O'REILLY®

Exam DP-420: Microsoft Azure Cosmos DB Developer Crash Course

Developing Cloud Native Applications Using Cosmos DB



Reza Salehi

Cloud Consultant and Trainer











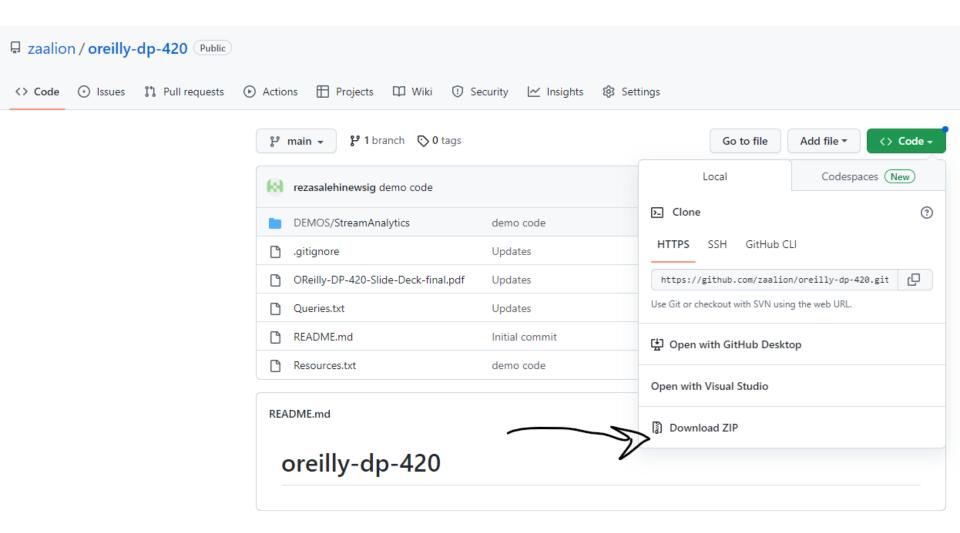


Course Overview

Course Repository

https://github.com/zaalion/oreilly-dp-420





Official DP-420 Exam Documentation

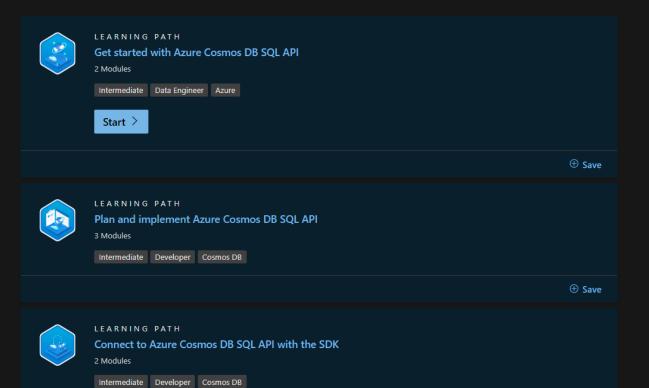
https://docs.microsoft.com/en-us/certifications/exams/dp-420



Two ways to prepare

Online - Free Instructor-led - Paid

Items in this collection





DP-420 Crash Course

- Design and implement data models
- Design and implement data distribution
- Integrate an Azure Cosmos DB solution
- Optimize an Azure Cosmos DB solution
- Maintain an Azure Cosmos DB solution



Design and Implement Data Models

Design a Non-relational Data Model for Azure Cosmos DB for NoSQL

- <u>Develop</u> a <u>design</u> by storing multiple entity types in the same container
- <u>Develop</u> a <u>design</u> by storing multiple related entities in the same document
- Develop a model that <u>denormalizes</u> data across documents
- Develop a <u>design</u> by <u>referencing</u> between documents
- <u>Identify primary</u> and <u>unique keys</u>
- Identify <u>data</u> and <u>associated access patterns</u>
- Specify a <u>default TTL on a container</u> for a transactional store



Cosmos DB is a Schema-free Database



Data Modeling in Azure Cosmos DB

- Embed data: Entities will be stored as self-contained items represented as JSON documents. (denormalized)
- Reference data: Entities will be stored as separate JSON documents with references to one another. (normalized)
- Hybrid data models: Combination of the above

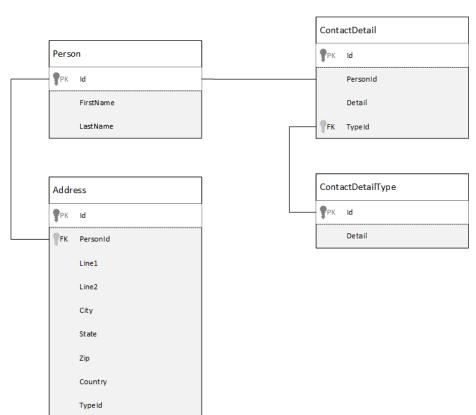


Embed Data





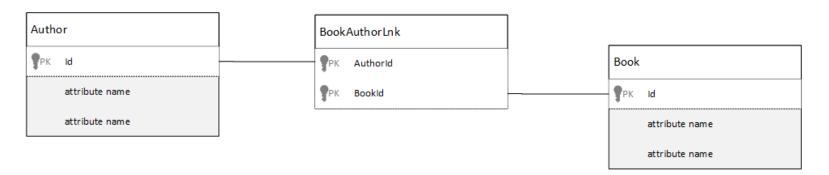
Embed Data



```
"id": "1",
"firstName": "Thomas",
"lastName": "Andersen",
"addresses": [
        "line1": "100 Some Street",
        "line2": "Unit 1",
        "city": "Seattle",
        "state": "WA",
        "zip": 98012
"contactDetails": [
    {"email": "thomas@andersen.com"},
    {"phone": "+1 555 555-5555", "extension": 5555}
```



Reference Data





Reference Data



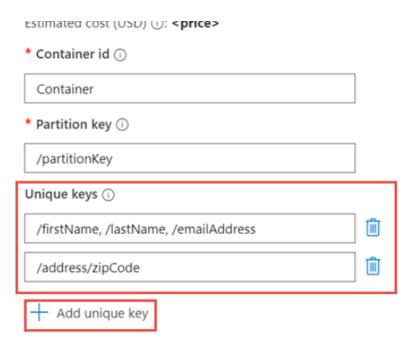
```
Publisher document:
{
    "id": "mspress",
    "name": "Microsoft Press",
    "books": [ 1, 2, 3, ..., 100, ..., 1000]
}

Book documents:
{"id": "1", "name": "Azure Cosmos DB 101" }
{"id": "2", "name": "Azure Cosmos DB for RDBMS Users" }
{"id": "3", "name": "Taking over the world one JSON doc at a time" }
...
{"id": "100", "name": "Learn about Azure Cosmos DB" }
...
{"id": "1000", "name": "Deep Dive into Azure Cosmos DB" }
```



Azure Cosmos DB Keys

- Partition Key (Primary Key)
- Unique Key





Time to Live (TTL) in Azure Cosmos DB

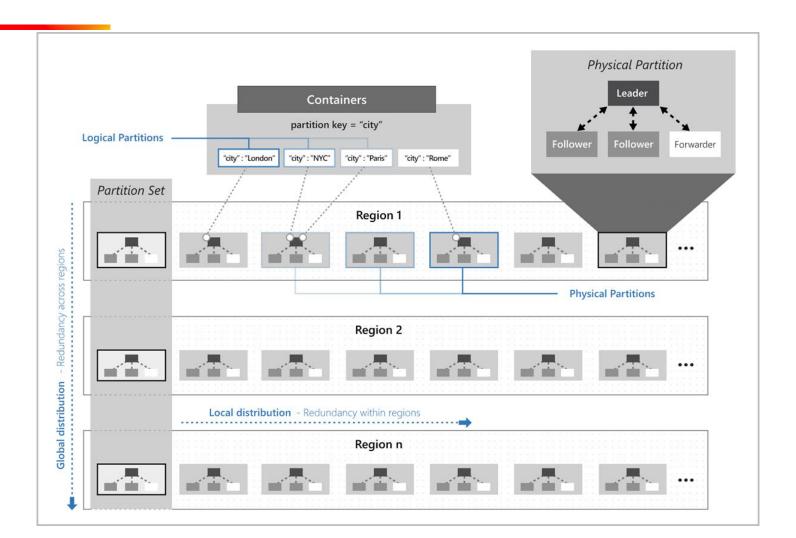
 Ability to delete items automatically from a container after a certain time period.



Design a Data Partitioning Strategy for Azure Cosmos DB for NoSQL

- Choose a <u>partitioning strategy</u> based on a specific workload
- Choose a <u>partition key</u>
- Plan for <u>transactions</u> when <u>choosing</u> a <u>partition</u> key
- Evaluate the cost of using a cross-partition query
- Calculate and evaluate data distribution based on partition key selection
- Calculate and evaluate <u>throughput</u> distribution based on partition key selection
- Construct and implement a synthetic partition key
- Design and implement a <u>hierarchical partition key</u>
- Design <u>partitioning</u> for workloads that require <u>multiple partition keys</u>







Azure Cosmos DB Partitions

- Logical partitions: A set of items that have the same partition key.
- <u>Physical</u> partitions: Smaller containers could have many logical partitions, but they might only need a single physical partition. Physical partitions are entirely managed by Azure Cosmos DB.



Azure Cosmos DB Partition Key

- Its value should not change
- Only "string" values
- This property should have a wide range of possible values to avoid "hot", or "cold" partitions.
- Spread process, and data storage evenly across all logical partitions as much as possible
- "A logical partition also defines the scope of database transactions"



Cosmos DB Query Types

<u>In-partition</u> query

Cross-partition query



Cosmos DB Query Types

In-partition query

```
SQL

SELECT * FROM c WHERE c.DeviceId = 'XMS-0001'
```

Cross-partition query

```
SQL

SELECT * FROM c WHERE c.Location = 'Seattle'
```



Synthetic Partition Key

- It is best to have a partition key with several values, in the hundreds or thousands
- This results in evenly distributing the data (and workload)
- In many cases, such a property doesn't exist in your data, so you can create a synthetic partition key.



Synthetic Partition Key

```
{
  "deviceId": "abc-123",
  "date": 2018
}
```

```
{
    "deviceId": "abc-123",
    "date": 2018,
    "partitionKey": "abc-123-2018"
}
```



Synthetic Partition Key Types

- Concatenate multiple properties of an item
- Use a partition key with a random suffix (e.g., 2022-09-09.24)
- Use a partition key with pre-calculated suffixes (e.g., Vehicle-Identification-Number, VIN)



Hierarchical Partition Key Types

 Configure up to a three-level hierarchy for your partition keys to further optimize data distribution and enable higher scale.

```
C#

// List of partition keys, in hierarchical order. You can have up to three levels of k
List<string> subpartitionKeyPaths = new List<string> {
    "/TenantId",
    "/UserId",
    "/SessionId"
};
```



Plan and Implement Sizing and Scaling for a Database Created with Azure Cosmos DB

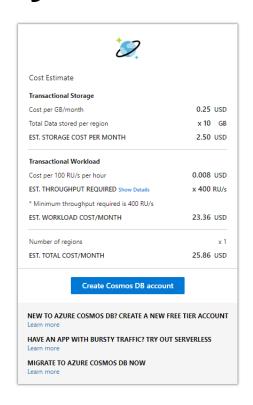
- Evaluate the <u>throughput</u> and data storage <u>requirements</u> for a <u>specific workload</u>
- Choose between <u>serverless and provisioned</u> models
- Choose when to use <u>database-level provisioned throughput</u>
- Design for granular <u>scale units</u> and <u>resource governance</u>
- Evaluate the cost of the global distribution of data
- Configure throughput for Azure Cosmos DB by using the <u>Azure portal</u>



Azure Cosmos DB Capacity Calculator

Azure Cosmos DB Account Settings The simplified Azure Cosmos DB calculator assumes commonly used settings for indexing policy, consistency, and other parameters. For a more accurate estimate, please sign in to provide your workload details. API 🚯 SQL (Core) Number of regions 1 Multi-region writes Workload per region For a more accurate cost estimate based on your own data, please sign in and upload your sample data. Total data stored in transactional store Off On Item size (upto 2048 KB) KB Creates/sec ① 10 10 Deletes/sec ① Oueries/sec @

Calculate





Cosmos DB Provisioning Models

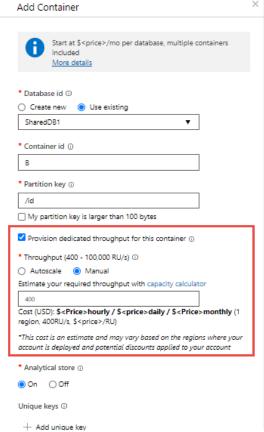
- Provisioned: You will set throughput (RUs) for your databases or/and containers
- <u>Auto-scale</u>: "Scale the throughput (RU/s) of your database or container automatically and instantly"
- Serverless: Use the Azure Cosmos account in a consumption-based mode

Auto scale vs. Serverless



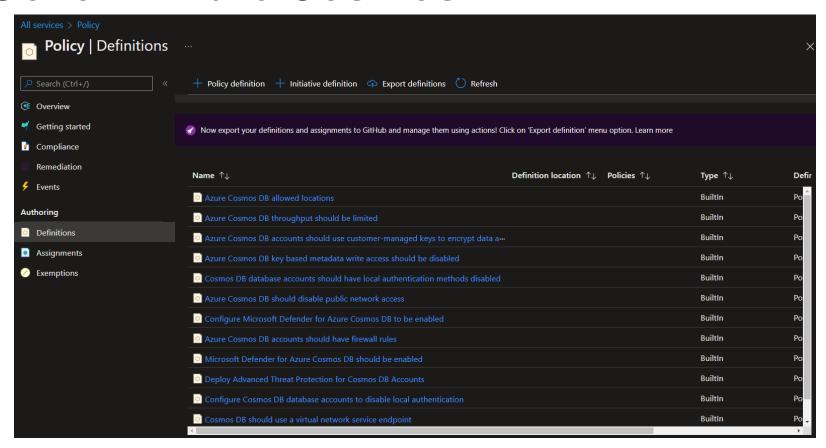
Provisioned throughput in Azure Cosmos DB

- At the Azure Cosmos DB containers level
- At the Azure Cosmos DB databases level





Govern Azure Cosmos DB





Implement Client Connectivity Options in the Azure Cosmos DB SDK

- Choose a <u>connectivity mode</u> (<u>gateway</u> versus <u>direct</u>)
- <u>Implement</u> a <u>connectivity mode</u>
- Create a connection to a database
- Enable offline development by using the <u>Azure Cosmos DB emulator</u>
- Handle connection errors (also see the these)
- Implement a <u>singleton</u> for the client
- Specify a region for global distribution
- Configure <u>client-side threading and parallelism</u> options
- Enable SDK logging

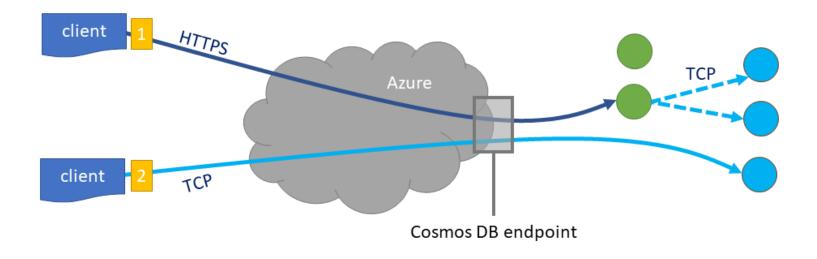


Azure Cosmos DB Connectivity Modes

1 Gateway mode (using HTTPS)
2 Direct mode over TCP

Gateway node

Backend node





Azure Cosmos DB Connectivity Modes

Gateway mode: Supported on all SDKs. Best option, if your application runs
within a corporate network with strict firewall restrictions. It uses the standard
HTTPS port and a single DNS endpoint.

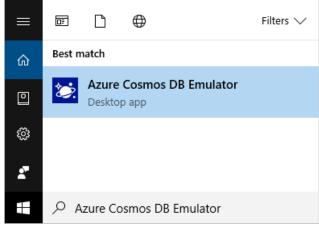
 <u>Direct mode</u>: Supports TCP connectivity, using TLS for authentication and traffic encryption. It offers better performance



Azure Cosmos DB Emulator

Allows developers to work offline without connecting to a live Azure Cosmos DB instance.

You can <u>download the emulator</u> from Microsoft





Checklist for Troubleshooting Issues

- Use the latest SDK.
- Review the performance tips and follow the suggested practices.
- Enable the SDK logging to help you troubleshoot an issue.
- Log metrics by using the Azure portal.
- Portal metrics show the Azure Cosmos DB telemetry
- Log the diagnostics string in the V2 SDK or diagnostics in V3 SDK from the point operation responses.
- Log the SQL Query Metrics from all the query responses



Connection Errors

- Request header too large
- Request timeout exceptions
- Slow requests
- Service unavailable exceptions



Implement Data Access by Using the Language for Azure Cosmos DB for NoSQL

- Implement <u>queries</u> that use <u>arrays</u>, <u>nested objects</u>, <u>aggregation</u>, and <u>ordering</u>
- Implement a <u>correlated subquery</u>
- Implement queries that use <u>array</u> and <u>type-checking functions</u>
- Implement queries that use <u>mathematical</u>, <u>string</u>, and <u>date functions</u>
- Implement queries based on <u>variable data</u>



Implement Data Access by Using Azure Cosmos DB for NoSQL SDKs

- Choose when to use a <u>point operation</u> versus a <u>query operation</u>
- Implement a point operation that creates, updates, and deletes documents
- Implement an update by using a <u>patch operation</u>
- Manage multi-document <u>transactions</u> using SDK <u>Transactional Batch</u>
- Perform a multi-document load using Bulk Support in the SDK
- Implement optimistic concurrency control using ETags
- Implement <u>session consistency</u> by using <u>session tokens</u>
- Implement a query operation that includes <u>pagination</u>
- Implement a query operation by using a continuation token
- Handle transient errors and 429s
- Specify <u>TTL</u> for a document
- Retrieve and use <u>query metrics</u>



Point Operation vs. Query Operation

- Ways to read data in Cosmos DB
- Point reads (SDK): A key/value lookup on a single item ID and partition key.
- Query reads

	Point read (assumes 1 KB item)	Query
Latency	Typically less than 10 ms	Variable
RU charge	1 RU	At least 2.3 RUs, variable
Number of items returned	1 item	Unlimited (if results size is too large, results are split across multiple pages)
Include partition key?	Required	Recommended



Document Update in Azure Cosmos DB

- Ways to update data in Cosmos DB
- Update the whole JSON record
- Partial document update



Azure Cosmos DB Consistency Levels





Specify TTL Using Cosmos DB SDK

```
.NET SDK v3
             Java SDK v4
                           Node SDK
                                       Python SDK
  C#
                                                                                                Copy
  Database database = client.GetDatabase("database");
  ContainerProperties properties = new ()
      Id = "container",
      PartitionKeyPath = "/customerId",
      // Never expire by default
      DefaultTimeToLive = -1
  };
  // Create a new container with TTL enabled and without any expiration value
  Container container = await database
      .CreateContainerAsync(properties);
```



Specify TTL Using Cosmos DB SDK

```
.NET SDK v3
             Java SDK v4
                           Node SDK
                                       Python SDK
  C#
                                                                                                Copy
  Database database = client.GetDatabase("database");
  ContainerProperties properties = new ()
      Id = "container",
      PartitionKeyPath = "/customerId",
      // Expire all documents after 90 days
      DefaultTimeToLive = 90 * 60 * 60 * 24
  };
  // Create a new container with TTL enabled and without any expiration value
  Container container = await database
      .CreateContainerAsync(properties);
```



Implement Server-side Programming in Azure Cosmos DB for NoSQL by Using JavaScript

- Write, deploy, and call a <u>stored procedure</u>
- Design stored procedures to work with multiple items <u>transactionally</u>
- Implement and call <u>triggers</u>
- Implement a <u>user-defined function</u>



Cosmos DB Stored Procedures

```
JavaScript
                                                                                                  Copy
var helloWorldStoredProc = {
    id: "helloWorld",
    serverScript: function () {
        var context = getContext();
        var response = context.getResponse();
        response.setBody("Hello, World");
```



Cosmos DB Pre-triggers

```
JavaScript
                                                                                                  Copy
function validateToDoItemTimestamp() {
    var context = getContext();
    var request = context.getRequest();
    // item to be created in the current operation
    var itemToCreate = request.getBody();
    // validate properties
    if (!("timestamp" in itemToCreate)) {
        var ts = new Date();
        itemToCreate["timestamp"] = ts.getTime();
    request.setBody(itemToCreate);
```



Cosmos DB Post-triggers

```
JavaScript
                                                                                                 Copy
function updateMetadata() {
    var context = getContext();
   var container = context.getCollection();
   var response = context.getResponse();
    // item that was created
   var createdItem = response.getBody();
   var filterQuery = 'SELECT * FROM root r WHERE r.id = " metadata"';
   var accept = container.queryDocuments(container.getSelfLink(), filterQuery,
        updateMetadataCallback);
    if(!accept) throw "Unable to update metadata, abort";
   function updateMetadataCallback(err, items, responseOptions) {
        if(err) throw new Error("Error" + err.message);
            if(items.length != 1) throw 'Unable to find metadata document';
            var metadataItem = items[0];
           metadataItem.createdItems += 1;
           metadataItem.createdNames += " " + createdItem.id;
            var accept = container.replaceDocument(metadataItem._self,
                metadataItem, function(err, itemReplaced) {
                        if(err) throw "Unable to update metadata, abort";
            if(!accept) throw "Unable to update metadata, abort";
```



Cosmos DB UDF

```
Copy
JavaScript
function tax(income) {
    if (income == undefined)
        throw 'no input';
    if (income < 1000)</pre>
        return income * 0.1;
    else if (income < 10000)
        return income * 0.2;
    else
        return income * 0.4;
```



Design and Implement Data Distribution

Design and implement a replication strategy for Azure Cosmos DB

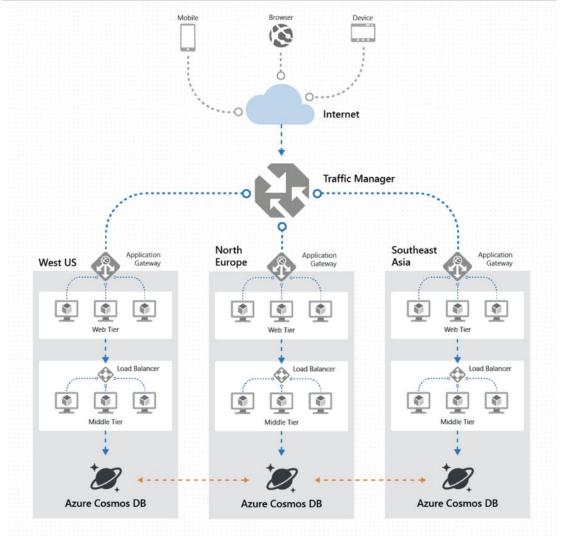
- Choose when to distribute data
- Automatic failover policies for regional failure for <u>Azure Cosmos DB</u> Core API
- Perform <u>manual failovers</u> to move <u>single master</u> write regions
- Choose a consistency model
- Identify <u>use cases</u> for different consistency models
- Evaluate the impact of consistency model choices on <u>availability</u> and associated RU cost
- Evaluate the impact of consistency model choices on <u>performance and latency</u>
- Specify <u>application</u> connections to <u>replicated</u> data



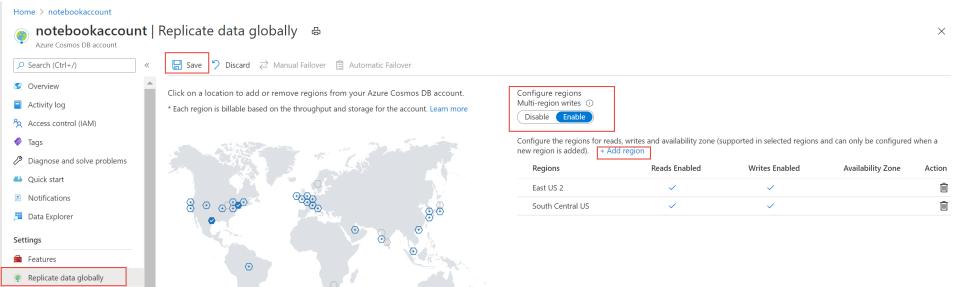
Distribute Data

- Azure Cosmos DB is a globally distributed database system
- Allows you to read and write data from the local region replicas of your database.
- Azure Cosmos DB transparently replicates the data to all the regions associated with your Cosmos account.









■ Default consistency



Design and Implement Multi-region Write

- Choose when to use <u>multi-region write</u>
- Implement multi-region write
- Implement a <u>custom conflict resolution policy</u> for <u>Azure Cosmos DB for NoSQL</u>



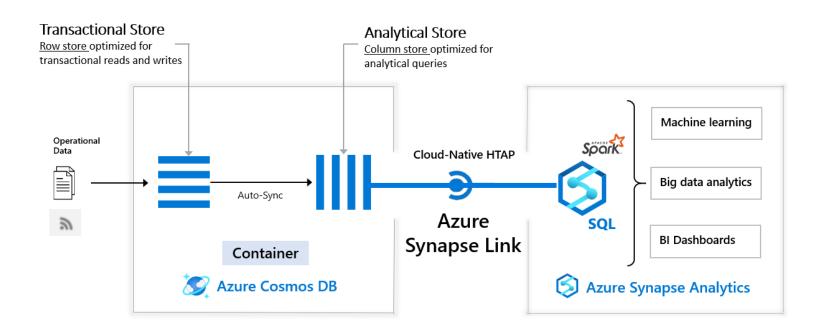
Integrate an Azure Cosmos DB Solution

Enable Azure Cosmos DB Analytical Workloads

- Enable <u>Azure Synapse Link</u>
- Choose between <u>Azure Synapse Link</u> and <u>Spark Connector</u>
- Enable the <u>analytical store</u> on a container
- Connection to an analytical store and query from <u>Azure Synapse Spark</u> or <u>Azure Synapse SQL</u>
- Perform a query against the <u>transactional</u> store from Spark
- Write data back to the transactional store from Spark

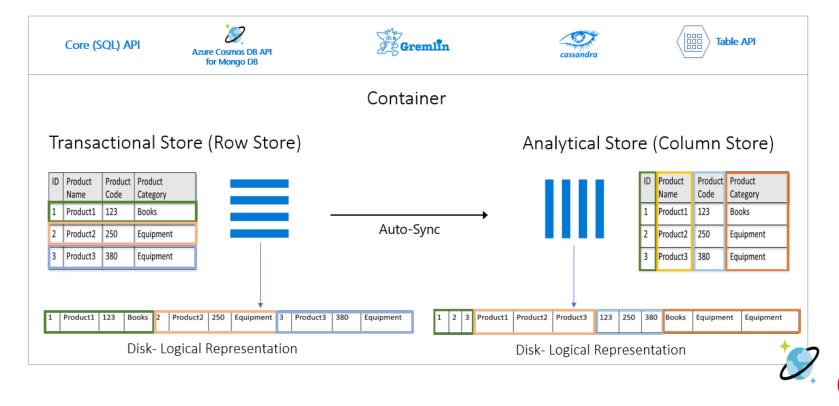


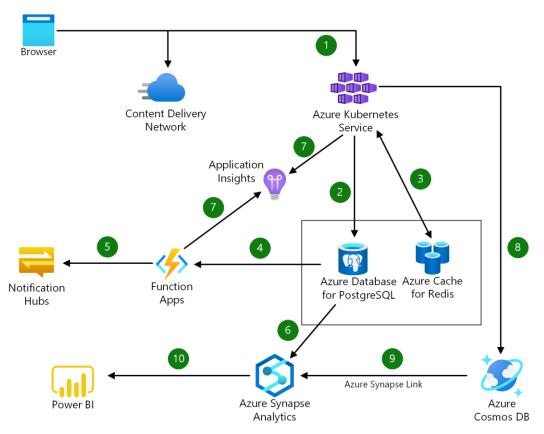
Azure Synapse Link





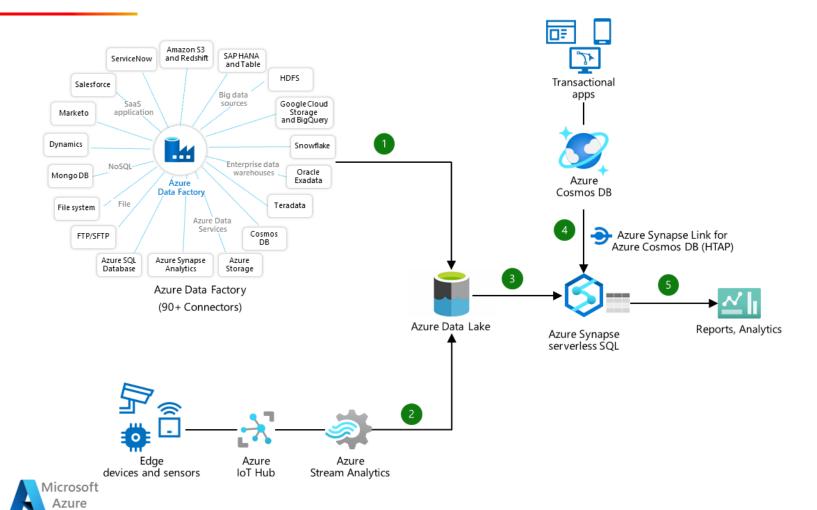
Azure Cosmos DB Analytical Store







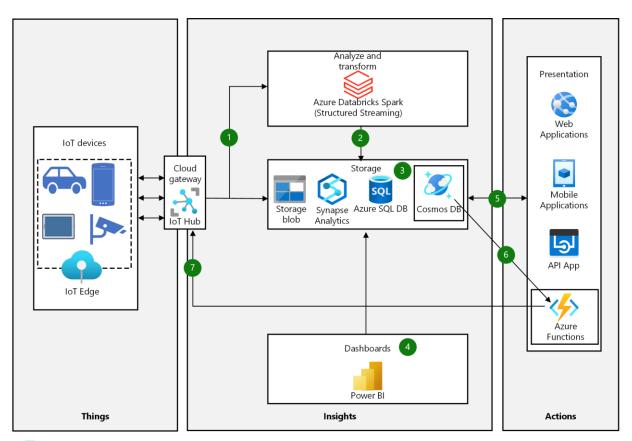




Write Spark DataFrame to Azure Cosmos DB

```
Copy
Java
// To select a preferred list of regions in a multi-region Azure Cosmos DB account, add option("spark.cos
import org.apache.spark.sql.SaveMode
df.write.format("cosmos.oltp").
    option("spark.synapse.linkedService", "<enter linked service name>").
    option("spark.cosmos.container", "<enter container name>").
    option("spark.cosmos.write.upsertEnabled", "true").
    mode(SaveMode.Overwrite).
    save()
```



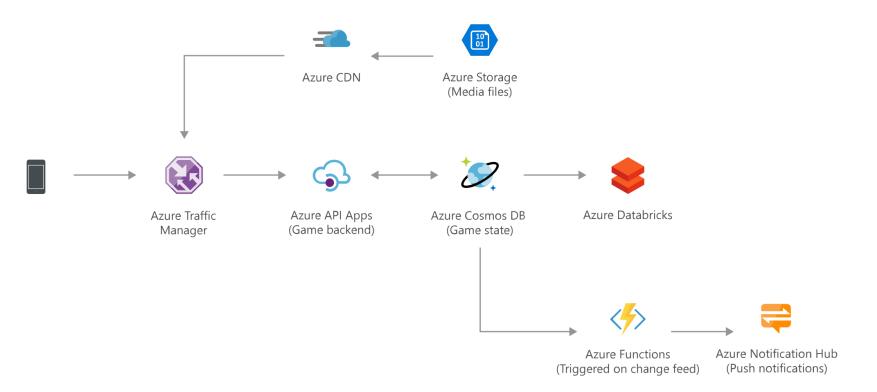






Online Recommendations Service **Azure Container Service** Azure Cosmos DB (Recommendations API) (Product + User Vectors) **Order Transactions** Shoppers Apache Spark on Azure Databricks E-commerce Store Azure API Apps Azure Cosmos DB Change Feed (Customer Order) (Customer Order)



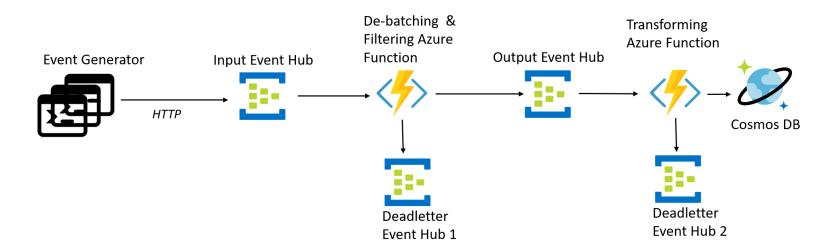




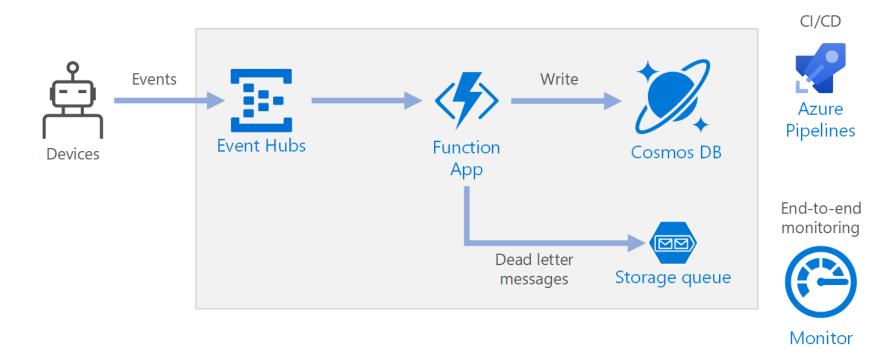
Implement Solutions Across Services

- Integrate events with other applications by using <u>Azure Functions</u> and <u>Azure</u>
 <u>Event Hubs</u>
- <u>Denormalize</u> data by using <u>Change Feed</u> and Azure <u>Functions</u>
- Enforce referential integrity by using Change Feed and Azure Functions
- Aggregate data by using Change Feed and Azure Functions, including reporting
- Archive data by using Change Feed and Azure Functions
- Implement <u>Azure Cognitive Search</u> for an Azure Cosmos DB solution











Optimize an Azure Cosmos DB Solution

Optimize Query Performance in Azure Cosmos DB for NoSQL

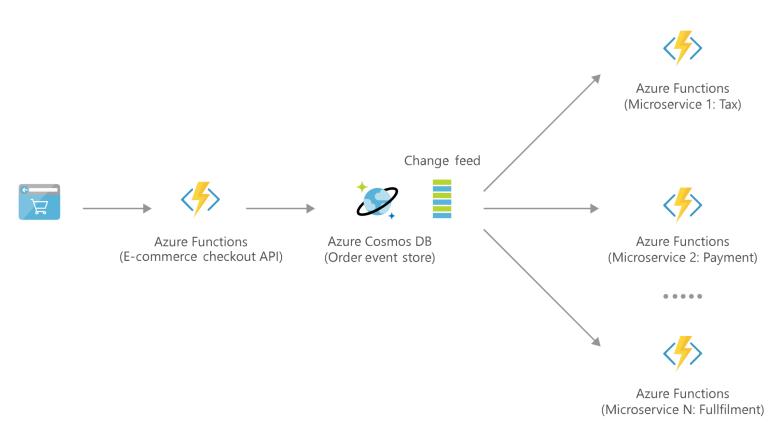
- Adjust indexes on the database
- Calculate the cost of the query
- Retrieve request unit cost of a point operation or query
- Implement Azure Cosmos DB <u>integrated</u> cache



Design and Implement Change Feeds for an Azure Cosmos DB for NoSQL

- Develop an <u>Azure Functions trigger to process a change feed</u>
- Consume a change feed from within an application by using the SDK
- Manage the number of change feed instances by using the <u>change feed estimator</u>
- Implement <u>denormalization</u> by using a change feed
- Implement <u>referential enforcement</u> by using a <u>change feed</u>
- Implement aggregation persistence by using a change feed
- Implement <u>data archiving</u> by using a change feed







Define and Implement an Indexing Strategy for an Azure Cosmos DB for NoSQL

- Choose when to use a <u>read-heavy versus write-heavy index strategy</u>
- Choose an <u>appropriate index type</u>
- Configure a <u>custom indexing policy</u> by using the Azure portal
- Implement a <u>composite index</u>
- Optimize index performance



Maintain an Azure Cosmos DB Solution

Monitor and Troubleshoot an Azure Cosmos DB Solution

- Evaluate response status code and <u>failure</u> metrics
- Monitor metrics for normalized throughput usage by using <u>Azure Monitor</u>
- Monitor server-side latency metrics by using Azure Monitor
- Monitor data replication in relation to latency and availability
- Configure Azure Monitor alerts for Azure Cosmos DB
- Implement and query Azure Cosmos DB logs (and see this)
- Monitor throughput <u>across</u> partitions
- Monitor distribution of data across partitions
- Monitor security by using logging and auditing



Implement Backup and Restore for an Azure Cosmos DB Solution

- Choose between <u>periodic</u> and <u>continuous</u> backup
- Configure periodic backup
- Configure continuous backup and recovery
- Locate a recovery point for a <u>point-in-time recovery</u>
- Recover a database or container from a recovery point



Implement Security for an Azure Cosmos DB Solution

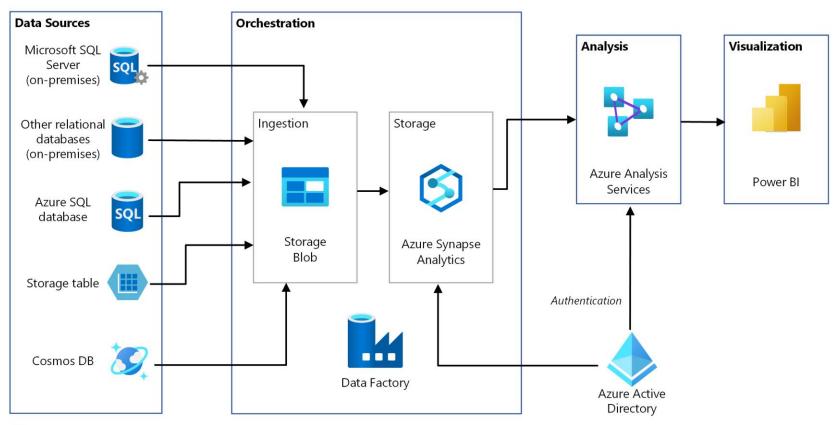
- Choose between <u>service-managed and customer-managed encryption keys</u>
- Configure <u>network-level access control for Azure Cosmos DB</u>
- Configure data encryption for Azure Cosmos DB
- Manage control plane access to Azure Cosmos DB by using <u>Azure role-based access</u> control (RBAC)
- Manage data plane access to <u>Azure Cosmos DB by using keys</u>
- Manage data plane <u>access to Azure Cosmos DB by using Azure Active Directory</u>
- Configure Cross-Origin Resource Sharing (CORS) settings
- Manage account keys by using Azure Key Vault
- Implement customer-managed keys for <u>encryption</u>
- Implement <u>Always Encrypted</u>



Implement Data Movement for an Azure Cosmos DB Solution

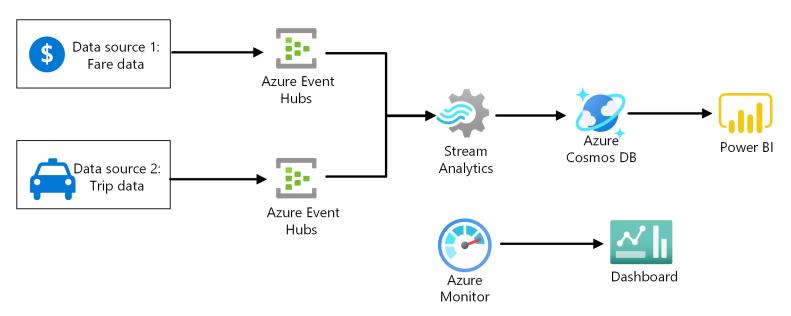
- Choose a data movement strategy
- Move data by using client SDK <u>bulk</u> operations
- Move data by using Azure Data <u>Factory</u> and Azure Synapse pipelines
- Move data by using a <u>Kafka</u> connector
- Move data by using <u>Azure Stream Analytics</u>
- Move data by using the Azure Cosmos DB <u>Spark</u> Connector





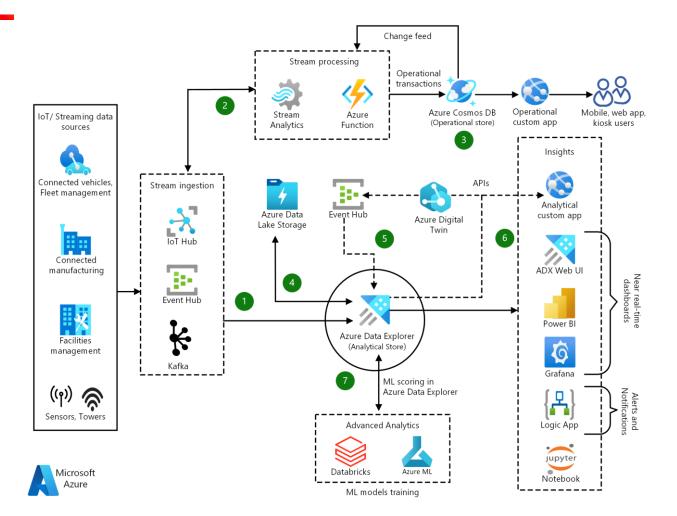




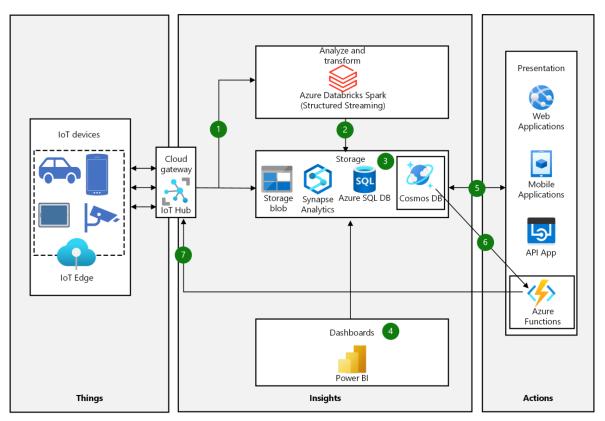
















Implement a DevOps Process for an Azure Cosmos DB Solution

- Choose when to use <u>declarative</u> <u>versus</u> imperative operations
- Provision and manage Azure Cosmos DB resources by using <u>ARM templates</u>
- Migrate between standard and <u>auto scale throughput</u> by using PowerShell or Azure CLI
- Initiate a regional failover by using PowerShell or Azure CLI
- Maintain index policies in production by using ARM templates



The Exam

DP-420

- Exam DP-420
- Skills measured

$\ \ \ \mathsf{Tip}$

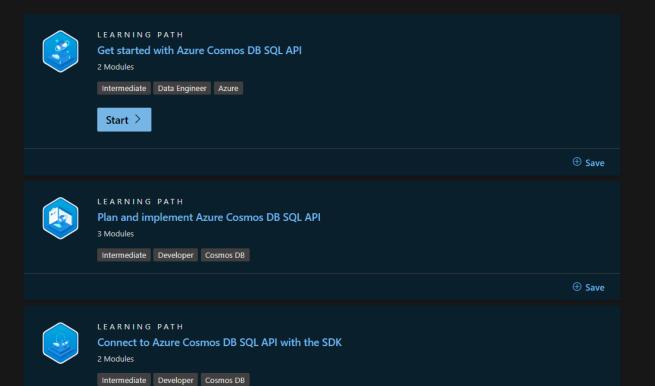
- Download the **DP-420 study guide** ☑ to help you prepare for the exam
- Demo the exam experience by visiting our Exam Sandbox ☑



Two ways to prepare

Online - Free Instructor-led - Paid

Items in this collection





Questions in DP-420

- Number of Questions 40-60 Questions
- Questions
 - Multiple choice
 - Drag and drop
 - Scenario based
- Pass Score 700 (on a scale of 1-1000)
- See <u>exam sandbox</u>



Schedule exam

Exam DP-420: Designing and Implementing Cloud-Native Applications Using Microsoft Azure Cosmos DB

Languages: English, Japanese, Chinese (Simplified), Korean, German, French, Spanish, Portuguese (Brazil), Arabic (Saudi Arabia), Russian, Chinese (Traditional), Italian, Indonesian (Indonesia)

Retirement date: none

This exam measures your ability to accomplish the following technical tasks: design and implement data models; design and implement data distribution; integrate an Azure Cosmos DB solution; optimize an Azure Cosmos DB solution; and maintain an Azure Cosmos DB solution.

Schedule exam >

Official practice test for Designing and Implementing Cloud-Native Applications Using Microsoft Azure Cosmos DB All objectives of the exam are covered in depth so you'll be ready for any question on the exam.

United States

\$165 USD*

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



⊕ Save

Select exam options

AZ-104: Microsoft Azure Administrator





Where do you want to take your exam?



At a test center



I have a Private Access Code

Prepare for your online exam at your home or office



Your computer

Use a personal computer that has a reliable webcam and internet connection.

Run system test.



Your testing space

The room should be a distractionfree, private place.

See <u>acceptable spaces</u> and view permitted <u>comfort aid list</u>.



Your photo ID

We'll verify your governmentissued identification (ID) when you arrive for your exam.

Review admission & ID policies



What to expect

Check in for your OnVUE exam 30 minutes before your appointment time.

Watch our <u>short video</u> to get familiar with the process.

Questions?

Check out the OnVUE FAQs and minimum technical requirements.



Cart

Review and confirm contact information to avoid issues on test day.

Description	Details	Price Actions
		165.00 Remove

Available Products

In addition to scheduling your exam, you might be interested in the following products.



Microsoft Official Practice Test powered by MeasureUp - 30 day online access Get a discount on available Microsoft Official Practice Test for Microsoft certification exams (Fundamentals, Role-based, or Specialty) 30-day online access. USD 80.00 Add to Order

Special offer: Regularly priced at USD 99.00! Click here for details

More Details





It's time to test your system

Order #: 0064-8802-7606

Your appointment is confirmed! An order confirmation containing important exam day information has been sent to: zaalion@gmail.com

What's next?

Run a system test

We need to verify that the computer and internet connection you plan to use on exam day meet the <u>minimum requirements</u> for online testing. It'll just take 5 minutes to run:

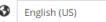
Equipment and internet connection checks

Exam simulation

Description	Details	Order Information	Price
			165.00







System Test

☐ I confirm that on my exam day I will be using this same testing space, computer, and internet connection.

Alert! Work computers generally have more restrictions that may prevent a successful test. Ensure you are not behind a corporate firewall, and shut down any Virtual Private Networks (VPNs) or Virtual Machines.

1. Copy Access Code

Click 'Copy Access Code'.

This code will authorize you to perform a system test.

690-635-235

Copy Access Code

2. Download OnVUE

Click 'Download'.

Download

3. Run OnVUE

Run the OnVUE application from your Downloads folder.



Course Repository

https://github.com/zaalion/oreilly-dp-420



O'REILLY® Thank you!

Reza Salehi



