O'REILLY®

Implementing an Azure Data Solution

Crash Course

Microsoft Certified: Azure Data Engineer Associate

December/2020

Reza Salehi

Cloud Consultant and Trainer

O'Reilly Media Instructor, Pluralsight Author

@zaalion

Microsoft CERTIFIED Trainer 2008 - 2018

Microsoft CERTIFICATION EXAM DP-200

Microsoft CERTIFIED AZURE DATA ENGINEER ASSOCIATE

Questions & Resources

- Post questions in the QnA box
- Resources are in the course repository
  - https://github.com/zaalion/oreilly-dp-200-201

- Reach out:
  - Twitter: @zaalion

Course Overview

DP-200 Candidate Profile

- Microsoft Azure data engineers who
  - Collaborate with business stakeholders to identify and meet the data requirements
  - To implement data solutions that use Azure data services.

Azure Data Engineers

- Responsible for data-related implementation tasks
  - Include provisioning data storage services
  - Ingesting streaming and batch data, transforming data.
  - Implementing security requirements
  - Implementing data retention policies
  - Identifying performance bottlenecks, and
  - Accessing external data sources.

DP-200 Candidates

- Must be able to implement data solutions that use
  - Azure Cosmos DB, Azure SQL Database, Azure Synapse Analytics (formerly Azure SQL DW), Azure Data Lake Storage, Azure Data Factory, Azure Stream Analytics, Azure Databricks, and Azure Blob storage.

DP-200 Skills Measured

- Skills measured:
  - Implement data storage solutions (40-45%)
  - Manage and develop data processing (25-30%)
  - Monitor and optimize data solutions (30-35%)

DP-201 Skills Measured

Exam DP-200: Implementing an Azure Data Solution skills

Implement Data Storage Solutions

Implement Data Storage Solutions

- Implement non-relational data stores
- Implement relational data stores
- Manage data security

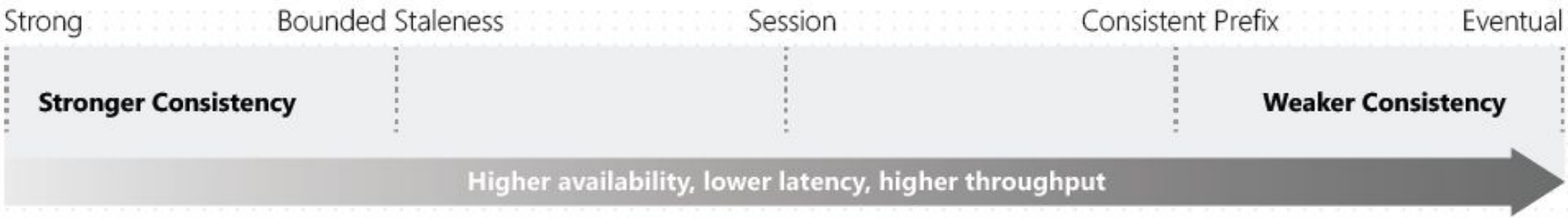
Implement Non-relational Data Stores

- Implement a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage
- Implement data distribution and partitions
- Implement a consistency model in Cosmos DB
- Provision a non-relational data store
- Provide access to data to meet security requirements
- Implement for high availability, disaster recovery, and global distribution

Plan for Secure Endpoints

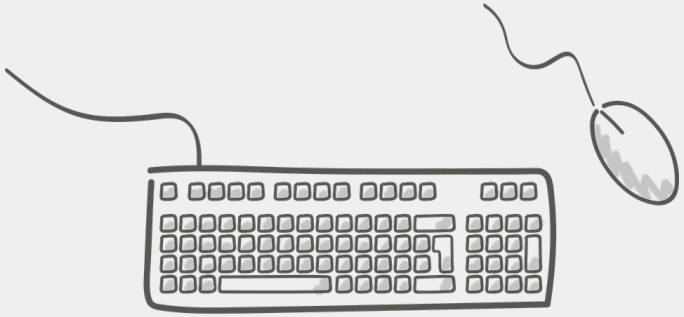
- Secure endpoints:
  - Azure Cosmos DB
  - Azure Storage Account
  - Azure Synapse Analytics
  - Azure Data Factory
  - Azure Databricks

Cosmos DB Security



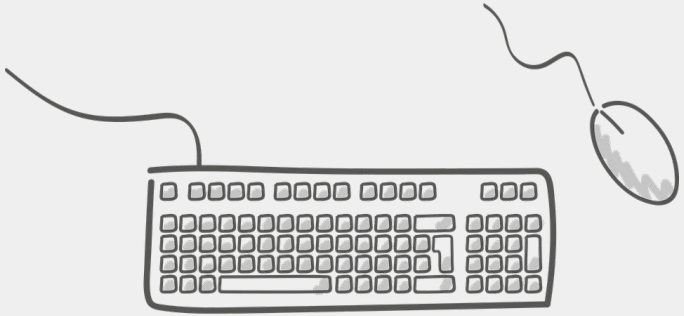
https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

Demo



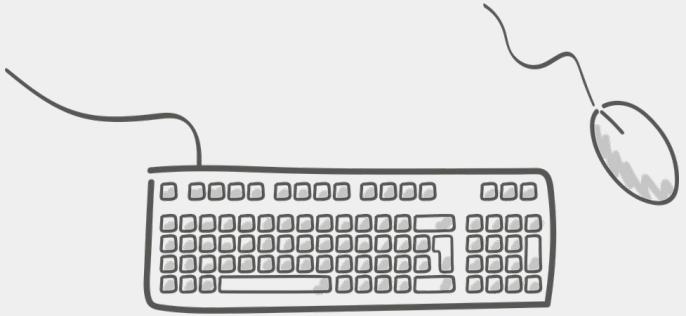
- Azure Cosmos DB
  - Provisioning
  - Data explorer
  - Throughput
  - Security
  - Disaster recovery
  - Consistency levels

Demo



- Azure Blob Storage
  - Provisioning
  - Data explorer
  - Security
  - Tiers

Demo



- Azure Data Lake Gen 2

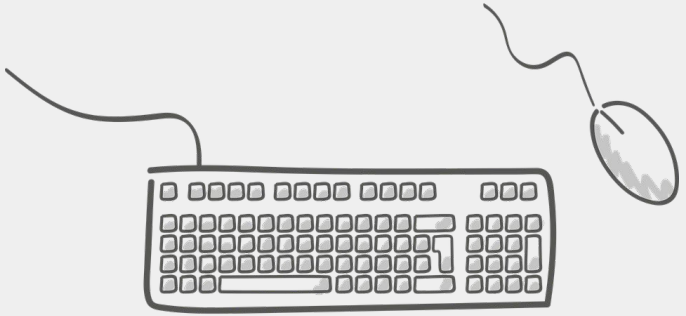
Implement relational data stores

- Provide access to data to meet security requirements
- Implement for high availability and disaster recovery
- Implement data distribution and partitions for Azure Synapse Analytics
- Implement PolyBase

Plan for Secure Endpoints

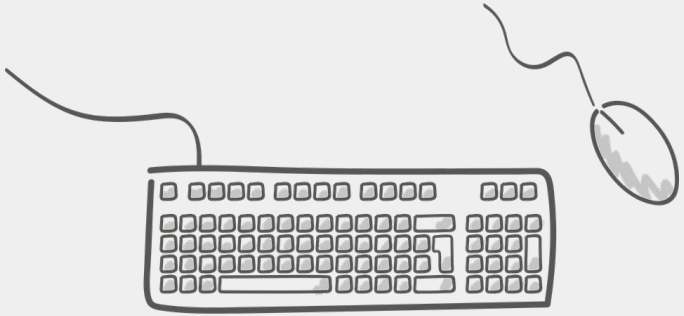
- Azure Synapse access:
  - Firewall
  - Azure Active Directory
  - SQL authentication

Demo



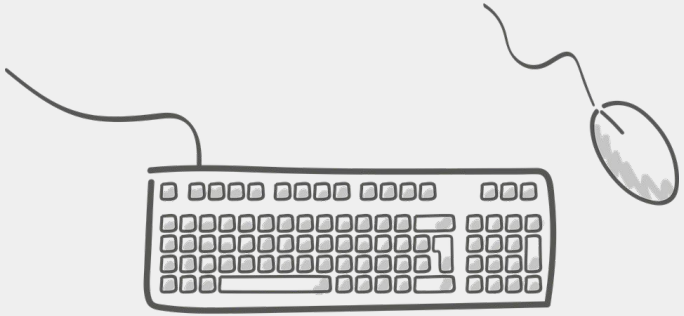
- Azure Synapse Analytics
  - Provisioning and access
  - Querying/Analytics

Demo



- Azure Synapse Analytics
  - High availability and disaster recovery

Demo



- Azure Synapse Analytics
  - Partitioning and data distribution

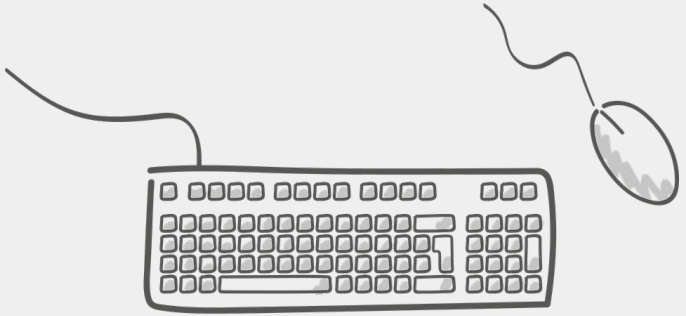
Manage Data Security

- Implement data masking
- Encrypt data at rest and in motion

Plan for Secure Endpoints

- Azure Synapse security options
  - Dynamic data masking
  - Row level security
  - Transparent Data Encryption
  - Always Encrypted

Demo



- Azure Synapse Analytics
  - Securing data

Manage and Develop Data Processing

Manage and Develop Data Processing

- Develop batch processing solutions
- Develop streaming solutions

Develop Batch Processing Solutions

- Develop batch processing solutions by using Data Factory and Azure Databricks
- Ingest data by using PolyBase
- Implement the integration runtime for Data Factory
- Create linked services and datasets
- Create pipelines and activities
- Create and schedule triggers
- Implement Azure Databricks clusters, notebooks, jobs, and autoscaling
- Ingest data into Azure Databricks

Azure Data Factory

Code-Free ETL as a Service

INGEST

- Multi-cloud and on-prem hybrid copy data
- 90+ native connectors
- Serverless and auto-scale
- Use wizard for quick copy jobs

CONTROL FLOW

- Design code-free data pipelines
- Generate pipelines via SDK
- Utilize workflow constructs: loops, branches, conditional execution, variables, parameters, ...

DATA FLOW

- Code-free data transformations that execute in Spark
- Scale-out with Azure Integration Runtimes
- Generate data flows via SDK
- Designers for data engineers and data analysts

SCHEDULE

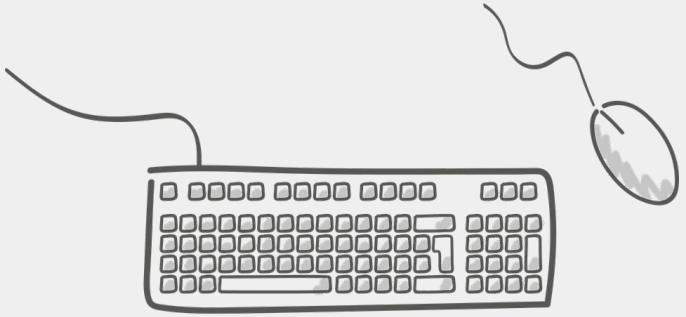
- Build and maintain operational schedules for your data pipelines
- Wall clock, event-based, tumbling windows, chained

MONITOR

- View active executions and pipeline history
- Detail activity and data flow executions
- Establish alerts and notifications

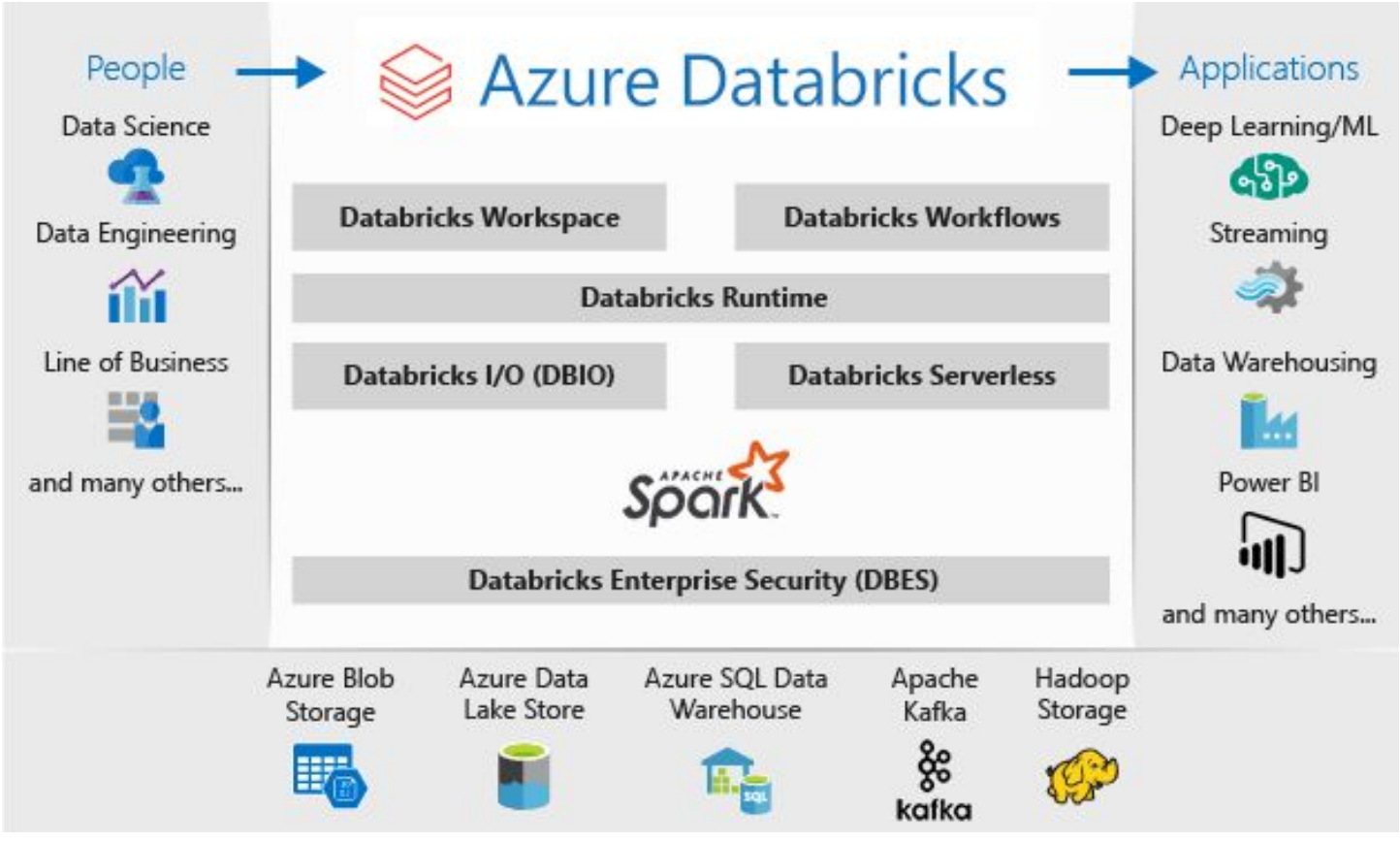
https://docs.microsoft.com/en-us/azure/data-factory/introduction

Demo

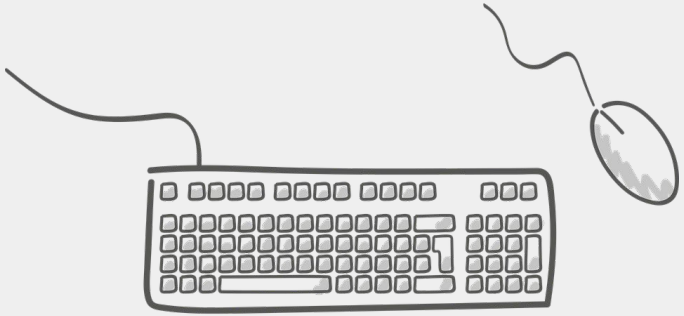


- Azure Data Factory
  - Pipelines
  - Linked services
  - Datasets
  - Schedules
  - Run history

Azure Databricks



Demo

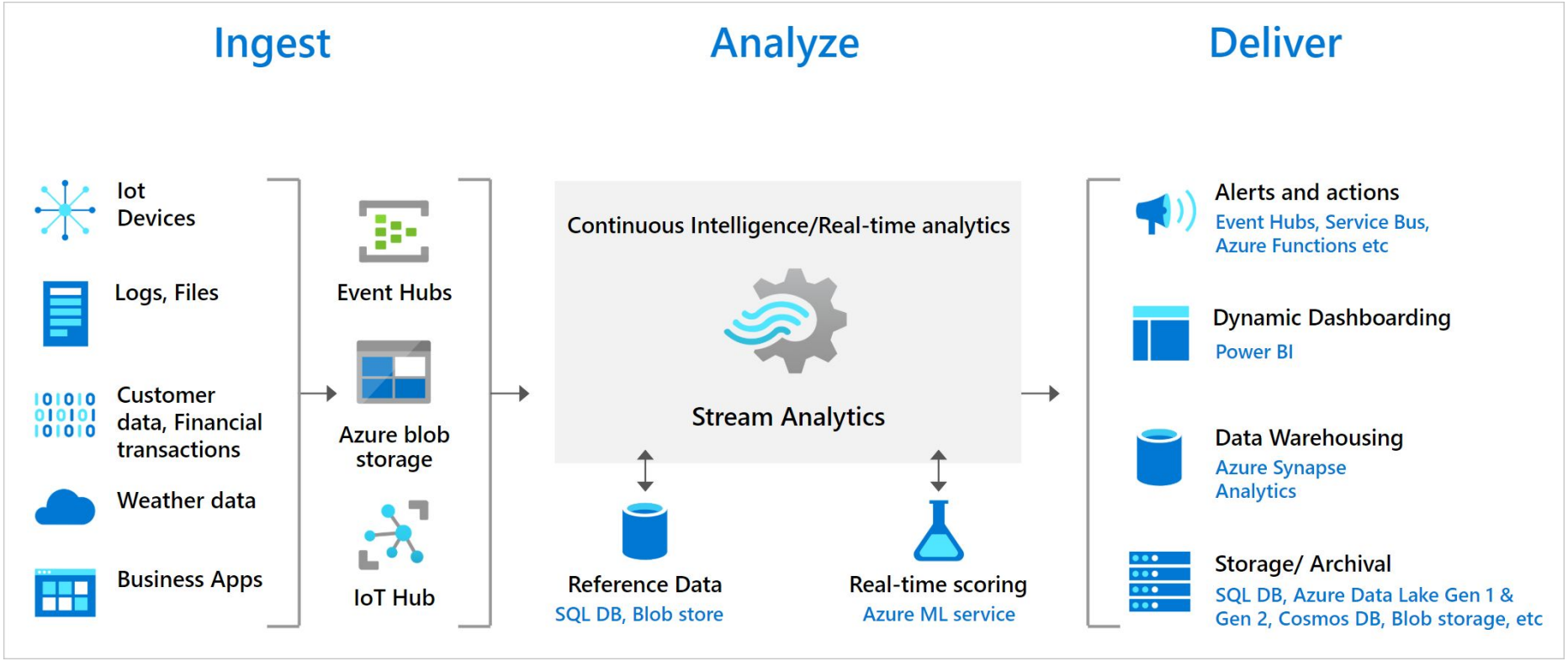


- Azure Databricks
  - Clusters
  - Notebooks
  - Jobs

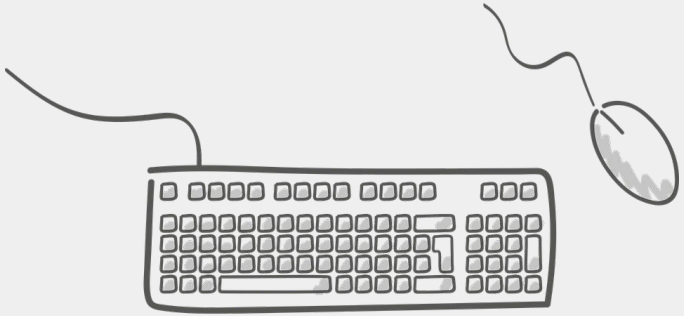
Develop Streaming Solutions

- Configure input and output
- Select the appropriate built-in functions
- Implement event processing by using Stream Analytics

Azure Stream Analytics



Demo



- Azure Stream Analytics
  - Inputs
  - Outputs
  - Functions
  - Windows
  - Jobs

Monitor and Optimize Data Solutions

Monitor and Optimize Data Solutions

- Monitor data storage
- Monitor data processing
- Optimize of Azure data solutions

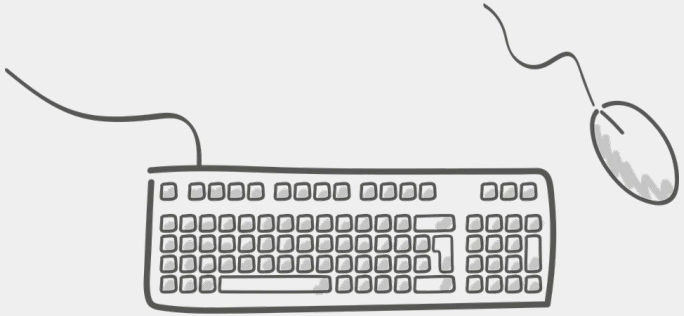
Monitor data storage

- Monitor relational and non-relational data stores
- Implement Blob storage monitoring
- Implement Data Lake Storage Gen2 monitoring
- Implement Azure Synapse Analytics monitoring
- Implement Cosmos DB monitoring

Monitor data processing

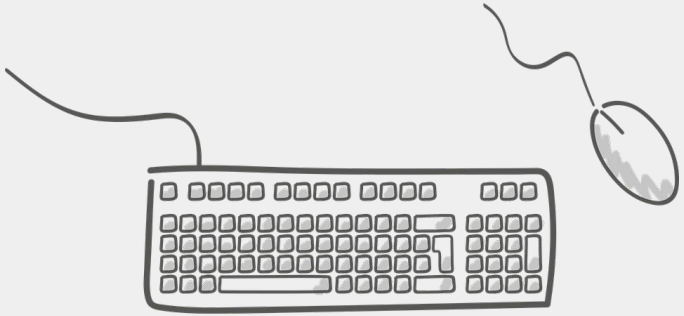
- Monitor Data Factory pipelines
- Monitor Azure Databricks
- Monitor Stream Analytics
- Configure Azure Monitor alerts
- Implement auditing by using Azure Log Analytics

Demo



- Configure Monitoring
  - Azure Storage
  - Azure Data Lake Gen 2
  - Azure Cosmos DB
  - Azure Synapse Analytics
  - Azure Stream Analytics

Demo



- Configure Monitoring
  - Alerts

Optimize of Azure Data Solutions

- Troubleshoot data partitioning bottlenecks
- Optimize Data Lake Storage Gen2
- Optimize Stream Analytics
- Optimize Azure Synapse Analytics
- Manage the data lifecycle

The Exam

Questions in DP-200

- Multiple choice
- Drag and drop
- Scenario based
- No hands-on labs (as of December 10, 2020)

DP-200

- Exam DP-200 : https://docs.microsoft.com/en-us/learn/certifications/exams/dp-200
- Skills measured : https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3Vzx2

storage services, ingesting streaming and batch data, transforming data, implementing security requirements, implementing data retention policies, identifying performance bottlenecks, and accessing external data sources.

Candidates for this exam must be able to implement data solutions that use the following Azure services: Azure Cosmos DB, Azure SQL Database, Azure Synapse Analytics (formerly Azure SQL DW), Azure Data Lake Storage, Azure Data Factory, Azure Stream Analytics, Azure Databricks, and Azure Blob storage.

Part of the requirements for: Microsoft Certified: Azure Data Engineer Associate
Related exams: 1 related exam
Important: See details
Go to Certification Dashboard

Schedule exam

Exam DP-200: Implementing an Azure Data Solution

United States

Languages: English, Japanese, Chinese (Simplified), Korean
Retirement date: none

This exam measures your ability to accomplish the following technical tasks: implement data storage solutions; manage and develop data processing; and monitor and optimize data solutions.

$165 USD*

Price based on the country in which the exam is proctored.

Schedule exam

Official practice test for Implementing an Azure Data Solution
All objectives of the exam are covered in depth so you'll be ready for any question on the exam.

Verify exam discount eligibility

For Microsoft employees

Microsoft employees are eligible for discounted exams. The discount will be reflected at the end of the checkout process. For MOS exams at Certiport, please request a voucher through the Microsoft Employee Voucher Portal.

To verify you are a Microsoft employee, link your Microsoft work account (alias@microsoft.com).

Link account

For Microsoft event attendees

If you recently attended a Microsoft event, you may be eligible for a discounted Microsoft Certification exam. To check eligibility, select an event you attended and verify the account used to register for the event.
Terms and Conditions apply.

Microsoft Ignite 2019, Orlando

Verify account

Continue scheduling exam

Proceed to the Pearson VUE website to complete the exam scheduling process.

Go to Pearson VUE

Pearson VUE - Select exam optio

wsr.pearsonvue.com/testtaker/registration/PrivateAccessCodePage/MICROSOFT?conversationId=4780073

Incognito

Microsoft

Select exam options

DP-200: Implementing an Azure Data Solution

All fields are required.

How do you want to take your exam? Exam delivery option descriptions

- ○ At a local test center
- ◉ At my home or office
- ○ I have a Private Access Code

Are you going to be testing on this device and network?

If so, perform a quick pre-check to verify compatibility of your device and network before planning to take this exam in your home or office.
If you skip, be sure to do a full system test before test day to avoid lost exam fees and launch delays.

Run pre-check

Next

Pearson VUE - Select exam optio | System Check

systemcheck.proctorcam.com/systemcheck?candidateId=247036059

Incognito

Pearson | VUE

System check - Checking your requirements

Microphone

Default - Microphone (SI

Internet speed

Webcam

Integrated Webcam (0c4

Next

Course Repository

https://github.com/zaalion/oreilly-dp-200-201

Q&A

O'REILLY®

Thank you!

Reza Salehi

@zaalion