AWS
Databases

AWS Database Options

RDS

Non-RDS

In-Memory

Data Warehousing

Custom

Running Databases on AWS:
http://aws.amazon.com/running_databases/

	SQL	NoSQL
Data Storage	Rows and Columns	Key-Value
Schemas	Fixed	Dynamic
Querying	Using SQL	Focused on collection of documents
Scalability	Vertical	Horizontal

SQL

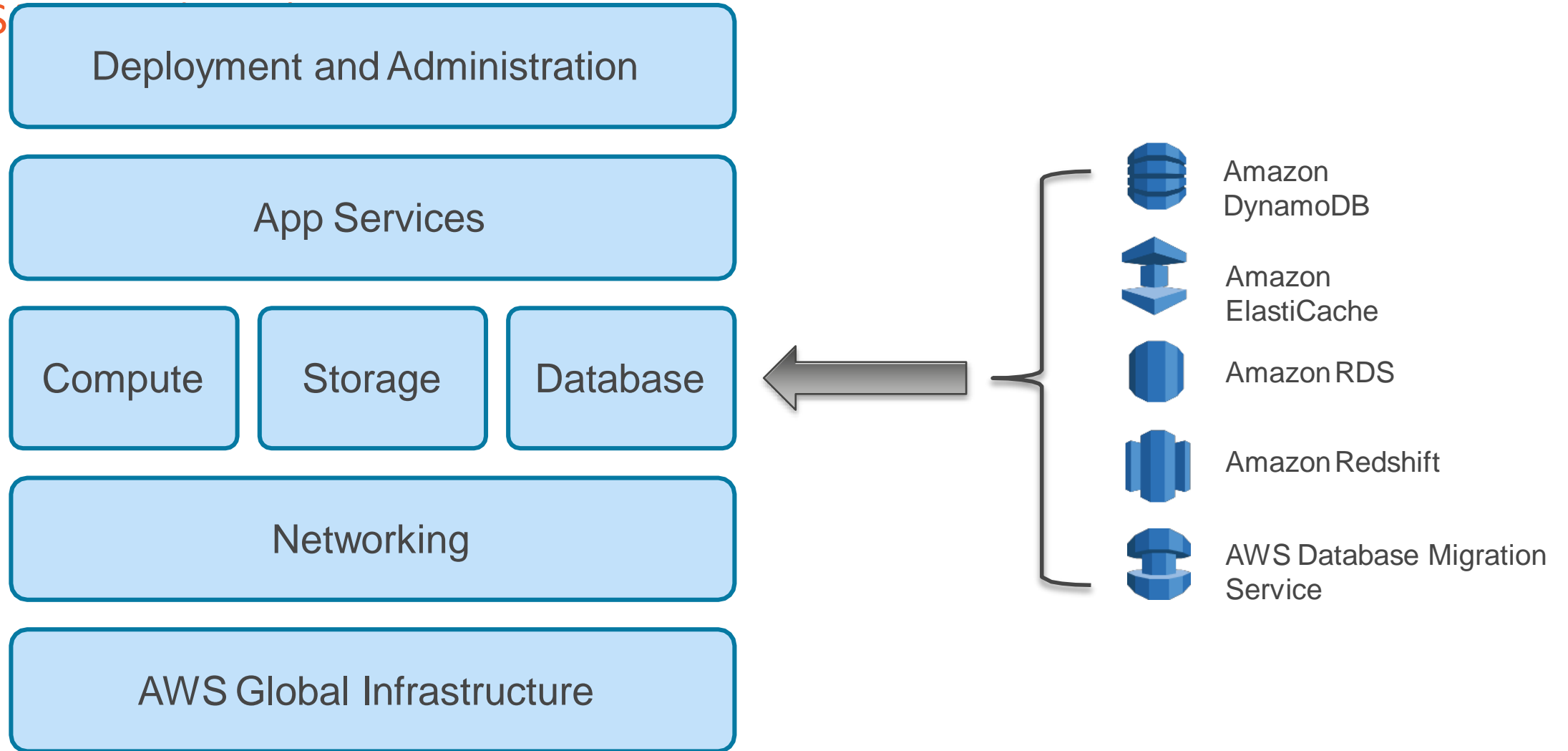| ISBN	Title	Author	Format
918293245265	Cloud Computing Concepts	Wilson, Joe	Paperback
3142536475869	The Database Guru	Gomez, Maria	eBook

NoSQL

```
{
    ISBN: 9182932465265,
    Title: "Cloud Computing
    Concepts",  Author: "Wilson, Joe",
    Format: "Paperback"
}
```

Data Storage Considerations

- No one size fits all.

- Analyze your data requirements by considering:
    - Data formats
    - Data size
    - Query frequency
    - Data access speed
    - Data retention period

AWS

Deployment and Administration

App Services

Compute | Storage | Database

Networking

AWS Global Infrastructure

Amazon DynamoDB

Amazon ElastiCache

Amazon RDS

Amazon Redshift

AWS Database Migration Service

Amazon RDS Options

MySQL | Oracle

Microsoft SQL | PostgreSQL | Aurora

www.scmGalaxy.com

Non-RDS

SimpleDB

DynamoDB

MongoDB

Couchbase

In - Memory

Elastic Cache4

Datebase warehouse

Amazon Redshift

Amazon Relational Database Service (RDS)

Amazon RDS

- Cost-efficient and **resizable capacity**
- Manages time-consuming **database administration** tasks
- Access to the full capabilities of **Amazon Aurora**, **MySQL**, **MariaDB**, **Microsoft SQL Server**, **Oracle**, and **PostgreSQL** databases
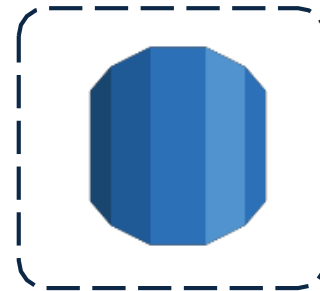
Amazon RDS

- Simple and **fast to deploy**
- Manages common database administrative tasks
- **Compatible** with your applications
- Fast, predictable performance
- Simple and **fast to scale**
- Secure
- Cost-effective

DB Instances

- DB Instances are the basic building blocks of Amazon RDS.

- They are an **isolated database environment** in the cloud.

- They can **contain multiple user-created databases**.

Automatic Backups:

- Restore your database to a point in time.

- Are enabled by default.

- Let you choose a retention period up to 35 days.

Manual Snapshots:

- Let you build a new database instance from a snapshot.

- Are initiated by the user.

- Persist until the user deletes them.

- Are stored in Amazon S3.

Cross-Region Snapshots

- Are a **copy** of a **database** snapshot stored in a **different AWS Region**.

- Provide a backup for disaster **recovery**.

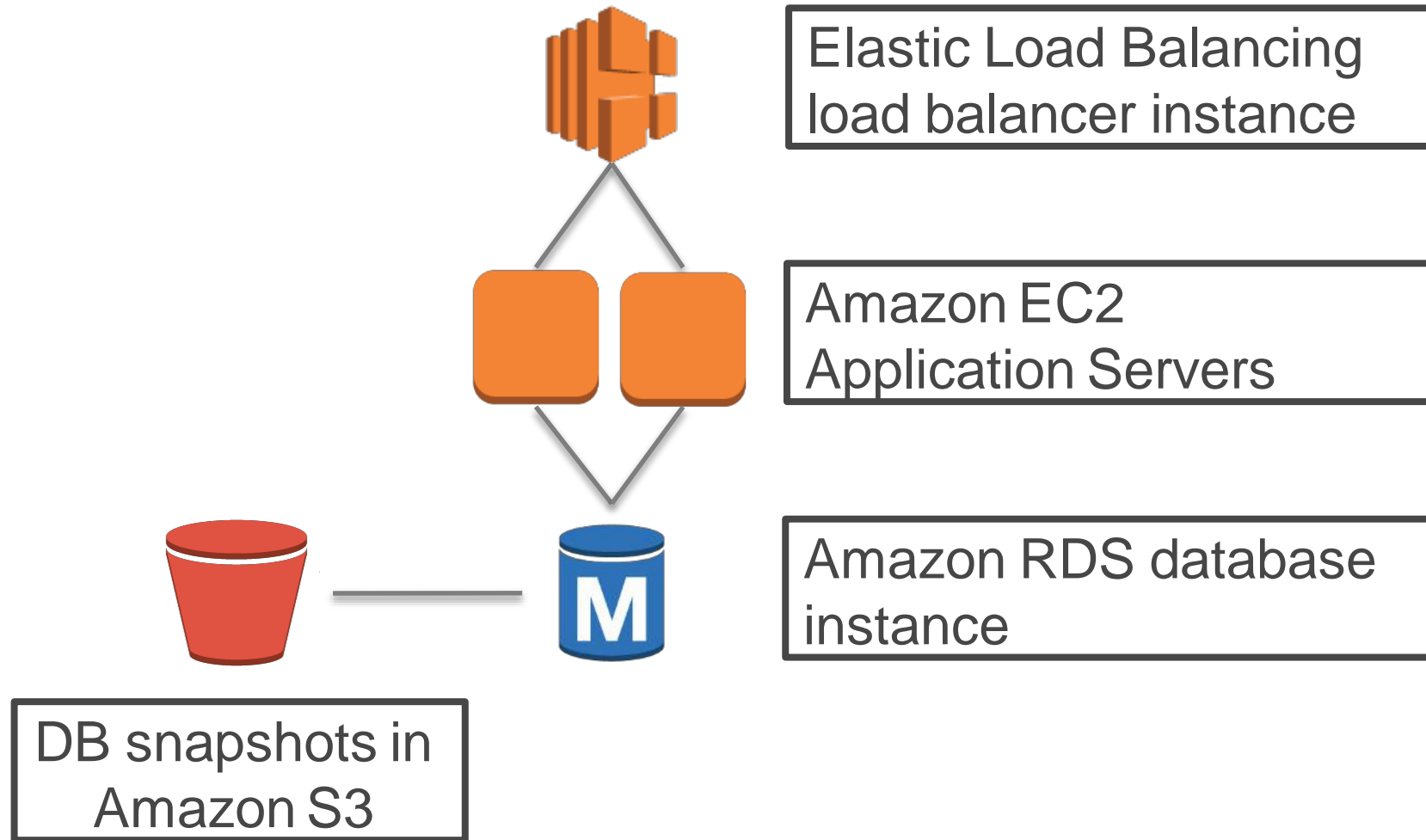- Can be used as a **base** for **migration** to a different region.

Amazon RDS Security

- Run your DB instance in an **Amazon VPC**.
- Use **IAM policies** to grant access to Amazon RDS resources.
- Use **security groups**.
- Use Secure Socket Layer (**SSL**) connections with DB instances (Amazon Aurora, Oracle, MySQL, MariaDB, PostgreSQL, Microsoft SQL Server).
- Use Amazon RDS **encryption** to secure your RDS instances and snapshots at rest.
- Use network encryption and transparent data encryption (**TDE**) with Oracle DB and Microsoft SQL Server instances.
- Use the security features of your DB engine to **control access** to your DB instance.

A Simple Application Architecture

Elastic Load Balancing load balancer instance

Amazon EC2 Application Servers

Amazon RDS database instance

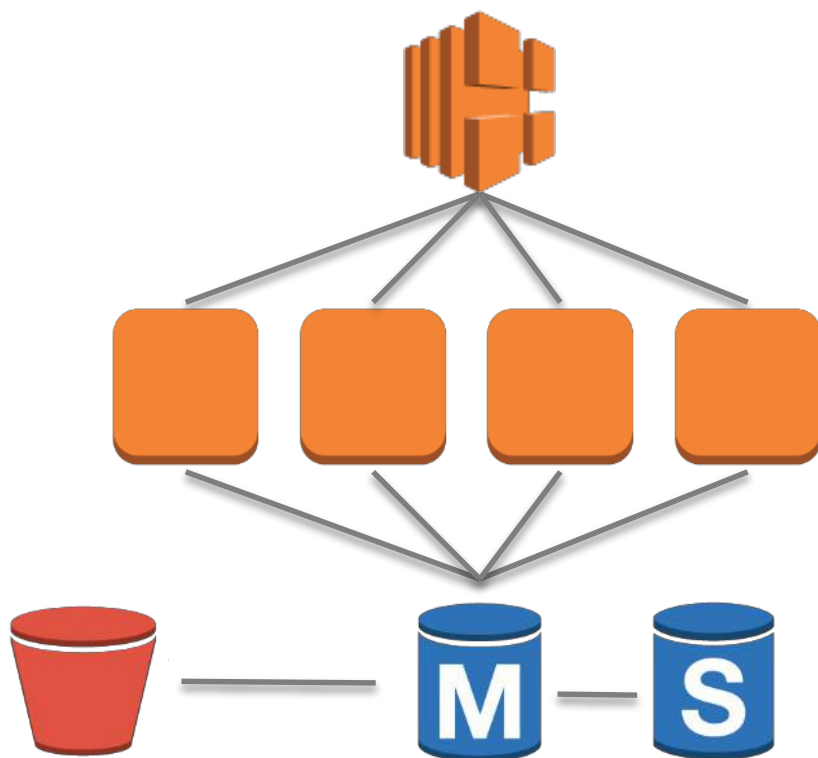DB snapshots in Amazon S3

Multi-AZ RDS Deployment

- With **Multi-AZ** operation, your database is **synchronously replicated to another Availability Zone** in the same AWS Region.

- **Failover** to the standby **automatically** occurs in case of master database failure.

- Planned maintenance is applied first to standby databases.

A Resilient, Durable Application Architecture

Elastic Load Balancing load balancer instance

Application, in Amazon EC2 instances

Amazon RDS database instances: Master and Multi-AZ standby

DB snapshots in Amazon S3

Amazon RDS Best Practices

- **Monitor** your memory, CPU, and storage usage.
- Use **Multi-AZ** deployments to automatically provision and maintain a synchronous standby in a different Availability Zone.
- Enable **automatic backups**.
- Set the **backup window** to occur during the daily low in WriteIOPS.
- To increase the I/O capacity of a DB instance:
  - Migrate to a DB instance class with high I/O capacity.
  - Convert from standard storage to provisioned IOPS storage and use a DB instance class optimized for **provisioned IOPS**.
  - Provision additional throughput capacity (if using provisioned IOPS storage).
- **Test** failover for your DB instance.
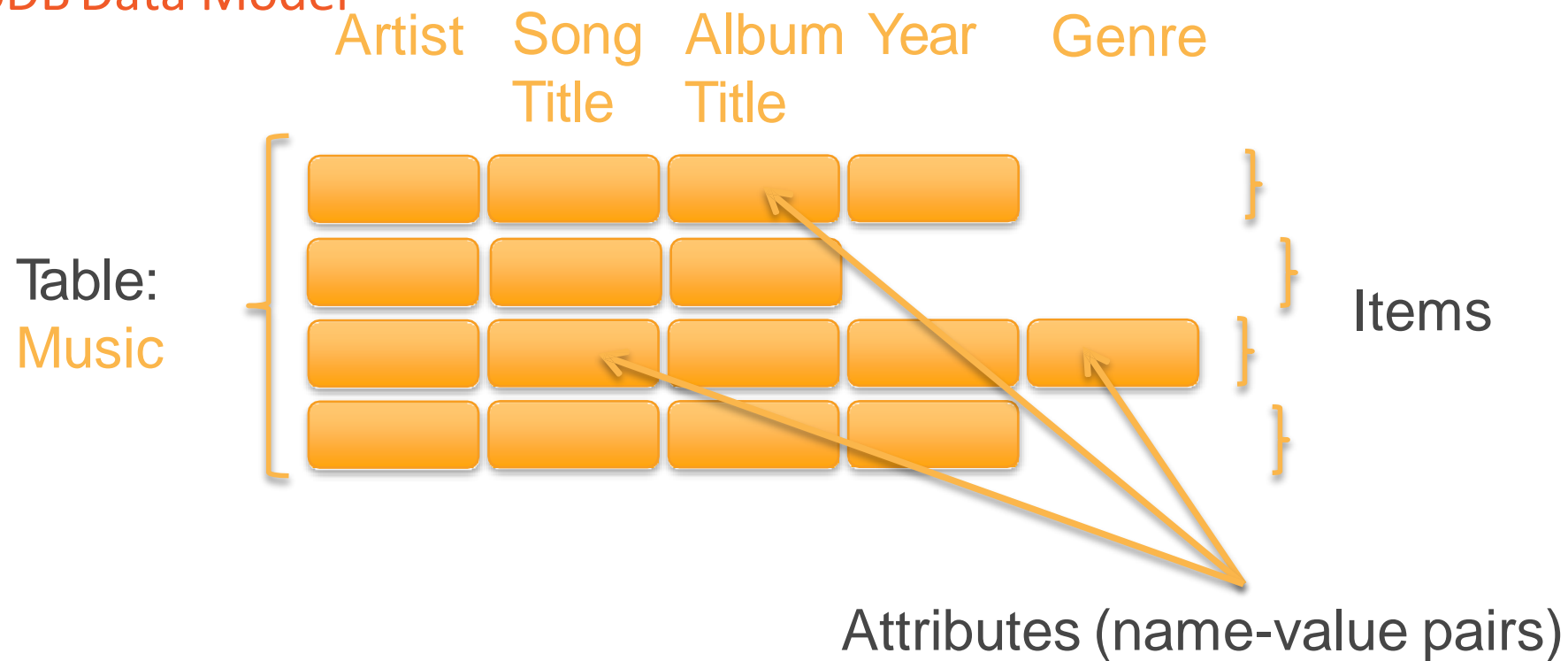
Amazon DynamoDB

Amazon DynamoDB

- Allows you to store any amount of data with **no limits.**
- Provides fast, predictable performance using **SSDs.**
- Allows you to easily provision and change the **request capacity** needed for each table.
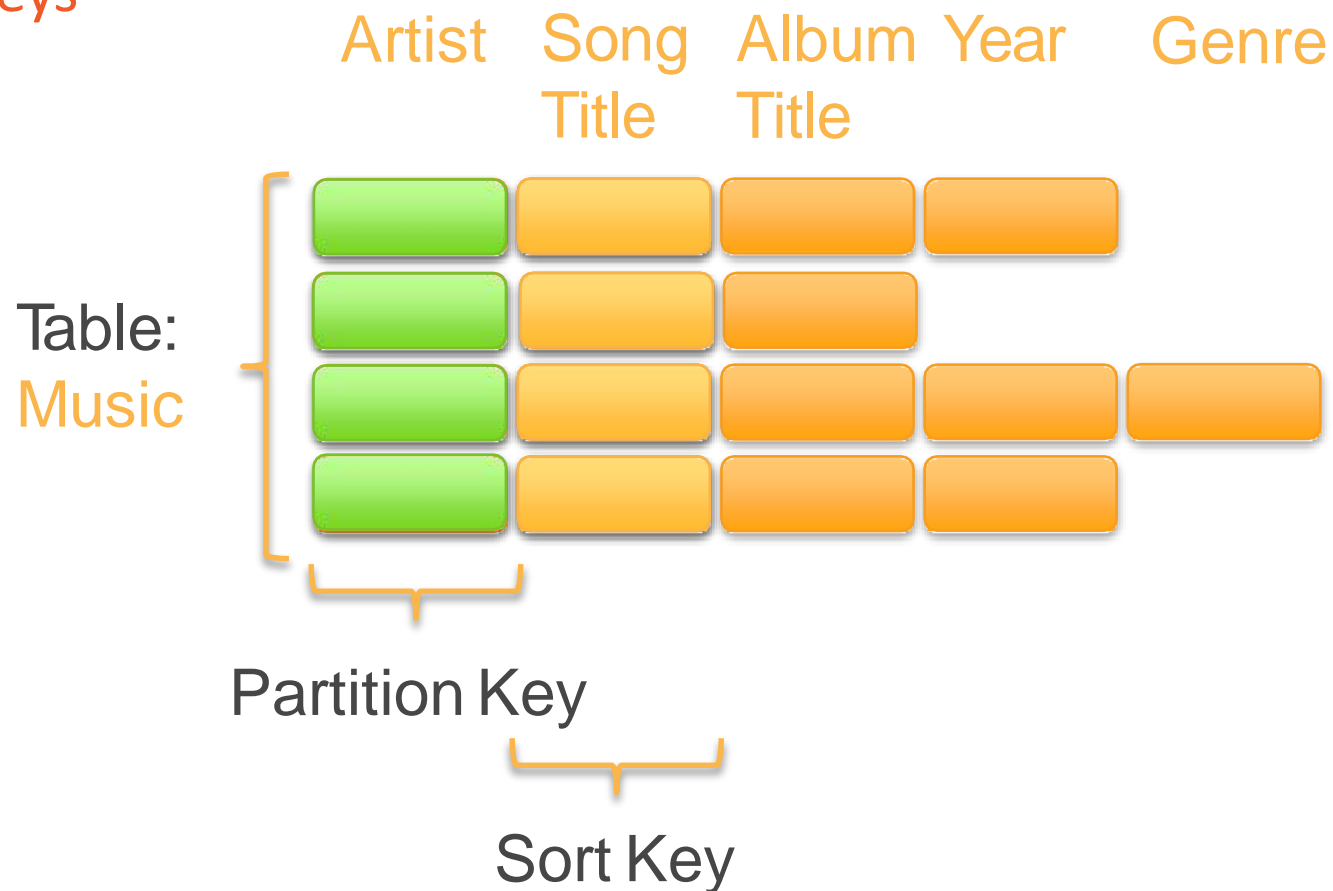- Is a **fully managed**, **NoSQL** database service.

DynamoDB Data Model

Artist Song Title Album Title Year Genre

Table: Music

Items

Attributes (name-value pairs)

Primary Keys

Artist   Song Title   Album Title   Year   Genre

Table: Music

Partition Key

Sort Key

(DynamoDB maintains a sorted index for both keys)

Table: Music
Partition Key: Artist
Sort Key: Song Title

Provisioned Throughput

You specify how much **provisioned throughput capacity** you need for reads and writes.

Amazon DynamoDB allocates the necessary machine resources to meet your needs.

- **Query**:
  - Query a table using the partition key and an optional sort key filter.
  - If the table has a secondary index, query using its key.
  - It is the **most efficient way to retrieve items** from a table or secondary index.
- **Scan**:
  - You can scan a table or secondary index.
  - Scan reads every item – **slower than querying**.
- You can use conditional expressions in both Query and Scan operations.

Simple Application Architecture

AWS

Business logic

Clients

Elastic Load
Balancing

Amazon EC2
app instances
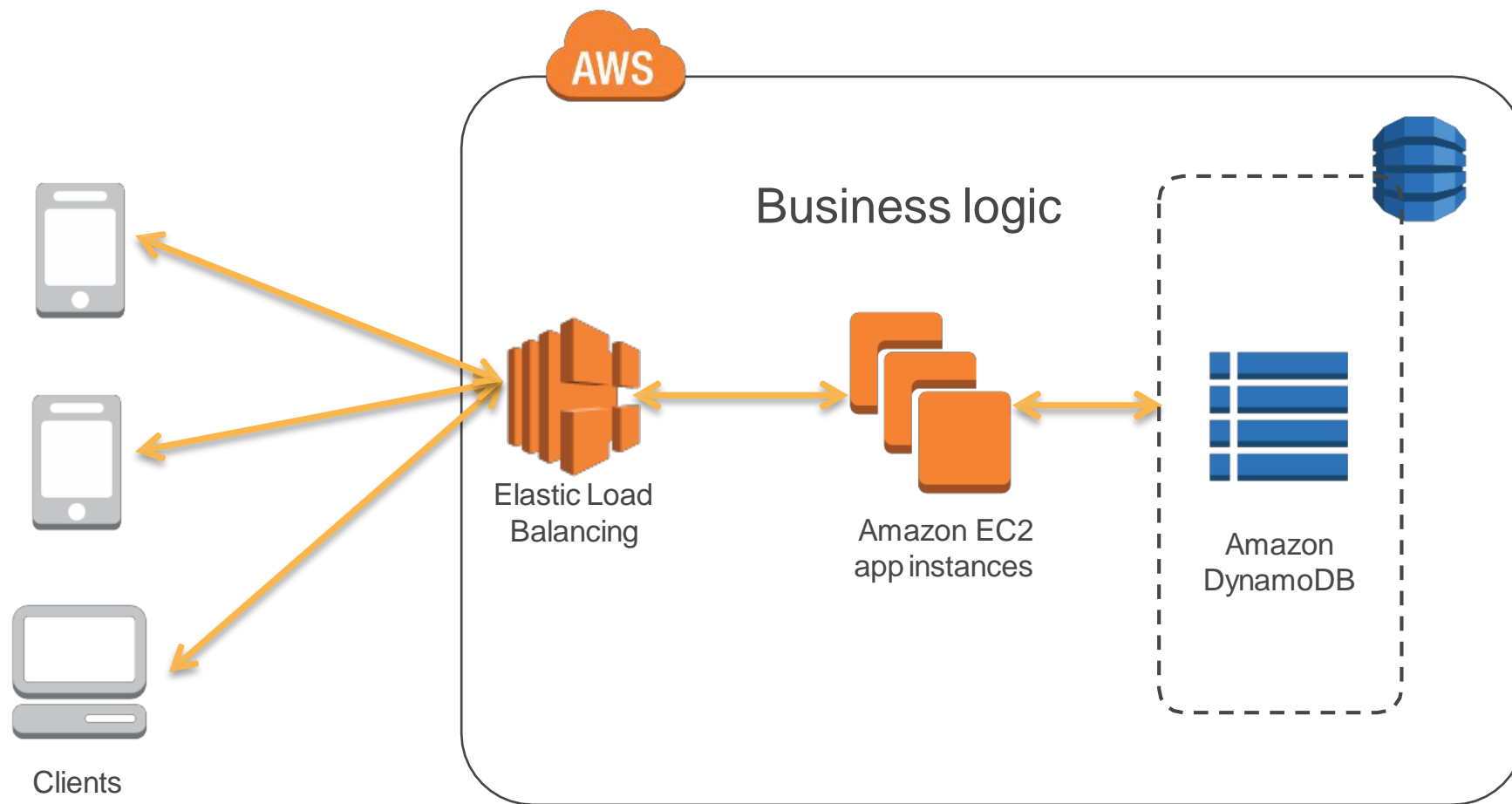
Amazon
DynamoDB

Factors	Relational (Amazon RDS)	NoSQL (Amazon DynamoDB)
Application Type	• Existing database apps • Business process–centric apps	• New web-scale applications • Large number of small writes and reads
Application Characteristics	• **Relational** data models, transactions • **Complex** queries, joins, and updates	• Simple data models, transactions • Range queries, simple updates
Scaling	Application or **DBA–architected** (clustering, partitions, sharding)	**Seamless, on-demand scaling** based on application requirements
QoS	• Performance–depends on data model, indexing, query, and storage optimization • Reliability and availability • Durability	• Performance–**Automatically optimized** by the system • Reliability and availability • Durability

If You Need	Consider Using	
A relational database service with minimal administration	**Amazon RDS** • Choice of Amazon Aurora, MySQL, MariaDB, Microsoft SQL Server, Oracle, or PostgreSQL database engines • Scale compute and storage • Multi-AZ availability	
A fast, highly scalable NoSQL database service	**Amazon DynamoDB** • Extremely fast performance • Seamless scalability and reliability • Low cost	
A database you can manage on your own	Your choice of **AMIs** on Amazon EC2 and Amazon EBS that provide scale compute and storage, complete control over instances, and more.	

Knowledge Check

Q: What are the basic building blocks of Amazon Relational Database Service (RDS)?

DB Instances

True or False: Amazon DynamoDB allows you to store any amount of data with no limits.

True

True or False: Scan is the most efficient way to retrieve items from a DynamoDB table.

False

Q: You are creating a resilient, durable application using Amazon RDS.   In addition to Amazon RDS's automatic backups, what feature should you use to ensure that your backups are durable retained?

Manual Snapshots

Understanding AWS Relational Database Services (RDS)

RDS Characteristics

Database engine managed by AWS

MySQL, Oracle, Microsoft SQL, PostgreSQL, MariaDB, or Amazon Aurora

Multi-AZ deployment options

On-demand and reserved instance pricing

Magnetic, GP-SSD, or PIOPS

Oracle and Microsoft SQL licensing:
- Included Licenses
- Bring your own licensing

Automated or manual backups

RDS Automated Backups

Continuously tracks changes and backs up your DB

Volume snapshot of your entire DB instance, not just DBs

One day of backups retained by default but can be configured up to 35 days

Backup retention period defined during configuration

When you delete an RDS instance, all automated snapshots are deleted

- Manual snapshots are preserved

RDS Automated Backups

Automated backups occur daily during a 30 minute configurable backup window

Automated backups are preserved for a configurable number of days (retention period)

RDS Restore

You cannot restore from a DB snapshot to an existing DB instance

- A new DB instance is created when you restore

Only default DB parameters and security groups are restored

- You must manually associate all other DB parameters and SGs

RDS combines daily backups in conjunction with transaction logs to restore the DB Instance to any point during the retention period

Up to the last five minutes

Multi-AZ Failover

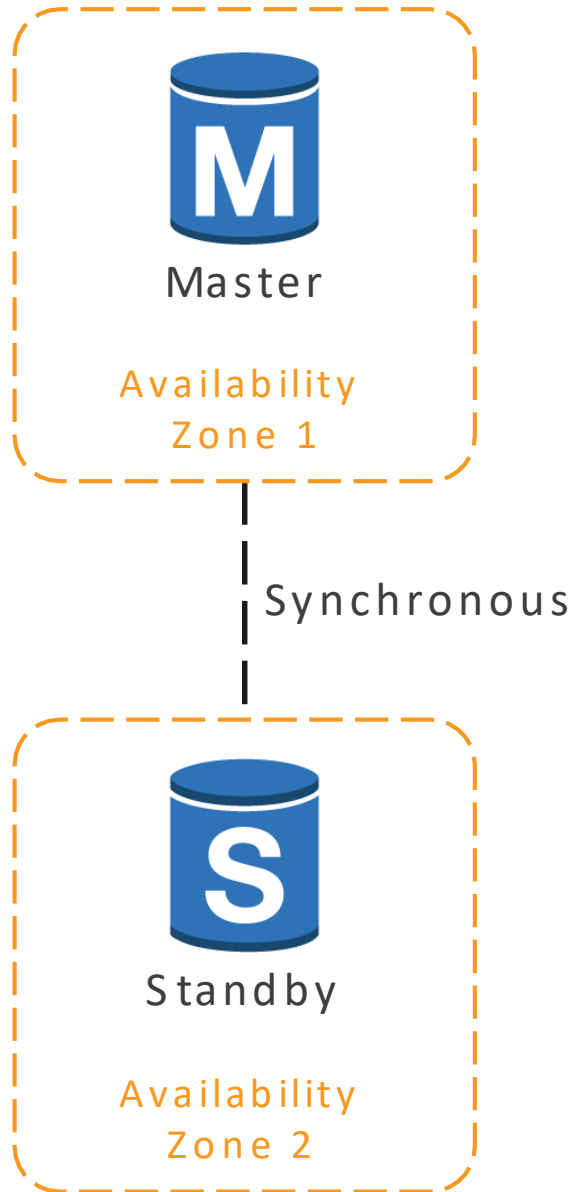Multi-AZ RDS deployment designed for HA
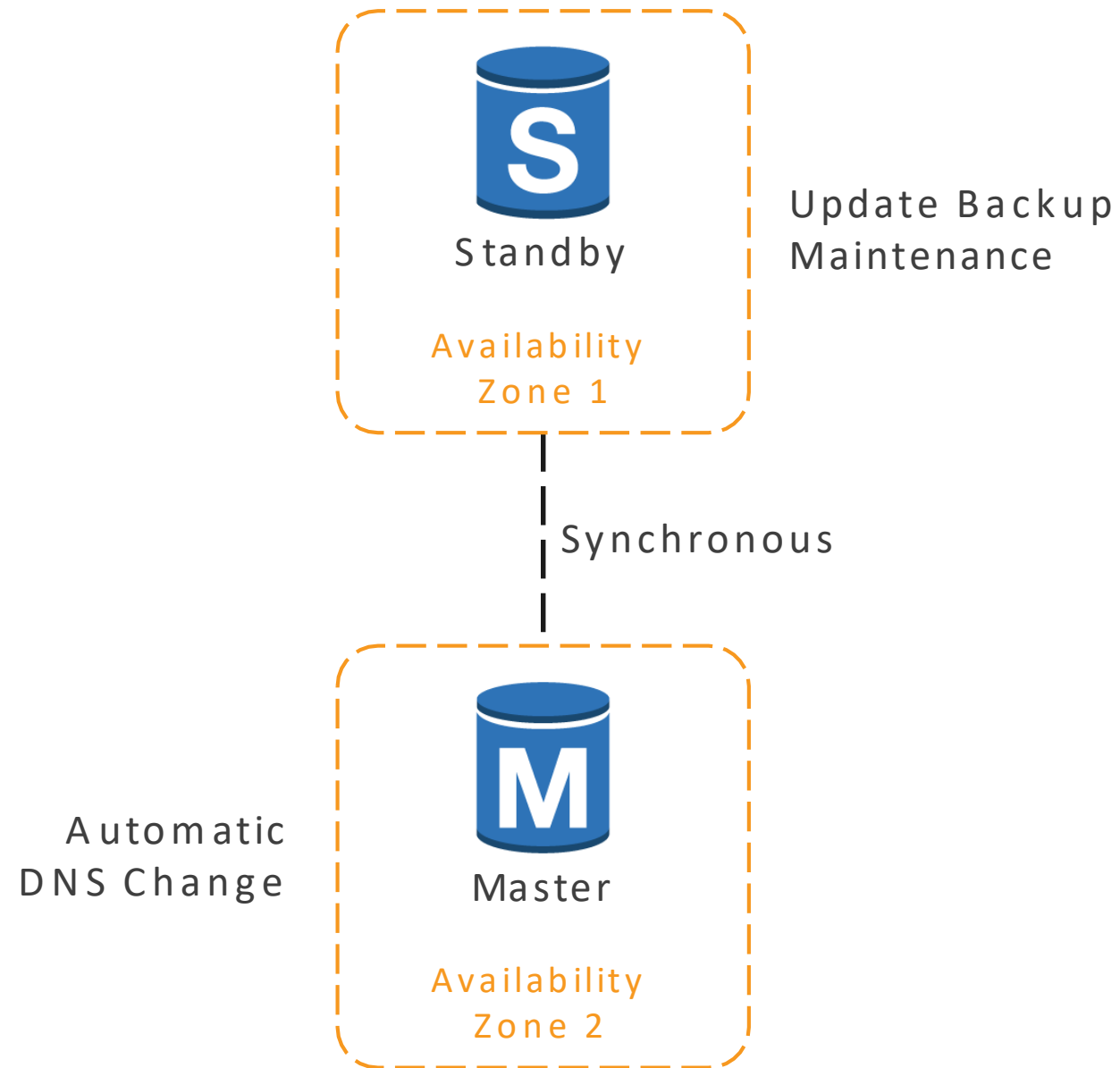
Synchronous replica in secondary AZ

Standby replica RDS instance is invisible
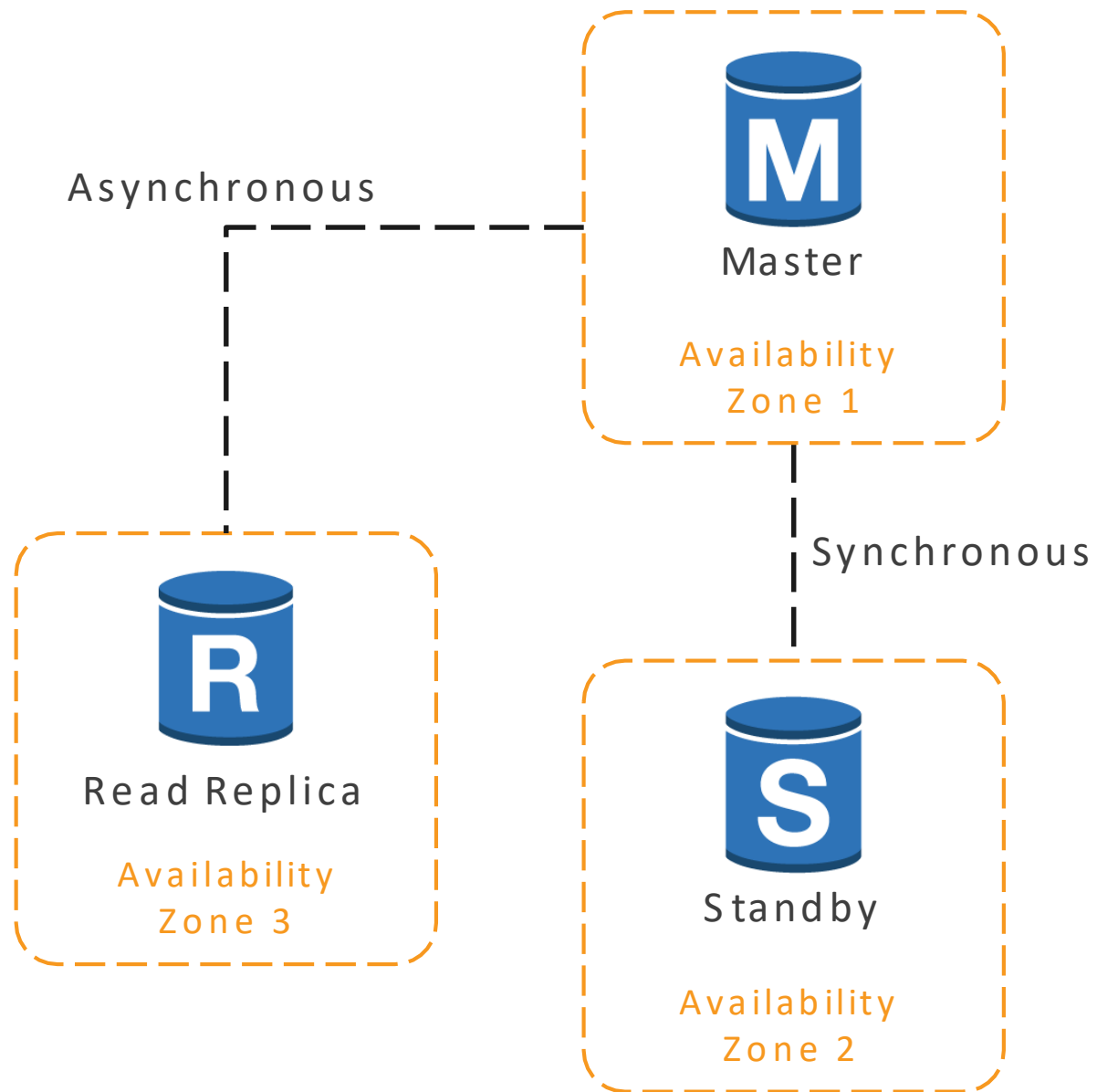
DB snapshots always taken against standby instance

AWS automatically adjusts DNS record when needed

Multi-AZ is different from a RDS read replica

RDS Read Replicas

Read replicas designed for workload sharing / offloading

Created from a snapshot of the master instance

Asynchronous replication / Read-only connections

Read-only disaster recovery

RDS Reserved Instances

DB engine

DB instance class

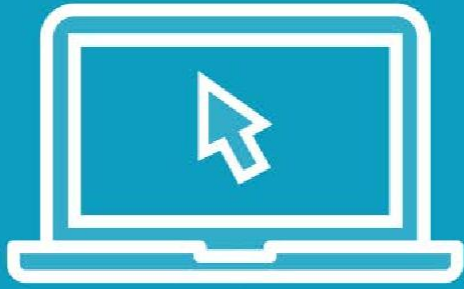Deployment type

License model

Region

RDS Reserved Instances

Move between Azs in the same Region

Are available for Multi-AZ deployments

Can be applied to Read Replicas provided the DB Instance class and Region are the same

Demo

Creating an RDS instance

Windows Integrated Authentication

Choose one of the AWS offered directory services

Establish a trust relationship

Windows integrated authentication only works with a domain created using AWS directory service

Alternatively, you can use SQL authentication

Summary

RDS characteristics

Multi-AZ failover

RDS read replicas

RDS reserved instances

Understanding Amazon DynamoDB and Redshift

Amazon DynamoDB

Fully managed, highly available and scalable NoSQL

    Automatically and synchronously replicates data across three Availability Zones

SSDs and limiting indexing on attributes provides high throughput and low latency

ElastiCache can be used in front of DynamoDB

- Offload high amounts of reads for non- frequently changed data

Amazon DynamoDB

Ideal for existing or new applications that need:

- Flexible NoSQL database with low read and write latencies
- Ability to scale storage and throughput up or down as needed without code changes or downtime
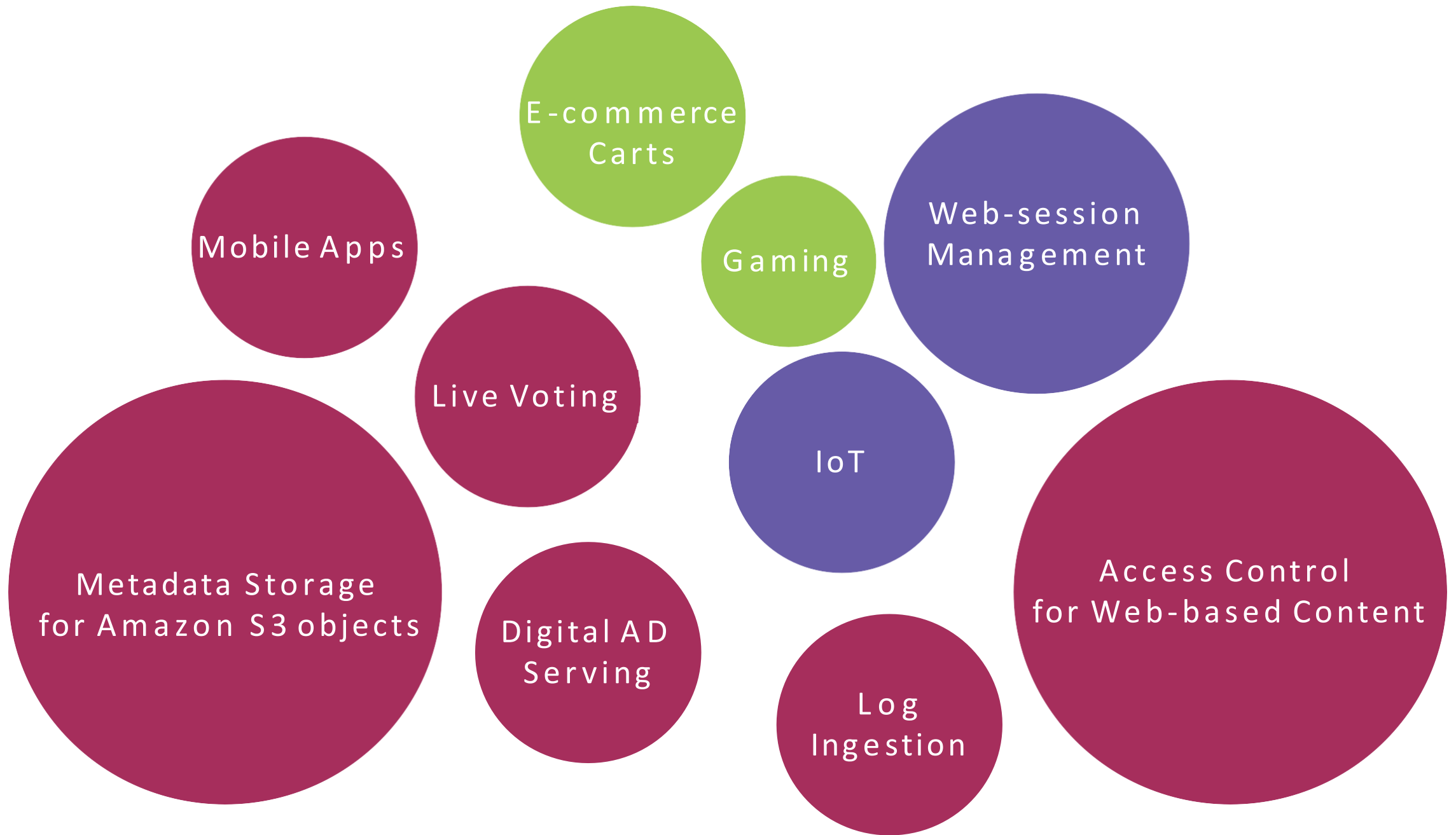
Non-ideal DynamoDB Scenarios

Prewritten application tied to a traditional relational database

Joins and/or complex transactions

BLOB data

Large data with low I/O rate

DynamoDB Integration

Amazon Elastic MapReduce	Amazon Redshift	Amazon Data Pipeline
Amazon S3	Management Console and APIs	

Amazon DynamoDB

Stores structured data in tables, indexed by a primary key

Tables are a collection of items and Items are made up of attributes (columns)

Primary key can be:
- Single-attribute hash key
- Composite hash-range key
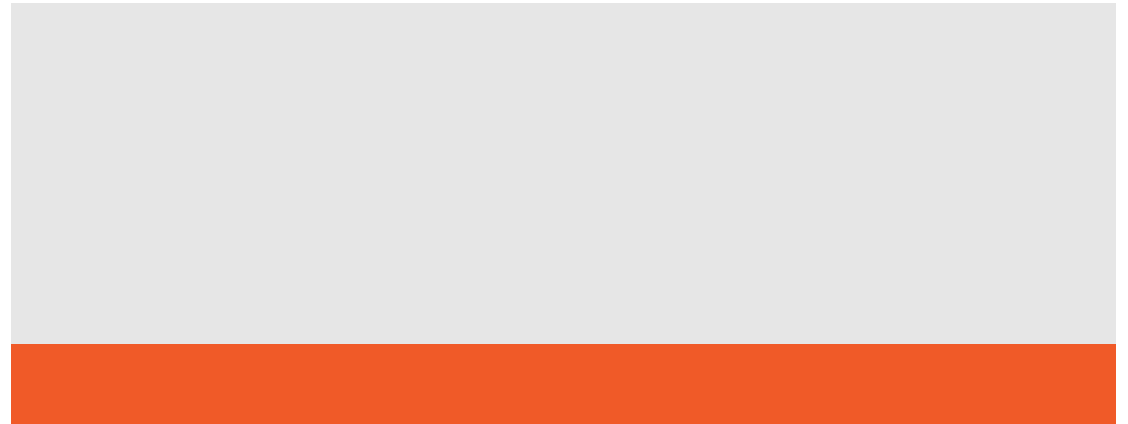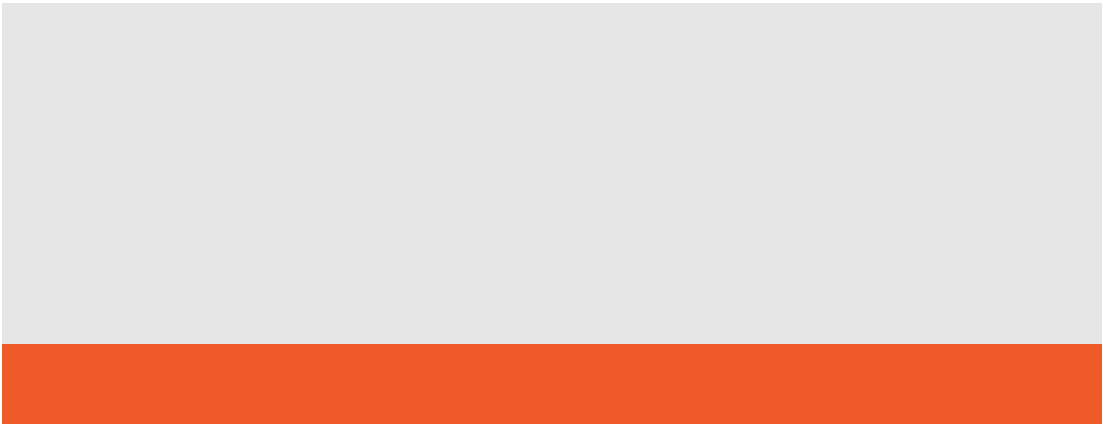
DynamoDB Features

Secondary indexes

Streams

Cross-region replication

Triggers

Schema-less

Two Ways to Search

ElastiCache

Open-source in-memory caching engines

- Memcached
  - Widely adopted memory object caching system
- Redis
  - Popular open-source in-memory key-value store
  - Supports data structures such as sorted sets and lists

Master / Slave replication and Multi-AZ

- Can be used to achieve cross AZ redundancy

Memcached vs. Redis

	Memcached	Redis
Cache to offload DB	√	√
Multithreaded performance	√	X
Horizontal scaling	√	X
Multi-AZ	X	√
Backup and restore	X	√
Pub/sub functionality	X	√
Sorting and ranking	X	√
Advanced data types	X	√
Persistence	X	√

Suggested Reading

Performance at Scale with Amazon ElastiCache:

https://d0.awsstatic.com/whitepapers/performance-at-scale-with-amazon-elasticache.pdf

Amazon Redshift

Fast and fully managed petabyte-scale relational data warehouse service

Analyze all your data using your existing business intelligence tools

HDD and SSD Platforms

Starts at $0.25/hour

Scale to $1,000/TB/Year
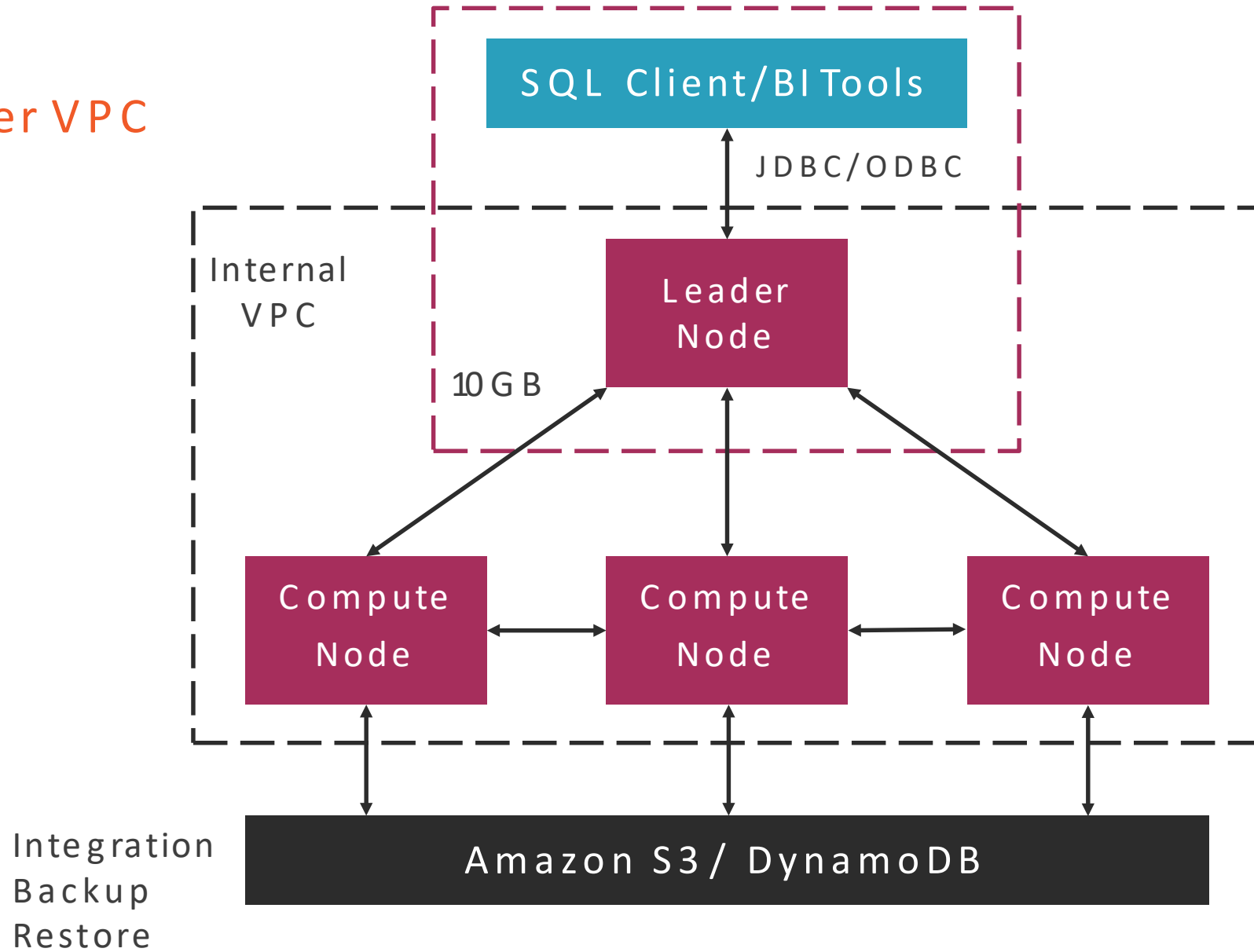
Amazon Redshift Architecture

Leader Node

- Simple SQL end point
- Stores metadata
- Optimizes query plan
- Coordinates query execution

Compute Nodes

- Local columnar storage
- Parallel/distributed execution of all queries, loads, backups, restores, resizes

Customer VPC

SQL Client/BI Tools

JDBC/ODBC

Internal VPC

10 GB

Leader Node

Compute Node

Compute Node

Compute Node

Integration Backup Restore

Amazon S3 / DynamoDB

Backup and Fault Tolerance

Continuous/incremental backups

- Multiple copies within cluster
- Continuous and incremental backups to S3
- Continuous and incremental backups across regions
- Streaming restore

Backup and Fault Tolerance

Fault tolerance
- Disk failures
- Node failures
- Network failures
- Availability Zone/Region level disasters

Security

Load encrypted from S3

SSL to secure data in transit Amazon VPC for network

isolation Encryption to secure

data at rest

Audit logging and AWS CloudTrail integration

SOC 1/2/3, PCI-DSS, FedRAMP, BAA

Amazon DynamoDB

Amazon Redshift