

Designing and Implementing an Azure Data Solution Crash Course

O'REILLY MEDIA

December 7 and 10 2020

Instructor: Reza Salehi

@zaalion

Resources

Exam DP-201: Designing an Azure Data Solution

<https://docs.microsoft.com/en-us/learn/certifications/exams/dp-200>

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3Vzx2>

Exam DP-201: Designing an Azure Data Solution

<https://docs.microsoft.com/en-us/learn/certifications/exams/dp-201>

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VRMb>

Understand data store models

<https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview>

<https://docs.microsoft.com/en-ca/azure/storage/common/storage-redundancy>

Types of storage accounts

<https://docs.microsoft.com/en-us/azure/storage/common/storage-account-overview#types-of-storage-accounts>

Quickstart: Build a .NET console app to manage Azure Cosmos DB SQL API resources

<https://docs.microsoft.com/en-us/azure/cosmos-db/create-sql-api-dotnet>

Quickstart: Build a .NET Framework or Core application using the Azure Cosmos DB Gremlin API account

<https://docs.microsoft.com/en-us/azure/cosmos-db/create-graph-dotnet>

Quickstart: Migrate an existing MongoDB Node.js web app to Azure Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/create-mongodb-nodejs>

Azure Cognitive Search documentation

<https://docs.microsoft.com/en-us/azure/search/>

Azure Cache for Redis

<https://docs.microsoft.com/en-us/azure/azure-cache-for-redis/cache-overview>

Azure SQL documentation

<https://docs.microsoft.com/en-us/azure/azure-sql/>

Other Relational options

<https://docs.microsoft.com/en-us/azure/mysql/>

<https://docs.microsoft.com/en-us/azure/postgresql/>

<https://docs.microsoft.com/en-us/azure/mariadb/>

Azure Table storage overview

<https://docs.microsoft.com/en-us/azure/cosmos-db/table-storage-overview>

Azure Table Storage vs. Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/table-support>

Horizontal, vertical, and functional data partitioning

<https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning#why-partition-data>

Designing partitions

<https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning#designing-partitions>

Azure Architecture Center - Databases

<https://docs.microsoft.com/en-us/azure/architecture/browse/#databases>

Azure Architecture Center - Storage

<https://docs.microsoft.com/en-us/azure/architecture/browse/#storage>

Serverless web application on Azure (design ideas)

<https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/serverless/web-app>

Azure Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/introduction>

<https://docs.microsoft.com/en-us/azure/cosmos-db/introduction#next-steps>

<https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally>

<https://docs.microsoft.com/en-us/azure/cosmos-db/partitioning-overview>

<https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability>

<https://docs.microsoft.com/en-us/azure/cosmos-db/request-units>

<https://docs.microsoft.com/en-us/azure/architecture/framework/resiliency/backup-and-recovery#azure-cosmos-db>

<https://docs.microsoft.com/en-us/azure/cosmos-db/online-backup-and-restore>

Introduction to Azure Data Lake Storage Gen2

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>

Introduction to Azure Blob storage

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>

Azure Storage redundancy

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

Data cache

<https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/data-cache-with-redis-cache>

Gaming using Cosmos DB (design ideas)

<https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/gaming-using-cosmos-db>

Configure multiple write-regions for Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-manage-database-account#configure-multiple-write-regions>

Distribute your data globally with Azure Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally>

Consistency levels in Azure Cosmos DB

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

Azure Storage redundancy

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

How does Azure Cosmos DB provide high availability

<https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability>

Partitioning tables in dedicated SQL pool - Azure Synapse

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition>

Manage compute for dedicated SQL pool (formerly SQL DW) in Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-manage-compute-overview>

Data Warehouse Units (DWUs) for dedicated SQL pool (formerly SQL DW) in Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/what-is-a-data-warehouse-unit-dwu-cdwu>

What is a dedicated SQL pool (formerly SQL DW) in Azure Synapse Analytics?

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-what-is>

Backup and restore in Azure Synapse Dedicated SQL pool

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/backup-and-restore>

<https://docs.microsoft.com/en-ca/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture#azure-storage>

Azure Synapse Hash-distributed tables

<https://docs.microsoft.com/en-ca/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture#hash-distributed-tables>

Azure Synapse Round-robin distributed tables

<https://docs.microsoft.com/en-ca/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture#round-robin-distributed-tables>

Azure Synapse Replicated Tables

<https://docs.microsoft.com/en-ca/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture#replicated-tables>

QuickStart: Use serverless SQL pool

<https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-sql-on-demand>

Azure Synapse Analytics terminology

<https://docs.microsoft.com/en-us/azure/synapse-analytics/overview-terminology>

Transact-SQL features supported in Azure Synapse SQL (Dedicated vs. Serverless SQL pools)

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/overview-features>

Load data from Azure Data Lake Storage into dedicated SQL pools in Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store>

Securely load data using Synapse SQL

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-examples>

How to Access Synapse Studio

<https://docs.microsoft.com/en-us/azure/synapse-analytics/get-started-create-workspace#open-synapse-studio>

Download and install Azure Data Studio

<https://docs.microsoft.com/en-us/sql/azure-data-studio/download-azure-data-studio?view=sql-server-ver15>

Connect to Synapse SQL with Azure Data Studio

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/get-started-azure-data-studio>

Azure Storage Explorer

<https://azure.microsoft.com/en-us/features/storage-explorer/>

Use PolyBase to load data into Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse#use-polybase-to-load-data-into-azure-synapse-analytics>

High availability for Azure SQL Database and SQL Managed Instance

<https://docs.microsoft.com/en-us/azure/azure-sql/database/high-availability-sla>

High availability for Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/migrate/azure-best-practices/analytics/azure-synapse>

What is Azure Data Factory?

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

Linked services in Azure Data Factory

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-linked-services>

Pipelines and activities in Azure Data Factory

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities>

Datasets in Azure Data Factory

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services>

Mapping data flows in Azure Data Factory

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-data-flow-overview>

Create a trigger that runs a pipeline on a schedule.

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger>

Copy data from and to a REST endpoint by using Azure Data Factory

<https://docs.microsoft.com/en-us/azure/data-factory/connector-rest>

Batch processing

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/batch-processing>

Azure Data Lake Best practices

<https://docs.microsoft.com/en-us/azure/databricks/delta/best-practices>

Azure Architecture Center - Analytics

<https://docs.microsoft.com/en-us/azure/architecture/browse/#analytics>

Azure Stream Analytics

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction>

Understand time handling in Azure Stream Analytics

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling>

What is Azure Databricks?

<https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks>

Azure Databricks Workspace concepts

<https://docs.microsoft.com/en-ca/azure/databricks/getting-started/concepts>

Azure Databrick Clusters

<https://docs.microsoft.com/en-us/azure/databricks/clusters/>

Cluster size and autoscaling

<https://docs.microsoft.com/en-us/azure/databricks/clusters/configure#autoscaling>

Azure Databricks Notebooks

<https://docs.microsoft.com/en-us/azure/databricks/notebooks/>

Azure Databricks Jobs

<https://docs.microsoft.com/en-us/azure/databricks/jobs>

Transformation with Azure Databricks

<https://docs.microsoft.com/en-us/azure/data-factory/solution-template-databricks-notebook>

Grant limited access to Azure Storage resources using shared access signatures (SAS)

<https://docs.microsoft.com/en-us/azure/storage/common/storage-sas-overview>

Azure role-based access control in Azure Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/role-based-access-control>

Manage storage account access keys

<https://docs.microsoft.com/en-us/azure/storage/common/storage-account-keys-manage?tabs=azure-portal>

Auditing for Azure SQL Database and Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview>

Cosmos DB Consistency models

<https://www.youtube.com/watch?v=t1--kZjrG-o>

Dynamic data masking

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

Data Discovery & Classification

<https://docs.microsoft.com/en-us/azure/azure-sql/database/data-discovery-and-classification-overview>

Optimize costs by automating Azure Blob Storage access tiers

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal>

Soft delete for blobs

<https://docs.microsoft.com/en-us/azure/storage/blobs/soft-delete-blob-overview>

Data encryption in Azure Cosmos DB

<https://docs.microsoft.com/en-us/azure/cosmos-db/database-encryption-at-rest>

Azure Storage encryption for data at rest

<https://docs.microsoft.com/en-us/azure/storage/common/storage-service-encryption>

Transparent data encryption for SQL Database, SQL Managed Instance, and Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/azure-sql/database/transparent-data-encryption-tde-overview?tabs=azure-portal>

Azure SQL Always Encrypted

<https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15>

Azure SQL Dynamic Data Masking

<https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking?view=sql-server-ver15>

Tutorial: Extract, transform, and load data by using Azure Databricks

<https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse>

Notebooks Magic Commands

<https://ipython.readthedocs.io/en/stable/interactive/magics.html>

Azure Databricks pricing

<https://azure.microsoft.com/en-ca/pricing/details/databricks/>

Azure Databricks Clusters

<https://docs.microsoft.com/en-ca/azure/databricks/clusters/configure#cluster-mode>

Parquet file

<https://docs.databricks.com/data/data-sources/read-parquet.html>

Tutorial: Load the New York Taxicab dataset

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/load-data-from-azure-blob-storage-using-copy>

Use the Azure Data Lake Storage Gen2 URI

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction-abfs-uri>

Understand data store models

<https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview>

Choosing a big data storage technology in Azure

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-storage>

Choose a data storage approach in Azure

<https://docs.microsoft.com/en-us/learn/modules/choose-storage-approach-in-azure/>

Criteria for choosing a data store

<https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-considerations>

Horizontal, vertical, and functional data partitioning

<https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning>

Choose the appropriate API for Azure Cosmos DB

<https://docs.microsoft.com/en-us/learn/modules/choose-api-for-cosmos-db/>

High availability for Azure Synapse Analytics

<https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/migrate/azure-best-practices/analytics/azure-synapse>

What is Azure Synapse and how is it different from Azure Data Bricks and SQL?

<https://blog.bismart.com/en/azure-synapse-difference-from-azure-data-bricks-and-sql>

Tutorial: Stream data into Azure Databricks using Event Hubs

<https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-stream-from-eventhubs>

Introduction to Azure Blob storage

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>

Understanding block blobs, append blobs, and page blobs

<https://docs.microsoft.com/en-us/rest/api/storageservices/understanding-block-blobs--append-blobs--and-page-blobs>

Block blob pricing

<https://azure.microsoft.com/en-us/pricing/details/storage/blobs/>

Azure Pricing calculator

<https://azure.microsoft.com/en-ca/pricing/calculator/>

Azure Cosmos DB Gremlin graph support and compatibility with TinkerPop features

<https://docs.microsoft.com/en-us/azure/cosmos-db/gremlin-support>

HDInsight

<https://azure.microsoft.com/en-ca/services/hdinsight/>

Azure Cosmos DB capacity calculator

<https://cosmos.azure.com/capacitycalculator/>

Stream processing pipeline with Azure Databricks

<https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/data/stream-processing-databricks>

Introduction to Stream Analytics windowing functions

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

Built-in Functions (Azure Stream Analytics)

<https://docs.microsoft.com/en-us/stream-analytics-query/built-in-functions-azure-stream-analytics>

What is Azure Event Grid?

<https://docs.microsoft.com/en-us/azure/event-grid/overview>

What is PolyBase?

<https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver15>

Course Repository

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

My Related Pluralsight Course

Implementing NoSQL Databases in Microsoft Azure

<https://app.pluralsight.com/library/courses/microsoft-azure-nosql-databases-implementing/table-of-contents>

Building Streaming Data Pipelines in Microsoft Azure

<https://app.pluralsight.com/library/courses/microsoft-azure-building-streaming-data-pipelines/table-of-contents>

Implementing a Relational Database in Microsoft Azure SQL Database

<https://app.pluralsight.com/library/courses/microsoft-azure-implementing-relational-database-solutions/table-of-contents>

Microsoft Azure Developer: Securing Data

<https://app.pluralsight.com/library/courses/microsoft-azure-data-securing/table-of-contents>

Design and Document Data Flows with Microsoft Azure

<https://app.pluralsight.com/library/courses/microsoft-azure-data-flows-document-design/table-of-contents>

Thank you!

Reza Salehi

@zaalion