**Microsoft**

Faster
deployment

Reduced
risk

$

Value

Flexibility
and choice

# Introducing Microsoft Data Warehouse Fast Track for SQL Server 2016

November 11, 2016

# Contents

# SQL Server 2016: The Microsoft Data Warehouse platform that has everything built-in

**Organizations positioned to use data to support strategic business decisions will be more successful than those that lag in their use of data[1]. Microsoft Data Warehouse Fast Track for SQL Server 2016 is an advanced data platform reference architecture that works with breakthrough hardware improvements to deliver superior performance — both on premises and in the cloud — and help position your organization to use data more effectively. With speed, scalability, and agility, SQL Server 2016 enables customers to break free from vendor lock-in and drive business transformation.**

Benchmark testing results tell the performance story: SQL Server 2016 offers up to 40 percent improvement in query performance over SQL Server 2014. TPC benchmarks that were run on data sets up to 145 TB demonstrate that you can achieve superlative performance on increasingly less-expensive industry-standard hardware. Current benchmarks are testing in the 150-200 TB range; improved results are expected soon[2].

SQL Server 2016 data warehouse capabilities build on this high-performance foundation to power solutions that span the range of requirements from small to massive. For organizations that want to get started with a data warehouse, the SQL Server 2016 Data Warehouse Fast Track program enables customers to work with Microsoft partners to deploy a customized, prebuilt data warehouse 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.

As a company focused on enabling business intelligence (BI), Microsoft provides a trusted SQL Server 2016 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.

Building on such advanced capabilities, the SQL Server 2016 Data Warehouse Fast Track platform lets you query both traditional relational data and new data types with common T-SQL commands by using

---

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

[2] TCP™ performance results:
http://www.tpc.org/tpch/results/tpch_perf_results.asp?resulttype=noncluster&amp;version=2%&amp;currencyID=0

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 in-memory processing. 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.

The following pages will help explain how a combination of the new SQL Server 2016 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 2016 enhancements. Then, within this context, the following sections examine the Data Warehouse Fast Track program and Reference Architecture in detail.

# SQL Server 2016: More than just a database

Building on the innovations introduced in SQL Server 2014, the latest version of SQL Server delivers a platform with everything built in: Organizations can develop mission-critical applications for online transaction processing (OLTP), with breakthrough scalability, in-memory performance, and high validity. Customers can protect data at rest and in motion with new Always Encrypted functionality. (See Table 1 for a brief overview of the most significant security advances). For additional information on each of these enhanced security/privacy technologies, refer to Appendix A: Most Secure SQL Server Ever.

| SQL Server 2016: Most Secure Server Ever | |
|---|---|
| Always Encrypted | • Protects data at rest and in motion.<br>• 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.<br>• 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.<br>• Greatly simplifies data access restriction techniques with no changes needed to the application layer.<br>• Allows companies to meet privacy standards and achieve regulatory compliance. |

*Table 1: SQL Server 2016 security features.*

In addition, users can transform data into actionable insights that can be delivered on any device — online or offline — at one-fifth 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 # G00271405[3]. SQL Server 2014 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 2016 that deliver breakthrough performance, advanced security, and rich integrated reporting and analytics, see the SQL Server 2016 site[4].



*Figure 1: Gartner Magic Quadrant for Operational Database Management Systems, SQL Server 2014.*

---

[3] Gartner "Magic Quadrant for Operational Database Management Systems," by Donald Feinberg, Merv Adrian,  Nick Heudecker,  Adam M. Ronthal, Terilyn Palanca, October 12, 2015. 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.

[4] https://www.microsoft.com/en-us/cloud-platform/sql-server

# The Microsoft Big Data and Data Warehousing Portfolio

The innovations and strengths built into SQL Server 2016 provide context for a discussion of the Microsoft data warehousing portfolio. SQL Server 2016 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 provides options to help achieve the size and type of data warehouse for their particular needs, as shown in Figure 2.
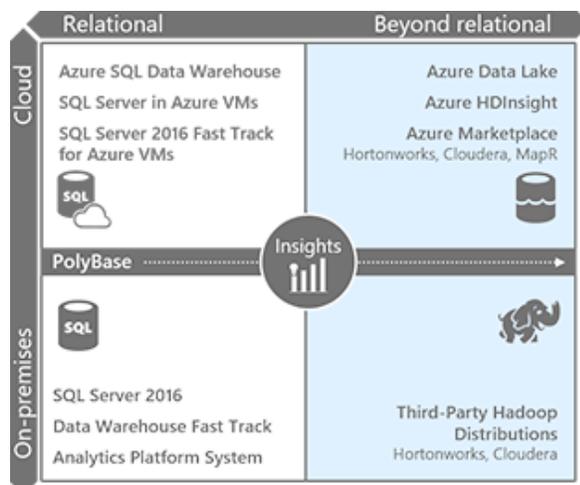


*Figure 2: The Microsoft data warehouse portfolio.*

Microsoft options for data warehouse workloads 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 145+ TB compute capability and 1.2 PB of storage capability. Customers can work with Microsoft partners to select the Microsoft-certified solution for their needs.

- **Cloud-based Fast Track:** SQL Server 2016 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 mem, 64 TB storage, 4.2 TB local cache) has been certified for Fast Track.

- The 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 2016 data warehouse can integrate with non-relational data, such as that stored in 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 have the ability to 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 2016, 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.
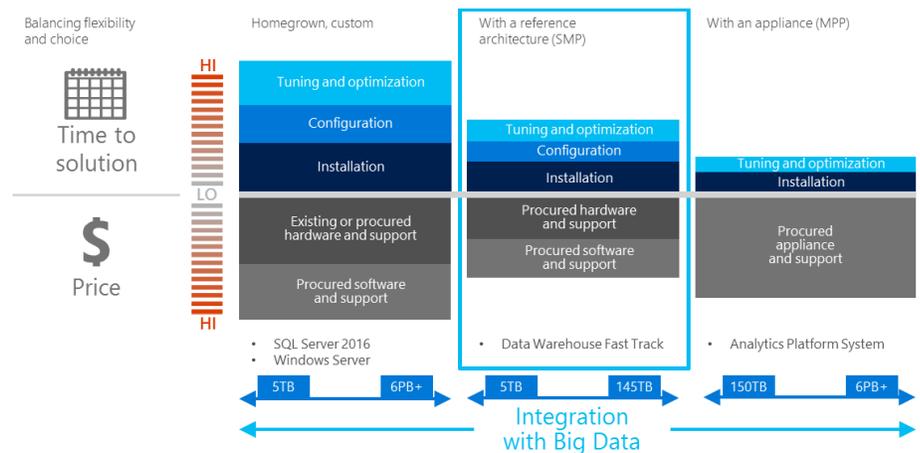
*Figure 3: Cost and time-to-solution comparisons between home-grown, Data Warehouse Fast Track, and the APS appliance. Above estimates are based on Microsoft internal research; actual cost and time to solution will vary, depending on hardware configuration and volume of data in storage.*

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 2016 Data Warehouse Fast Track options, refer to Appendix C: Resource Changes for Data Warehouse Fast Track, SQL Server 2014 to 2016.

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 seen 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 2016 Enterprise edition and well-tuned configurations from certified hardware vendors. Fast Track is an SMP data warehouse option that has a current on-premises data compute capacity range that seamlessly scales from 5 TB to 145 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. 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 2016 enhanced 10:1 compression ratio, coupled with columnstore technology, boosts performance impressively over SQL Server 2014. Independent verification by Intel, documented in a white paper[5]  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.

# 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 two key attributes of Fast Track, as illustrated in Table 2, which follows.

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

| MORE CHOICE | BETTER PERFORMANCE |
|---|---|
| New configurations from major hardware partners use the latest industry hardware and software engineering. | Balanced capacity and performance, validated with SQL Server 2016. |
| Organizations can gain immediate value from latest-generation servers and storage. | Scales well on emerging high-core technology. SQL Server 2016 on a 4-socket box with enhanced hardware shows 80 percent+ gain in query performance over SQL Server 2014, based on TPC benchmarks. |
| Separate performance metrics for rowstore and columnstore help in choosing a reference platform. | 70+ percent of the certifications (SQL Server 2014 and SQL Server 2016) offer storage subsystems using solid state technology. |

*Table 2: More choice, better performance with SQL Server 2016 as part of 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 the Windows Server operating system and SQL Server to deliver a balanced SMP data warehouse with optimized performance.

**Out-of-the-box offerings:** Data Warehouse Fast Track is designed for data warehousing. Rather than being a one-size-fits-all approach to database configuration, the Data Warehouse Fast Track approach is optimized specifically for a data warehouse workload.

**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 that are 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 2016

In accordance with Moore's Law[6], CPUs today offer significantly more power than two years ago. The SQL Server 2016 software enhancements are designed to take advantage of this power. This means that SQL Server 2016 Fast Track implementations will benefit from this boost in processing power. Today, SQL Server 2016 — running on two CPU sockets that support 1.5x more cores than were supported in SQL Server 2014 — will perform twice as well as SQL Server 2014 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, provides a more balanced configuration for SQL Server 2016 and the Fast Track Reference Architecture.

SSD prices continue to drop, and Microsoft recommends using SSDs as the primary storage for database data and TempDB files for SQL Server 2016. 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.

**SQL Server 2016 improves performance/ core, handles more cores and more RAM**

1,106,832 QphH

606,821 QphH

+82%

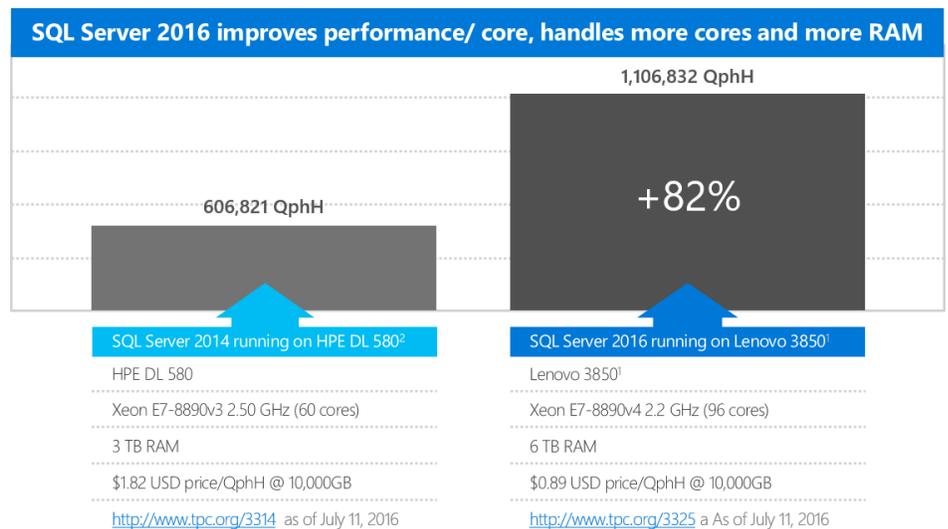| SQL Server 2014 running on HPE DL 580[2] | SQL Server 2016 running on Lenovo 3850[1] |
| --- | --- |
| HPE DL 580 | Lenovo 3850[1] |
| Xeon E7-8890v3 2.50 GHz (60 cores) | Xeon E7-8890v4 2.2 GHz (96 cores) |
| 3 TB RAM | 6 TB RAM |
| $1.82 USD price/QphH @ 10,000GB | $0.89 USD price/QphH @ 10,000GB |
| http://www.tpc.org/3314  as of July 11, 2016 | http://www.tpc.org/3325 a As of July 11, 2016 |

*Figure 4: SQL Server 2016 Data Warehouse Fast Track performance, as compared to SQL Server 2014 Fast Track, both on-premises. The chart compares two TPC-H 10 TB results, each on a four-socket server, as of July 11, 2016. SQL Server 2016 running on more powerful hardware showed tremendous performance improvement.*

Comparing an on-premises implementation of the SQL Server 2014 Fast Track Reference Architecture to an on-premises implementation of the SQL Server 2016 Fast Track Reference Architecture underscores the differences (and performance gains) between the two. Figure 4, which compares SQL Server 2014[7] to SQL Server 2016[8] performance, illustrates that SQL Server 2016, running on certified Fast Track hardware and

---

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

[7] http://www.tpc.org/3314

[8] http://www.tpc.org/3325

software, beat the SQL Server 2014 record by more than a 2:1 margin. Both the Lenovo and HPE servers are 4-socket units; the Lenovo has more cores per socket — a benefit of the recent advancements in CPU engineering. Two amazing outcomes: SQL Server 2016 performance improved by 82 percent, and the QphH[9] (query-per-hour) price/ performance measure dropped by more than half, from $1.82 to $0.89 USD.

For a more detailed drill-down of hardware performance enhancements and how SQL Server 2016 takes advantage of these enhancements for greater performance, 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," which includes a link to a research paper from Intel.

# In-memory Performance Improvements

The SQL Server 2016 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 — and enhanced data compression.

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 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 2016 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 2016 has the capability to add (B-Tree) indices to columnstore-based tables; this enables efficient single-row lookups.

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

- Massive compression, as a result of 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.

---

[9] http://www.dba-oracle.com/t_tpc_benchmarks.htm

- 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 2016, 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[10]:

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 30 Fast Track 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 https://www.microsoft.com/en-us/cloud-platform/data-warehouse-fast-track

---

[10] https://www.microsoft.com/en-us/cloud-platform/data-warehouse-fast-track

# Appendix A:
## SQL Server 2016, Most Secure Server Ever

| Always Encrypted | Dynamic Data Masking | Row-Level Security |
|---|---|---|
| **Supports regulatory compliance.** | **Supports privacy standards and regulatory compliance.** | **Supports privacy standards and regulatory compliance.** |
| Sensitive data is encrypted in the database at all times. | Sensitive data can be masked from non-privileged users. | Control read/write to specific rows of data based on flexible access criteria. |
| Encryption-decryption done transparently within the client driver. | Data masked on-the-fly; completely application-transparent. | Access logic stored within database; minimize "data leakage." |
| Basic equality queries supported. | | Works transparently at query time; no application changes needed. |

Privacy and security features have been enhanced in Microsoft® SQL Server 2016. Built into the engine are the following features:

### Always Encrypted

Protecting data at rest and in motion has been a challenge. In the past, data could be encrypted or decrypted in one of two places:

1. In the database — but that means the DBA team had access to the plaintext.

2. In the application — in which case the developers would be deploying their own type of encryption code.

Both scenarios gave security professionals nightmares.

SQL Server 2016 addresses this concern with Always Encrypted, which ensures that

1. Sensitive data is encrypted in the database at all times.

2. Encryption and decryption are done transparently within the client driver (ADO.NET, JDBC, ODBC — some small application changes will be necessary)

3. Basic equality queries (JOIN, GROUP BY, DISTINCT, equality searches) are supported on encrypted data (more complex queries will require decryption prior to query execution).

Some limitations and restrictions still apply (i.e., replication may require decryption, under certain circumstances; range queries will need the data decrypted in order to run; in order to show analytics and some reports on-screen, apps will have to decrypt the data first), but SQL Server 2016 Always Encrypted allows companies to meet privacy standards and achieve full regulatory compliance.

**Dynamic Data Masking**

Obfuscation of specific bits of data (SSNs, birthdays, account numbers, etc.) has always been a challenge, but SQL Server 2016 has risen to this challenge with Dynamic Data Masking technology. Companies that work with healthcare or financial data are very concerned about exposing sensitive and personally identifiable data, especially when restoring production databases into a development or quality assurance environment. The accepted solution prior to SQL Server 2016 was to scramble the data immediately following a restore. However, that added time and complexity to the restore process, often resulting in nonsensical data and making application development and QA that much more difficult.

SQL Server 2016 Dynamic Data Masking can hide sensitive data from non-privileged users by employing built-in or custom masking rules. Data can be masked on-the-fly while the underlying data remains untouched. The process is completely transparent to the application or query, no changes needed. SQL Server 2016 Dynamic Data Masking allows companies to meet privacy standards and achieve regulatory compliance.

**Row-Level Security**

Not every user needs to have (or should have) access to all the data in the database. 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 2016, Microsoft moved the onerous task of limiting row access from the application or the DBA to the engine. All it takes is this straightforward 1-2-3 approach:

1. Create a security predicate function to do the security check.

2. Create a security policy on a table that points to your new security function.

3. Enforce that security policy based on whoever's logged in — without changing queries.

Microsoft has given organizations a solution that controls reads and writes to the data, based on flexible access criteria (user identity, role/group memberships, connection data, time of day, etc.), and a solution that meets privacy standards and regulatory compliance. This solution works transparently (no application or query changes needed), and the logic is stored in one central repository — within the database, making maintenance of access policy consistent and easy. SQL Server 2016 Row-Level Security allows companies to meet privacy standards and achieve regulatory compliance.

# Appendix B:
## Make Smarter Decisions Faster: In-memory Built in

A key SQL Server innovation is in-memory processing capabilities. Microsoft introduces real-time Operational Analytics used with the in-memory columnstore technology for operational OLTP systems in SQL Server 2016. In the past, in order 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 a few hours. So being able to easily get real-time or even near-real-time analytics for the operational database was extremely difficult.
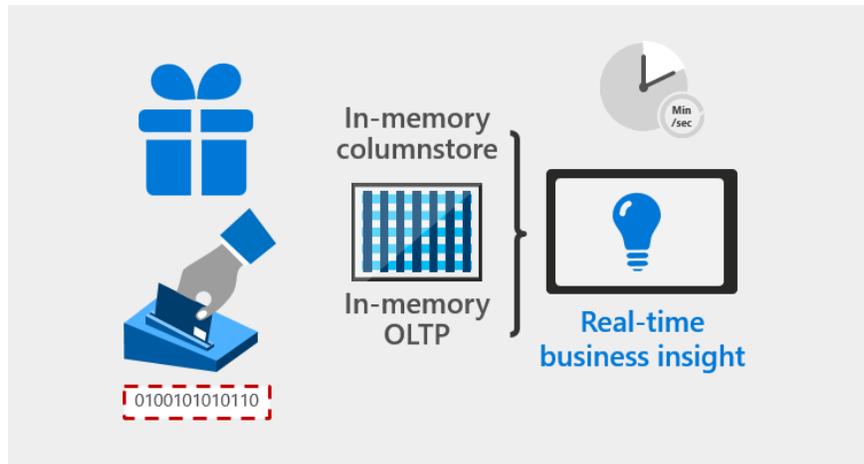


To get to real-time in-memory operational analytics (HTAP) involves building a nonclustered columnstore index in memory on a traditional table of the OLTP database. This technology, introduced to the SQL Server 2016 OLTP environment, gives many customers the ability to efficiently do analytical queries 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 retail operation, and it's now 3:00 PM. You have a one-hour window in which to place your inventory reorder. That's not nearly enough time to offload today's sales transactions to the data warehouse and run the inventory reorder reports — at one time it was, but your company has grown so much that the window is now not long enough. So what do you do?

On your operational SQL Server instance, you do have headroom with both memory and CPU, so you create an in-memory nonclustered columnstore index on the sales table from the OLTP database. All you want to know is what inventory you sold today, and how much to reorder for tomorrow.

With SQL Server 2016, this merging of technologies allows you to extract fast-changing information, using analytical queries against the

---

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

transactional database, without having to port (using ETL) the OLTP data into a data warehouse and then run analytical queries against the data warehouse. Within a matter of minutes, you can extract real-time insight from your operational platform.

# Appendix C:
## Resource Changes for Data Warehouse Fast Track, SQL Server 2014 to 2016

**Minimum Transaction Log Space**

**SQL Server 2014:**
400 GB or 1.2% of rated capacity

**SQL Server 2016:**
400 GB or 2% of rated capacity

**TempDB size**

**SQL Server 2014:**
25%-30%

**SQL Server 2016:**
No change

**Minimum memory**

**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: 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 2016 for on-premises data warehouse operations. These are reflected in the 2016 Fast Track Reference Architecture. The rationale for the increase: 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 2016
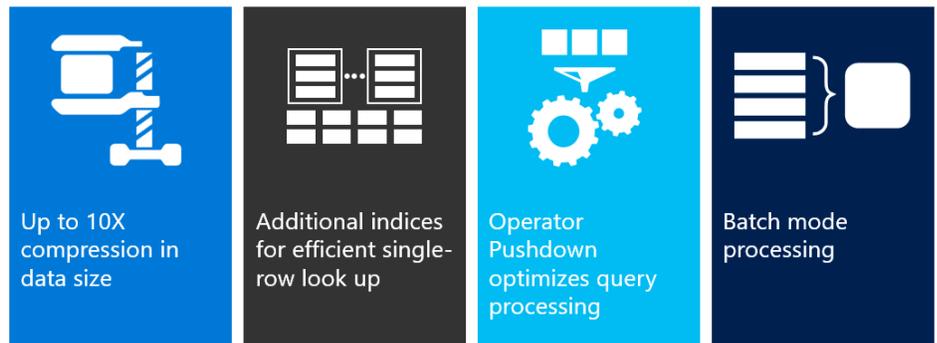
The white paper detailing the on-premises installation can be found at http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/microsoft-sql-database-analytics-paper.pdf.

| 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 2016 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 2016 to deliver stunning performance, processing 100TB worth of data to bring key business insights in a matter of minutes. SQL Server 2016, 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 2016 Columnstore Technology?

| | | | |
|---|---|---|---|
| Up to 10X compression in data size | Additional indices for efficient single-row look up | Operator Pushdown optimizes query processing | Batch mode processing |

Microsoft SQL Server 2016 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:

- **Up to 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:** SQL Server 2016 adds the capability to add additional (B-Tree) indices to columnstore-based tables, which enables efficient single-row lookups.

In addition to these architectural features, SQL Server 2016 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 2016 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.