Activity 04: Security

Wide World Importers is hosting in their data warehouse a plethora of data coming from many disparate sources. The idea of bringing all of their data together into Azure Synapse Analytics for them to query, gain insights, and consume in ways they have never done before is exhilarating! As much as it is an exciting game-changer for this business, it opens up a large surface area for potential attack. Security must be established at the forefront at the time of design of this solution.

With your help they have already secured the data in the data lake. In this activity, you will help design a solution for securing the data available within the SQL tables that make up their data warehouse.

Requirements

- Want to make sure that the connection information to the data warehouse is always securely maintained, by the components that need it to access the data.
- Want to ensure that Azure Synapse is secured in depth at the network level.
- Will need the flexibility to assign users to groups who should have access to the workspace, as well those that might have elevated permissions, such as those who can administer the entire Synapse Workspace, or manage just the Spark or SQL Pools or use Pipelines.
- Want to maintain exclusive control over the keys to used to encrypt the data warehouse data at rest. They do not want Microsoft or any other entity to provide or have access to these keys.

Whiteboard

Open your whiteboard for the event, and in the area for Activity 4 provide your answers to the following challenges.

The following challenges are already present within the whiteboard template provided.

Challenges

1. The Synapse Pipelines that WWI is creating will need to access both the data in the data lake and in the data warehouse. Diagram and document the steps they should, and the Azure services they should use, to secure pipelines and pipeline runs.

2. Building upon the architecture you provided for the previous challenge, how would you address WWI's requirement to maintain exclusive control over the keys to used to encrypt the data warehouse data at rest?

3. WWI would like to understand how they will manage access control to the Synapse workspace with your proposed design. What four security groups should they create and what is the purpose of each group? How would you structure the group inheritance? Complete the following diagram and add your justifications.

4. Diagram how you would recommend WWI secure the network boundary around Azure Synapse Analytics? If they wanted to ensure that access to Synapse Studio is only possible from a VM on the approved virtual network (or from a computer connected to that virtual network using VPN), how would they configure that? How would they monitor for suspicious traffic against the storage account and receive alerts to the same?

5. Complete the following diagram to illustrate how they would secure access to data in the data warehouse. They have the following scenarios in mind:

    1. Hide specific columns from view: The CEO, who is an authorized personnel with access to all the information in the `Sale` table. However, a Data Analyst, only should never be able to see the Revenue column.
    2. Hide certain rows from view: The CEO has access to all the information in the `Sale` table, but a data analyst is restricted to only seeing the rows for the region she supports.
    3. Hide details of specific fields from view: All users should default to seeing only partial credit card numbers and emails in the `CustomerInfo` table.