Tuning and Optimizing a Data Warehouse in Microsoft Azure Synapse Analytics

Gary Grudzinskas

AZURE CONSULTANT AND AUTHOR

@garygrudzinskas

Overview

- Backup and restoring a data warehouse
- Managing costs of a data warehouse
- Workload management
- Implementing security
- Monitoring Synapse Analytics
- Deleting the data warehouse

Data Warehouse Backup

- Snapshots offer a point of time that can be used to recover or copy data

- Consists of many files from distributed system located in Azure storage

- Captures incremental changes

- Automatic Restore Points taken several times per day and cannot be deleted

- User defined snapshots can be taken anytime

Data Warehouse Restore

Restore to a point of time

Deleted at 7-day mark and when there are 42 restore points

Will not be available with a paused DW

Available after a DW is dropped for 7 days

Can keep original or restored DB

Charged at Premium Storage rates rounded to the nearest TB
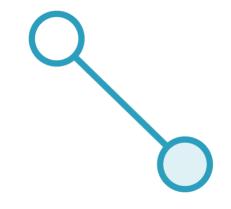
Geo-redundant

Managing Costs in Azure Synapse Analytics

Webpage for pricing

Compute pricing

Storage pricing

Get help with pricing details

Connection Security

Firewall rules
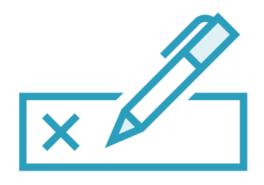
Allow and block IP addresses

Uses port 1433

Server firewall rules only
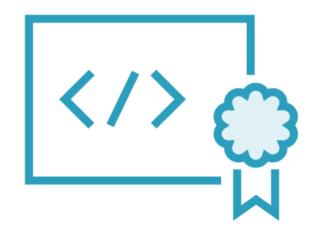
Connections encrypted by default

Authentication Security

Who has access to data warehouse

SQL Server Authentication

Azure Active Directory Authentication

Users should use unique credentials

Create logins to specific roles

Authorization Security

What users can do once logged in

Privileges are determined by role memberships and permissions

**Granular permissions on individual columns, views, tables, schemas, objects and stored procedures**
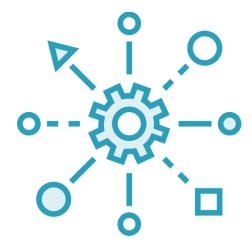
Transparent
Data Encryption

Encrypts and decrypts data at rest

Encrypts the storage of an entire database by using a symmetric key

Uses AES-256 encryption algorithm

Workload Management

All operations that transpire in relation to a data warehouse

The process of loading data

Running analysis and reporting

Managing the data in the data warehouse

Exporting data from the data warehouse

Classification, importance, isolation
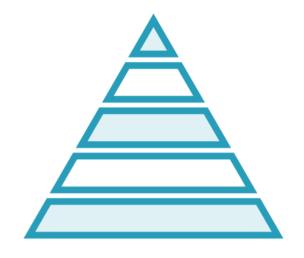
Workload Classification

Assign users to a role that has a corresponding resource class

CREATE WORKLOAD CLASSIFIER

A classification that is based on a database user will take precedence over a role membership

Workload Importance

5 Levels of importance: low, below_normal, normal, above_normal, and high

A request with higher importance will be run before a request with lower importance

Can prevent lock contention
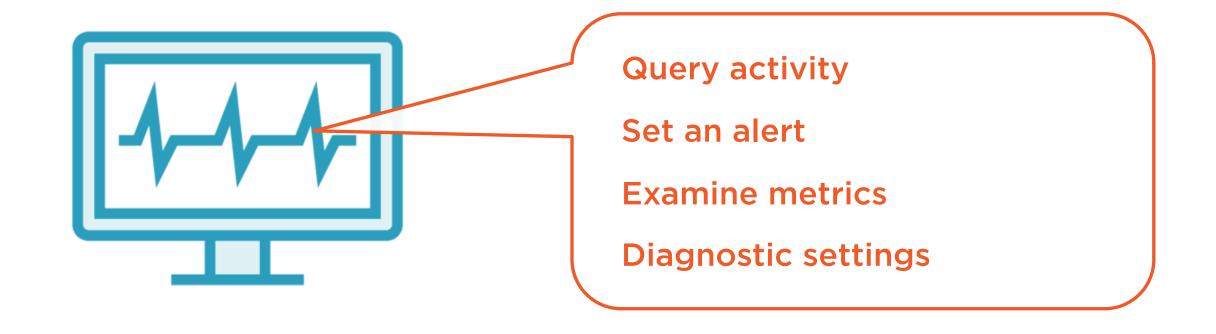
Workload Isolation

Uses workload groups to define isolation

Guarantees implicit level of concurrency with the MIN_PERCENTAGE_RESOURCE parameter

Define a maximum level of concurrency with the CAP_PERCENTAGE_RESOURCE parameter

Implementing Azure Synapse Analytics Monitoring

Query activity

Set an alert

Examine metrics

Diagnostic settings

Deleting an Azure Synapse Analytics SQL Pool

Delete button and fill in name

Irreversible

Suggested at end of course