

# Module 4

# Database & Machine Learning Intro

V I R T U A L  
**AWS**OME DAY

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# SQL and NoSQL 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
9182932465265	Cloud Computing Concepts	Wilson, Joe	Paperback
3142536475869	The Database Guru	Gomez, Maria	eBook

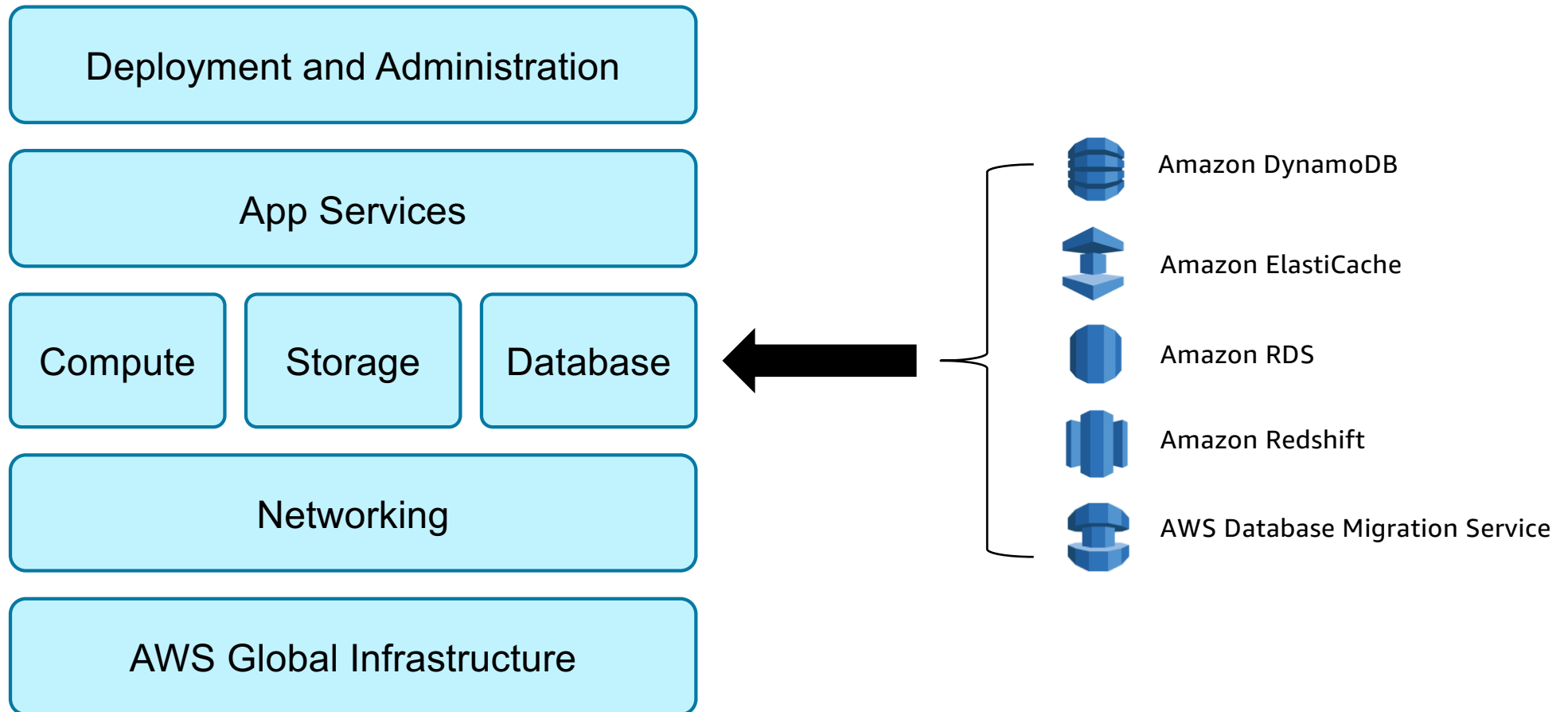
## NoSQL

{ 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 Managed Database Services



# 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



# How Amazon RDS Backups Work



## Automatic Backups:

- Restore your database to a point in time.
- Are enabled by default.
- Let you choose a retention period up to 35 days.
- Are stored in Amazon S3

## Manual Snapshots:

- Let you build a new database instance from a snapshot.
- Are initiated by the user.
- Persist until the user deletes them.



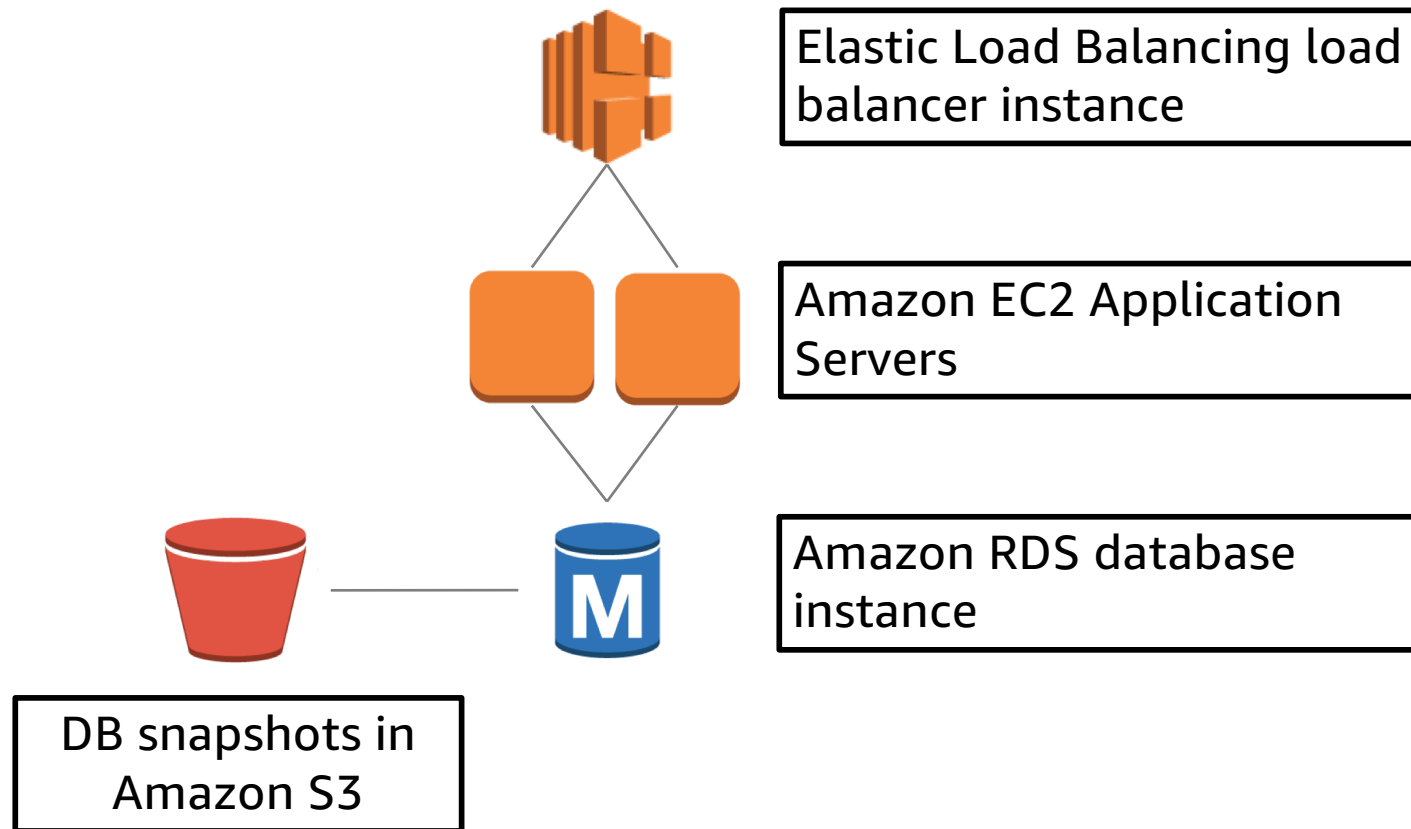
# Cross-Region Snapshots

- 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.





# A Simple Application Architecture

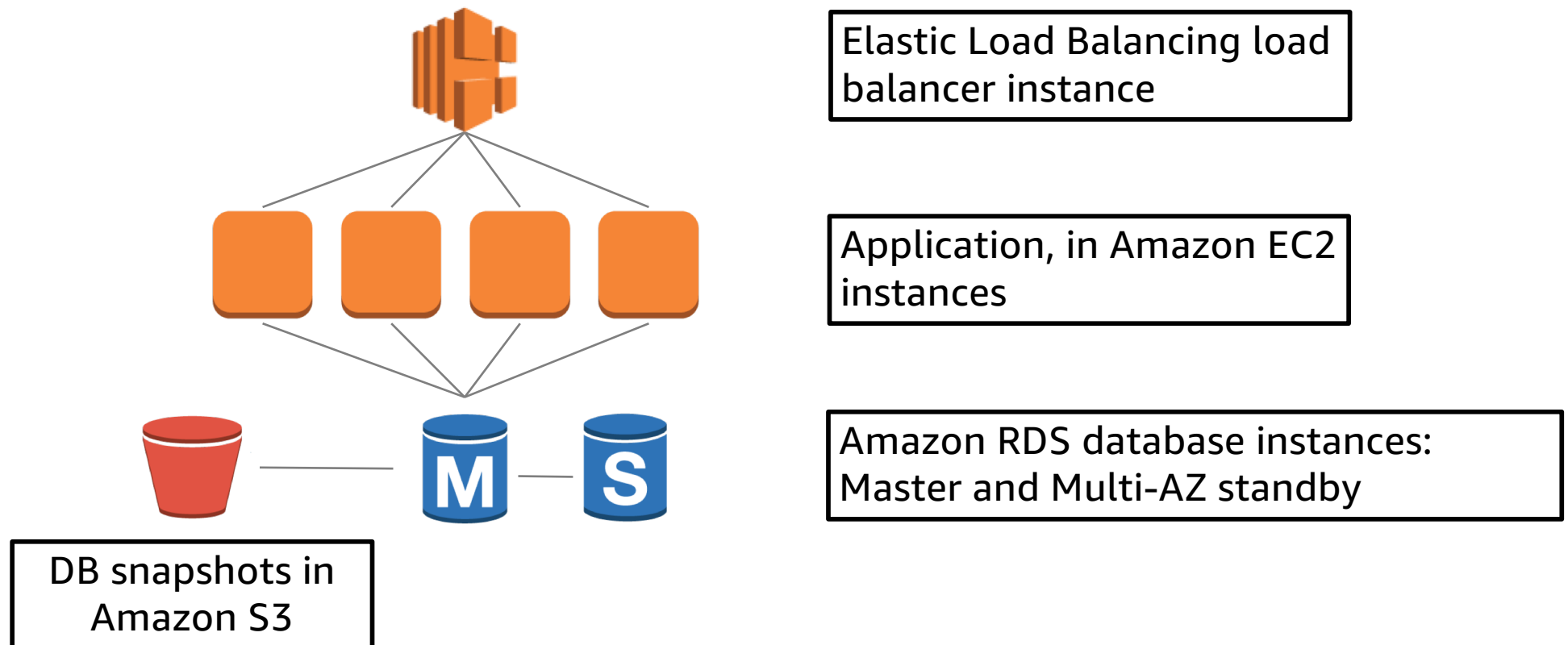


# Multi-AZ RDS Deployment



- 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



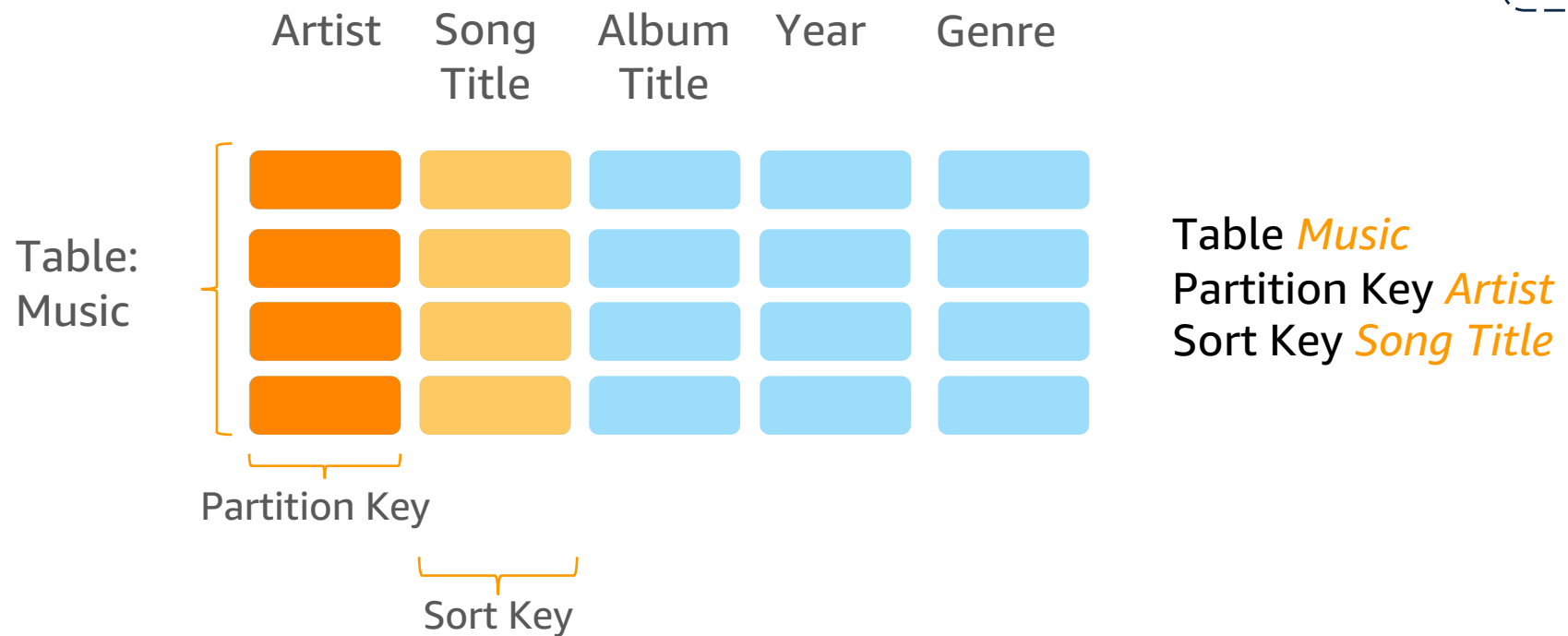
# Amazon DynamoDB



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.

# Primary Keys






(DynamoDB maintains a sorted index for both keys)

# Supported Operations



- **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.

# Database Considerations

If You Need	Consider Using
A relational database service with minimal administration	<b>Amazon RDS</b> <ul style="list-style-type: none"><li>• Choice of Amazon Aurora, MySQL, MariaDB, Microsoft SQL Server, Oracle, or PostgreSQL database engines</li><li>• Scale compute and storage</li><li>• Multi-AZ availability</li></ul> 
A fast, highly scalable NoSQL database service	<b>Amazon DynamoDB</b> <ul style="list-style-type: none"><li>• Extremely fast performance</li><li>• Seamless scalability and reliability</li><li>• Low cost</li></ul> 
A database you can manage on your own	Your choice of <b>AMIs</b> on Amazon EC2 and Amazon EBS that provide scale compute and storage, complete control over instances, and more. 



# Amazon Aurora

Fastest-growing AWS service, ever

MySQL and PostgreSQL compatible

---

5X faster than standard MySQL and 3X PostgreSQL

---

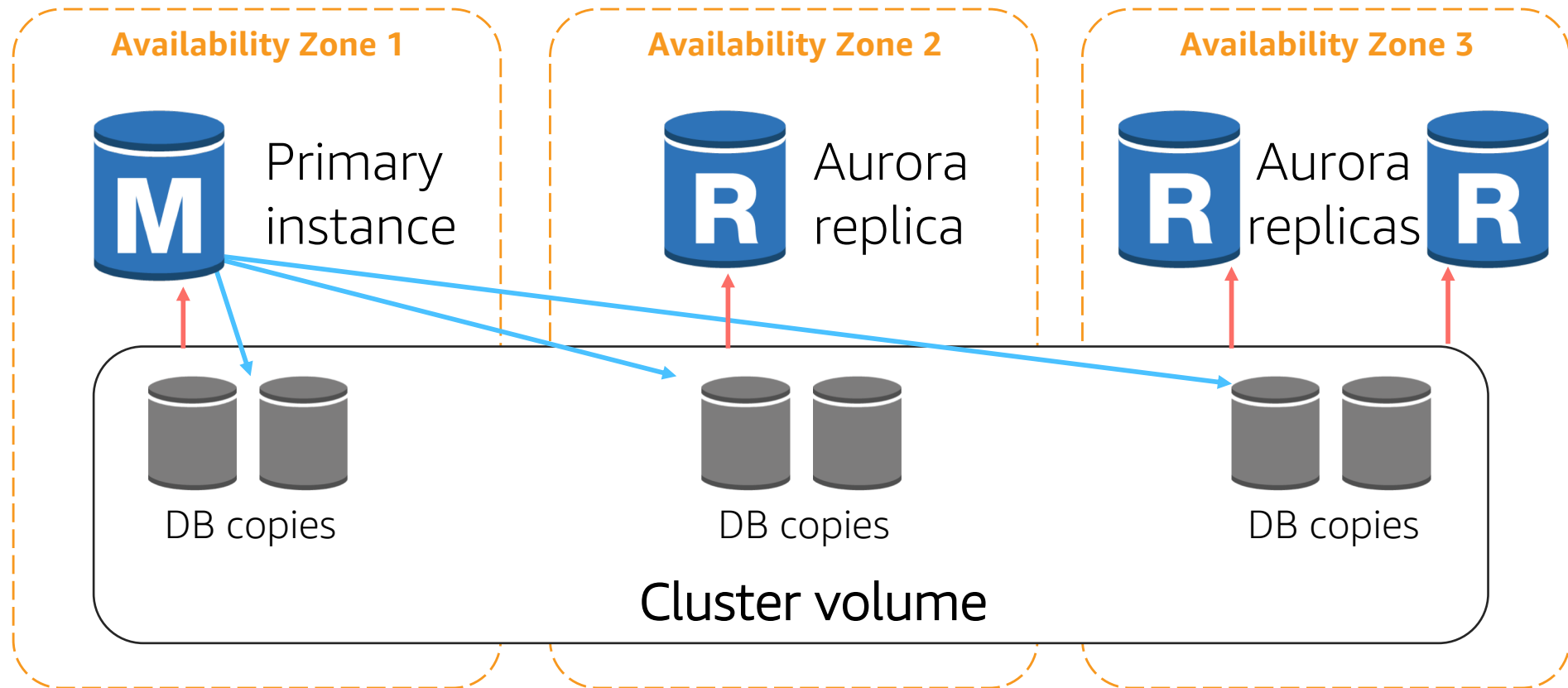
Highly available and durable

---

1/10<sup>th</sup> the cost of commercial grade databases



## Each Aurora DB cluster can have up to 15 Aurora replicas



# Aurora Serverless

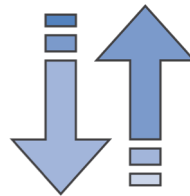


**Responds to your application automatically**

- Scales capacity
- Shut down
- Start up



Pay for number of ACUs used



Good for spiky, unpredictable workloads

# Machine Learning

V I R T U A L  
**AWS**OME DAY

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# AWS deep experience differentiates services



1995

Amazon has invested in AI/ML since our inception, and we share our knowledge and capabilities with our customers



2017



**Product recommendation engine**



**Robot-enabled fulfillment centers**



**New product categories**



**ML-driven supply chain and capacity planning**



**Natural language processing-supported contact centers**



**Checkout-free shopping using deep learning**

# The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

## AI SERVICES

### VISION



Amazon  
Rekognition

### SPEECH



Amazon  
Polly



Amazon  
Transcribe  
*+Medical*

### TEXT



Amazon  
Comprehend  
*+Medical*



Amazon  
Translate



Amazon  
Textract

### SEARCH



Amazon  
Kendra

### CHATBOTS



Amazon  
Lex

### PERSONALIZATION



Amazon  
Personalize

### FORECASTING



Amazon  
Forecast

### FRAUD



Amazon  
Fraud Detector

### DEVELOPMENT



Amazon  
CodeGuru

### CONTACT CENTERS



Contact Lens  
*For Amazon Connect*

## ML SERVICES



Amazon SageMaker

Ground  
Truth

ML  
Marketplace

SageMaker Studio IDE

Built-in  
algorithms

Notebooks

Experiments

Model  
training &  
tuning

Debugger

Autopilot

Model  
hosting

Model Monitor

Neo

Augmented  
AI

## ML FRAMEWORKS & INFRASTRUCTURE



PYTORCH



DeepGraphLibrary

Deep Learning  
AMIs & Containers

GPUs &  
CPUs

Elastic  
Inference

Inferentia

FPGA

# Neural Text To Speech with Polly

## US English Matthew voice

"The season begins anew this Sunday as the Yankees will be in Fenway Park to continue their age-old rivalry with the Boston Red Sox."



## US English Joanna voice

"Born August 19th, 1932 in Jamaica, New York, he was the son of Martha Pessolano Vacheron and the late Edwin F. Vacheron, formerly of Garden City, New York."



*I bought 2lbs of meat and 16oz of potatoes*

Justin (US)



Amy (UK)



Raveena (IN)



Plain text

SSML

?

I bought 2lbs of meat and 16oz of potatoes.

957 characters remaining (1000 maximum)

Show default text

Clear text

Language and Region

English, Indian

Voice

☒ Raveena, Female

▶ Listen to speech

⬇ Save speech as MP3

Change file format

# Key Stakeholders – ML use case



## Data scientists

Try out ideas quickly

Need high productivity

“Let me focus on models,  
not infrastructure”



## Ops engineers

Secure access to data

Max out utilization of expensive compute

Facilitate build, train, deploy

“Give me a secure, scalable, flexible platform”



## Business leaders

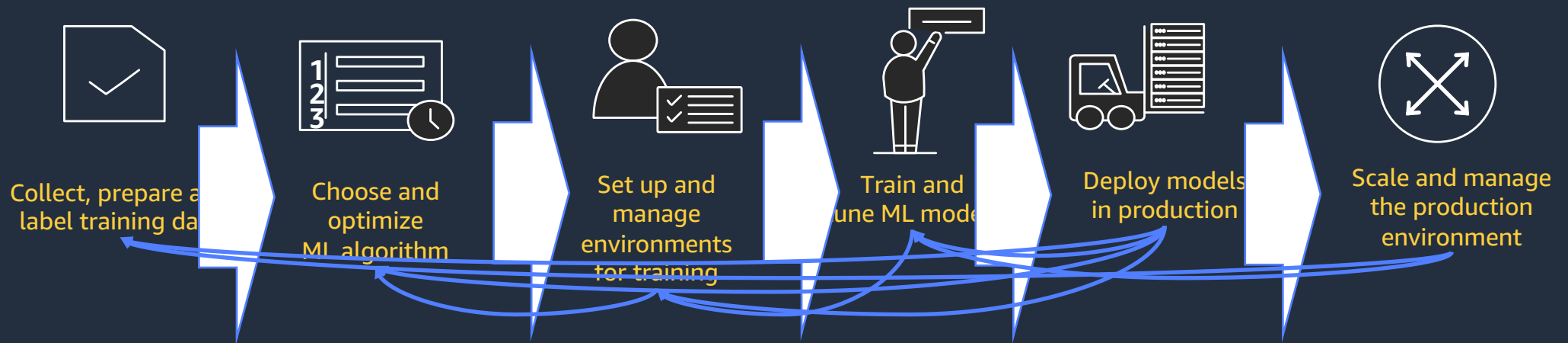
Deliver tangible business outcomes

Scale ML across all business units

“Keep the costs down”



# Traditional Machine learning process is hard



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# What is the right ML platform?

## Build-your-own ML platform

Custom tooling, security

Operations



Amazon  
EC2

**Data scientists ~ \$300/hr.**

**Infrastructure ~ \$3/hr.**

## Use a managed ML Service

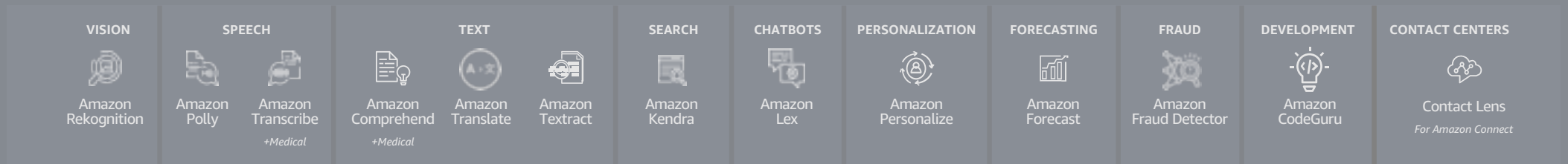


Amazon SageMaker

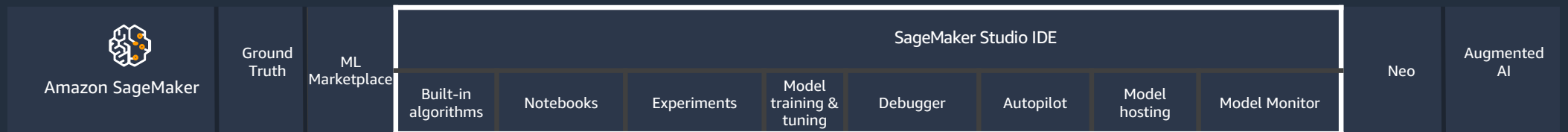
# The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

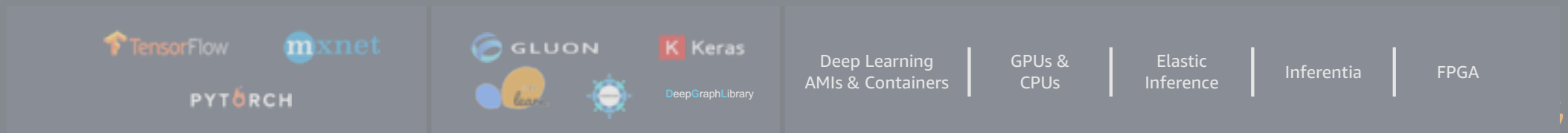
## AI SERVICES



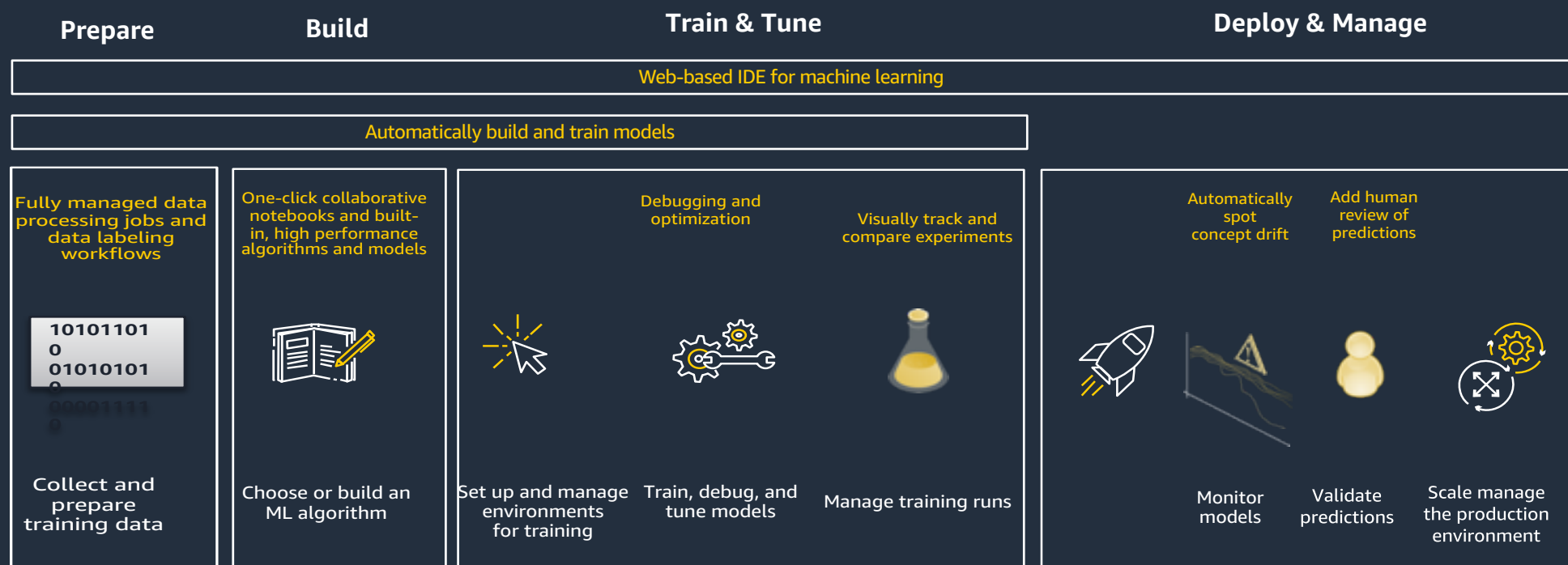
## ML SERVICES



## ML FRAMEWORKS & INFRASTRUCTURE



# Amazon SageMaker helps you build, train, and deploy models



# Amazon SageMaker

Build, train, deploy machine learning models quickly at scale

## Prepare

## Build

## Train & Tune

## Deploy & Manage

### Amazon SageMaker Studio

*Integrated Development environment(IDE) for Machine Learning*

### Amazon SageMaker Autopilot

*Automatically build and train models*

### One Click Deployment

*Supports real-time, batch & multi-model*

### Amazon SageMaker GroundTruth

*Build and manage training dataset*

### Amazon SageMaker Notebooks

*One-click notebooks with elastic compute*

### One Click Training

*Supports supervised, unsupervised & RL*

### Amazon SageMaker Model Monitor

*Automatically detect concept drift*

### Processing Job

*Supports Python or Spark*

### AWS Marketplace

*Pre-built algorithms, models, and data*

### Automatic Model Tuning

*One-click hyperparameter optimization*

### Amazon SageMaker Neo

*Train once, deploy anywhere*

### Amazon SageMaker Experiments

*Capture, organize, and compare every step*

### Amazon Elastic Inference

*Auto scaling for 75% less*

### Amazon SageMaker Debugger

*Debug and profile training runs*

### Amazon Augmented AI

*Add human review of model predictions*