

SESSION-3: COMPUTE and DATABASE

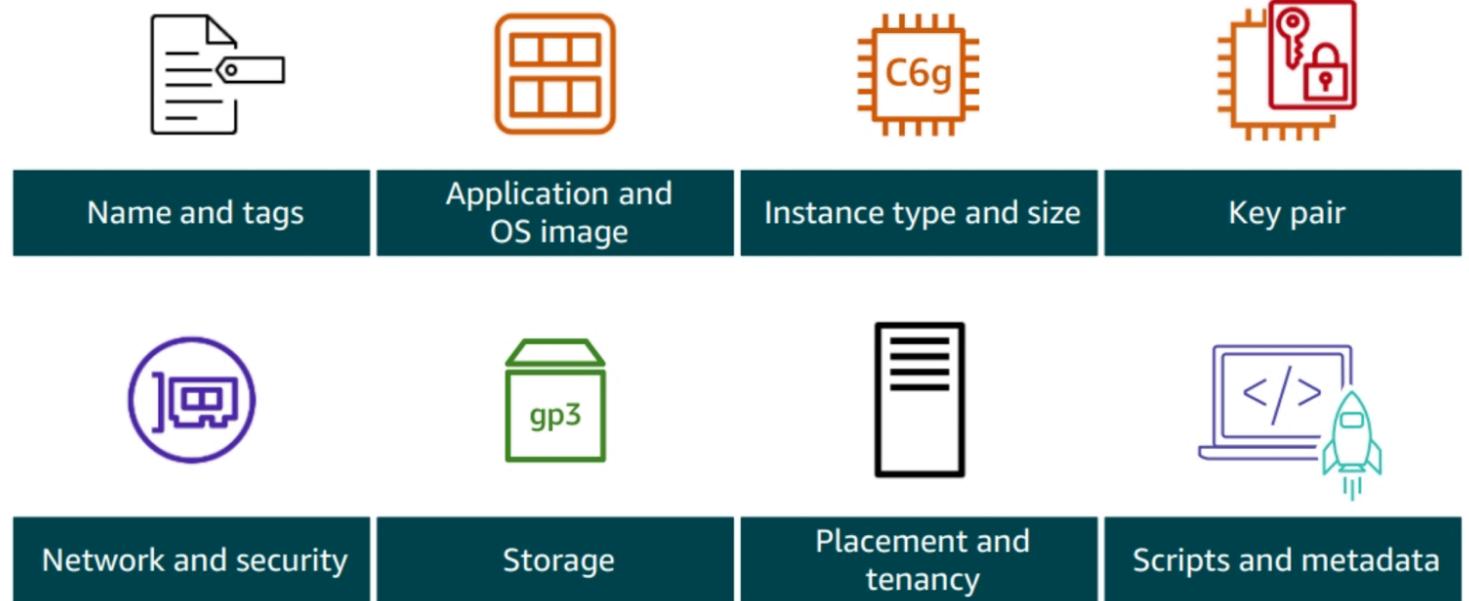
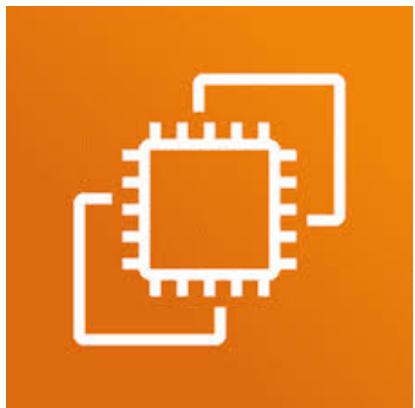


AWS Compute History

- In 2006, Amazon EC2 for Virtualization.
- In 2014, Amazon Elastic Container Service (Amazon ECS) for Containerization.
- In 2014, AWS Lambda for Serverless.
- In 2017, AWS Fargate for Serverless Containerization.
- In 2022, AWS Inferentia, Trainium and Graviton processors for AWS custom-built and specialized processors.

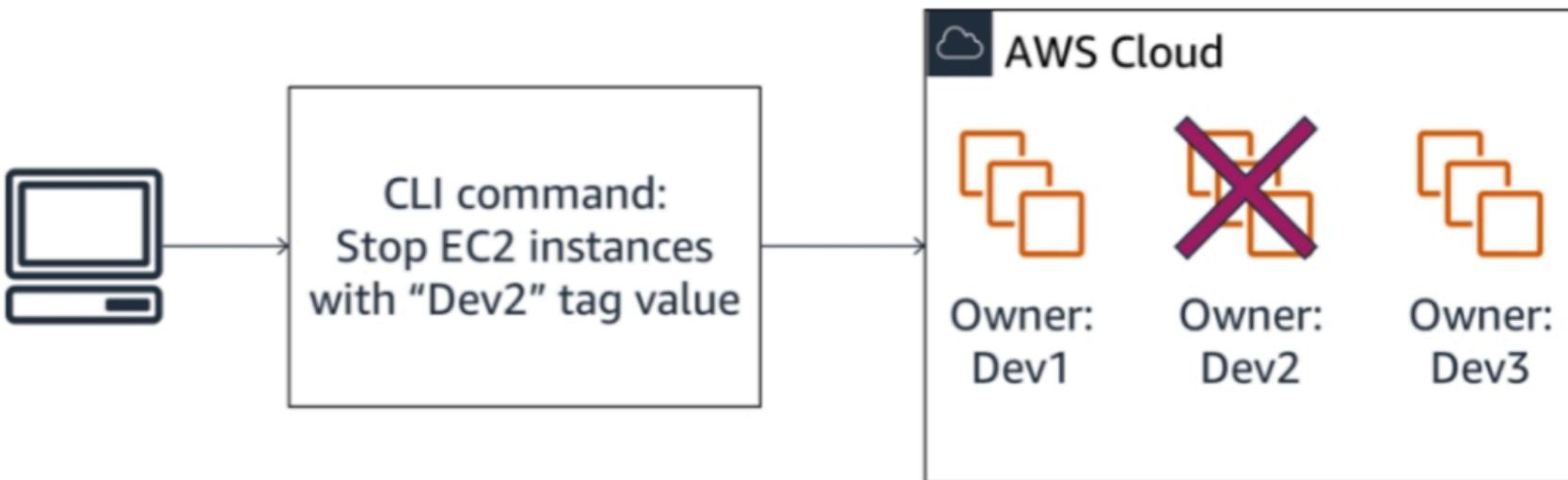
Amazon EC2 Instances

- Physical servers host EC2 instances in AWS Regions.
- EC2 instances give secure and resizable compute capacity in the cloud.
- The number of EC2 instances can be changed based on demand.



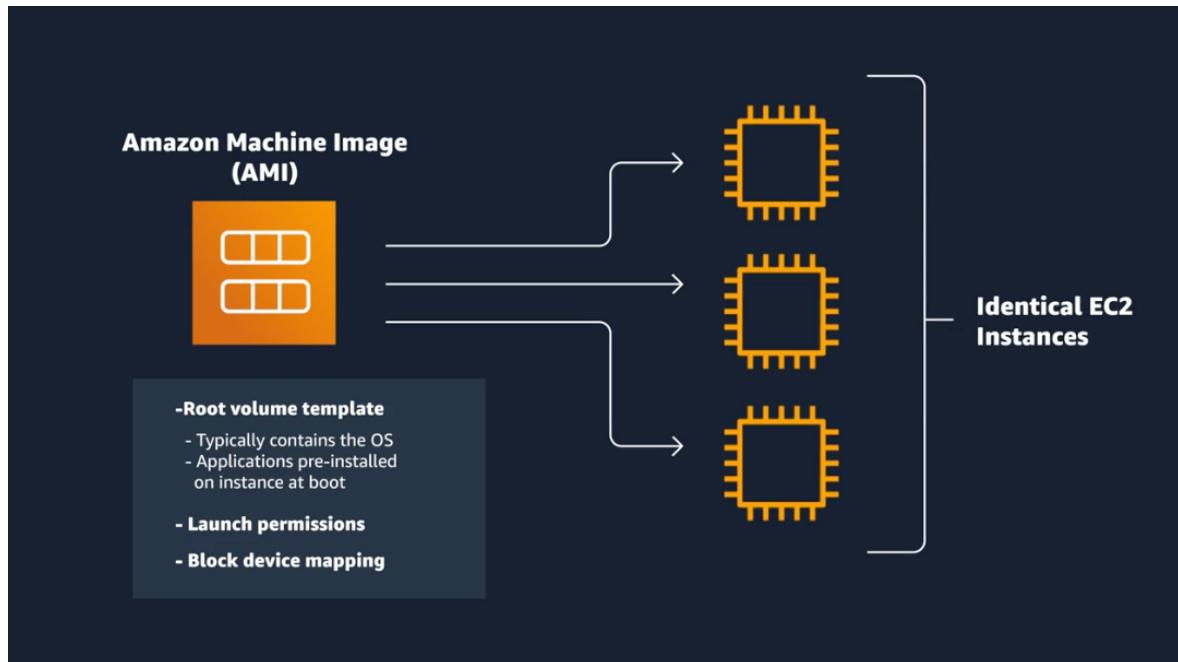
Tags

- Assigning names and tags is important for managing AWS resources.
- Also, tags provide filtering the resources and they are case-sensitive.



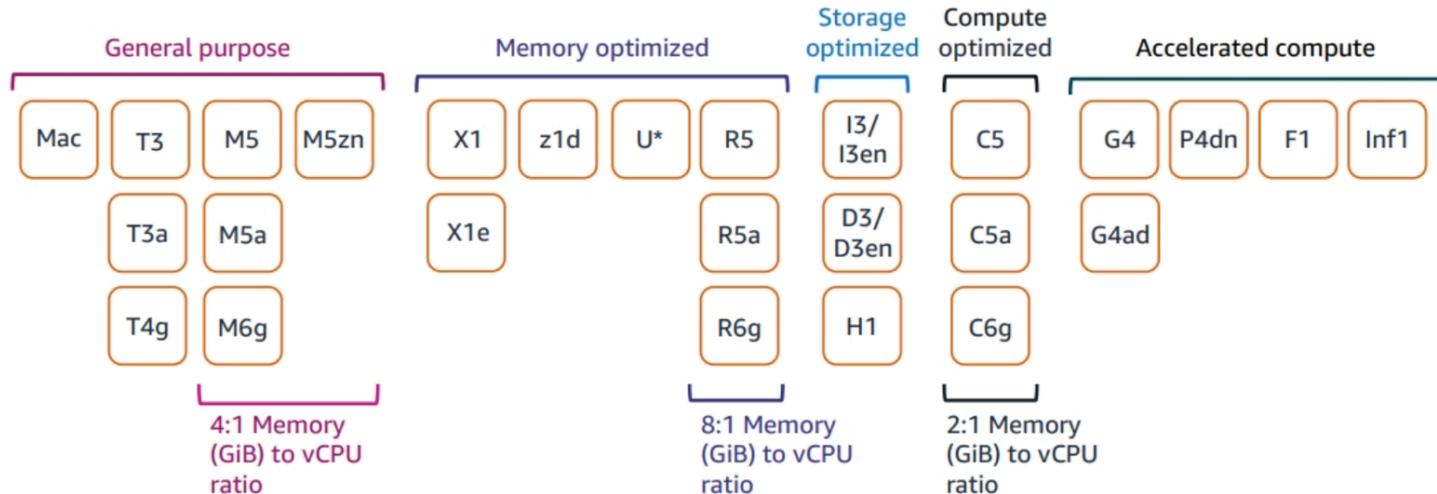
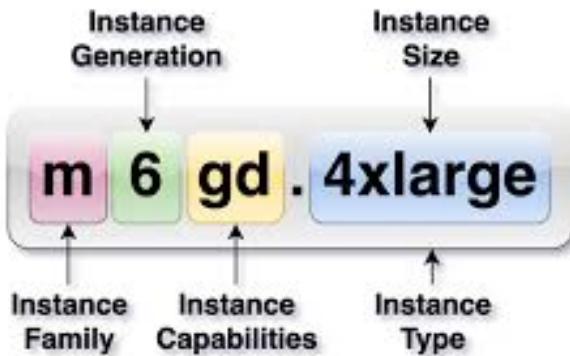
Amazon Machine Image (AMI)

- AMI is repeatable, reusable and recoverable.
- There are different resources for AMIs: AWS prebuilt AMIs, AWS Marketplace for catalog of AMIs, AMIs created by user or Amazon EC2 Image Builder



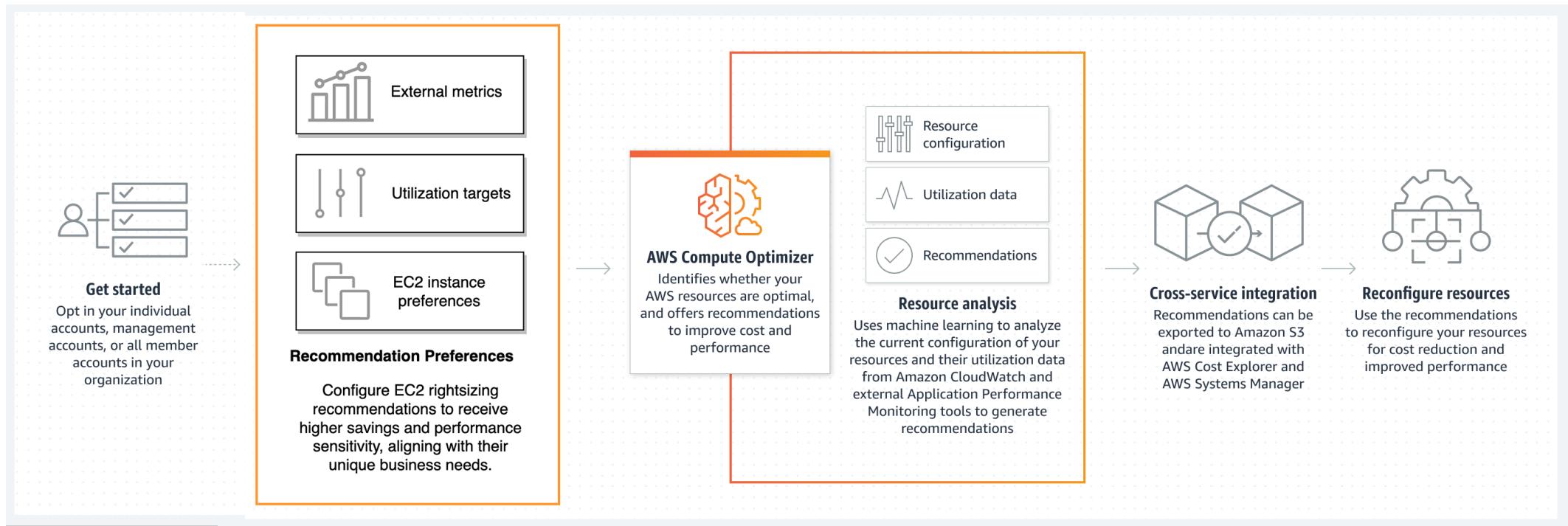


Instance Type Names

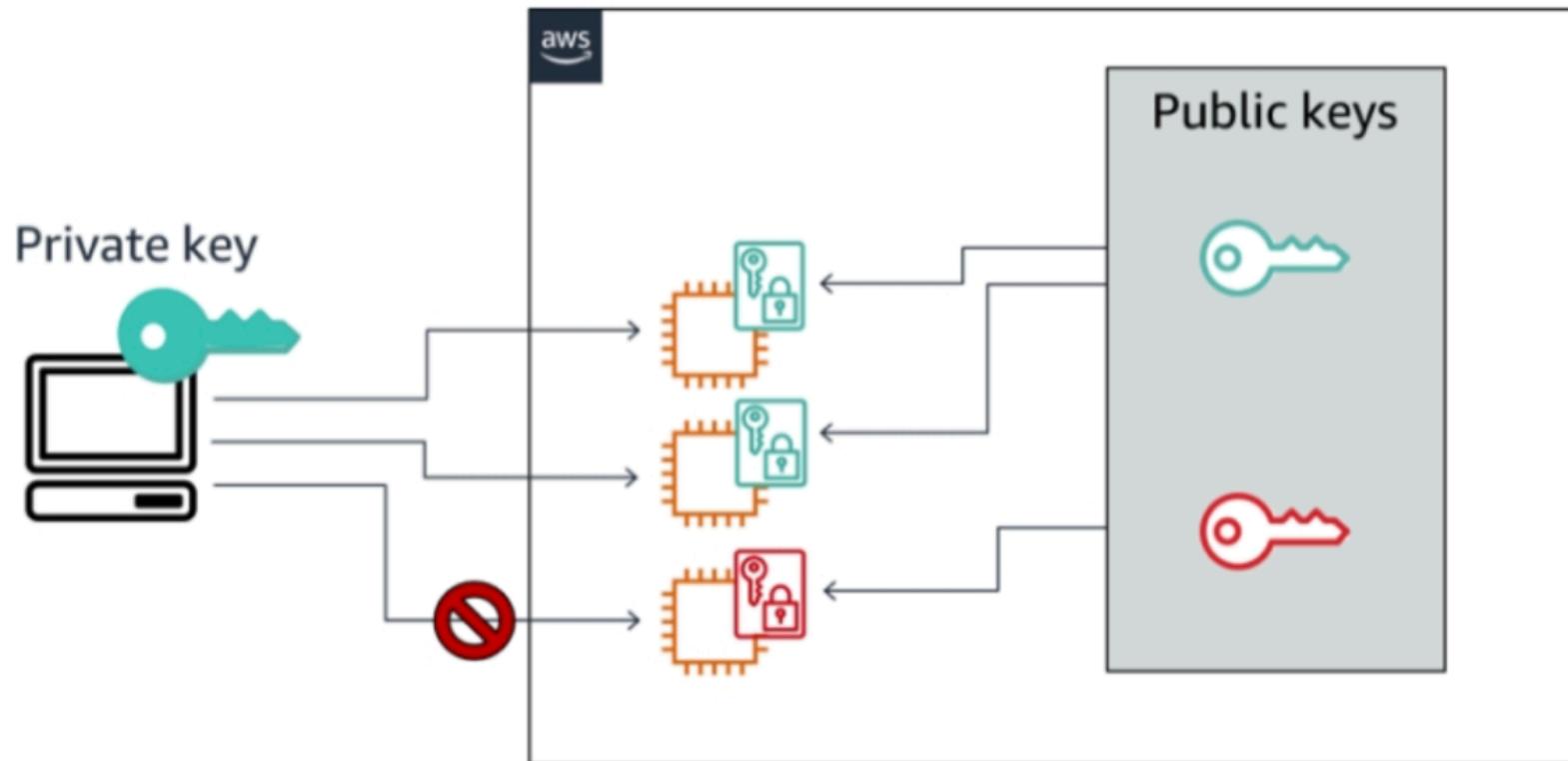




AWS Compute Optimizer

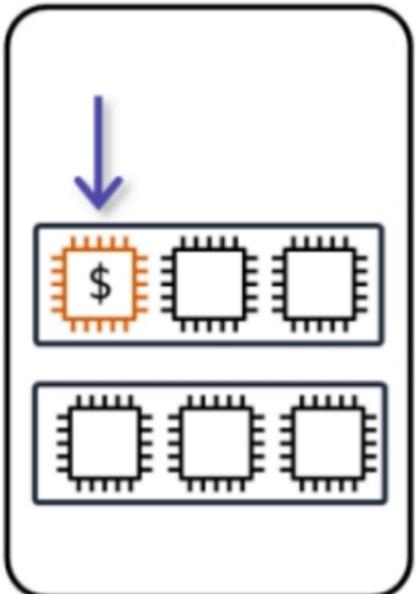


Amazon EC2 Key Pairs



Shared tenancy

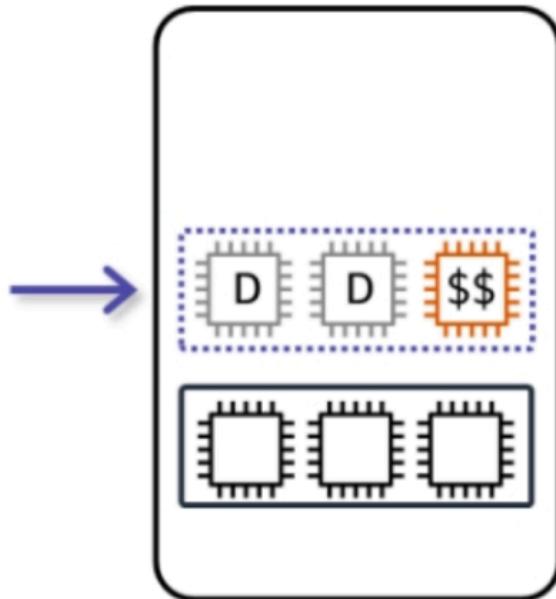
Share your hardware.



D = potential compute available
\$ = purchased compute

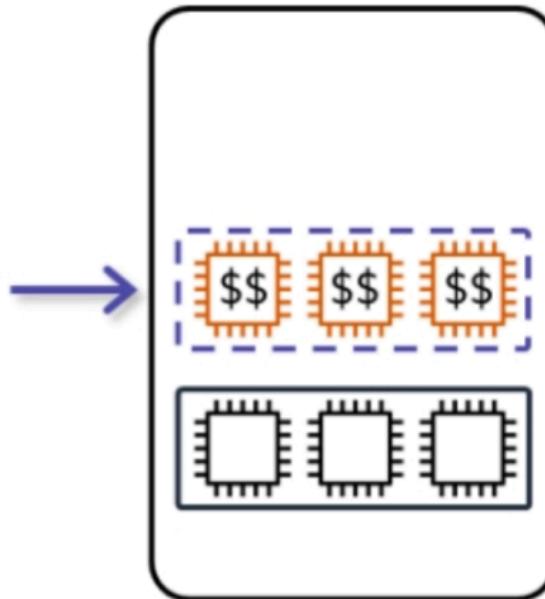
Dedicated Instance

Isolate your hardware.



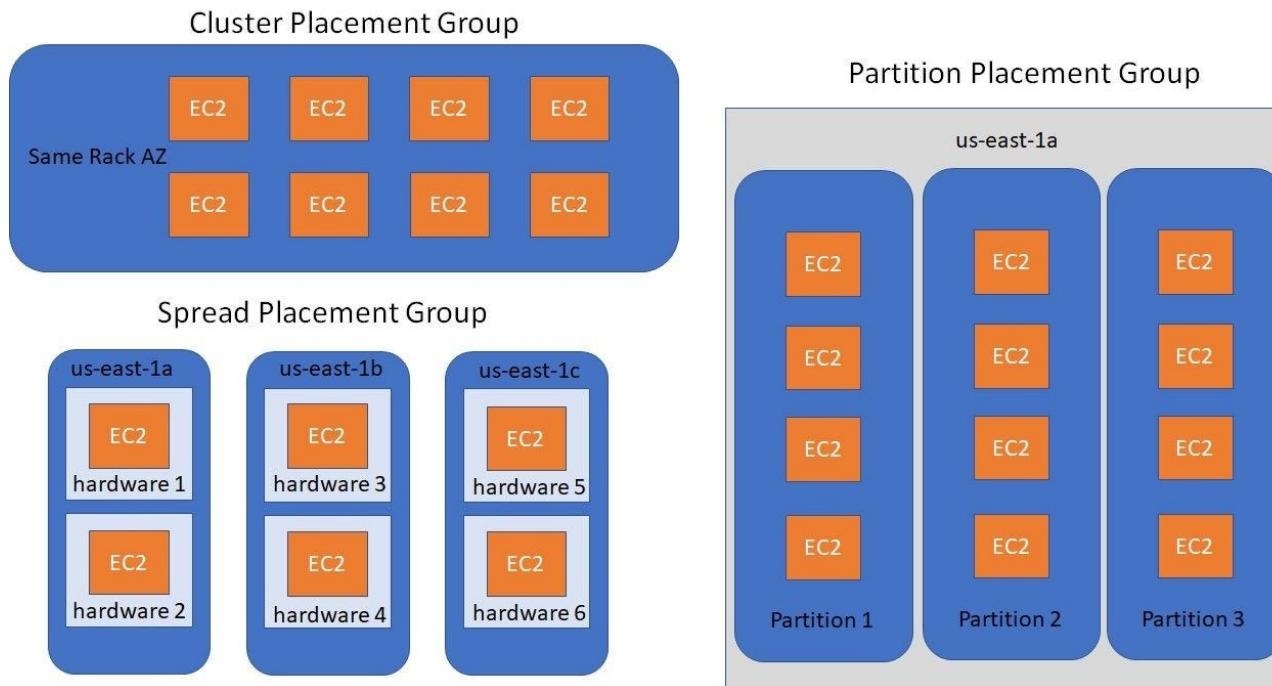
Dedicated Host

Control your hardware.



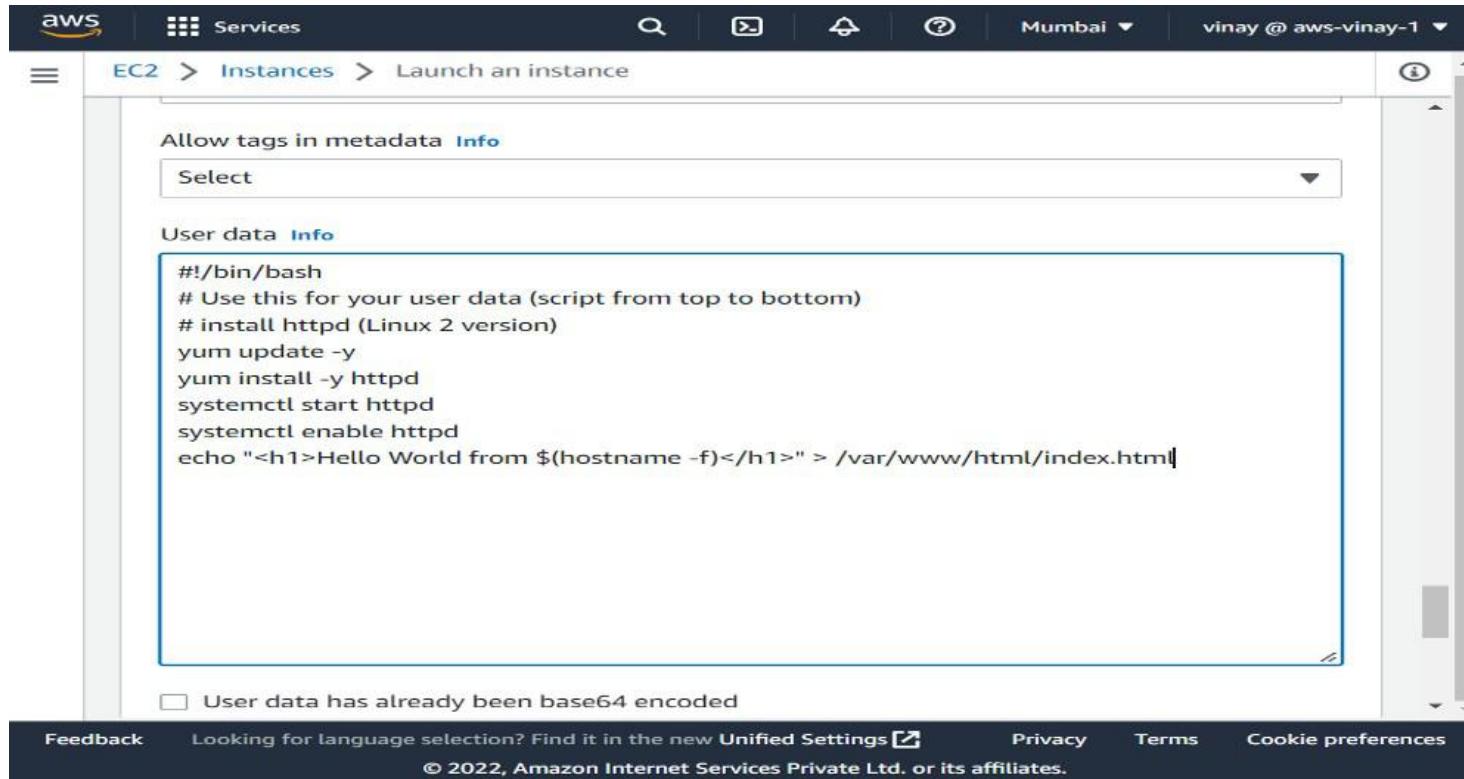
Placement Groups

- **Cluster** for low network latency and high network throughput.
- **Spread** for critical instances that should be fault-tolerant.
- **Partition** for avoiding correlated hardware failures.



User Data

- Running scripts as root user in the instance launch.
- Useful for automation tasks.



Instance Metadata

- It is only reachable from EC2 instance.

EC2 instance metadata can be used for automation.

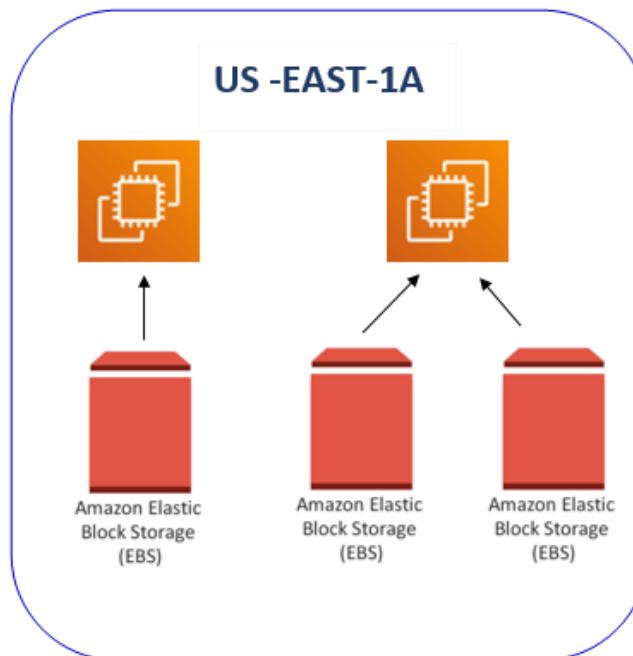
```
TOKEN=`curl -X PUT "http://169.254.169.254/latest/api/token"  
-H "X-aws-ec2-metadata-token-ttl-seconds: 21600"  
  
curl http://169.254.169.254/latest/meta-data/public-hostname  
-H "X-aws-ec2-metadata-token: $TOKEN"
```

| Metadata | Value |
|-----------------|--|
| instance-id | i-1234567890abcdef0 |
| mac | 00-1B-63-84-45-E6 |
| public-hostname | ec2-203-0-113-25.compute-1.amazonaws.com |
| public-ipv4 | 203.0.113.25 |
| local-hostname | ip-10-251-50-12.ec2.internal |
| local-ipv4 | 10.251.50.12 |



Amazon Elastic Block Store (Amazon EBS)

- Block-level storage in AZ.
- One or more volumes can be attached to a single EC2 instance.
- Can be move between instances.



Amazon EBS Volume Types

- SSD for high-performance and general purpose, HDD for big or infrequently accessed data.

| | Solid State Drives (SSD) | | Hard Disk Drives (HDD) | | |
|-------------|--------------------------------|--|--|---|------------------|
| Volume Type | General Purpose | Provisioned IOPS SSD | Throughput Optimized HDD | Cold HDD | EBS Magnetic |
| API Names | gp2 | io1 | st1 | sc1 | standard |
| Description | Balances price and performance | Highest SSD performance for Mission-critical low latency or high throughput | Low-cost. Designed for frequently accessed, throughput intensive workloads | Lowest HDD cost. Less frequently used workloads | |
| Use Cases | Most Workloads | Large Databases IOPS greater than 16,000 or Throughput greater than 250 MiB | Data Warehouses Big Data Log Processing | File Storage | Archival Storage |
| Volume Size | 1GiB - 16TiB | 4GB - 16 TiB | 500GiB - 15TiB | 500GiB - 15TiB | 500GiB - 15TiB |
| Max IOPS | 16,000 | 64,000 | 500 | 250 | 40-200 |

1. **General Purpose (SSD)** (gp2) for general usage without specific requirements
2. **Provisioned IOPS (SSD)** (io1) when you require really fast input & output
3. **Throughput Optimized HDD** (st1) magnetic drive optimised for quick throughput
4. **Cold HDD** (sc1) Lowest cost HDD volume for infrequently access workloads
5. **EBS Magnetic** (standard) previous generation HDD

Instance Store Volumes

- It is instance-local and non-persistent volume option. Also, it doesn't support snapshots.





Amazon EC2 Purchase Options

| Purchasing Method | Commitment Structure | Exchanging Options | % Savings over On-Demand |
|-----------------------|---|--|--------------------------|
| On-Demand | Charged on a \$/second basis for instances that are deployed | Not Applicable | Not Applicable |
| Compute Savings Plans | Commitments are made on \$/hour basis for term lengths of 1 or 3 years | Not Applicable | Up to 66% |
| EC2 Savings Plans | Commitments are made on \$/hour basis for term lengths of 1 or 3 years | Not Applicable | Up to 72% |
| Standard RI | Commitments are made to specific instance configurations for term lengths of 1 or 3 years | Can be bought and sold in the AWS RI Marketplace | Up to 72%* |
| Convertible RI | Commitments are made to specific instance configurations for term lengths of 1 or 3 years | Can be exchanged for other instances of same or different properties | Up to 66%* |

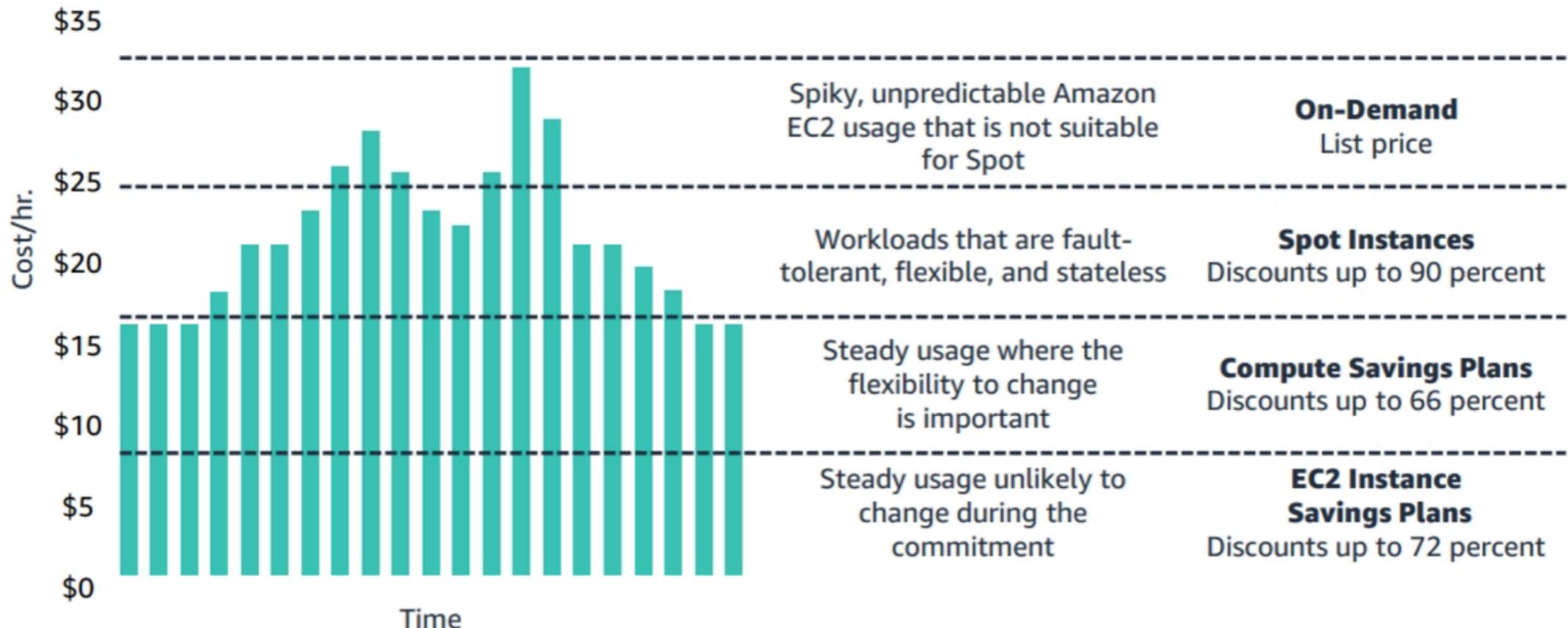
* RIIs save 5-10% over their Savings Plan counterparts



Savings Plan and Spot Instances

- Compute Savings Plans - Provides up to %66 off On-Demand rates, and applies to AWS Fargate and AWS Lambda.
- EC2 Instance Savings Plans - Provides the lowest prices, up to %72 off On-Demand rates on the selected instance family in a Region.
- Spot Instances - Decide what you can pay for compute, and save up to %90 from On-Demand price. Note that the interruptions should be taken into account.

Instance Purchase Options



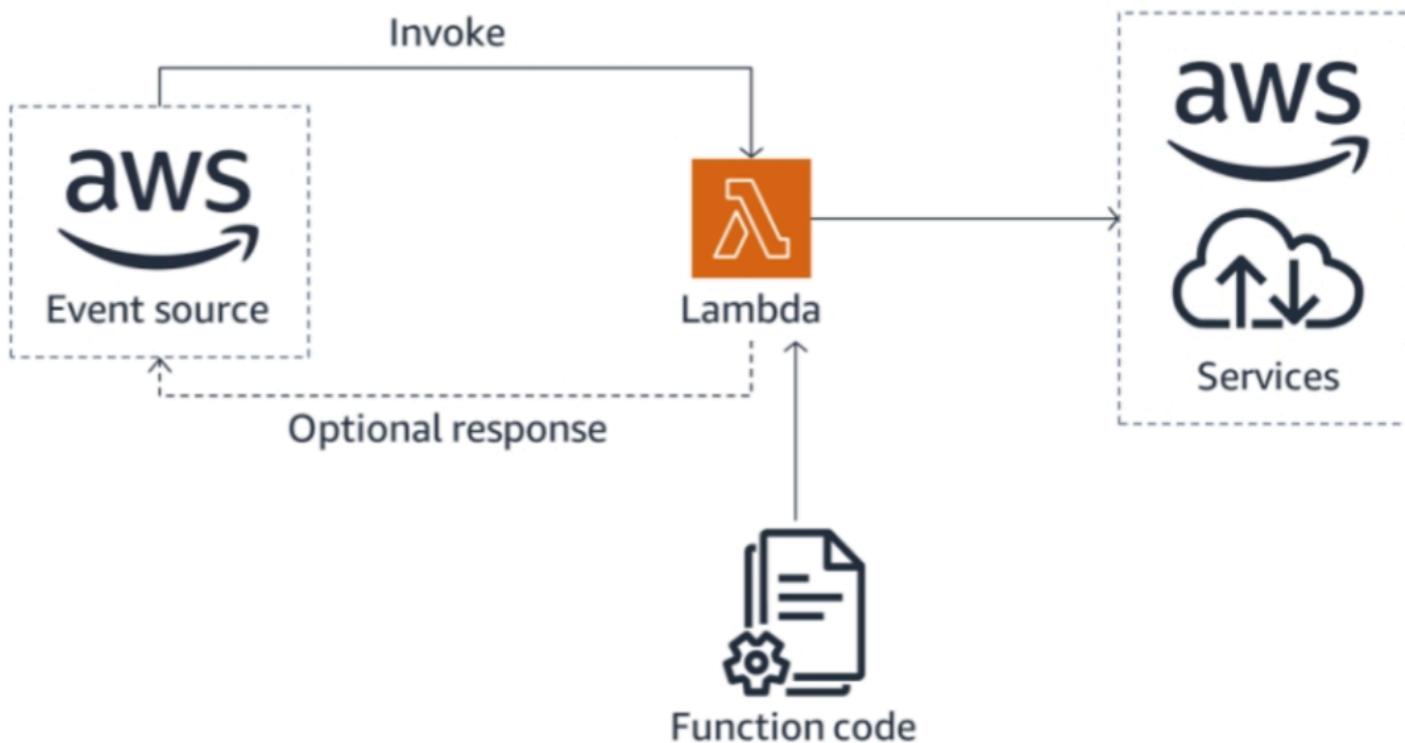
Serverless Computing

- It is highly available.
- It is fully managed by AWS. (No need for servers.)
- Event-source examples:



AWS Lambda

- Way of serverless computing by supporting different programming languages.
- Runs up to 15 minutes and supports up to 10 GB memory.



AWS database services



Amazon Relational Database
Service (Amazon RDS)



Amazon Aurora



Amazon Redshift



Amazon DocumentDB
(with MongoDB compatibility)



Amazon
DynamoDB



Amazon ElastiCache



Amazon MemoryDB
for Redis



Amazon Keyspaces
(for Apache Cassandra)



Amazon Timestream



Amazon Neptune



Amazon Quantum Ledger
Database (Amazon QLDB)

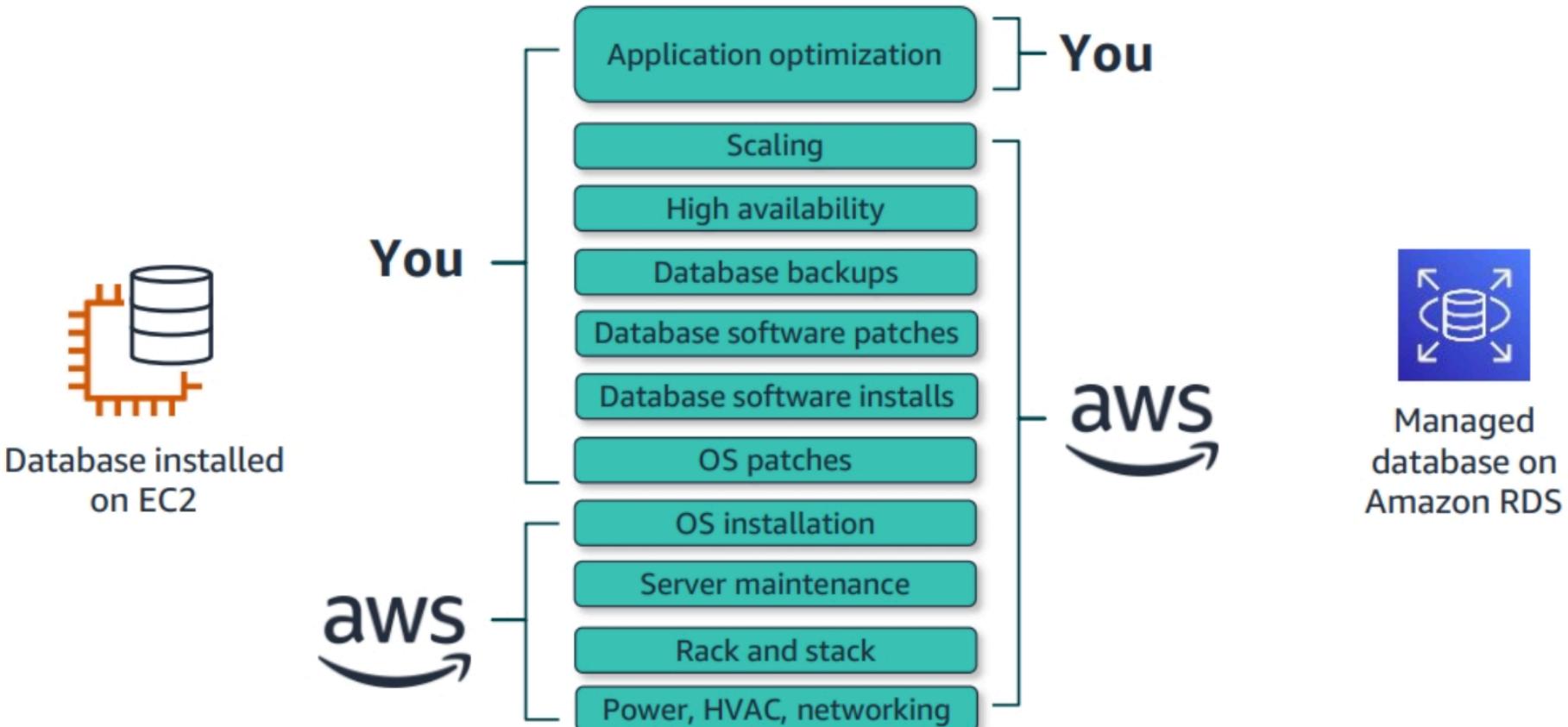
Relational and nonrelational databases

| | Relational (SQL) databases | Nonrelational (NoSQL) databases |
|---------------------------|---|--|
| Data storage | Tables with rows and columns | Key-value, wide-column, graph, document, or other models |
| Schemas | Fixed | Dynamic |
| Example database services |  Amazon RDS  Aurora |  DynamoDB  ElastiCache |

Choosing the right database

| Relational database | Nonrelational (NoSQL) database |
|--|---|
| You require strict schema rules and data quality enforcement. | You need your database to scale horizontally. |
| Your database doesn't need extreme read/write capacity. | Your data does not lend itself well to traditional schemas. |
| If you have a relational dataset that does not require extreme performance, a relational database management system can be the best, lowest effort solution. | Your read/write rates exceed the rates that can be economically supported through a traditional structured query language (SQL) database. |

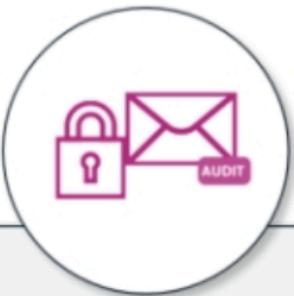
Managed and unmanaged services



Amazon RDS features



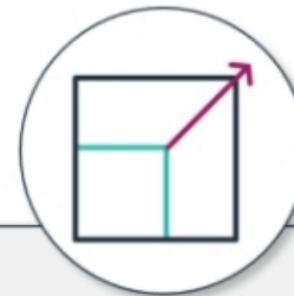
- Hardware, OS, and database software deployment and maintenance
- Built-in monitoring



- Data encryption at rest and in transit
- Industry compliance

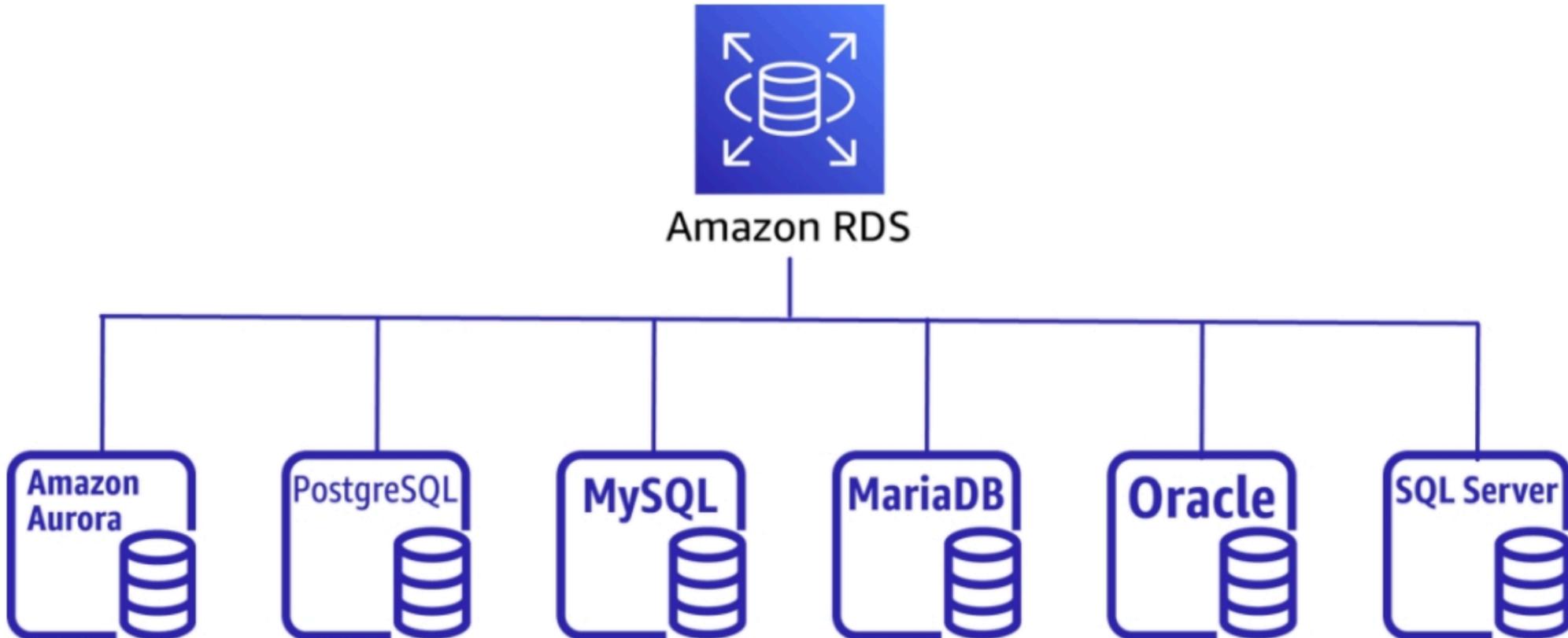


Automatic Multi-AZ data replication



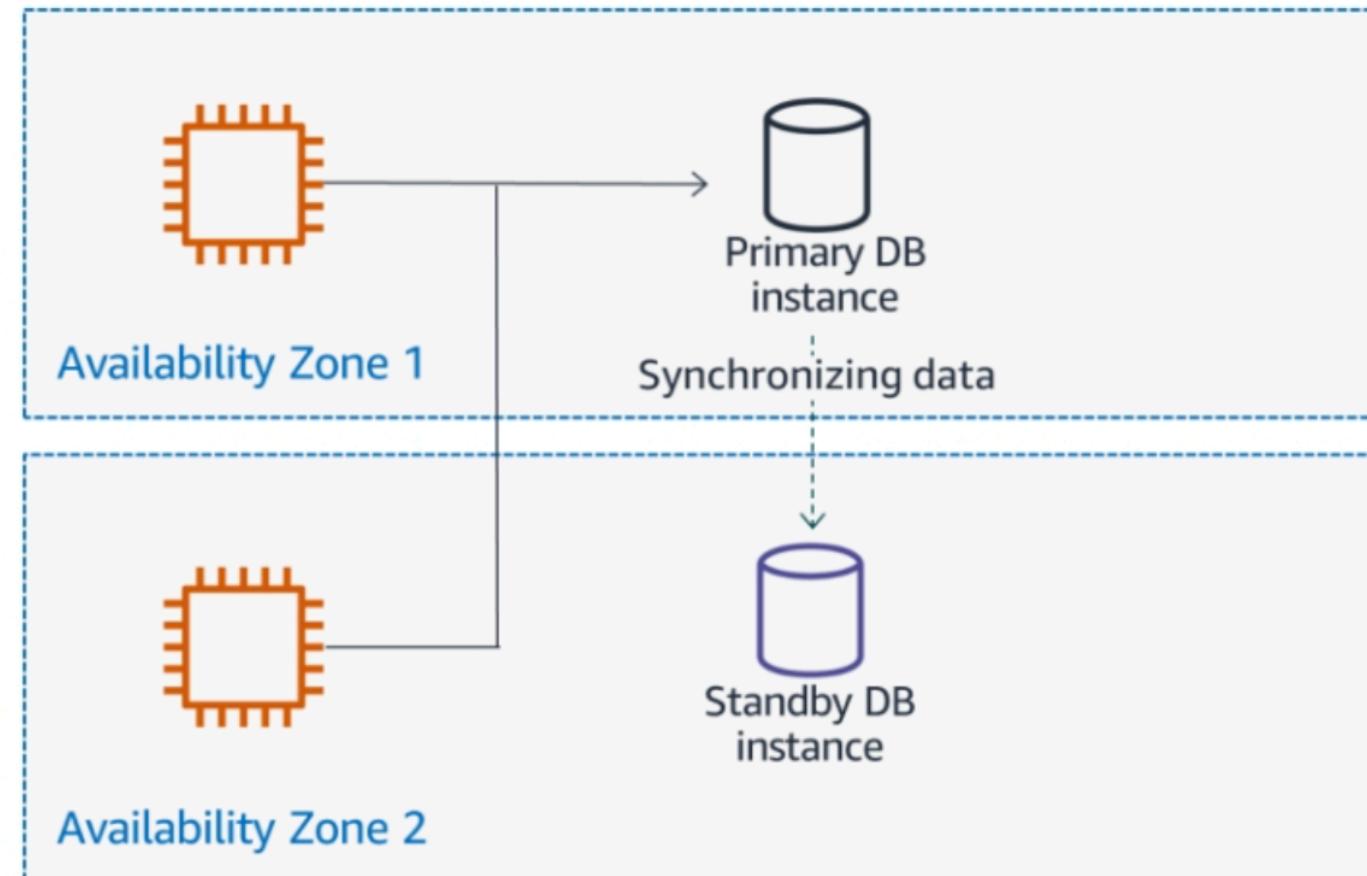
- Compute and storage scaling
- Minimal application downtime

Amazon RDS database engines



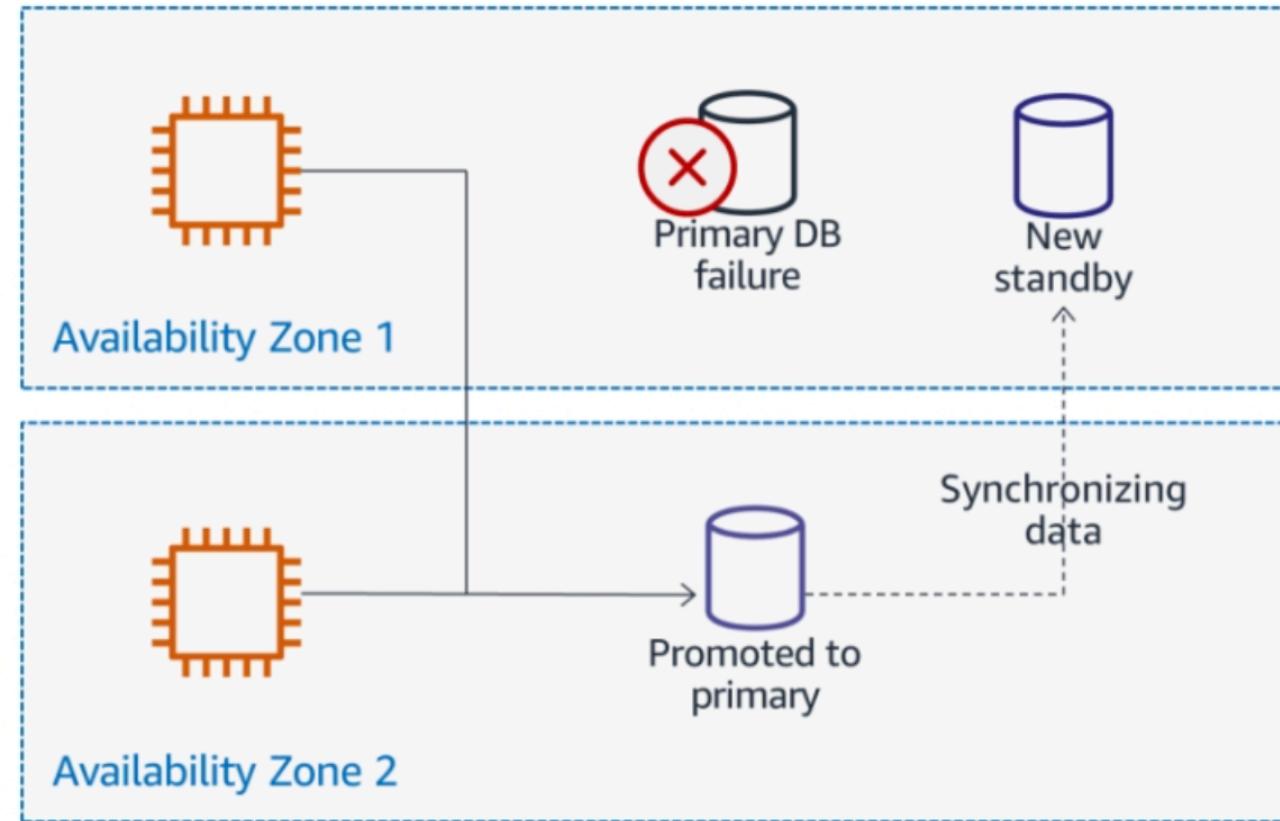
Amazon RDS Multi-AZ deployments

- Multi-AZ deployments:
- Replicate data to a standby DB instance in another Availability Zone
 - Not used for read-only scenarios



Amazon RDS Multi-AZ failover

Upon failure, the standby DB instance picks up the load.

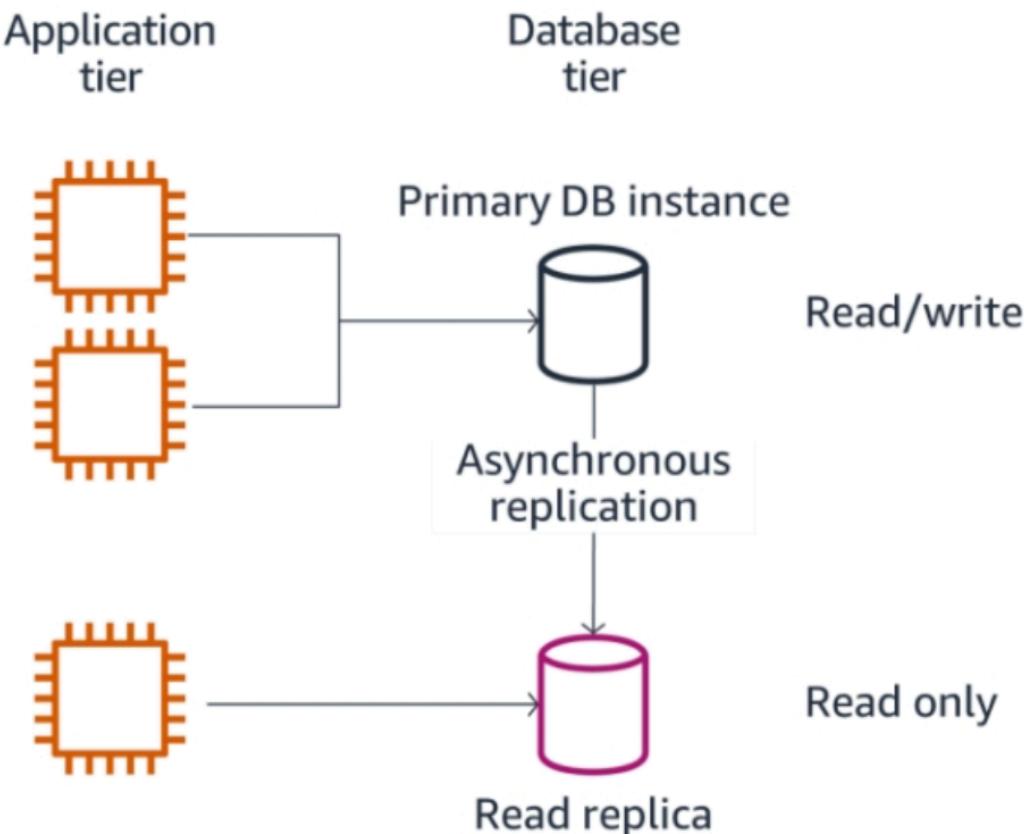


Read replicas

With read replicas, you can:

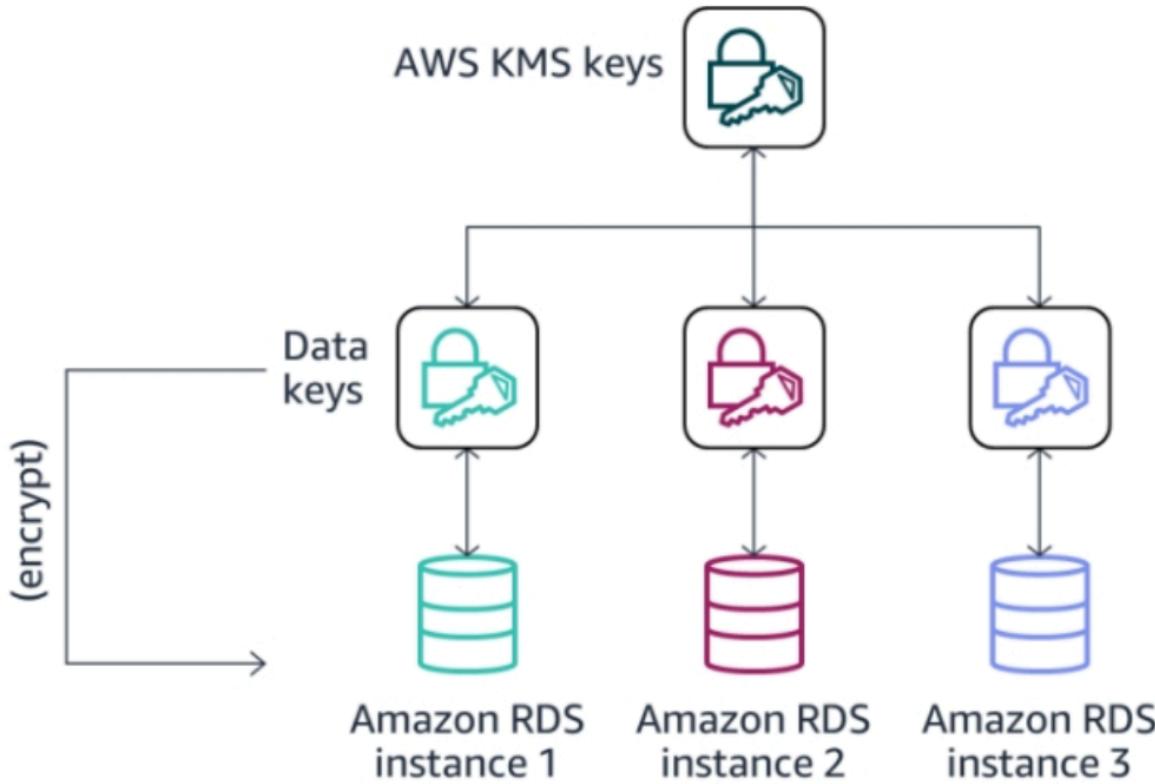
- Horizontally scale for read-heavy workloads
- Offload reporting
- Replicate across AWS Regions

aws



Data encryption at rest

- Managed by AWS KMS
- Unique data key encrypts your data
- AWS KMS key encrypts data keys
- Available for all RDS engines



Amazon Aurora

A MySQL and PostgreSQL compatible relational database built for the cloud



Performance
and scalability



Availability
and durability



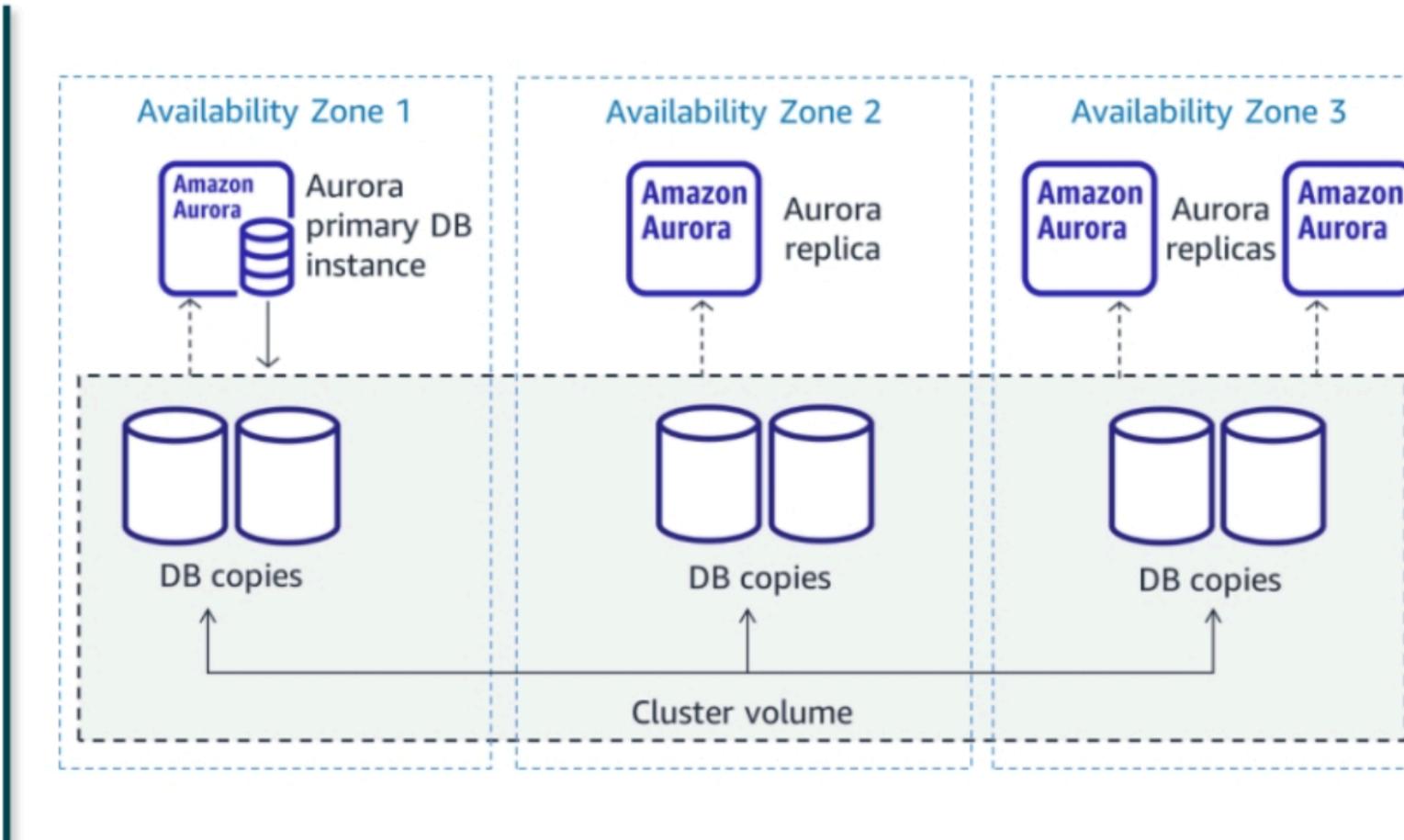
Highly
secure



Fully
managed

Aurora DB clusters

- A DB cluster consists of one or more DB instances and a cluster volume.
- Primary instances perform read/write operations.
- Aurora replicas are read-only.
- A cluster volume is a virtual database storage volume that spans multiple Availability Zones.



Aurora Serverless v2 for PostgreSQL and MySQL

Scaling configuration for Aurora that automatically scales capacity up or down based on your application's needs



Starts up on demand



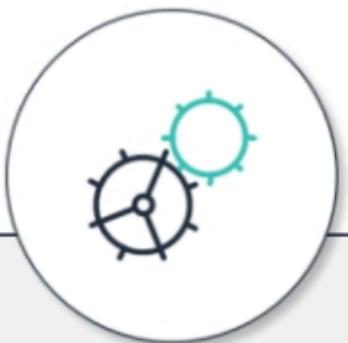
Only pay for what you use



No application impact when scaling

DynamoDB

A fully managed NoSQL AWS database service



Performance at scale



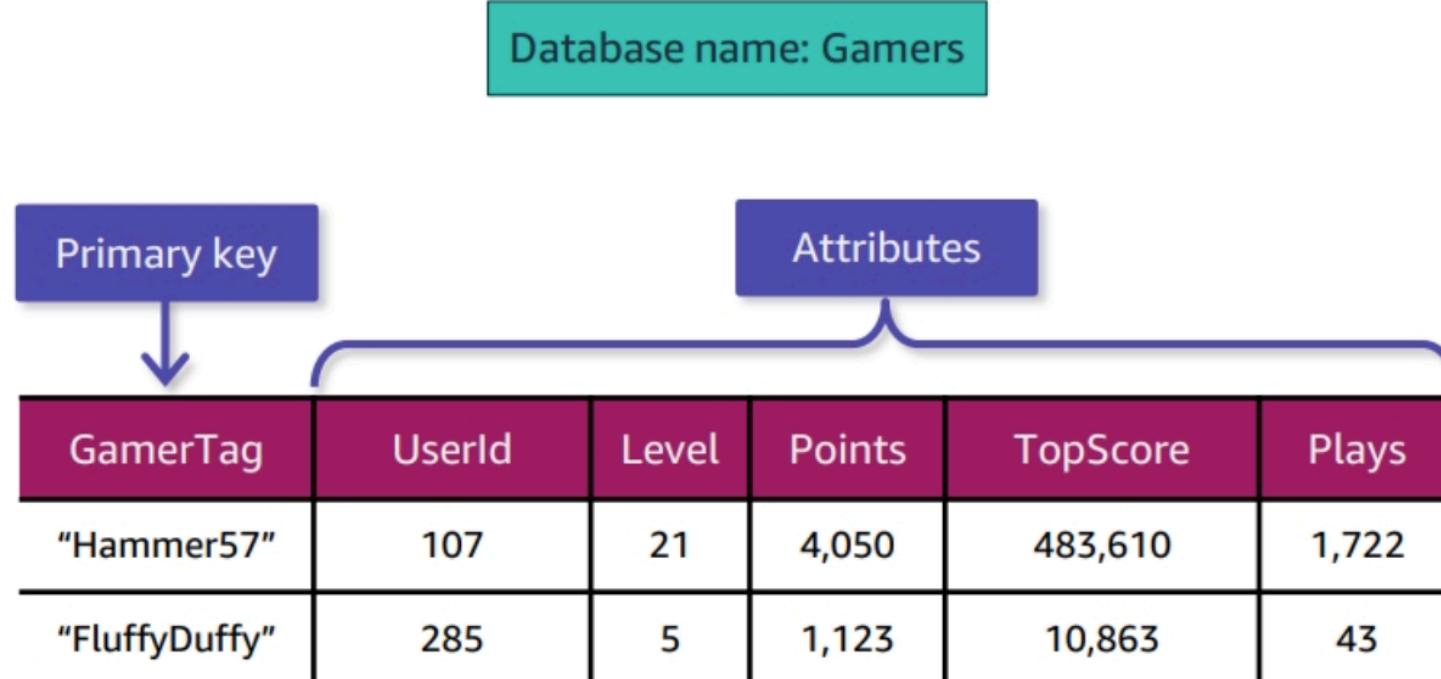
No servers to manage



Enterprise ready

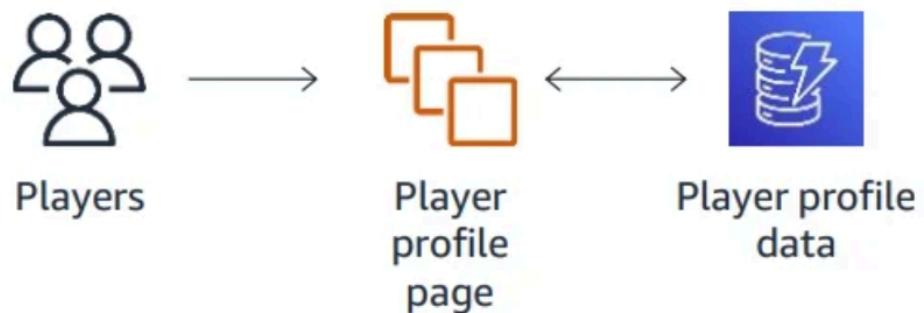
Key-value data

- Structured in simple key-value pairs with a flexible schema
- Ideal for uses where needed data can be mapped to a primary key
- Partitions data by key
- Delivers high-throughput, low-latency reads and writes



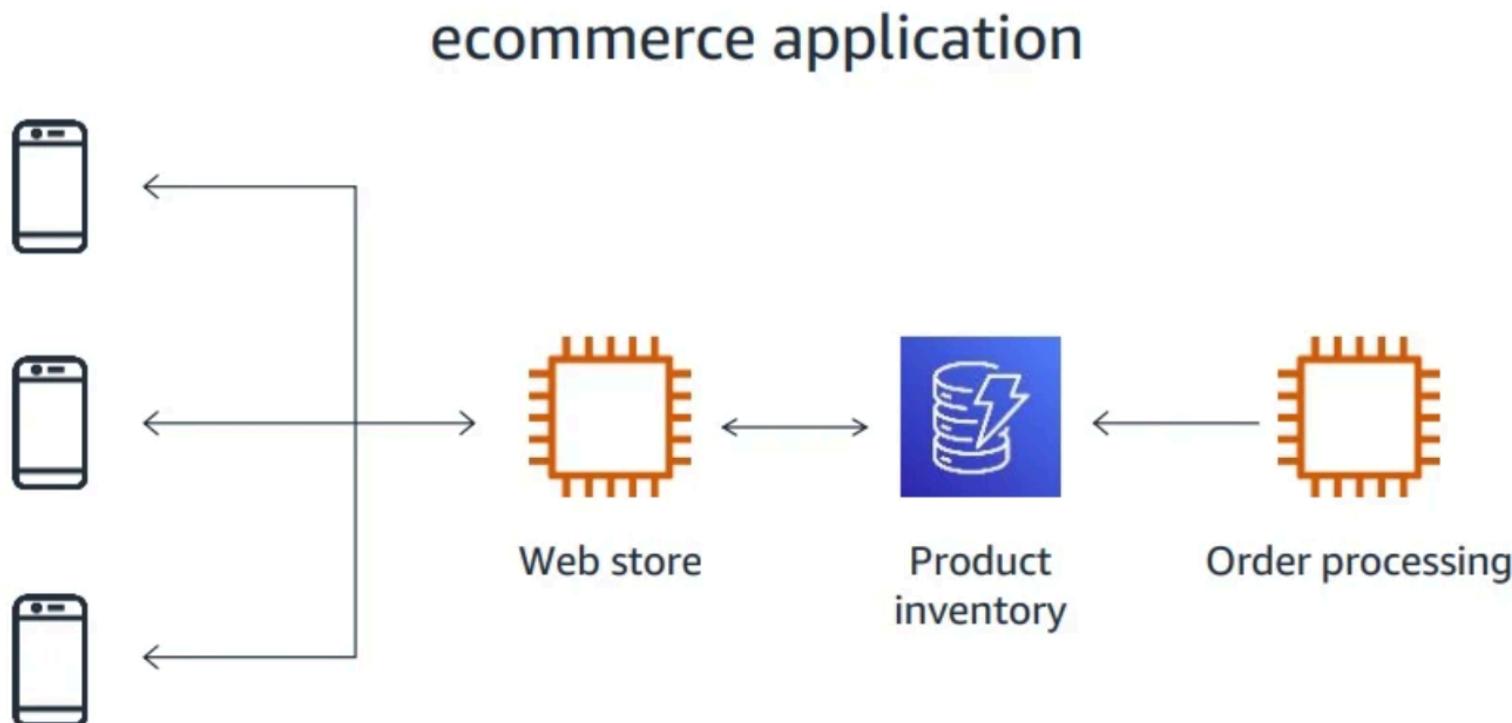
DynamoDB use case 1

Player profile page



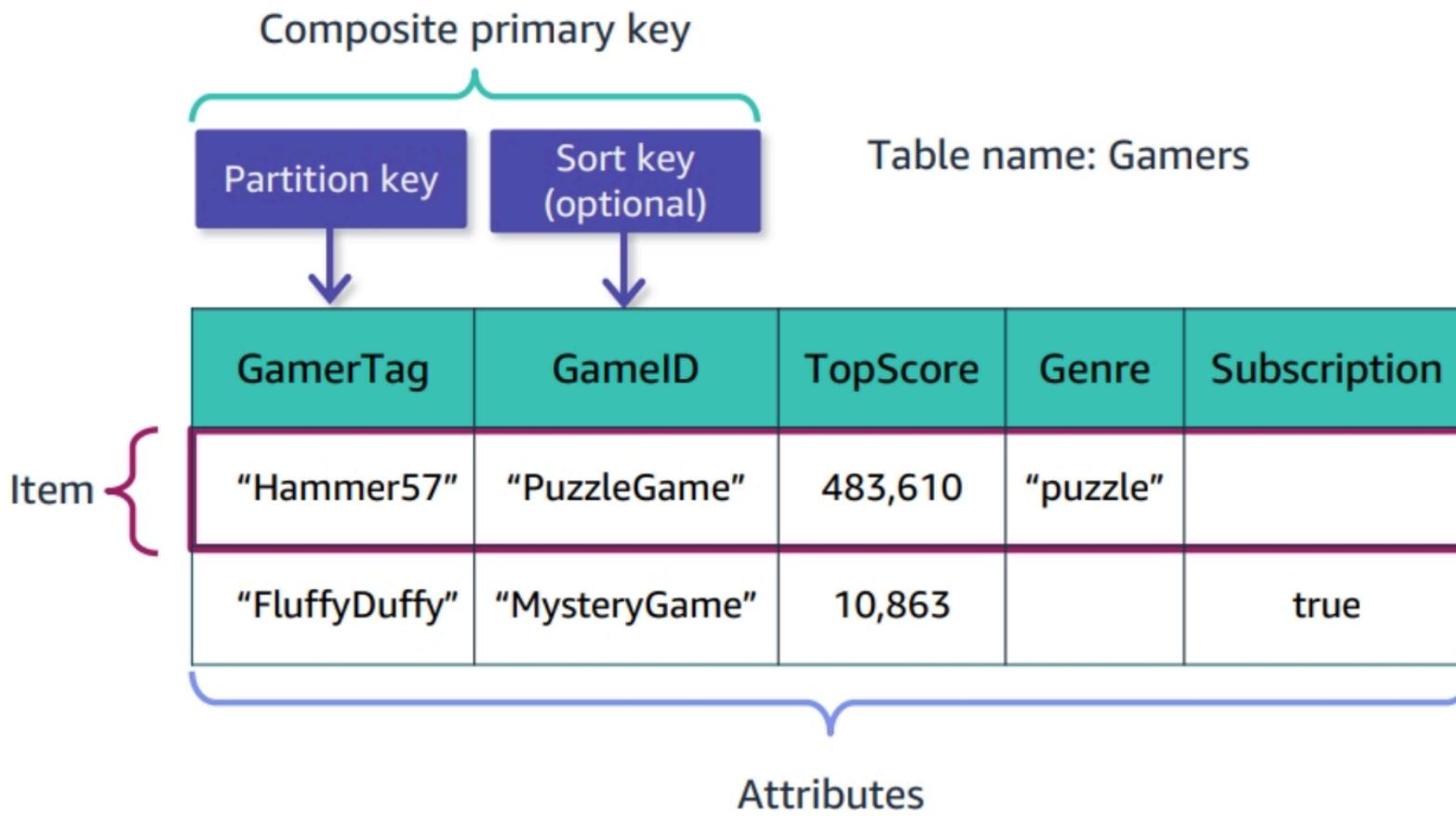
| GamerTag | UserId | TopScore | MemberSince | SubscriptionType |
|---------------|--------|----------|-----------------------|------------------|
| "Hammer57" | 101 | 5,842 | "2021-09-15:17:24:31" | "Gold" |
| "FluffyDuffy" | 243 | 1,024 | "2021-10-22:23:18:01" | "Platinum" |
| "NewPlayer" | 623 | 687 | "2021-10-22:23:22:01" | "Free" |

DynamoDB use case 2



DynamoDB tables

- Mandatory key-value access pattern
- Partition key determines data distribution
- Sort key permits rich query capabilities

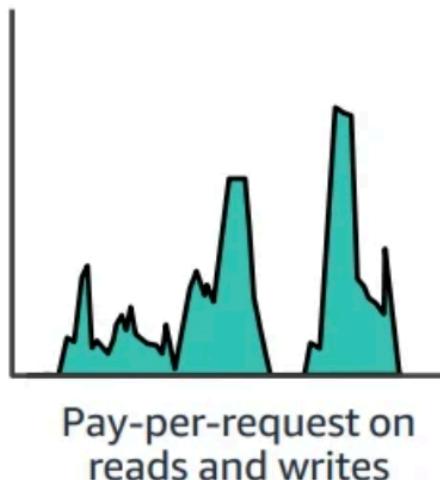


DynamoDB capacity and scaling

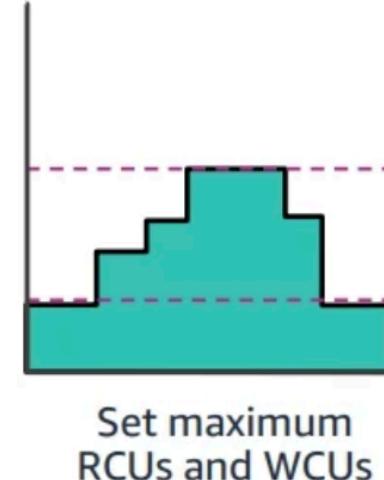
- DynamoDB measures read capacity in read capacity units (RCUs).
 - Read requests for up to a 4-KB item
- DynamoDB measures write capacity in write capacity units (WCUs).
 - Number of write requests per second for up to a 1-KB item

DynamoDB has two options for managing capacity:

On-demand



Provisioned



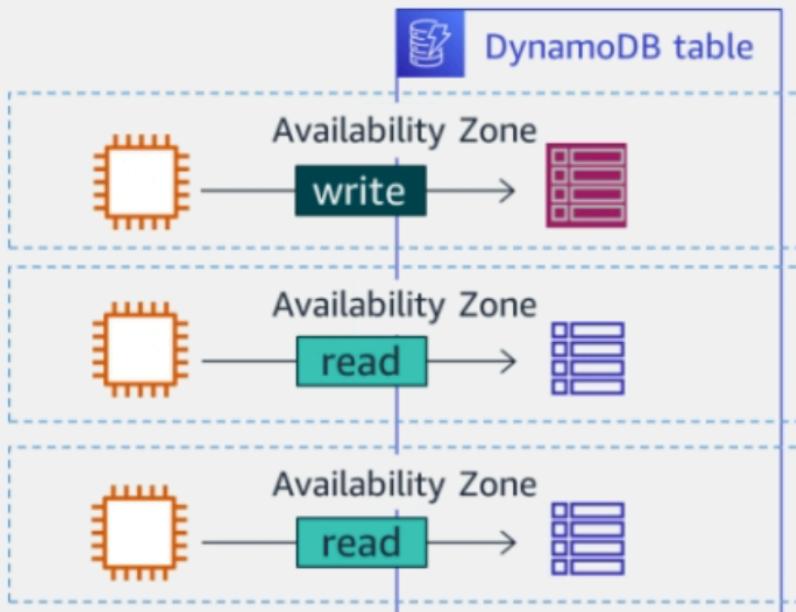
Use auto scaling to adjust your provisioned capacity to match demand

DynamoDB consistency options

DynamoDB replicates table data across three Availability Zones in a Region usually within one second.

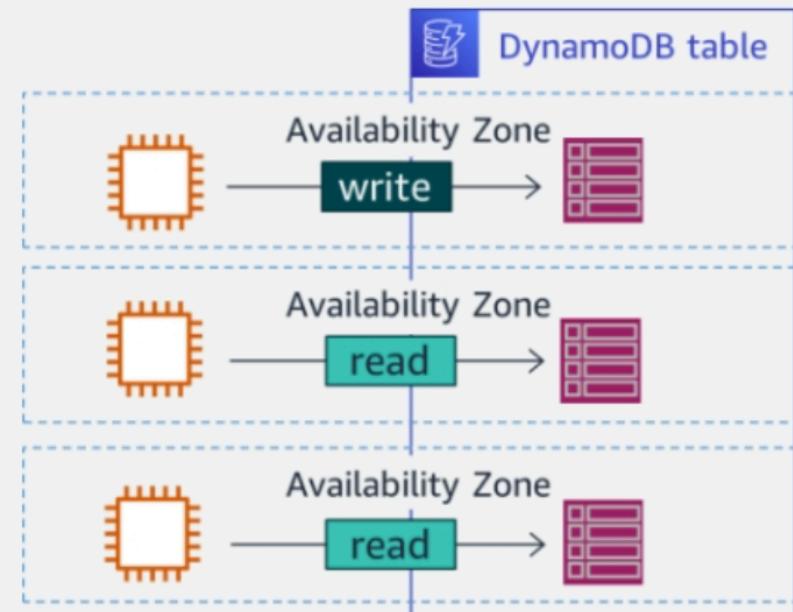
Eventually consistent read

Uses 0.5 read capacity unit



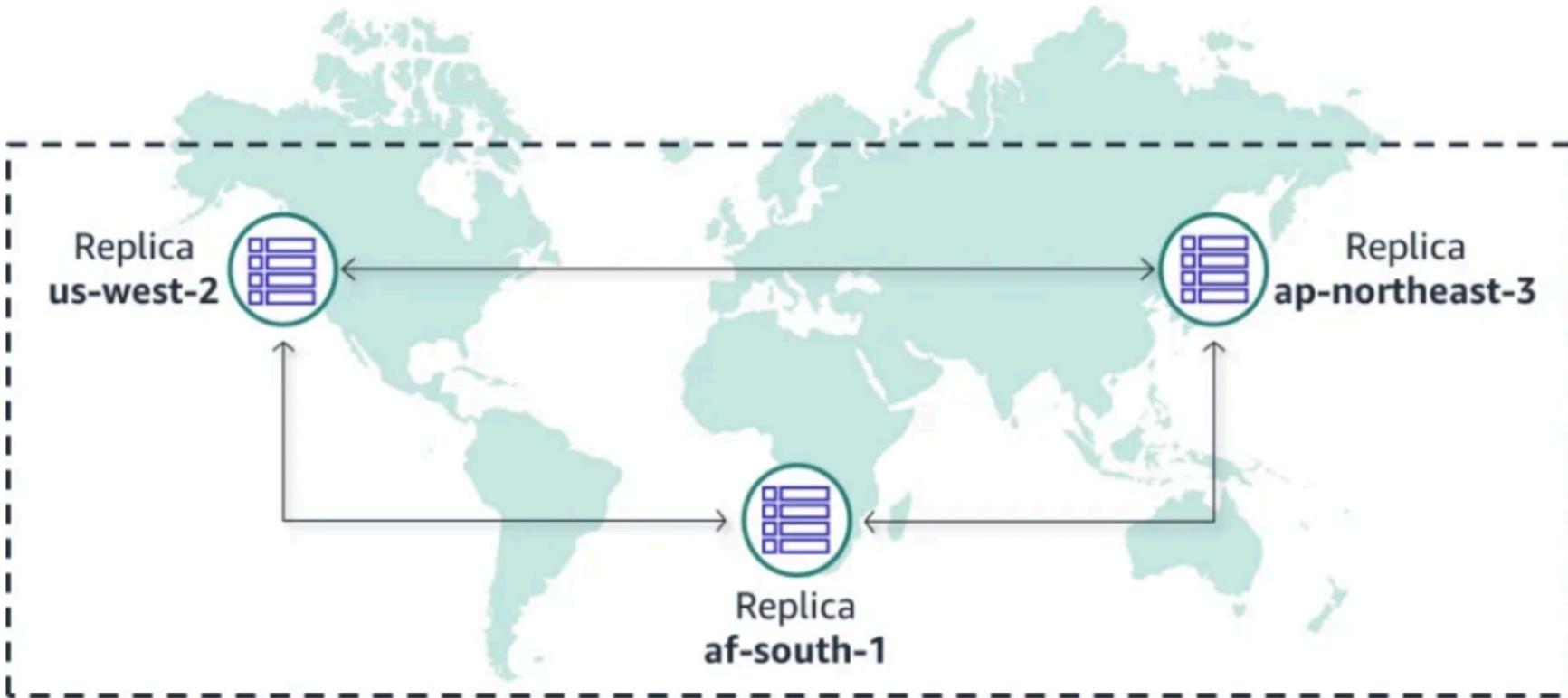
Strongly consistent read

Uses 1 read capacity unit



DynamoDB global tables

Global tables automate replication across Regions.



What should you cache?



Data that requires a slow and expensive query



Frequently accessed data

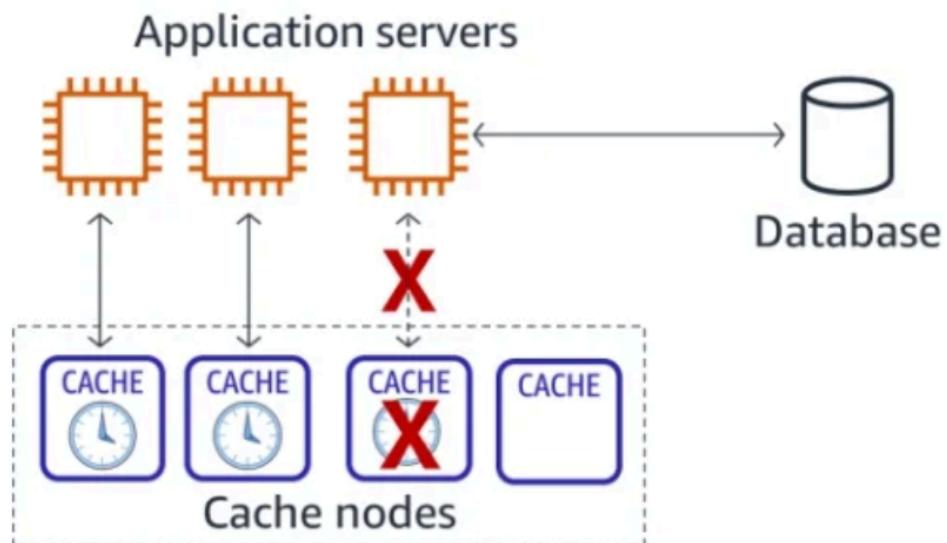


Information that is relatively static

Managing your cache

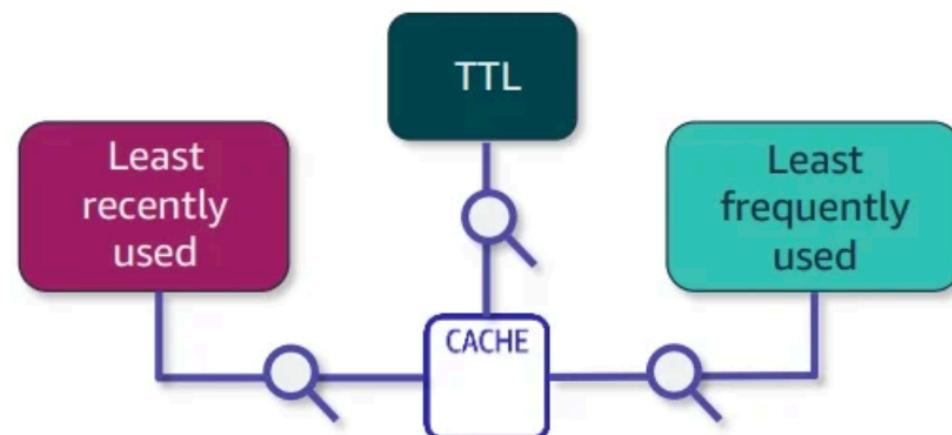
Cache validity

To minimize stale data, you can add a time to live (TTL) value to each application write.



Managing memory

When your cache memory is full, your cache evicts data based on your selected eviction policy. Eviction policies can evaluate any combination of the following:



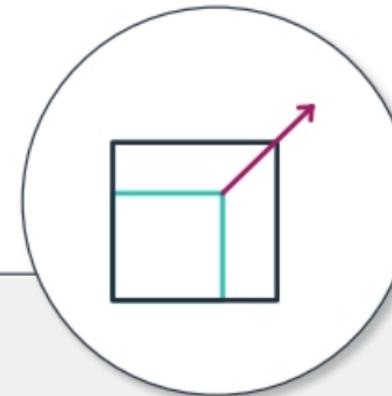
Amazon ElastiCache



Extreme performance



Fully managed



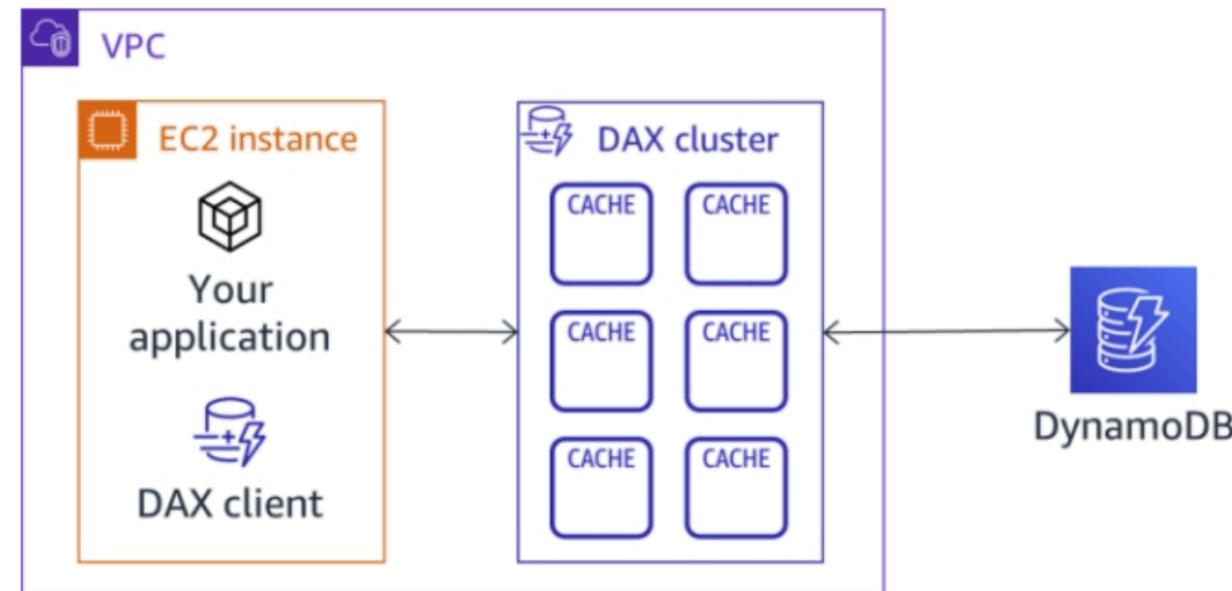
Easily scalable

ElastiCache engines

| |  ElastiCache for Memcached |  ElastiCache for Redis |
|--|---|---|
| Simple cache to offload database burden | Yes | Yes |
| Ability to scale horizontally for writes and storage | Yes | Yes (when using cluster mode) |
| Multi-AZ deployments | Yes | Yes |
| Multi-threaded performance | Yes | |
| Advanced data types | | Yes |
| Sorting and ranking datasets | | Yes |
| Publish and subscribe capability | | Yes |
| Backup and restore | | Yes |

Amazon DynamoDB Accelerator (DAX)

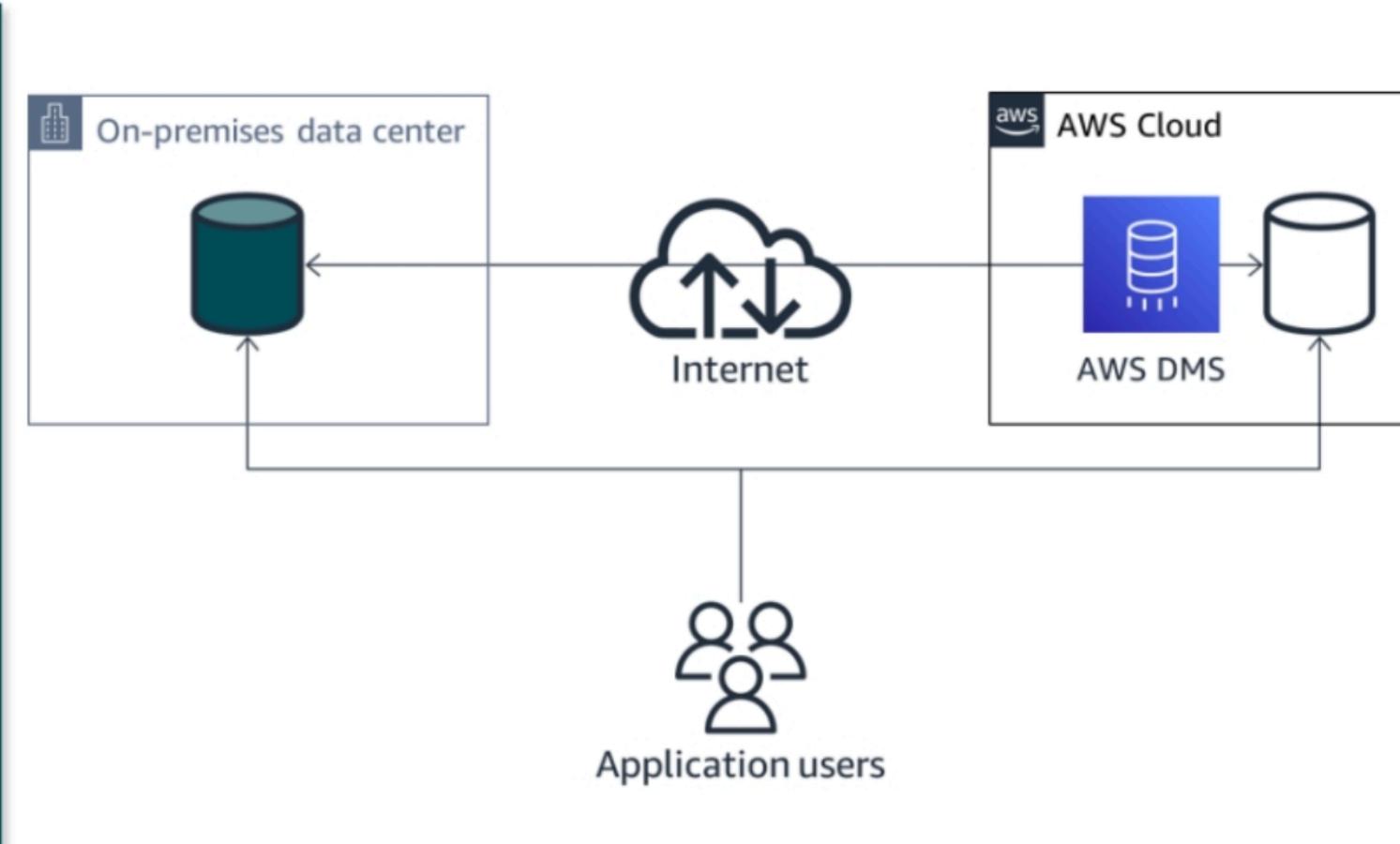
- A fully managed, highly available cache for DynamoDB
- Can deliver microsecond response times
- Can scale to millions of read requests per second



AWS Database Migration Service

AWS Database Migration Service (AWS DMS)

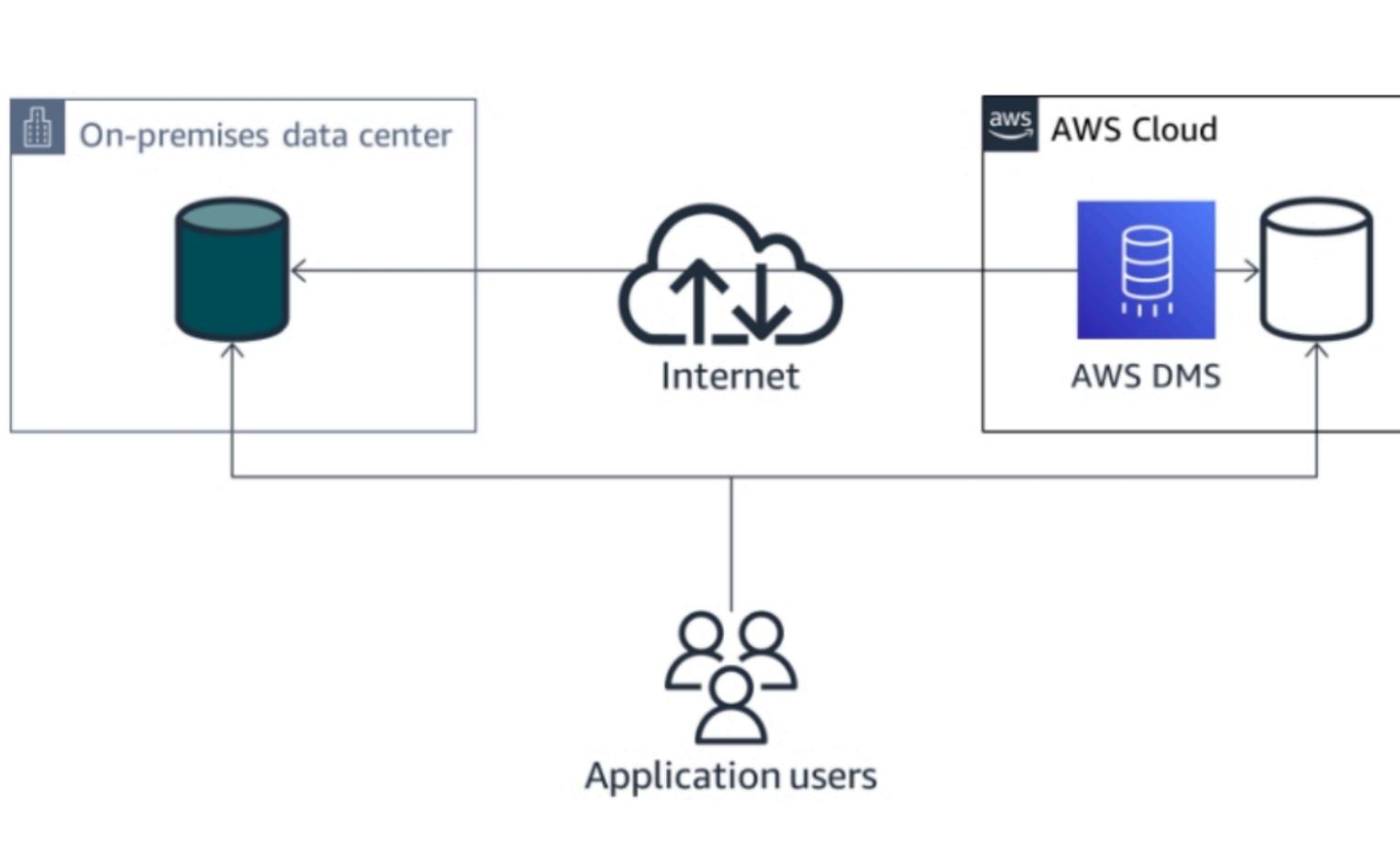
- Heterogeneous database migrations
- Database consolidation
- Continuous data replication
- Can point to a database, Amazon S3, AWS Snowball Edge, or other services



AWS Database Migration Service

AWS Database Migration Service (AWS DMS)

- Heterogeneous database migrations
- Database consolidation
- Continuous data replication
- Can point to a database, Amazon S3, AWS Snowball Edge, or other services



THANKS FOR LISTENING