Microsoft

Faster deployment

Reduced risk

$

Value

Flexibility and choice

Build a Modern Data Warehouse

with the Microsoft Data Warehouse Fast Track Program with SQL Server 2017

December 15, 2017

Contents

SQL Server 2017: The Microsoft Data Warehouse Platform is Built in

Increasingly, organizations are using data for a competitive advantage. Organizations positioned to use data to support strategic business decisions are likely to be more successful than organizations that do not[1]. With the introduction of the latest version of SQL Server, Microsoft continues to find new ways for customers to unleash the power of their data across their entire Data Estate, wherever it lives.

In response to the massive and growing amounts of data, Microsoft enables customers to run SQL Server on their favorite platform, both on-premises and in the cloud. With breakthrough hardware improvements, including Persistent Memory for in-memory databases, SQL Server positions your organization to use data more effectively. With speed, scalability, and agility, SQL Server enables customers to break free from vendor lock-in and drive business transformation.

Benchmark testing results show how SQL Server versions have improved over time. The top-performing SQL Server 2017 result shows a 90 percent gain in query performance over the top-performing SQL Server 2014 result, based on TPC benchmarks. TPC benchmarks run on data sets up to 150 TB demonstrate that you can achieve superlative performance on increasingly less-expensive industry-standard hardware. Current benchmarks are testing in the 200 TB range; improved results are expected soon.

SQL Server data warehouse capabilities build on this high-performance foundation to power solutions from small to massive, across operating system platforms, stretching from ground to cloud. For organizations that want an on-premises data warehouse fast, the SQL Server Data Warehouse Fast Track program enables customers to work with Microsoft partners to deploy a customized, prebuilt solution on the trusted Microsoft data platform. For larger data warehouse needs, Microsoft offers a Massively Parallel Processing (MPP) appliance that can scale from tens of terabytes up to multiple petabytes when incremental nodes are added to existing infrastructure.

---

[1] *The Digital Business Divide*, Keystone Research http://www.microsoft.com/en-us/cloud-platform/data-maturity-model-assessment

As a company focused on enabling business intelligence (BI), Microsoft provides a trusted SQL Server infrastructure that gives you confidence and consistency and enables you to incorporate a wider variety of data sources, including mobile, social, scanners, photos, videos, sensors, devices, RFID, web logs, advanced analytics, click streams, machine learning, and third-party data sources. With SQL Server, customers can query both traditional relational data and new data types with common T-SQL commands by using PolyBase, from either on-premises or in the cloud. Queries that used to take hours can now be done in minutes or seconds with the use of Persistent Memory and in-memory processing. (Persistent Memory, unlike dynamic RAM, does not lose information when the power goes off.) This means you can get results from queries in near real time, using streaming technologies. In addition, with change-tracking functionality, you can identify configuration changes in your environment to help pinpoint operational issues.

Just as BI helps organizations discover trends based on what's already happened, AI helps them look forward. Microsoft SQL Server is the only commercial database with AI built in. Recent advances give users the ability to connect to data sources across Windows, Linux, and Linux-based Docker platforms, using a variety of languages while using tools like SQL Server R Services (introduced in SQL Server 2016) to run deep analysis of the data in near-real time. Supported languages include T-SQL, Java, C/C++, C#/VB.NET, PHP, Node.js, Python, and Ruby.

The following pages will help explain how a combination of the new SQL Server capabilities, inexpensive hardware, the Data Warehouse Fast Track Reference Architecture, and your preferred Microsoft partner can help you deploy a modern data warehouse solution that enables you to use data to benefit your organization. The first section covers SQL Server enhancements. Then, within this context, the following sections examine the Data Warehouse Fast Track program and Reference Architecture in detail.

SQL Server 2017: Everything Built in

Building on the innovations introduced in previous versions, the latest edition of SQL Server delivers a platform with everything built in: Organizations can develop mission-critical applications for online transaction processing (OLTP) across multiple operating system platforms, with breakthrough scalability, in-memory performance, high validity, and the fewest vulnerabilities of any database management system on the market. SQL Server security is built in layers, using the principle of defense in-depth. The world-class security developed for the Windows platform works the same way on Linux. Customers can protect data at rest, in motion, and at the client site with new Always Encrypted functionality. (See Table 1 for a brief overview of the most significant security advances). For additional information on these enhanced security/privacy technologies, refer to Appendix A: Most Secure SQL Server Ever.

SQL Server: Most Secure Server Ever	
Always Encrypted	• Protects data at rest and in motion. • Allows companies to meet privacy standards and achieve full regulatory compliance.
Dynamic Data Masking	• Obfuscates data such as Social Security numbers, birthdays, and account numbers on-the-fly with no changes needed to the application layer. • Allows companies to meet privacy standards and achieve regulatory compliance.
Row-Level Security	• Limits access to data that is needed by a specific type of user (read-only, read-write) on differing data sets. • Greatly simplifies data access restriction techniques with no changes needed to the application layer. • Allows companies to meet privacy standards and achieve regulatory compliance.

Table 1: Some of the SQL Server security features.

In addition, users can transform data into actionable insights that can be delivered on any device — online or offline — by bringing the power of SQL Server to Linux, Linux-based Docker containers, and Windows at one-tenth the cost of other BI solutions. Organizations can use R Services to analyze operational data in real time and at scale, directly within a SQL Server database — without moving the data for analysis. Real-time Operational Analytics (also known in the industry as hybrid transaction and analytical processing, or HTAP) can now be used with in-memory columnstore technology for online transaction processing systems. (See Appendix B: Make Smarter Decisions Faster.) And as organizations take advantage of the cloud, whether data is in an on-premises datacenter, private cloud, or Microsoft Azure, the experience is consistent across server and database as a service (DBaaS).

Acknowledging the Microsoft vision and capabilities for SQL Server, Gartner Group positioned Microsoft in the coveted spot in the upper right quadrant of its "Magic Quadrant for Operational Database Management Systems", ID # G00317993.[2] Microsoft ranked at the top of the charts, both in completeness of vision and ability to execute (see Figure 1). For details on features and enhancements in SQL Server that deliver breakthrough performance, advanced security, and rich integrated reporting and analytics, also see the SQL Server website.[3]

Figure 1: Gartner Magic Quadrant for Operational Database Management Systems, SQL Server 2017.[2]

---

[2] Gartner "Magic Quadrant for Operational Database Management Systems," by Nick Heudecker, Donald Feinberg, Merv Adrian, November 2017. The above quadrant graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. Gartner does not endorse any vendor, product, or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

[3] http://www.microsoft.com/sql-server/sql-server-2017

The Modern Data Estate and Microsoft Data Warehousing Portfolio

The innovations and strengths built into SQL Server provide a foundation to discuss the Microsoft data warehousing portfolio. SQL Server is a highly secure mission-critical database that comes with everything built in — including data warehouse functionality. For customers focused on a data warehousing workload, Microsoft offers options to help achieve the size, type, and location of a data warehouse for different needs, as shown in Figure 2.
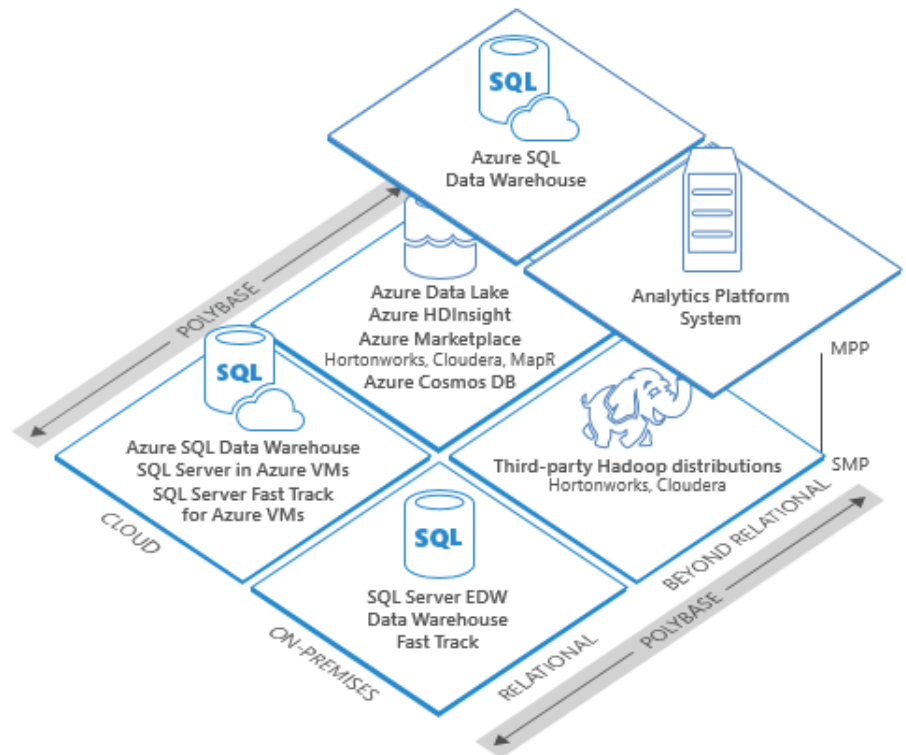

Figure 2: Microsoft solutions and capabilities include options for customers based on size, type, and data warehouse location.

Microsoft options for data warehouse venues include:

- **On-premises Fast Track:** The Data Warehouse Fast Track program, built on a Symmetric Multiprocessing (SMP) Reference Architecture, is an on-premises solution for a data warehouse with up to 150+ TB compute capability and 1.2 PB of storage capacity. Customers can work with Microsoft partners to select the Microsoft-certified solution for their needs.

- **Azure-based Fast Track:** SQL Server Fast Track for Azure VMs is a hosted cloud solution built on a SMP reference architecture. Azure offers varying VM capacities, from A0 (1 core, 1 GB mem, 100 GB) to G5 (32 cores, 448 GB mem, 64 TB). At the time of publication, the largest VM (the GS5: 32 cores, 448 GB memory, 64 TB storage, 4.2 TB local cache) has been certified for Fast Track.

- **Analytics Platform System** is an on-premises MPP appliance for larger data warehouses that delivers linear scalability to 6+ PB.

- **Azure SQL Data Warehouse** is a hosted cloud MPP solution for larger data warehouses. It can quickly grow or shrink storage and compute as needed. Compute and storage are separated, resulting in predictable and scalable performance.

All variants of the SQL data warehouse can integrate with non-relational data, such as data stored in graph databases and data lakes powered by the NoSQL Hadoop, by means of queries written in Transact-SQL (T-SQL) and using the PolyBase "bridge" to translate to non-SQL Server repositories.

Both the on-premises Data Warehouse Fast Track program and the Analytics Platform System appliance can stretch to the cloud by extending to SQL Server implemented in Azure virtual machines (VMs) or by using Azure SQL Data Warehouse, respectively.

Options for On-premises Solutions

With SQL Server, Microsoft offers data warehouse solutions that extend from terabytes to petabytes; but balancing flexibility and choice isn't always straightforward. Microsoft has worked hard to package data warehousing solutions that let customers determine the best and most cost-effective option for their situation and to reduce the time to a working solution, while reining in costs. Figure 3 illustrates the options and their relative cost implications.
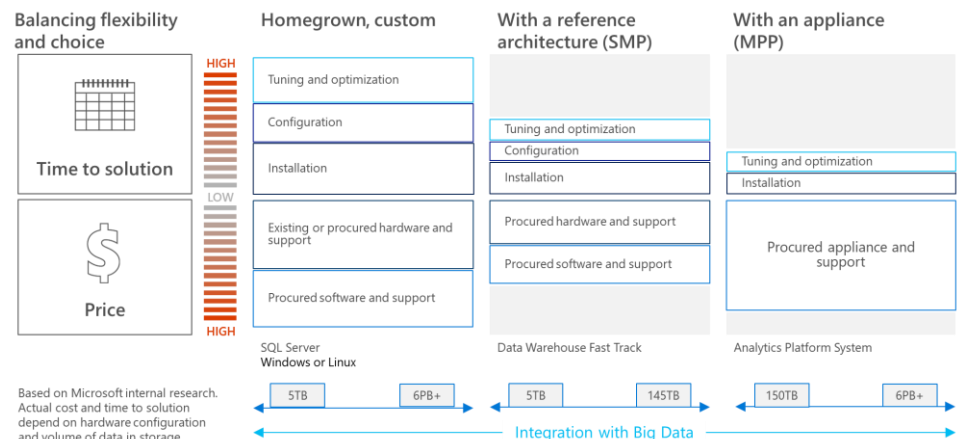

Figure 3: Cost and time-to-solution comparisons between home-grown, Data Warehouse Fast Track solutions, and the Analytics Platform System appliance. Estimates are based on Microsoft internal research; actual cost and time to solution will vary, depending on hardware configuration and volume of data in storage.

For smaller (up to 150 TB compute capability and 1.2 PB storage capability) data warehouses, the SMP data warehouse is the most commonly used architecture. These systems are characterized by a single instance of a relational database management system (DBMS) sharing all resources (CPU/Memory/Disk – i.e., "shared everything"). This is most easily achieved with the help of the Microsoft Data Warehouse Fast Track program and its Reference Architecture.

Companies with extensive time and resources can use the Fast Track Reference Architecture to build their own homegrown data warehouse solution (assembly is definitely required). Figure 3 illustrates the cost and time requirements. For a speedier, proven solution requiring fewer resources and less time, however, organizations can work with a certified Microsoft partner (no assembly required). Contracting with a partner ensures that specific size and use requirements are met without exceeding necessary capacities. While hardware continues to drop in price, there's no need to overspend.

In addition to correctly sizing the hardware and processing capabilities, the Data Warehouse Fast Track option, deployed in conjunction with a Microsoft partner, will eliminate the need for IT staff to have deep comprehension of the data warehouse file system layouts, and — since the Microsoft partner must be certified in order to offer Data Warehouse Fast Track services — customers can be confident that the physical implementation of the data warehouse complies with Data Warehouse Fast Track Reference Architecture guidelines. For more information on the resource changes or enhancements incorporated in the SQL Server Data Warehouse Fast Track options, refer to Appendix C: Resource Changes for Data Warehouse Fast Track, SQL Server 2016 to 2017.

The third option, for data capacities of 6 PB or more, is the Microsoft Analytics Platform System. It offers the shortest time-to-value, with costs roughly comparable to the homegrown solution. While it's not impossible to roll-your-own MPP data warehouse, many organizations will prefer to start with a preconfigured appliance that has been customized to specific needs while reducing the risk, cost, and complexity of deploying a very large data warehouse. Expertise is required because within the MPP universe, queries are different, modeling is different, data structures are different, even partitioning of data is different from the comparable structures in an SMP environment. The Microsoft partners that offer MPP solutions have undergone grueling certification processes and will deliver a solution that outperforms its market rivals.

To fully realize the value of data, you need a complete platform that can manage both structured and unstructured data with security, consistency, and credibility. Data warehouse and big data solutions from Microsoft provide a trusted infrastructure that can handle all types of data and can scale from terabytes to petabytes while providing real-time business insights and fast performance.

Data Warehouse Fast Track Program

Recent years have resulted in a data explosion with accompanying challenges in data storage and information retrieval. Web sites are streaming data directly into corporate databases at a rate unthinkable just a few years ago. Databases are swelling to sizes unmanageable with currently technology. Extracting information from this massive amount of data is getting more complicated as the questions get more and more sophisticated. Organizations are missing opportunities and wasting effort. Microsoft has stepped up to the challenge by introducing the Data Warehouse Fast Track Reference Architecture, a set of guidelines that partners can use to help customers build medium-to-large data warehouse solutions based on SQL Server Enterprise edition and well-tuned configurations from certified hardware vendors. Fast Track is an SMP data warehouse option running on Windows Server or Linux that has a current on-premises data compute capacity range that seamlessly scales from 5 TB to 150 TB, with a data storage capacity that reaches past 1 PB. Microsoft is constantly pushing the capacity envelope, and the Data Warehouse Fast Track program is rapidly extending this range with the publication of new benchmarks.

This mature deployment, available from certified vendors, works best when matched to customer data and application requirements. Data can be stored as either a set of rows (the traditional database method) or it can be stored as a set of columns, using the enhanced clustered columnstore technology. Columnstore can result in a storage savings that varies between 35-90 percent over row store, depending on the data. The Fast Track program has many administrative, operational and programming capabilities and is a recommended best experience for the smaller enterprise data warehouse. Customers can upgrade from SMP, either on-premises or in the Azure cloud, to the massive multi-processor Analytics Platform System (formerly Parallel Data Warehouse) or Azure SQL Data Warehouse, respectively.

All versions of the Microsoft data warehouse offer in-memory capability; it's built into SQL Server. In-memory databases perform significantly faster than their disk-bound counterparts. The SQL Server enhanced 10:1 compression ratio, coupled with columnstore technology, boosts performance impressively over previous versions of SQL Server. Independent verification by Intel, documented in a white paper,[4] demonstrates the vast scalability of a Data Warehouse Fast Track solution using newer hardware (CPUs, solid state devices — SSDs — and RAID controllers). Working with a 100 TB data set, a single Xeon® processor-based platform (4 sockets/72 cores) recorded significant gains (up to 30 percent) with SQL Server 2016 and Windows Server 2016. Intel later repeated the test using the same workload and same configuration

---

[4] http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/microsoft-sql-database-analytics-paper.pdf

– the only difference was upgraded processors. With the newer processors, Intel observed up to 36 percent performance gains on average for complex ad-hoc queries against the 100 TB data set. [5]

Data Warehouse Fast Track Advantages

A Data Warehouse Fast Track solution is designed to deliver not only ease and speed of deployment and cost effectiveness, but also ensure that customers get the experience they prefer. Greater choice and better performance are key attributes of Fast Track, as shown in Table 2.

More choice

New configurations from major hardware partners use the latest industry hardware and software engineering.

Organizations can gain immediate value from latest-generation servers and storage.

Separate performance metrics for rowstore and columnstore help in choosing a reference platform.

More choice in platform (Windows, Linux) and language (T-SQL, Java, C/C++, C#/VB.Net, PHP, Node.js, Python, Ruby)

Better performance

Balanced capacity and performance, validated with the latest version of SQL Server.

70%+ of the certifications offer storage subsystems using solid state technology. Scales well on emerging high-core technology. The top-performing SQL Server 2017 result shows nearly 100% gain in query performance over the top-performing SQL Server 2014 result, based on TPC benchmarks.

Persistent Memory technology enables the in-memory diskless database, backed by high-speed SSDs in case of power loss.

Table 2: More choice, better performance with a Data Warehouse Fast Track solution.

Choice and performance are not the only advantages offered by a Fast Track solution. Additional benefits include the following:

Faster deployment: The Fast Track program uses the core capabilities of SQL Server on Windows or Linux servers to deliver a balanced SMP data warehouse with optimized performance.

Out-of-the-box offerings: Data Warehouse Fast Track solutions are optimized specifically for data warehouse workloads.

Reduced risk: Predetermined balance across key system components minimizes the risk of overspending for CPU or storage resources that will never be realized at the application level.

Predictable out-of-the-box performance: Fast Track configurations are built to a capacity that already matches the capabilities of the SQL Server application for the selected system and target workload.

Flexibility and choice: Top industry hardware and software engineering are at the heart of the Fast Track solution. Choose from multiple partner configurations and Windows Server or Linux deployments that are

---

[5] http://itpeernetwork.intel.com/performance-just-next-level-business-analytics-new-intel-xeon-processor-e7-v4-microsoft-sql-server-2016-100tb-dataset

certified for Data Warehouse Fast Track use. Get the advanced capabilities of latest-generation servers and storage.

Value: Fast Track solutions are prebuilt, eliminating the necessity of having to research and test multiple hardware components. Cost per QphH (query per hour, a price/performance measure) drops significantly when Fast Track is deployed with columnstore and in-memory technologies. Organizations can gain immediate value from latest-generation servers and storage provided with Fast Track solutions.

Improved Hardware and Storage Options for SQL Server

In accordance with Moore's Law,[6] CPUs today offer significantly more power than two years ago. The SQL Server software enhancements are designed to take advantage of this power. This means that SQL Server Fast Track implementations will benefit from this boost in processing power. Today, SQL Server— running on two CPU sockets that support 1.5x more cores than were supported in previous versions of SQL Server— will perform twice as well as previous versions of SQL Server Fast Track. The performance boost delivered by more cores per socket and the availability of more (and less-expensive) RAM to accommodate more in-memory operations, including the option of using Persistent Memory and circumventing storage I/O altogether provides a more balanced configuration for modern versions of SQL Server and the Fast Track Reference Architecture.

Microsoft recommends that reference architectures use SSDs as the primary storage for database data and TempDB files for SQL Server. New and better compression technologies (compression ratios have been increased, from 3.5:1 to as high as 10:1) mean that columnstore technology used with SSDs should be the primary on-disk structure, enabling more data to fit on a drive and ensuring that more data can be retrieved during each physical IO.

To better understand the differences between SQL Server versions, it's worth comparing on-premises implementations of the SQL Server Fast Track Reference Architectures. Figure 4, which follows, compares a SQL Server 2014 implementation[7] to two implementations running SQL Server 2017[8] in the TPC-H 10 TB nonclustered category. The figure illustrates that SQL Server 2017, running on certified Fast Track hardware and software, beat the SQL Server 2014 record by nearly a 2:1 margin. Although SQL 2014 running on the HPE SuperDomeX has twice the number of sockets and 28 percent more cores than the hardware

---

[6] http://en.wikipedia.org/wiki/Moore%27s_law

[7] http://www.tpc.org/3329

[8] http://www.tpc.org/3328 and http://www.tpc.org/3317

running SQL 2017, both SQL Server 2017 configurations outstripped the SQL Server 2014 configuration in terms of performance. Also QphH cost dropped by nearly two-thirds (0.93 and 0.95$ vs 2.27$) with SQL Server 2017. While the Lenovo system came in second to the HPE implementation, the trade-off is based on a less expensive price point.
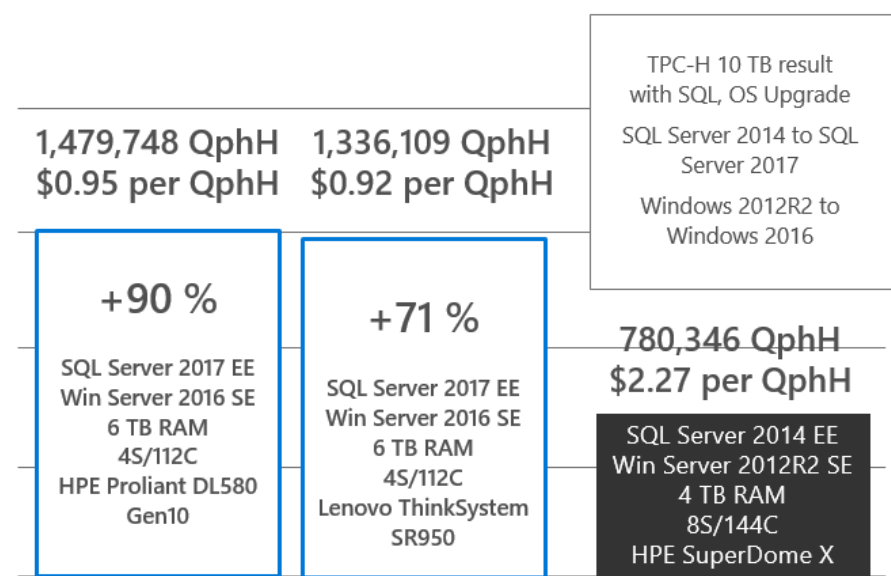

Figure 4: SQL Server 2017 improves performance with in-memory columnstore and more RAM per core. The chart compares two top SQL Server 2017 performers as of December 2017 against the top SQL Server 2014 performer in the TPC-H 10 TB nonclustered category. The two SQL Server 2017 configurations offer sizzling performance at two different price points, compared to the SQL Server 2014 configuration.

For a more detailed drill-down of how SQL Server takes advantage of hardware enhancements, refer to Appendix D, "Accelerating Large-scale Business Analytics: Using a 100TB Dataset Powered by the Intel® Xeon® Processor E7 Family and Microsoft® SQL Server 2016."

In-memory Performance Improvements

The SQL Server Data Warehouse Fast Track program not only takes advantage of the latest hardware advances, it also employs the latest in software innovations, such as columnstore technology — SQL Server database indices created from columns of data rather than from rows of data — enhanced data compression and Persistent Memory.

Columnstore indices offer great advantages over traditional row stores for analytics and data warehousing queries. Columnstore is ideally suited for star schemas and for tables with billions of rows. The average 10x compression created by columnstore indices results in space and cost savings, significant performance increases due to dramatically reduced IO requirements, and faster performance, as SQL Server need only scan the specific columns required by each query. Columnstore indices also require less RAM to hold data in memory, resulting in additional performance boosts. In addition, SQL Server supports the presence of a non-persisted computed column in a columnstore index, and additional

B-Tree indices to columnstore-based tables; the B-Tree indices enable efficient single-row lookups.

In brief, in-memory columnstore offers the following benefits:

- Massive compression from storing data in columnar format.

- Next-generation performance with ability to load data into or out of memory.

- Efficient single-row lookup with columnstore support for additional (B-Tree) indices.

- Real-time results since indexes are updateable and clustered for real-time trickle loading.

For a deeper look into enhancements and improvements to columnstore technology in SQL Server, see Appendix E: What's New About SQL Server 2016 Columnstore Technology?

Fast Track Evaluation Process

Working with Microsoft partners, customers can evaluate the Data Warehouse Fast Track Reference Architecture to find the best fit for their specific circumstances, based on existing partner relationships and multiple hardware configuration options. Participating partners undergo an extensive certification process to participate in the Data Warehouse Fast Track program. As a result, customers can be confident that the solution will offer the performance and quality they need.

The main high-level process for evaluating Data Warehouse Fast Track options with vendors generally follows these steps:[9]

1. Select the certified reference architecture that suits your needs.

2. Test a sample workload on the selected configuration.

3. Validate the scenario.

4. Fine tune for performance.

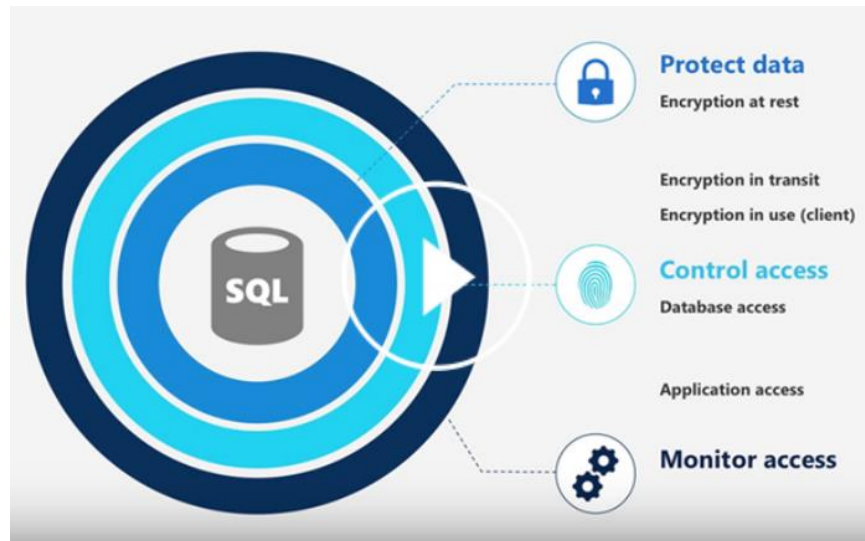5. Make the final decision to go or no-go.

While in-depth evaluation obviously requires greater effort than these steps imply, working with Fast Track Partners makes it much easier to select a data warehouse solution compared to trying to figure out everything unassisted. Customers can evaluate more than 40 Fast Track

---

[9] http://www.microsoft.com/sql-server/data-warehousing

Reference Architectures to find the best fit for their specific circumstances. Knowing that participating partners must pass an extensive certification process to qualify for the Data Warehouse Fast Track program gives customers the certainty that the solution they choose will give them the performance and quality they need and will help them get to an in-place solution fast.

Find out more about Microsoft's Data Warehouse Fast Track program at http://www.microsoft.com/sql-server/data-warehousing.

Appendix A:
SQL Server 2017, Most Secure Server Ever

SQL Server Security is built in layers, using the principle of defense in-depth. The world-class security developed for the Windows platform works the same way on Linux.

The first layer of security is protecting the data itself. Most of this is done using various types of encryption. Data is protected at rest, while it's in transit, and while it's being used by the client application.

> SQL Server protects the data when it's at rest with **Transparent Data Encryption** (TDE), backup encryption, and cell-level encryption. TDE encrypts the entire database – data files, log files, backup files – without requiring any application-level changes. TDE works the same on Windows and Linux.
>
> Encryption in transit requires encrypting communication with SQL Server. Encryption in transit uses the latest industry standard **TLS 1.2 protocol**, which is supported by both Windows and Linux. Encryption in transit can also be configured to be FIPS 140-2 compliant, if the Linux machine is running in FIPS mode.
>
> Encryption in use is client-side encryption, which is SQL Server's **Always Encrypted**. Always Encrypted encrypts the data on the client side, with minimal changes to the application. Always Encrypted ensures that sensitive data is encrypted in the database at all times; that encryption and decryption are done transparently within the client driver (ODBC/JDBC) with few if any application code changes required. Basic equality queries, such as JOIN, GROUP BY, DISTINCT, and equality searches, are supported on encrypted data. More complex queries will require decryption prior to query execution.

The second layer is access control – who is allowed to access which parts of the data. SQL Server authentication is highly granular, so access control can be achieved through authentication and authorization. Active Directory (AD) authentication is available for Windows and Linux,

creating a password-less single-sign on using the Kerberos protocol. Authentication is managed centrally using AD groups, allowing a single user to have the same identity across the entire organization and across both Windows & Linux machines.

Application access, which is part of this layer, involves SQL Server's client-side programmability features:

Row-Level Security limits access to specific rows in a table based on the querying user's role and authorization. Previous solutions to this quandary involved enforcement at the application side, or creating views, triggers, or other database devices to prohibit users from accessing or updating data that they shouldn't be allowed to view or update, or both. For SQL Server, Microsoft moved the onerous task of limiting row access from the application or the database administrator to the SQL Server engine, which attaches access logic to the database table itself. Microsoft has given organizations a solution that controls reads and writes to the data, based on flexible access criteria, including user identity, role/group memberships, connection data, time of day, and more – a solution that meets privacy standards and regulatory compliance. This solution works transparently with no application or query changes required, for consistent maintenance of access policy and reduced risk of accidental data exposure.
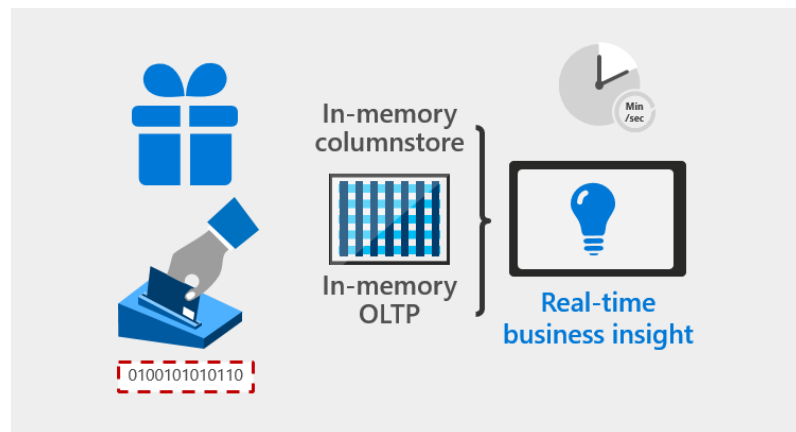
Dynamic Data Masking can hide query results from non-privileged users by employing a centralized policy that is stored within the database and enforced by SQL Server. Data is masked on-the-fly without touching the underlying data. The process is completely transparent to the application or query, no changes needed. This simplifies the application code and minimizes the risk of accidental data exposure, allowing companies to meet privacy standards and achieve regulatory compliance.

The third, outermost layer is monitoring access – keeping track of who's doing what inside the database. SQL Server's fine-grained audit feature allows tracking of users, actions, and objects, and the database engine logs all activity based on the audit policy. This audit trail allows SQL Server to meet stringent compliance standards.

Azure Virtual network service enables you to securely connect Azure resources to each other with virtual networks. Azure Virtual network is a logical isolation of the Azure cloud dedicated to your Azure subscription. You can think of Azure Virtual network as a representation of your own network in the cloud, and you can connect your Azure Virtual network to your on-premises network. Read more about Azure Virtual Network capabilities.

Appendix B:

Make Smarter Decisions Faster: In-memory Built in

New to SQL Server 2017 is the combination of in-memory processing capabilities and Persistent Memory (PMEM) technology[10], which—when combined with the in-memory columnstore technology for operational OLTP systems in SQL Server—delivers real-time Operational Analytics out of the box. In the past, to get Operational Analytics, you had to extract data from the OLTP system, load it into a data warehouse, and use analysis tools that were connected to the data warehouse. The extract, transform, and load (ETL) was typically an overnight process, unless you had a scheme to trickle changes into the data warehouse—but this latter process still took hours. So being able to easily get real-time or even near-real-time analytics for the operational database was extremely difficult.

To get to real-time in-memory operational analytics (HTAP) involves building a nonclustered columnstore index in memory on a traditional table of the OLTP database. This technology, introduced to the SQL Server OLTP environment, gives many customers the ability to efficiently do analytical queries directly on their operational data with no ETL lag or complexity and minimal impact on OLTP performance.[11]

For instance, let's say that you're a retailer. It's now 3:00 PM., and you have one hour to place your inventory re-order. That's not nearly enough time to offload today's sales transactions to the data warehouse and run the inventory reorder reports—perhaps it was at one time, but your company has grown so much that one hour is no longer enough time. What do you do?

On your operational SQL Server instance, you have headroom with both memory and CPU. This enables you to create an in-memory nonclustered columnstore index on the sales table from the OLTP

---

[10] HPE Persistent Memory (PMEM) technology is a system configured to use high speed memory as disks. The memory is fully backed up by high-speed SSDs in case the system loses power. All the data is kept in memory at all times.

[11] https://channel9.msdn.com/Events/DataDriven/SQLServer2016/Real-Time-Operational-analytics

database. All you want to know is what inventory you sold today, and how much to reorder for tomorrow.

With SQL Server, you can extract fast-changing information, using analytical queries against the transactional database, without having to use ETL to port the OLTP data into a data warehouse and then run analytical queries against the data warehouse. Within a matter of minutes, or seconds if you've invested in a persistent RAM hardware setup, you can extract real-time insight from your operational platform.

Appendix C:
Resource Changes for Data Warehouse Fast Track, SQL Server 2014 to SQL Server 2017

Minimum Transaction Log Space	TempDB size	Minimum memory
SQL Server 2014: 400 GB or 1.2% of rated capacity	**SQL Server 2014:** 25%-30%	**SQL Server 2014:** 1/2/4/8 Sockets **8/9/12/15 GB RAM per TB of rated capacity respectively** • Minimum 540 GB RAM for 2-socket 60 TB certification. • Minimum 720 GB RAM for 4-socket 60 TB certification.
SQL Server 2016, 2017: 400 GB or 2% of rated capacity	**SQL Server 2016, 2107:** No change	**SQL Server 2016, 2017: 1/2/4/8 Sockets** **10/12/15/18 GB RAM per TB of rated capacity respectively** • Minimum 720 GB RAM for 2-socket 60 TB certification. • Minimum 900 GB RAM for 4-socket 60 TB certification.

The figure above illustrates the changing configuration requirements for upgrading from SQL Server 2014 to SQL Server 2016 or SQL Server 2017 for on-premises data warehouse operations. These are reflected in the Fast Track Reference Architecture.

The rationale for the increase in minimum memory requirements: The new generation of processors has more cores per socket, and RAM costs are a smaller portion of the total system cost. Additional RAM provides a significant performance boost when the data can become fully resident in RAM.

Appendix D:
Accelerating Large-scale Business Analytics: Using a 100TB Dataset Powered by the Intel® Xeon® Processor E7 Family and Microsoft SQL Server

Find the white paper detailing the on-premises installation at http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/microsoft-sql-database-analytics-paper.pdf.

Performance Characteristics Of Concurrent Runs			
Avg. CPU Utilization	84%	Avg. Read Size	130KB
Avg. Write Size	64KB	Read/Write Ratio	3:2
Avg. Read Bandwidth	4.4GB/s	Avg. Write Bandwidth	1.3GB/s
Avg. I/O Bandwidth	5.6GB/s	Peak I/O Bandwidth	15.3GB/s
Avg. IOPS (Read + Write)	55,740	Peak IOPS (Read + Write)	179,484
Avg. I/O Latency	6ms	Peak I/O Latency	23ms

Summary

"In this paper, we presented our findings on running a large (100TB) data warehousing application using Microsoft® SQL Server on the Intel® Xeon® processor E7-8800 v3 and enterprise-class Intel® SSDs and RAID controllers. We demonstrated that it is possible to build and run a 100TB data warehouse on a symmetric multiprocessor (SMP) configuration using technology that is readily available today from Intel and Microsoft. Our system with four Intel® Xeon® processor E7-8800 v3 product family processors fuels the advanced business analytics capabilities of Microsoft® SQL Server to deliver stunning performance, processing 100TB worth of data to bring key business insights in a matter of minutes. SQL Server, optimized for Intel® architecture, makes use of large system memory that can fit terabytes of data, eliminating costly I/O operations. Enterprise-level Intel® SSDs ensure fast data access, delivering performance as needed."

Appendix E:
What's New About SQL Server 2017 Columnstore Technology?

More than 10X compression in data size

Additional indices and in-memory support for efficient single-row look up

Operator Pushdown optimizes query processing

Batch mode processing

Microsoft SQL Server has made significant improvements in data warehousing columnstore technology. Columnstore indices (indices created from columns of data) offer great advantages over traditional row stores for analytics and data warehousing queries. Columnstore indices are ideally suited for the star schema and tables populated with

billions of rows, which are common to data warehouses. Among the advantages for analytics are:

- **More than 10X compression in data size:** Data warehouses are very large by nature, and the compression offered by columnstore index technologies offers both space and cost savings. Columnstore also significantly increases performance due to the dramatically reduced IO requirements, which happen as a result of 10:1 compression coupled with the ability to scan only the specific columns required by each query. This compression reduces the amount of memory required to hold a given number of rows from the source data warehouse, thus enabling more queries to run in memory, from start to finish.

- **Additional Indices:** Beginning with SQL Server 2016, you can have one or more rowstore (B-Tree) nonclustered indices on a columnstore-based table, which enables efficient single-row lookups. By using the UNIQUE constraint on a rowstore index, you can enforce a primary key constraint on the columnstore.

- **Computed Columns:** SQL Server supports a non-persistent computed column in a columnstore index; this feature will be welcomed by financial and banking applications, among others.

In addition to these architectural features, SQL Server has further optimized the processing of queries in columnstore indices in the following ways:

- **Operator Pushdown:** Pushdown refers to moving both filter and aggregation query operations closer to the data, so that many of the filters and calculations can be done in the scan operators, essentially pre-filtering the data prior to the execution of more complex query operations. This dramatically reduces the volume of data that needs to be handled by JOIN and GROUP BY operators.

- **Batch Mode Processing:** SQL Server includes enhancements in batch-mode processing (processing many rows at a time rather than serially calculating values on each individual row). These batch operations are further optimized by leveraging Single Instruction Multiple Data (SIMD) vector processing CPU instructions in the Intel® architectures.