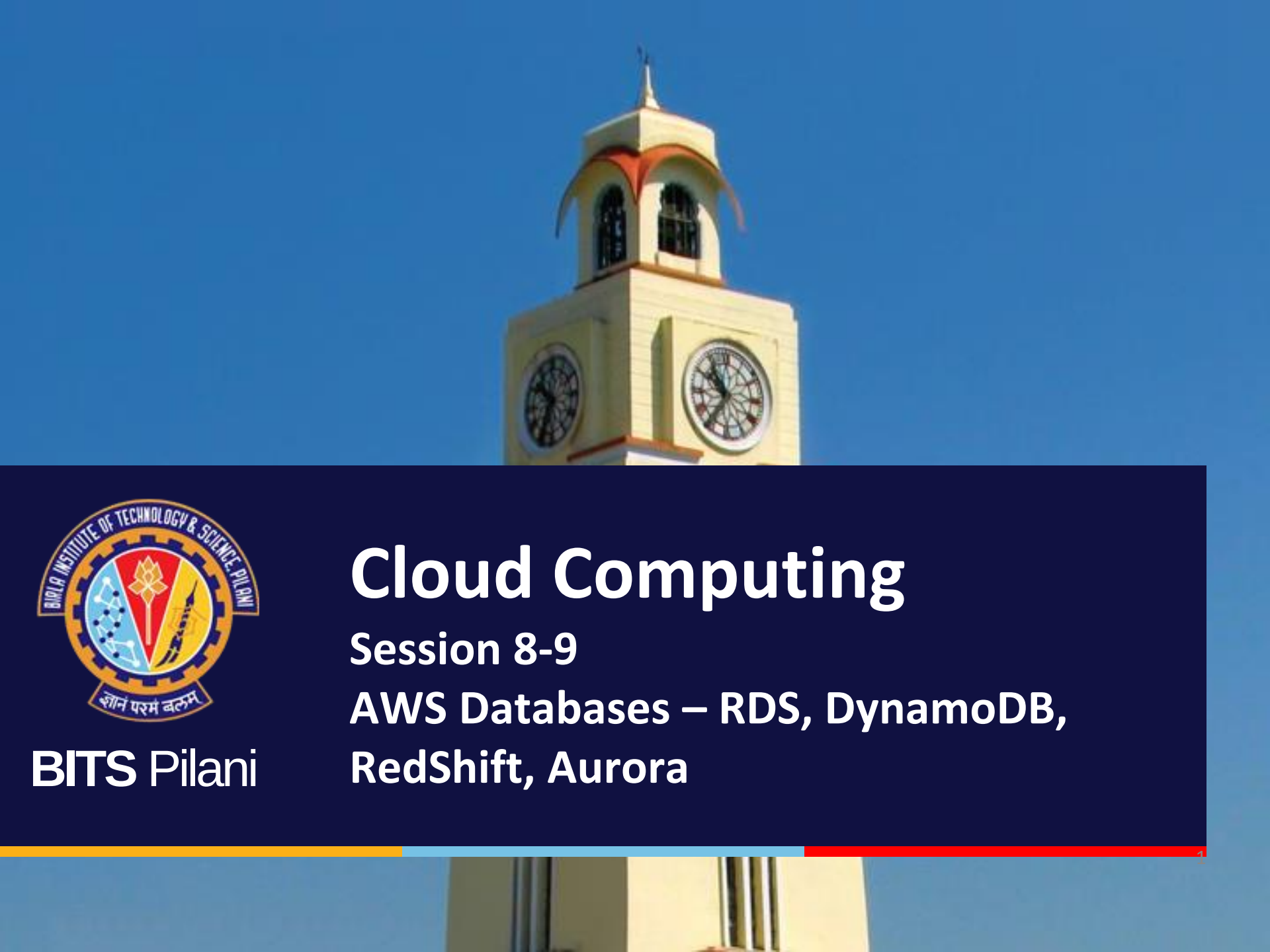


BITS Pilani

Cloud Computing

Session 8-9

AWS Databases – RDS, DynamoDB, RedShift, Aurora

Agenda

- Amazon Relational Database Service (Amazon RDS)
- Amazon DynamoDB
- Amazon Redshift
- Amazon Aurora

Revisit Database

- Organized collection of data / information
- Electronically on a computer system
- Two Broad Types
  - Relational Database
    - E.g: SQL
  - Non Relational Database
    - E.g: MongoDB

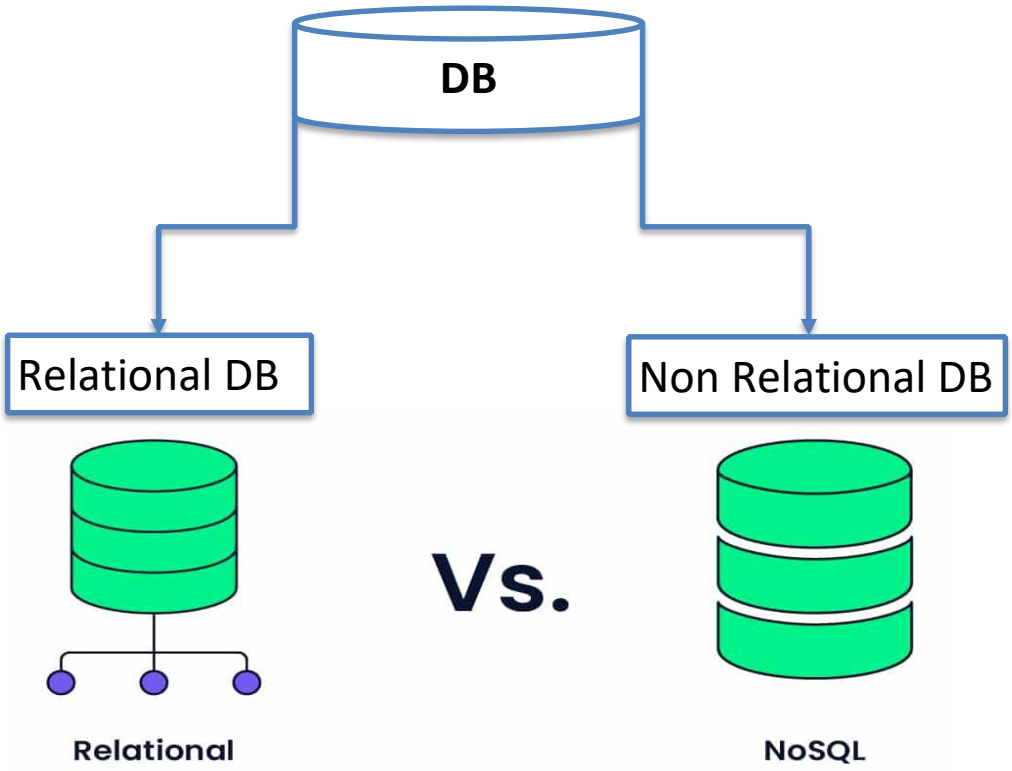


Image Courtesy: Data-sleek.com

Relational Database

- Schema oriented
- Structured collection of data
- Data organized into tables
- Tables are linked together by relationships
- Relationships are defined as
  - one-to-one,
  - one-to-many, or
  - many-to-many.
- Supports **ACID** Properties.
  - Atomicity, Consistency, Isolation, and Durability

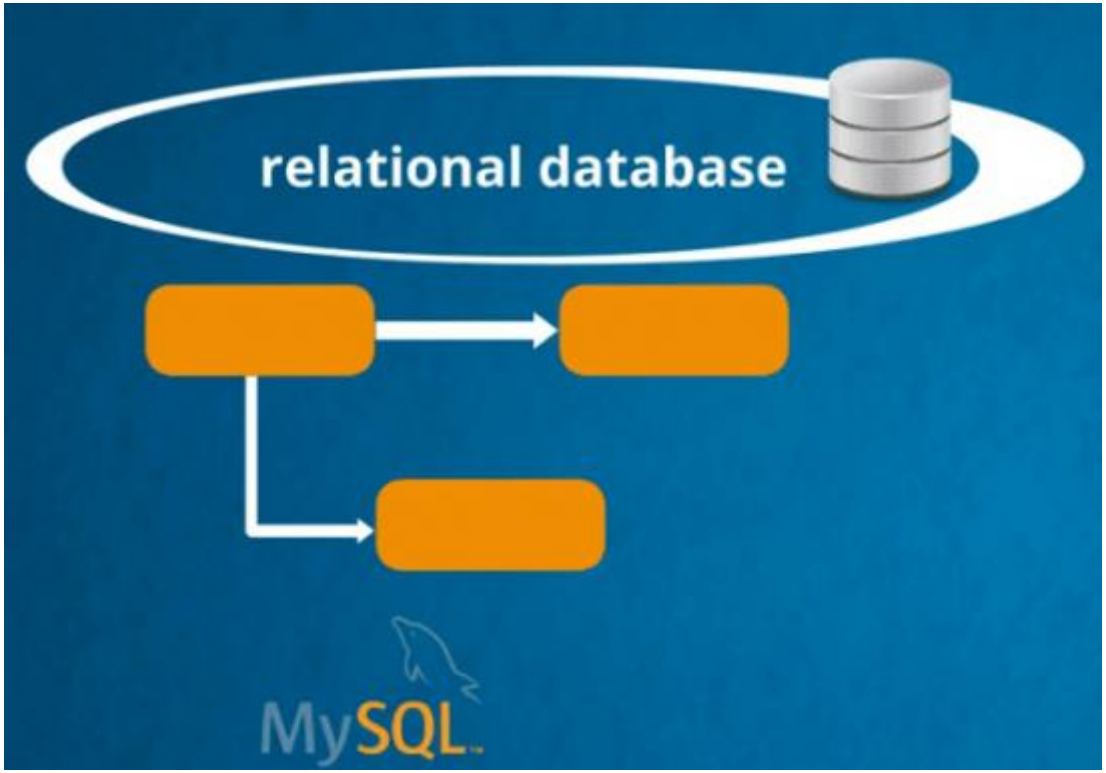


Image Courtesy: 365datascience.com

Non Relational Database

Image Courtesy: 365datascience.com

- Independent collection of data / information
- No fixed schema / dynamic schema
- Documents, key-value stores
- Large volumes of unstructured data and real-time web applications
- Inclined towards evaluating and supporting CAP Theorem
  - Consistency
  - Availability, and Partition Tolerance

Relational Vs Non Relational Database

	Relational (SQL)	Non-Relational
Data Storage	Rows and columns	Key-value, document, graph
Schemas	Fixed	Dynamic
Querying	Uses SQL	Focuses on collection of documents
Scalability	Vertical	Horizontal
Example	<table><tr><th>ISBN</th><th>Title</th><th>Author</th><th>Format</th></tr><tr><td>3111111223439</td><td>Withering Depths</td><td>Jackson, Mateo</td><td>Paperback</td></tr><tr><td>3122222223439</td><td>Wily Willy</td><td>Wang, Xiulan</td><td>Ebook</td></tr></table>	{ ISBN: 3111111223439, Title: "Withering Depths", Author: "Jackson, Mateo", Format: "Paperback" }

Database Systems

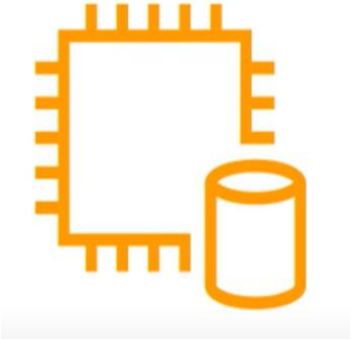
On-Premise setup requires

- Setup the infrastructure (CAPEX)
- Maintenance (OPEX)
- Upgrades
- Scalability
- High Availability and recovery

Database (DB) on EC2 Vs DB as a Service?

Benefits

- Customization
- Controlling Infrastructure
- High Availability



Problems ? What ?

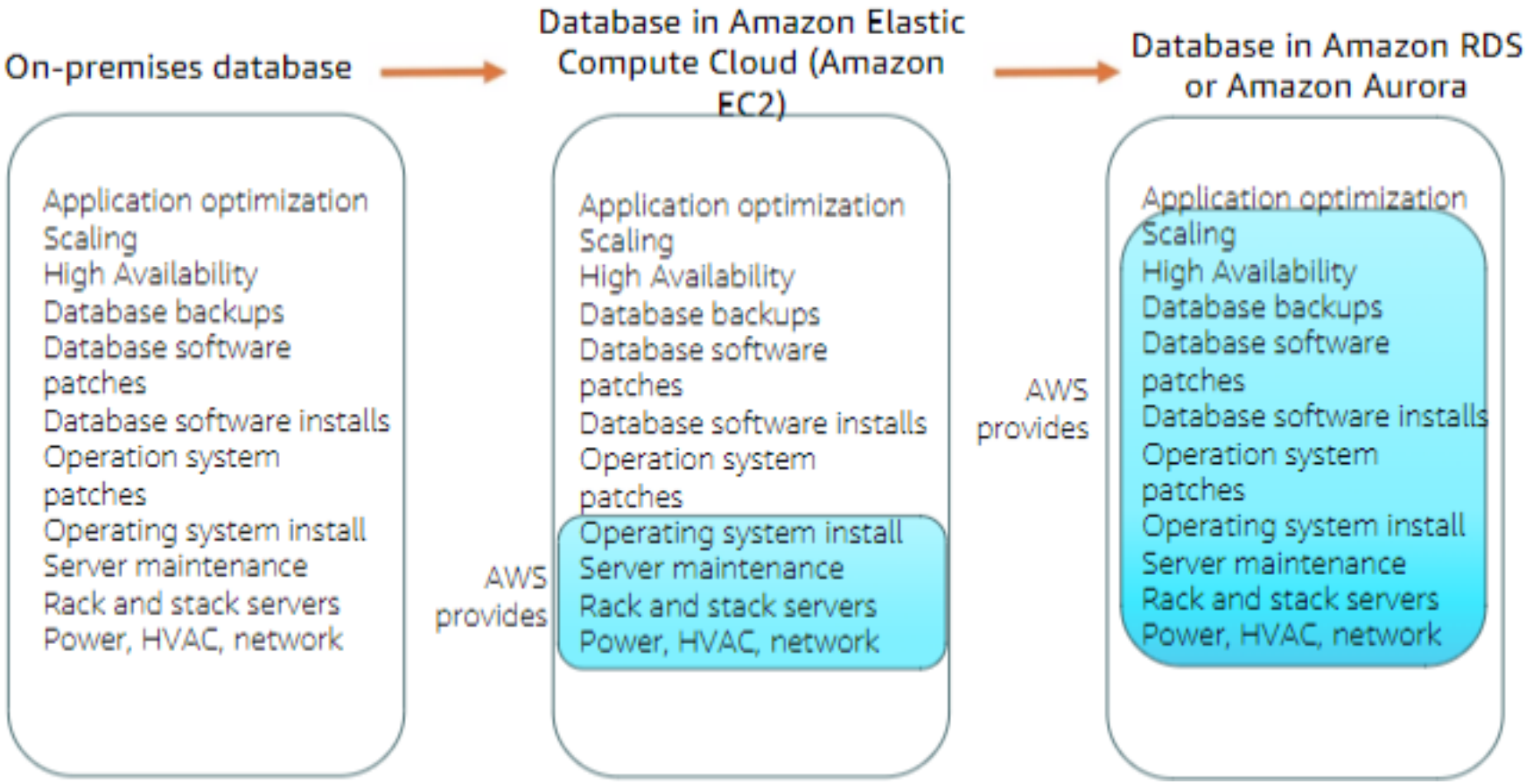
- Manually setup and configure
- Operational overhead
- Resource management

Managed Service for DB Functions

- Scaling,
- Fault tolerance, and
- Availability

Are typically built into the service.

From On-premises DBs to Amazon RDS



Managed Services Responsibilities

You manage:

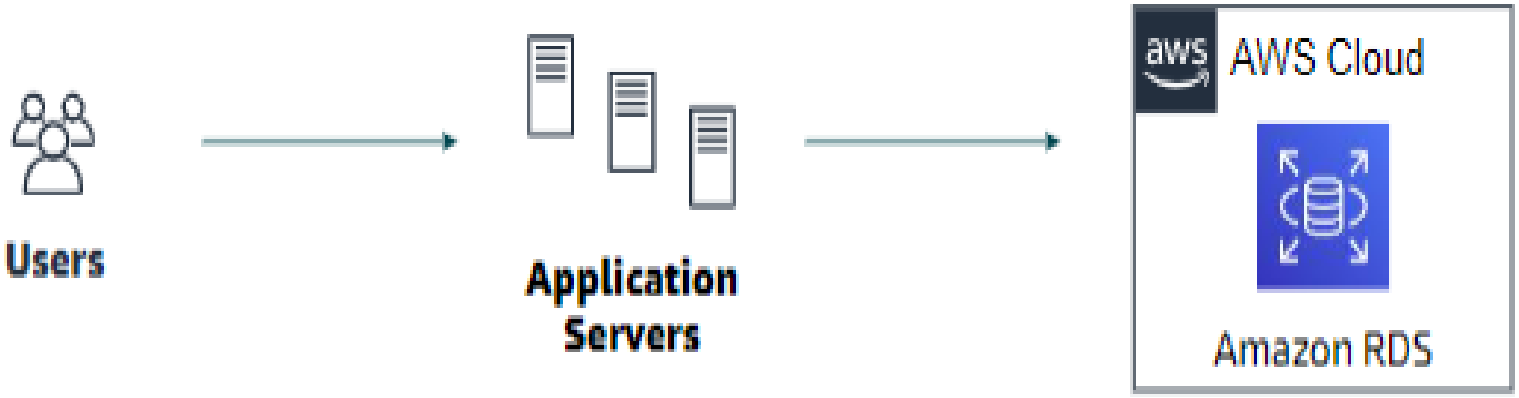
- Application optimization

AWS manages:

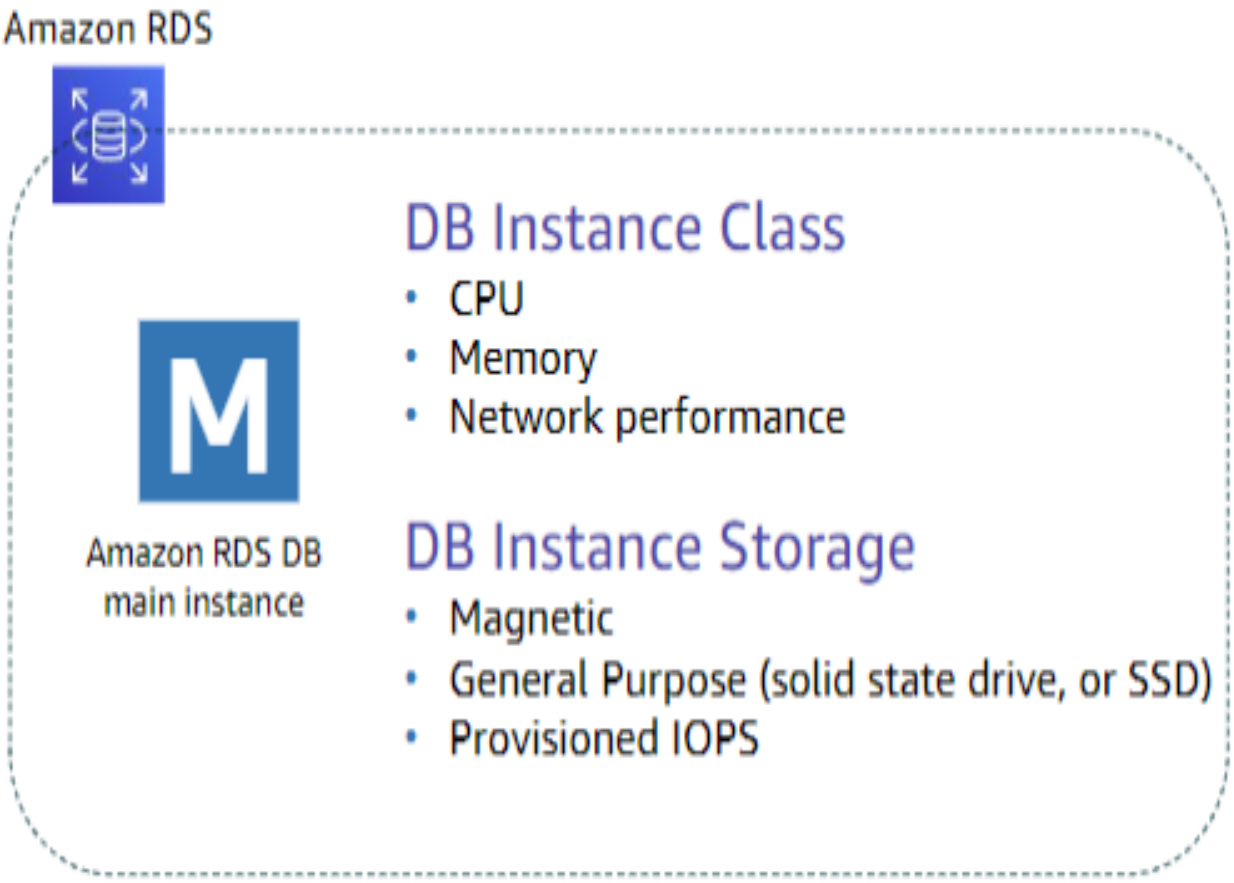
- OS installation and patches
- Database software installation and patches
- Database backups
- High availability
- Scaling
- Power and racking and stacking servers
- Server maintenance

Amazon RDS

Managed service that sets up and operates a relational database in the cloud.



Amazon RDS DB Instances



MySQL

Amazon Aurora

Microsoft SQL Server

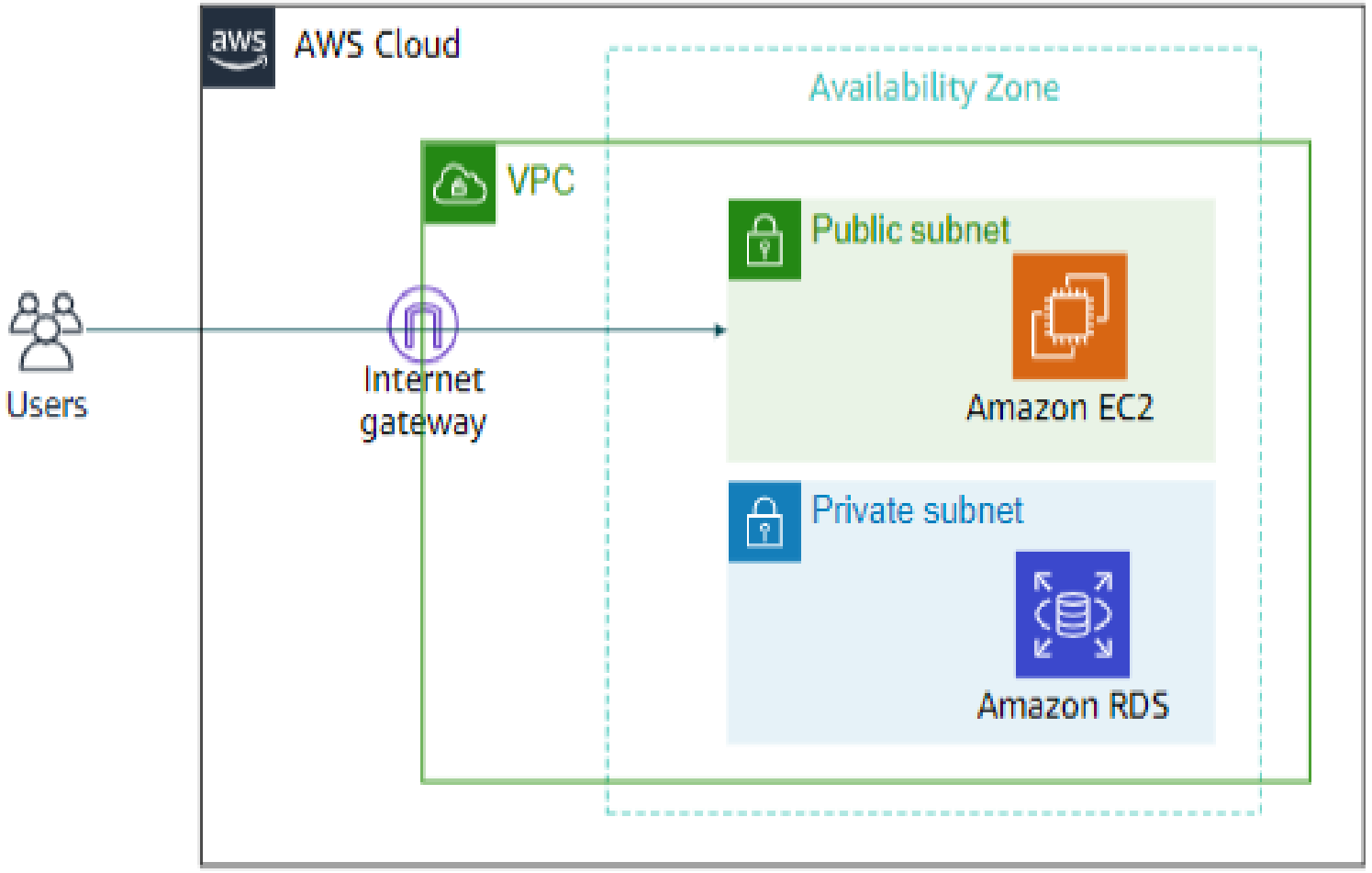
PostgreSQL

MariaDB

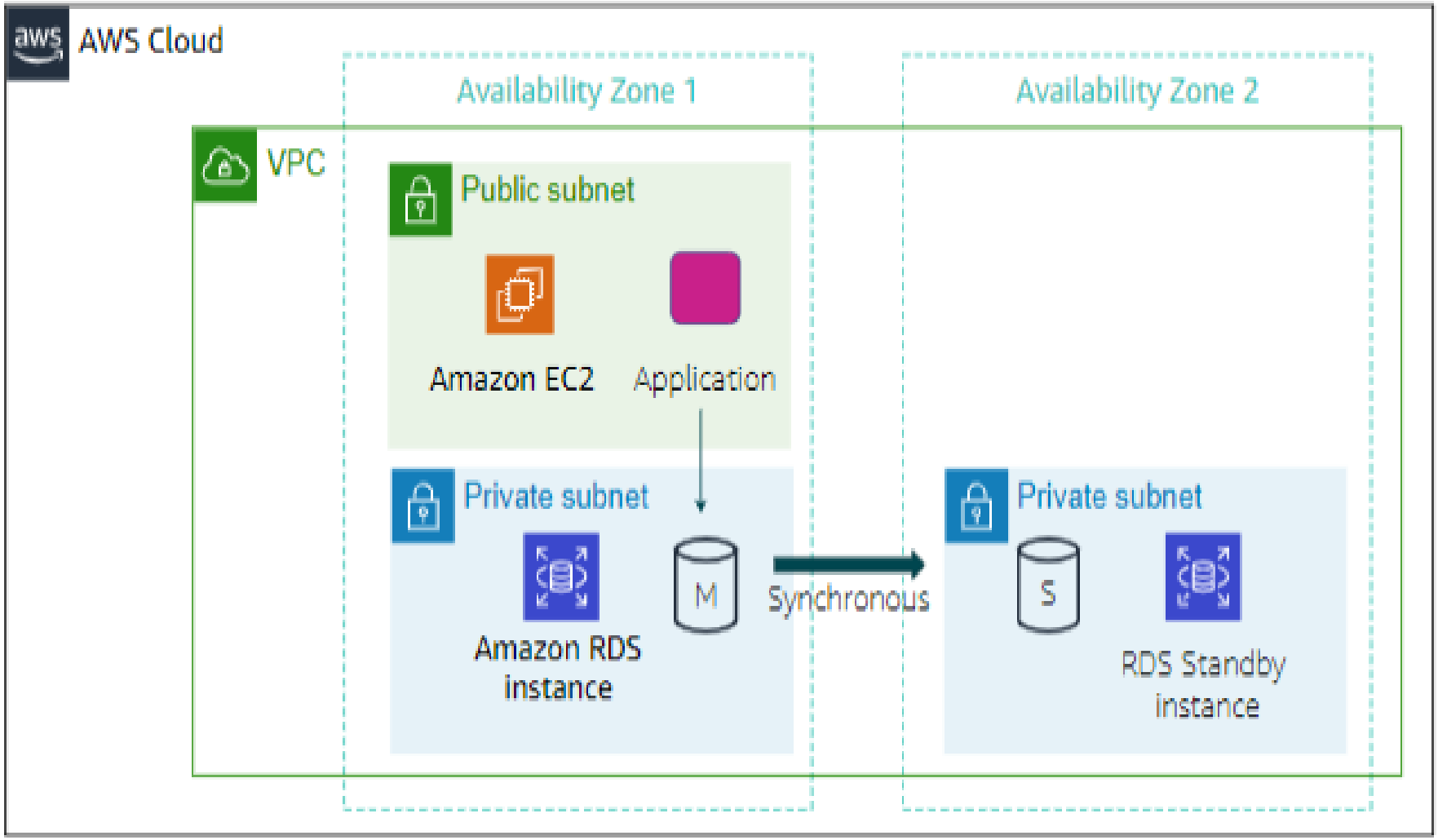
Oracle

DB engines

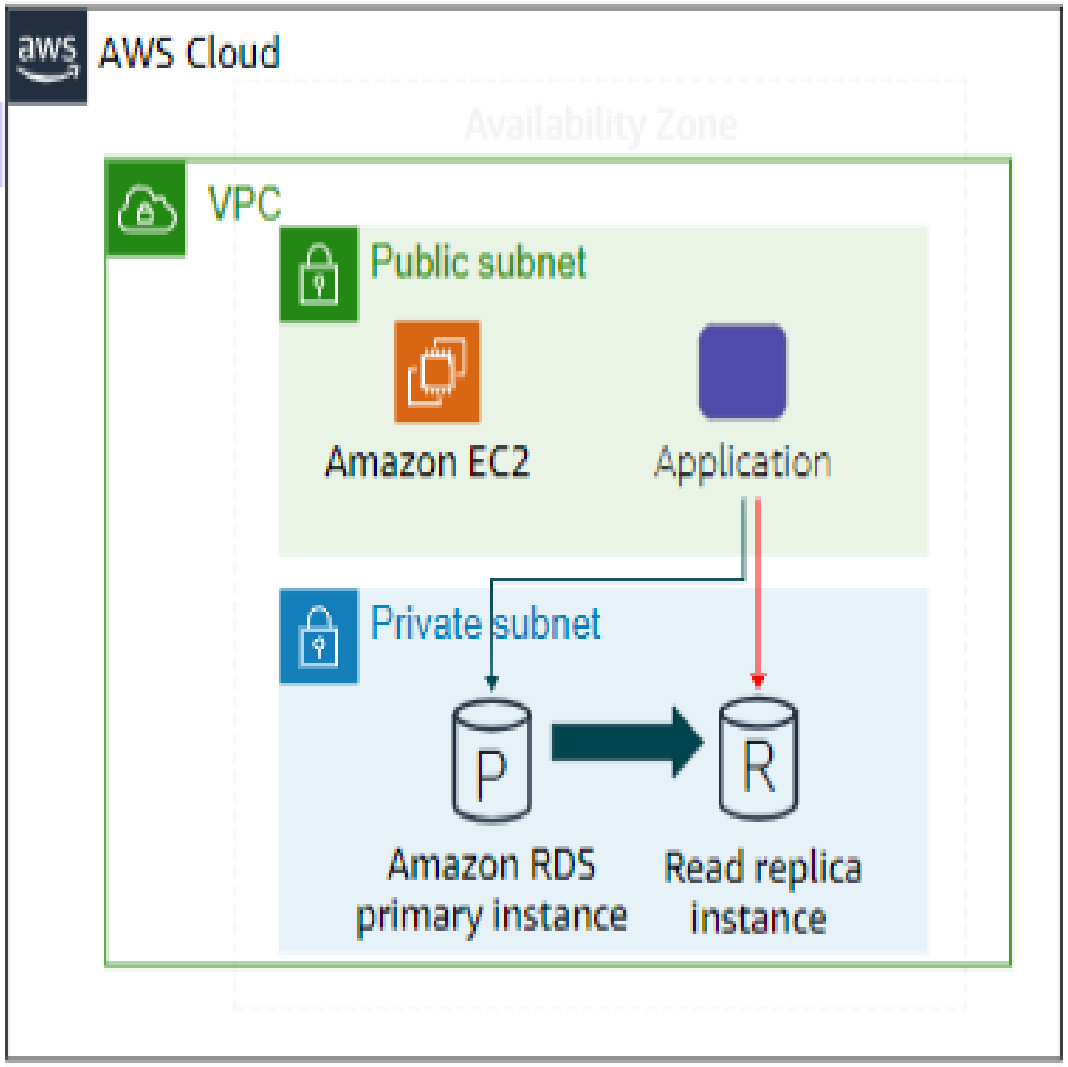
Amazon RDS in a virtual private cloud (VPC)



High Availability with Multi-AZ deployment(1/2)



Amazon RDS Read Replicas



Features & Functions

- Offers asynchronous replication
- Use for read-heavy database workloads
- Offload read queries
- Can be promoted to primary if needed

Amazon RDS: Billing & Storage

Clock-hour billing: Resources incur charges when running

- **Provisioned Storage**
  - No charge
    - Backup storage of up to 100 percent of database storage for an active database
  - Charge (GB/month)
    - Backup storage for terminated DB instances

- **Additional Storage**
  - Charge (GB/month)
    - Backup storage in addition to provisioned storage

Amazon RDS: Deployment type and data transfer

- **Requests**
  - The number of input and output requests that are made to the database
- **Deployment type**
  - Storage and I/O charges vary, depending on whether you deploy to
    - Single Availability Zone
    - Multiple Availability Zones
- **Data transfer**
  - No charge for inbound data transfer
  - Tiered charges for outbound data transfer

Use cases

Web and mobile applications	**✓High throughput** **✓Massive storage scalability** **✓High availability**
E-commerce applications	✓Low-cost database ✓Data security ✓Fully managed solution
Mobile and online games	✓Rapidly grow capacity ✓Automatic scaling ✓Database monitoring

What is Amazon DynamoDB?

Fast and flexible NoSQL database service for any scale

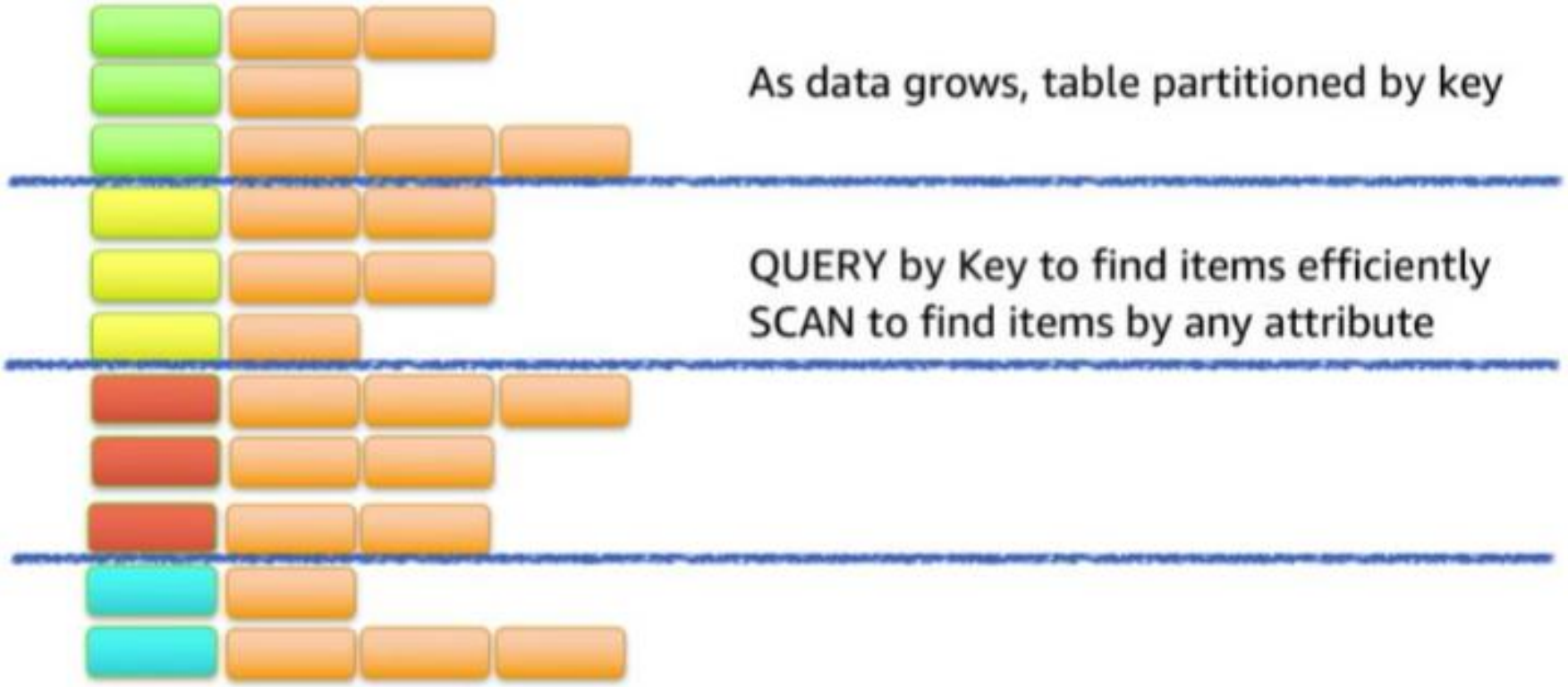
- NoSQL database tables
- Virtually unlimited storage
- Items can have differing attributes
- Low-latency queries
- Scalable read/write throughput

Amazon DynamoDB Core Components

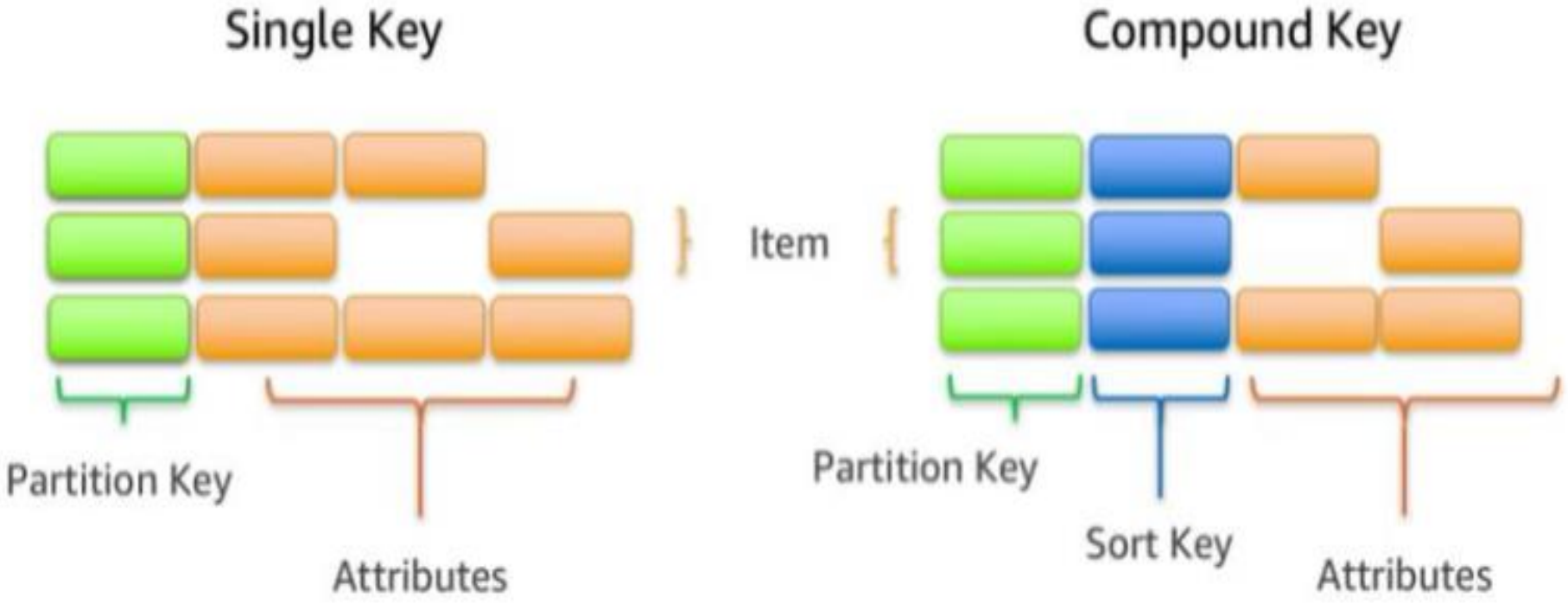
- Tables,
- Items, and
- Attributes

- Two different kinds of primary keys:
  - Partition key and
  - Sort key - optional

Partitioning

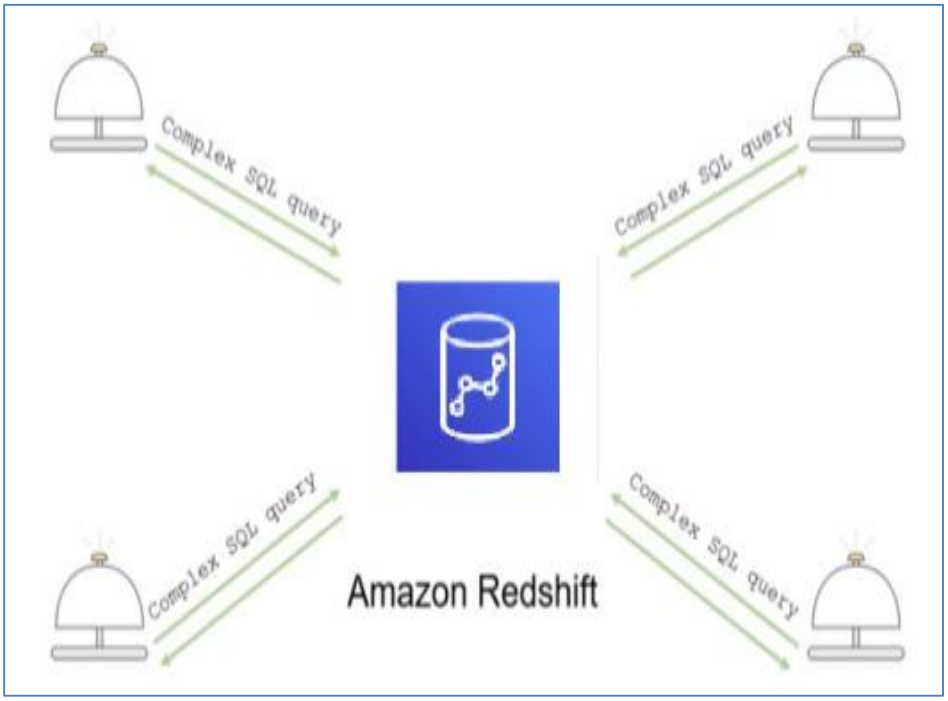


Items in a table must have a key



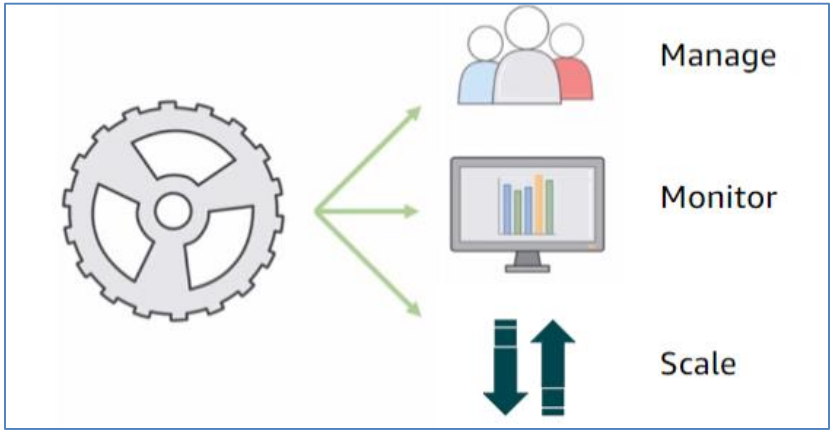
Amazon Redshift

Amazon Redshift



- It is a relational database service
- Peta bytes scale data warehouse service in cloud
- Massive Parallel Processing (MPP)

- Business intelligence



Row-wise Storage

Amazon Redshift

SSN	Name	Age	Addr	City	St
101259797	SMITH	88	899 FIRST ST	JUNO	AL
892375862	CHIN	37	16137 MAIN ST	POMONA	CA
318370701	HANDU	12	42 JUNE ST	CHICAGO	IL

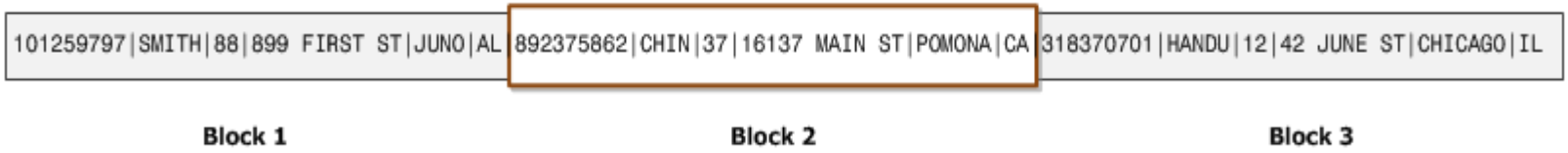


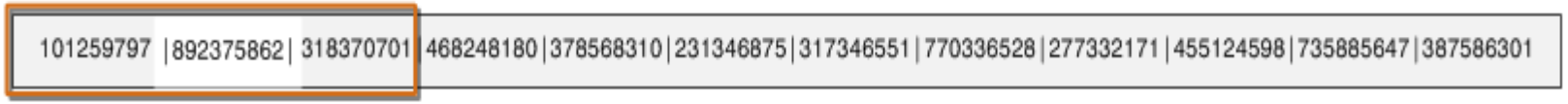
Image Courtesy: Columnar storage - Amazon Redshift

- Data blocks store values sequentially for each consecutive column of a row
- Row-wise storage okay for On Line Transaction Processing – reading/writing all values of a record/row.

Columnar Storage

Amazon Redshift

SSN	Name	Age	Addr	City	St
101259797	SMITH	88	899 FIRST ST	JUNO	AL
892375862	CHIN	37	16137 MAIN ST	POMONA	CA
318370701	HANDU	12	42 JUNE ST	CHICAGO	IL



Block 1

Image Courtesy: [Columnar storage - Amazon Redshift]

- Each data block stores values of a single column for multiple rows
- Efficient
  - Less no of I/O operations to fetch a specific column values for huge rows,
  - Saves space as well.
- Aids OLAP

Amazon Redshift

Amazon Redshift

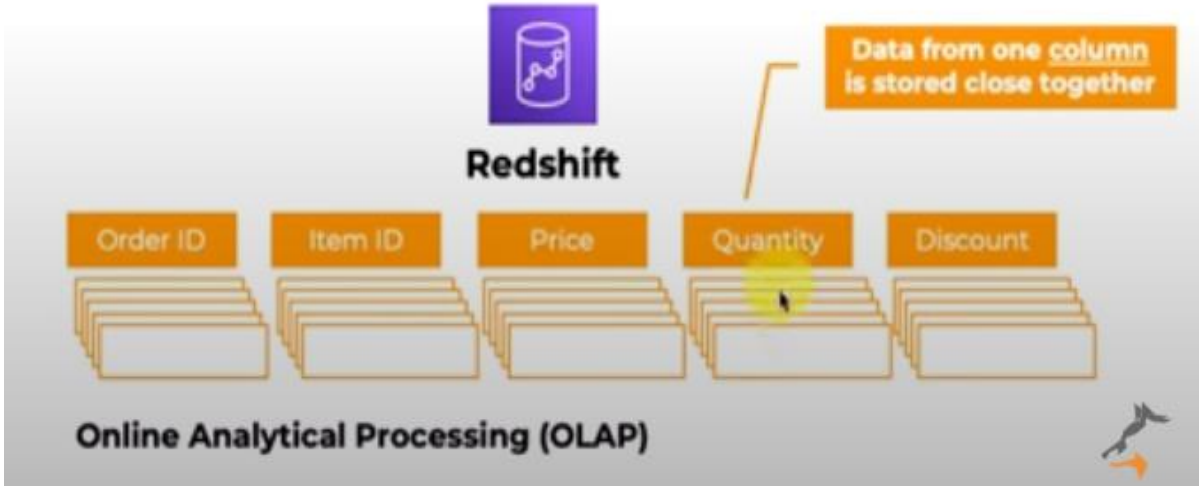


Image Courtesy: cloudwolf.com

- Based on columnar storage - Data from each column stored closer
- **Redshift** designed for On Line Analytical Processing (OLAP).
- **Redshift** provides SQL capability designed for fast OLAP of very large datasets that are stored in both **Redshift** clusters and AWS S3 data lakes.

Amazon Redshift use cases

Enterprise DataWare house (EDW)

- Migrate at a pace that customers are comfortable with
- Respond faster to business needs

Big data

- Low price point for small customers
- Managed service for ease of deployment and maintenance
- Focus more on data and less on database management

Software as a service (SaaS)

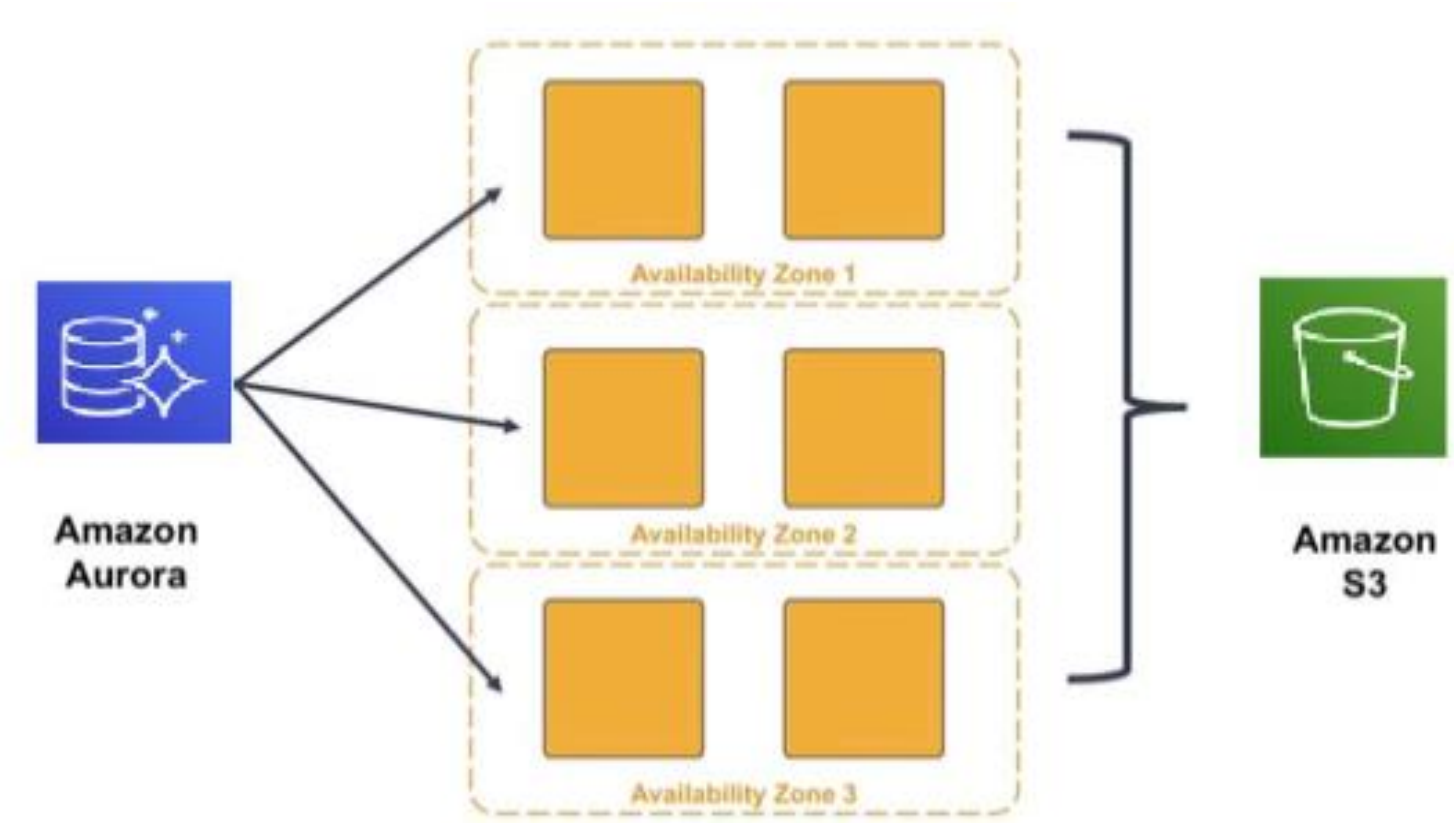
- Scale the dataware house capacity as demand grows
- Add analytic functionality to applications
- Reduce hardware and software costs

Amazon Aurora

- Enterprise-class relational database
- Compatible with MySQL or PostgreSQL
- 5x the throughput of MySQL and 3x of PostgreSQL
- 99.999% multi-Region availability.
- Automate time-consuming tasks
  - Provisioning,
  - Patching,
  - Backup,
  - Self-healing: Recovery, failure detection, and repair.

Amazon Aurora Service Benefits

High Availability



The right tool for the right job

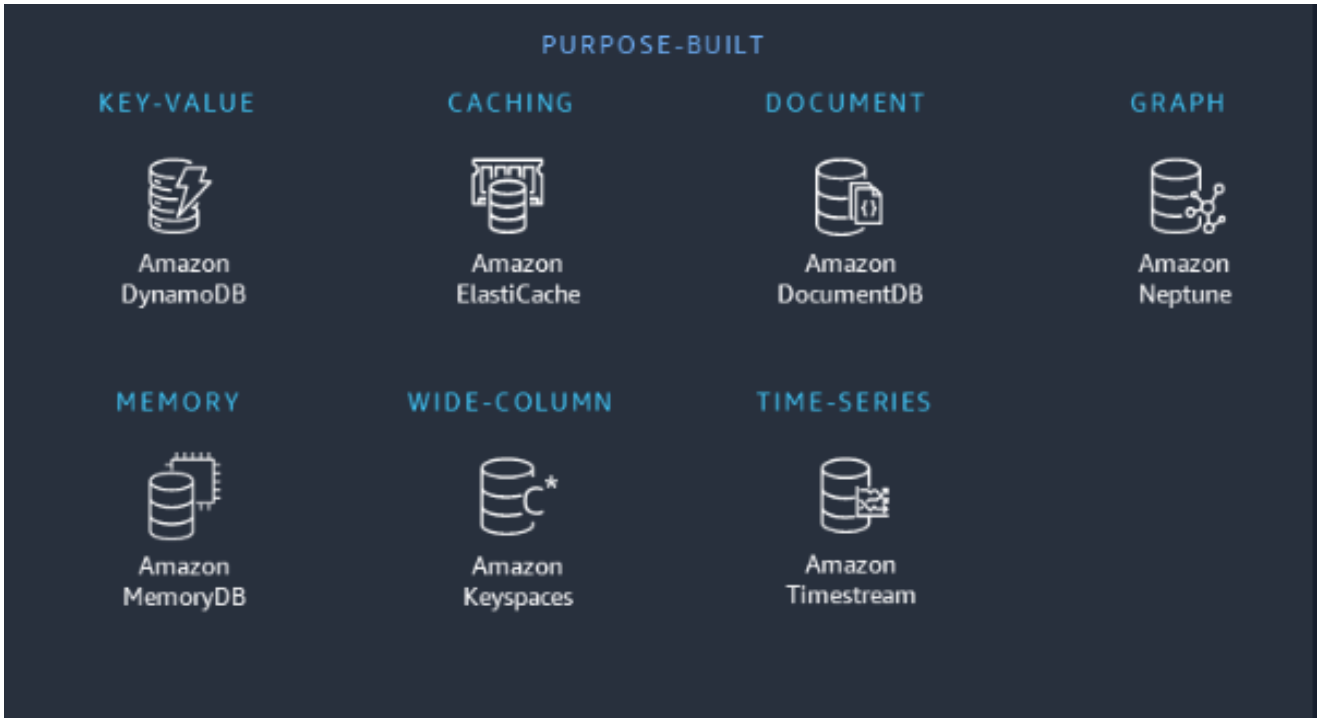
- What are my requirements?
- Enterprise-class relational database
  - **Amazon RDS**
- Fast and flexible NoSQL database service for any scale
  - **Amazon Dynamo DB**
- Operating system accessor application features that are not supported by AWS database services
  - **Databases on Amazon EC2**

When to Use Amazon RDS

- **Use Amazon RDS when your application requires:**
  - Complex transactions or complex queries
  - A medium to high query or write rate –Up to 30,000 IOPS (15,000 reads + 15,000 writes)
  - No more than a single worker node or shard
  - High durability

- **Do not use Amazon RDS when your application requires:**
  - Massive read/write rates (for example, 150,000 write/second)
    - Huge volumes of data or throughput demands
  - Simple GET or PUT requests and queries that a NoSQL database can handle – without schema
  - Relational database management system (RDBMS) customization

The right tool for the right job



Choosing an AWS database service - Choosing an AWS database service

- Specific case-driven requirements
  - Machine learning, data warehouse, graphs
  - **AWS purpose-built database services**

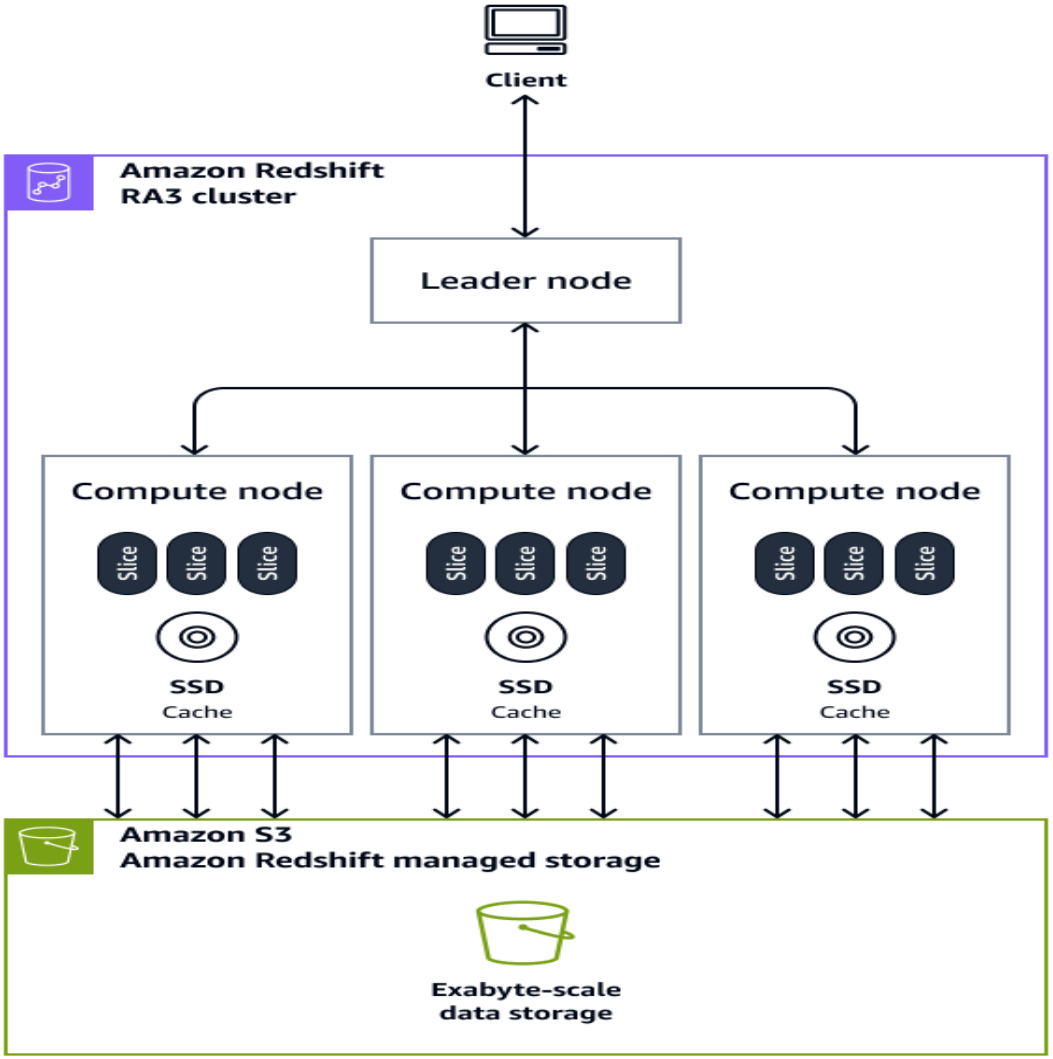
Summary

- AWS Database Services
  - RDS – SQL
  - Dynamo DB – No SQL Key Value Store
  - Redshift – SQL – Data Warehousing
  - Aurora – SQL Compatibility with advanced services

References

- docs.aws.amazon.com
- Purpose-built databases
- Choosing an AWS database service - Choosing an AWS database service
- aws.academy

Redshift Architecture



Architecture components of an Amazon Redshift data warehouse - AWS Prescriptive Guidance