Cloud Computing Architecture

Scaling Data Stores

SWINBURNE UNIVERSITY OF TECHNOLOGY

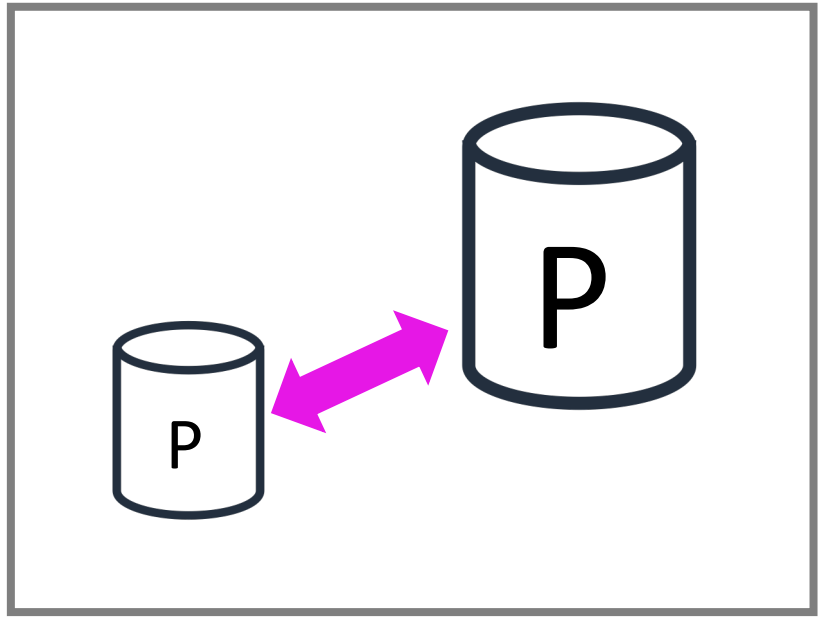
Scaling Data Stores

This presentation:

- Vertical Scaling – RDS
- Horizontal Scaling – RDS
- Auto Scaling – Dynamo DB

Vertical scaling with Amazon RDS: Push-button scaling

- Scale DB instances vertically up or down
- From micro to 24xlarge and everything in between
- Scale vertically with minimal downtime



Caching

For increased read performance we can put a cache in front of our RDS database.

For example we can use the non-relational database Elasticache for Memcached.

Horizontal scaling: Database sharding

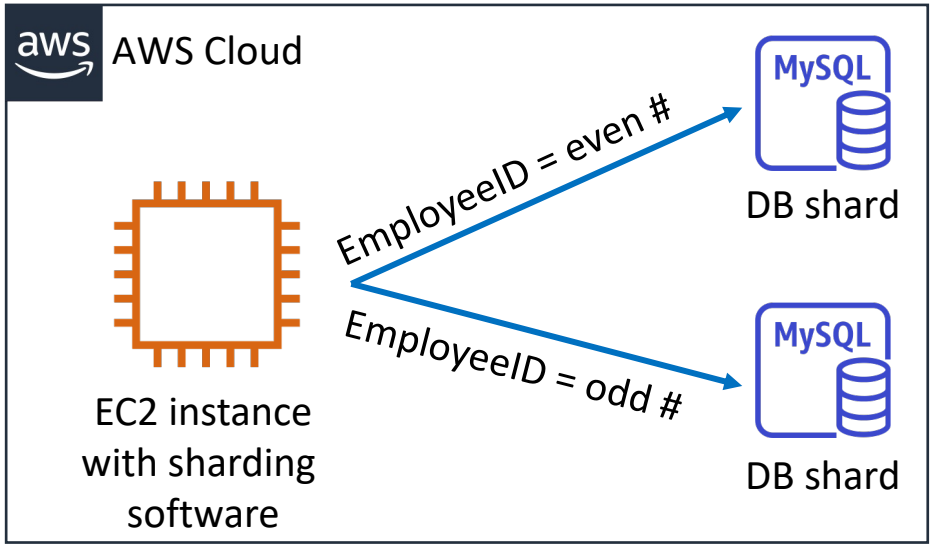
Without shards, all data resides in one partition.

- Example: Employee IDs in one database

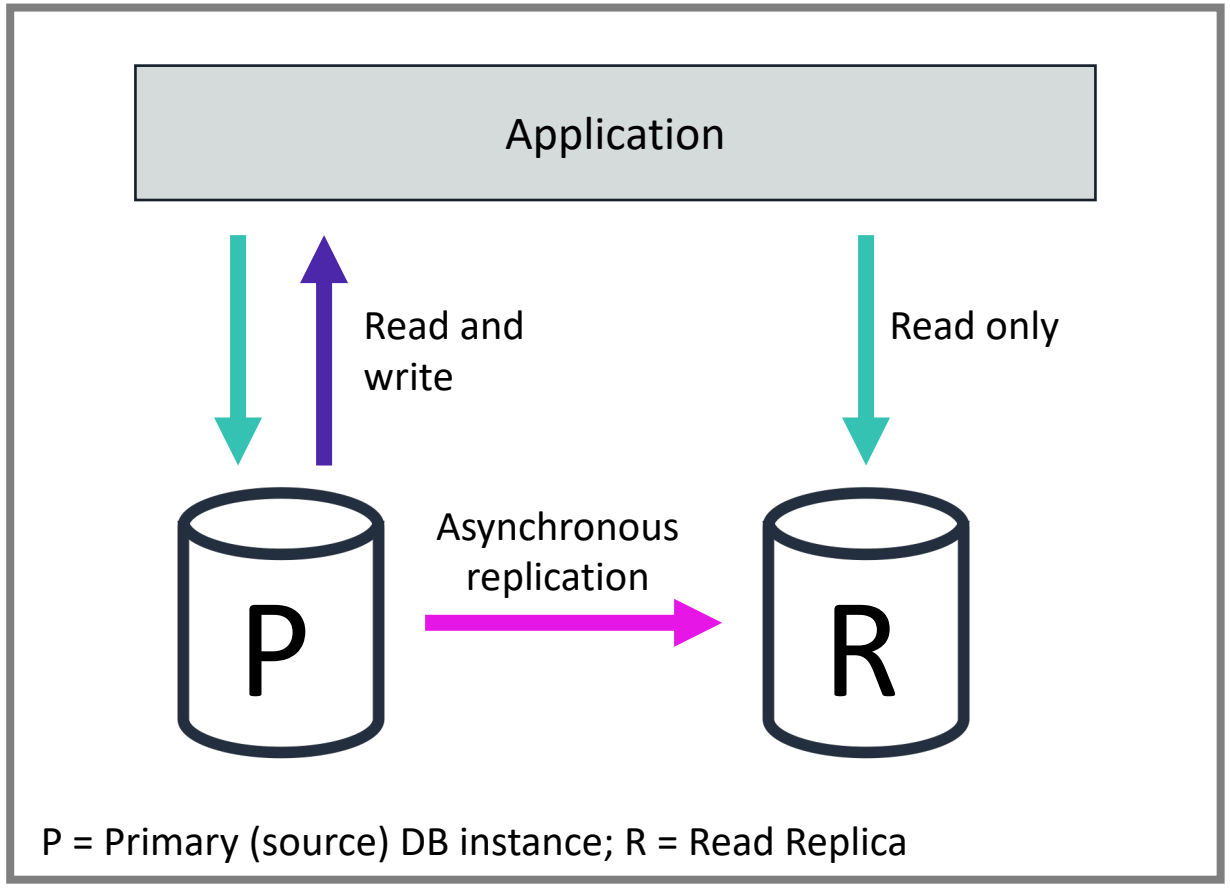
With sharding, data is split into large chunks (shards).

- Example: Even-numbered employee IDs in one database, and odd-numbered employee IDs in another database

In many circumstances, sharding improves write performance.



Horizontal scaling with Amazon RDS: Read replicas



- Horizontally scale for read-heavy workloads
- Up to five read replicas and up to 15 Aurora replicas
- Replication is asynchronous
- Available for Amazon RDS for MySQL, MariaDB, PostgreSQL, and Oracle

Scaling with Amazon DynamoDB: Auto scaling

On-Demand

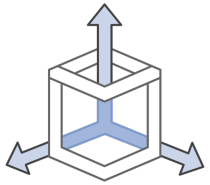
Pay per request

No more provisioning

Use case: Spiky, unpredictable workloads. Rapidly accommodates to need.

Auto scaling

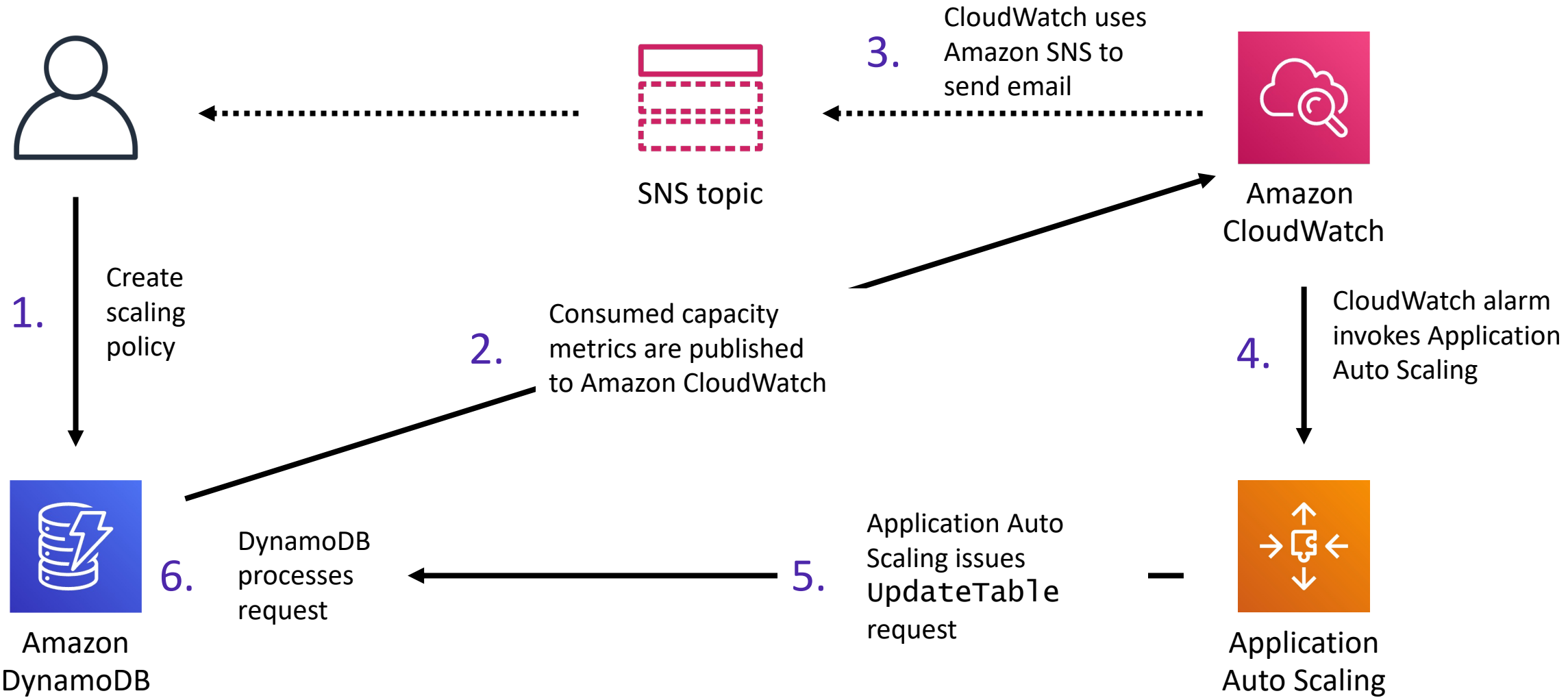
Default for all new tables



Specify upper and lower bounds

Use case: General scaling, good solution for most applications.

How to implement DynamoDB auto scaling



Lecture References

Recommend Viewing

Swinburne Lecture – High Level Overview

AWS Academy – Deeper dive

ACA Module 9