



# Deeper Insights across Data

Technical White Paper

**Published:** June 2015

**Applies to:** SQL Server 2016

**Summary:** Data warehousing, analytics, and business intelligence must adapt to a whole new scope, scale, and diversity of information and data. They also must continue to embrace the new ways that we discover and collaborate on information every day. Today, data comes from relational and non-relational sources, from on-premises environments and the cloud, and from Big Data and other sources. Microsoft SQL Server can help you to access, integrate, and store this data. It also can help you to scale and manage it, as well as gain deeper insights from it.

# Copyright

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

This white paper is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2015 Microsoft Corporation. All rights reserved.

Microsoft, Active Directory, Microsoft Azure, Bing, Excel, SharePoint, Silverlight, SQL Server, Visual Studio, Windows, and Windows Server, are trademarks of the Microsoft group of companies.

All other trademarks are property of their respective owners.

# Contents

SQL Server evolution .....	4
Deeper insights across data with SQL Server .....	6
Access any data .....	6
SQL Server FileStream .....	6
Seamless connection and analysis of Big Data through Hadoop connectors .....	7
PolyBase .....	7
Azure Data Factory .....	9
Integration Services .....	9
Scale and manage .....	10
Microsoft Big Data solution .....	10
Massive data warehousing.....	10
Analytics Platform System .....	11
In-memory data warehousing (DW): Columnstore index.....	11
Azure HDInsight and Apache projects .....	13
Analysis Services .....	13
Master Data Services.....	14
Powerful insights .....	15
Power BI.....	15
SQL Server Data and BI Tools in Visual Studio 2015 .....	16
Azure Stream Analytics.....	16
Azure Machine Learning .....	17
R integration .....	17
Reporting Services .....	18
Mobile BI .....	19
Conclusion .....	19
More information .....	19

## SQL Server evolution

The increasing volume, velocity, diversity, and locations of enterprise data make it increasingly challenging to discover, connect to, move, transform, integrate, and analyze it all. In fact, with the compound annual growth rate of data from 2013 to 2020 estimated at 41 percent<sup>1</sup>, this can be termed a data explosion. This is the result of more and more devices combined with a new hunger on the part of businesses for more data to better understand and predict their customers' needs to make smarter decisions with data.

Yet without the ability to analyze this data, the data alone are not valuable. Microsoft SQL Server and the related Microsoft data platform can deliver the tools you need to gain deeper insights from all of your data. Our customers and industry analysts agree that SQL Server is a leader here (Figures 1 and 2).

Figure 1: Gartner Magic Quadrant for data warehouse and data management solutions for analytics<sup>2</sup>



<sup>1</sup> IDC, "Digital Universe," December 2012.

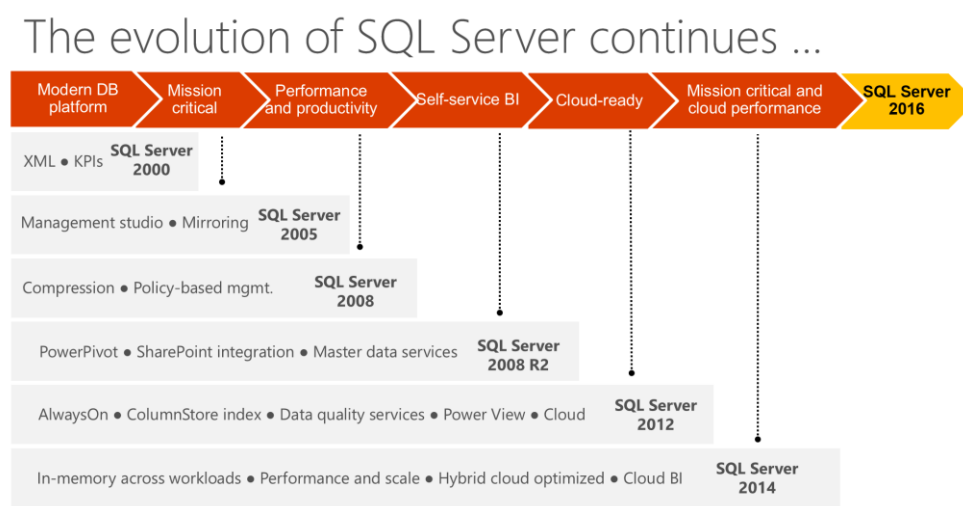
<sup>2</sup> Gartner, <http://www.gartner.com/technology/reprints.do?id=1-29WVKZC&ct=150213&st=sb>, February 2015.

Figure 2: Gartner Magic Quadrant for business intelligence and analytics platforms<sup>3</sup>



SQL Server has been evolving along with the explosion in data sources and continues to innovate to facilitate the management of data (Figure 3).

Figure 3: Major SQL Server functionalities added across releases



SQL Server 2016 introduces many new features and enhancements, including:

- PolyBase technology allows you to query relational SQL Server and Hadoop data through a single T-SQL query;

<sup>3</sup> Gartner, <http://www.gartner.com/technology/reprints.do?id=1-2AELNBP&ct=150223&st=sb>, February 2015.

- A number of enhancements to Analysis Services, including enterprise grade tabular models, more efficient parallel processing, and in-memory data management;
- Additional enhancements to Reporting Services, including support for modern browsers and devices, updated support for a broad array of data sources, new chart types, and more;
- SQL Server development tools in Visual Studio for building BI models and Analysis Services models and Reporting Services reports;
- Integration of the R language, bringing predictive analytic capabilities to your relational database;
- Reporting for mobile.

## Deeper insights across data with SQL Server

Big Data, both on the cloud and on-premises; new types of non-relational data; and the continuing importance of data warehouses and transactional systems have been some of the key trends that are impacting how we are designing the Microsoft data platform.

In this technical white paper, we will examine these issues in the context of accessing more types of data, scaling resources to handle the increasing volume, and using these data to provide powerful insights for your business.

### Access any data

Data can be in many formats and locations, whether relational or non-relational, on-premises or on the cloud, and Big Data or traditional tabular data. SQL Server has a comprehensive set of tools to make accessing all of these data easy.

### SQL Server FileStream

#### Embracing unstructured data through support for complex data types

With the dramatic shift from structured to complex data types, the new mission critical requires support for complex data types to be built-in, not added on at extra cost and effort. SQL Server supports a growing number of types and volumes of complex data with FILESTREAM, Remote BLOB Storage (RBS), and spatial support—enhancements on top of the already robust and built-in foundation that extend beyond relational capabilities. This support allows organizations to build rich and innovative applications without paying extra premiums.

SQL Server FileTable builds on FILESTREAM to bring Win32 namespace support and application compatibility to the file data stored in SQL Server. Countless applications maintain their data in two “worlds”: *unstructured* (documents, media files, and other unstructured data in file servers) and *structured* (related, structured metadata in relational systems). FileTable lowers the barrier to entry for organizations that have files on servers that currently run Win32 applications, while reducing the effort caused by maintaining two disparate systems and keeping them in sync.

#### High availability for unstructured data

With SQL Server, complex data types are handled with the same attention as common data types. Organizations can use FILESTREAM to store and manage complex data in a variety of ways, as if it were part of the database. Additionally, with SQL Server, organizations can enjoy the high-availability benefits of AlwaysOn for complex data managed through FILESTREAM—even when they take advantage of RBS and FileTable.

## Seamless connection and analysis of Big Data through Hadoop connectors

Hadoop connectors for SQL Server and Analytics Platform System (APS) are available for download to organizations that have licenses for SQL Server and APS. These connectors enable bidirectional data movement across SQL Server and Hadoop, so users can work effectively with both structured and unstructured data. Additionally, people can use the leading business intelligence (BI) platform from Microsoft to perform analysis on Hadoop data sets. Users can gain access to and create mashups of Hadoop data sets by using familiar productivity tools such as Microsoft Excel and award-winning BI clients such as PowerPivot and Power View to perform analysis in an immersive and interactive way.

### New in SQL Server 2016:

#### PolyBase

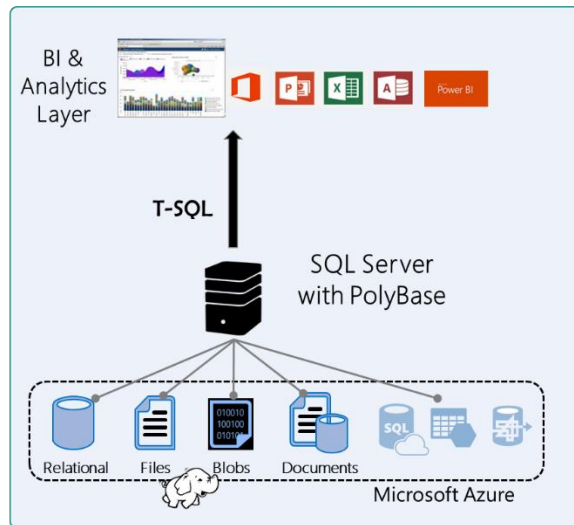
The reality of modern data warehousing is complex. Monolithic single stores for the enterprise's data are becoming ever rarer. Instead, it is increasingly likely that enterprises will have multiple relational databases, Hadoop data, document-oriented NoSQL databases, and so on.

PolyBase allows users to query non-relational data in Hadoop, blobs, files, and combine it on-the-fly with their existing relational data in SQL Server. It also provides the option for users to import Hadoop data for persistent storage in SQL Server and to export aged relational data into Hadoop. Similar functionality has been available in Microsoft Analytics Platform System and this has now been incorporated into the SQL Server database engine.

PolyBase also provides the ability to access and query data that is either on-premises or on the cloud and run analytics and BI on that data. SQL Server 2016 and PolyBase therefore can help you to build out a hybrid solution that delivers deeper insights on your data, wherever it may be located.

While PolyBase does allow you to move data in a hybrid scenario, it is also common to leave the data where it resides and query it wherever it may be (Figure 4). This ties into the concept of a data lake. A data lake can be thought of as providing full access to raw Big Data without moving it. This may be viewed as an alternate approach to processing the Big Data to make its analysis easier, then moving/synchronizing it in a data warehouse. The advantages of not moving the data are many. It generally means that beyond setting up the connectivity in the data lake, you are not developing anything new, so it can be more cost effective. Also there may be organizational limits to moving or modifying the data that become irrelevant with this approach. Finally, data processing and synchronization can be complex operations, and you may not know in advance how to process the data to deliver the best insights. SQL Server 2016 and PolyBase can be an important component in setting up a data lake, combining it with your relational data, and doing analysis and BI on it.

Figure 4: Using PolyBase to query data in a hybrid scenario



After installing PolyBase in SQL Server (by adding an instance feature), you can begin working with PolyBase inside of Visual Studio (VS) by connecting to SQL Server by using the SQL Server Object Explorer in VS.

Working with external data is generally a three-step process:

1. Defining the external data sources, such as those on Windows Azure Blob Storage, or in Hadoop (either in Microsoft HDInsight or external Hadoop); passing in security credentials.
2. Defining the external file/data format (e.g. delimited text files, Hive RC, Hive ORC, or Parquet).
3. Defining an external table. Supported data definition language (DDL) statements include CREATE/DROP, ALTER, and CREATE STATISTICS.

Once PolyBase and the external tables are set up, you can begin querying the data. You can join tables in SQL Server with external tables and use a mature set of T-SQL statements for querying. You can also execute split-based query processing, called “push-down computation” using new query hints. Push-down computation allows you push processing to the source level, as shown here.

```
SELECT DISTINCT C.FirstName, C.LastName, C.MaritalStatus
FROM Insurance_Customer_SQL -- table in SQL Server
INNER JOIN ( -- external tables referring to data in
    --2 HDP Hadoop clusters
    SELECT * FROM SensorData_ExternalHDP WHERE Speed > 35
    UNION ALL
    SELECT * FROM SensorData_ExternalHDP2 WHERE Speed > 35
) AS SensorD
ON C.CustomerKey = SensorD.CustomerKey
OPTION (FORCE EXTERNALPUSHDOWN) -- push-down computation
```

PolyBase can be used with Microsoft BI tools as a data source and it can be used by many third-party BI tools as well, such as Tableau, Cognos, and the like. PolyBase is also integrated with AlwaysOn and



failover. You can also scale out PolyBase by adding multiple SQL Server 2016 instances to a PolyBase farm.

### **Improved in SQL Server 2016:**

#### **Azure Data Factory**

Azure Data Factory (ADF) enables you to process on-premises data like SQL Server, together with cloud data like Azure SQL Database, blobs, and tables. These data sources can be composed, processed, and monitored through simple, highly available, fault-tolerant data pipelines. ADF supports Hive, Pig, and C# processing, along with key processing features such as automatic Hadoop (HDInsight) cluster management, re-tries for transient failures, configurable timeout policies, and alerting. ADF helps you quickly assess end-to-end data pipeline health, pinpoint issues, and take corrective action if needed. You can also use data pipelines to deliver transformed data from the cloud to on-premises sources like SQL Server or keep it in your cloud storage sources for consumption by BI and analytics tools, including Azure Machine Learning and other applications.

#### **Integration Services**

Microsoft Integration Services is a platform for building enterprise-level data integration and data transformations solutions. You use Integration Services to solve complex business problems by copying or downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data. The packages can work alone or in concert with other packages to address complex business needs. Integration Services can extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations.

### **Improved in SQL Server 2016:**

SQL Server 2016 contains a number of enhancements that can improve the development, management, and monitoring of your SQL Server Integration Services (SSIS) data packages and delivers a number of benefits to your on-premises and on-the-cloud operations in the areas of cloud integration, connectivity improvements, and product improvements.

#### **Cloud integration**

Azure Data Factory (ADF) can now orchestrate on-premises SSIS execution. Further, SSIS can read from ADF as a data source via the ADF data flow task. SSIS developers can also now leverage the Azure Storage Connector to move data from on-premises to Azure Storage, or vice versa. SSIS developers can also trigger Azure HDInsight jobs directly from SSIS, so they can better integrate with HDInsight to process data already in the cloud without needing to move the unprocessed cloud data on-premises. SSIS developers can manage and control other Azure services/assets generically within SSIS pipeline by using an Azure commandlet task to issue a commandlet to Azure directly.

#### **Connectivity improvements**

In addition to the ADF and HDInsights connectors mentioned above, SQL Server 2016 also has a wide range of new and enhanced data connectors including: OData Version 4, Hadoop File System (HDFS), JavaScript Object Notation (JSON), and Oracle/Teradata connector V4 by Attunity. **Power Query can also be used as a data source.**

## Product improvement

SQL Server 2016 makes a number of significant enhancements to SSIS usability. For example, the package designer itself has been improved with many enhancements in the areas of resizing, dragging and dropping, and other features. SQL Server 2016 also supports package templates to facilitate code reuse and package setup. Also, SSIS packages can now be deployed incrementally, rather than having to deploy the entire project. Custom logging levels can also be configured on top of the current log level. Developers can easily and fully upgrade their projects to SQL Server 2016 SSIS compatibility with a click of a button. Further, you can now easily configure high-availability AlwaysOn for the SSIS catalog database directly in SQL Server Management Studio (SSMS) without the need to set it up manually.

## Scale and manage

With a diverse data portfolio generated in unprecedented volumes, it is increasingly important for organizations to have fast, redundantly available global access to data scaling to petabytes and beyond. SQL Server will help you scale and manage your data to give you deeper insights, regardless of the location, volume, velocity, or form of the data.

## Microsoft Big Data solution

Microsoft strategy for Big Data embraces Hadoop for activating ambient data that come into existence outside the traditional data platform. Hadoop is the open-source implementation of MapReduce parallel computation engine and environment, and it is used for processing streams of data that go well beyond the size of even the largest enterprise data sets. Whether the data are from sensors, clickstreams, social media, geographical locations, or are generated and collected in large masses, Hadoop is often in the service of processing and analyzing them.

The Big Data solution from Microsoft enables customers to augment their analysis with publicly available data from social media sites such as Twitter and Facebook and hundreds of trusted data providers on Azure Marketplace. Azure Marketplace also exposes hundreds of applications and data mining algorithms to help organizations unlock new business insights.

To complement Microsoft strategy for Big Data overall, PolyBase offers breakthrough technology on the data processing engine in Microsoft's Analytics Platform System. As noted above, PolyBase is designed to be a simple way to combine non-relational data and traditional relational data for analysis. While organizations would normally burden IT to prepopulate the data warehouse with Hadoop data or would undergo extensive training in MapReduce to query non-relational data, PolyBase does these tasks seamlessly to give users the benefits of Big Data without the complexities.

Also, through deep integration with PowerPivot, Power View, and enterprise data warehouse tools, the Microsoft solution for Big Data offers organizations deep insights into all their structured and unstructured data with the tools they use every day.

## Massive data warehousing

Microsoft provides a range of solutions that help organizations address the challenges of Big Data with its family of data warehouse solutions—SQL Server, SQL Server Fast Track Data Warehouse, and Analytics Platform System—that provide a robust and scalable platform for storing and analyzing data in a traditional data warehouse. SQL Server 2014 provides enhanced features such as RBS and partitioned tables that scale to 15,000 partitions to support large, sliding-window scenarios. (In a *sliding-window scenario*, partitioned tables are managed for efficiency to maintain the same number of partitions over time by adding a new partition to accommodate the newest data and removing the partition that contains

the oldest data.) SQL Server 2014 also has increased support for as many as 640 logical cores to enable high performance for very large workloads and consolidation scenarios.

## Analytics Platform System

The Microsoft Analytics Platform System (APS) is a turnkey Big Data analytics appliance, combining massively parallel processing (MPP) data warehouse technology from Microsoft together with HDInsight, (the 100% Apache Hadoop distribution from Microsoft) and delivering it as a turnkey appliance. To integrate data from SQL Server with data from Hadoop, APS offers the PolyBase data querying technology.

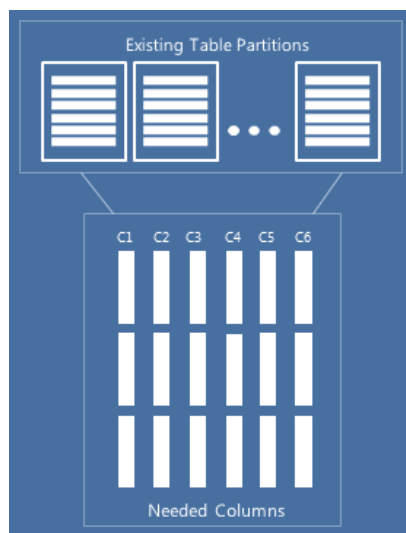
Benefits of APS:

- Up to 100x performance gains over legacy data warehouses
- Relational and non-relational data in one appliance
- Seamless integration of the relational data warehouse and Hadoop with PolyBase
- Linear scale-out to 6 petabytes of user data capacity
- The lowest price per terabyte for a data warehouse appliance in the industry

## In-memory data warehousing (DW): Columnstore index

The SQL Server in-memory columnstore index stores and manages data by using column-based data storage and column-based query processing (Figure 5). Columnstore indexes can transform the DW experience for users by enabling faster performance for common DW queries such as filtering, aggregating, grouping, and star-join queries.

Figure 5: Columnstore index in-memory columnar storage



Columnstore index is built on top of an existing row-level table and provides a view of the data that puts the index on specific columns. It translates the data based on only those required columns and then stores this view, which results in dramatic performance gains. The level of performance gains is dependent on the data and the nature of the query.

Database administrators also can update the clustered columnstore index online to accommodate real-time data warehouse queries without the need to drop and recreate the index. To save disk space, they

can apply a new option called `COLUMNSTORE_ARCHIVE` for higher compression and storage space savings of as much as 90 percent.

Columnar storage format provides significant data compression by storing each column separately. Since the data within a column is similar and often repeated, SQL Server can achieve a very high level of data compression. Higher compression rates improve query performance by using a smaller in-memory footprint.

Analytic queries often select only a few columns from the `FACT` table. With columnar storage, only the required columns are read into memory, reducing I/O even further, unlike row-based storage format where all columns get loaded into memory as part of the rows. Optimized query execution is achieved using techniques such as applying predicates in compressed format, pushing down predicates to storage layer when possible, leveraging new processor architectures, and utilizing a new batch execution mode.

Batch-mode query processing is basically a vector-based query execution mechanism, which is tightly integrated with the columnstore index. Queries that target a columnstore index can use batch-mode to process up to 900 rows together, which enables efficient query execution, providing 3–4x in query performance improvement. In SQL Server, batch-mode processing is optimized for columnstore indexes to take full advantage of their structure and in-memory capabilities.

SQL Server can scale out on a single server since columnstore tables do not need to fit in memory. That is, there is no limitation to fit all of the columnstore indexes in memory since these indexes reside on disk and only the required data is brought into memory on as-needed basis as part of running queries. Therefore, one can have a much larger DW configuration on a single server, like a server with 2 TB memory and 40 TB of data running on it. For even larger data warehouses, Microsoft and Dell have teamed up to offer the Cloud Platform System (CPS) appliance.

Columnstore indexing can be used to achieve up to 100x query-performance gains over traditional row-oriented storage and significant (typically 10x) data compression for common data patterns.

### **Improved in SQL Server 2016:**

#### **Operational analytics**

Traditionally, customers have run operational (online transaction processing) and analytics (DW) workloads on different boxes connected through extract, transform, and load (ETL). The ETL process periodically migrates the data from the operational store to the data warehouse to enable businesses to make informed and timely decisions. The drawbacks of this topology include that operational data is not immediately available for analytics due to delays in ETL, the complexity of this kind of solution, as well as higher costs.

SQL Server 2016 introduces some significant enhancements in this area, including the ability to create updateable nonclustered columnstore indexes and to run a columnar index over your in-memory or on-disk row store. This means you can gain the speed of in-memory OLTP while also conducting operational analytics.

#### **Enhancements to columnstore indexing**

There are many enhancements to columnstore indexing. Tables with a columnstore index on them can now enforce referential integrity through foreign key relations. These tables now also can have additional indexes on them to make querying more efficient. Plus, it is now possible to add data to a table that has a nonclustered columnstore index without having to drop and recreate the index. You can create

columnstore indexes on both on-disk tables and memory-optimized tables, too. SQL Server 2016 columnstore also supports improved monitoring and troubleshooting with new DMVs, XEvents, and PerfMon counters.

### **Analytics on OLTP workloads**

In addition to changes in columnstore indexing, SQL Server 2016 also minimizes the impacts of analytical queries being run against an OLTP workload. These analytical queries can be run on an in-memory OLTP workload with no application changes and a minimal impact on the performance of the OLTP workload. You also have the option of offloading analytical workloads to a readable secondary in a high availability environment. SQL Server 2016 columnstore provides the best performance and scalability available and can be applied in a flexible manner to both analytical and OLTP workloads.

### **Azure HDInsight and Apache projects**

Microsoft is delivering an enterprise-class implementation, or distribution, of Hadoop called HDInsight for Azure that is integrated with SQL Server, Active Directory, and Microsoft System Center to make it dramatically easier, more efficient, and more cost effective for organizations to capitalize on the opportunity Big Data can bring. HDInsight is the Hadoop distribution from and supported by Microsoft that is 100-percent compatible with Apache. HDInsight empowers organizations with new insights into previously untouched unstructured data, while connecting to widely used business intelligence tools.

HDInsight incorporates a series of tools designed to facilitate working with Big Data. These tools include the following:

- **Sqoop**: works with structured data, such as that in a SQL Server database or a data warehouse and imports it into, or exports out of, HDInsight clusters
- **Hbase**: NoSQL database for unstructured and semi-structured data
- **Oozie**: Workflow management
- **Hive**: SQL-like querying of Big Data
- **PIG**: scripting tools for MapReduce transformations
- **Storm**: Processes data in real-time

These tools are important building blocks in architecting Big Data solutions.

### **Analysis Services**

Analysis Services is an online analytical data engine used in decision support and business intelligence solutions, providing the analytical data for business reports and client applications such as Excel, Reporting Services reports, and other third-party BI tools. A typical workflow for Analysis Services includes building an online analytical processing (OLAP) or tabular data model, deploy the model as a database to an Analysis Services instance, process the database to load it with data, and then assign permissions to allow data access. When it's ready to go, this multi-purpose data model can be accessed by any client application supporting Analysis Services as a data source.

### **Improved in SQL Server 2016:**

For SQL Server 2016 several enhancements to Analysis Services are planned, including improvements in enterprise readiness, modeling platform, BI tools, SharePoint integration, and hybrid BI.

In the area of enterprise readiness, enhancements include improved performance with unnatural hierarchies, relational OLAP distinct count, drill-through queries, processing and query process separation as well as semi-additive measures. Database Console Commands are planned to also

supports detecting issues with multidimensional OLAP indexes. Additionally, Netezza is planned to be available as a data source. In Tabular mode, processing optimizations will enhance parallel partition processing.

Additional enhancements to the modeling platform in Tabular mode in SQL Server 2016 Analysis Services will make the user experience for creating data models easier and performant. Plans for the data model include support for bi-directional (many-to-many) cross filtering. Query engine optimizations to enhance performance for Direct Query and also new Data Analysis Expressions functions are planned, such as:

- DATEDIFF
- GEOMEAN
- PERCENTILE
- PRODUCT
- XIRR
- XNPV

In the area of hybrid connectivity, Power BI now provides seamless connectivity to on-premises Analysis Services models. This is discussed in more detail in the section on Powerful Insights below.

SQL Server 2016 Analysis Services will provide functional parity with SharePoint vNext and Excel vNext.

## Master Data Services

Master Data Services (MDS) continues to make it easier for organizations to manage master data structures (object mapping, reference data, and dimensions and hierarchies) used in data integration operations. With Entity Based Staging, database administrators (DBAs) can load all members and attribute values for an entity at one time. Additionally, the Explorer and Integration Management functional areas of the Master Data Manager web application have been updated with a new look and feel based on the Microsoft Silverlight browser development tool. DBAs can add and delete members more quickly, and can move them into a hierarchy more easily.

The MDS Add-in for Excel democratizes data management, so information workers have the ability to build data management applications directly in Excel. Information workers can use this add-in to load a filtered set of data from the MDS database, work with the data in Excel, and then publish the changes back to the MDS database. Administrators also can use the add-in to create new entities and attributes.

### Improved in SQL Server 2016:

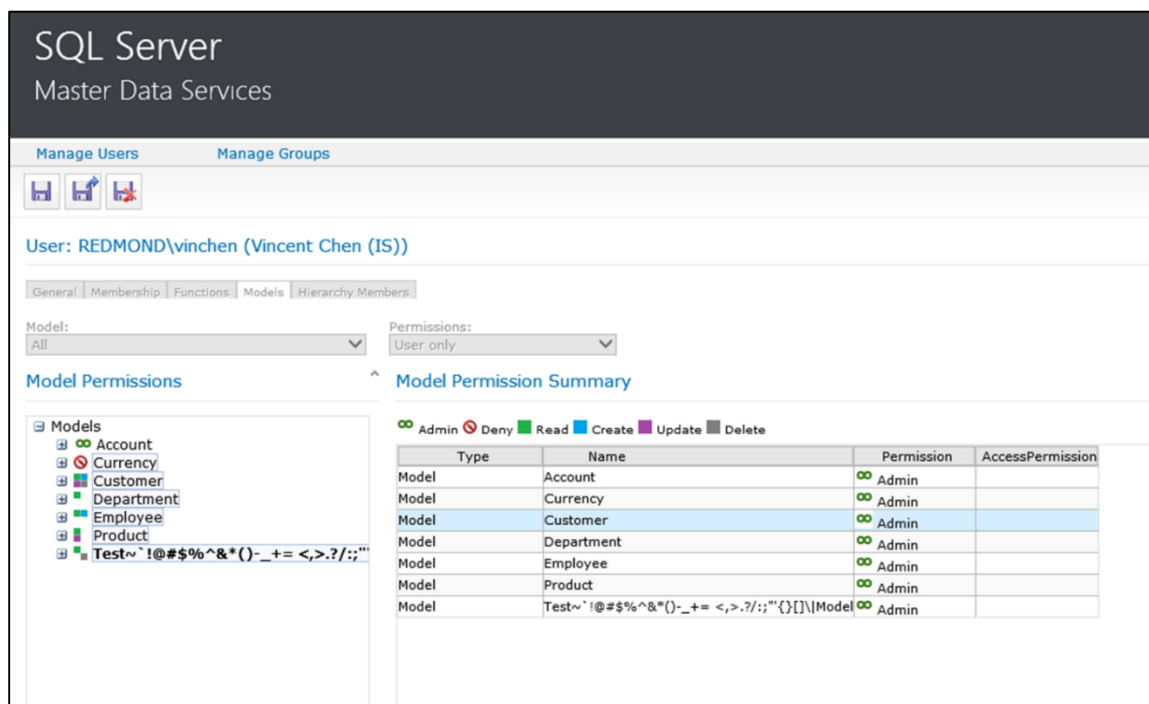
Enhancements in MDS can be organized around performance and scale, manageability and monitoring, and security.

In the areas of performance and scale, there have been improvements made in the Excel add-in as well as in bulk entity-based staging operations. For scalability, the MDS model deployment has been enhanced to support bigger models. Also, optional row-level compression per entity is now supported.

In the areas of manageability and monitoring, configurable retention settings for the transaction logs and reuse of entities across models is now supported. Each attribute now has a viewable display name. There are also enhancements in hierarchy management, as well as in troubleshooting and logging.

In the area of security, MDS in SQL Server 2016 now has more granular security permissions around read, write, delete, and create (Figure 6). There is also support for multiple system administrators and an explicit model admin permission property.

Figure 6: Enhancements to security permissions in Master Data Services



## Powerful insights

Data are the new currency and the use of data has become a major competitive differentiator. International Data Corporation (IDC) recently conducted a study that looked at companies that have a data-centric culture versus those without and the study calculates a massive data dividend for those companies have a data-centric culture. These companies were harnessing diverse data, not just in relational or located internally, but both relational and non-relational, internal and external. They were also using new analytical models to not just look at historical data but to use historical data to predict the future. The study also noted that insights from data were not limited to a few in the companies but were shared more broadly in those organizations. Finally, they were doing all of this at speed, in near real-time in many cases. They tend to be far more productive, be more efficient with their operations, and innovate faster, and the combination of these factors leads to higher sales versus their competition.<sup>4</sup>

## Power BI

Power BI is a collection of online services and features that enables you to find and visualize data, share discoveries, and collaborate in intuitive new ways.

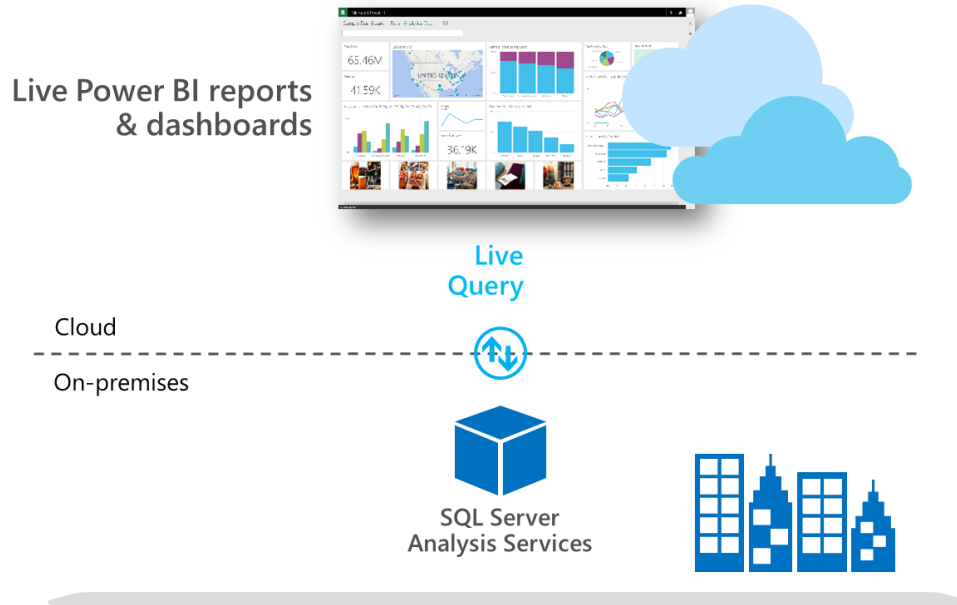
### Improved in SQL Server 2016:

#### Power BI Analysis Services Connector

SQL Server 2016 enables you to connect Power BI to your on-premises Analysis Services models. This allows you to leverage your existing investments in on-premises SQL Server Analysis Services while gaining the ability to utilize the rich toolset of Power BI without having to move those data sets to the cloud (Figure 7).

<sup>4</sup> IDC and Microsoft, April 2014.

Figure 7: Taking advantage of the combined benefits of Power BI and SQL Server Analysis Services



Further, as discussed above, SQL Server 2016 Analysis Services has enhancements in several areas, including the ability to better work with large tabular models, and you can access these from Power BI. Power BI can access the existing permission model setup, which supports fine-grained permissions, and use this security for your Power BI dashboards and reports.

### Improved in SQL Server 2016:

#### SQL Server Data and BI Tools in Visual Studio 2015

SQL Server Data Tools have been consolidated and improved in a number of ways in the SQL Server Data Tools for Visual Studio 2015. SQL Server Data Tools includes all SQL Server project types for Relational Databases, Integration Services, Analysis Services, and Reporting Services. Furthermore, the setup experience has been streamlined along with the process for importing from Office vNext. There is also support for a new Analysis Services Tabular scripting language and object model.

#### Azure Stream Analytics

Azure Stream Analytics service in the cloud enables you to rapidly develop and deploy a low-cost, real-time analytics solution to uncover insights from devices, sensors, infrastructure, and applications. It enables various opportunities including Internet of Things (IoT) scenarios, real-time remote management and monitoring, or gaining insights from devices like mobile phones or connected cars.

Azure Stream Analytics uses a SQL-based syntax and automatically distributes queries for scale, performance, and resiliency. It is most often used for small numbers of high-volume real-time queries. It uses Azure Event Hubs to ingest millions of events per second. Azure Stream Analytics will process ingested events in real-time, comparing multiple real-time streams or comparing real-time streams together with historical values and models. This enables the detection of anomalies, transformation of incoming data, the ability to trigger an alert when a specific error or condition appears in the stream, and



the ability to power real-time dashboards. It can handle complex event processing, including aggregation, reduction, and cleanup.

## Azure Machine Learning

Azure Machine Learning is a fully managed cloud service that enables data scientists and developers to efficiently embed predictive analytics into their applications, helping organizations benefit from massive datasets.

Azure Machine Learning can analyze Big Data to help identify patterns in the data and predict future events and user behavior. Azure Machine Learning offers a streamlined experience for all data scientist skill levels, from setting up with only a web browser, to using drag and drop gestures and simple data flow graphs to set up experiments. Machine Learning Studio features a library of time-saving sample experiments, R and Python packages and best-in-class algorithms from Microsoft businesses like Xbox and Bing. Azure ML also supports R and Python custom code, which can be dropped directly into your workspace. Experiments are easily shared, so others can pick up where you left off.

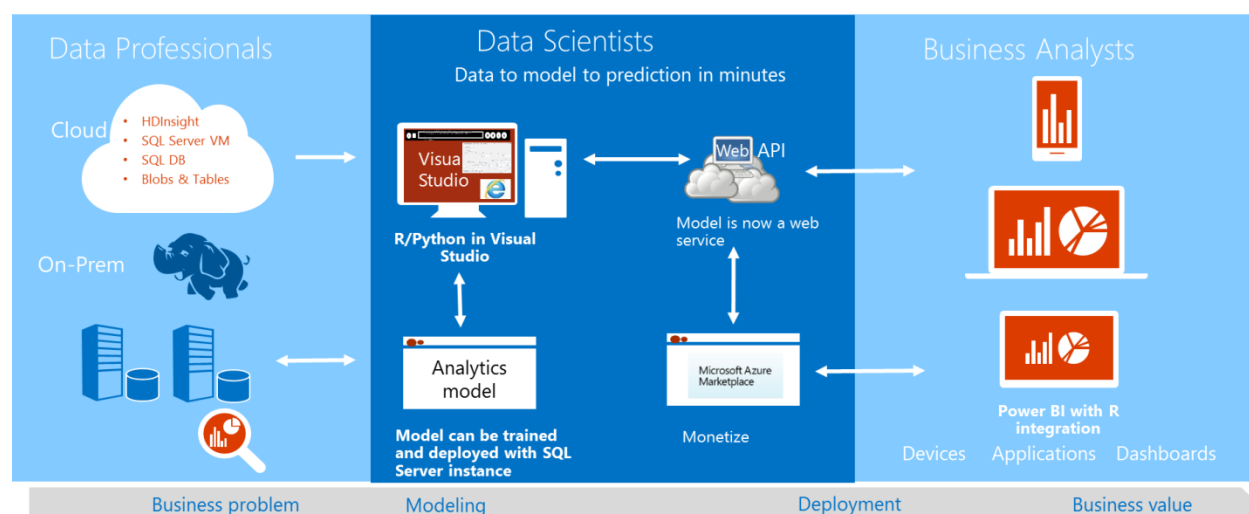
### New in SQL Server 2016:

## R integration

R is the most popular language for predictive analytics available today. However, R as an open source language has not scaled well for Big Data analytics. Microsoft's purchase of Revolution Analytics, as the leading provider for commercial software and services built on top of R, has brought this functionality to the Microsoft data platform.

Predictive Analytics is a key Big Data capability. R allows you to bridge the gap between the database and data science. SQL Server 2016 allows you to manage R models in SQL Server. This will help you use the power of R and data science to unlock Big Data insights with advanced analytics (Figure 8).

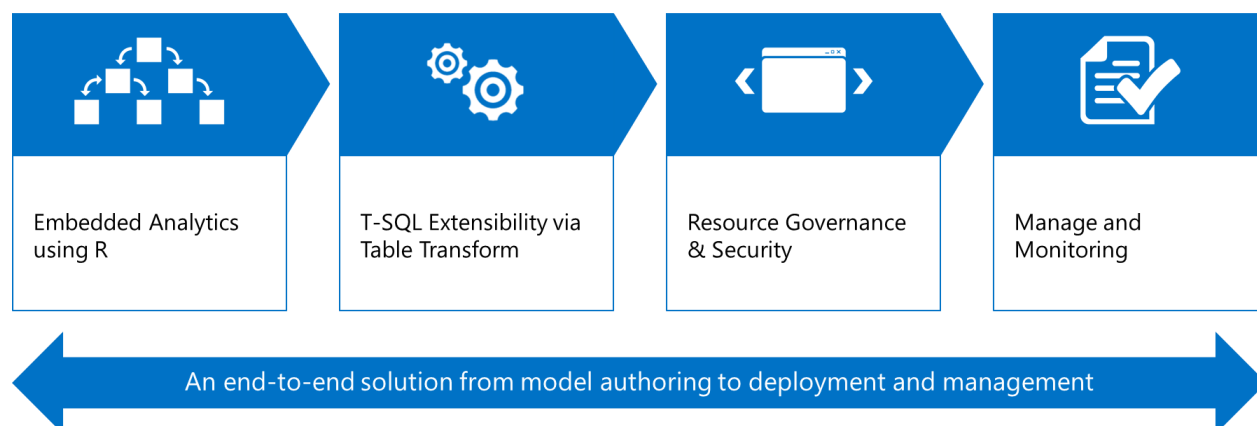
Figure 8: Using advanced analytics to gain insight into Big Data



This integration means that database professionals can use T-SQL for advanced analytics on operational data and models and can secure and ensure their availability. Data scientists can work in their favorite analytics environment, such as R in Visual Studio or Python in Visual Studio, while taking advantage of

the computational power, memory, and parallelism of the database engine and increasing model fidelity (Figure 9).

Figure 9: Flexibility in model authoring, deployment, and management with R integration



This integration of R will facilitate many Big Data scenarios, such as using Big Data for better audience targeting, churn forecasting, anomaly detection, and fraud and risk analysis. While business users can access the results from anywhere and on any device. Further, once models have been developed and trained, they can be deployed as web services to the Azure Marketplace.

## Reporting Services

SQL Server Reporting Services (SSRS) provides a full range of ready-to-use tools and services to help you create, deploy, and manage reports for your organization. Reporting Services includes programming features that enable you to extend and customize your reporting functionality.

Reporting Services is a server-based reporting platform that provides comprehensive reporting functionality for a variety of data sources. Reporting Services includes a complete set of tools for you to create, manage, and deliver reports and APIs that enable developers to integrate or extend data and report processing in custom applications. Reporting Services tools work within the Microsoft Visual Studio environment and are fully integrated with SQL Server tools and components.

### Improved in SQL Server 2016:

SQL Server 2016 brings a number of enhancements to Reporting Services that enable you to create modern reports over all your data and consume them on any device. With native connectors for the latest versions of Microsoft data sources such as SQL Server and Analysis Services; third-party data sources such as Oracle Database, Oracle Essbase, SAP BW, and Teradata; and ODBC and OLEDB connectors for many more data sources, you can create reports over a broad array of data sources.

New report themes and styles are planned to make it easier than ever to create modern-looking reports, and new chart types enable you to visualize your data in new ways. Greater control over parameter prompts give you enhanced ability to design dynamic, parameterized reports.

With support for modern browsers on multiple platforms, Reporting Services provides a platform for viewing and interacting with your reports on today's multitude of modern devices.

## Mobile BI

SQL Server 2016 supports mobile business intelligence on Windows, iOS, and Android devices (Figure 10). This allows users to visualize and interact with data on their mobile devices, using the native mobile apps available at no charge at the respective app stores. You can use these tools to connect to enterprise data sources, integrate with Active Directory for user authentication, deliver live data updates to mobile devices, and personalize data queries for each user.

Figure 10: Mobile BI on Windows, iOS, and Android devices



## Conclusion

Data warehousing, analytics, and business intelligence must adapt to a whole new scope, scale, and diversity of information and data—and they must continue to embrace the new ways that we discover and collaborate on information every day. The data that enterprises care about are expanding at an explosive speed. Whether data are coming from relational or non-relational sources, on-premises or on the cloud, Big Data, or other sources, SQL Server can help you access, integrate, and store these data as well as scale and manage them and gain deeper insights from them. SQL Server delivers the tools you need to gain deeper insights from your data more easily than ever before.

## More information

The following websites offer more information about topics discussed in this white paper:

- <http://www.microsoft.com/en-us/server-cloud/products/sql-server-2016/>: SQL Server website
- <http://www.microsoft.com/en-us/bi/default.aspx>: Microsoft business intelligence website
- <http://www.microsoft.com/bigdata/>: Microsoft Big Data solutions

## Feedback

Did this paper help you? Please give us your feedback by telling us on a scale of 1 (poor) to 5 (excellent) how you rate this paper and why you have given it this rating. More specifically:

- Are you rating it highly because of relevant examples, helpful screen shots, clear writing, or another reason?
- Are you rating it poorly because of examples that don't apply to your concerns, fuzzy screen shots, or unclear writing?

This feedback will help us improve the quality of white papers we release.

Please send your feedback to: <mailto:sqlfback@microsoft.com>