NVMe.

Wide area network (WAN)

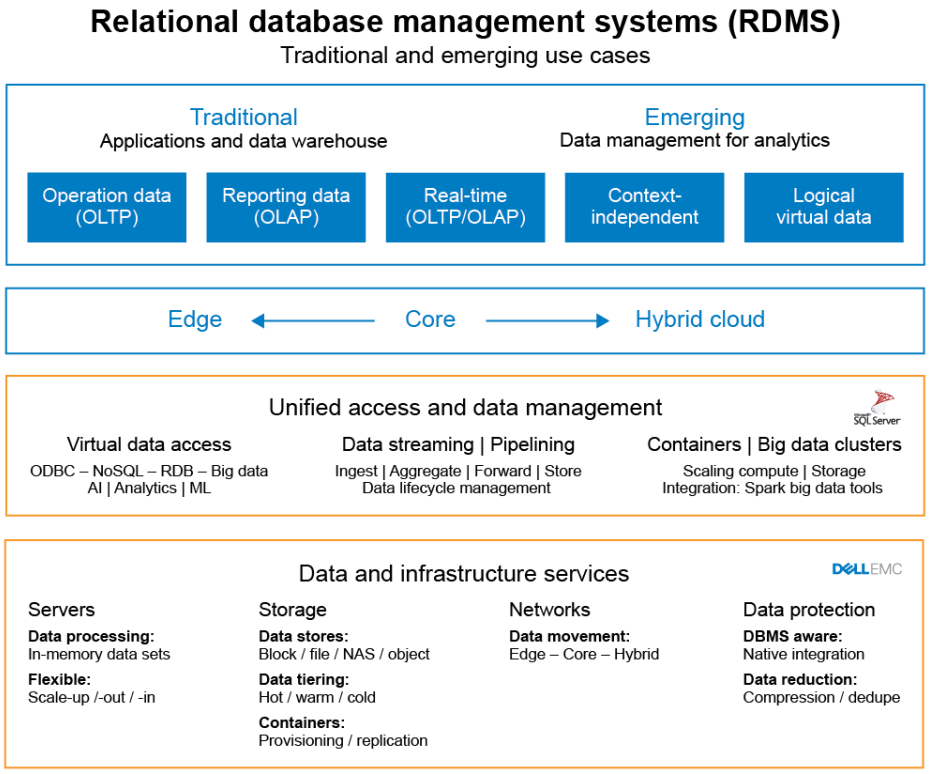
Replication for WAN is a given, but consider the increasing role of WAN as SQL Server environments adapt to distributed environments; the WAN will play a more important role in collecting data from both cloud and edge environments. As more organizations look to move toward a hybrid cloud model, they will need an infrastructure connected to an ecosystem that enables a path to the cloud and investment protection for the next hardware that they decide to purchase. This way, when they are ready to move the workload to the cloud, the connection is already in place. SDN in a wide area network (SD-WAN) technologies allow companies to modernize how branch and remote offices connect to the data they need. These organizations can radically reduce costs by moving away from expensive dedicated lines (MPLS) toward lower-cost broadband connectivity, or a mix of multiple connection options.

The network, whether in the data center or in access and edge networks, is a critical starting point and foundation to digital transformation and supporting the next evolution in data management solutions, such as SQL Server 2019. Dell EMC's vision[10] for the network is based on a disaggregated model offering an open ecosystem in which organizations can pick and choose from a wide range of innovative, industry-standard network applications, OSes and hardware platforms. This approach gives customers maximum choice and control over the technologies they select and the architectures they adopt, resulting in measurable cost-savings and increases in service agility.

Summary

Now that we have reviewed a framework, as illustrated by the image below, for emerging data management solutions for virtualized data, including aspects of SQL Server 2019 capabilities and examples of complementary data services and infrastructure use cases enabled by Dell EMC, in the next section we will introduce considerations in planning your journey.

[10] For more information on Dell EMC Open Networking refer to the Dell EMC Networking Overview.



What to take away

We are in the midst of a data management paradigm shift. All the IT "truths" we have come to know over the last several decades are changing — and in a good way. Foundationally, the way we collectively look at data and what we do with that data is shifting to a more nimble and accessible future state. Savvy IT leaders know that this new era requires a new approach to data management. One that replaces monolithic, siloed databases with new models for using and sharing data across the organization. Ensuring your planning is actively addressing this journey will allow you to make decisions that continue to enable that progress. These decisions will become additive and interactive versus standalone siloed builds that become islands of neglected technology.

The new truth is:

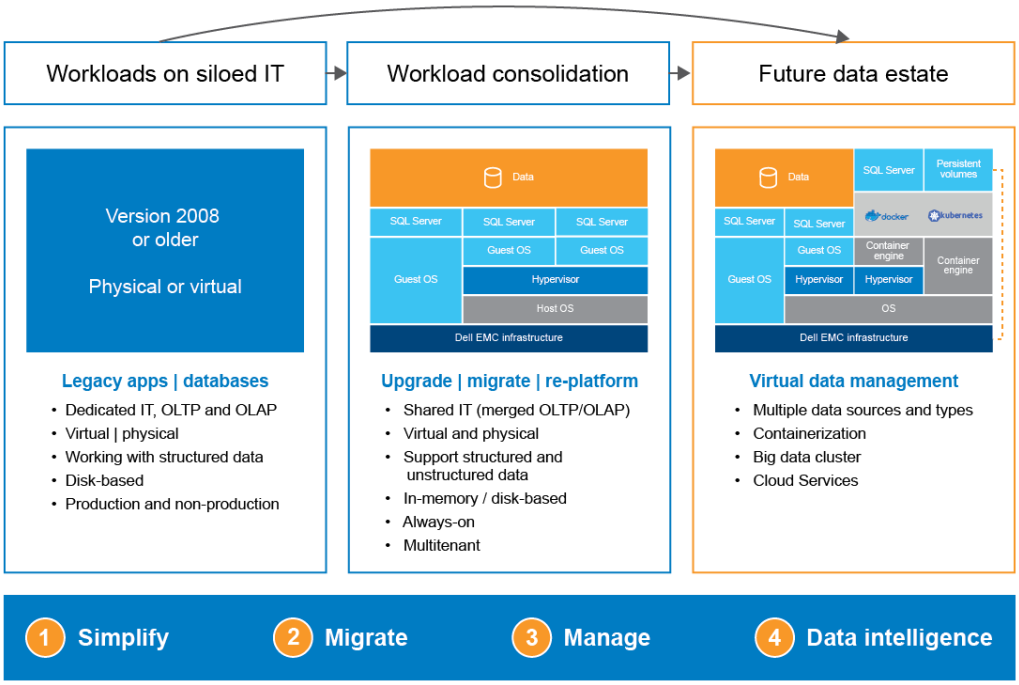
- Data has gravity and applications and analytics orbit around that data, wherever it resides.
- Virtualizing data creates a totally new paradigm where data is democratized and available to all, without the need for time-consuming, complex ETL and ever-increasing amounts of storage.
- The forklift mentality is at an end. It's no longer an "or" story, but an "and" story. You have the freedom to do what makes sense for your business.

Relational databases, such as SQL Server 2019, have emerged with architectures and tools designed to bring business and operational data together with AI, ML and DL. This evolution into unified data management platforms enables organizations to harness more data from more sources, providing actionable insights more quickly and cost-effectively than the competition.

Where will your organization start?

Digital transformation enabled by IT transformation is a journey. While everyone's path to the future data estate is unique, as illustrated below, most share some common objectives. These include sustaining older versions of SQL Server deployments (including the applications) while migrating to SQL Server 2019 and ultimately enabling virtual data access and management leveraging containers, big data clusters and hybrid cloud services.

Planning your journey



Let's look at some of these objectives:

- The typical business today is running hundreds of applications and supporting multiple databases and mixed workloads. Consolidation strategies must take into consideration ensuring the availability and performance of business-critical applications while maintaining low latency with fewer resources. The introduction of faster more powerful CPUs and new storage technologies is making it possible for businesses to consolidate databases without the traditional associated risks. Consolidating and simplifying mixed workloads running on Microsoft SQL siloed IT environments will not only lower TCO and increase productivity, but establish a modern infrastructure foundation ready for the future data estate.
- Companies are looking for more choices of development languages, data types and OSes in order to unlock the value from their data estate and maximize its value. Re-platforming SQL Server to Linux (or a mixed Windows®/Linux platform) will open doors to greater flexibility as it pertains to data integration, containerization and analysis. There are also potential cost benefits including licenses, training time and hardware costs.[11]
- Containerization provides a flexible platform for even the most complex applications and databases. Containerization of SQL Server empowers companies to deploy updates and upgrades on the fly, and with portability to build locally, deploy to the cloud and run anywhere.
- Develop a roadmap and deployment strategy for moving to Microsoft SQL Server 2019. Depending on your as-is state and desired to-be state, planning includes profiling and assessments for applications and SQL Server datastores to prioritize projects — upgrades, application-centric agile migrations and potential re-platforming — to execute in phases over the journey.

With a modern data management platform in place and complementary data and infrastructure services, IT will be ready to support the business with IT services for new data-driven intelligent applications, business processes and analytics powered from the edge to the core.

[11] Microsoft, "Re-platforming and modernizing your data workloads with SQL Server on Linux", July 2018

Dell Technologies is ready to support you on the journey

Microsoft and Dell Technologies have partnered for over 30 years to provide pre-tested, pre-validated and fully integrated solutions that are engineered to deliver optimized performance for Microsoft SQL Server, Microsoft Azure Stack and Microsoft Azure Stack HCI. Microsoft grants competencies to partners who demonstrate proficiency by having both skilled employees as well as satisfied customers; Dell Technologies has over 16,000 certified employees across customer support, maintenance, deployment, training and consulting and has been granted 18 global competencies.

Dell Technologies has the expertise to help you make the transition to virtualized data that resides in the optimum location without the forklift upgrade. SQL Server manages your data across platforms, on-premises and cloud. The goal of Dell Technologies Consulting Services is to meet you where you are, on any platform, with the tools and languages of your choice.

Dell Technologies Consulting Services can help you plan a solid, foundational set of goals and develop a roadmap for modernization. Our SQL modernizations teams will perform a comprehensive assessment that includes discovery workshops, interviews and thought leadership to provide guidance for your data management strategy. We will identify your future-state goals and create an actionable roadmap, benefits analysis and initial migration priority for your most important workloads. We work to help your business meet its full potential with a data strategy that may include SQL Server, Azure SQL Database and Azure SQL Data Warehouse.

We discover the as-is SQL Server environment including the current state of all in-scope servers, associated workloads and configurations. We then:

- Inventory and classify applications that align to SQL Server databases and all dependencies, thinking through all the connections, reporting, ETL processes and so on.
- Group and prioritize the SQL Server databases or instances by application group and develop a near-term modernization plan and roadmap for modernization. This is also an excellent time to consider database consolidation.
- Identify the rough order of magnitude for future-state compute, storage and software requirements to support a modernization plan.

If you are in a situation where some of your data still resides on end of support (EoS) databases such as SQL Server 2008 and you are unable to upgrade but don't want to risk being without security updates, we can migrate you to Azure or Azure Stack without changing your applications, extending security updates for three years.

In addition, we can migrate your data to your target of choice, understanding that it needs to be flexible enough to build intelligent applications on any data, any platform, any language on-premises and in the cloud. We will also help you upgrade, re-platform and/or consolidate to make it happen.

Learn more

If you're ready to consider a certified, award-winning Microsoft partner who understands your SQL Server endeavors for modernization, Dell EMC's holistic approach can help you minimize risk and business disruption.

To find out more, visit www.dellemc.com/sql and contact your Dell EMC representative.

Additional reading

Dell EMC

- Consolidate and Simplify Mixed Database Workloads with Dell EMC
- Dell EMC Ready Solutions for Microsoft SQL: Design for Dell EMC XtremIO: With PowerEdge R840, Windows Server 2016, and RHEL 7.6
- SQL Server Containers on Linux: Software Development Use Cases Using Dell EMC Infrastructure
- dellemc.com/sql

Microsoft

- SQL Server on Linux: A guide to re-platforming and modernising your data workloads
- Microsoft SQL Server 2019 Big Data Clusters

Intel

- Intel Optane DC persistent memory

DELL EMC