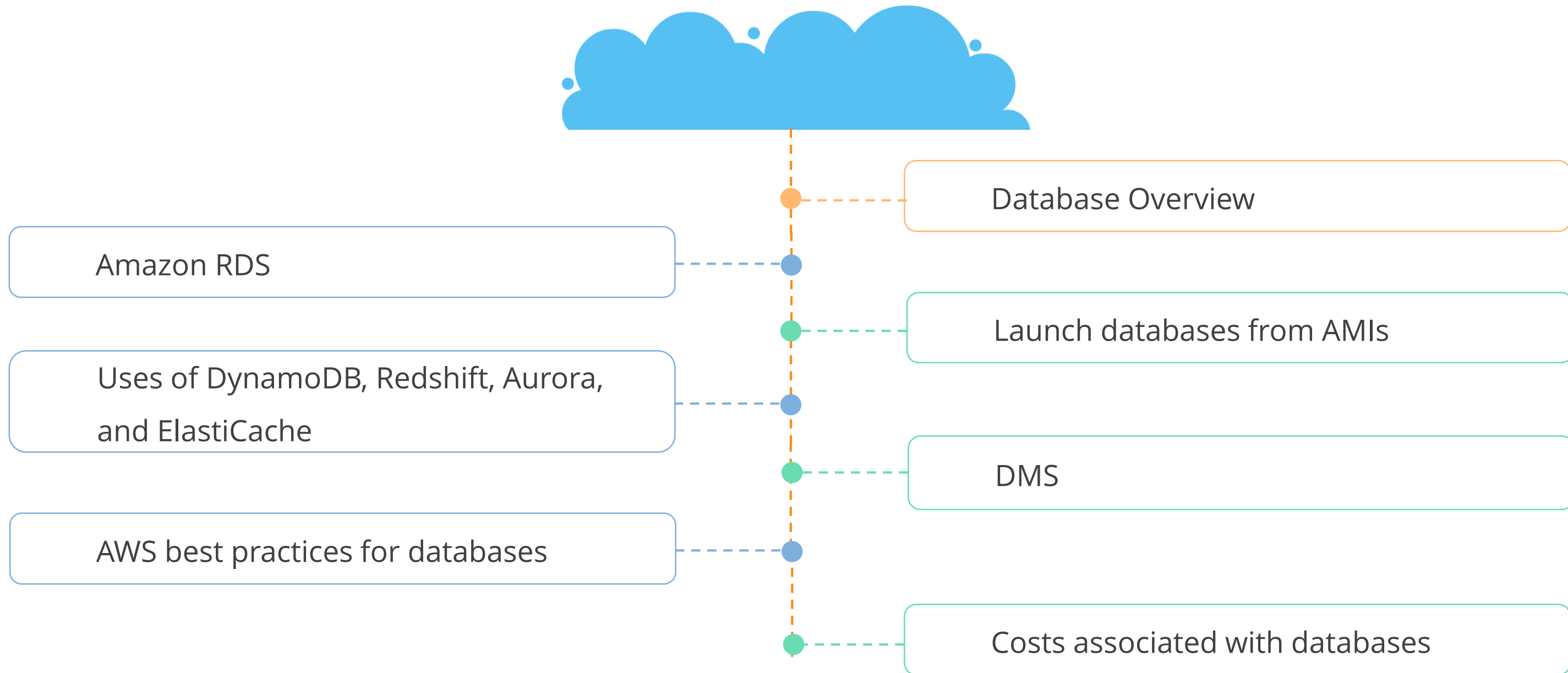


AWS Solutions Architect–Associate Level

Lesson 8: Databases



What You'll Learn



Databases Overview

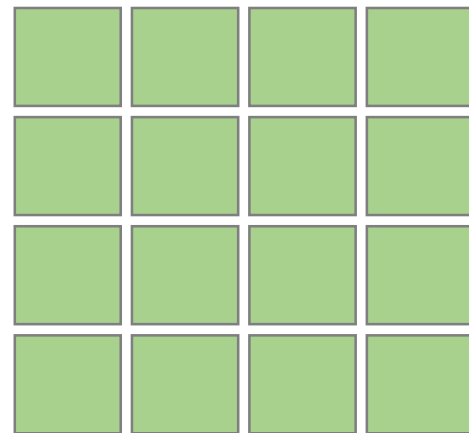
Overview of the different types of databases available within AWS

Relational Databases

The most common form of databases are relational databases or SQL databases. It is used to store structured data.

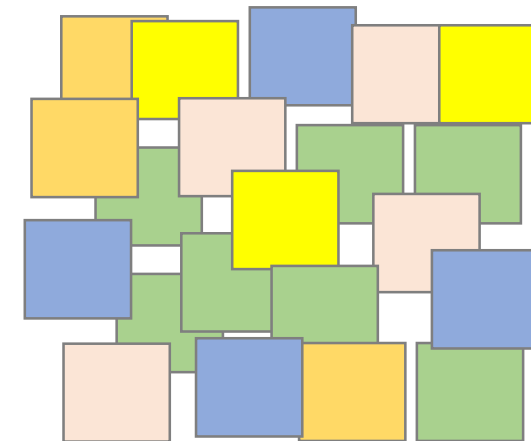
It is a collection of data items organized as a set of formally described tables. It is also known as the relational model.

Structured Data



Relational Database

Unstructured Data



Relational Databases (contd.)

Relational databases store data in rows and columns.

Columns

Rows

Employee Number	Surname	Forename	Sex	DOB	Salary	Section Name
EN1200	Gilbert	Trevor	M	20/10/1967	20,000	Salary
EN1244	Singh	Raj	M	01/11/1971	15,250	Pensions
EN1232	MacDonald	Janet	F	12/06/1966	25,000	Pensions
EN1445	Hussain	Khalid	M	23/08/1982	17,000	Stores
EN1653	Varo	Linda	F	11/03/1977	9,500	IT

OLTP

Relational databases are mainly used for OLTP (Online Transaction Processing) which involves business transactional processing using real time data.

An example of an OLTP database is a database that stores the transactions for an ATM.



ATM

Name	Time	Amount	Account No:
David	11.30 AM	500	12345678

RDBMS

The most popular relational database management systems are:



RDBMS (contd.)

AWS offers a number of relational databases.



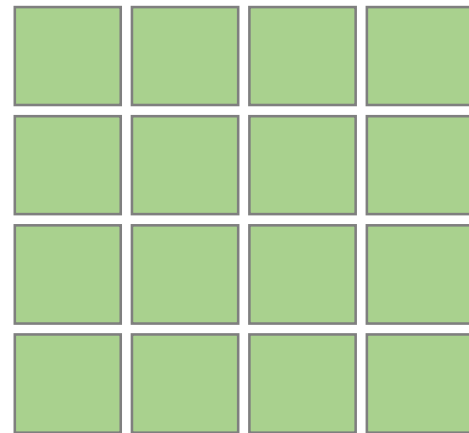
ORACLE®



Non-Relational Databases

Non-relational databases are referred to as NoSQL databases or Not Only SQL databases.

Structured Data



Relational Database

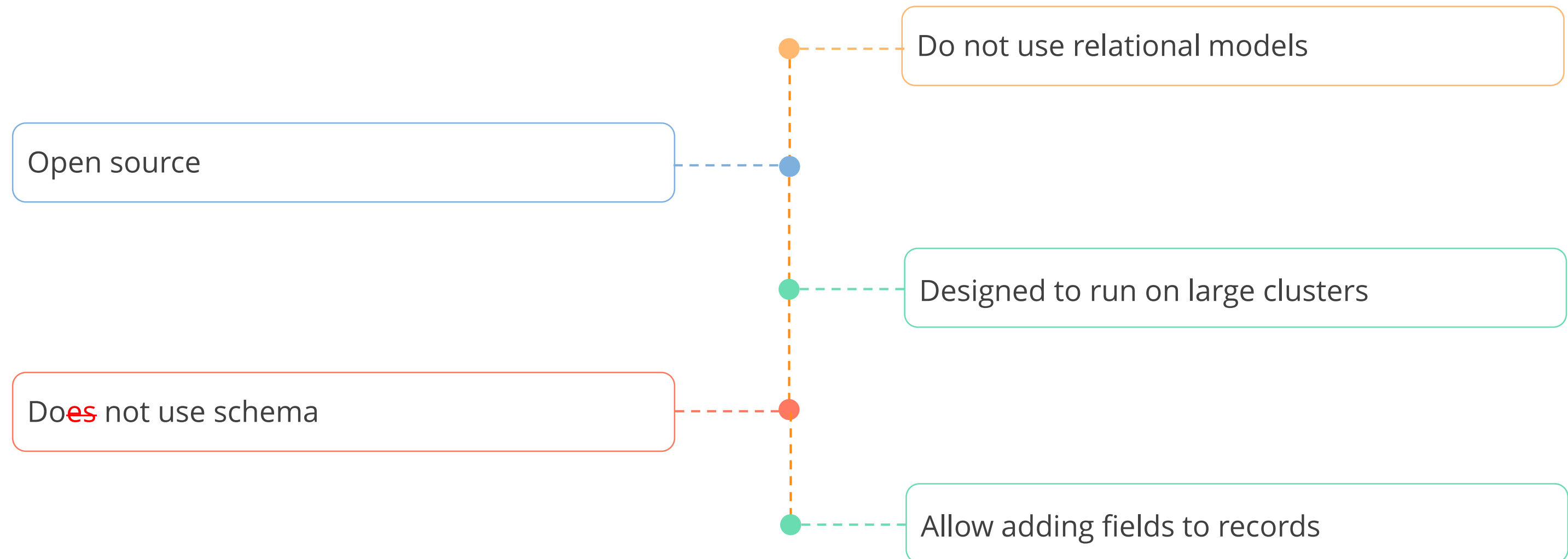
Unstructured Data



Non-relational Database

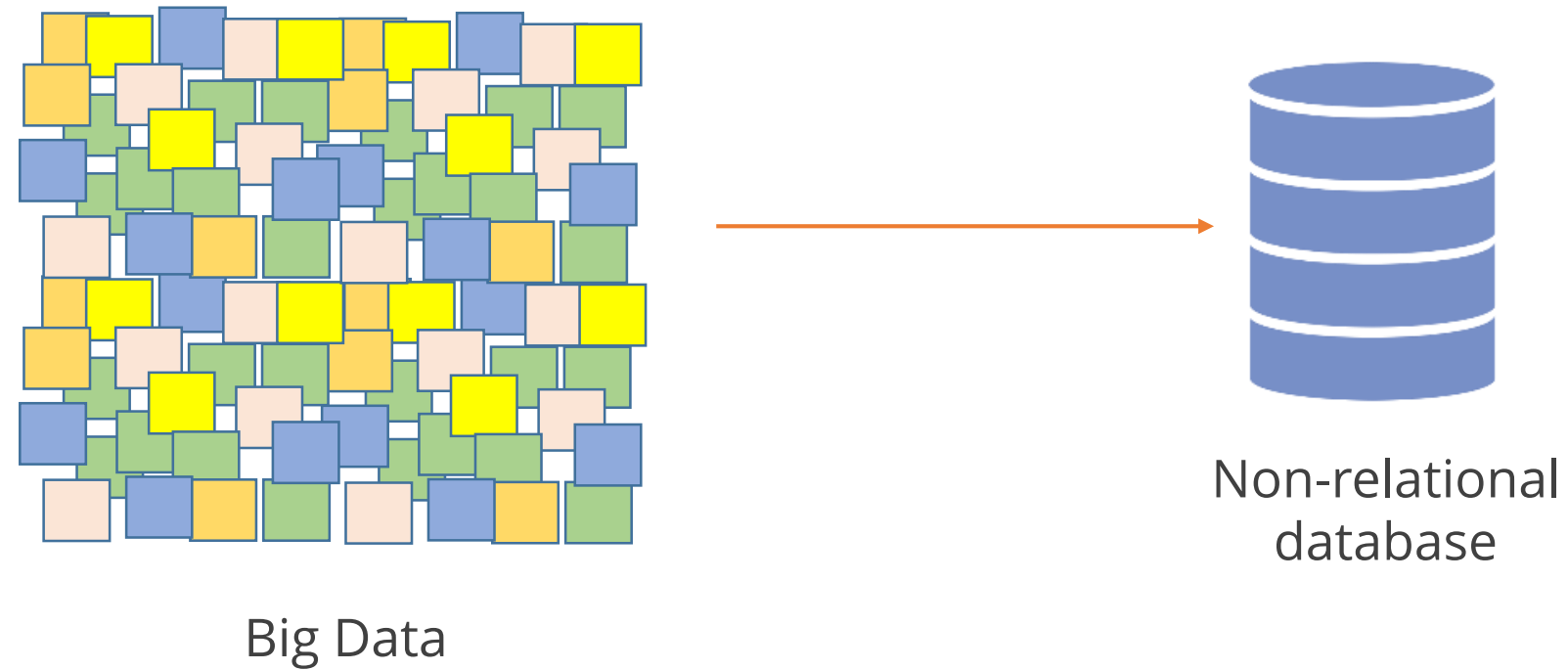
Non-Relational Databases (contd.)

Following are the features of non-relational databases:



Big Data

NoSQL databases are commonly used where huge amounts of unstructured data needs to be delivered.



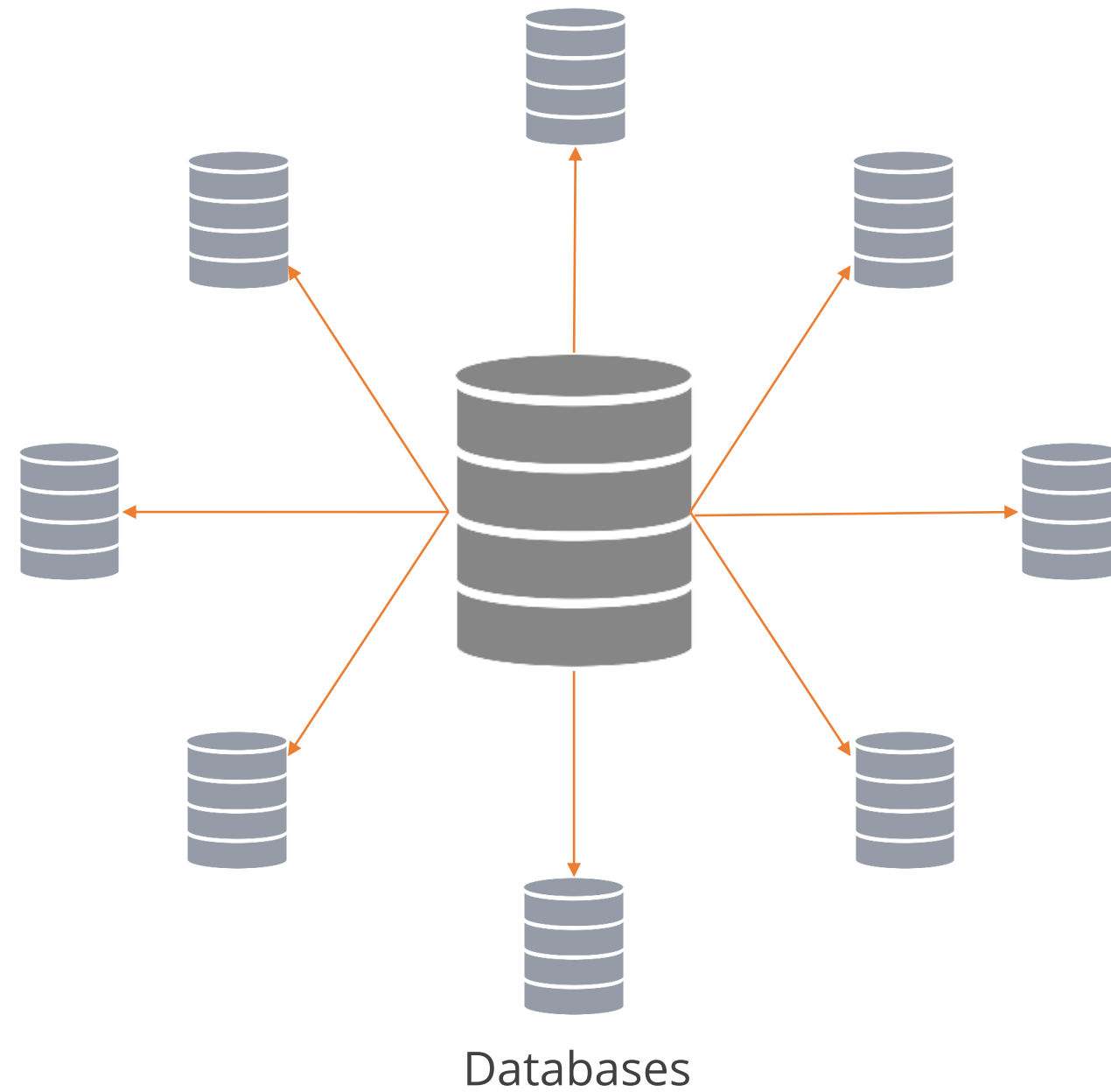
NoSQL Platforms

The most popular NoSQL databases are Cassandra, MongoDB, CouchDB, and so on.



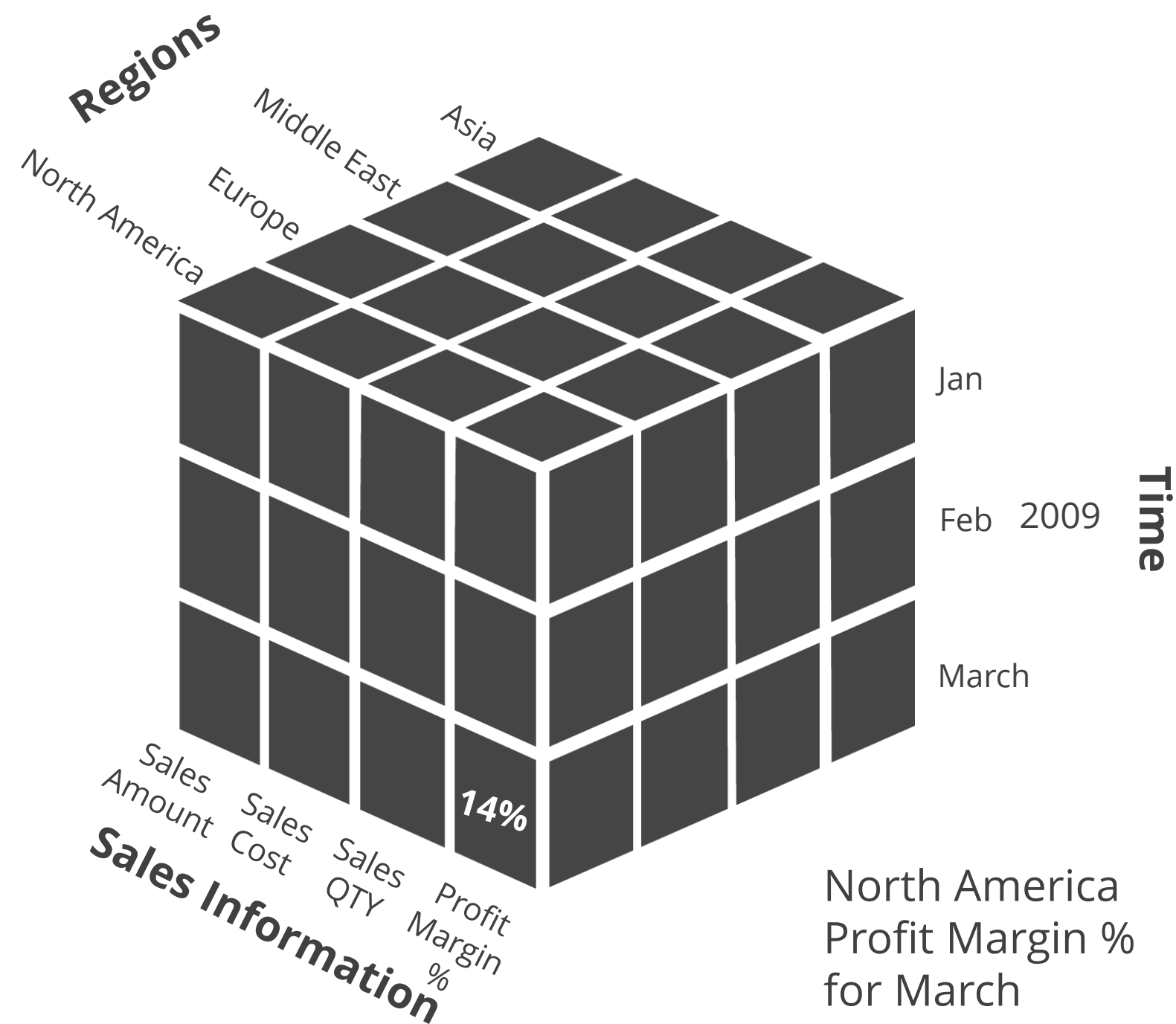
Data Warehousing

A data warehouse stores data from multiple sources for analytical purposes.



OLAP

Data warehouses are designed for OLAP (Online Analytics Processing), which is used to integrate copies of transactional data from other systems and use it for analytical purposes.



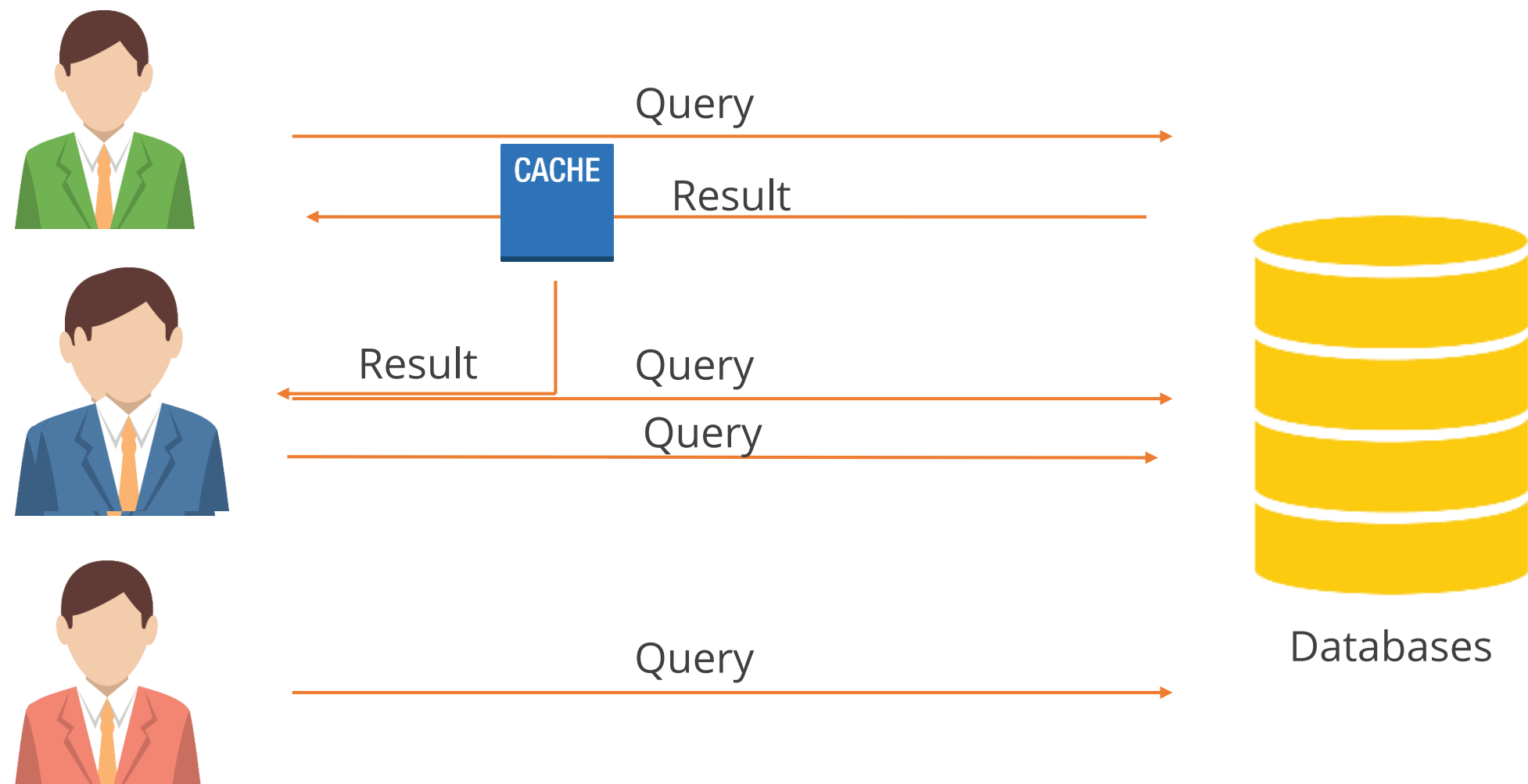
Data Warehouse Platforms

The most popular data warehouses are Cognos, SQL Server Reporting Services, and Oracle Hyperion.



In-Memory Databases

In-memory databases are cached based databases that store results in memory to reduce the load on your database infrastructure and to improve user response time.



In-Memory Database Platforms

The most popular cache based databases are MEMCACHED and Redis.
AWS offers its own version called ElastiCache.



ElastiCache



Knowledge Check

KNOWLEDGE
CHECK

Which are the four types of database platforms in AWS?

- a. Relational, Non-relational, Data warehouse, and In-memory
- b. Structured, Non structured, Data storehouse, and Cached
- c. Relational, Non-relational, Data storehouse, and Cached
- d. Structured, Non structured, Data Warehouse, and In-memory



KNOWLEDGE
CHECK

Which are the four types of database platforms in AWS?

- a. Relational, Non-relational, Data warehouse, and In-memory
- b. Structured, Non structured, Data storehouse, and Cached
- c. Relational, Non-relational, Data storehouse, and Cached
- d. Structured, Non structured, Data warehouse, and In-memory



The correct answer is **a)**

The four types of database platforms available on AWS are Relational, Non-relational, Data warehouse, and In-memory.

Amazon Relational Database Service (RDS)

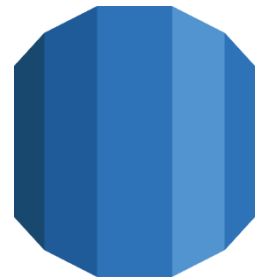
Details about Amazon RDS

Amazon RDS

Amazon Relational Database Service (RDS) is a managed service that allows you to run databases in the AWS cloud on EC2 instances, without having to worry about database administration management tasks.

Amazon RDS Platforms

Amazon RDS is available in Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB.



Aurora DB



ORACLE®



Amazon RDS



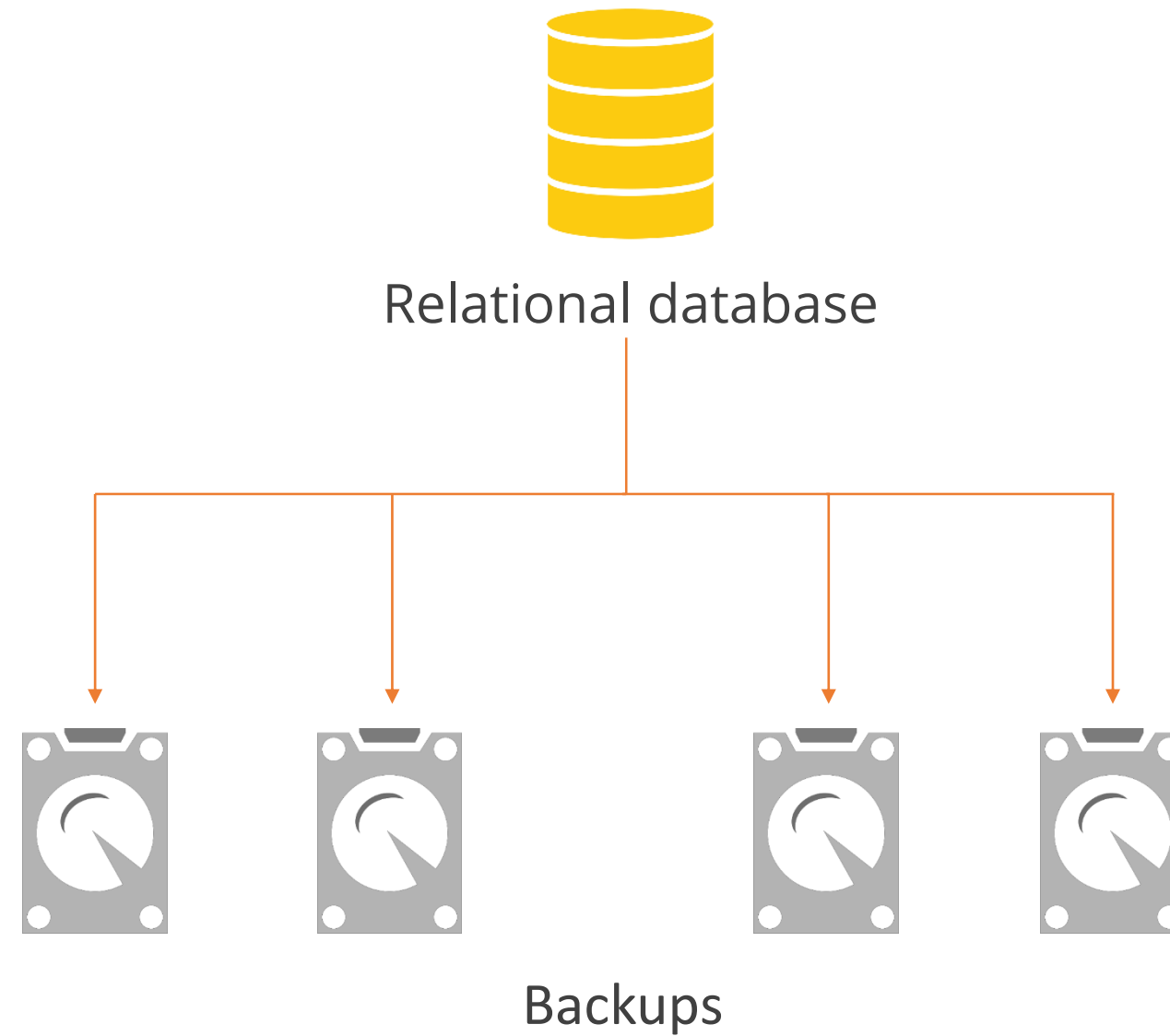
Preconfigured with Suitable Settings

RDS instances are preconfigured with a setting suitable for the type of EC2 instance you select. There is no database installation process required.



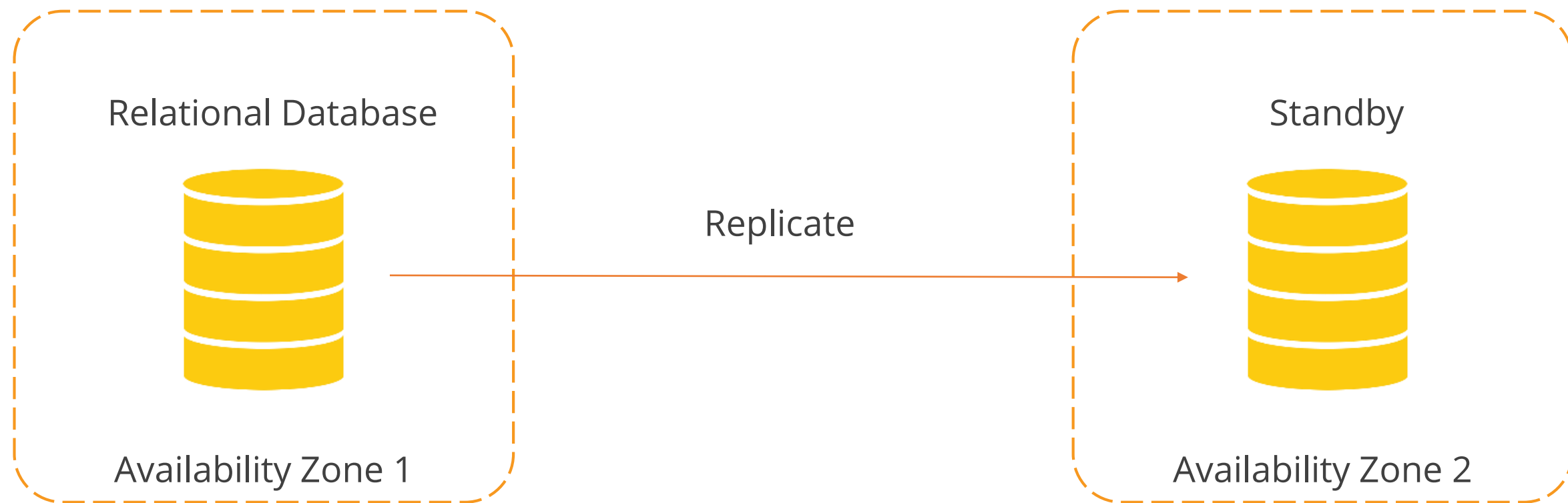
Automated Backups

Automated backups are enabled by default. It retains both database and transaction log backups for up to 35 days for point-in-time recovery.



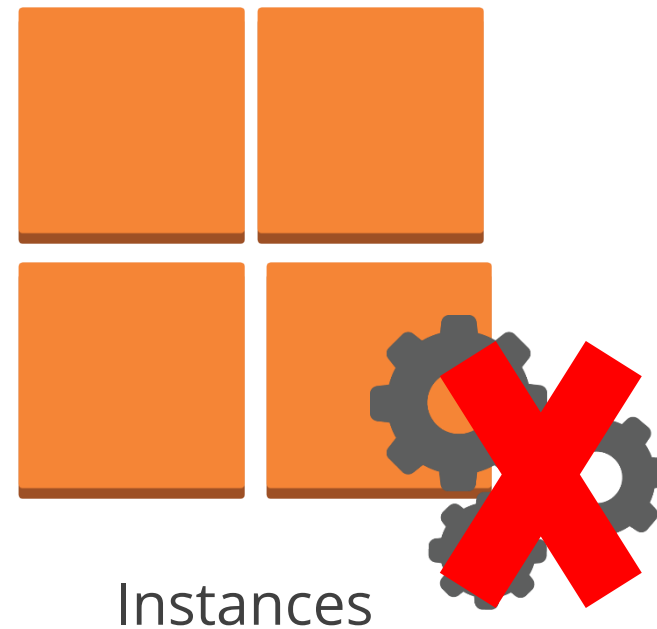
Multi-Availability Zone Deployments

Multi-Availability Zone deployments synchronously replicate the data to a standby instance in a different Availability Zone.



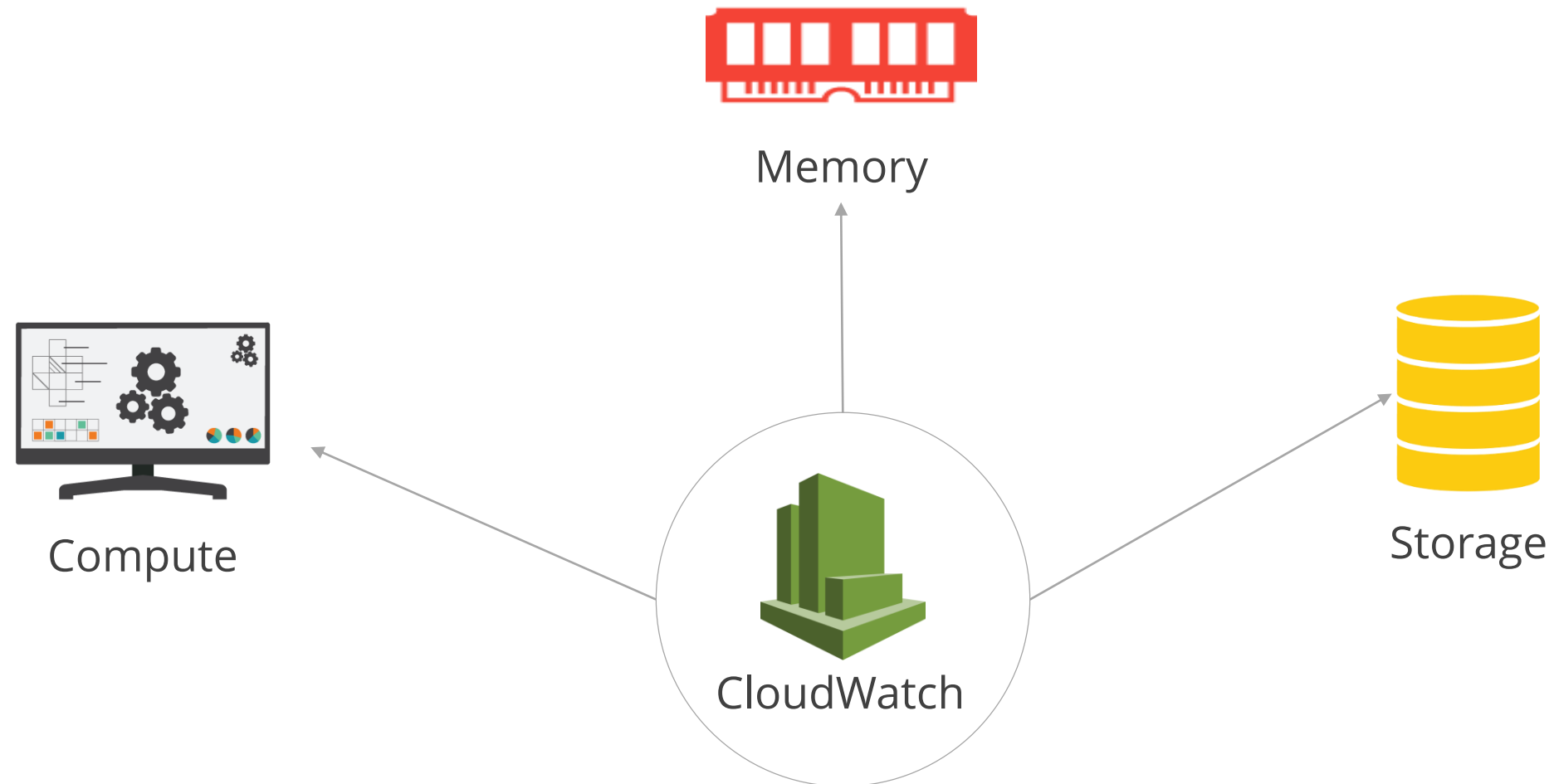
Automatic Hardware Replacement

If a database instance fails due to hardware failure, RDS automatically replaces the instance with no manual interaction required.



Monitoring and Metrics

Monitoring and metrics are available using CloudWatch at no extra charge; so you can monitor compute, memory, and storage.



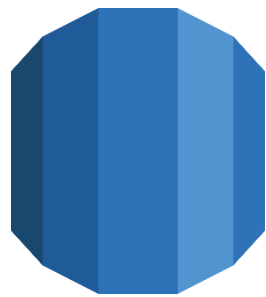
Monitoring and Metrics (contd.)

RDS offers automatic software patching to ensure that database software stays up-to-date without any manual intervention.



Scalability

RDS instances can be scaled up or down in just a few minutes depending on your requirements. The maximum size you can scale up is 32 virtual CPUs and 244 GB of RAM.



RDS Instance



RDS Instance



RDS Instance

32 virtual CPUs
and
244GB of RAM

Storage Increase

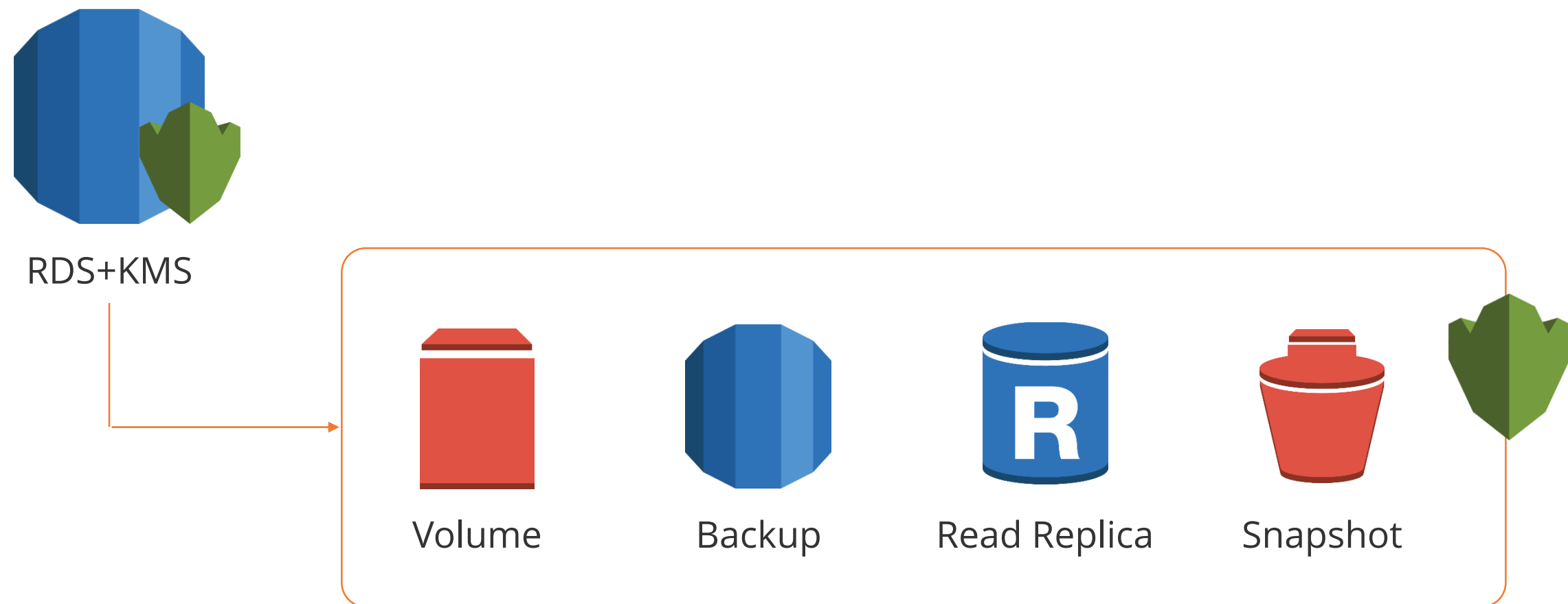
Some RDS instances allow you to increase the storage with zero downtime. Amazon Aurora automatically extends the size of the database per requirement.



RDS Instance

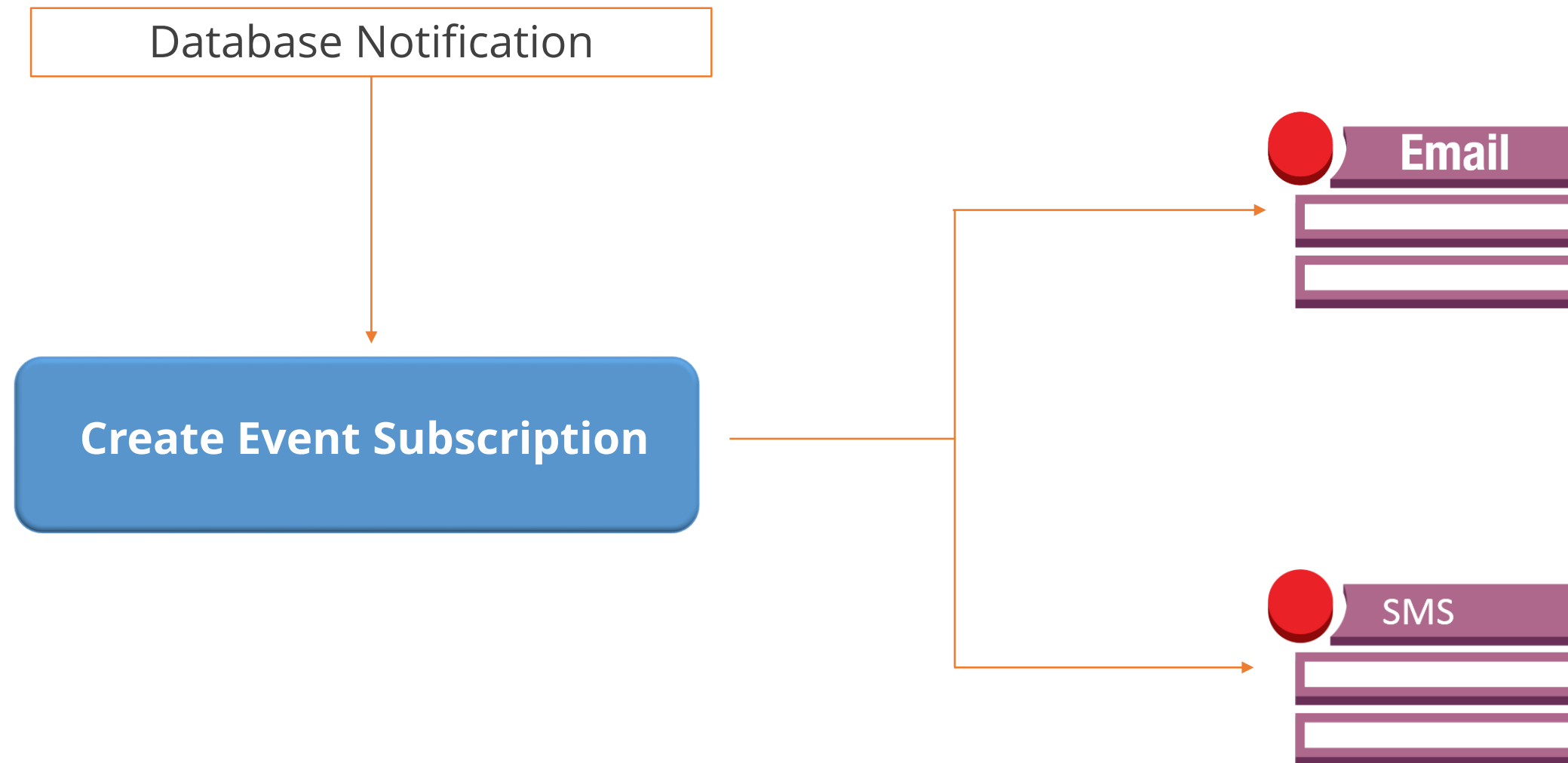
Encryption

You can launch your RDS instance to encrypt your data at rest and in transit using AWS Key Management Service (KMS).



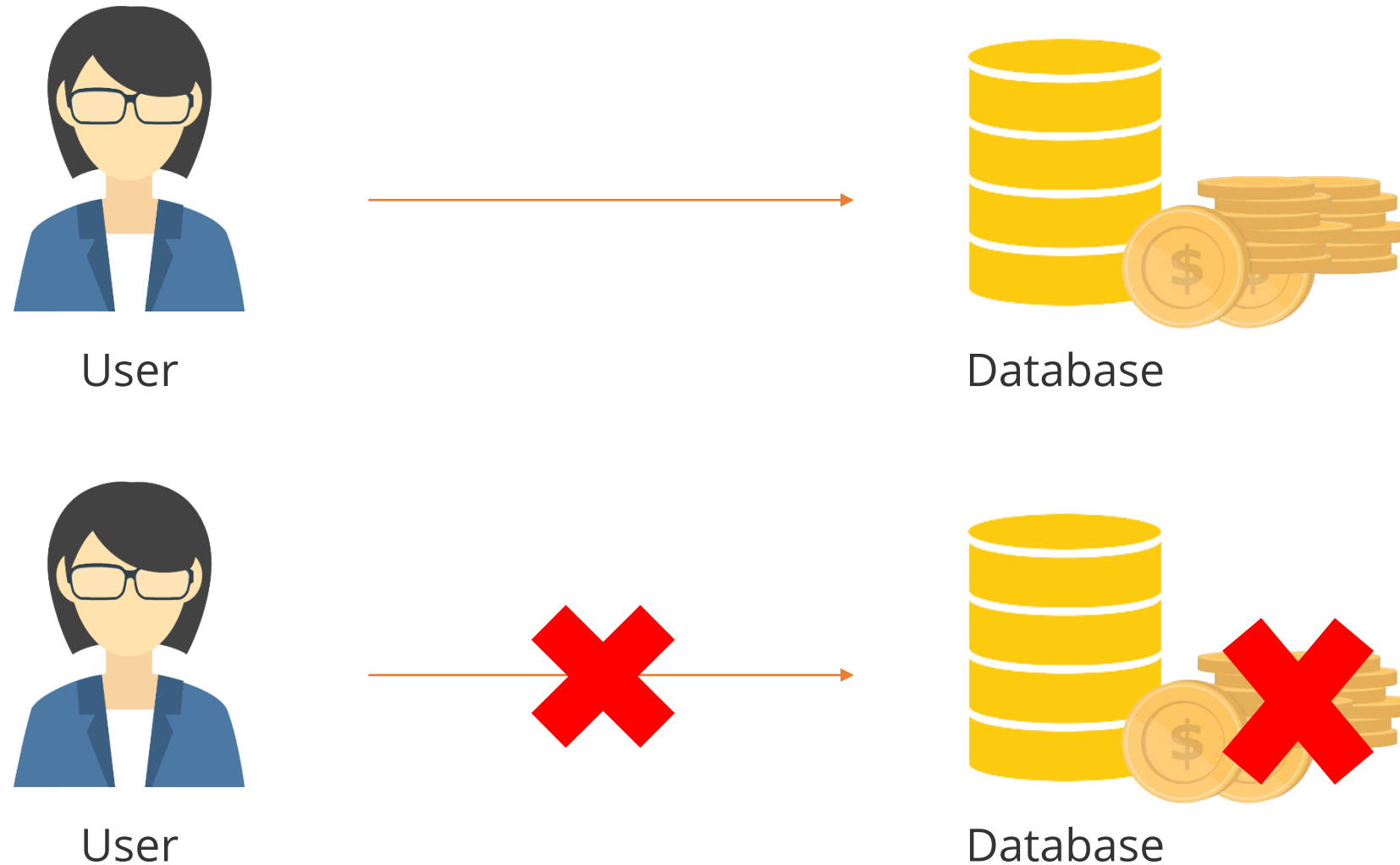
Event Notifications

Database Event Notifications can be configured so you are informed via email or SMS about database events.



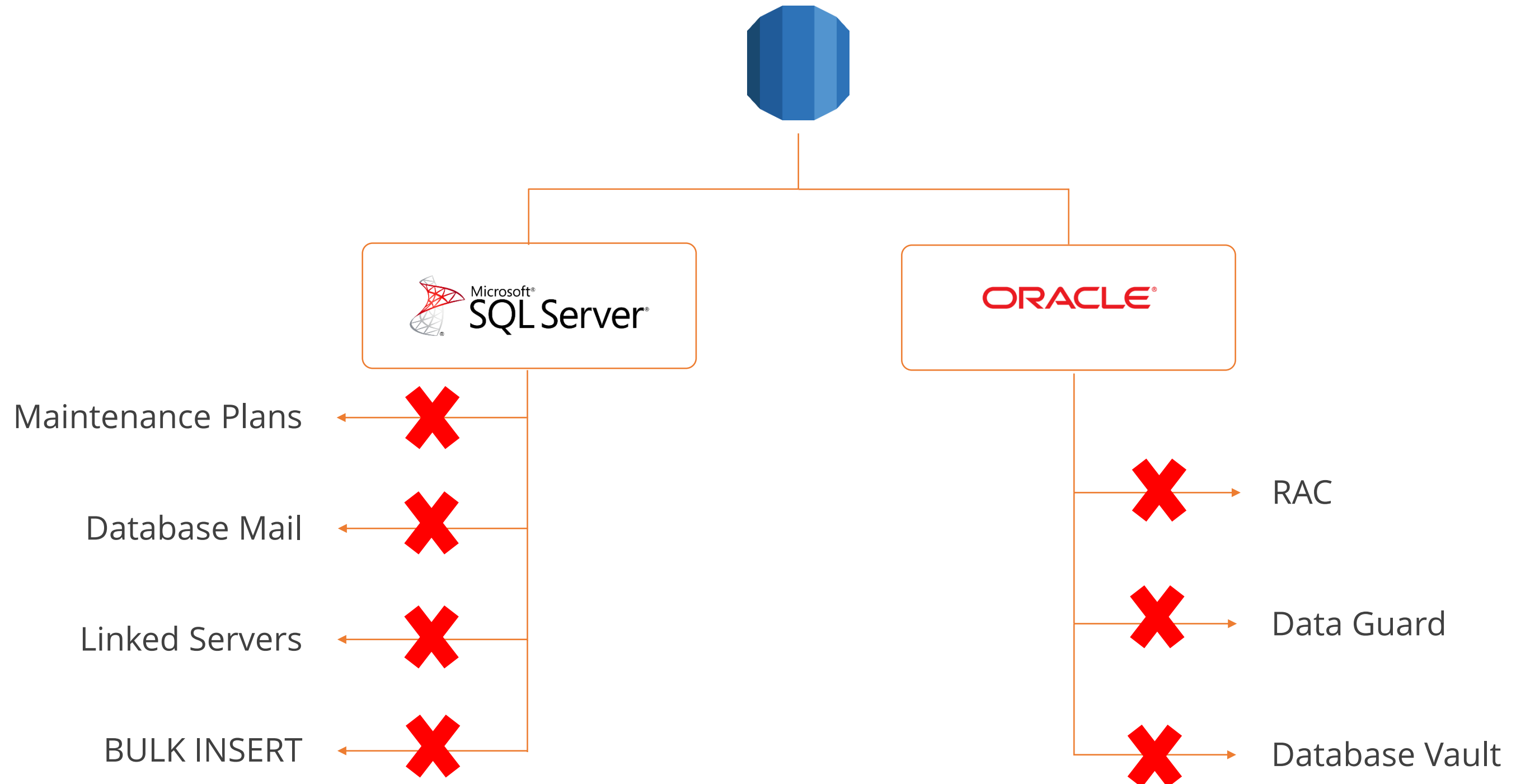
Cost Effective

As with most AWS products, you only pay for what you use. When you stop or terminate the instance, you don't have to pay.



RDS Limitations

RDS has a few limitations. For AWS to be able to provide a fully managed solution, some of the functionalities have to be removed.



RDS Limits

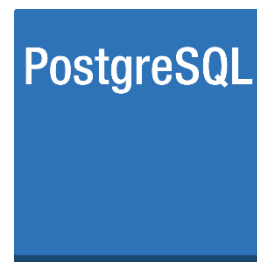
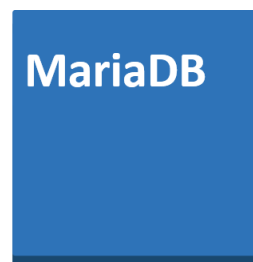
You cannot use Secure Socket Shell (SSH) or Remote Desktop Protocol (RDP) with RDS instances.



Maximum RDS Volume size=6 Terabytes



SQL Server maximum RDS Volume size=4 Terabytes



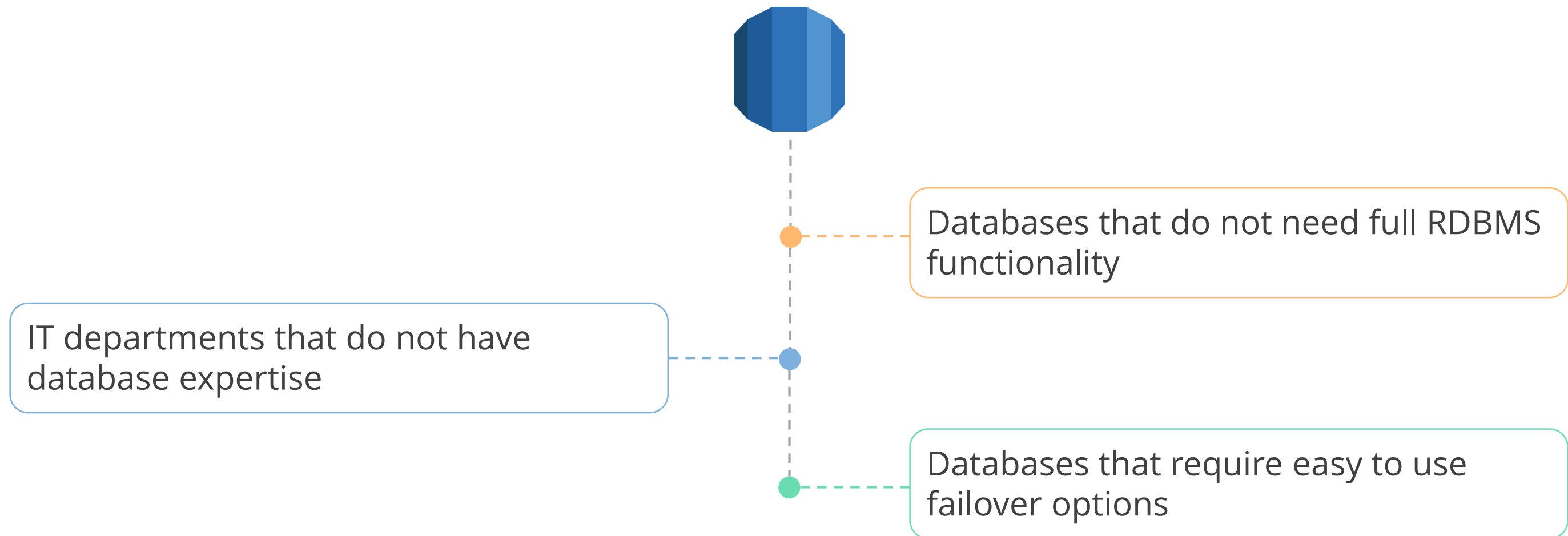
30000 IOPS



20000 IOPS

Uses of RDS

You can use RDS in the following scenarios:





Demo 1: Amazon RDS

Demonstrate how to launch an RDS instance



Knowledge Check

KNOWLEDGE
CHECK

Which of the following platforms are available on RDS?

- a. Amazon Aurora, SAP ASE, DynamoDB, RedShift, MySQL, and MariaDB
- b. Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB
- c. Amazon Aurora, Oracle, Microsoft SQL Server, RedShift, MySQL, and ElastiCache
- d. Amazon Aurora, Oracle, DynamoDB, PostgreSQL, MySQL, and ElastiCache



KNOWLEDGE
CHECK

Which of the following platforms are available on RDS?

- a. Amazon Aurora, SAP ASE, DynamoDB, RedShift, MySQL, and MariaDB
- b. Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB
- c. Amazon Aurora, Oracle, Microsoft SQL Server, RedShift, MySQL, and ElastiCache
- d. Amazon Aurora, Oracle, DynamoDB, PostgreSQL, MySQL, and ElastiCache



The correct answer is **b)**

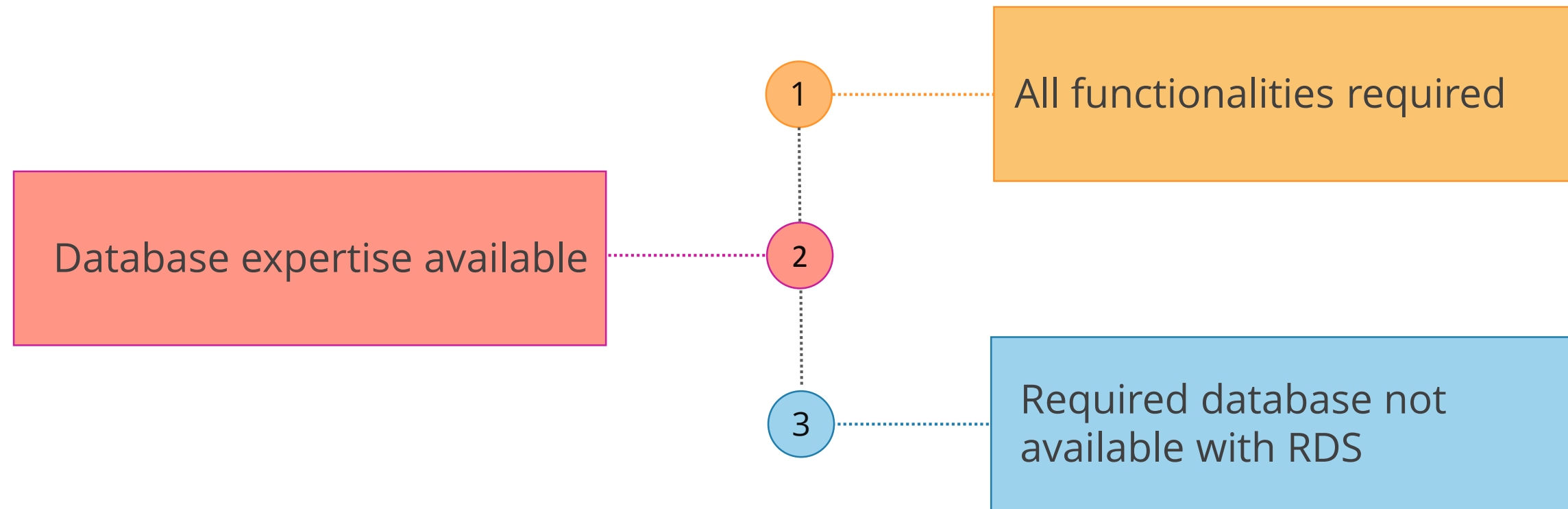
Amazon RDS is available for Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB.

AMI Databases

Details of AMI database installation

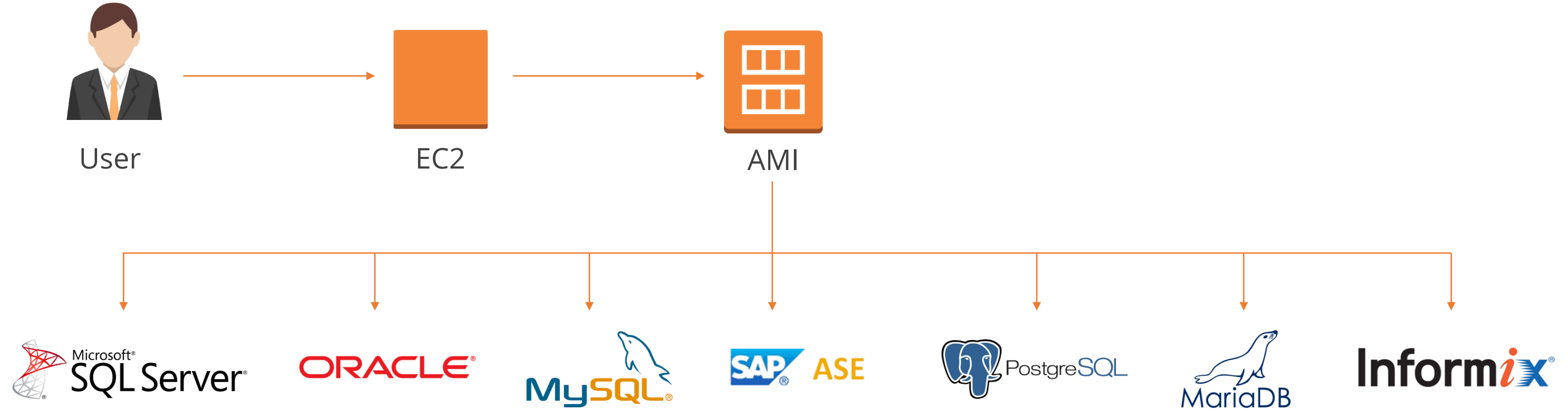
Alternatives to RDS

In the following scenarios you might require an alternative to RDS:



AMI Databases

When you launch an EC2 instance, there are a number of AMIs with preinstalled database.



Cost and Licenses

You are charged for the cost of the EC2 instance plus a license fee for the database software based on an hourly basis.



EC2



License



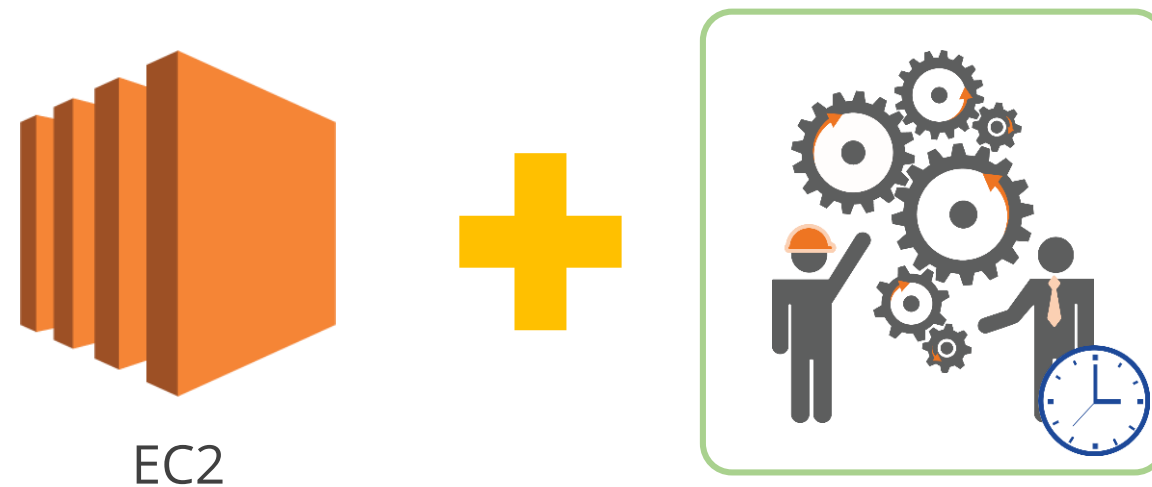
Free Database Platform



Charged Database Platform

Manual Installation

Your alternative to AMIs is a manual installation. You can launch the EC2 instance type of your choice and then install the database software yourself.





Demo 2: Launching databases from an AMI

Demonstrate how to launch a database instance using an AMI



Knowledge Check

KNOWLEDGE
CHECK

Why would you choose NOT to use RDS?

- a. To avoid the 12-36 month commitment AWS demands
- b. RDS performance is significantly slower than installing your own RDBMS
- c. Because certain RDBMS functionalities are not available on RDS
- d. RDS instances take too long to launch



KNOWLEDGE
CHECK

Why would you choose NOT to use RDS?

- a. To avoid the 12-36 month commitment AWS demands
- b. RDS performance is significantly slower than installing your own RDBMS
- c. Because certain RDBMS functionalities are not available on RDS
- d. RDS instances take too long to launch



The correct answer is **c)**

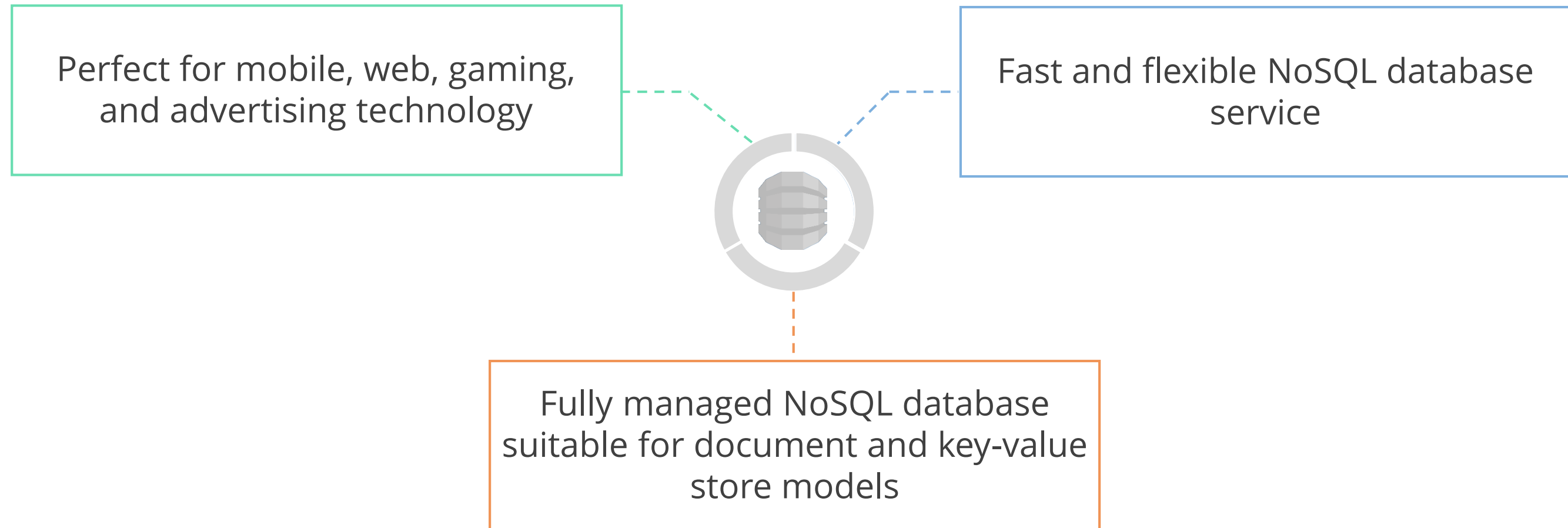
RDS has to remove certain functionalities from the RDBMS to provide a fully managed solution. If you need a particular functionality that is removed, then RDS is not suitable for you.

Amazon DynamoDB

Overview of Amazon DynamoDB

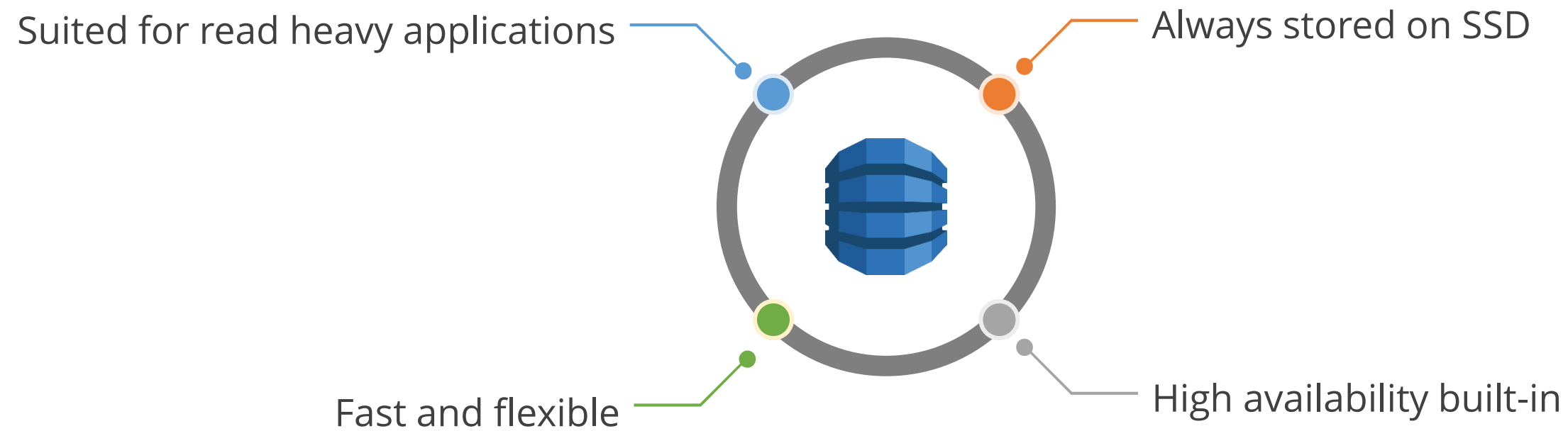
Amazon DynamoDB

Following are the features of Amazon DynamoDB:



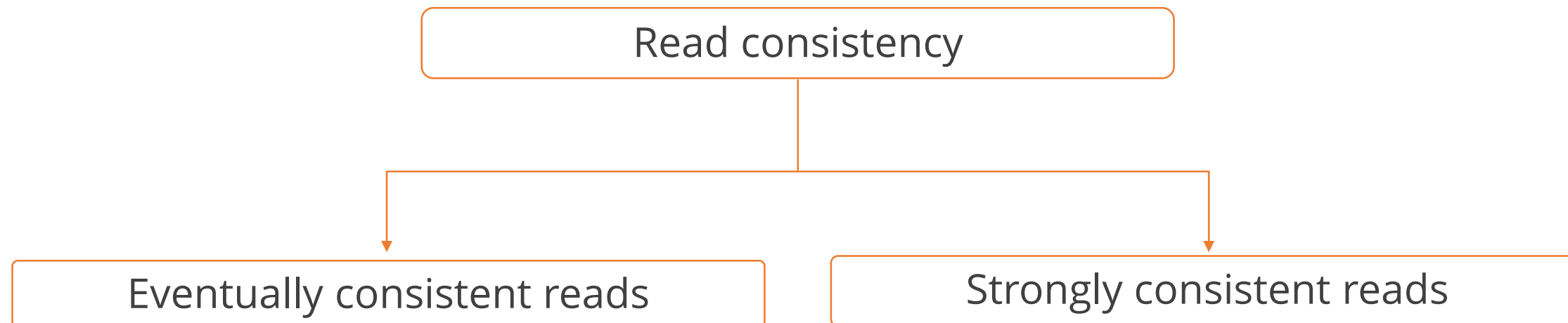
Amazon DynamoDB Benefits

Following are the benefits of Amazon DynamoDB:



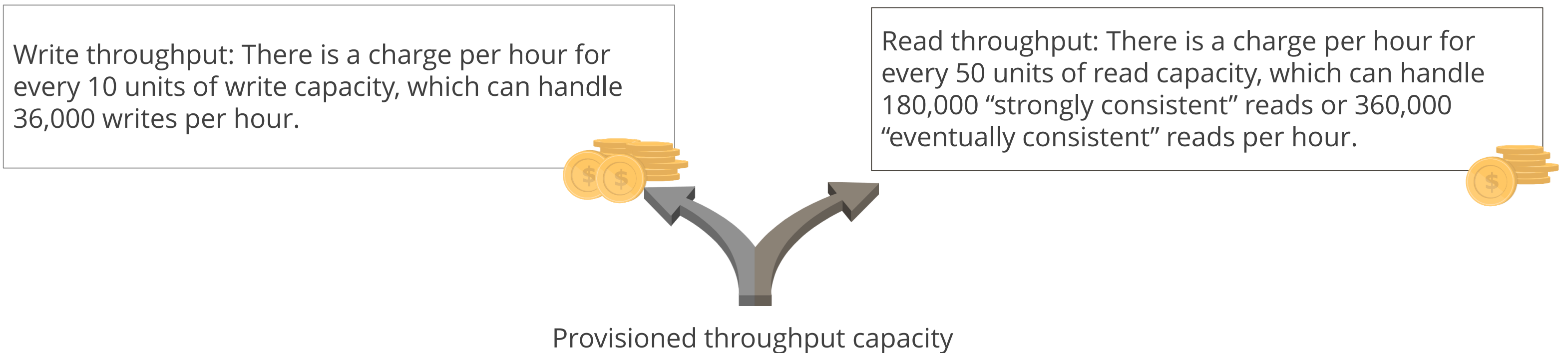
Read Consistency

DynamoDB offers two different read consistency types:



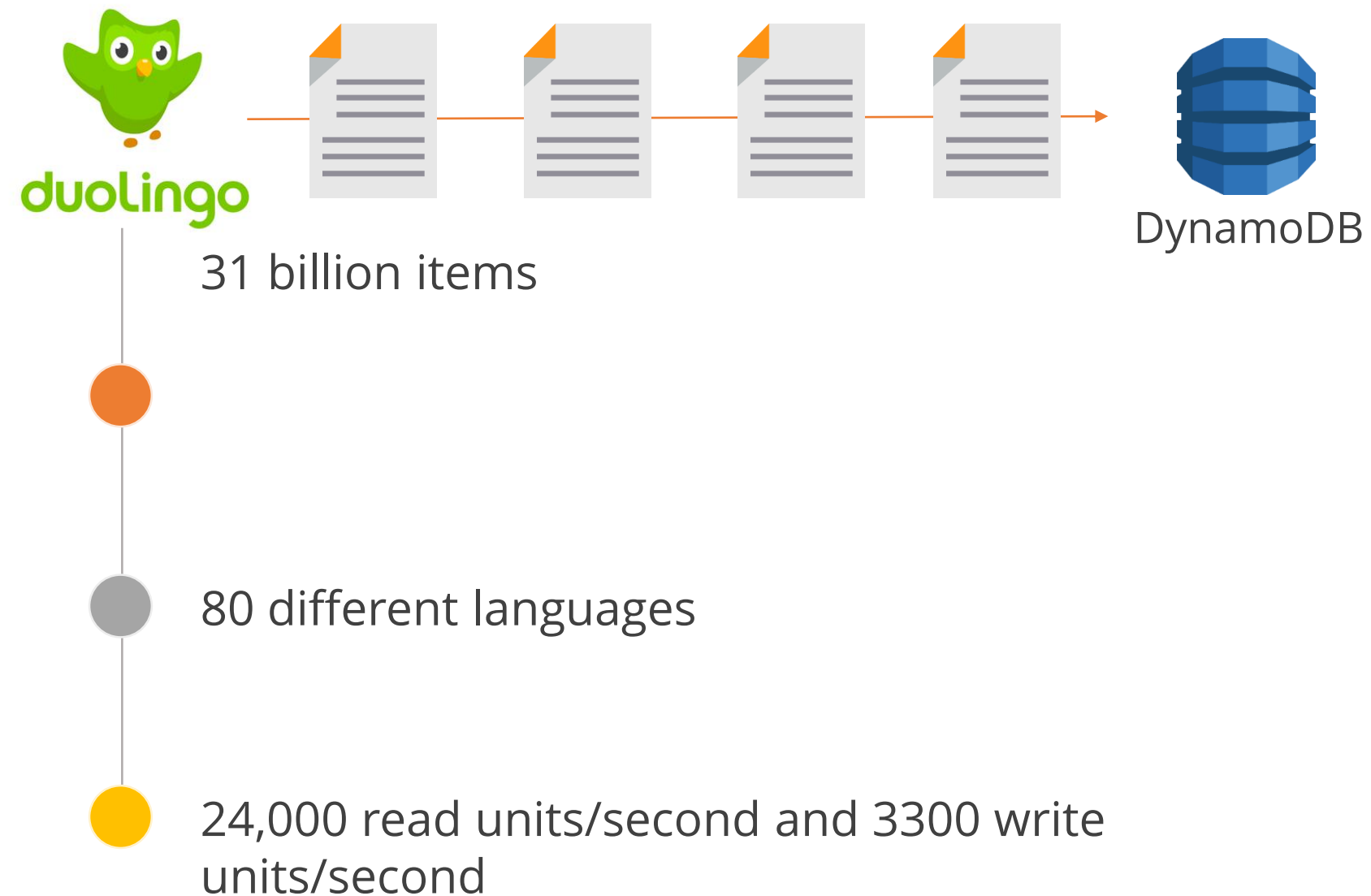
Provisioned Throughput Capacity

DynamoDB is charged using provisioned throughput capacity. When you create a DynamoDB table, you can define the capacity that you want to reserve for reads and writes.



Case Study: Duolingo

Duolingo uses DynamoDB to store 31 billion items for their online learning site that delivers lessons in 80 languages.





Demo 3: Amazon DynamoDB

Demonstrate how to use and scale DynamoDB



Knowledge Check

KNOWLEDGE
CHECK

Which application type is best suited for DynamoDB?

- a. Structured data with heavy write activity
- b. Unstructured data with heavy write activity
- c. Structured data with heavy read activity
- d. Unstructured data with heavy read activity



KNOWLEDGE
CHECK

Which application type is best suited for DynamoDB?

- a. Structured data with heavy write activity
- b. Unstructured data with heavy write activity
- c. Structured data with heavy read activity
- d. Unstructured data with heavy read activity



The correct answer is **d)**

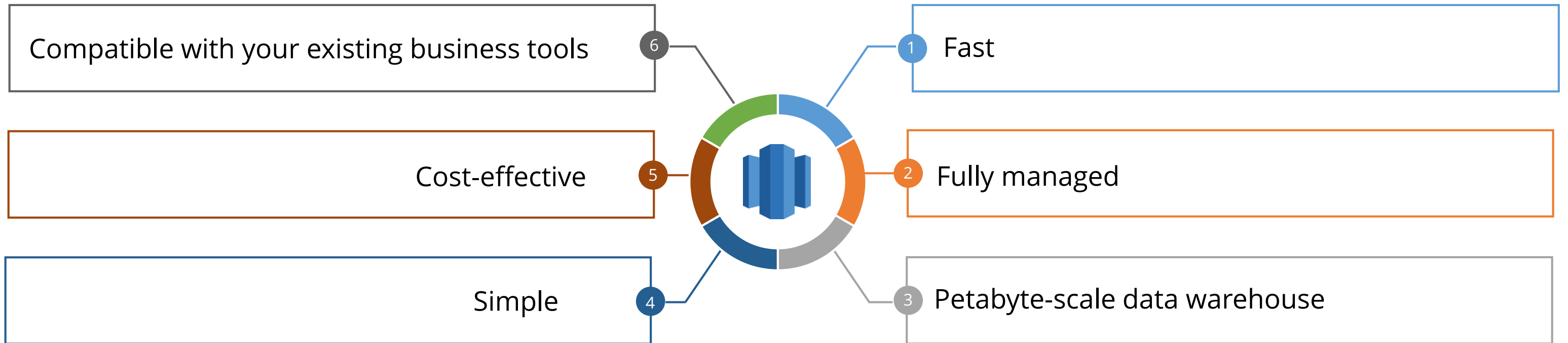
DynamoDB is a NoSQL database best suited for applications with heavy read and low write activity.

Amazon Redshift

Overview of Amazon Redshift

Amazon Redshift

Following are the features of Amazon Redshift:



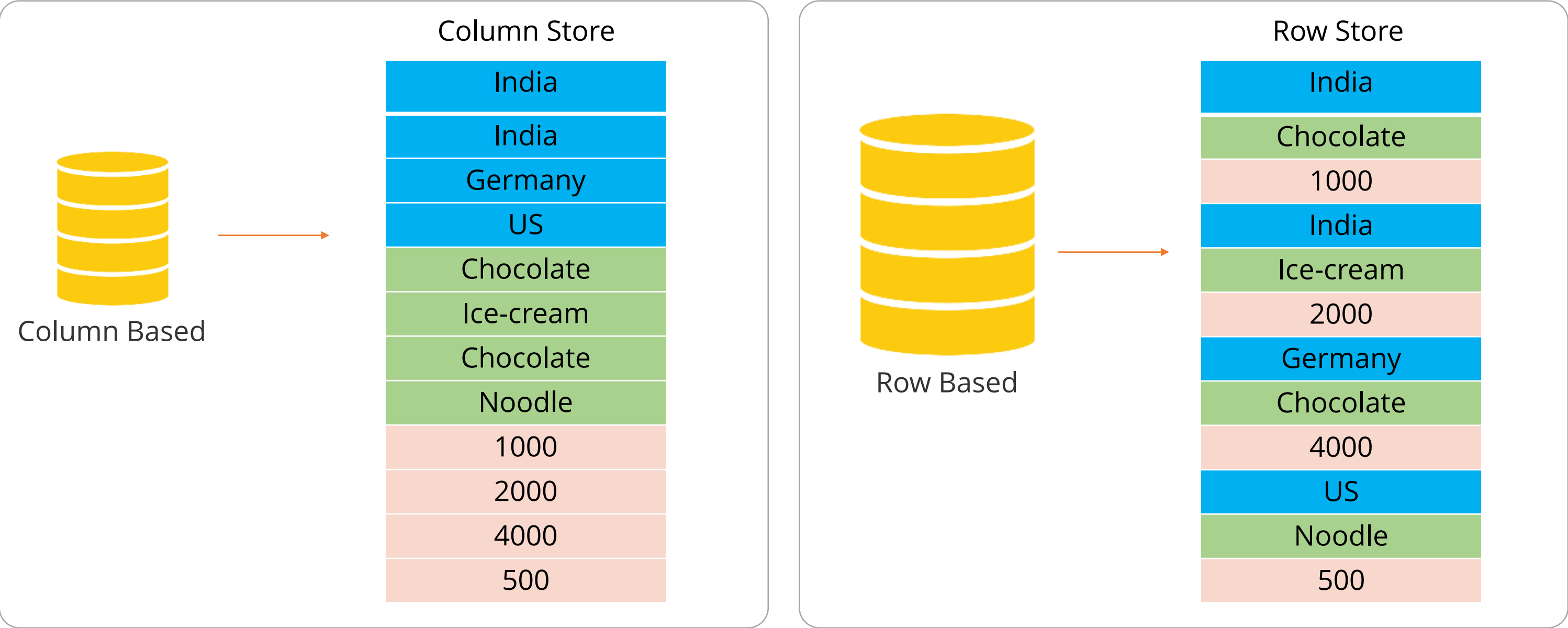
OLAP and Columns

Redshift is a column based database designed for OLAP, which allows you to combine multiple complex queries to provide answers.

		Country	Product	Sales					
Row 1	Row Store	India	Chocolate	1000	Country	India			
		India	Ice-cream	2000		India			
		Germany	Chocolate	4000		Germany			
		US	Noodles	500		US			
Row 2		Chocolate				Product	Chocolate		
		1000						Ice-cream	
		India						Chocolate	
Row 3		Ice-cream					Noodle		
		2000						1000	
Row 4		Germany				Sales	2000		
		Chocolate						4000	
		4000						500	
		US							
		Noodle							
		500							

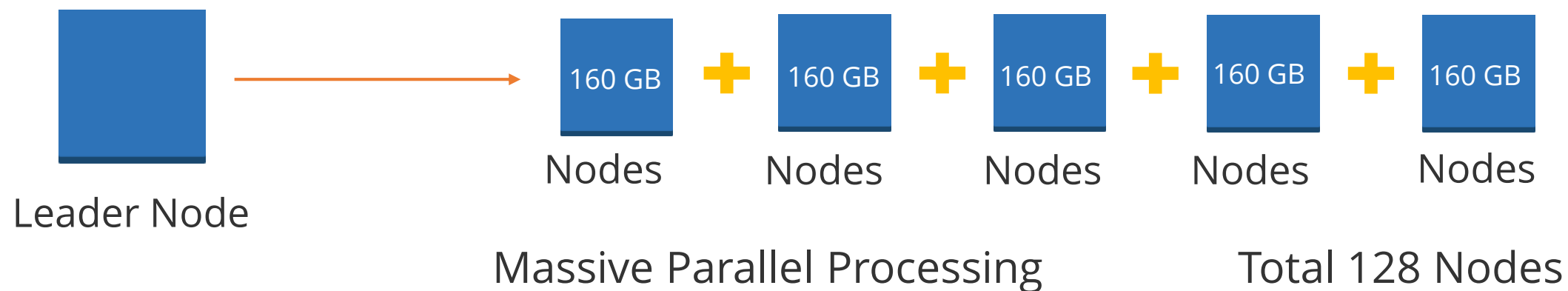
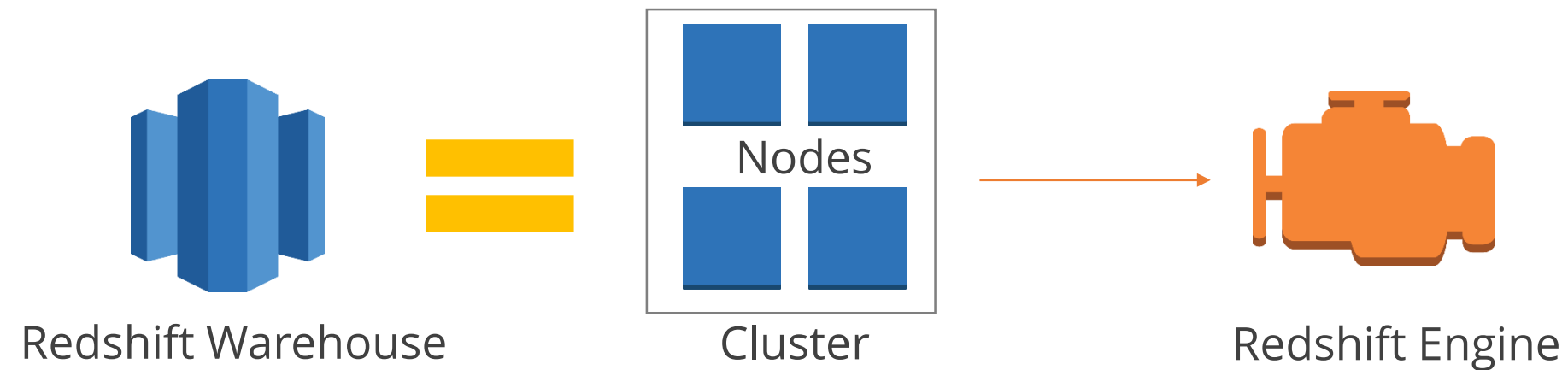
Columnar Storage

Columnar data is stored sequentially in the storage. It requires less reads to get all the data.



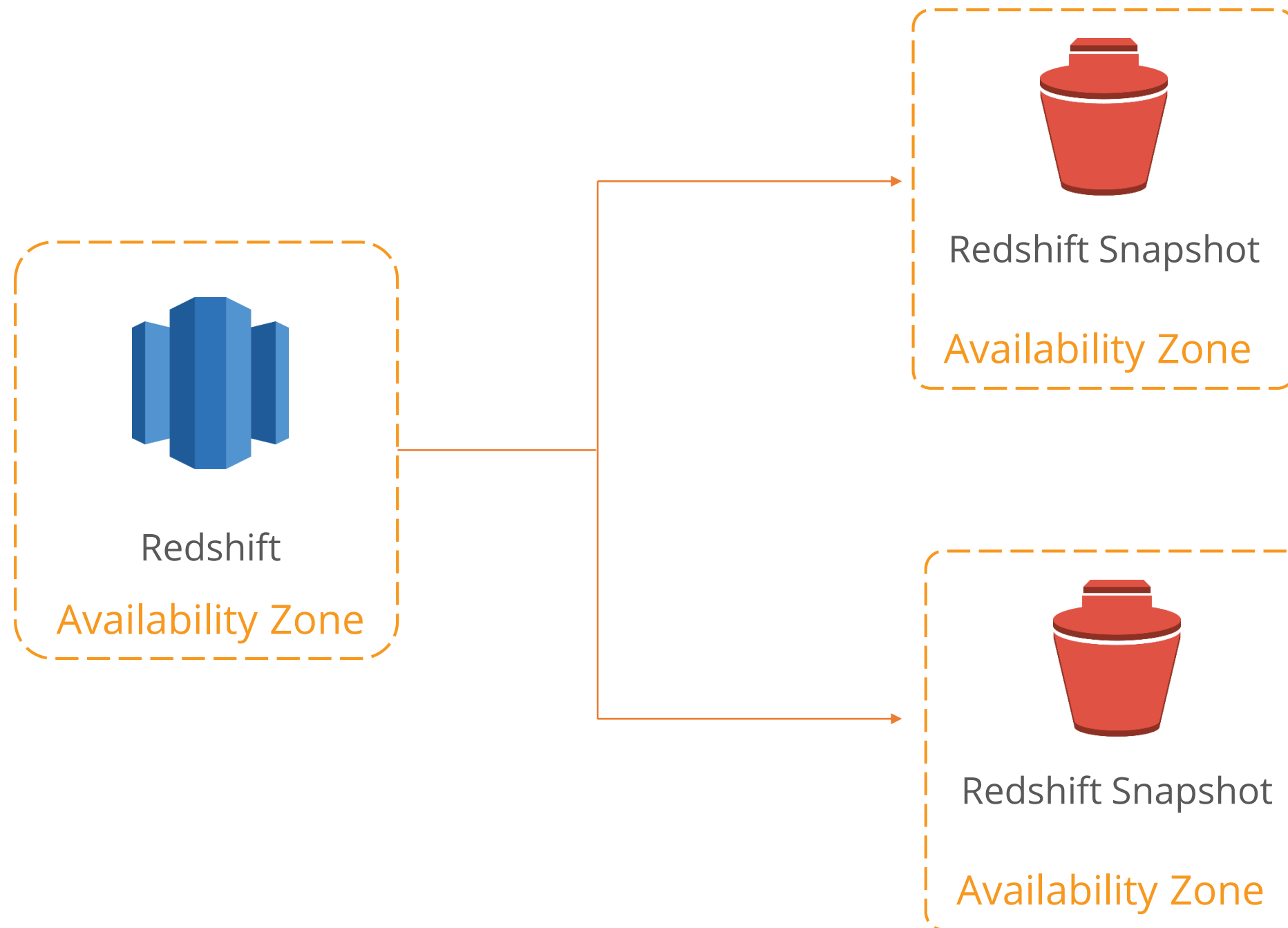
Nodes

An Amazon Redshift data warehouse is a collection of computing resources called nodes. Nodes are organized into a group called a cluster. Each cluster runs in Amazon Redshift engine that contains one or more databases.



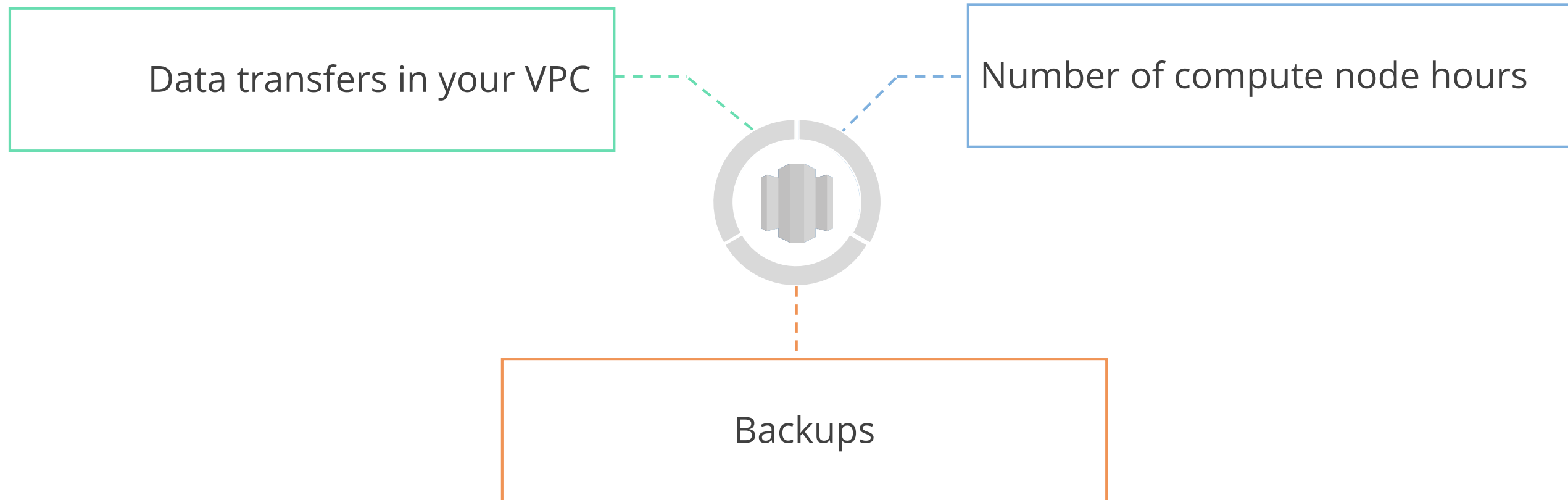
Limitation

Redshift does not have high availability as it is available only in one availability zone.



Redshift Costs

Following are the cost associated with Redshift:





Knowledge Check

KNOWLEDGE
CHECK

Which of the following is NOT a benefit of Redshift?

- a. Ability to reduce storage costs
- b. Requires less reads to get data from storage
- c. Provides high availability
- d. Ability to process OLAP queries quickly and efficiently



KNOWLEDGE
CHECK

Which of the following is NOT a benefit of Redshift?

- a. Ability to reduce storage costs
- b. Requires less reads to get data from storage
- c. Provides high availability
- d. Ability to process OLAP queries quickly and efficiently



The correct answer is **c)**

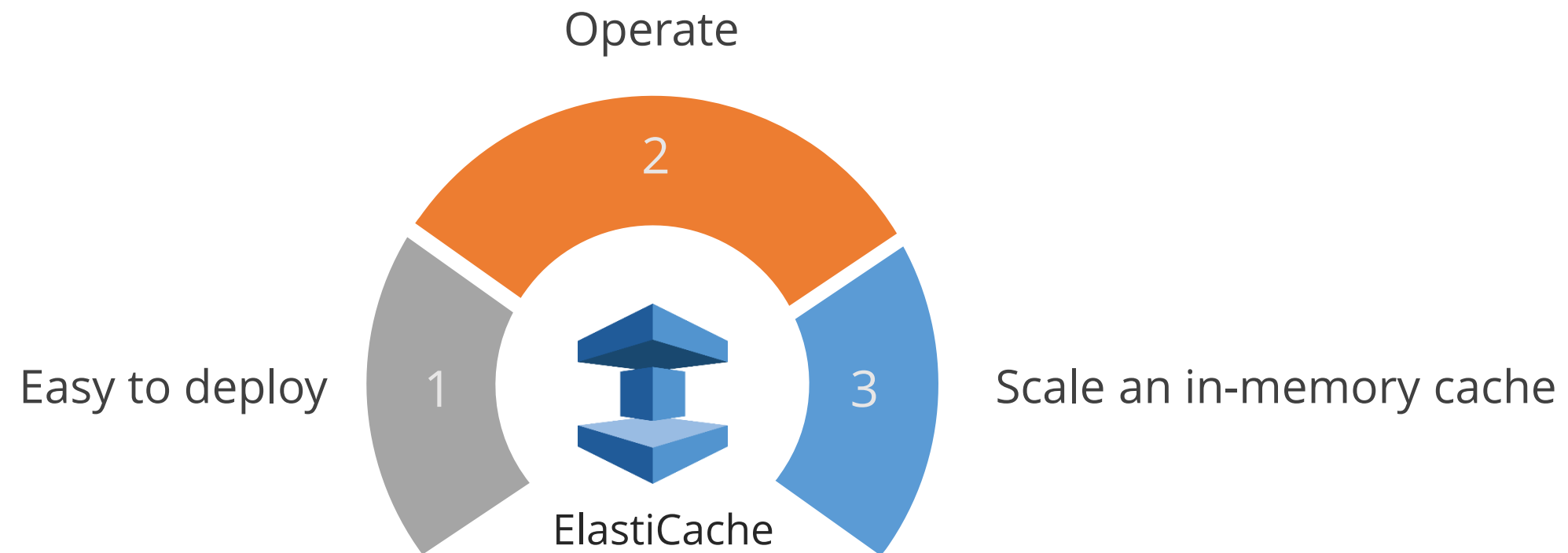
Redshift is not designed for high availability as it is only available in one Availability Zone.

Amazon ElastiCache

Overview of Amazon ElastiCache

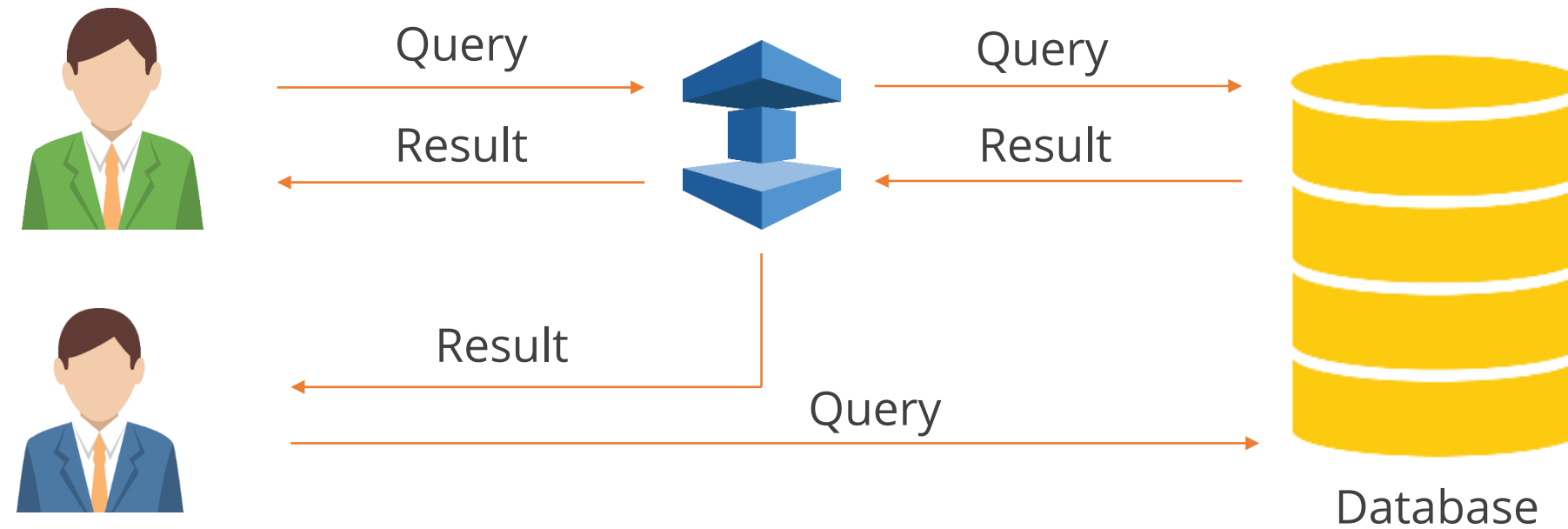
Amazon ElastiCache Overview

Following are the features of Amazon ElastiCache:



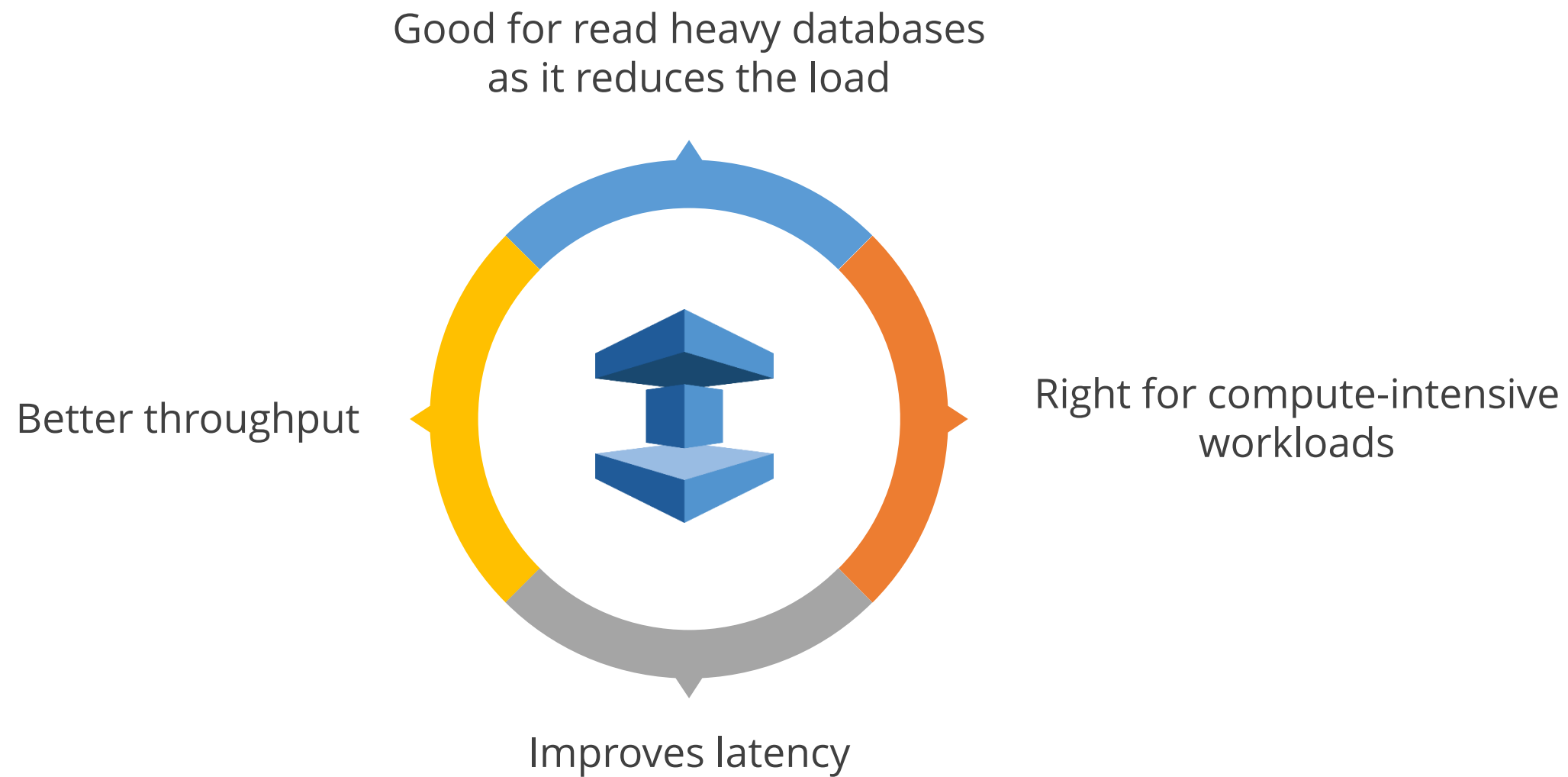
Cached Query Results

ElastiCache caches can store input/output intensive database query results or compute intensive calculations.



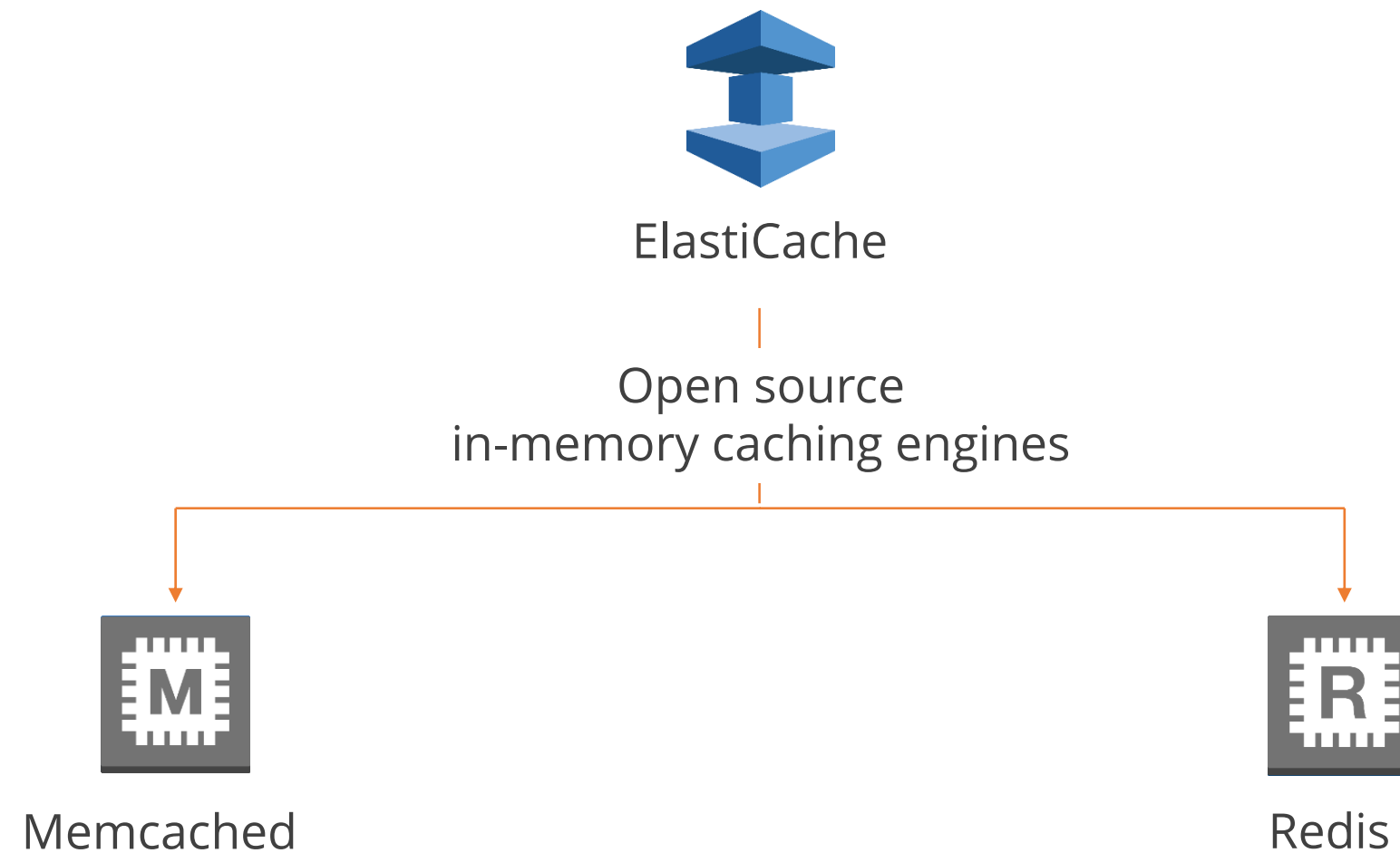
Amazon ElastiCache Benefits

Following are the benefits of Amazon ElastiCache:



Memcached and Redis

Amazon ElastiCache supports both Memcached and Redis, which are open source in-memory caching engines.





Knowledge Check

KNOWLEDGE
CHECK

Which type of application is ElastiCache suitable for?

- a. Write heavy with low read activity
- b. Read heavy with low write activity
- c. Read heavy with high write activity
- d. Write heavy with high read activity



KNOWLEDGE
CHECK

Which type of application is ElastiCache suitable for?

- a. Write heavy with low read activity
- b. Read heavy with low write activity
- c. Read heavy with high write activity
- d. Write heavy with high read activity



The correct answer is **b)**

ElastiCache is suitable for applications that are read heavy with low write activity so users' queries can be cached for future use.

Amazon Aurora

Overview of Amazon Aurora

Amazon Aurora Overview

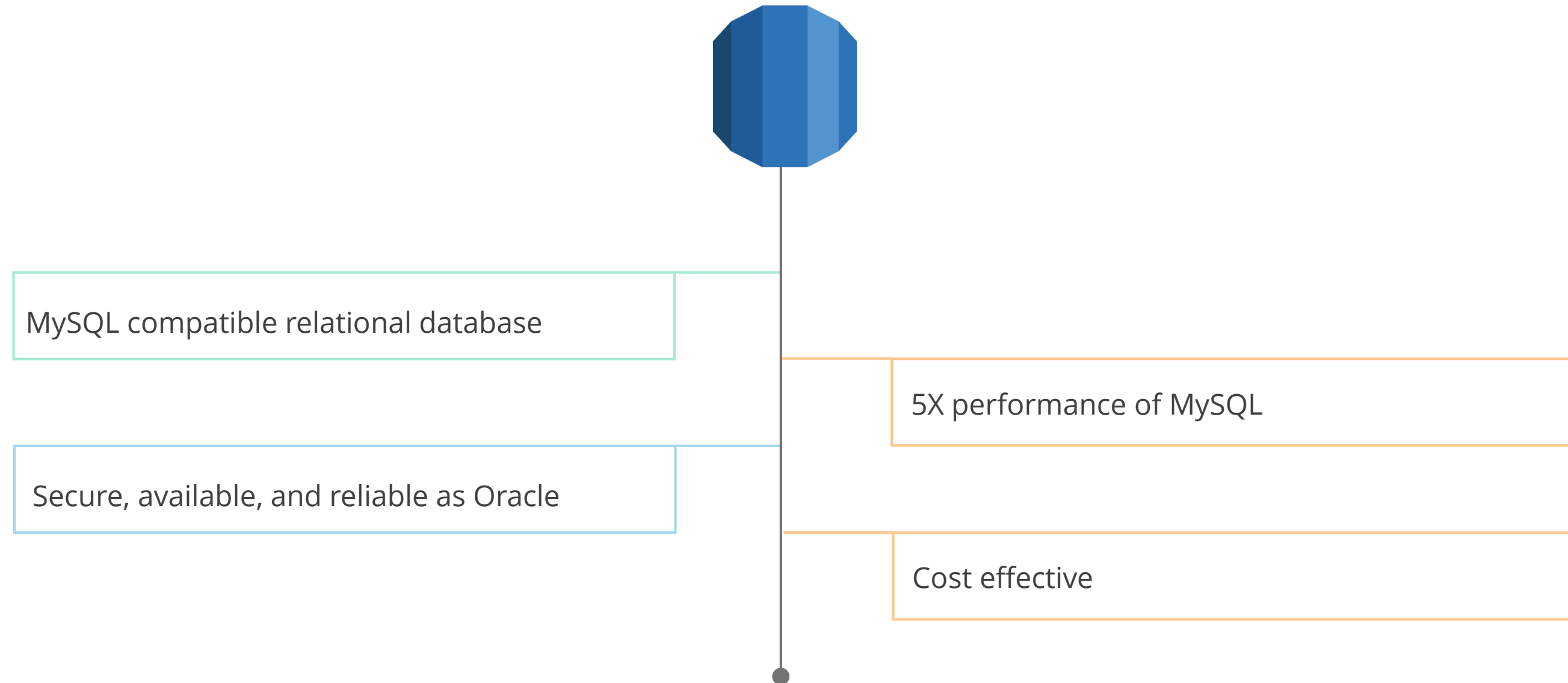
Amazon Aurora is a MySQL and Postgres compatible relational database engine that combines the speed and availability of high-end commercial databases with the simplicity and cost-effectiveness of open source databases.



Amazon Aurora runs only on AWS.

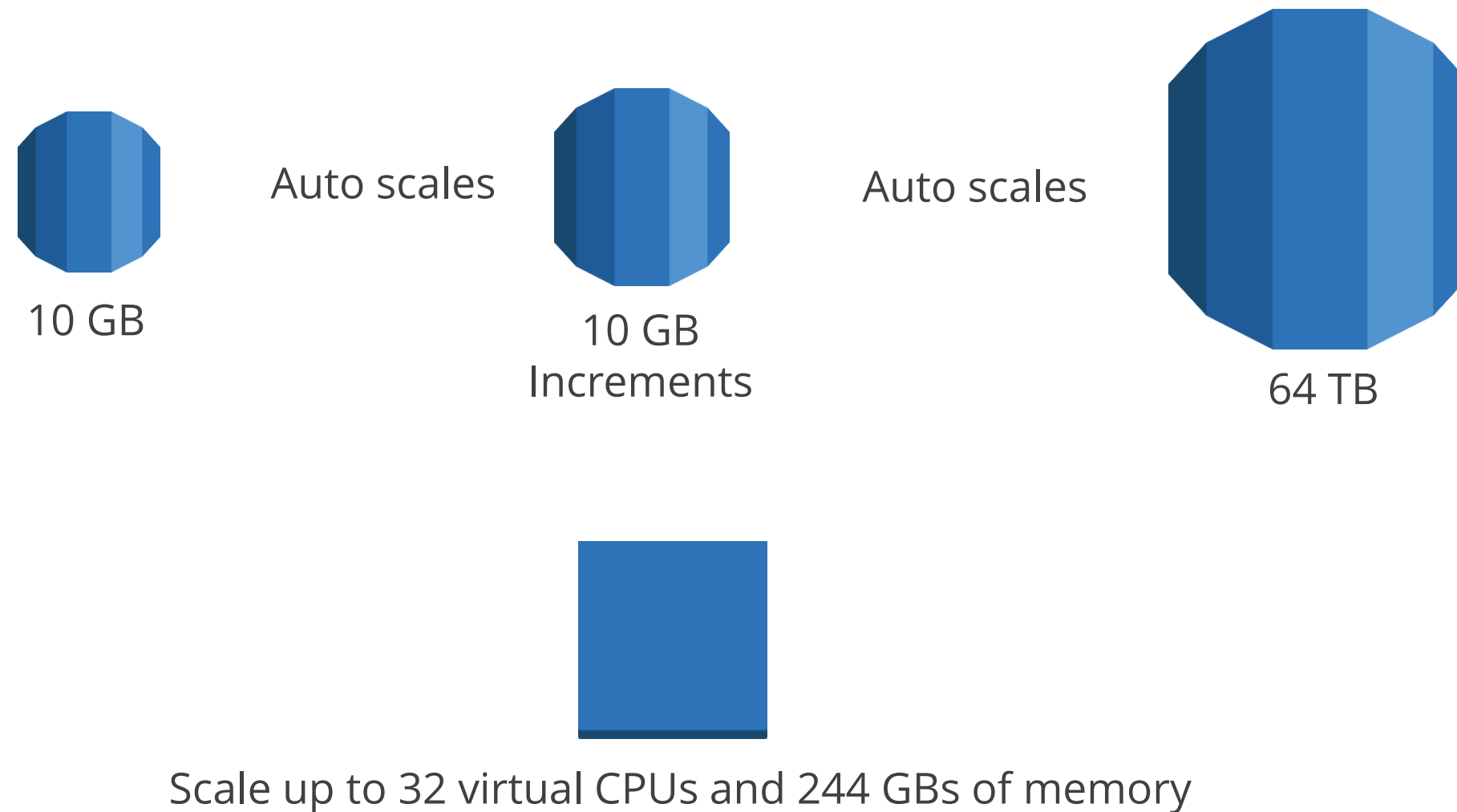
Amazon Aurora Benefits

Following are the benefits of Amazon Aurora:



Amazon Aurora Architecture

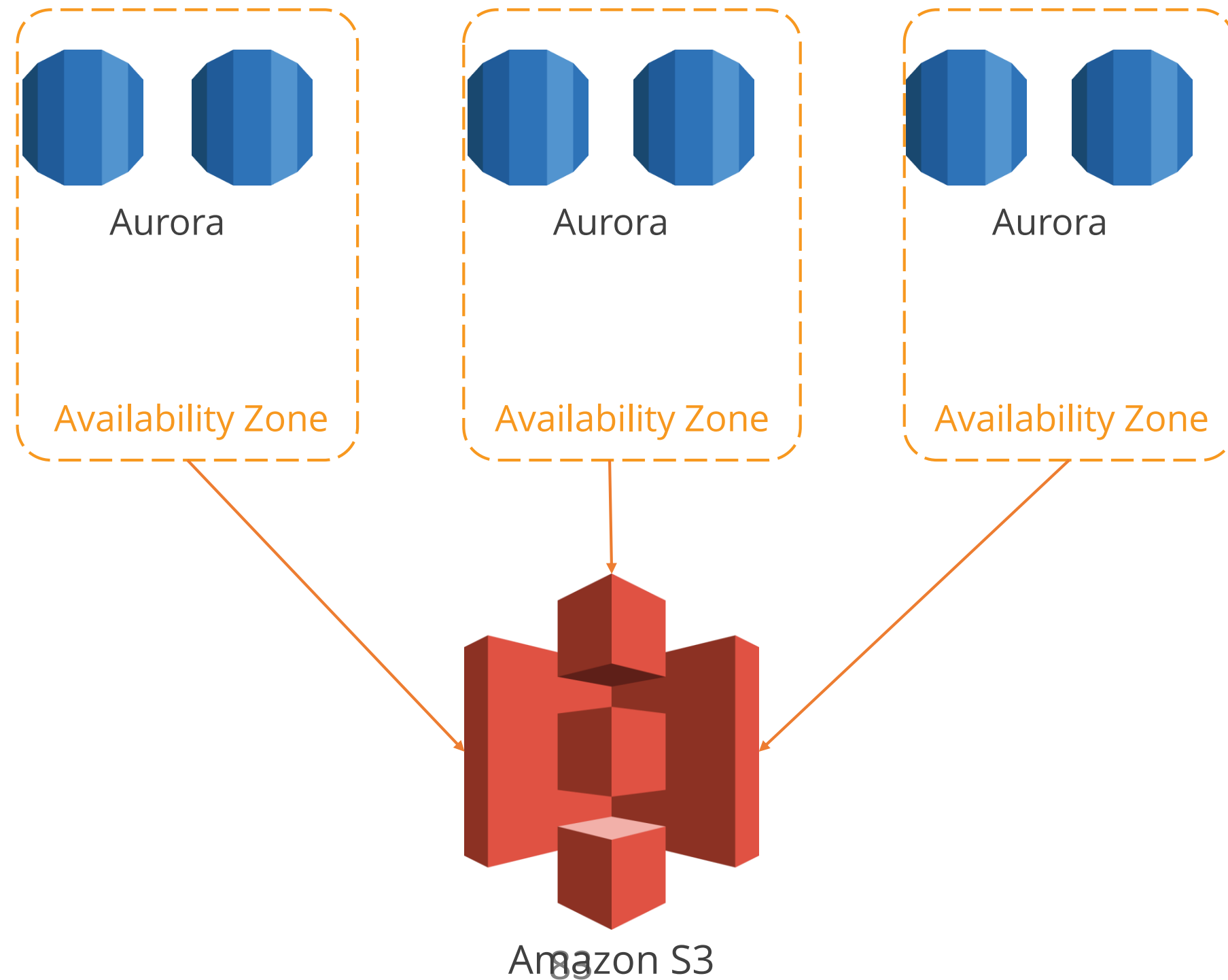
Amazon Aurora architecture starts with a 10 GB database that scales in 10 GB increments up to 64 TB. You can scale instance size up to 32 virtual CPUs and 244 GB of memory.



The auto scaling process takes a couple of minutes as it is not instant.

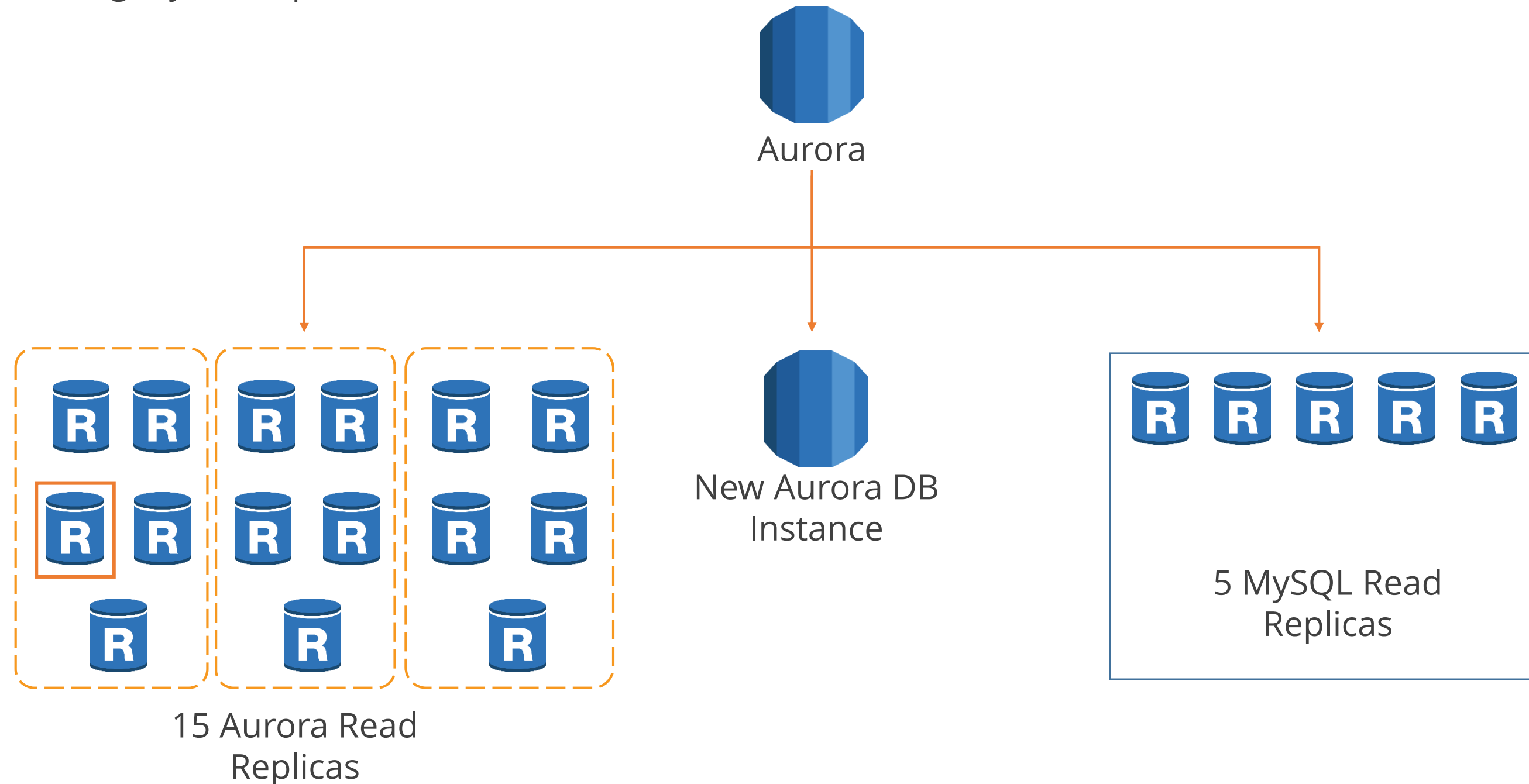
Amazon Aurora Architecture (contd.)

Aurora offers high-availability, it maintains six copies of data which is replicated across three Availability Zones.



Amazon Aurora Replicas

Amazon Aurora creates up to 15 Aurora read replicas and up to five MySQL read replicas which can be created using MySQL replication.





Knowledge Check

KNOWLEDGE
CHECK

Amazon Aurora maintains how many copies of data?

- a. 6 copies in 2 Availability Zones
- b. 3 copy in 3 Availability Zones
- c. 9 copies in 3 Availability Zones
- d. 6 copies in 3 Availability Zones



KNOWLEDGE
CHECK

Amazon Aurora maintains how many copies of data?

- a. 6 copies in 2 Availability Zones
- b. 3 copy in 3 Availability Zones
- c. 9 copies in 3 Availability Zones
- d. 6 copies in 3 Availability Zones



The correct answer is **d)**

Amazon Aurora maintains 6 copies of data which is replicated across three Availability Zones.

AWS Database Migration Services (DMS)

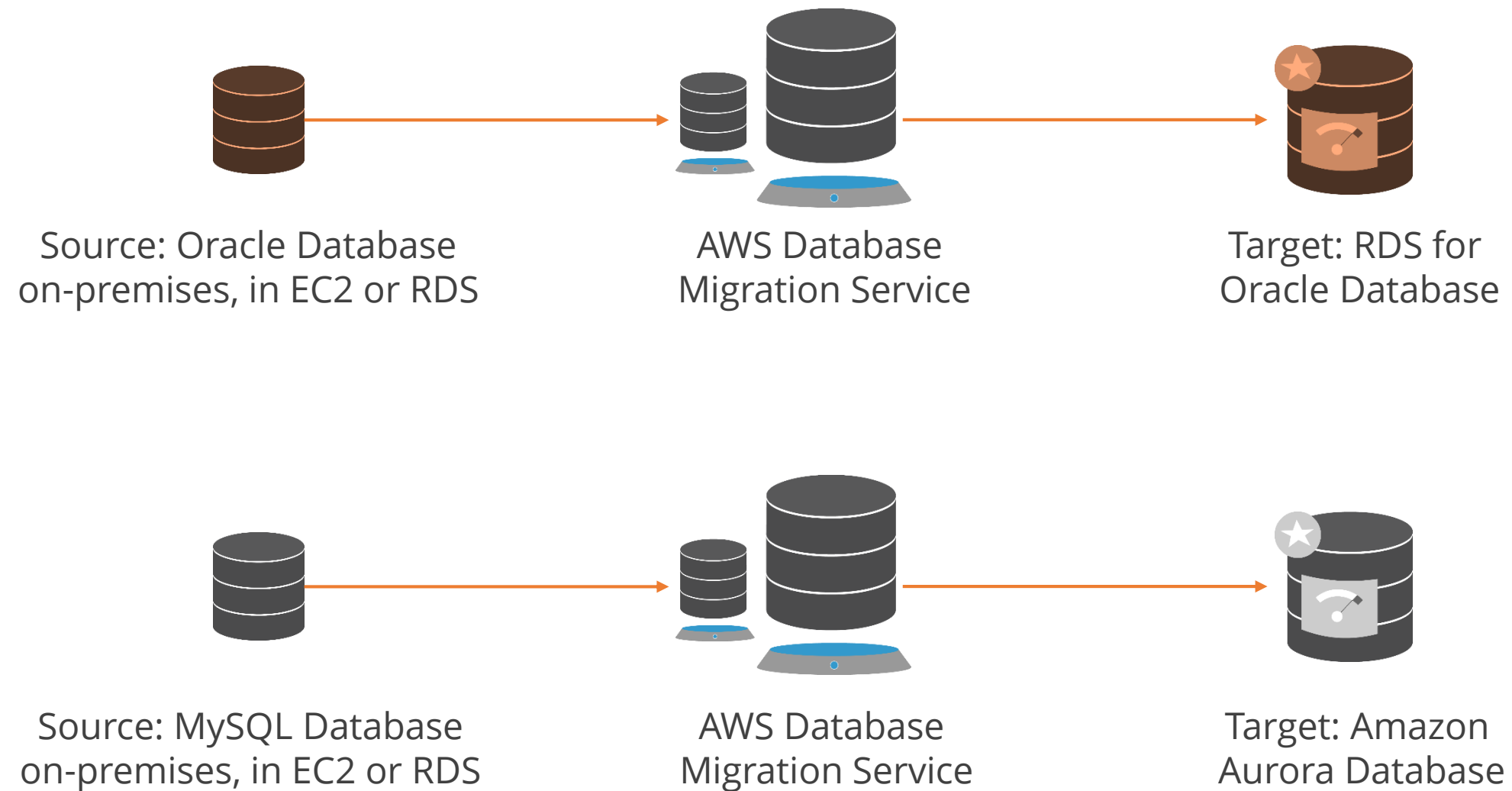
Overview of AWS DMS

AWS DMS Overview

According to Amazon, “AWS Database Migration Service allows you to migrate your databases to AWS infrastructure with minimal downtime.”

Homogenous Migrations

You can perform homogenous migrations, such as Oracle to Oracle or MySQL to Amazon Aurora.



Heterogeneous Migrations

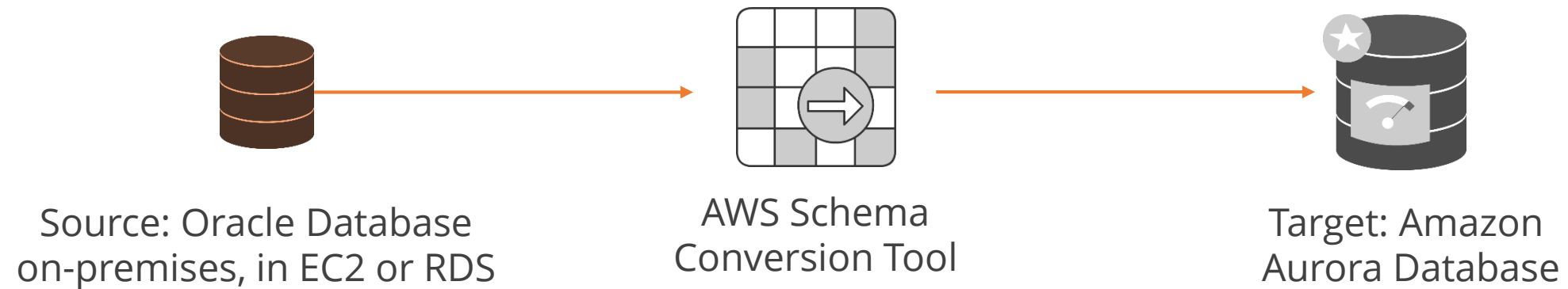
Using the AWS Schema Conversion Tool you can perform heterogeneous migrations between different RDBMS.

Source Database	Target Database
Microsoft SQL Server	Amazon Aurora, MySQL, PostgreSQL, MariaDB
MySQL and MariaDB	PostgreSQL
Oracle	Amazon Aurora, MySQL, PostgreSQL, MariaDB
PostgreSQL	Amazon Aurora, MySQL, MariaDB
Amazon Aurora	PostgreSQL

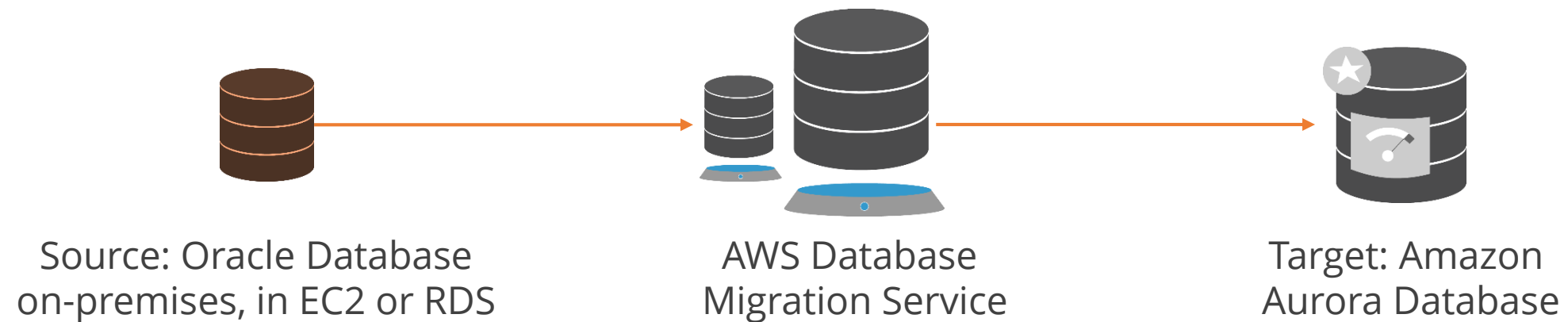
AWS Schema Conversion Tool

The Schema Conversion Tool is a two step conversion process:

STEP 1:

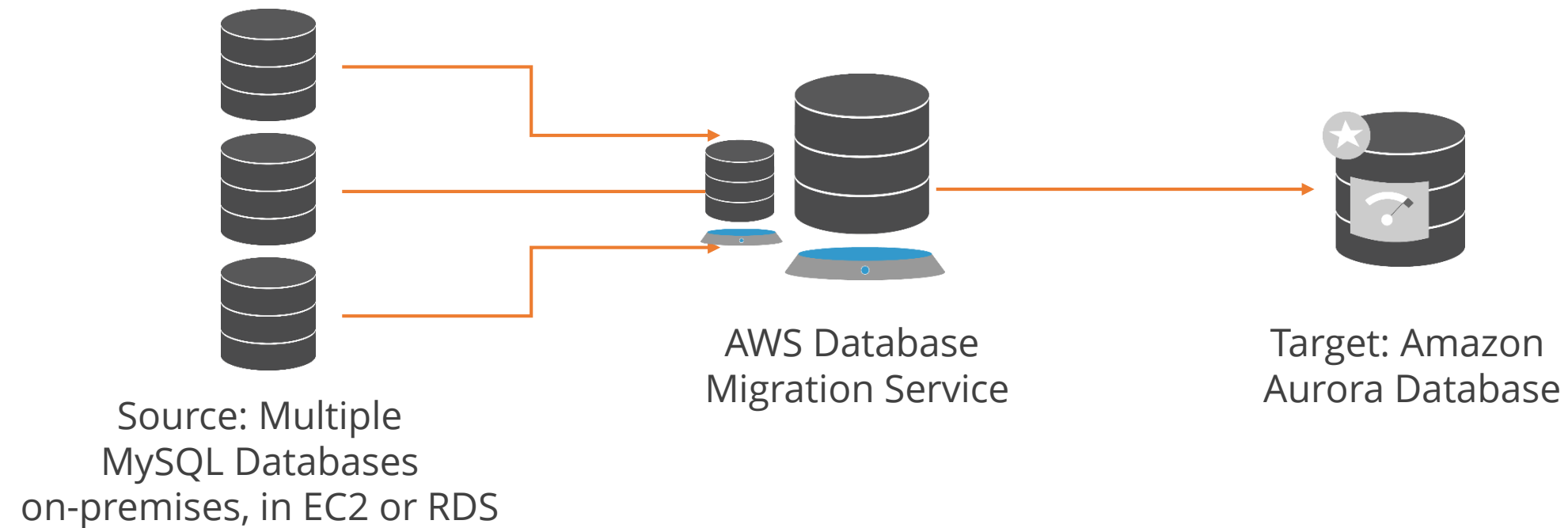


STEP 2:



AWS Schema Conversion Tool (contd.)

DMS has the ability to stream data from any supported database source to Aurora or Redshift so you can easily consolidate your data or create a data warehouse.





Knowledge Check

KNOWLEDGE
CHECK

Which of the following two types of migration does AWS DMS support?

- a. Heterogeneous and Homogenous
- b. Transactional and Replication
- c. Synchronous and Asynchronous
- d. Transient and Migratory



KNOWLEDGE
CHECK

Which of the following two types of migration does AWS DMS support?

- a. Heterogeneous and Homogenous
- b. Transactional and Replication
- c. Synchronous and Asynchronous
- d. Transient and Migratory



The correct answer is **a)**

AWS DMS supports Heterogeneous and Homogenous migration.

Database Best Practices

Overview of AWS recommended database best practices

RDS

RDS

OLTP

NoSQL

Data Warehousing
and OLAP

Use Amazon RDS instances to automatically provide a fully managed RDBMS with the following benefits:

- Built-in redundancy between Availability Zones
- Automatic backups
- Automatic patching and updates

OLTP

RDS

OLTP

NoSQL

Data Warehousing
and OLAP

For OLTP applications use DynamoDB or RDS with provisioned IOPS.

NoSQL

RDS

OLTP

NoSQL

Data Warehousing
and OLAP

For NoSQL databases use DynamoDB as it is a fast and flexible database service.

Data Warehousing and OLAP

RDS

OLTP

NoSQL

Data Warehousing
and OLAP

For Data Warehousing and OLAP use Amazon Redshift.

In-Memory

In-memory

Migration

For In-memory or caching database requirements use Amazon ElastiCache.

Migration

In-memory

Migration

When migrating your databases to AWS use the AWS Database Migration Services. Check if you can reduce your database costs by migrating to a lower cost RDBMS.

Database Costs

Overview of costs associated with AWS databases

Database Costs

The following costs are associated with the AWS databases:

- Amazon RDS:
 - Cost per hour depending on the instance size and RDBMS licensing cost (if required)
 - Per GB cost for storage
- Amazon DynamoDB:
 - Write Throughput: fee per hour for every 10 units of Write Capacity
 - Read Throughput: fee per hour for every 50 units of Read Capacity
 - Per GB fee for indexed storage
 - Fee for DynamoDB streams
 - Data transfer out fee
- Amazon Redshift:
 - Cost per hour depending on the instance size

Database Costs (contd.)

The following costs are associated with the AWS databases:

- Amazon ElastiCache:
 - Cost per hour depending on the cache size selected
 - Data transfer cost between EC2 instances and ElastiCache in different AZs
- AWS Database Migration Services:
 - Cost per hour for the compute resources use during migration
 - Data transfer out to the Internet

Key Takeaways

Key Takeaways

- RDS allows you to run databases in the AWS cloud on EC2 instances without worrying about database administration management tasks.
- In situations where RDS isn't suitable, you can launch an EC2 instance from AMIs with pre-installed databases.
- Amazon DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit millisecond latency at any scale.
- Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse that makes it simple and cost-effective to analyze all your data using your existing business intelligence tools.
- ElastiCache is an AWS service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud.
- Amazon Aurora is a MySQL and Postgres compatible relational database engine that combines the speed and availability of high-end commercial databases with the simplicity and cost-effectiveness of open source databases.
- AWS Database Migration Service allows you to migrate your databases to AWS infrastructure with minimal downtime.



Practice Assignment: Amazon RDS

Launch an Amazon RDS instance and connect to a webserver

Launch Amazon RDS Instance



Your company wants to use an RDS instance as the data repository for their webserver.

You will need to complete the following tasks:

1. Create an additional private subnet in your custom VPC.
2. Verify the inbound rules for your DBSERVER security group.
3. Create a DB Subnet Group to host your RDS instance.
4. Launch a MySQL RDS instance into the DB subnet group.
5. Launch a new webserver into a public subnet of your custom VPC.
6. Configure connectivity between the webserver and the RDS instance.

You can use demonstration one from this lesson as a reference for this Practice Assignment.



QUIZ

1

What is the charge when replicating data between your primary and secondary RDS instances?

- a. The AWS standard charge
- b. Half the AWS standard charge
- c. Depends on the amount of data transferred
- d. No charge; it's free



QUIZ

1

What is the charge when replicating data between your primary and secondary RDS instances?

- a. The AWS standard charge
- b. Half the AWS standard charge
- c. Depends on the amount of data transferred
- d. No charge; it's free



The correct answer is **d**

Explanations: RDS synchronously replicates your data to a standby instance in a different Availability zone for free.

QUIZ

2

Which AWS service is most suitable for non-relational databases?

- a. ElastiCache
- b. DynamoDB
- c. RDS
- d. EC2



QUIZ

2

Which AWS service is most suitable for non-relational databases?

- a. ElastiCache
- b. DynamoDB
- c. RDS
- d. EC2



The correct answer is **b**

Explanations: DynamoDB is a NoSQL database that is best suited for non relational databases.

QUIZ

3

What is the maximum-sized RDS volume you can have for MySQL and Oracle?

- a. 6TB
- b. 4TB
- c. 2TB
- d. 3TB



QUIZ

3

What is the maximum-sized RDS volume you can have for MySQL and Oracle?

- a. 6TB
- b. 4TB
- c. 2TB
- d. 3TB



The correct answer is **a**

Explanations: MySQL, PostgreSQL, and Oracle RDS instances are limited to 6TB in size. For SQL Server, it is 4TB, and for SQL Server Express, it is 300GB.

QUIZ

4

What is the maximum IOPS capacity of an Oracle and MySQL RDS instance?

- a. 10,000 IOPS
- b. 20,000 IOPS
- c. 30,000 IOPS
- d. 50,000 IOPS



QUIZ

4

What is the maximum IOPS capacity of an Oracle and MySQL RDS instance?

- a. 10,000 IOPS
- b. 20,000 IOPS
- c. 30,000 IOPS
- d. 50,000 IOPS



The correct answer is **c**

Explanations: MySQL, MariaDB, PostgreSQL, and Oracle RDS instances have a maximum of 30000 IOPS. SQL Server has a maximum of 20,000 IOPS.

QUIZ

5

Which of the following ports needs to be open to use RDP to a SQL Server RDS instance?

- a. 3389
- b. 1433
- c. 22
- d. You cannot RDP to an RDS instance



QUIZ

5

Which of the following ports needs to be open to use RDP to a SQL Server RDS instance?

- a. 3389
- b. 1433
- c. 22
- d. You cannot RDP to an RDS instance



The correct answer is **d**

Explanations: You cannot use RDP (or SSH) to RDS instances.



This concludes “Databases.”

The next lesson is “Application Services.”