|  |  |
| --- | --- |
|  |  |
|  | What’s new for Data Warehouse Fast Track for SQL Server 2014 |

|  |
| --- |
| © 2014 Microsoft Corporation. All rights reserved. This document is provided “as-is.” Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes. You may modify this document for your internal, reference purposes. |

Contents

3 Executive summary

4 Part of the Microsoft modern data warehouse

5 Improving DW Fast Track with SQL Server 2014

6 Clustered columnstore indexes

9 Additional SQL Server 2014  
advantages

10 Changes in hardware specification

11 Conclusion

11 Get started today

11 For more information

|  |
| --- |
| Executive summary |

Designing an effective data warehouse solution requires the right configuration of software and hardware, but not every organization has the time, resources, and expertise to develop such a solution. The Data Warehouse Fast Track (DW Fast Track) reference architecture is designed to help you build an optimal data warehouse solution that meets your unique needs.

With a step-by-step guide from Microsoft and its hardware partners, you can build medium to large data warehouse solutions by taking advantage of Microsoft SQL Server 2014 Enterprise in combination with well-tuned configurations from approved hardware partners.

With SQL Server, you can build a data warehouse solution that delivers mission-critical performance by achieving faster transactions—up to 30 times faster—with in-memory online transaction processing (OLTP), in addition to enhanced security, scalability, and high availability. You can also gain faster insights into virtually any data through enhanced self-service business intelligence (BI) with familiar tools like Microsoft Excel. SQL Server offers a platform for hybrid cloud by providing a common set of tools across the entire application lifecycle. Finally, SQL Server 2014, using In-Memory Columnstore, currently holds the world record TPC-H benchmark in price and performance for data warehousing.[[1]](#footnote-1)

Through extensive co-engineering and testing, Microsoft supports qualified partners in the development and validation of data warehouse reference architectures that achieve high performance at a cost-effective price. DW Fast Track delivers one of the most balanced configurations of CPU, IO, and storage resources through the use of templates, tools, and tests shared with partners that are audited and validated by Microsoft. Ultimately, the reference architecture can help you reduce the time and cost of setting up data warehouse workloads, while allowing you to work with a hardware vendor of your choice.

This white paper is intended for database architects and database administrators. It explains how DW Fast Track fits within the Microsoft modern data warehouse vision, while providing an overview of the improvements that SQL Server 2014 brings to the DW Fast Track offering, which can deliver up to 10 times faster queries with up to 15 times better disk compression. Finally, this paper covers changes in the reference configurations from the Data Warehouse Fast Track for SQL Server 2012 implementation, which provide more memory and processing power for your data warehouse workloads.

|  |
| --- |
| Part of the Microsoft modern  data warehouse |

The Microsoft modern data warehouse delivers a complete logical data and analytics platform with a suite of fully supported solutions and technologies. It is designed to meet the needs of even the most demanding modern enterprise—on-premises, in the cloud, or within nearly any hybrid scenario (Figure 1).

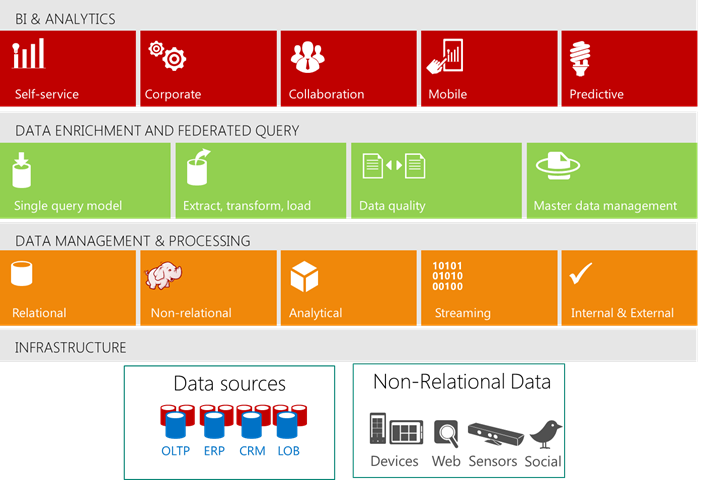


Figure 1: Layered framework of the modern data warehouse

### Data management and processing

A modern data warehouse starts with the ability to handle both relational and non-relational (like Hadoop) data sources seamlessly. It can manage data in real time using complex event processing (CEP) technologies, as well as easily integrate internal and external data. In addition, a modern data warehouse provides an analytics engine for predictive analysis and interactive exploration of aggregated data from different perspectives.

### Data enrichment and federated query

A modern data warehouse can enrich data with extract, transform, and load (ETL) capabilities; it also can better ensure credible and consistent data through support resources like Microsoft Data Quality Services and Master Data Services. Plus, a modern data warehouse provides a single query mechanism across different types of data through a federated query service.

### Business intelligence and analytics

A modern data warehouse must support the breadth of tools available to help organizations act on data. This includes familiar self-service tools that simplify data analysis for business users and help them create and share analytics in a team environment across a variety of devices. Further, some companies need tools to help operationalize self-service solutions for broader use, as well as a platform that supports predictive analytical models to assist with real-time decision making.

### Fitting within the modern data warehouse

The Data Warehouse Fast Track for SQL Server 2014 offering is a reference architecture and set of prescriptive guides delivered through Fast Track partners, which for the previous version included HP, IBM, Dell, Cisco, Hitachi Data Systems, X-IO Storage, Violin-Fujitsu, NEC, Huawei, Fujitsu–NetApp, and EMC. This single server symmetric multiprocessing (SMP) based architecture allows SQL Server 2014 to take advantage of all the cores within the server to distribute queries across the cores for greater performance. SQL Server 2014 can also take advantage of shared server memory to cache data into memory to improve performance by reducing read and write IO operations to disk. DW Fast Track provides guidance to vendors for using solid-state memory storage devices for additional optimizing of IO operations when data cannot fit completely into memory.

To integrate DW Fast Track with Hadoop, you need to use ETL processes or tools like Apache Sqoop to import and export Hadoop data.

Since DW Fast Track is built on SQL Server 2014 Enterprise Edition, all of the data enrichment and business intelligence tools just work.

|  |
| --- |
| Improving DW Fast Track with SQL Server 2014 |

SQL Server 2014 introduces new features for data warehouse workloads, improving on the advances made in the SQL Server 2012 release. These new features can provide greater performance, greater compression of data on disk and within memory, greater availability with online operations, and greater control with resource governor. SQL Server 2014 provides greater capabilities for the DW Fast Track reference architecture as shown in Figure 2.

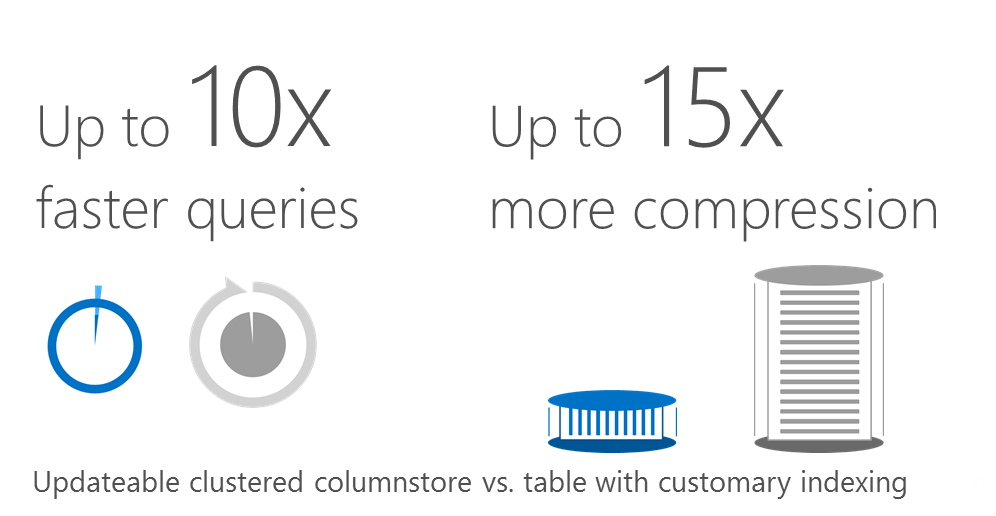


Figure 2: Improved performance and greater compression with In-Memory Clustered Columnstore

|  |
| --- |
| Clustered columnstore indexes |

The SQL Server In-Memory Columnstore index stores and manages data by using column-based data storage and column-based query processing. Columnstore indexes give high performance gains for queries that use full table scans, and are not well-suited for queries that search for a particular value in the data.

### Clustered columnstore index becomes the table

When you create a clustered columnstore index on a table, SQL Server transforms the physical layout of the table from rowstore to columnstore format. You get greater compression over rowstore page compression. Additionally, you do not have an extra index on the table like you have with a non-clustered columnstore index. Here are some of the benefits you can get with clustered columnstore indexed tables:

* Columns often have similar data, which results in high compression rates. This is especially true for fact tables in a star schema.
* High compression rates improve query performance by using a smaller in-memory footprint. In turn, query performance can improve because SQL Server can perform more query and data operations in-memory.
* A new query execution mechanism called batch-mode execution has been added to SQL Server that reduces CPU usage by a large amount. Batch-mode execution is closely integrated with, and optimized around, the columnstore storage format. Batch-mode execution is sometimes known as *vector-based* or *vectorized* *execution*.
* Queries often select only a few columns from a table, which reduces total I/O from the physical media.

SQL Server provides even greater compression for less frequently accessed partitions with the COLUMNSTORE\_ARCHIVE feature. It comes with a trade-off of slower query performance for an additional 37 percent to 66 percent reduction depending on the data.[[2]](#footnote-2) Nasdaq, as the case study below reveals, was able to take advantage of clustered columnstore indexes.

Case Study: NASDAQ OMX

NASDAQ OMX Group Reduces 2 PB of Data to 500 TB with Microsoft In-Memory Technology.[[3]](#footnote-3)

*“The database itself has a smaller storage footprint on disk due to in-memory columnstore, and encrypted backups reduce the need for swap space on disk used by third-party tools to encrypt data before it’s moved to storage…SQL Server 2014 is a game changer for us. Its in-memory columnstore, advanced statistics, and cardinality estimation are the tools we need to manage our very large databases for the long term.” – Stanley Orlowski, Director of Database Structures, NASDAQ OMX*

### In-Memory Columnstore performance

In a traditional data warehouse, data is stored in records or rowstores. However, rowstores are not optimal for many star schema-based queries. Columnstore technology on fact tables within a star schema improves query performance for large tables by reducing the amount of data that needs to be processed through I/O.

In-Memory Columnstore changes the primary storage engine to an updateable and indexed in-memory columnar format, which groups, stores, and indexes data in compressed column segments. In-Memory Columnstore improves query performance over traditional data warehouses because only the columns needed for the query must be read. Also, columns are heavily compressed, reducing the number of bytes to be read or moved. In addition, In-Memory Columnstore maximizes the use of the CPU by accessing data held in-memory when processing a query.

Combined, these techniques result in compression (up to 15 times), as well as performance gains (up to 10 times). In-Memory Columnstore can improve query performance, even based on existing hardware investments as discovered by Nexon.

Case Study: Nexon

Online gaming company Nexon found that clustered columnstore indexes improved query performance for large fact tables from 4 minutes to 9 seconds.[[4]](#footnote-4)

*“Data analysts change query patterns in this way and that, and work with all columns because they want to size up all the possibilities from various angles and reason out the right solution, rather than just sticking within narrow confines. Clustered columnstore indexes (CCI) enables them to look into all columns and find the answer in a faster manner. CCI will thus help data analysts work easier and faster without having to contact DB managers about the problem of slow queries and/or tune queries when handling large volumes of data.” – NEXON member*

SQL Server 2014 delivers exceptional in-memory performance as recent TPC-H 10 terabyte and 3 terabyte benchmarks reveal. For example, the combination of SQL Server 2014 using clustered columnstore indexes and Windows Server 2012 R2 Standard Edition running on an HP ProLiant DL580 Gen8 Server has achieved record-breaking results in data warehousing price and performance. At 10 terabytes, the Microsoft and HP configuration achieved a world record non-clustered performance of 404,005 query-per-hour (QphH), with the SQL Server implementation soaring past the previous price/performance record to reach $2.34 USD dollars per QphH. At 3 terabytes, it achieved a world record non-clustered performance of 461,837 QphH, with the SQL Server implementation reaching $2.04 USD dollars per QphH.[[5]](#footnote-5)

### Updateable tables with clustered columnstore index

SQL Server 2014 improves the non-clustered columnstore technology introduced with SQL Server 2012 by enabling insert, update, and delete operations against a table with a clustered columnstore index. To improve performance of update actions, SQL Server 2014 implements a deltastore architecture that uses traditional temporary B-tree rowstore objects. Once the deltastore is ready for insert, SQL Server periodically processes the deltastores in the background and converts them into columnar storage format. When queries are made against the table, SQL Server automatically performs a union query of the columnstore table with any unprocessed deltastores on the table.

Bulk insert operations that process between 102,400 and 1,048,576 rows are loaded directly into the columnstore table. For more information on how SQL Server inserts, updates, and deletes records, see the help topic [Using Clustered Columnstore Indexes](http://msdn.microsoft.com/en-us/library/dn589807(v=sql.120).aspx).

|  |
| --- |
| Additional SQL Server 2014 advantages |

There are additional enhancements in SQL Server 2014 that can improve the performance and the availability of your data warehouse workloads.

### Query processor cardinality estimation updates

The cardinality estimation logic, called the *cardinality estimator*, is re-designed in SQL Server 2014 to improve the quality of query plans and query performance. The new cardinality estimator incorporates assumptions and algorithms that work well on modern online transaction processing (OLTP) and data warehousing workloads. It is based on in-depth cardinality estimation research on modern workloads as well as Microsoft’s learnings over the past 15 years of improving the SQL Server cardinality estimator. Feedback from customers shows that while most queries benefit from the change or remain unchanged, a small number might show regressions compared to the previous cardinality estimator.

### Managing the lock priority of online operations

The ONLINE = ON option now contains a WAIT\_AT\_LOW\_PRIORITY option which permits you to specify how long the rebuild process should wait for the necessary locks. The WAIT\_AT\_LOW\_PRIORITY option also allows you to configure the termination of blocking processes related to that rebuild statement. For more information, see [ALTER TABLE (Transact-SQL)](http://msdn.microsoft.com/en-us/library/ms190273(v=sql.120).aspx) and [ALTER INDEX (Transact-SQL)](http://msdn.microsoft.com/en-us/library/ms188388(v=sql.120).aspx). Troubleshooting information about new types of lock states is available in [sys.dm\_tran\_locks (Transact-SQL)](http://msdn.microsoft.com/en-us/library/ms190345(v=sql.120).aspx) and [sys.dm\_os\_wait\_stats (Transact-SQL)](http://msdn.microsoft.com/en-us/library/ms179984(v=sql.120).aspx).

### Resource Governor enhancements for physical IO control

The Resource Governor enables you to specify limits on the amount of CPU, physical IO, and memory that incoming application requests can use within a resource pool. In SQL Server 2014, you can use the new MIN\_IOPS\_PER\_VOLUME and MAX\_IOPS\_PER\_VOLUME settings to control the physical IOs issued for user threads for a given resource pool. For more information, see [Resource Governor Resource Pool](http://msdn.microsoft.com/en-us/library/hh510189(v=sql.120).aspx) and [CREATE RESOURCE POOL (Transact-SQL)](http://msdn.microsoft.com/en-us/library/bb895329(v=sql.120).aspx).

The MAX\_OUTSTANDING\_IO\_PER\_VOLUME setting of the ALTER RESOURCE GOVENOR sets the maximum outstanding I/O operations per disk volume. You can use this setting to tune IO resource governance to the IO characteristics of a disk volume and to limit the number of IOs issued at the SQL Server instance boundary. For more information, see [ALTER RESOURCE GOVERNOR (Transact-SQL)](http://msdn.microsoft.com/en-us/library/bb934013(v=sql.120).aspx).

### Incremental statistics

A major problem with updating statistics in large tables in SQL Server is that the entire table always has to be scanned (for example, when using the WITH FULLSCAN option), even if only recent data has changed. This is also true with partitioning: even if only the newest partition changed since the last statistics update, the entire table, including all the partitions that did not change, has to be scanned again.

Using SQL Server 2014 incremental statistics you can update only the partition or partitions that you need, and the information on these partitions will be merged with the existing information to create the final statistics object. Another advantage of incremental statistics is that the percentage of data changes required to trigger an automatic update of statistics now works at the partition level, which basically means that now only 20 percent of rows changed (changes on the leading statistics column) per partition are required. Related statements allow or report incremental statistics. Affected syntax includes UPDATE STATISTICS, sp\_createstats, CREATE INDEX, ALTER INDEX, ALTER DATABASE SET options, DATABASEPROPERTYEX, sys.databases, and sys.stats. For more information, see [CREATE STATISTICS (Transact-SQL)](http://msdn.microsoft.com/en-us/library/ms188038(v=sql.120).aspx).

|  |
| --- |
| Changes in hardware specification |

Microsoft and its hardware partners realize that the price of memory is dropping with 16-GB dual in-line memory modules (DIMMs) becoming the standard and that 2-socket server performance has doubled over the last two years with 50 percent more cores. As such, the DW Fast Track reference architecture needs to account for these advances.

Table 1 shows the updated minimum memory recommendations based on the number of sockets in the server configuration.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SQL Server version | 1-socket | 2-socket | 4-socket | 8-socket |
| 2012 | 64 | 128 | 256 | 512 |
| 2014 | 128 | 256 | 512 | 1024 |

Table 1: Minimum memory recommendations

SQL Server 2014 in-memory updateable columnstore technology will be able to take advantage of the additional memory for even better performance of your data warehouse workloads.

By using clustered columnstore indexes on tables as the default configuration, the new DW Fast Track specification compression ratio will change from 3.5 to 5.

|  |
| --- |
| Conclusion |

The Data Warehouse Fast Track for SQL Server 2014 reference architecture provides an alternative to a build-it-yourself data warehouse approach, and it is delivered by a wide variety of hardware partners for a faster time to solution. Innovations within SQL Server 2014, such as In-memory updateable clustered columnstore, provide up to 10 times faster query performance over traditional rowstore tables and up to 15 times better compression of data depending on your data.

DW Fast Track fits within the Microsoft modern data warehouse vision by offering a standalone data warehousing platform or the option to integrate, as part of a larger enterprise Big Data strategy, with the Microsoft Analytics Platform System as part of a well-designed hub and spoke architecture.

|  |
| --- |
| Get started today |

Get started building your modern data warehouse today by completing the following:

1. Learn more about Fast Track data warehousing at <http://www.microsoft.com/sqlserver/en/us/solutions-technologies/data-warehousing/fast-track.aspx>.
2. Download the trial version of SQL Server 2014 at <http://www.microsoft.com/en-us/sqlserver/sql-server-2014.aspx>.
3. Sign up for a free architectural design session for data warehousing with your Microsoft representative.

|  |
| --- |
| For more information |

Explore the following resources for more information on DW Fast Track for SQL Server 2014 and supporting technologies:

SQL Server: <http://www.microsoft.com/sqlserver/>

SQL Server TechCenter: <http://technet.microsoft.com/en-us/sqlserver/>

SQL Server Developer Center: <http://msdn.microsoft.com/en-us/sqlserver/>

Microsoft modern data warehouse: <http://www.microsoft.com/sqlserver/en/us/solutions-technologies/data-warehousing.aspx>

Microsoft Analytics Platform System: <http://www.microsoft.com/>aps

Microsoft Big Data solution: <http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence/big-data-solution.aspx>

1. SQL Server 2014 and HP Sets Two World Records for Data Warehousing Leading In Both Performance and Price/Performance, <http://blogs.technet.com/b/dataplatforminsider/archive/2014/04/16/sql-server-2014-and-hp-sets-two-world-records-for-data-warehousing-leading-in-both-performance-and-price-performance.aspx>, April 2014 [↑](#footnote-ref-1)
2. Microsoft Research, Enhancements to SQL Server Column Stores, <http://research.microsoft.com/pubs/193599/Apollo3%20-%20Sigmod%202013%20-%20final.pdf>, June 2013 [↑](#footnote-ref-2)
3. Microsoft Case Studies*, NASDAQ OMX Group Reduces 2 PB of Data to 500 TB with Microsoft In-Memory Technolog*y, <http://www.microsoft.com/casestudies/Microsoft-SQL-Server-2014-Enterprise/NASDAQ-OMX/NASDAQ-OMX-Group-Reduces-2-PB-of-Data-to-500-TB-with-Microsoft-In-Memory-Technology/710000004283>, April 2014 [↑](#footnote-ref-3)
4. Microsoft Case Studies, *A new vision for DW in the age of Big Data: Clustered Columnstore Index*, <http://www.microsoft.com/casestudies/Case_Study_Detail.aspx?CaseStudyID=710000004340>, April 2014 [↑](#footnote-ref-4)
5. SQL Server 2014 and HP Sets Two World Records for Data Warehousing Leading In Both Performance and Price/Performance, <http://blogs.technet.com/b/dataplatforminsider/archive/2014/04/16/sql-server-2014-and-hp-sets-two-world-records-for-data-warehousing-leading-in-both-performance-and-price-performance.aspx>, April 2014 [↑](#footnote-ref-5)