SQL Server
2016

Session 20

PolyBase, Query Store, and Stretch Database

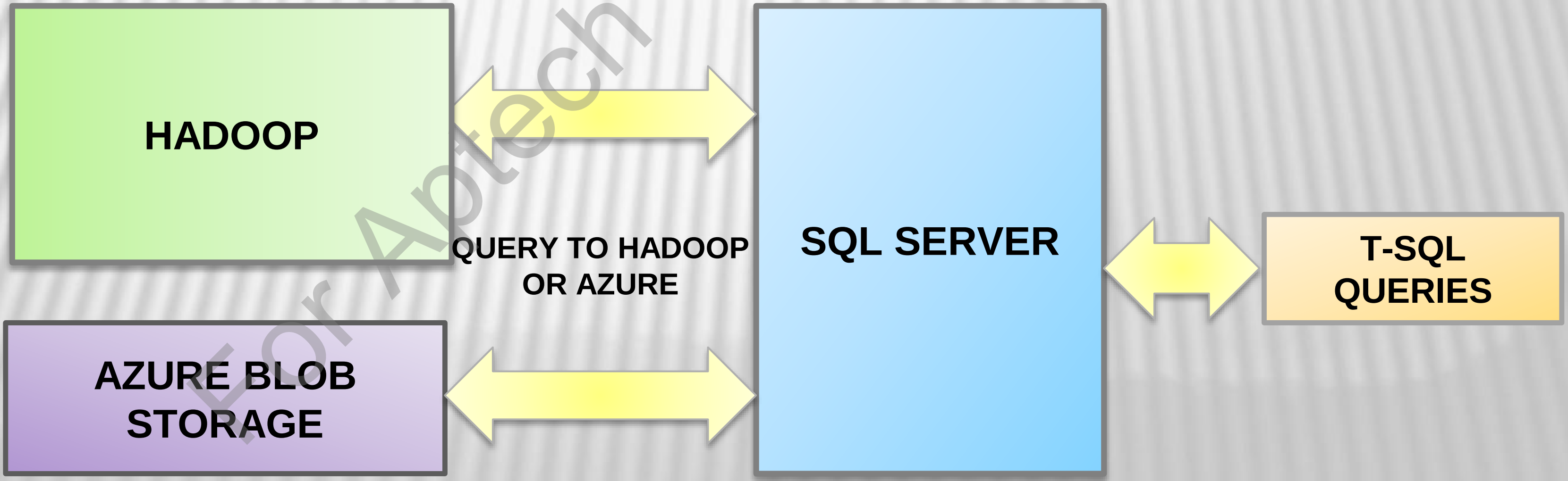
Objectives

- Describe PolyBase
- Explain features and advantages of PolyBase
- Define and describe Query Store
- Explain how to dynamically stretch warm and cold transactional data from SQL Server to Azure
- Describe how to tune workload performance with Query Store

SQL Server 2016

Understanding Polybase 1-2

- Increase in storage of data in Hadoop or Azure creates the need for applications to connect with them and skills to work with this data.
- Polybase is a built-in feature in SQL Server 2016.
- Connect from SQL Server to Hadoop or Azure Storage with T-SQL.
- No need for any importing tools or in-depth knowledge of any other Hadoop concepts like MapReduce or Hive.



Understanding Polybase 2-2

SQL Server 2016

- Simplifies access to big data on Hadoop and Azure Blob Storage using simple T-SQL
- Enables import and export of data between SQL Server and Hadoop or Azure Blob Storage
- Integrates with Business Intelligence tools to generate reports
- Improves performance by processing data on Hadoop or where the data resides
- Integrates with different systems to work with data, eliminating the need to customize any existing applications
- Incorporates external data into the database schema, thereby keeping the operations transparent to the end user and querying application

Polybase Architecture

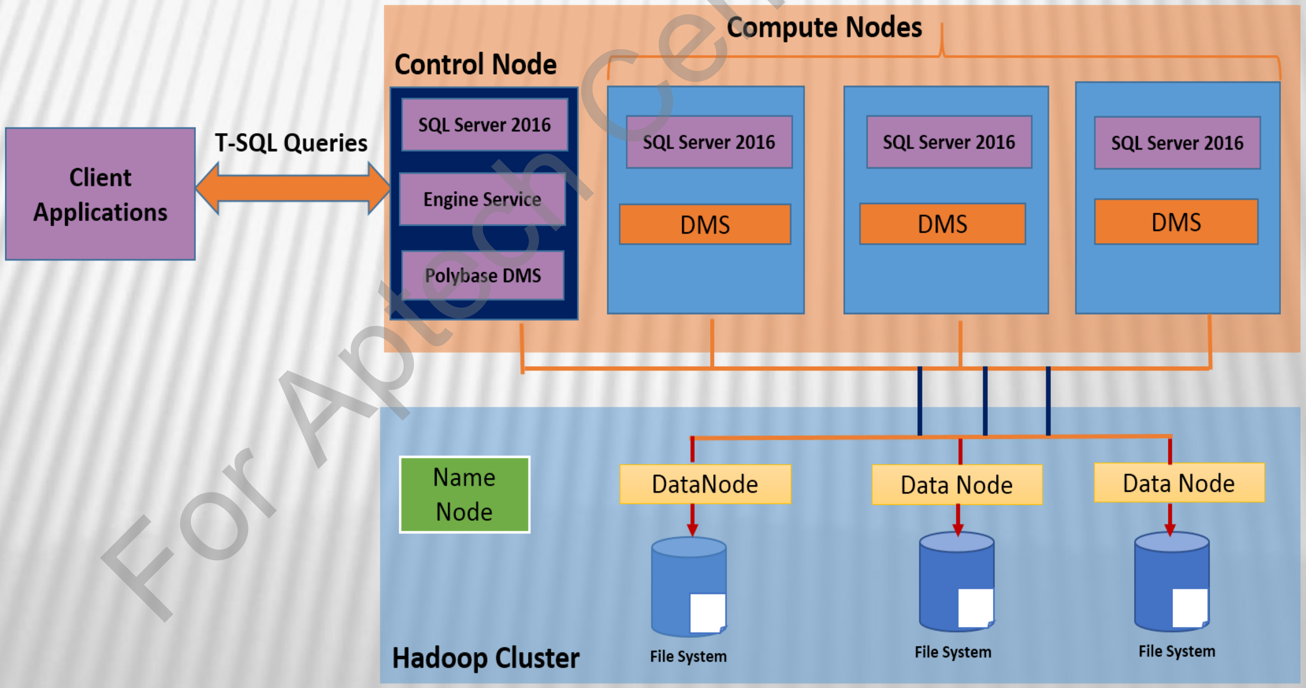
SQL Server 2016

CONTROL NODE

- Similar to Hadoop's Name Node/Job tracker and SQL Server instance acts as the Control Node
- Runs the Polybase Engine Service
- Manages queries between applications and external data sources

COMPUTER NODE

- Similar to Hadoop's Data Node/Task tracker
- Runs Polybase Data Movement Service
- Brings data temporarily to SQL Server to handle the parallel query executions



SQL Server 2016

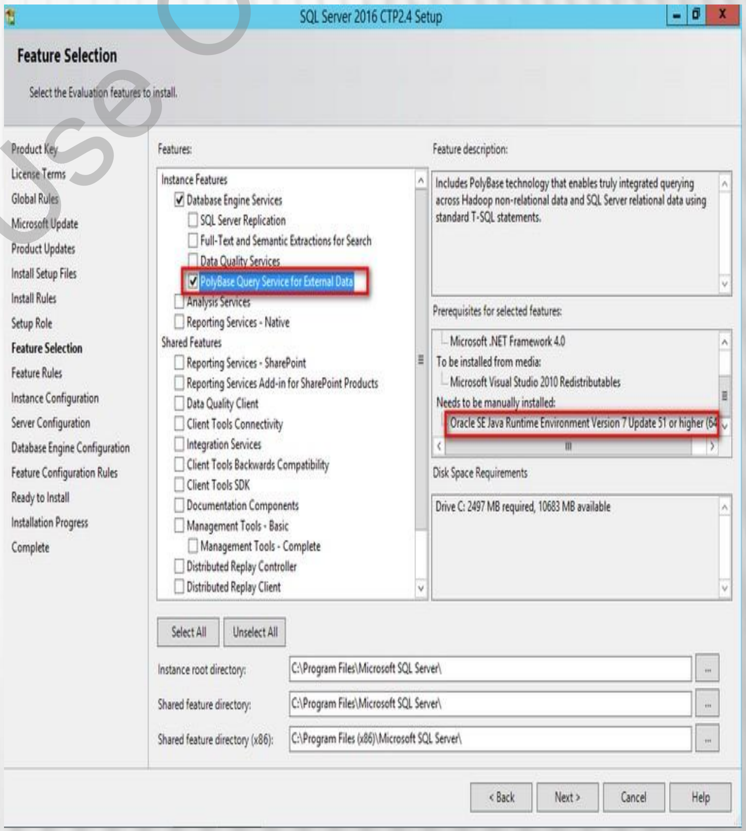
Installing and Setting up Polybase 1-2

1. Select Polybase feature during SQL Server set up
2. Ensure JRE installation
3. Configure interaction with Hadoop or Azure Storage

```
EXEC sp_configure , hadoop connectivity, 5;
GO
RECONFIGURE;
GO
```

4. Create a data source object that defines the connection to the external data source:

```
CREATE EXTERNAL DATA SOURCE HDP
WITH
(
TYPE='HADOOP'
Location='hdfs://sandbox.hortonworks.com:80
20';
```



SQL Server 2016 CTP2.4 Setup

Feature Selection

Select the Evaluation features to install.

Features:

Instance Features
- ☑ Database Engine Services
  - □ SQL Server Replication
  - □ Full-Text and Semantic Extractions for Search
  - □ Data Quality Services
  - ☑ PolyBase Query Service for External Data
- □ Analysis Services
- □ Reporting Services - Native

Shared Features
- □ Reporting Services - SharePoint
- □ Reporting Services Add-in for SharePoint Products
- □ Data Quality Client
- □ Client Tools Connectivity
- □ Integration Services
- □ Client Tools Backwards Compatibility
- □ Client Tools SDK
- □ Documentation Components
- □ Management Tools - Basic
  - □ Management Tools - Complete
- □ Distributed Replay Controller
- □ Distributed Replay Client

Feature description:

Includes PolyBase technology that enables truly integrated querying across Hadoop non-relational data and SQL Server relational data using standard T-SQL statements.

Prerequisites for selected features:

Microsoft .NET Framework 4.0

To be installed from media:

Microsoft Visual Studio 2010 Redistributables

Needs to be manually installed:

Oracle SE Java Runtime Environment Version 7 Update 51 or higher (64

Disk Space Requirements

Drive C: 2497 MB required, 10683 MB available

Instance root directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory (x86): C:\Program Files (x86)\Microsoft SQL Server\

SQL Server 2016

Installing and Setting up Polybase 2-2

5. Define a file format object to specify the format of external data by using the statement: CREATE EXTERNAL FILE FORMAT.

```
CREATE EXTERNAL FILE FORMAT TSV
WITH (
        FORMAT_TYPE = DELIMITEDTEXT,
         FORMAT_OPTIONS (
               FIELD_TERMINATOR = , \T’ ,
               DATE_FORMAT = , MM/DD/YYYY’
```

6. Create external tables using the CREATE EXTERNAL TABLE statement to facilitate a schema-based structure.

```
CREATE EXTERNAL TABLE sample_1
(
Code nvarchar (255),
Description nvarchar (255),
Total_emp int,
Salary nvarchar (255)
)
WITH
(
LOCATION = ‘/APPS/HIVE/WAREHOUSE/SAMPLE_07’,
DATA_SOURCE = HDP2,
FILE_FORMAT=TSV,
REJECT_TYPE=value,
REJECT_VALUE=0
)
```

7. Verify creation of External Resources and External tables folders.

Query Store

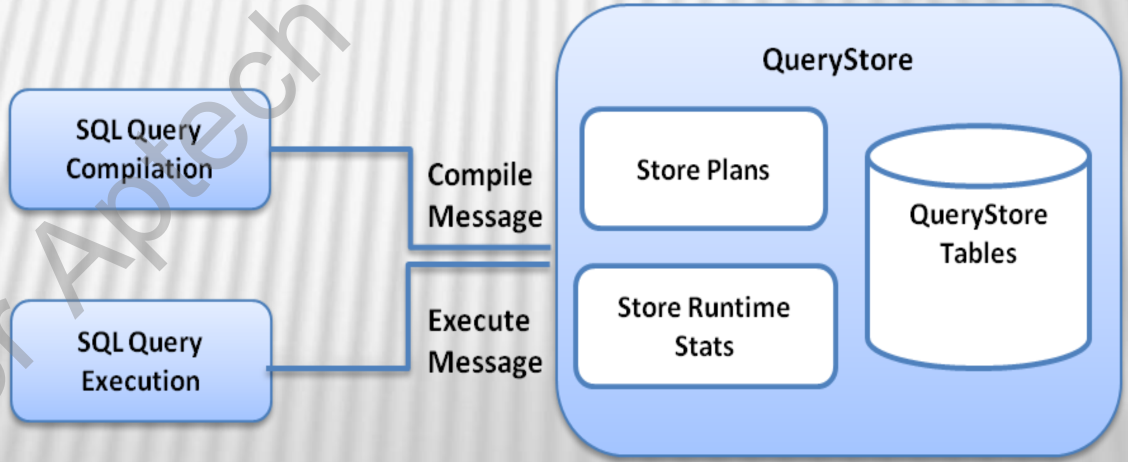
SQL Server 2016

- ❑ Is a built-in tool in SQL Server 2016 to aid performance tuning
- ❑ Maintains a complete history of execution plans of all queries including all the changes to the plans
- ❑ Monitors and stores the performance information of each execution plan over a period
- ❑ Analyzes information regarding query execution times, memory consumption
- ❑ Identifies and alerts about the plans causing performance issues
- ❑ Facilitates quick 'roll-back' to any old plan and thus helps to troubleshoot performance issues

SQL Server 2016

Query Store Architecture

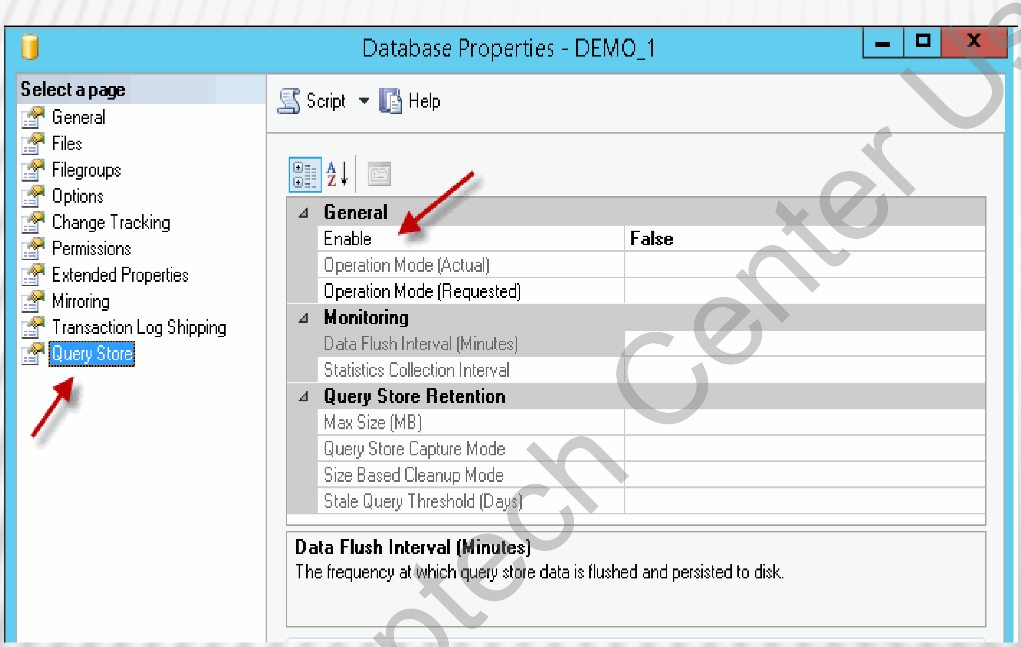
- ❑ Each query compilation or execution by SQL Server sends a message to the Query store.
- ❑ Compilation and execution information is first stored in cache and then on disk.
- ❑ INTERVAL_LENGTH_MINUTES parameter and the DATA_FLUSH_INTERVAL_SECONDS parameter specified during Query Store configuration determine the frequency of cache storage and flush to disk.



SQL Server 2016

Enabling Query Store

- ❑ Two methods to enable the Query Store on a database:
  - ❑ Using GUI inside SQL Server Management Studio (SSMS)



Database Properties - DEMO_1

General	
Enable	False
Operation Mode (Actual)	
Operation Mode (Requested)	
Monitoring	
Data Flush Interval (Minutes)	
Statistics Collection Interval	
Query Store Retention	
Max Size (MB)	
Query Store Capture Mode	
Size Based Cleanup Mode	
Stale Query Threshold (Days)	

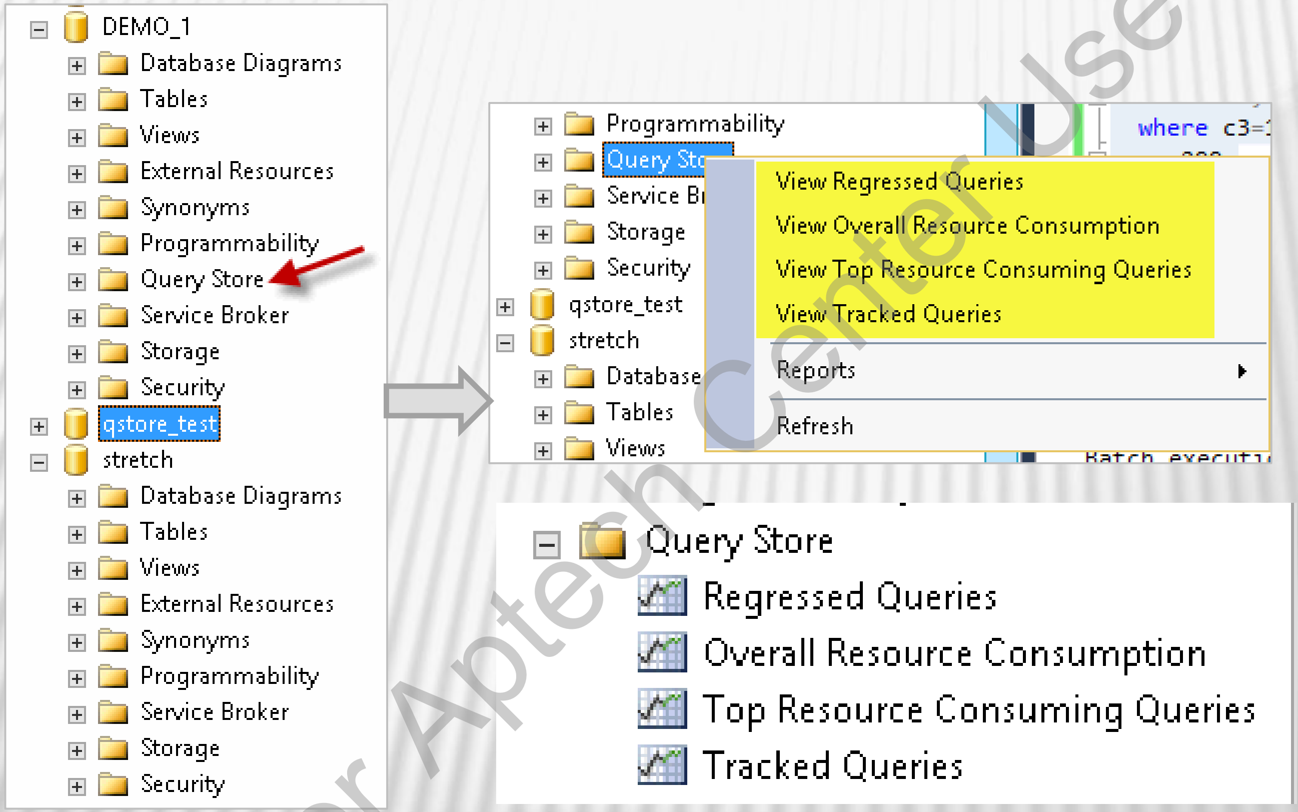
Data Flush Interval (Minutes)
The frequency at which query store data is flushed and persisted to disk.

  - ❑ Using T-SQL:

```
ALTER DATABASE [DEMO] SET QUERY_STORE = ON
GO
```

Viewing Query Information

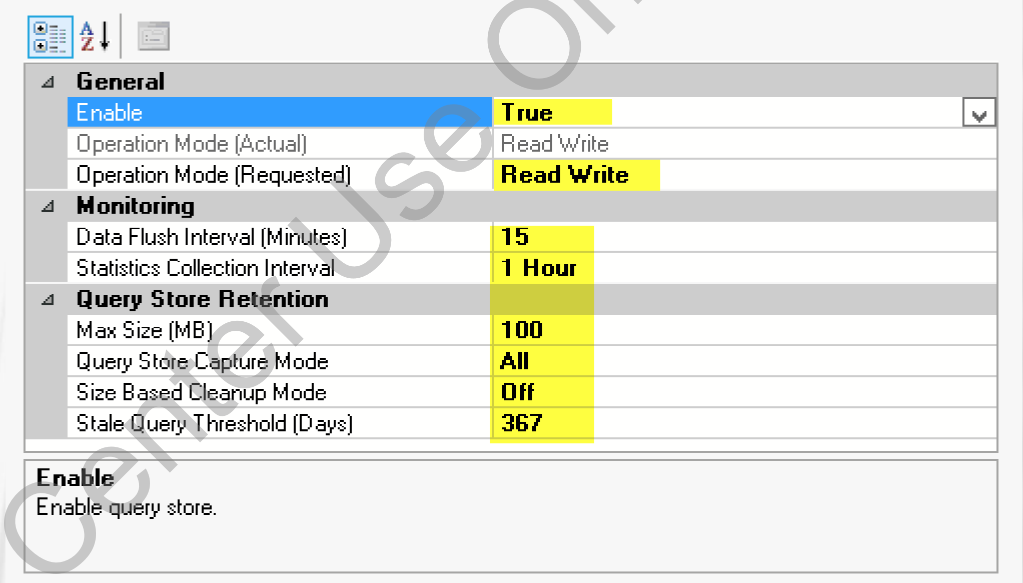
❑ To view information about queries right-click or expand the Query Store container and select the required options.



SQL Server 2016

Configuring Query Store

- ❑ Setup parameters using SSMS



General	
Enable	**True**
Operation Mode (Actual)	Read Write
Operation Mode (Requested)	**Read Write**
Monitoring	
Data Flush Interval (Minutes)	**15**
Statistics Collection Interval	**1 Hour**
Query Store Retention	
Max Size (MB)	**100**
Query Store Capture Mode	**All**
Size Based Cleanup Mode	**Off**
Stale Query Threshold (Days)	**367**

Enable
Enable query store.

- ❑ Setup parameters using T-SQL

```
ALTER DATABASE (DEMO_1)
SET QUERY_STORE (OPERATION_MODE = READ ONLY,
                 CLEANUP_POLICY = STALE_QUERY_THRESHOLD_DAYS = 367),
                 DATA_FLUSH_INTERVAL_SECONDS = 900,
                 INTERVAL_LENGTH_MINUTES = 60,
                 MAX_STORAGE_SIZE_MB = 100,
                 QUERY_CAPTURE_MODE = AUTO,
                 SIZE_BASED_CLEANUP_MODE = AUTO)
GO
```

Applications of Query Store

- ❑ Identify and rectify performance impact due to plan changes
- ❑ Alter queries to ensure optimal resource usage
- ❑ Analyze and allocate resources based on query frequency
- ❑ Troubleshoot performance issues due to upgrades or new versions of applications

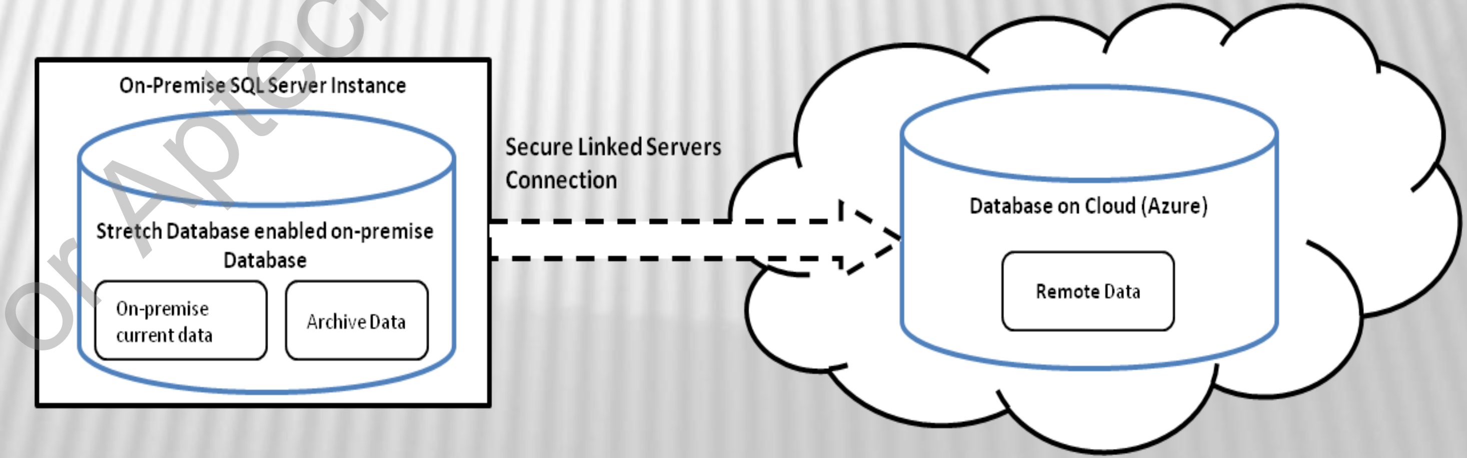
Stretch Database

- ❑ Enables storage of a part of a database in the cloud.
- ❑ Facilitates secure migration of tables and data to the cloud.
- ❑ Allows to access and query the data on the cloud at any time, without any changes to existing applications or queries.
- ❑ Reduces storage needs of on-premise data by using the vast storage capacity of the cloud.
- ❑ Reduces costs as cloud-storage is economical.

SQL Server 2016

Stretch Database Architecture

- ❑ A secure linked server definition establishes a connection to the remote or cloud database.
- ❑ When Stretch Database is enabled on a table, the resources on cloud are prepared and if migration is enabled, eligible data is migrated to the cloud database.

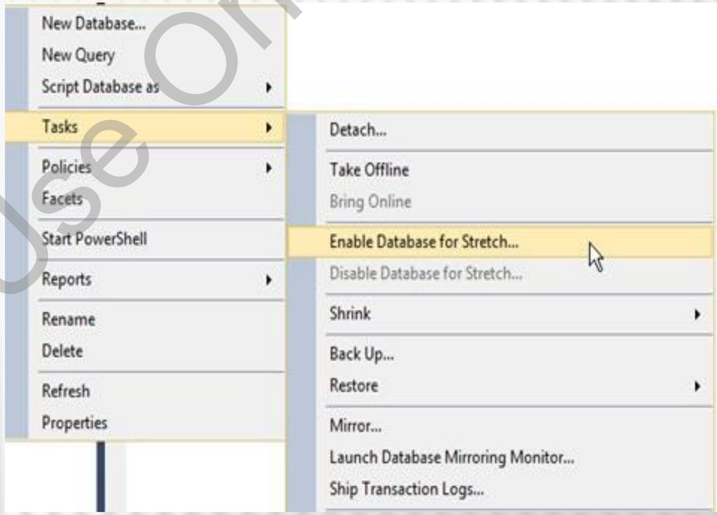

On-Premise SQL Server Instance
Stretch Database enabled on-premise Database
On-premise current data
Archive Data
Secure Linked Servers Connection
Database on Cloud (Azure)
Remote Data

SQL Server 2016

Setting up Stretch Database

1. Create an Azure account.
2. Enable Stretch Database.

```
EXEC sys.sp_configure N'remote data
archive', '1';
RECONFIGURE;
GO
```


New Database...
New Query
Script Database as
Tasks
Policies
Facets
Start PowerShell
Reports
Rename
Delete
Refresh
Properties
Detach...
Take Offline
Bring Online
Enable Database for Stretch...
Disable Database for Stretch...
Shrink
Back Up...
Restore
Mirror...
Launch Database Mirroring Monitor...
Ship Transaction Logs...

3. Select the database and the table that needs to be ‘Stretched’ to the cloud.
4. On the Object Explorer, right-click the database, select Tasks and then select the Enable Database for Stretch option.
5. Complete the actions on the wizard to create secure link.
6. Enable Remote Data Archive on the table.

```
USE [StretchDemo];
GO
ALTER TABLE [StretchSampleTable]
ENABLE REMOTE_DATA_ARCHIVE WITH
 (MIGRATION_STATE = ON );
GO
```

7. Query the Dynamic Management View (DMV) **sys.dm_db_rda_migration_status** to confirm and monitor data migration.

SQL Server 2016

Stretch Database Limitations

Stretch Database does not support:

- ❑ Tables that are replicated
- ❑ Memory-optimized tables
- ❑ Tables that contain FILESTREAM data
- ❑ Tables that use Change Tracking or Change Data
- ❑ Tables with columns that are always-encrypted data
- ❑ Tables with data-types such as XML, filestream, and sql_variant
- ❑ It also does not support some features such as:
- ❑ Foreign key constraints on a table
- ❑ ColumnStore Indexes and full-text indexes
- ❑ UPDATE statements and DELETE statements and the CREATE INDEX and ALTER INDEX operations

SQL Server 2016

Summary

- The PolyBase feature provides seamless integration with external data sources, such as Hadoop or Azure Blob Storage.
- PolyBase eliminates the need to specialized skills on Hadoop internals by enabling query-runs on external data sources with simple Transact-SQL commands.
- Query Store is a built-in tool to improve performance by maintaining historical information of every query and execution plan.
- Query Store tracks the performance of queries and triggers alerts on poorly performing plans.
- Stretch Database is a feature in SQL Server 2016 that enables stretching some part of a database to the Azure cloud, thereby lowering long-term storage costs as well as maintenance efforts.
- With Stretch Database, archive data as well as any current data can be migrated to the cloud securely and also accessed as easily as accessing local data at any time.