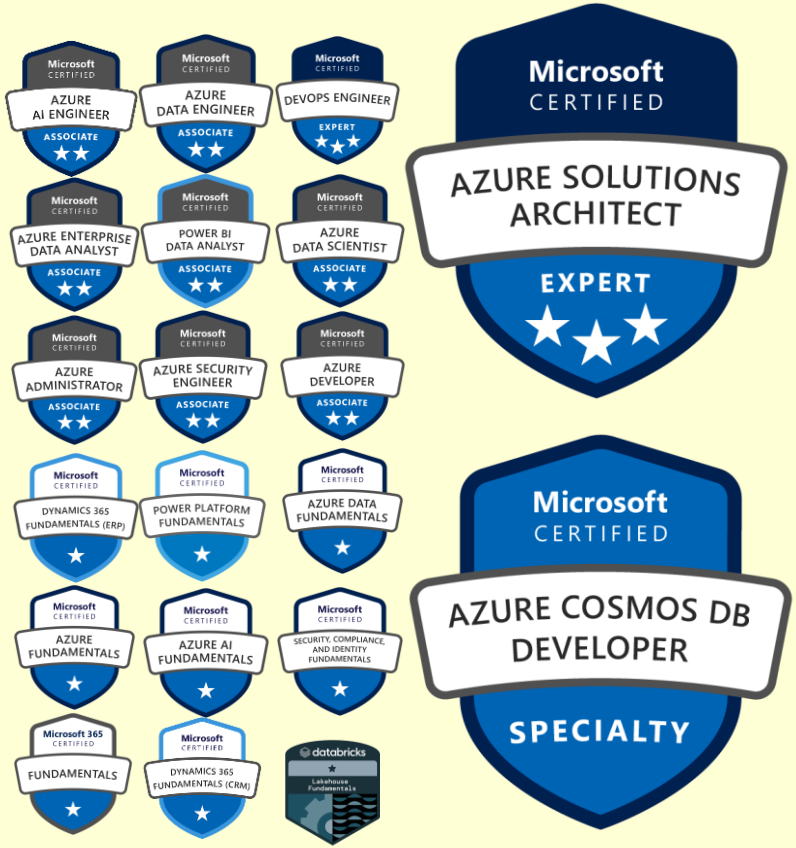
COSMOS 101

How you became a COSMONAUT in 60 minutes or less

WELCOME



Scott Bell

- Senior Consultant at Avanade and Previously Altius.
- Avanade UK&I Databricks SME
- Interested in Azure Data Platforms, Architecture and Design Pattens
- Masters Degree in Computer Science Focusing on Machine Learning in the Cloud
- Run @DailyDatabricks Twitter for Tips, Tricks and Hacks

QUEST INTO THE COMSOS

01 **WHAT IS COSMOS DB**

A brief intro

02 **KEY CONCEPTS**

03 **WHEN TO USE IT?**

What can it do for you?

Demo Time!

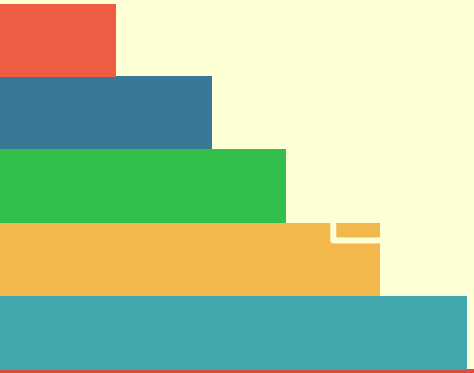
04 **HOW DO WE BUILD?**

01

WHAT IS COSMOS

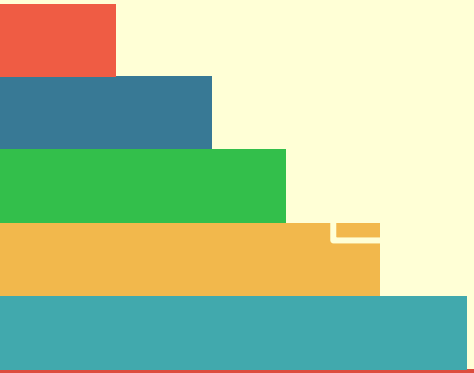
WHAT IS IT?

Azure Cosmos DB is a fully managed NoSQL database service that has **“Planet Scale”**



NO SQL

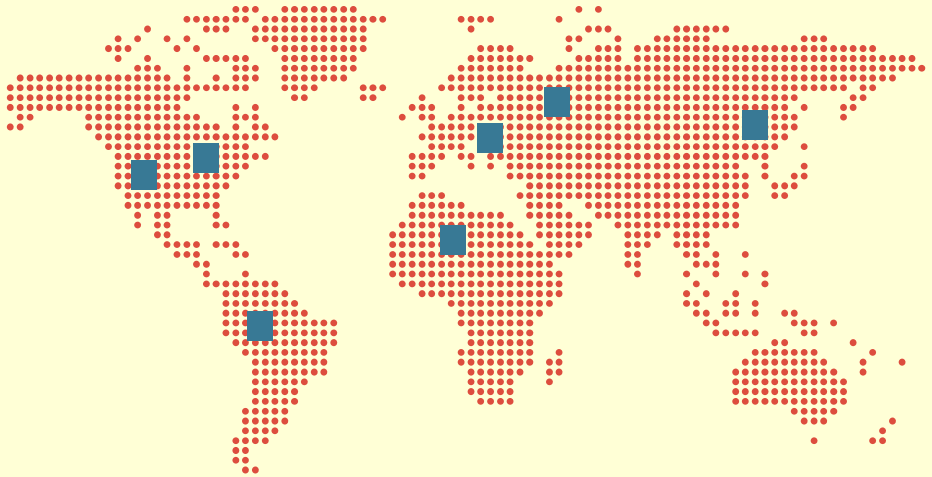
NoSQL databases are “**non-relational**,” **schemaless**, datastores that handle unstructured data



MANAGED

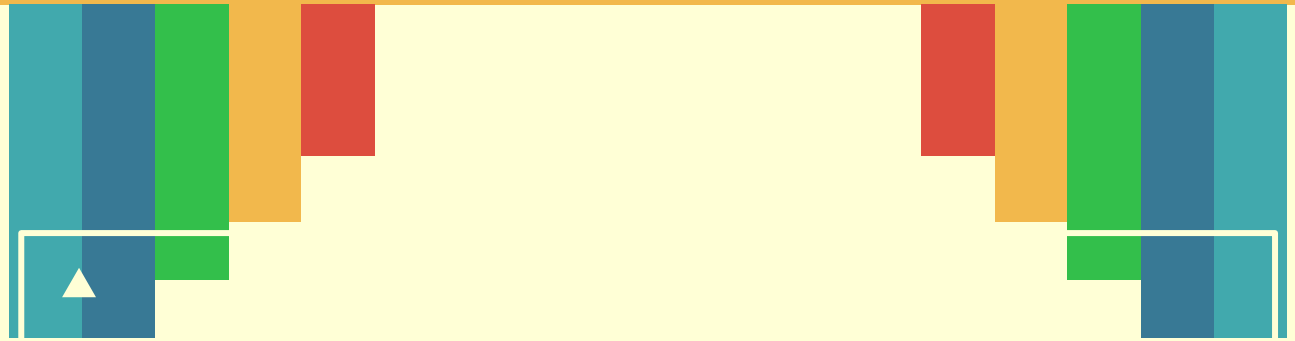
SLA backed PaaS offering with the ability to instantly scale, be highly available, distribute data workloads and have enterprise security.

PLANET SCALE



REPLICATE
Your data geolocated to your applications and users

Distrubute
Read and Write between different regions to increase scalablity



99.999%

SLAs or your money back*

Multi Modal

SQL API

Stores JSON Documents in collections and gives you a SQL interface

Document Store that uses BSON

MONGO DB

Cassandra

OSS Distrubuted Wide Column Store

Graph database that stores data as edges and vertices

GREMLIN API

Table API

Key Value Table driven from Azure Table Storage

RECAP

DISTRUBUTED

DATA IS REPLICATED BETWEEN DATA CENTRES WITH LOW LATENCY

HIGHLY SCALABLE

PLANET SCALE PERFORMANCE ACROSS THE GLOBLE

MULTI MODAL

Lots of different APIs to support different USECASES (i.e GRAPH, JSON, KEY VALUE PAIRS)

NO SQL

NO FIXED SCHEMAS and IS NONE RELATIONAL

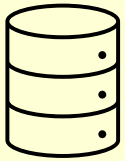
HIGH THROUGHPUT

HANDLES LARGE VOLUMES & VELOCITIES OF DATA ACROSS THE WORLD

02

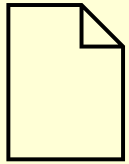
KEY CONCEPTS

COMPONENTS



DATABASE

1



ITEMS

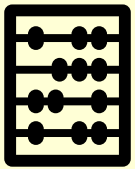
3

2

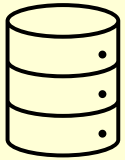
CONTAINERS

4

Request Units

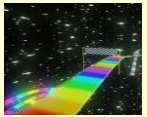


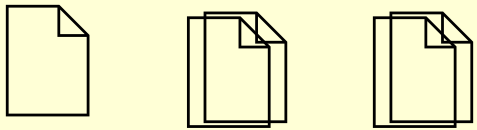
COMPONENTS



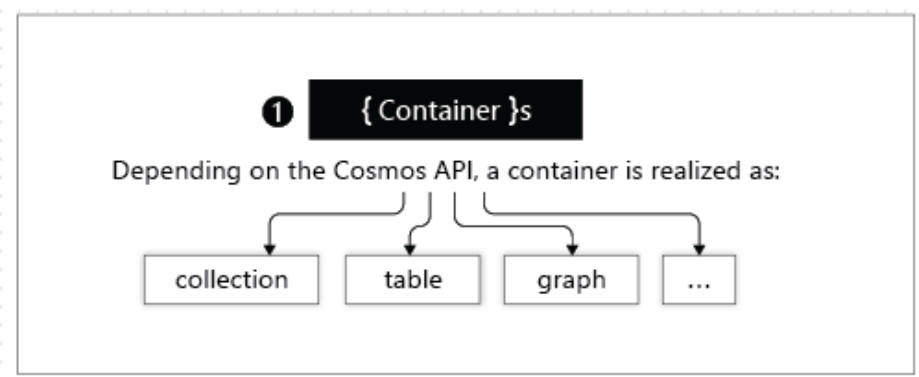
DATABASE

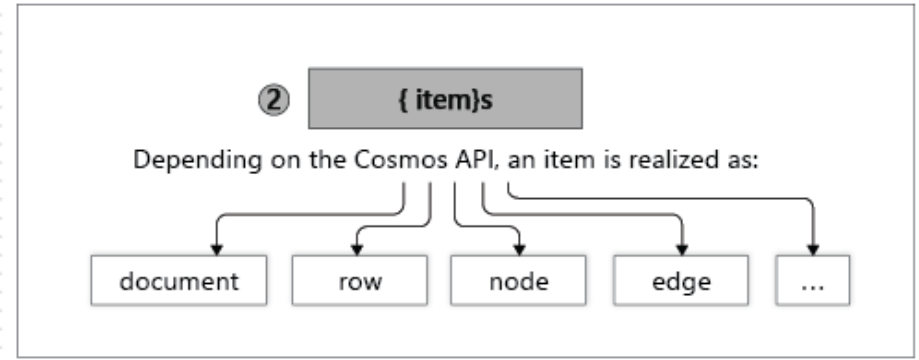
CONTAINERS



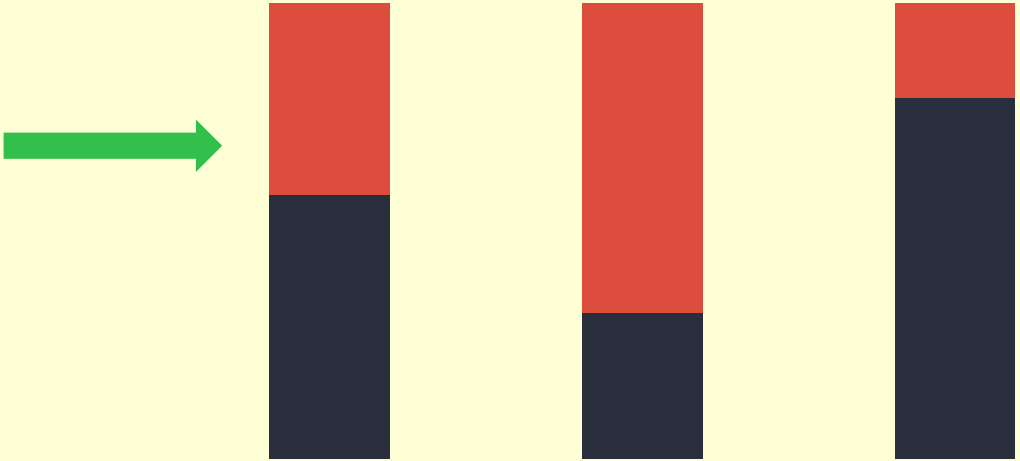


ITEMS


❶ { Container }s
Depending on the Cosmos API, a container is realized as:
collection
table
graph
...

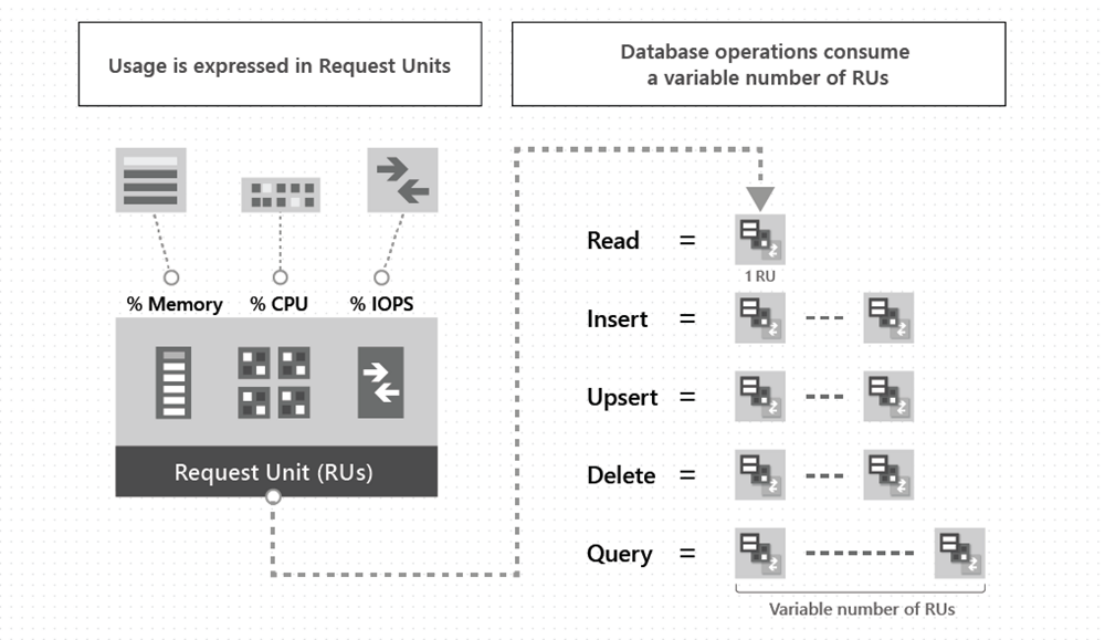

② { item}s
Depending on the Cosmos API, an item is realized as:
document
row
node
edge
...

PARTITIONING



Partition 1 Partiion 2 Partition 3

Cost



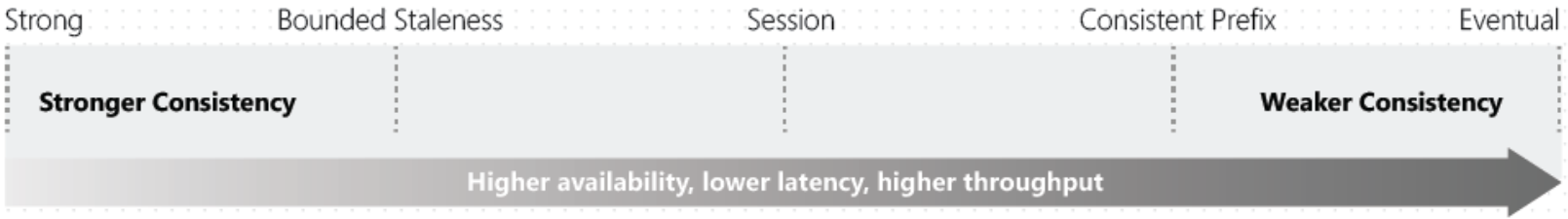
Complexity

They represent a cost per operation or provision throughput

REQUEST UNITS

A combination of the Memory, CPU and IO required to perform an operation

CONSISTENCY MODELS



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

Scaling

Serverless

Only charged for RU's you consume. Useful for irregular access patterns and low traffic applications

Provisioned

Reserved Capacity for at a database or container level. You can increase throughput at any point

Autopilot

Allows you scale your database or containers automatically based upon demand

Change FEED

- persistent record of changes to a container in the order they occur!

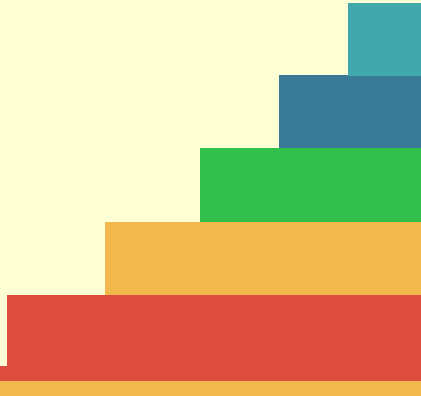
- The persisted changes can be processed asynchronously and incrementally, and the output can be distributed across one or more consumers for parallel processing.

- Can operate in a PUSH or PULL Method with support for most APIs

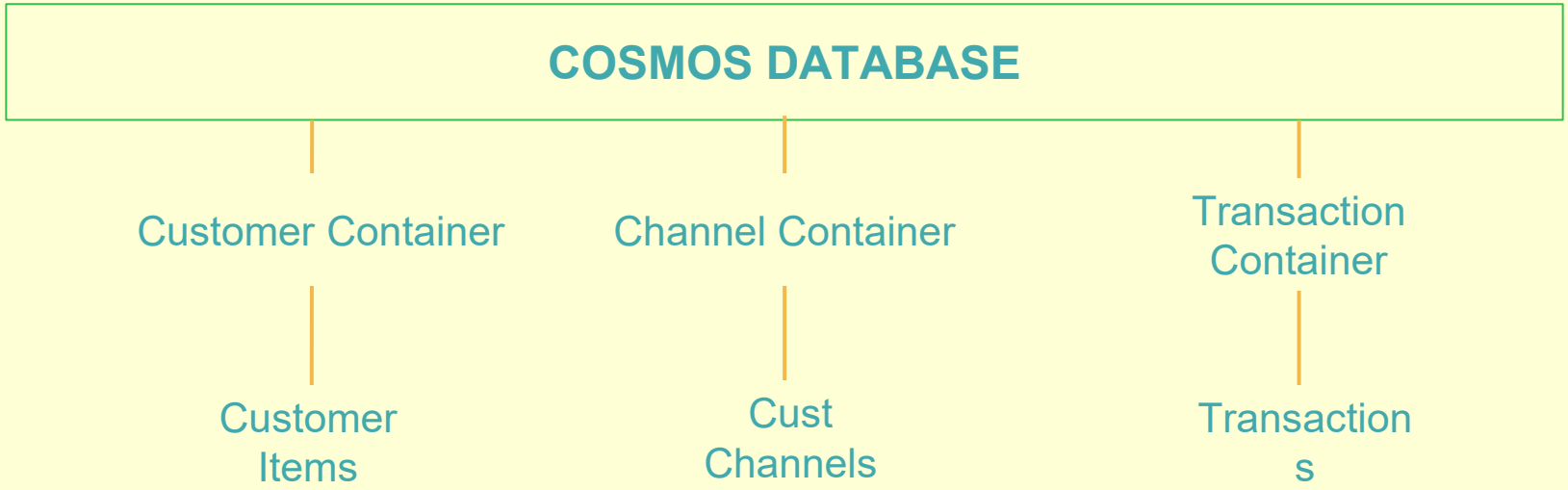
- Change feed doesn't log deletes

03

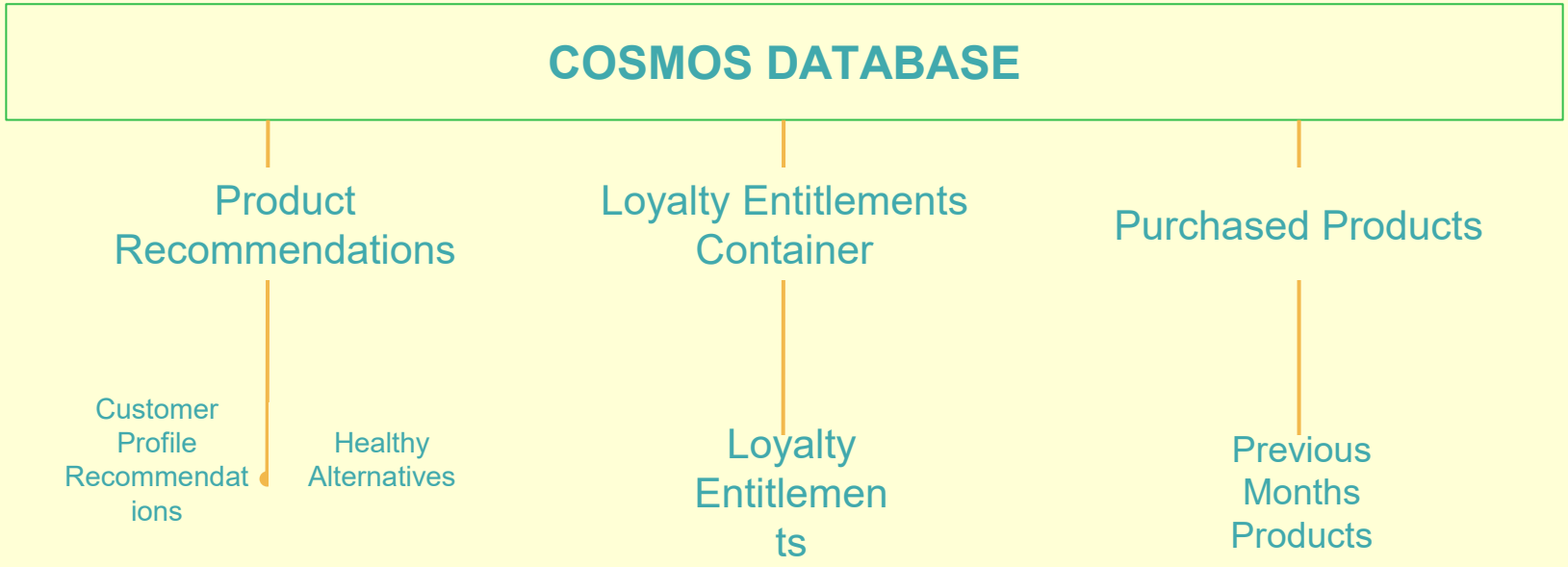
WHEN TO USE IT?



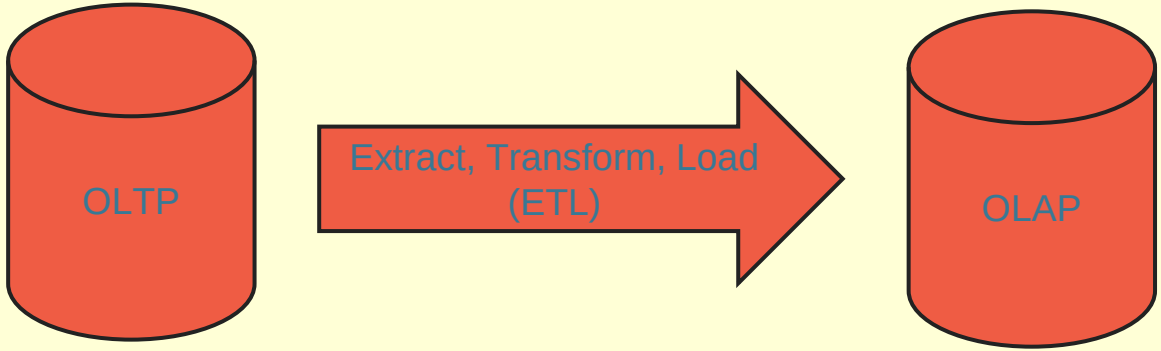
EXAMPLE


COSMOS DATABASE
Customer Container
Channel Container
Transaction Container
Customer Items
Cust Channels
Transaction s

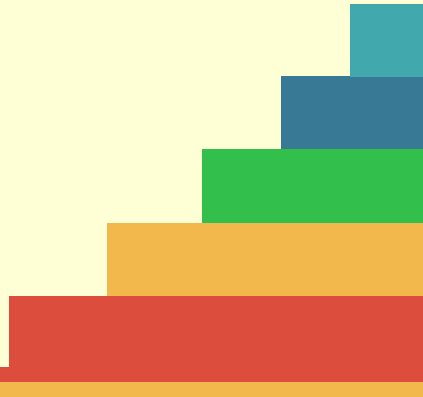
EXAMPLE 2


COSMOS DATABASE
Product Recommendations
Customer Profile Recommendations
Healthy Alternatives
Loyalty Entitlements Container
Loyalty Entitlements
Purchased Products
Previous Months Products

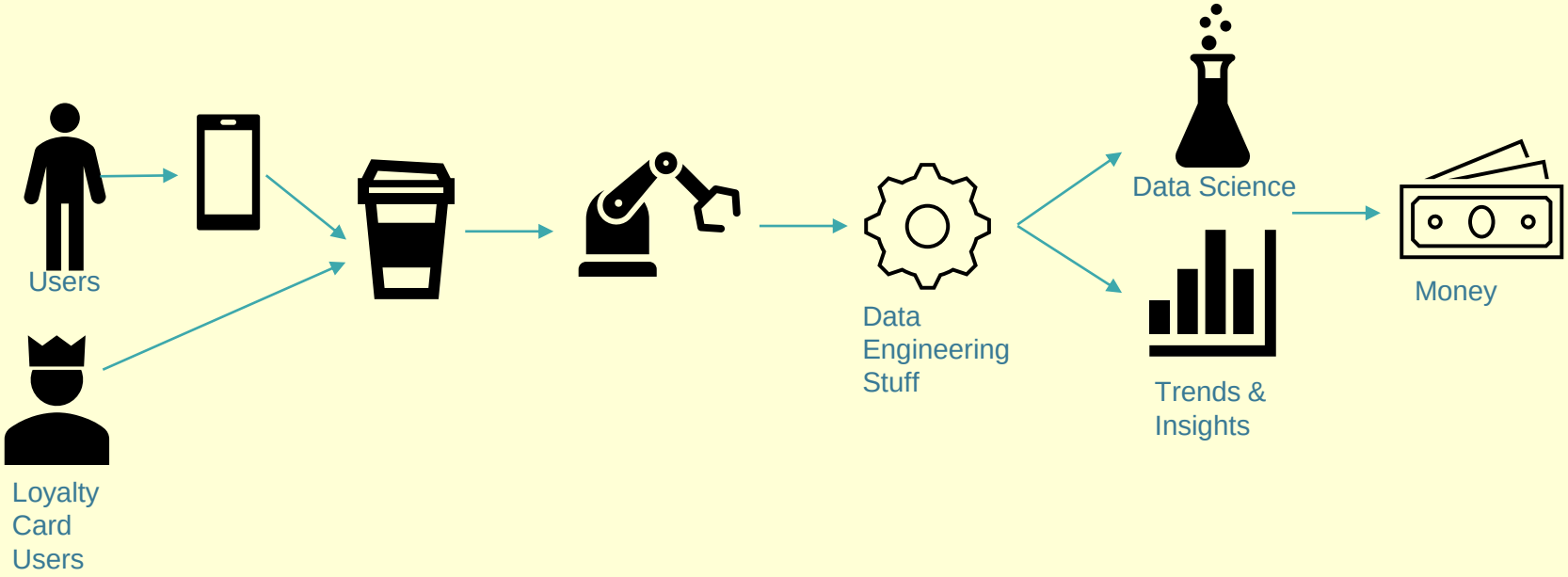
HTAPS



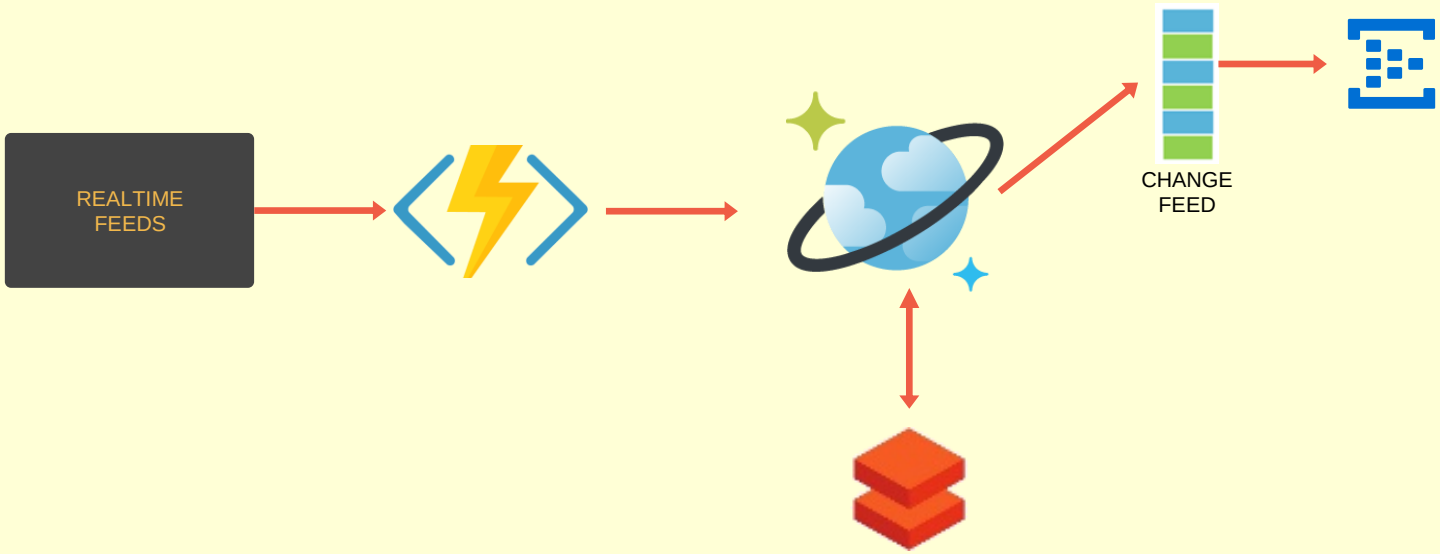
04
LETS BUILD



Problem


Users
Loyalty Card Users
Data Engineering Stuff
Data Science
Trends & Insights
Money

Architecture



Sample Document


```
{
    "id": "612375421376821390123",
    "partitionKey": "612375421376821390123",
    "typeOfDrink": "Espresso",
    "_rid": "9FE7ALxHlEMBAAAAAAAAAA==",
    "_self": "dbs/9FE7AA==/colls/9FE7ALxHlEM=/docs/9FE7ALxHlEMBAAAAAAAAAA=",
    "_etag": "\"0000c401-0000-0100-0000-622b663b0000\"",
    "_attachments": "attachments/",
    "_ts": 1647011387
}
```

DEMO TIME

THINGS WE DIDN'T COVER

TIME TO LIVE - https://docs.microsoft.com/en-us/azure/cosmos-db/sql/how-to-time-to-live#:~:text=%20Use%20the%20following%20steps%20to%20enable%20time,want%20to%20set%20time%20to%20live%2C...%20More%20feed#:~:text=%20Features%20of%20change%20feed%20%201%20Change,to%20an%20item%20appears%20exactly%20o...%20See%20More.

ETAGS- https://docs.microsoft.com/en-gb/azure/cosmos-db/account-databases-containers-items

Index Policies - https://learn.microsoft.com/en-us/azure/cosmos-db/index-policy

USEFUL RESOURCES

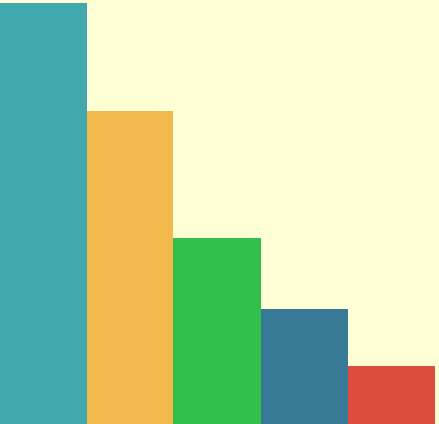
COSMOS EMULATOR - https://cosmos.azure.com/capacitycalculator/

COSMOS COST PROFILER - https://cosmos.azure.com/capacitycalculator/

Choose your Scale Model - https://docs.microsoft.com/en-gb/azure/cosmos-db/how-to-choose-offer

THANKS!

Do you have any questions?
talks@scottjamesbell.com
@fusionet24
www.myyearindata.com
@dailydatabricks



Azure Cosmos DB

Microsoft

Azure Cosmos DB

Add to Favorites

Microsoft

★ 3.6 (668 Azure ratings)

Plan

Azure Cosmos DB

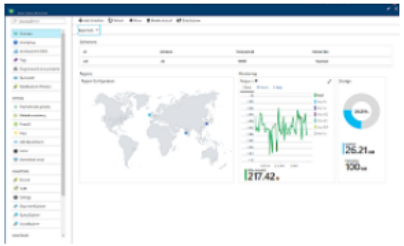
Create

Overview | Plans | Usage Information + Support | Reviews

Azure Cosmos DB is a fully managed, globally-distributed, horizontally scalable in storage and throughput, multi-model database service backed up by comprehensive SLAs. Azure Cosmos DB was built from the ground up with global distribution and horizontal scale at its core – it offers turn-key global distribution across any number of Azure regions by transparently scaling and replicating your data wherever your users are. You can elastically scale throughput and storage worldwide and pay only for the throughput and storage you need. Cosmos DB guarantees single-digit millisecond latencies at the 99th percentile anywhere in the world, offers multiple well-defined consistency models to fine-tune for performance and guaranteed high availability with multi-homing capabilities – all backed by industry leading service level agreements (SLAs).

Cosmos DB is truly schema-agnostic – it automatically indexes all the data without requiring you to deal with schema and index management. Cosmos DB is multi-model – it natively supports document, key-value, graph and columnar data models. With Cosmos DB, you can access your data using the APIs of your choice. Cosmos DB is a fully managed, enterprise ready and trustworthy service. All your data is fully and transparently encrypted and secure by default. Cosmos DB is ISO, FedRAMP, EU, HIPAA, and PCI compliant as well.

Media



Select API option ...

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL database service for building scalable, high performance applications. Learn more

To start, select the API to create a new account. The API selection cannot be changed after account creation.

Core (SQL) - Recommended

Azure Cosmos DB's core, or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.

Create Learn more

Azure Cosmos DB API for MongoDB

Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.

Create Learn more

Cassandra

Fully managed Cassandra database service for apps written for Apache Cassandra. Recommended if you have existing Cassandra workloads that you plan to migrate to Azure Cosmos DB.

Create Learn more

Azure Table

Fully managed database service for apps written for Azure Table storage. Recommended if you have existing Azure Table storage workloads that you plan to migrate to Azure Cosmos DB, but do not want to re-write your application to use the SQL API.

Create Learn more

Gremlin (Graph)

Fully managed graph database service using the Gremlin query language, based on Apache TinkerPop project. Recommended for new workloads that need to store relationships between data.

Create Learn more

Create Azure Cosmos DB Account - Core (SQL)

Basics | Global Distribution | Networking | Backup Policy | Encryption | Tags | Review + create

Azure Cosmos DB is a fully managed NoSQL database service for building scalable, high performance applications. Try it for free, for 30 days with unlimited renewals. Go to production starting at $24/month per database, multiple containers included. Learn more

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Visual Studio Enterprise – MPN

Resource Group * COSMOS101

Create new

Instance Details

Account Name * shdhsdhsdh

Location * (US) West US

Capacity mode ⓘ (•) Provisioned throughput ( ) Serverless

Learn more about capacity mode

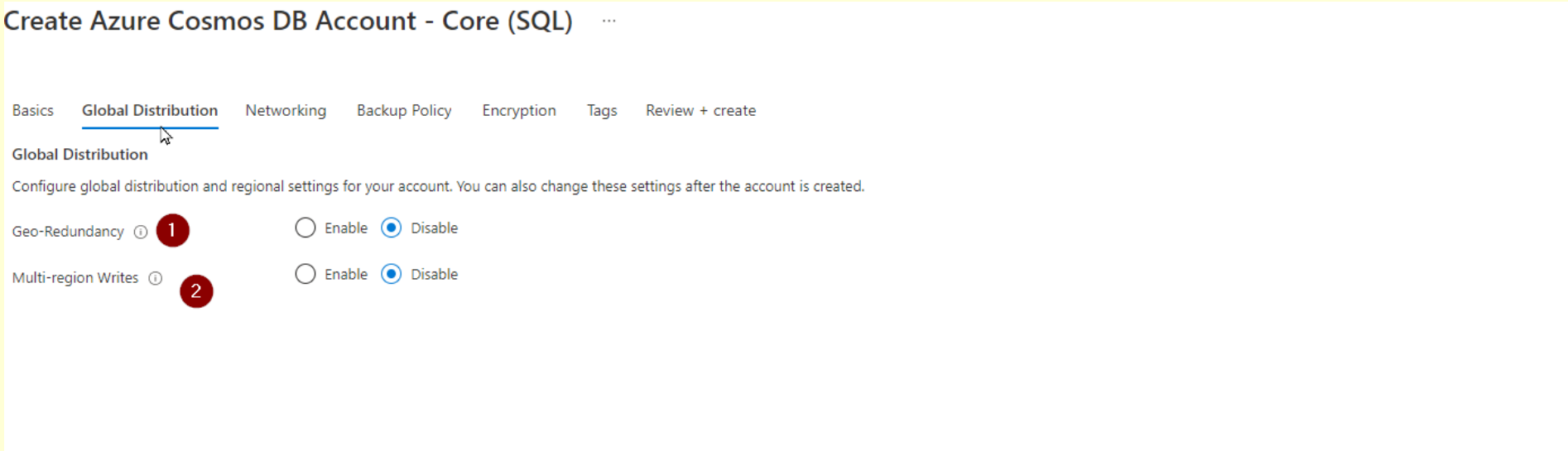
With Azure Cosmos DB free tier, you will get the first 1000 RU/s and 25 GB of storage for free in an account. You can enable free tier on up to one account per subscription. Estimated $64/month discount per account.

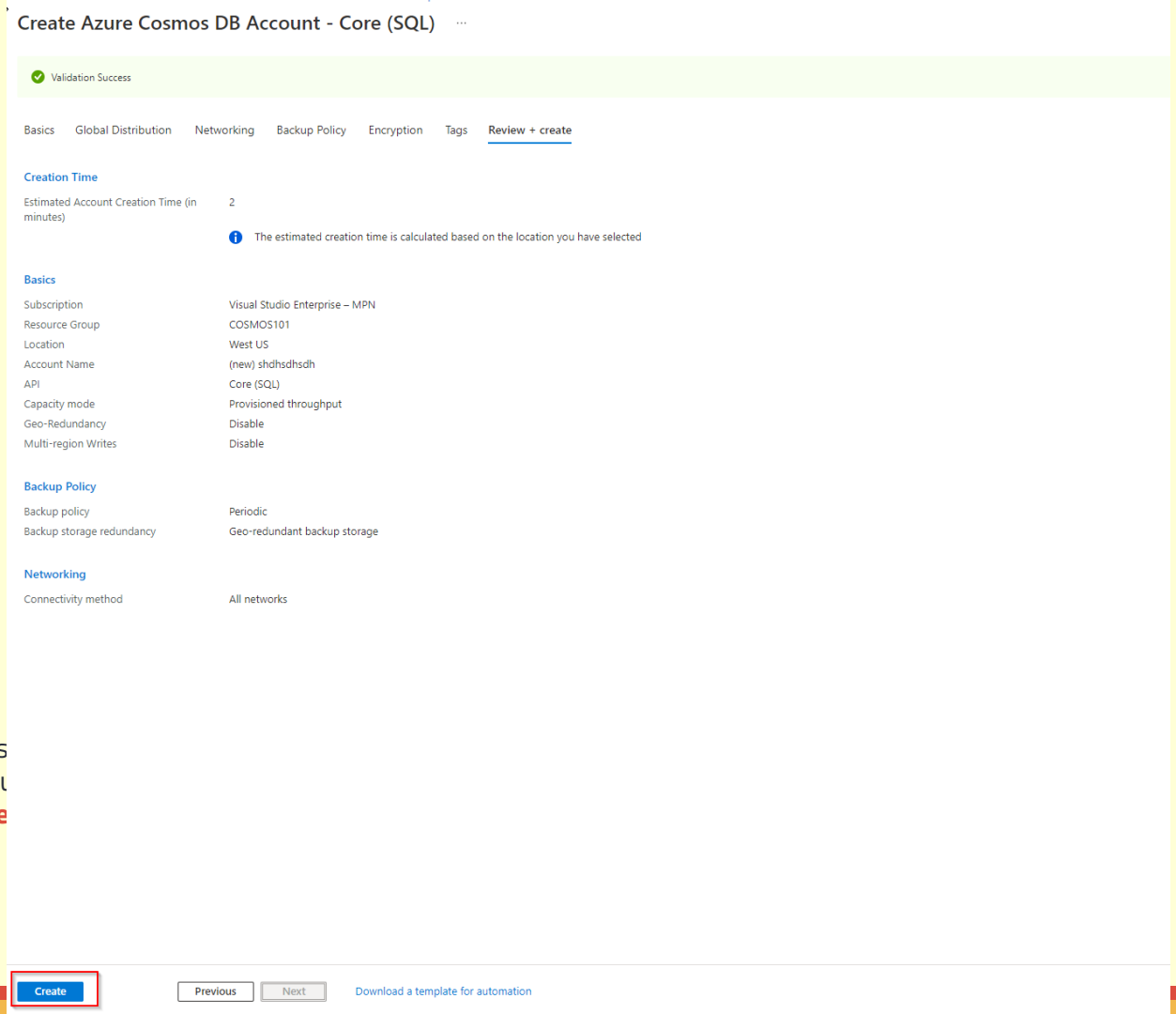
The subscription you have selected already has an account with free tier enabled.

Apply Free Tier Discount ( ) Apply (•) Do Not Apply

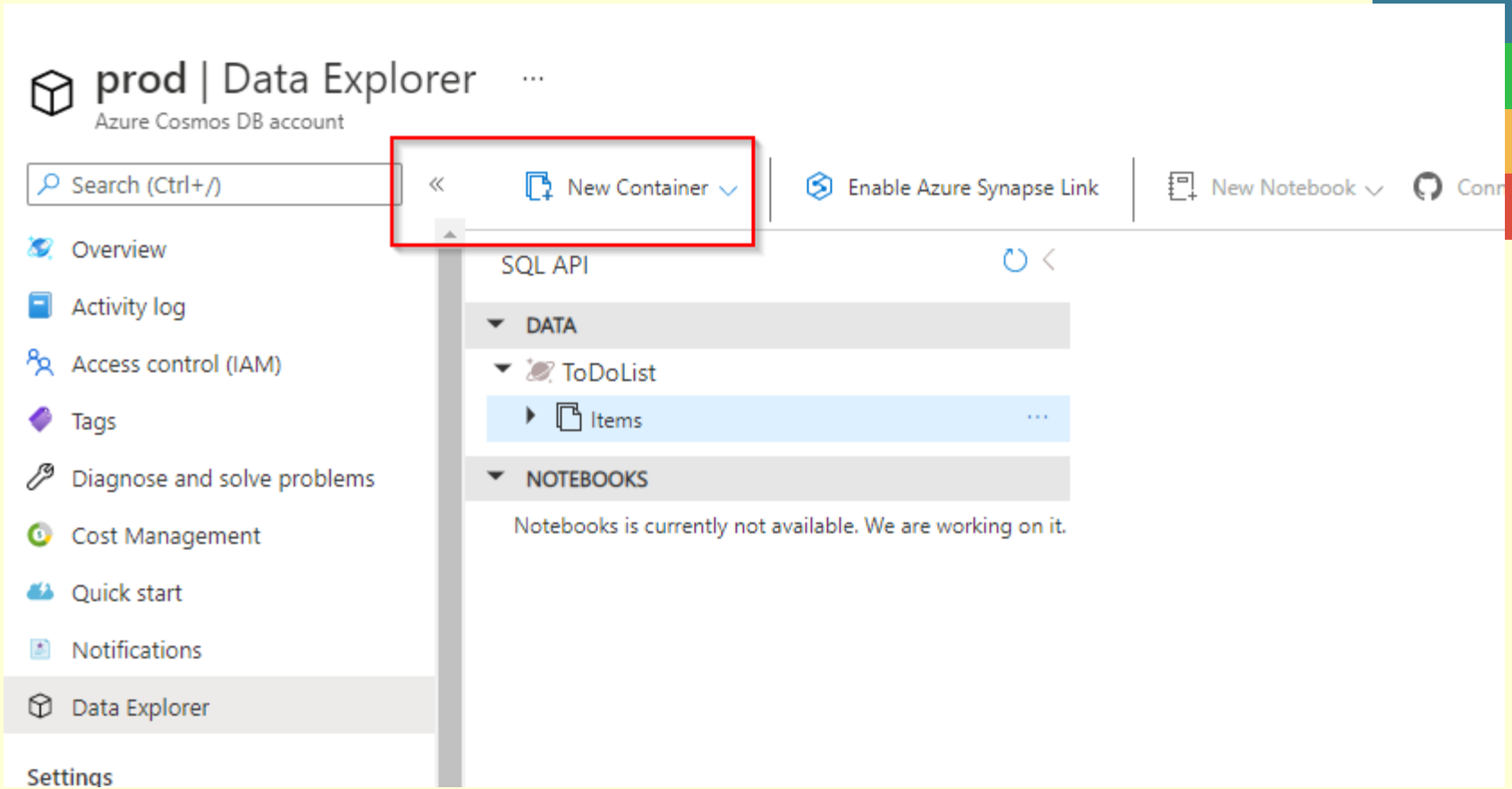
Limit total account throughput [ ] Limit the total amount of throughput that can be provisioned on this account

ⓘ This limit will prevent unexpected charges related to provisioned throughput. You can update or remove this limit after your account is created.


Create Azure Cosmos DB Account - Core (SQL)
Basics
Global Distribution
Networking
Backup Policy
Encryption
Tags
Review + create
Global Distribution
Configure global distribution and regional settings for your account. You can also change these settings after the account is created.
Geo-Redundancy
1
Enable
Disable
Multi-region Writes
2
Enable
Disable



CREDITS: This
Slidesgo, inclu
images by **Fre**


prod | Data Explorer
Azure Cosmos DB account
Search (Ctrl+/)
New Container
Enable Azure Synapse Link
New Notebook
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Cost Management
Quick start
Notifications
Data Explorer
Settings
SQL API
DATA
ToDoList
Items
NOTEBOOKS
Notebooks is currently not available. We are working on it.