



# DP-200T01: Monitoring and Troubleshooting Data Storage and Processing



# Agenda

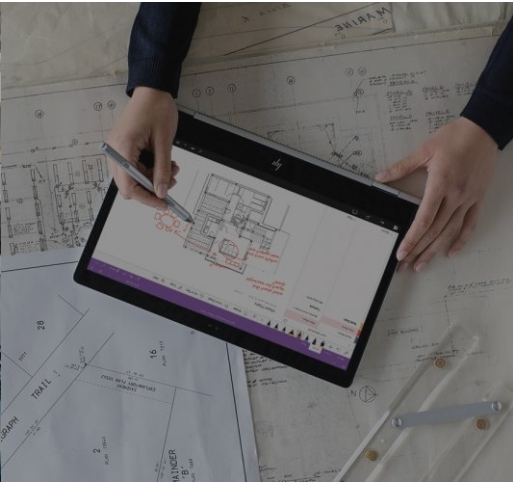
- L01 - General Azure monitoring capabilities
- L02 - Troubleshoot common data storage issues
- L03 - Troubleshoot common data processing issues
- L04 - Manage disaster recovery





# Lesson 01

## General Azure Monitoring Capabilities



# Lesson Objectives

- Azure Monitor
- Monitoring the network
- Diagnose and Solve Problems

# Azure Monitor

Azure Monitor provides a holistic monitoring approach by collecting, analyzing, and acting on telemetry from both cloud and on-premises environments

## Metric Data

Provides quantifiable information about a system over time that enables you to observe the behavior of a system.

## Log Data

Logs can be queried and even analyzed using Azure Monitor logs. In addition, this information is typically presented in the overview page of an Azure Resource in the Azure portal.

## Alerts

Alerts notify you of critical conditions and potentially take corrective automated actions based on triggers from metrics or logs.

# Monitoring the network

Azure Monitor logs within Azure monitor has the capability to monitor and measure network activity.

## Network Performance Monitor

Network Performance Monitor measures the performance and reachability of the networks that you have configured.

## Application Gateway Analytics

Application Gateway Analytics contains rich, out-of-the box views you can get insights into key scenarios, including:

- Monitor client and server errors.
- Check requests per hour

# Diagnose and Solve Issues

Home > ctocdb - Diagnose and solve problems

## ctocdb - Diagnose and solve problems

Azure Cosmos DB account

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Quick start

Notifications

Data Explorer

Settings

Replicate data globally

Default consistency

Firewall and virtual networks

CORS

### RESOURCE HEALTH

✔ Available

There aren't any known problems affecting this Cosmos DB database account [More details](#)

### RECENT ACTIVITY

Activity for the past 24 hours

[Quick Insights](#) | [See all activity](#)

### SOLUTIONS TO COMMON PROBLEMS

- My database is slow
- My request unit (RU) charging is unclear
- I need more storage/throughput
- My queries are slow
- MongoDB API Support
- Import MongoDB data into CosmosDB

# Review Questions

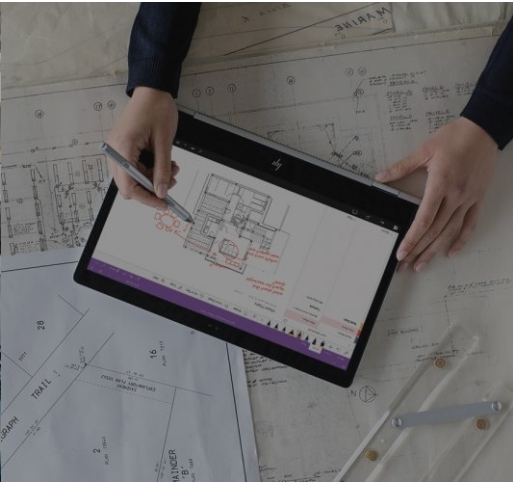
- Q01 – You want to respond to the critical condition and take corrective automated actions using Azure Monitor. Which feature would you use?
- A01 – Alerts
- Q02 – You are receiving an error message in Azure SQL Data Warehouse, You want to view information about the service and help to solve the problem, what can you use to quickly check the availability of the service?
- A02 – Diagnose and solve problem





# Lesson 02

## Troubleshoot Common Data Storage Issues



# Lesson Objectives

- Connectivity issues
- Performance issues
- Storage issues

# Connectivity Issues

There are a range of issues that can impact connectivity issues, including:

## Unable to connect to the data platform

- The first area that you should check is the firewall configuration.
- Test the connection by accessing it from a location external to your network.
- Check maintenance schedules

## Authentication Failures

- The first check is to ensure that the user name and password is correct.
- Check the storage account keys and ensure that they match in the connection string.

## Cosmos DB Mongo DB API errors

- Mongo client drivers establishes more than one connection.
- On the server side, connections which are idle for more than 30 minutes are automatically closed down.
- Check for timeouts

## SQL Database Failover

Should you receive an "unable to connect" message (error code 40613) in the Azure SQL Database, this scenario commonly occurs when a database has been moved because of deployment, failover, or load balancing.



# Performance Issues

Data Lake  
Storage

Cosmos DB

SQL Data  
Warehouse

SQL Database

Colocation of  
Resources

# Storage Issues

## Consistency

Consider the consistency levels of the following data stores that can impact data consistency:

- Cosmos DB
- SQL Data Warehouse
- SQL Database

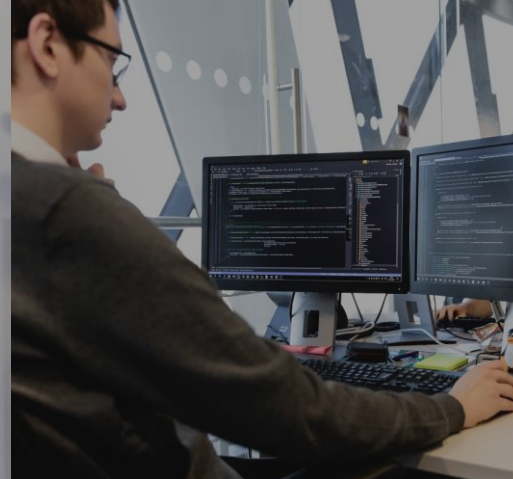
## Corruption

Data corruption can occur on any of the data platforms for a variety of reasons. You should have an appropriate disaster recovery strategy



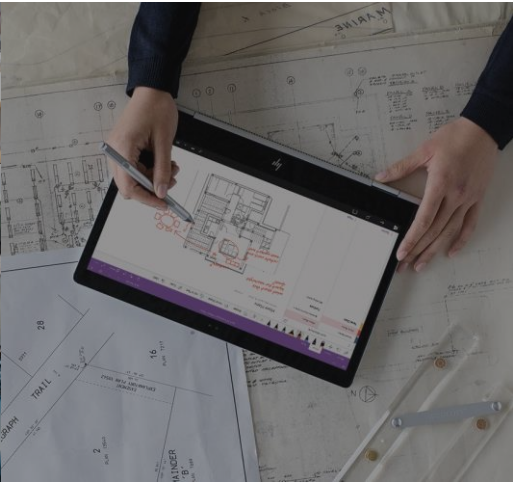
# Review Questions

- Q01 - You want to maximize the data integrity of data that is stored in a Cosmos DB. Which consistency level should you choose?
- A01 – Strong
- Q02 - You perform a daily SQL Data Warehouse using PolyBase with CTAS statements to load the data. User are reporting that the reports are running slow. What should you do to improve the performance of the query?
- A02 – Add table statistics are created and kept up to date



# Lesson 03

## Troubleshoot Common Data Processing Issues



# Lesson Objectives

- Troubleshoot streaming data
- Troubleshoot batch data loads
- Troubleshoot Azure Data Factory

# Troubleshoot streaming data

When using Stream Analytics, a Job encapsulates the Stream Analytic work and is made up of three components:

## Job input

The job input contains a **Test Connection** button to validate that there is connectivity with the input. However, most errors associated with a job input is due to the malformed input data that is being ingested.

## Job query

A common issue associated with Stream Analytics query is the fact that the output produced is not expected. In this scenario it is best to check the query itself to ensure that there is no mistakes on the code there.

## Job output

As with the job input, there is a **Test Connection** button to validate that there is connectivity with the output, should there be no data appearing. You can also use the **Monitor** tab in Stream Analytics to troubleshoot issues.

# Troubleshoot batch data loads

When trying to resolve data load issues, it is first pragmatic to make the holistic checks on Azure, as well as the network checks and diagnose and solve issue check. After that, then check:

## Azure Blob and Data Lake Store

Notwithstanding network errors; occasionally, you can get timeout or throttling errors that can be a symptom of the availability of the storage accounts.

## SQL Data Warehouse

- Make sure you are always leveraging PolyBase.
- Ensure CTAS statements are used to load data
- Break data down into multiple text files.
- Consider DWU usage

## Cosmos DB

- Check that you have provisioned enough RU's
- Review partitions and partitioning keys
- Check for client connection string settings

## SQL Database

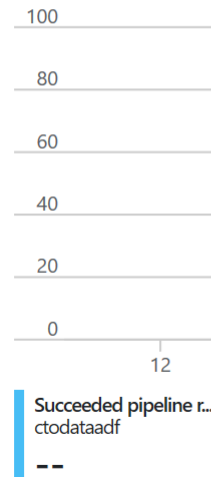
- Check that you have provisioned enough DTU's
- Review whether the database would benefit from elastic pools
- A wide range of tools can be used to troubleshoot SQL Database



# Troubleshoot Azure Data Factory

## Monitoring

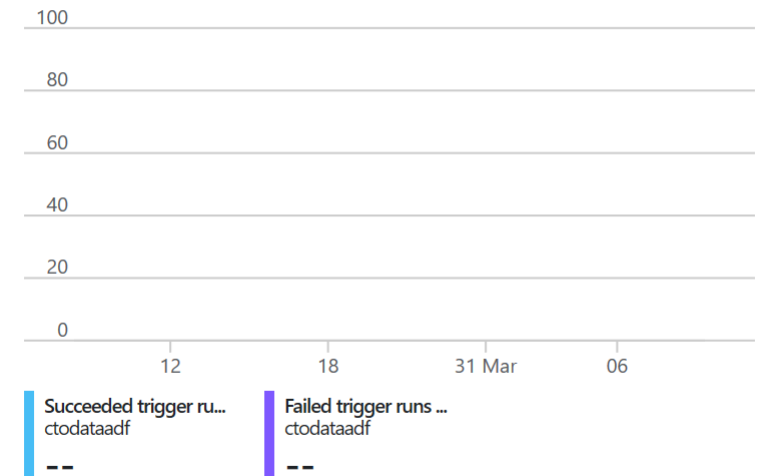
### PipelineRuns



### ActivityRuns



### TriggerRuns



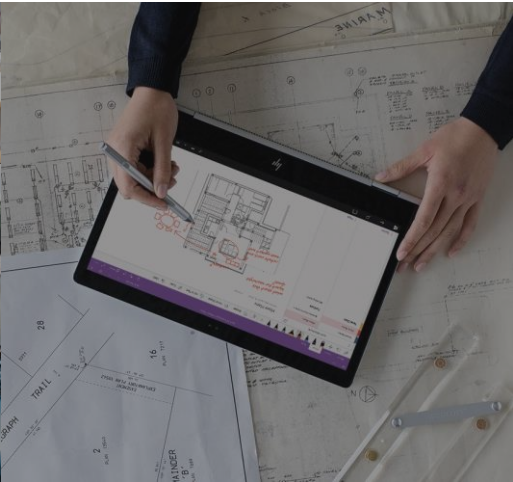
# Review Questions

- Q01 – What is the maximum number of activities per pipeline in Azure Data Factory?
- A01 – 40
- Q02 – You are monitoring the job output of a streaming analytics job. The monitor is reporting back that Runtime Errors > 0. What does this mean?
- A02 – The job can receive the data but is generating errors while processing the query.



# Lesson 04

## Managing Disaster Recovery



# Lesson Objectives

- Data redundancy
- Disaster recovery

# Data redundancy

Data redundancy is the process of storing data in multiple locations to ensure that it is highly available.

## Azure Blob and Data Lake Store

- Locally redundant storage (LRS)
- Zone-redundant storage (ZRS)
- Geo-redundant storage (GRS)
- Read-access geo-redundant storage (RA-GRS)

## SQL Data Warehouse

SQL Data Warehouse performs a **geo-backup** once per day to a paired data center. The RPO for a geo-restore is 24 hours.

## Cosmos DB

Azure Cosmos DB is a globally distributed database service. You can configure your databases to be globally distributed and available in any of the Azure regions.

## SQL Database

- Check that you have provisioned enough DTU's
- Review whether the database would benefit from elastic pools
- A wide range of tools can be used to troubleshoot SQL Database



# Disaster Recovery

There should be processes that are involved in backing up or providing failover for databases in an Azure data platform technology. Depending on circumstances, there are numerous approaches that can be adopted.

## Azure Blob and Data Lake Store

Supports account failover for geo-redundant storage accounts.

You can initiate the failover process for your storage account if the primary endpoint becomes unavailable.

## SQL Data Warehouse

SQL Data Warehouse performs a **geo-backup** once per day to a paired data center.

Data warehouse snapshot feature that enables you to create a restore point to create a copy of the warehouse to a previous state.

## Cosmos DB

Takes a backup of your database every **4 hours** and at any point of time

Only the latest 2 backups are stored.

## SQL Database

Creates database backups that are kept between 7 and 35 days

Uses Azure read-access geo-redundant storage (RA-GRS) to ensure that they preserved even if data center is unavailable.

# Review Questions

- Q01 – How long is the Recovery Point Objective for Azure SQL Data Warehouse?
- A01 – 8 hours
- Q02 – How often is a backup taken for Azure Cosmos DB?
- A02 - 4 hours

# Lab: Monitoring and Troubleshooting Data Storage and Processing



# Lab overview

The students will be able to define a broad monitoring solution that can help them monitor issues that can occur in their data estate. The student will then experience common data storage issues and data processing issue that can occur in cloud data solution. Finally they will implement a disaster recovery approach for a Data Platform technology.

## Lab objectives

After completing this lab, you will be able to:

1. Explain the monitoring capabilities that are available
2. Troubleshoot common data storage issues
3. Troubleshoot common data processing issues
4. Manage disaster recovery

# Lab scenario

As the Senior Data Engineer at AdventureWorks you have been tasked with defining the standard operating procedures for monitoring that data estate within the organization. You will start by defining the monitoring tools that will be used to support the approach.

You will then explore some of the common data storage and data processing issues that can occur during the normal operation of your infrastructure. Using the troubleshooting approach and the monitoring tools that you have defined in your standard operating procedures, you will solve the issue that are presented.

You have been also asked to define the disaster recovery approach for the data stores that you have created. You will document and implement the relevant changes noting the justification for the approach you have taken.

At the end of this lab, you will have:

1. Explain the monitoring capabilities that are available
2. Troubleshoot common data storage issues
3. Troubleshoot common data processing issues
4. Manage disaster recovery



# Lab review

- Exercise 1 – Who else should you involve when defining the monitoring capabilities for your organization?
- Exercise 2 – Are there other data storage issues that you face when working with data platform technologies on Azure?
- Exercise 3 – Are there other data processing issues that you face when working with data platform technologies on Azure?
- Exercise 4 – Do you have a Business Continuity/Disaster Recovery plan within your organization?

# Module Summary >

## In this module, you have learned about:

- General Azure monitoring capabilities
- Troubleshoot common data storage issues
- Troubleshoot common data processing issues
- Manage disaster recovery

## Next steps >

After the course, watch [Rahul Bagaria joins Lara Rubbelke discuss Azure monitor in this 16 minute video.](#)

