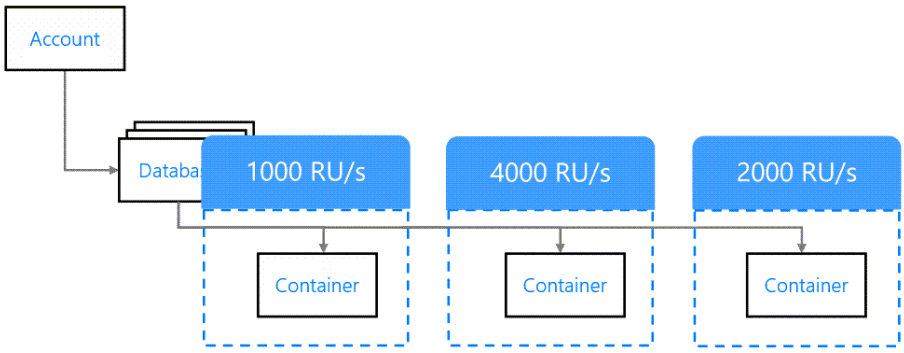
Module 02: Plan and implement Azure Cosmos DB SQL API

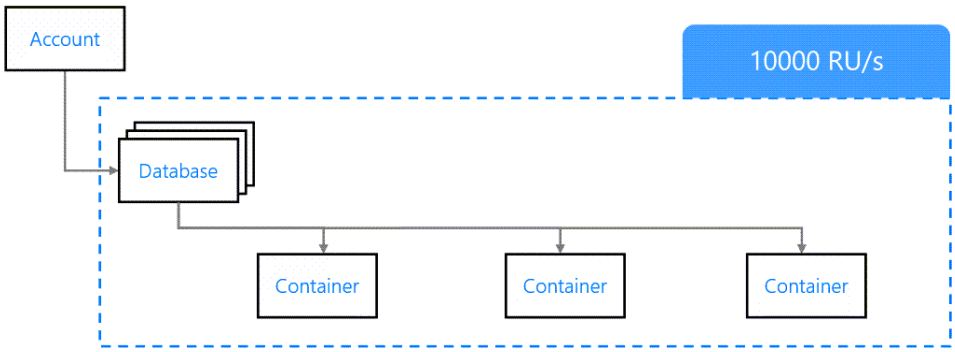
Plan Resource Requirements

Understand throughput

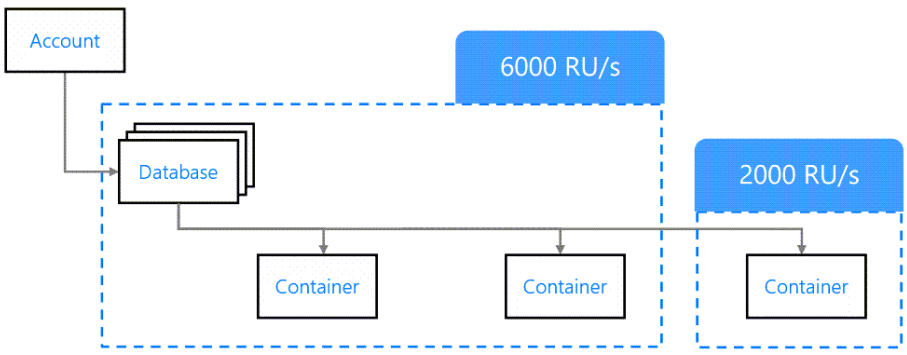
Container-level throughput provisioning



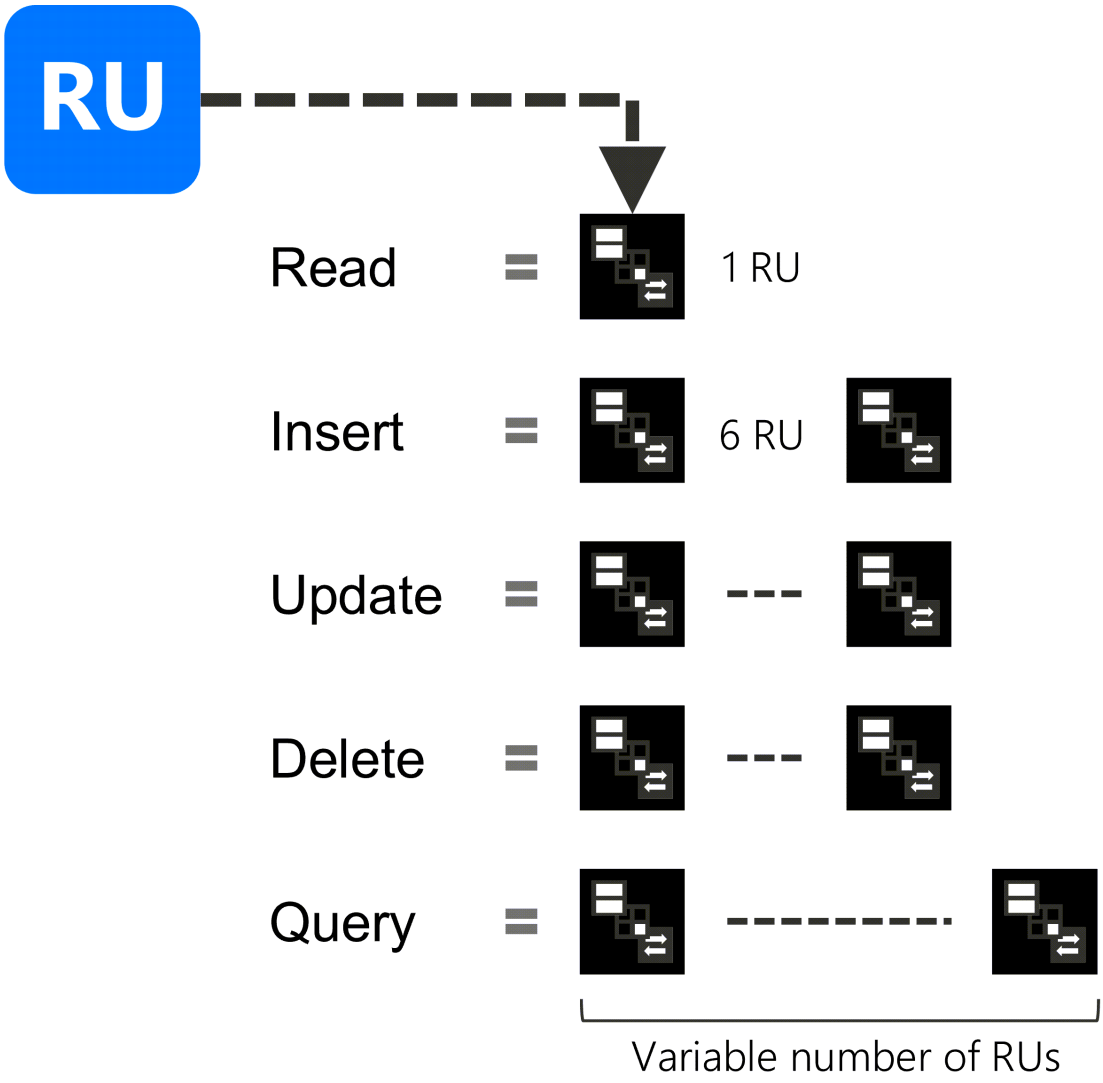
Database-level throughput provisioning



Mixed-throughput provisioning

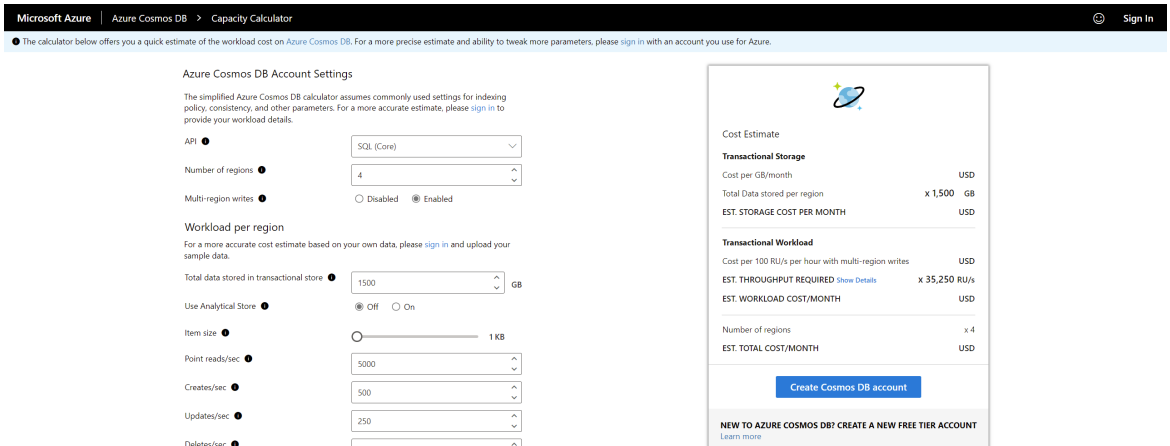


Evaluate throughput requirements



Operation type	Number of requests per second	Number of RU per request	RU/s needed
Write Single Document	10,000	10	100,000
Top Query #1	700	100	70,000
Top Query #2	200	100	20,000
Top Query #3	100	100	10,000
Total RU/s			200,000 RU/s

Evaluate data storage requirements



Time-to-live (TTL)

The TTL value for a container is configured using the `DefaultTimeToLive` property of the container's JSON object.

DefaultTimeToLive	Expiration
Does not exist	Items are not automatically expired
`-1`	Items will not expire by default
n	*n* seconds after last modified time

Example:

Container.DefaultTimeToLive	Item.ttl	Expiration in seconds
`1000`	*null*	`1000`
`1000`	`-1`	*This item will never expire*
`1000`	`2000`	`2000`

Another example:

Container.DefaultTimeToLive	Item.ttl	Expiration in seconds
null	*null*	*This item will never expire*
null	`-1`	*This item will never expire*
null	`2000`	*This item will never expire*

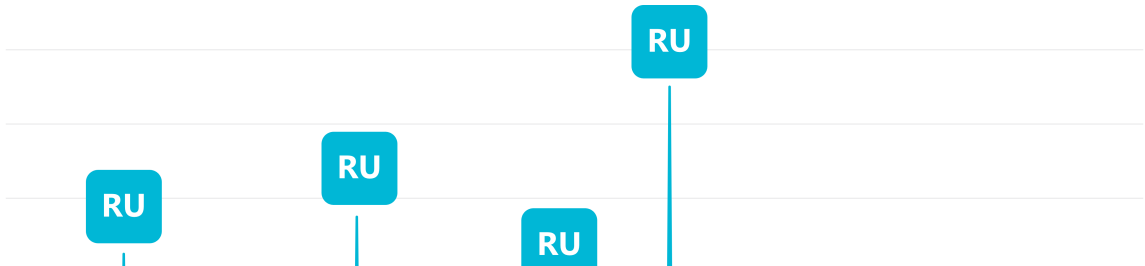
Plan for data retention with time-to-live (TTL)

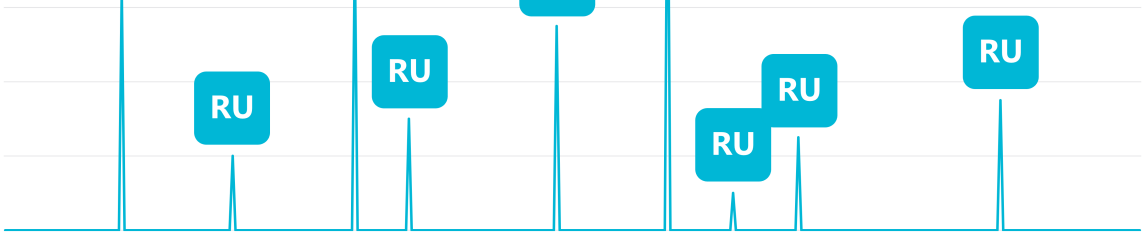
Consider solutions such to aggregate and migrate data such as:

- Change feed
- Azure Data Warehouse
- Azure Blob Storage

Configure Azure Cosmos DB SQL API throughput

Serverless

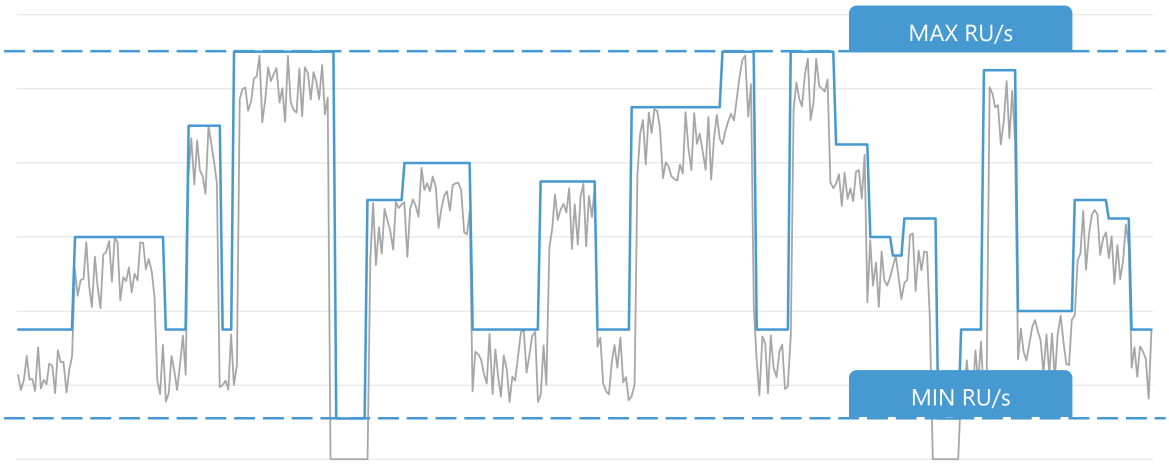




Compare serverless vs. provisioned throughput

Throughput	Workloads	RUs	Global Distribution	Compare storage limits
Provisioned	Ideal for workloads with predictable traffic patterns	Number RUs per second preset to each container	Can distribute data to an unlimited number of Azure regions	Unlimited data in a container
Serverless	Can handle workloads that have wildly varying traffic	Doesn't require any planning or automatic provisioning	Accounts can only run in a single Azure region	Up to 50 GB of data in a container

Autoscale



Compare autoscale vs. standard (manual) throughput

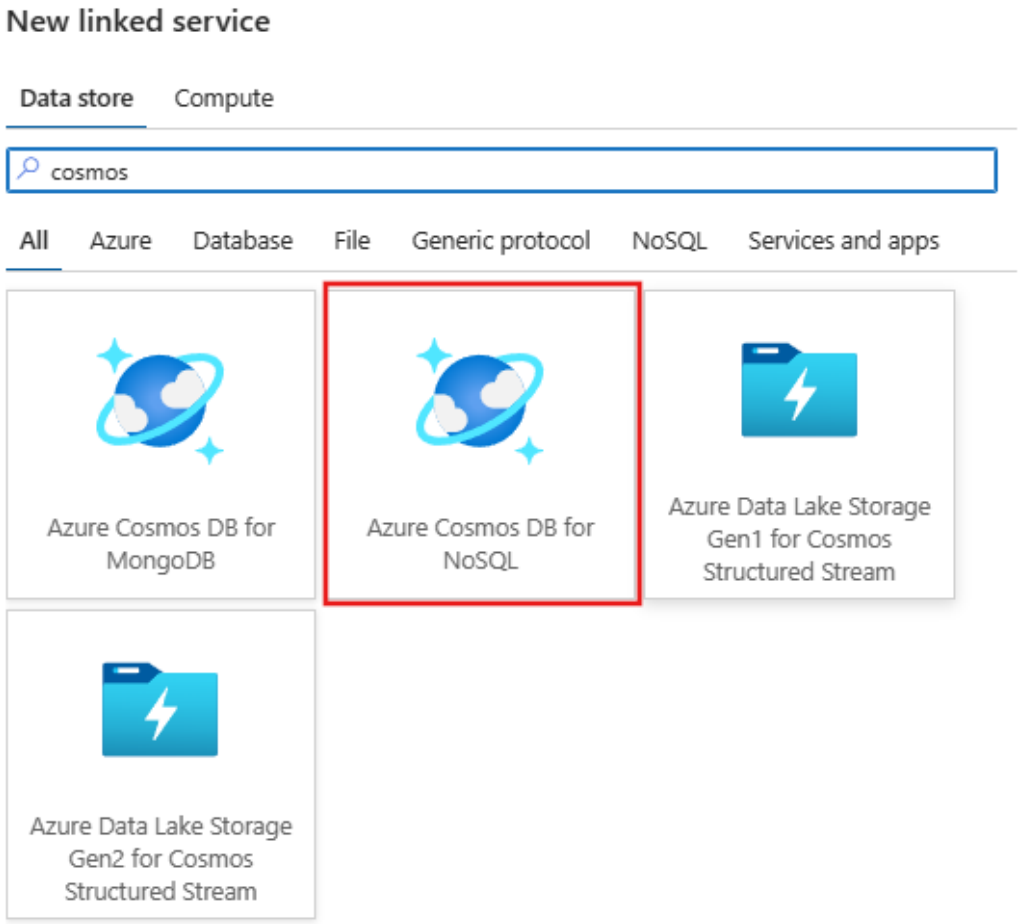
Scaling	Workloads	RUs	Scenarios	Rate-limiting
Standard	Suited for workloads with steady traffic	Requires a static number of request units to be assigned ahead of time	Where the application throughput can be accurately predicted	Since the RU/s are static, requests beyond this will be rate-limited
Autoscale	Suited for unpredictable traffic	You only set the maximum RUs	Where the application throughput can't be accurately predicted, but an acceptable max throughput can be assigned	Will scale up to the max RU/s before similarly rate-limiting responses

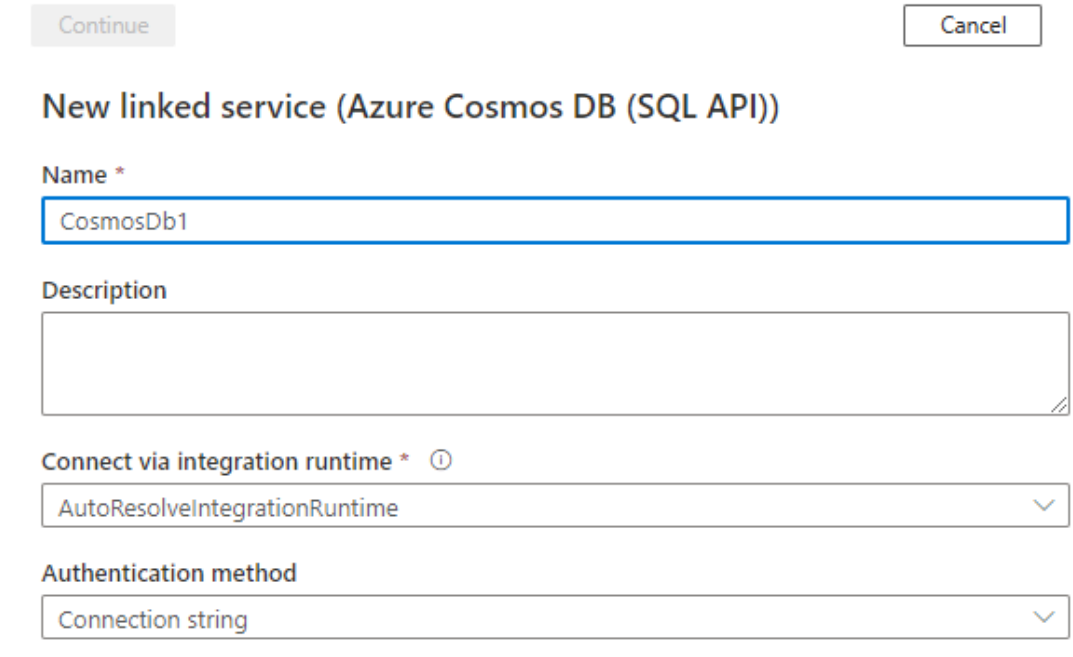
Moving data into and out of Azure Cosmos DB SQL API

Move data by using Azure Data Factory

Linked service

```
{
    "name": "<example-name-of-linked-service>",
    "properties": {
        "type": "CosmosDb",
        "typeProperties": {
            "connectionString": "AccountEndpoint=<cosmos-
endpoint>;AccountKey=<cosmos-key>;Database=<cosmos-database>"
        }
    }
}
```





Connection string | Azure Key Vault

Account selection method

From Azure subscription | Enter manually

Azure subscription

Select all

Azure Cosmos DB account name *

Database name *

No database

Additional connection properties

New

Annotations

New

Parameters

Advanced