Module 5

Implementing an Azure SQL Data Warehouse

Zyead Ahmed.
Aspiring To Learning Data Engineer.

Module Overview

- Advantages of Azure SQL Data Warehouse
- Implementing an Azure SQL Data Warehouse Database
- Developing an Azure SQL Data Warehouse
- Migrating to an Azure SQL Data Warehouse
- Copying Data with the Azure Data Factory

Lesson 1: Advantages of Azure SQL Data Warehouse

- What is Azure SQL Data Warehouse?
- Scalability and Cost
- Security and Availability
- PolyBase
- Hybrid Cloud

What is Azure SQL Data Warehouse?

- Cloud-based database
  - Relational and nonrelational
  - Enterprise workloads
  - Integrated with Azure
  - Fully managed service
- Benefits include:
  - Massive parallel processing
  - Advanced query optimization
  - Columnstore indexes
  - PolyBase integration
  - Auditing
  - Scalability

Scalability and Cost

- No upfront cost
- Storage
  - Adjusts automatically
  - Cost based on storage used
- Compute
  - Determines execution performance
  - Data Warehouse Unit (DWU)
  - Increase or decrease DWU
  - Cost based on DWU used
  - Pause and start

Security and Availability

- Security
  - Firewall
  - Add logins
  - Set authorisation
  - Auditing
- Availability
  - Can restore in different region
  - Can choose restore point in last seven days

PolyBase

- Can access unstructured data in other systems
- Set up external table to link to data source
- Query external table as normal table

Hybrid Cloud

- Can integrate between on-premises, cloud, and unstructured data sources
- Use PolyBase to query and copy data with Transact-SQL
- Schedule data copy using Azure Data Factory

Lesson 2: Implementing an Azure SQL Data Warehouse Database

- Creating a Server
- Creating a Database
- Configuring the Server Firewall
- Connecting to Azure Database Using SQL Server Management Studio
- Demonstration: Creating and Configuring an Azure SQL Data Warehouse Database

Creating a Server

- Logical server
- Specify:
  - Server name that has not been used
  - Server admin logon
  - Password
  - Location nearest to you
- Create database in same process

Creating a Database

- Create database
  - Name of database
  - Drag slider to change DWU performance
  - Create a new server or use existing server
  - Source
  - Create a new resource group or use existing resource group
- DWU settings
  - Scale
  - Pause/start

Configuring the Server Firewall

- Add client IP address before connecting
- Client IP address may change
- Specify rule:
  - Range of IP addresses to allow for change
  - IP addresses for other client computers

Connecting to Azure Database Using SQL Server Management Studio

- Fully qualified server name
- Connect to server using SSMS
- USE statement not supported
- Right-click database, New Query
- Most Transact-SQL supported in Azure SQL Data Warehouse databases

Demonstration: Creating and Configuring an Azure SQL Data Warehouse Database

In this demonstration, you will see how to:

- Create an Azure SQL Data Warehouse Database and server
- Change the performance settings
- Configure the Azure firewall
- Connect to the Azure server with SQL Server Management Studio

Lesson 3: Developing an Azure SQL Data Warehouse

- Concurrency and Memory Allocation
- Data Distribution
- CREATE TABLE AS SELECT
- GROUP BY Limitations
- Temporary Tables
- User Defined Schemas

Concurrency and Memory Allocation

- Resource class
- Concurrency slots
  - Query may use more than one concurrency slot
  - Dependent on resource class and DWU service level
- Concurrent queries
  - Maximum of 32 queries
  - Maximum slots dependent on DWU service level
- Memory allocation
  - Dependent on resource class and DWU service level

Data Distribution

- Data in tables allocated to distributions
- Round-robin distribution
  - Random distribution allocation
- Hash distribution
  - Choose hashed column
  - Distribution determined by function of column value
  - Ensure hashed column has even spread of data

CREATE TABLE AS SELECT

- Makes copy of a table
- Can set index properties and distribution type

```
CREATE TABLE Countries_New
WITH
(
 CLUSTERED COLUMNSTORE INDEX,
 DISTRIBUTION = HASH(CountryKey)
)
AS SELECT * FROM Countries
;
```

- Use to work around unsupported features

GROUP BY Limitations

- GROUP BY clause is supported
- GROUPING SETS, CUBE and ROLLUP subclauses are not supported
- UNION ALL operator is supported
- When migrating to Azure SQL Data Warehouse, ensure queries containing unsupported clauses are amended

Temporary Tables

- Local temporary tables can be accessed anywhere within session
- Global temporary tables are not supported

User Defined Schemas

- All data in one database
- Use schemas to identify legacy databases

Lesson 4: Migrating to an Azure SQL Data Warehouse

- The Data Warehouse Migration Utility
- Migrating Data with the Data Warehouse Migration Utility
- Other Migration Tools
- Differences Between SQL Server and Azure SQL Data Warehouse Schemas
- Updating Transact-SQL
- Demonstration: Migrating a Database to Azure SQL Data Warehouse

The Data Warehouse Migration Utility

- Advantages
  - Straightforward
  - Multiple tables
  - Specify distribution type
  - Notification of incompatibility
- Download from Internet
- Must have BCP and Excel installed

Migrating Data with the Data Warehouse Migration Utility

- Check compatibility
- Migrate schema
- Migrate data
  - bcp commands to export and import

Other Migration Tools

- Options for loading data into an Azure SQL Data Warehouse include:
  - **Azure Feature Pack for Integration Services (SSIS)**
    - Downloadable extension for SSIS that facilitates the movement of data between on-premises and cloud
  - **SSIS**
    - Add Azure SQL Data Warehouse connection in data flows
    - Use SQL Agent to schedule regular transfer of data
  - **Bulk Copy Program (bcp)**
    - Useful for small data, use bcp to copy data to flat files and load into the data warehouse destination
  - **AZCopy**
    - Copy data from flat files into Blob storage, and use PolyBase to load into data warehouse
  - **Import/Export**
    - For data larger than 10TB, bcp data to files, copy to disks and ship to Microsoft
  - **PolyBase and T-SQL**
    - Move UTF-8 formatted data in text files to Azure Blob storage or HDInsight, then use T-SQL command to load into the data warehouse
    - PolyBase uses the massively parallel processing (MPP) architecture for fast loading

Differences Between SQL Server and Azure SQL Data Warehouse Schemas

- Some table features not supported
  - Primary Keys
  - Foreign Keys
  - Unique Indexes
  - Constraints
- Some data types not supported
  - numeric
  - nvarchar(max)
  - varchar(max)

Updating Transact-SQL

- Some Transact-SQL not supported
- Rewrite to achieve same result

Demonstration: Migrating a Database to Azure SQL Data Warehouse

In this demonstration, you will see how to:

- Install the Data Warehouse Migration Utility
- Check compatibility of the legacy database
- Migrate the schema
- Migrate the data

Lesson 5: Copying Data with the Azure Data Factory

- The Azure Data Factory
- Creating a Data Factory
- Setting Up a Data Gateway for the On-Premises Server
- Setting up a Linked Service
- Setting Up a Dataset
- Setting Up a Pipeline Activity to Copy Data
- Data Factory Diagram

The Azure Data Factory

- Capabilities and application to Azure SQL Data Warehouse
- Entities
  - Activity
  - Pipeline
  - Dataset
  - Linked service
- Scheduling
- JSON templates
  - Edit parameters in script
  - Replace \ with \\

Creating a Data Factory

- Factory contains entities for activities
- Specify
  - Name
  - Resource group name
  - Region

Setting Up a Data Gateway for the On-Premises Server

- Access data factory from on-premises server
- Create new gateway
- Install on computer

Setting up a Linked Service

- New data store
- Edit parameters in JSON script
  - name
  - connectionString
    - Integrated Security
    - User ID
    - Password
  - gatewayName
  - userName: Use for Windows authentication
  - password: Use for Windows authentication

Setting Up a Dataset

- New dataset
- Edit parameters in JSON script
  - name
  - linkedServiceName
  - tableName
  - frequency
  - interval

Setting Up a Pipeline Activity to Copy Data

- New pipeline
- Edit parameters in JSON script
  - name
  - start
  - end
- Add activity script

Data Factory Diagram

- Shows data flow
- Pipeline properties
  - Activities
  - Datasets
- Dataset properties
  - Data slices