QCT's SQL Server Big Data Clusters to Bring Value to Your Business

We live in an age in which the amount of data we create and can access and use is increasing at a rate almost beyond comprehension: While 1.2 zettabytes of data was created in 2010, IDC predicts that figure will rise to 59 ZB in 2020 and 175 ZB by 2025.[1]

1   "IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data," IDC, May 8, 2020

QCT | intel.

Data is coming from more sources, faster than ever before. But this exponential growth in data volume is accompanied by a raft of challenges, including proving the veracity of the data. With so much data coming from different sources, much of it unstructured, businesses are struggling to know what data to trust and to understand how they can use the information to make wise and actionable decisions.

The next wave of data growth will come from the massive growth in the number of devices connected to networks and the Internet. The Internet of Things (IoT) is expected to connect tens of millions of devices each year to networks. With the number of sensors, cameras, drones and other devices increasing, businesses need to rethink how they manage data.

The challenge for business

Data is typically stored in application-specific silos. But getting insight from this large volume of constantly changing data has typically required a process called ETL (extract, transform, load). Data is copied from various sources into a data warehouse or data mart. As the data is ingested, it is transformed to make the data from the different sources structurally similar.

This model results in the duplication of data. And the extraction, staging and transformation processes need to be secured to mitigate the risk of unauthorized access.

ETL processes are very complex. Decisions about what data sources to use and the business logic of the transformation processes are held in software code. And if something changes in the source data or the business requirements and different data is needed, the cost of change is very high.

Today's businesses need to make decisions faster than ever before. But ETL processes are typically executed as batch processes, so the data being analyzed is historical, not real time or even near real time.

The streamlined flow of data is essential for machine learning (ML) and artificial intelligence (AI) applications. Siloed data adds complexity to workflows, as data engineers and data scientists require different tools. This slows things down and impacts the potential success of ML and AI projects.

Putting together a hardware platform to support a company's business analysis requirements is time consuming and requires significant hardware and software expertise. It can take weeks, or even months, to configure, deploy and tune a sophisticated data analysis and machine learning platform to deliver the best performance for different big data workloads.

A new type of data architecture is needed to support fast access to data, one that can power everything from decision-making to ensuring ML and AI applications have access to the information they need to deliver optimal outputs.

SQL Server Big Data Clusters is a complete platform for AI and ML

SQL Server Big Data Clusters gives businesses what they need to support their big data analytics and AI and ML applications. SQL Server Big Data Clusters has the ability to scale out compute and storage easily to allow businesses to harness the power of parallel processing. It also has the power to query data from data sources to break down data silos. Last, it provides a streamlined workflow for AI and ML by providing all the tools to ingest, store, process and train machine learning models.

SQL Server Big Data Clusters Architecture
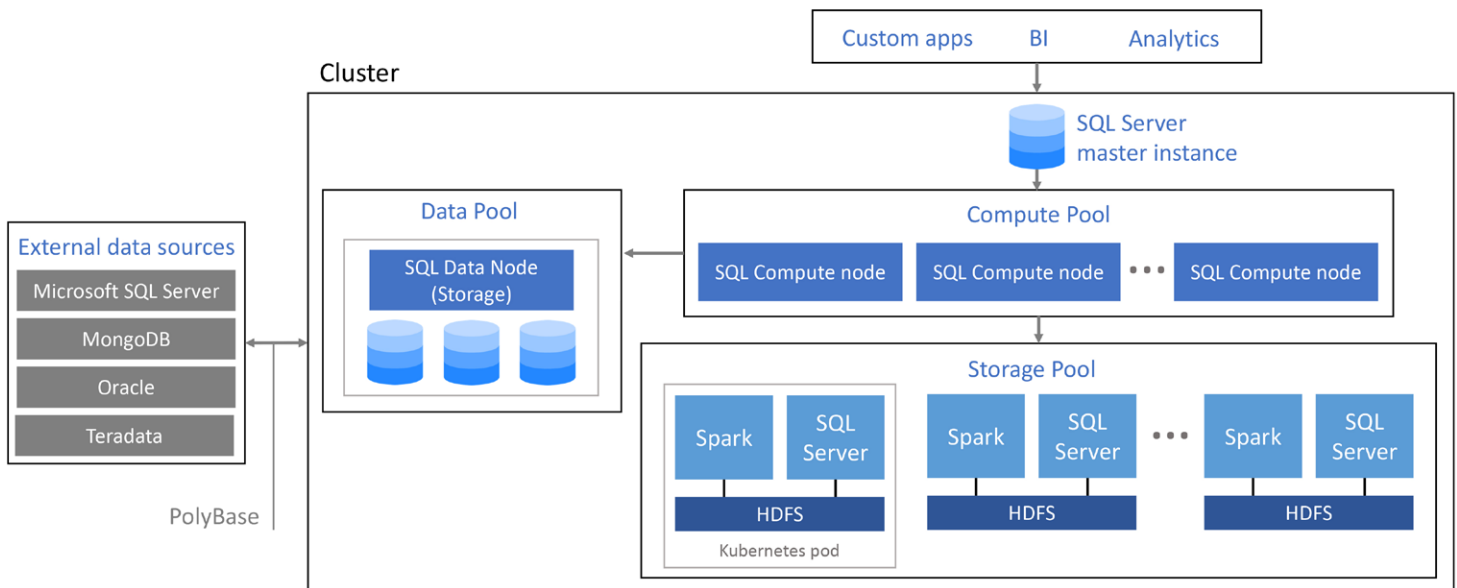
SQL Server has historically been a scale-up database. SQL Server Big Data Clusters was introduced with the Microsoft SQL Server 2019 release. It allows businesses to deploy scale-out clusters that enable businesses to read, write and process big data from high-value relational data and high-volume big data. SQL Server Big Data Clusters lets you deploy scale-out clusters of SQL Server, Spark and HDFS containers running on Kubernetes.

Organizations are able to use the SQL Server Big Data Clusters data virtualization feature to extract data from different sources without replicating and moving data, enabling unified data services for data engineers, data scientists and developers. This reduces the overhead associated with traditional ETL processes and streamlines access to end-to-end data to improve the flow of data into machine learning applications.

SQL Server Big Data Clusters provides the necessary tools for AI and ML applications. It incorporates the building blocks needed to create a robust and streamlined AI and ML workflow. Data ingestion, data transformation and ML model training and deployment can all be completed within the SQL Server Big Data Clusters environment.

However, putting together a solution to make this possible requires experience, knowledge and expertise.

QCT's approach starts with QPOD

The QCT SQL Server Big Data Clusters Solution is powered by QCT Platform on Demand (QPOD). This includes all the compute power, storage and networking capability needed to successfully deploy SQL Server Big Data Clusters.

QPOD is a workload-driven platform that provides an on-premises, rack-level system with best-practice hardware and software integration for specific workloads. It is flexible and scalable to meet business demands for different industries, such as manufacturing, healthcare, and higher education and research.

QPOD is prevalidated and preconfigured to save time and resources, enabling rapid deployment and easy management. This delivers the fastest possible time to value for the investment.

Choosing a best-practice hardware platform for SQL Server Big Data Clusters requires substantial hardware and software expertise.

While SQL Server Big Data Clusters provides the ability to process big data and meet the demand for advanced analytics and machine learning, QPOD provides the computing building blocks, such as deep learning and HPC for various workloads, to complement SQL Server Big Data Clusters. Together, SQL Server Big Data Clusters and QPOD provide a consolidated solution.

Choosing a best-practice hardware platform for SQL Server Big Data Clusters requires substantial hardware and software expertise. QPOD offers a wide range of hardware options with different compute and storage density. This level of flexibility can adapt to fit any workload.

The QPOD hardware portfolio includes an All-Flash server that delivers incredible IOPS performance. The four-socket CPU server supports workloads that require large memory capacity and high compute power. For HPC and AI training, QPOD offers accelerated servers that support up to eight dual-width GPUs or FPGAs. QPOD also offers other hardware platform choices, all of which support cutting-edge

technologies such as NVMe and persistent memory and are prevalidated to deliver a reliable system with the best performance.

Deploying SQL Server Big Data Clusters successfully requires that the right hardware is installed, configured and optimized by an experienced team. QCT's SQL Server Big Data Clusters solution is built using Intel® Xeon® Scalable processors and Intel® Optane™ DC Persistent Memory.

Intel® Xeon® Scalable processors enable enterprises to accelerate their OLTP and OLAP workloads as well as AI and ML applications. Intel® Optane™ DC Persistent Memory provides a large, affordable memory tier and support for data persistence, allowing customers to accelerate the performance of the in-memory OLTP of SQL Server and Apache Spark workloads on their SQL Server Big Data Clusters.

QCT chose to offer Microsoft SQL Server Big Data Clusters with the Intel® platform to provide a best-practice configuration optimized to run big data workloads. In addition, by leveraging QPOD Fast Deployment Service and QPOD Management Building Blocks, customers can speed up the deployment process to shorten time to value and lower operational cost.

SQL Server Big Data Clusters supports the next generation of healthcare

Many organizations across different industries are taking advantage of big data to build competitive advantage. Medicine and healthcare are at the forefront of that revolution.

According to Mordor Intelligence, the global precision medicine market will grow by almost 60% over the next four years.[2] Healthcare and medical organizations need to use AI and ML to derive insights from massive amount of structured and unstructured data collected from patients, diagnostic equipment, wearables and third parties. This is driving healthcare and medical organizations to look for new tools to store, manage and drive benefits from that data.

QCT's SQL Server Big Data Clusters Solution empowers the healthcare industry to deliver precision medicine. Organizations can use Spark in SQL Server Big Data Clusters to ingest and prepare large amounts of data for machine learning. Leveraging QPOD Compute Building Blocks, organizations can unlock the benefits of molecular dynamics simulations for drug analysis, image recognition with deep learning for cancer detection and numerous other use cases.

QCT delivers a consolidated platform with SQL Server Big Data Clusters complemented with QPOD, allowing healthcare organizations to derive value from their big data.

2 "Precision Medicine Market: Growth, Trends, and Forecast (2020-2025)," Mordor Intelligence, 2020

Manufacturing reaps the benefits of QPOD and SQL Server Big Data Clusters

Manufacturing companies are implementing IoT technologies to create smart factories. The large volume of data generated from thousands of IoT devices is an asset that can be used to enable predictive maintenance and optimize supply chain operation. However, manufacturing companies will need a new type of data platform.

A large manufacturing company has implemented SQL Server Big Data Clusters to handle its IoT data and deal with the challenge of data silos. To find a way to optimize the company's supply chain operation, data scientists conducted data mining using static data combined with IoT data. However, data engineers needed to constantly build and update complicated data pipelines to extract data from different data stores. This required a new technology with greater power to ingest its large volume of IoT data.

With the help of SQL Server Big Data Clusters' data virtualization, data scientists can now query data from different sources without data engineers having to build complicated data pipelines. Furthermore, large amounts of IoT data can be ingested into a storage pool of SQL Server Big Data Clusters and be pre-processed using Spark Streaming. This has allowed the company to achieve its goal of being a smart factory.

QCT's SQL Server Big Data Clusters is a perfect big data solution

Businesses will need to contend with the increasing volume, variety and velocity of data, putting a new strain on their entire technology stack, from compute to storage to network. At the same time, they will need to make decisions faster than ever before as the pace of the business world continues to accelerate.

QCT can deliver a SQL Server Big Data Clusters Solution that drastically reduces the time to value for businesses grappling with burgeoning data needs. This is possible because QCT has developed strong relationships over many years with Intel® and other major hardware partners, so it can design, build and deliver highly optimized big data solutions for a wide variety of industries and use cases.

Learn more about QCT's SQL Server Big Data Clusters Solution here.