

A cloud-native story de 0 à 1 million de procurations pour quelques euros avec Azure Cosmos DB

Christopher MANEU
Data Cloud Advocate @ Microsoft
@cmaneu

Agenda

Introduction – pourquoi faut-il d'autres options que SQL ?

Azure CosmosDB

Demo: Proxy vote. Supporter 1 million de demandes de procurations en une heure pour 2€

Comment apprendre & tester Azure Cosmos DB?

~~ Pause – 10 min ~~

Workshop

Pour aller plus loin : 1-day OpenHack Azure CosmosDB

Q&A

Pourquoi faut-il d'autres options que SQL ?

Les bases SQL peuvent faire beaucoup...

En termes de fonctionnalité

- JSON

- Geo-Data

- In-memory DBs

En termes de *scale*

- Azure PostgreSQL Hyperscale: > 100 Gb (40Tb by default)

- Azure SQL Business Critical: 16Tb storage

- Azure SQL Hyperscale: 100Tb / database

...mais elles ont leur limites

Performance à l'échelle

Multi-regions

Analytics

Développement d'applications (intégration avec les autres services Azure, schema-changes...)

Mais pourquoi ces limites ?

Atomicity

Correctness

Isolation

Durability

Mais pourquoi ces limites ?

Consistance

RDBMS

Disponibilité

NoSQL

Des bases de données “distribuées”

Besoins de scalabilité “illimités”

Faibles latences garanties (IoT, recommandations, ...)

Données distribuées géographiquement

Schéma “moins rigide”

Ces bases ne sont pas pour tous les usages

Attention si vous avez des :

- Transactions avec de nombreuses relations pointant vers la même entité
- Transactions nécessitant une forte consistance des données sur l'ensemble des datasets

L'effort de design est également plus conséquent que pour des bases relationnelles.

Choisir le bon service pour ses données

Azure SQL Database

Azure SQL Database for MySQL

Azure SQL Database for PostgreSQL

Azure SQL Database for MariaDB

Azure Cosmos DB

Azure Cache for Redis

Azure Table Storage

Azure Cosmos DB Gremlin

SQL Server

Azure Data Lake

Azure Data Explorer

Azure Databricks

Azure Synapse Analytics

Azure Search

Azure Cassandra Managed

Azure Time Series Insights

Azure Files

...

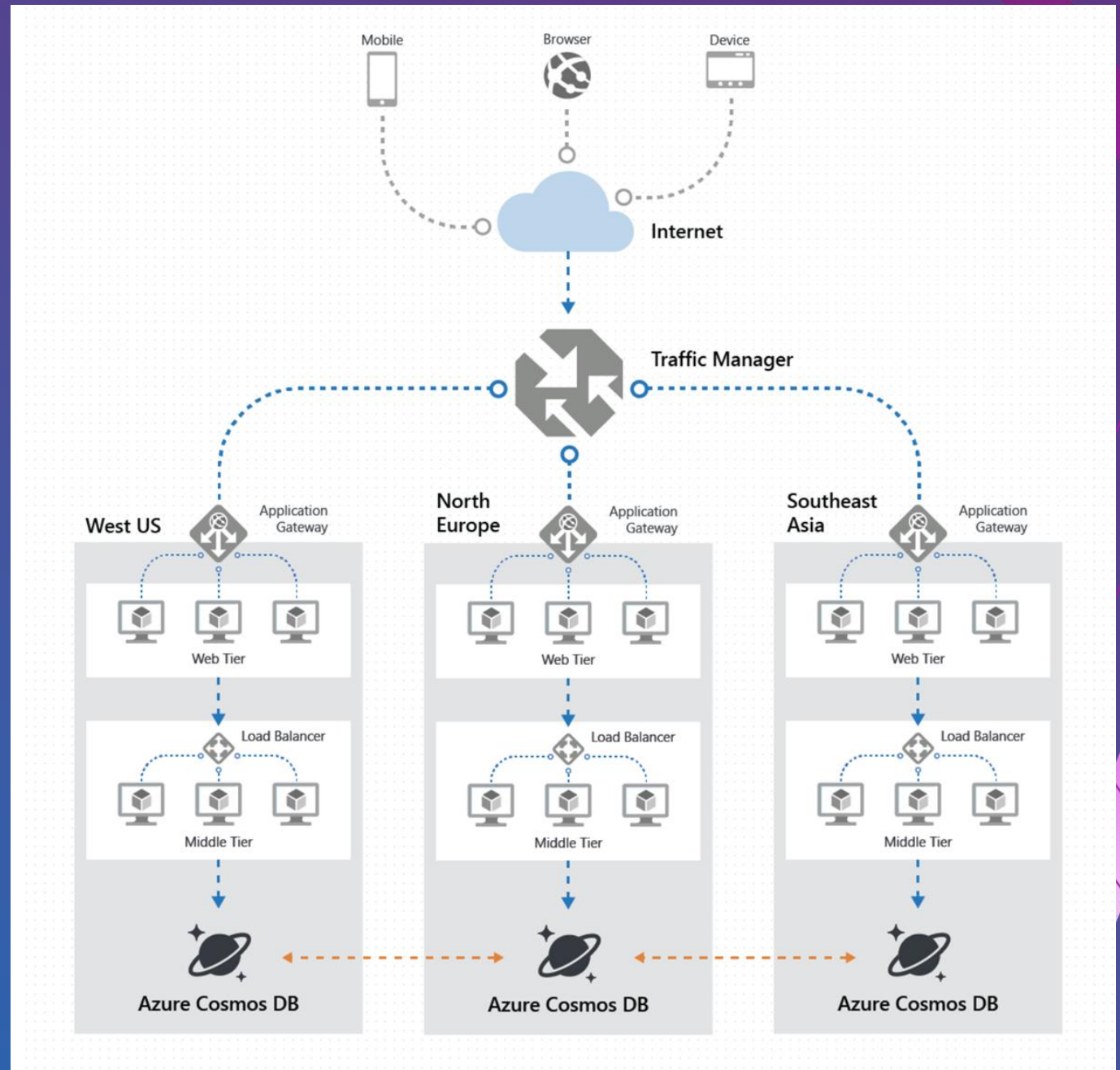
Azure Cosmos DB

Azure Cosmos DB est une base de données NoSQL, managée, multi-modèle, multi-region avec une scalabilité automatique instantanée et des temps de réponse à 1 chiffre tout en étant No-Ops et avec une disponibilité de "5 nines".

Multi-region

Cosmos DB est disponible dans **toutes** les regions Azure.

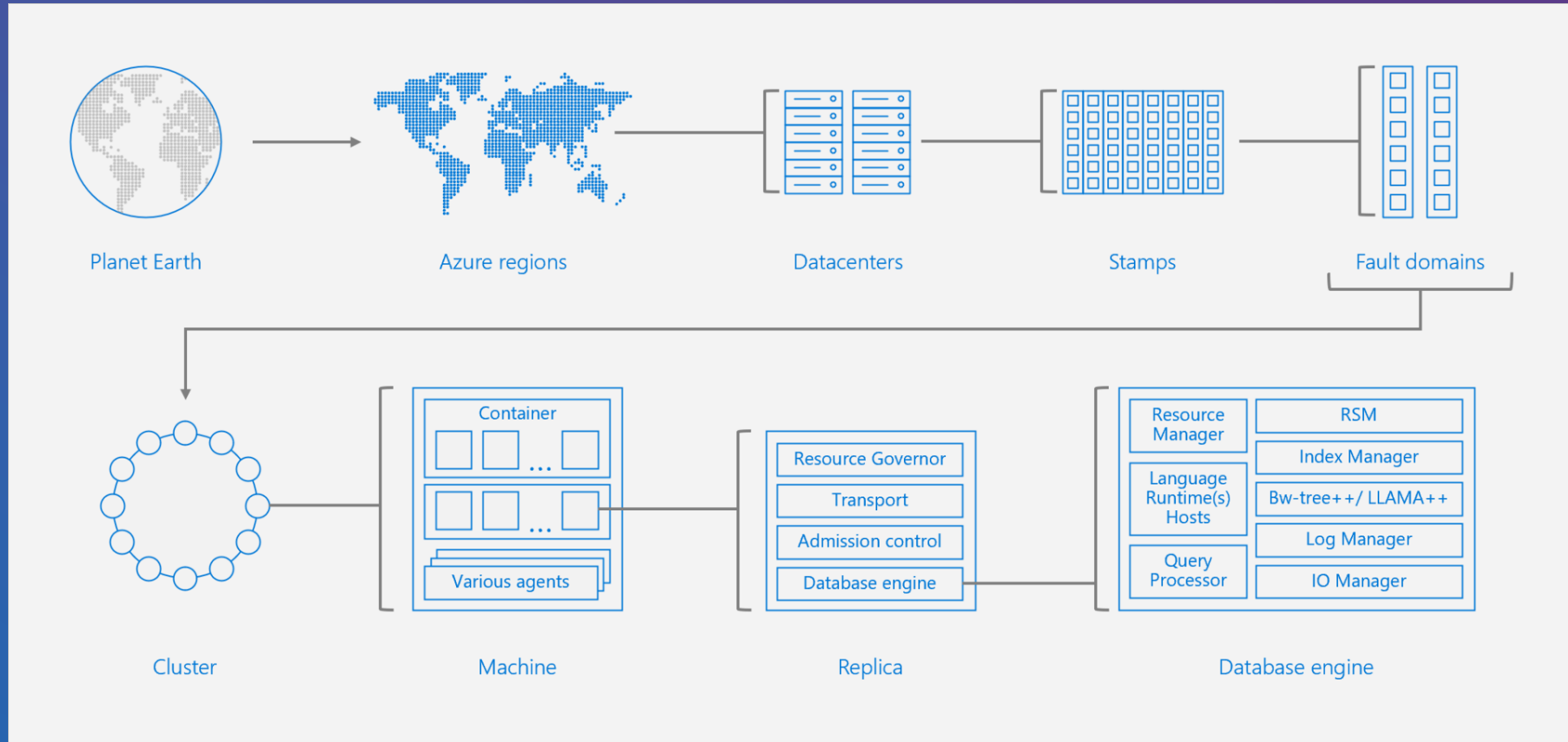
Failover automatique
Disponibilité 99.999%
 $R/W < 10\text{ms}$ 99th percentile



Multi-region



Multi-region



Multi-model

SQL

Table

Gremlin

MongoDB

Cassandra

Intro: dans une base CosmosDB

ACCOUNT

DATABASE

CONTAINER

A document icon representing a JSON document with the following content:

```
{  
  "id": "  
4sd48"  
}
```

A document icon representing a JSON document with the following content:

```
{  
  "id": "  
4sd48"  
}
```

CONTAINER

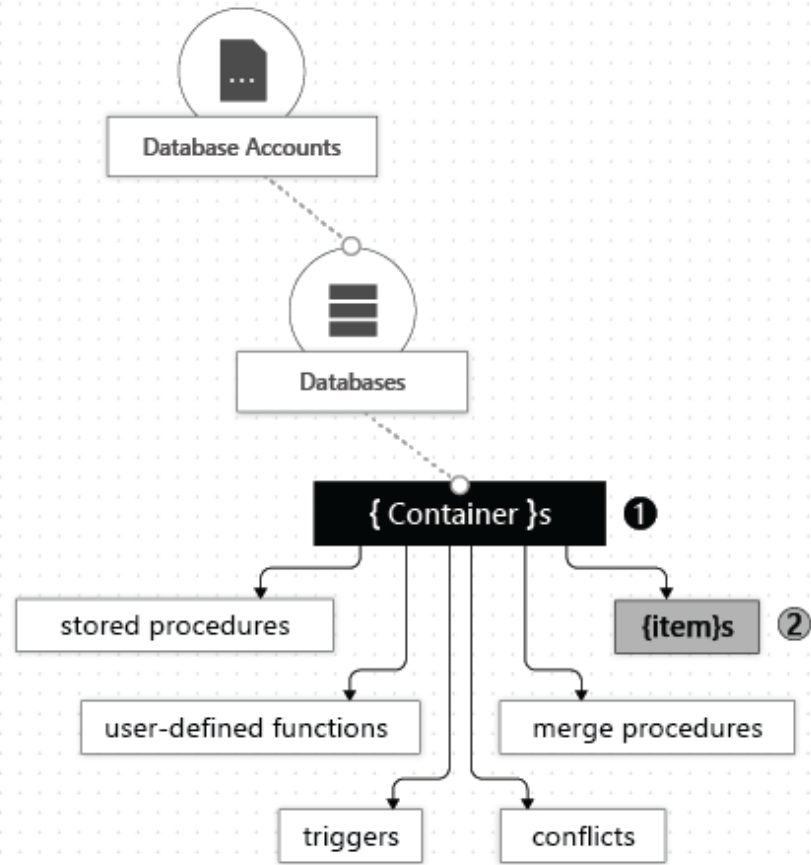
A document icon representing a JSON document with the following content:

```
{  
  "id": "  
4sd48"  
}
```

A document icon representing a JSON document with the following content:

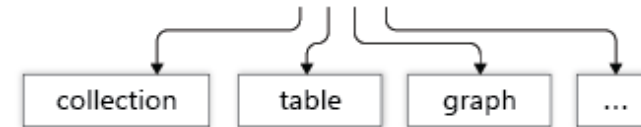
```
{  
  "id": "  
4sd48"  
}
```

Intro: dans une base CosmosDB



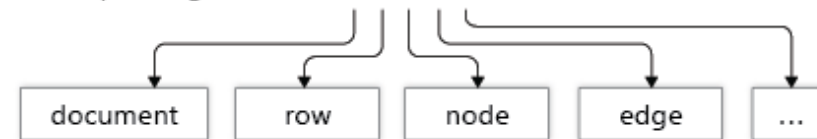
① { Container }s

Depending on the Cosmos API, a container is realized as:



② { item }s

Depending on the Cosmos API, an item is realized as:



Schema-less

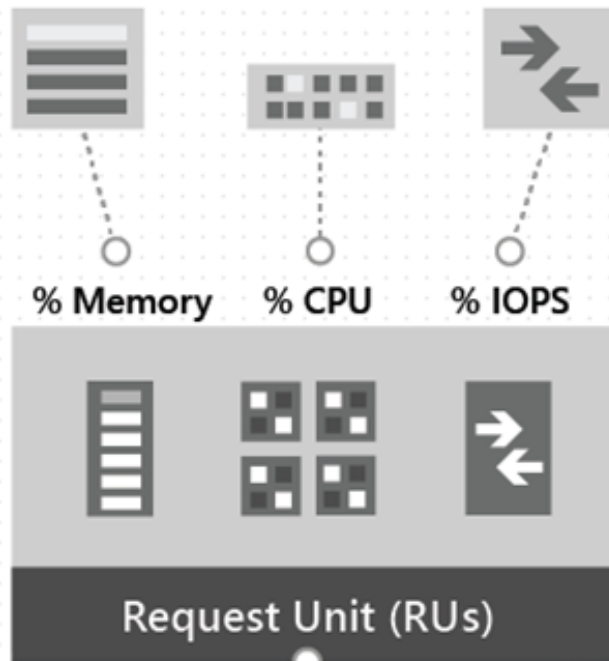
CONTAINER

```
{  
  "id":  
    "dh4h12Nf",  
  "firstName":  
    "Jean",  
  "lastName":  
    "DUPONT",  
  "requestDate":  
    "2022-04-12"  
}
```

```
{  
  "id":  
    "dh4h12Nf",  
  "postalCode":  
    "82000",  
  "city": "Montauban",  
  "integrationType":  
    "webhook",  
  "integrationConfig":  
    [  
      {"endpoint": "https://..."}  
    ]  
}
```

Object-relational mapping impedence

Scalabilité automatique – les RU/s



Read	=		1 RU	1kb = 1 RU
Insert	=	 --- 	5 RU mini	
Upsert	=	 --- 		
Delete	=	 --- 		
Query	=	 	Variable number of RUs	

Scalabilité automatique – les RU/s

Provisionned throughput

RU/s fixes

Autoscale throughput

10-100% RU/s
100 RU/s minimum

Serverless

0-5000 RU/s
50 GB / container max

1 000 000 RU/s max par conteneur *par défaut*

“No-Ops” (et No-ETL)

Cosmos DB est un service entièrement managé:

- Backup
- Failover automatique

Intégration à Azure Synapse Analytics, sans besoin d'ETL

Questions