



Optimizing performance and ROI with Azure Cosmos DB

Best practices for new customers and workloads

2019

Welcome to Azure Cosmos DB

To help you optimize your database performance and ROI, these are some recommended best practices to follow as you get started. Before you begin, have a look at [your account settings](#) (e.g., Location, Geo-Redundancy, Multi-region Writes, etc.) to ensure they're set as you want.

Data Modeling

In Azure Cosmos DB, you can significantly lower latency for common queries by optimizing your data model. We recommend spending time thinking about your workload (e.g., how the data will be queried, read/write ratio, etc.) before designing your data model, with the aim of reducing RU consumption in your most frequent queries.

To learn about data modeling in Azure Cosmos DB:

1. Read our [docs on data modeling](#)
2. Watch [a session on data modeling](#) from a recent Microsoft conference

Partitioning

For horizontal scalability, Azure Cosmos DB stores your data in smaller [partitions](#). Developers specify a partition key for each Azure Cosmos DB container at the time of creation. This critical design decision defines how your data will be stored in Azure Cosmos DB, as modifications are unavailable after initial creation, and non-optimized partition keys can lead to rate errors/throttling, misallocation of RU/s, and unnecessary costs.

To effectively manage partition keys:

1. Follow our best practices for [choosing a partition key](#) and [avoiding "hot" partitions](#)
2. Read this [real world example](#) of how to model and partition data

If you need to change your partition key later, create a new container with your desired key and [transfer data using Azure Data Factory](#). You can also use [change feed](#) to do a live migration into another partitioned collection.

Global Distribution

Azure Cosmos DB can replicate and distribute your data to any or all Azure regions, and supports [multi-region writes](#) ("multi-master"), providing low-latency by bringing data physically closer to your users. For enhanced SLAs including high availability of 99.999% and <10ms read and write latencies, and to enable multi-region writes, we strongly recommend that all production workloads replicate data to at least one additional region (within your country or elsewhere).

To manage multi-region settings, performance, and costs:

1. Watch our [global distribution overview video](#)

© 2019 Microsoft Corporation. All rights reserved.

AzureCosmosDB.com

2. Check your selected regions and monitor activity; you can [add or remove regions](#) at any time
3. Check [your account settings](#) to see if multi-master (“Multi-region Writes” in your settings) is on or off; It is possible to have a multi-region Azure Cosmos DB account without having multi-region writes enabled.
4. Understand [how to optimize costs for multi-region workloads](#)

Performance & Cost Optimization

Your account will be billed each month for the [throughput you provision \(RU/s\) and the storage you consume](#). Provisioned throughput can elastically scale up and down, giving you control over resources and costs: you are billed for the peak throughput that you provision each hour, so lowering provisioned throughput during low-demand periods (e.g., overnight) can provide savings.

To manage costs:

1. Watch a video on [Request Units](#), the Azure Cosmos DB measurement for throughput
2. Understand how to [optimize your development and testing costs](#) in Azure Cosmos DB
3. Learn [how to provision throughput at the database level](#) and have it shared among the database’s set of containers
4. Distribute request and storage volume to [avoid “hot” partitions](#)
5. Consider a [reserved capacity 1- or 3-year term](#) for savings of up to 65% on provisioned throughput

Thank you

We hope this helps you get started with Azure Cosmos DB. Should you have questions, please contact your sales or account manager. For technical questions, you can contact the Azure Cosmos DB engineering team at <mailto:askcosmosdb@microsoft.com>.