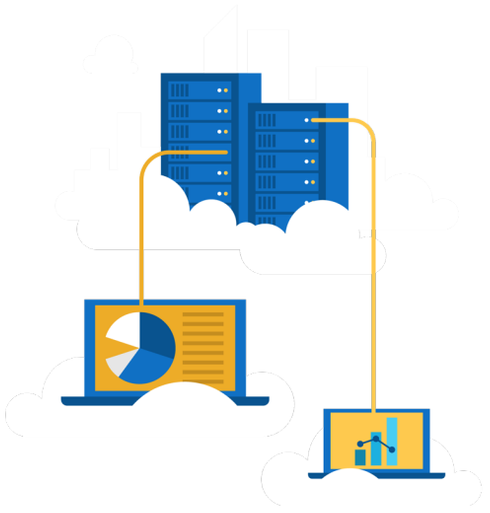
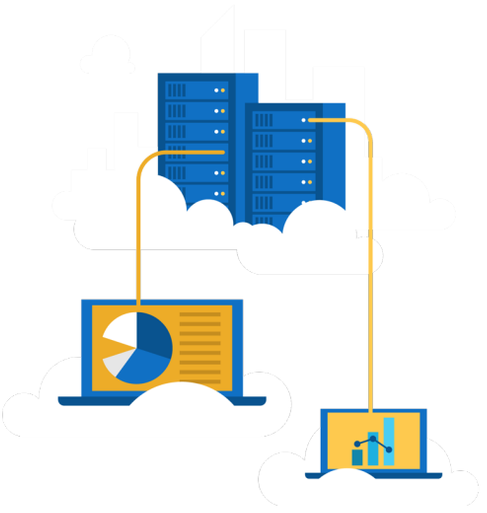
Azure Data Engineering – Day 3



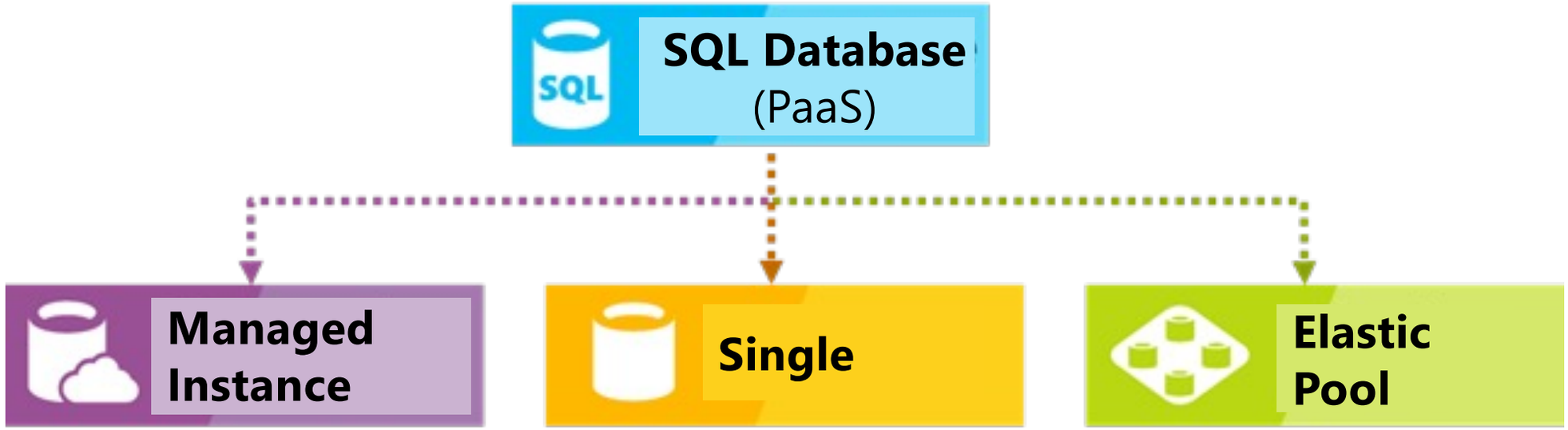
Lesson 01: Azure SQL Database



Azure SQL Database

Relational database managed service
Microsoft handles patching and updates
Shares code and features with SQL Server
Two purchasing models
vCore-based compute purchasing
DTU-based throughput purchasing

Azure SQL Database deployment options



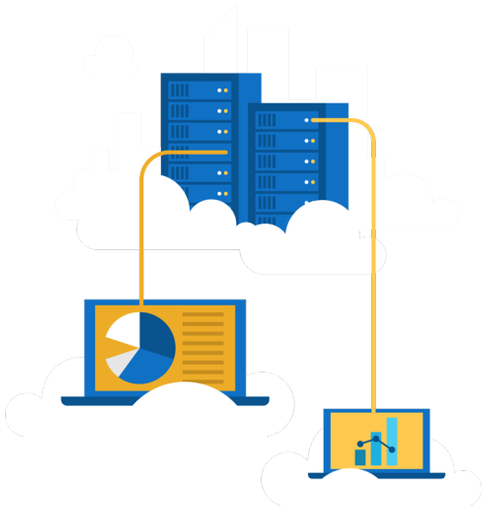
Choosing the right SQL option in Azure

SQL Server on a virtual machine (VM)	Azure SQL Database (Managed Instance)	Azure SQL Database (Logical server)
• You have full control over the SQL Server engine • Up to 99.95% availability • Full parity with the matching version of on-premises SQL Server • Fixed, well-known database engine version • Easy migration from SQL Server on-premises • Private IP address within Azure VNet • You have the ability to deploy application or services on the host where SQL Server is placed	• High compatibility with SQL Server on-premises • 99.99% availability guaranteed • Built-in backups, patching, recovery • Latest stable Database Engine version • Easy migration from SQL Server • Private IP address within Azure VNet • Built-in advanced intelligence and security • Online change of resources (CPU/storage)	• The most commonly used SQL Server features are available • 99.99% availability guaranteed • Built-in backups, patching, recovery • Latest stable Database Engine version • Ability to assign necessary resources (CPU/storage) to individual databases • Built-in advanced intelligence and security • Online change of resources (CPU/storage)

SQL option weaknesses

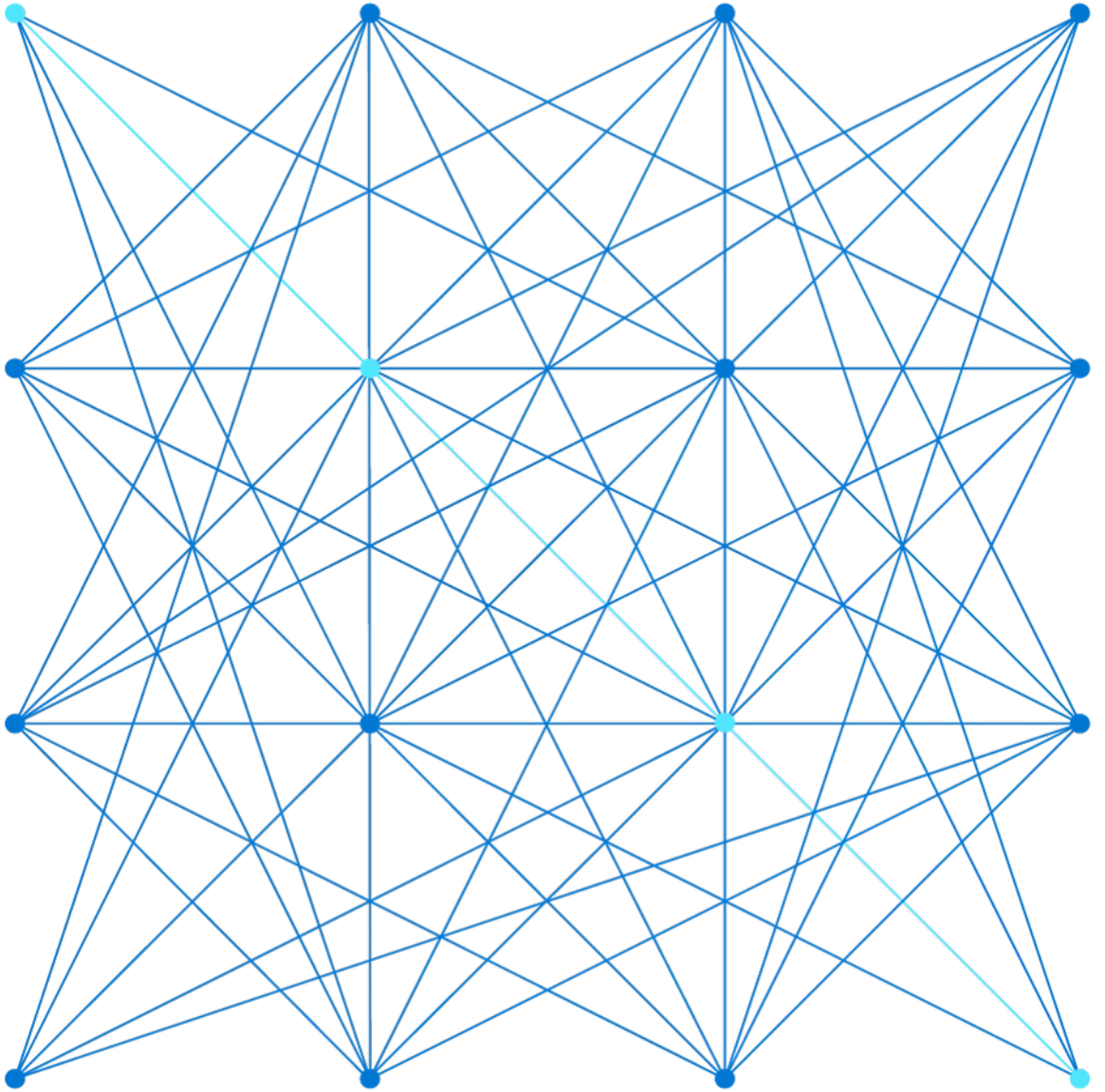
SQL Server on VM	Azure SQL Database (Managed Instance)	Azure SQL Database (Logical server)
• You need to manage your backups and patches • You need to implement your own high-availability solution • There is downtime while changing the resources (CPU/storage)	• There is still a minimal number of SQL Server features that are not available • No guaranteed exact maintenance time (but nearly transparent) • Compatibility with the SQL Server version can be achieved only by using database compatibility levels	• Migration from SQL Server might be difficult • Some SQL Server features are not available • No guaranteed exact maintenance time (but nearly transparent) • Compatibility with the SQL Server version can be achieved only by using database compatibility levels • Private IP address cannot be assigned (you can limit the access using firewall rules)

Demo: Creating an Azure SQL Database

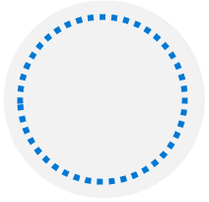


Microsoft Azure

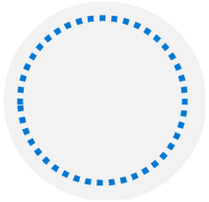
Run interactive queries using Azure Synapse Analytics serverless SQL pools



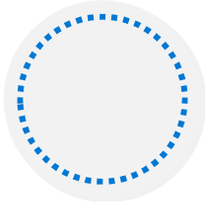
Agenda



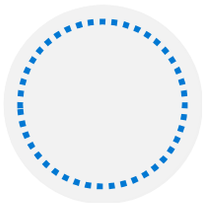
Lesson 01: Explore Azure Synapse serverless SQL pools capabilities



Lesson 02: Query data in the lake using Azure Synapse serverless SQL pools

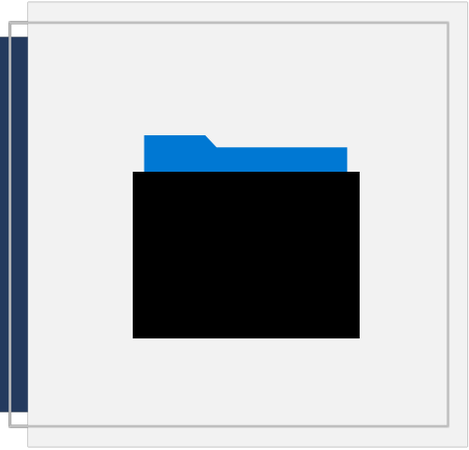


Lesson 03: Create metadata objects in Azure Synapse serverless SQL pools



Lesson 04: Secure data and manage users in Azure Synapse serverless SQL pools

Lesson 01: Explore Azure Synapse serverless SQL pools capabilities



Azure Synapse serverless SQL Pools

Every Azure Synapse Analytics workspace comes with serverless SQL pool endpoints so you can start querying data in seconds to minutes in a data lake as soon as the workspace is created. There's no infrastructure to setup or clusters to maintain.

Comparing dedicated SQL Pools with serverless SQL pools in Azure Synapse Analytics

Dedicated SQL pools

- Used for Data Warehouse operations
- Provides predictable performance and costs
- Reserves processing power for data stored in SQL tables

Serverless SQL pools

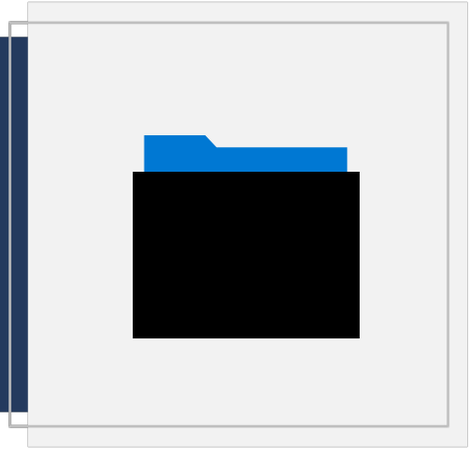
- Used for data preparation or ad-hoc queries against unstructured data.
- Provides an always available SQL endpoint for unplanned workloads
- Enables interactive querying

Explore Azure Synapse serverless SQL pools capabilities

Every Azure Synapse Analytics workspace comes with serverless SQL pool endpoints so you can start querying data in seconds to minutes in a data lake as soon as the workspace is created. There's no infrastructure to setup or clusters to maintain.

- **Data Exploration**
- **Data transformation**
- **Logical data warehouse**

Lesson 02: Query data in the lake using Azure Synapse serverless SQL pools



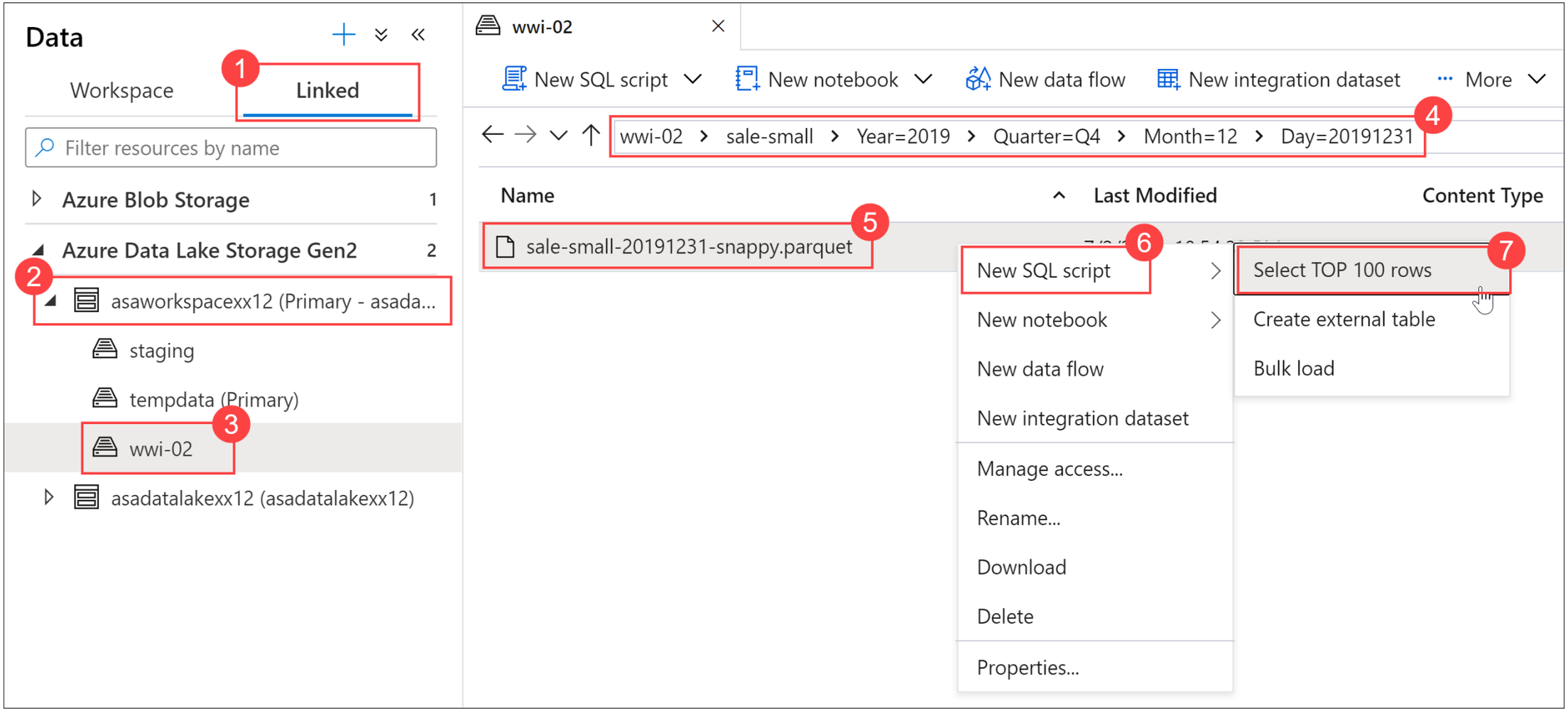
Common files to query

Parquet

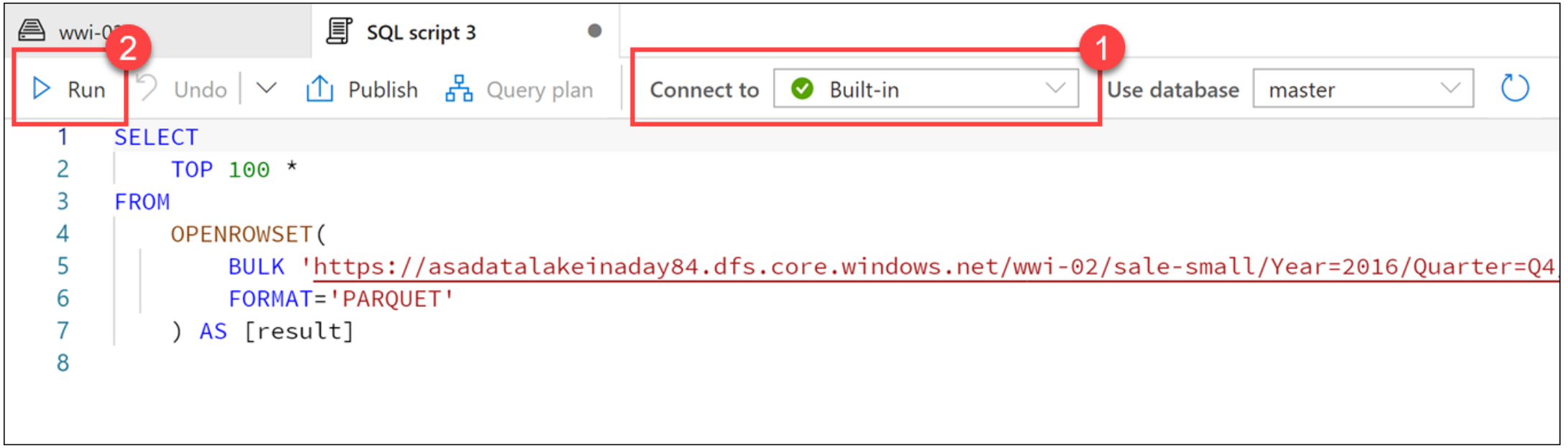
Json

DelimitedText

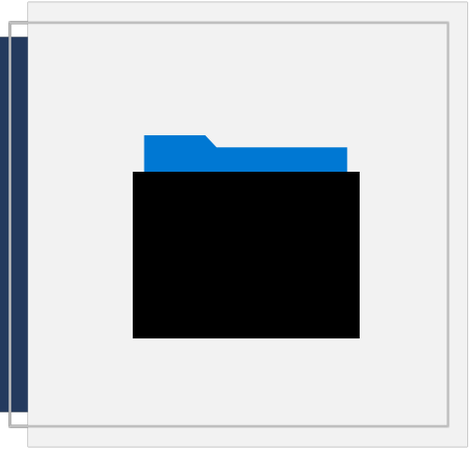
Using Azure Synapse Studio to view data



Querying parquet files in a data lake



Lesson 03: Create metadata objects in Azure Synapse serverless SQL pools



Create metadata objects in Azure Synapse serverless SQL pools

Database

```
CREATE DATABASE [YourDatabaseName]
```

Database scoped credential

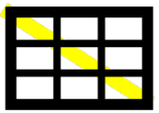
```
CREATE DATABASE SCOPED CREDENTIAL [sqlondemand]
WITH IDENTITY='SHARED ACCESS SIGNATURE',
SECRET = 'sv=2018-03-28&ss=bf&srt=sco&sp=rl&'
```

External data source

```
CREATE EXTERNAL DATA SOURCE SqlOnDemandDemo WITH (
    LOCATION = 'https://sqlondemandstorage.blob.core.windows.net',
    CREDENTIAL = sqlondemand );
```

External file format

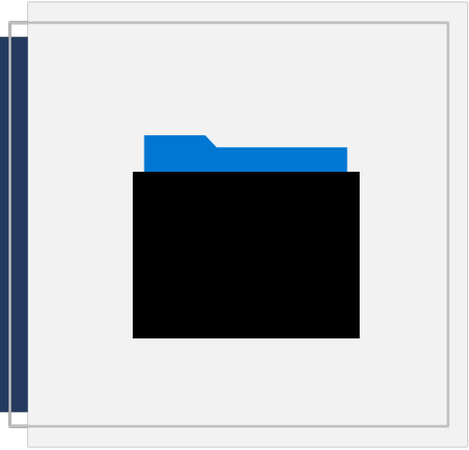
```
CREATE EXTERNAL FILE FORMAT QuotedCsvWithHeaderFormat
WITH
( FORMAT_TYPE = DELIMITEDTEXT,
FORMAT_OPTIONS ( FIELD_TERMINATOR = ',', STRING_DELIMITER = '"',
FIRST_ROW = 2 ) );
```



External Table

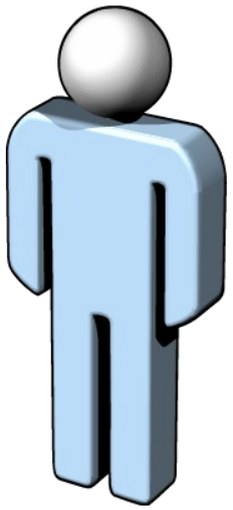
```
CREATE EXTERNAL TABLE populationExternalTable
( [country_name] VARCHAR (100) COLLATE Latin1_General_BIN2,
[year] smallint, [population] bigint )
WITH
( LOCATION = 'csv/population/population.csv',
DATA_SOURCE = sqlondemanddemo,
FILE_FORMAT = QuotedCSVWithHeaderFormat );
```

Lesson 04: Secure data and manage users in Azure Synapse serverless SQL pools



Securing access to data in a data lake when using Azure Synapse Analytics

Choose an authentication method



Azure Synapse Analytics dedicated SQL pool

Authentication options
SQL Authentication
Or
Azure Active Directory Authentication

Azure Data Lake Store

Authentication options
Anonymous access
Shared access signature
Managed identity
User identity

Manage users in Azure Synapse serverless SQL pools

User

Bob

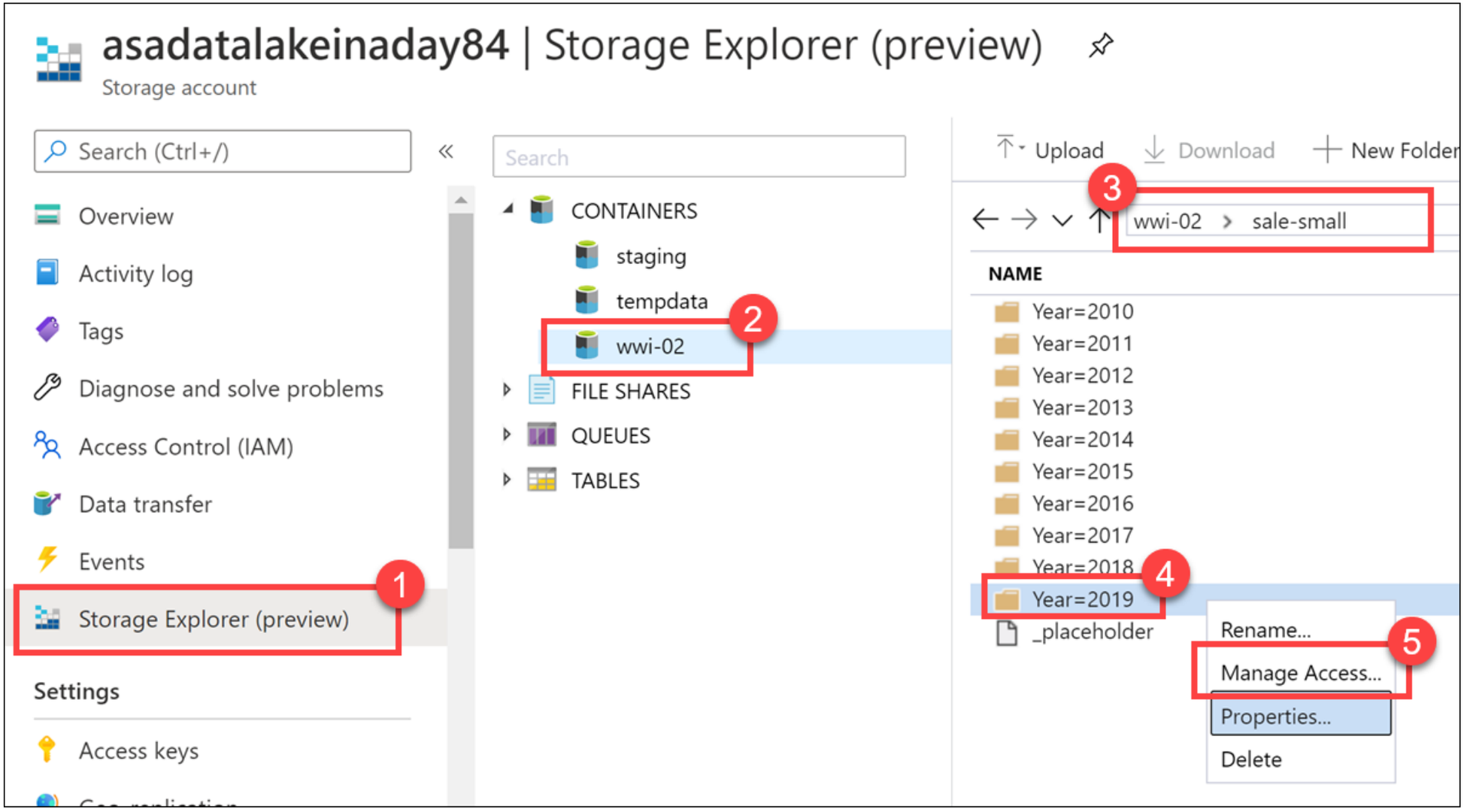
Role assignment

Storage Blob Data Reader

Storage account name

asadatalakeinaday84

Manage user access to data lake files



Review questions

Q01 – Which SQL function enables you to access files in Azure storage and read the contents of a remote data source?

A01 – OPENROWSET

Q02 – Which metadata object allows you to reuse the queries that you create and enable applications to view data in a serverless SQL pool?

A02 – Views

Q03 – Name the three permissions that can be set on a container object?

A03 – Read, write and execute

Microsoft Azure