# CLOUD COMPUTING

# Session 2 : AWS Core Services

APROBAT

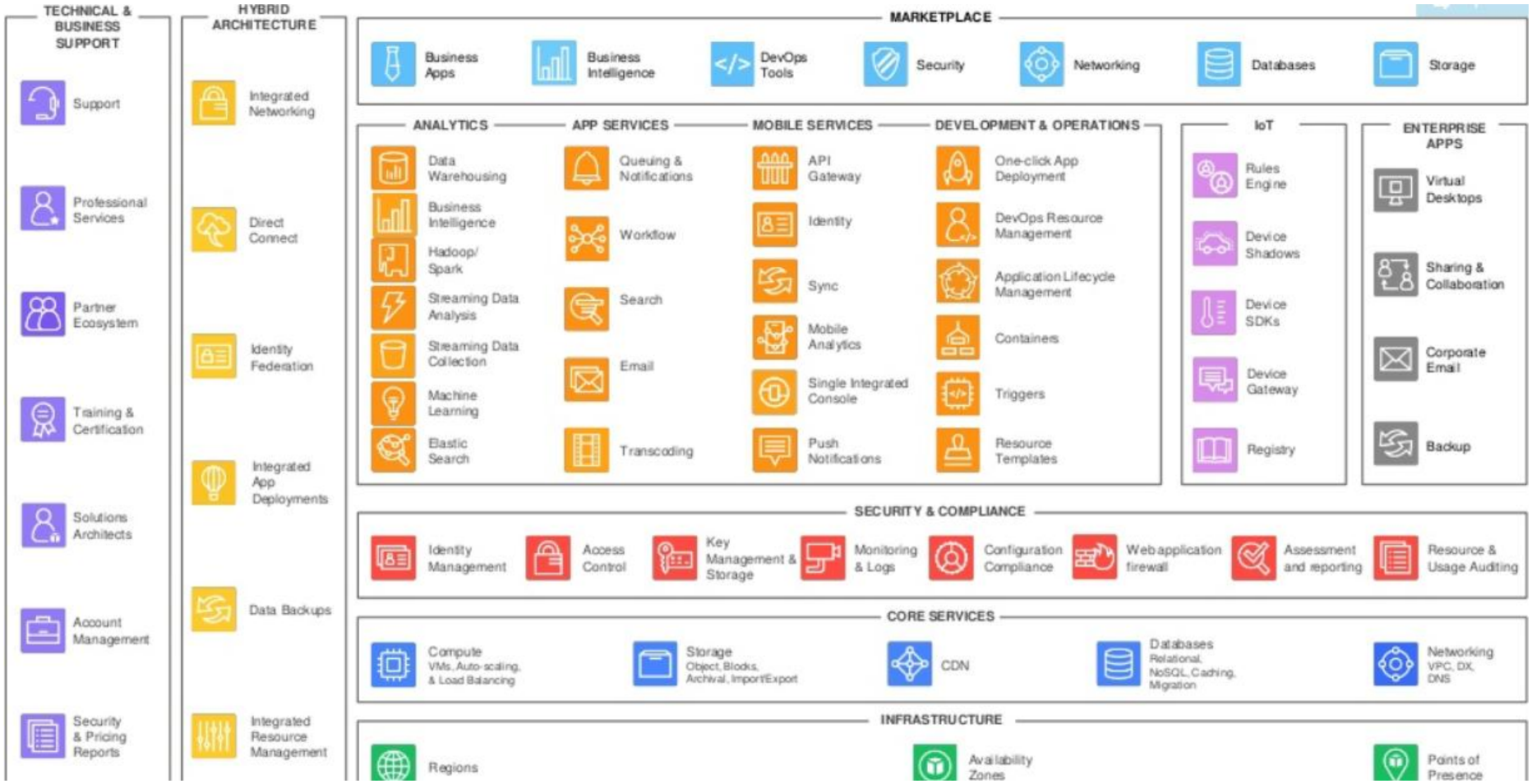
Conf. univ. dr. ing. IUSTIN PRIESCU - UTM
Dr. ing. Sebastian NICOLAESCU - Verizone Bussines-US

# What is AWS

Amazon Web Services is a secure cloud services platform with over 50 different services that include solutions for:


as well as:

# AWS Services



# AWS Regions


MONTREAL (Coming soon)
OHIO (Coming soon)
IRELAND
FRANKFURT
OREGON
UK (Coming soon)
N. CALIFORNIA
N. VIRGINIA
AWS GOVCLOUD
BEIJING
SEOUL
NINGXIA (Coming soon)
TOKYO
INDIA (Coming soon)
SINGAPORE
SÃO PAULO
SYDNEY
Region &
Number of Availability Zones
New Region
Coming Soon

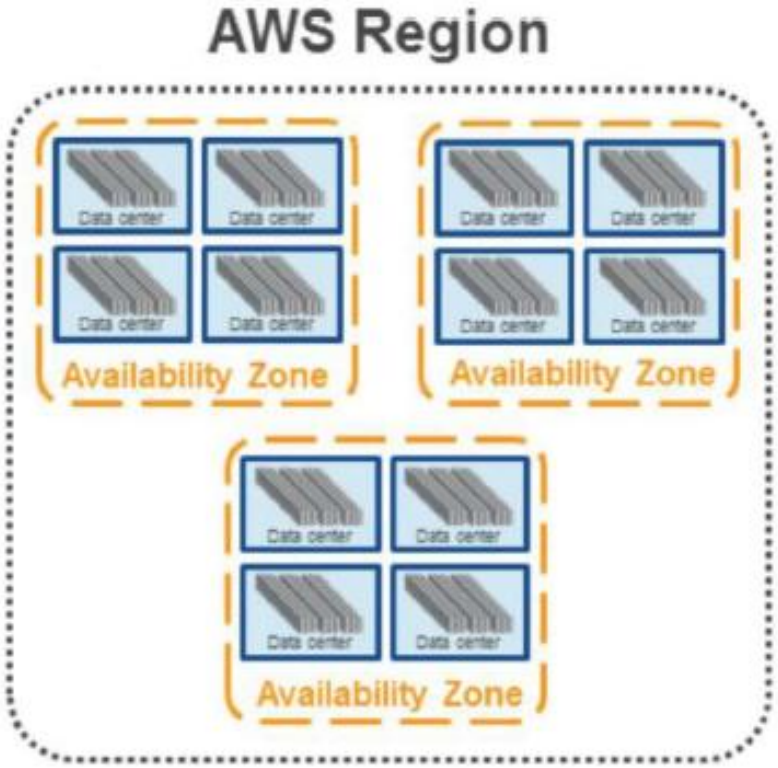
# AWS Region

Each region is made up of two or more Availability Zones.

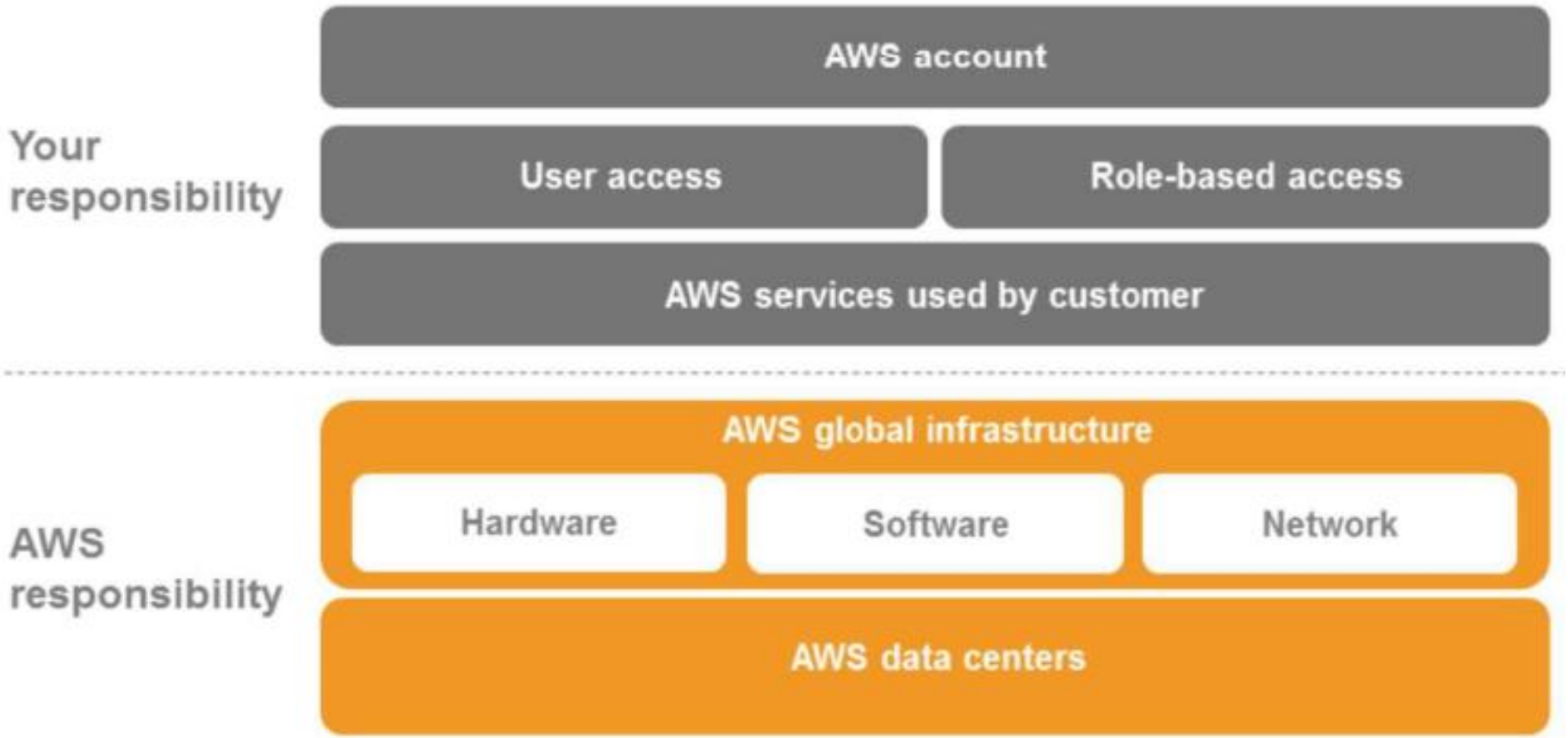
AWS has 15 regions worldwide.

You enable and control data replication across regions.

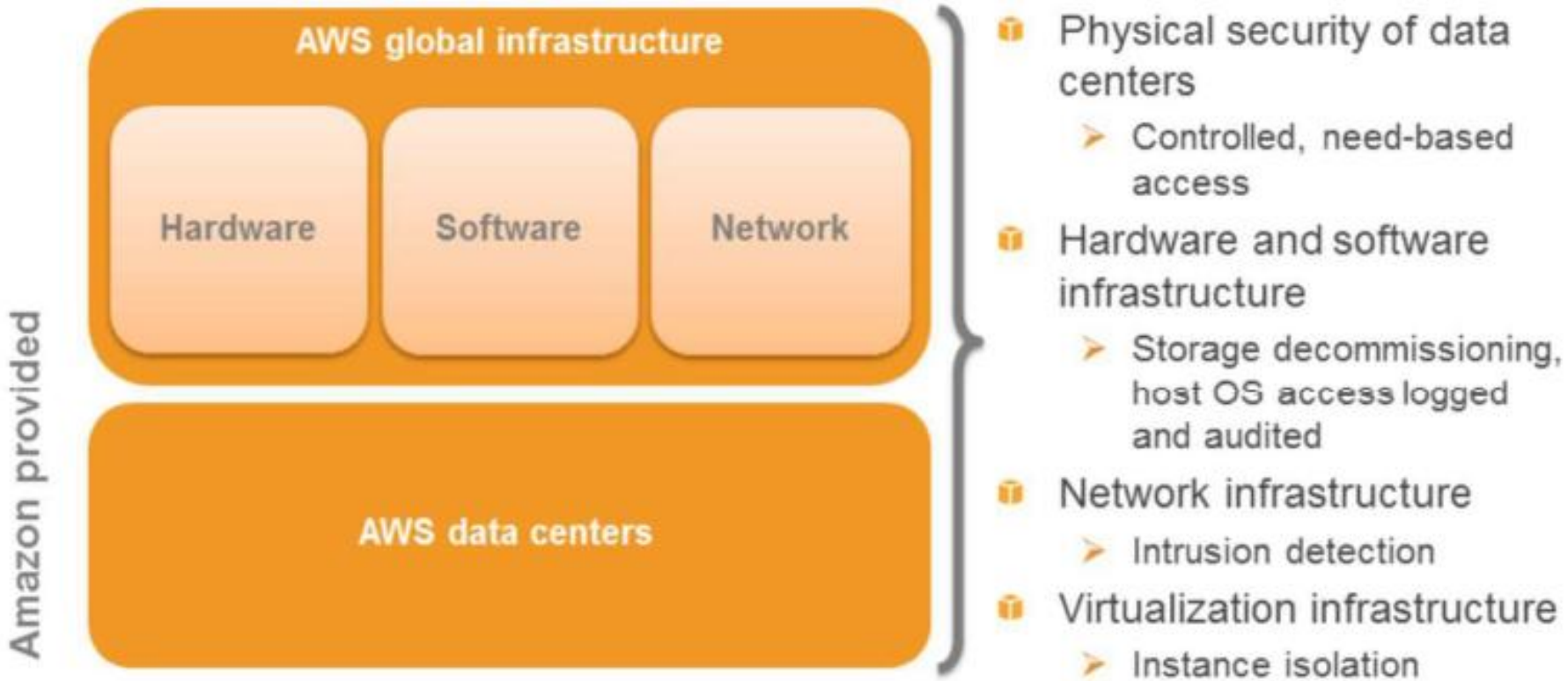
Communication between regions uses public Internet infrastructure.



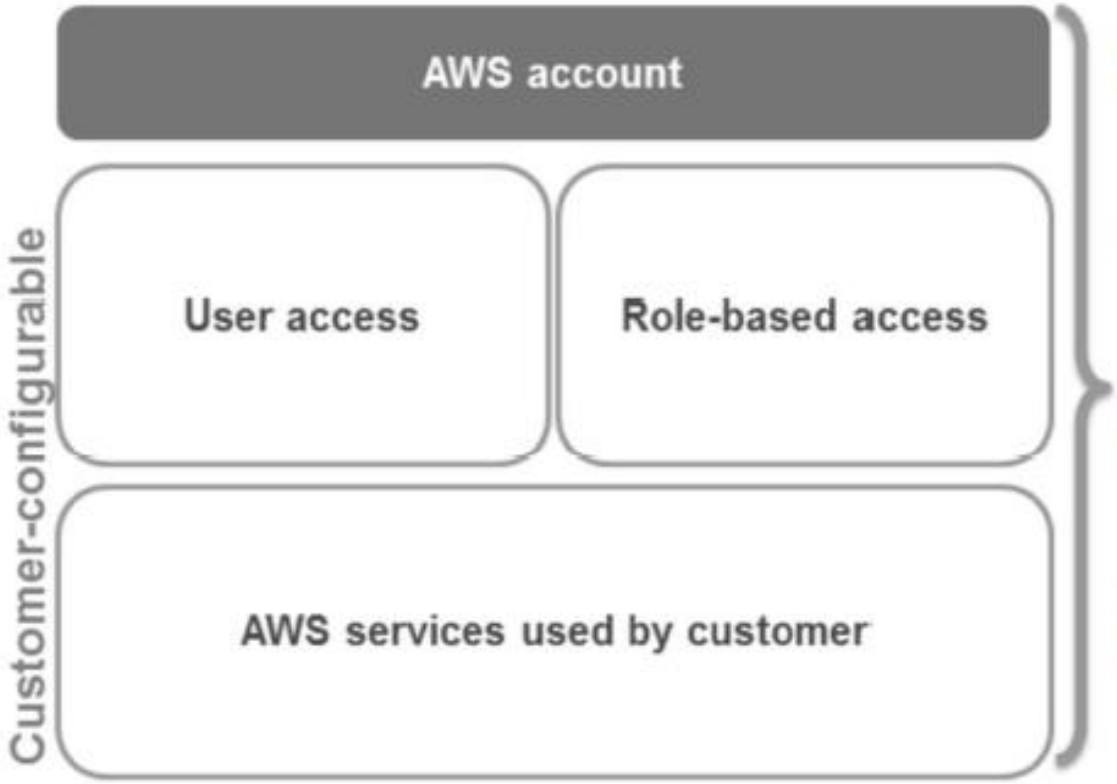
# Shared Responsibility Model



# AWS Security Responsibility – Security OF the Cloud



# Customer Security Responsibility – Security IN the Cloud




- Instance operating system
  - Including patching, maintenance
- Application
  - Passwords, role-based access etc.
- Security groups
- OS/host-based firewalls
  - Including intrusion detection/prevention systems
- Network configuration
- Account management
  - Separation of access

# Architecting for the cloud

# Best Practices

# 1. Enable Scalability

*Ensure that your architecture can handle changes in demand.*

# 2. Automate Your Environment

Removing manual processes to improve your system's stability and consistency, and the efficiency of your organization.

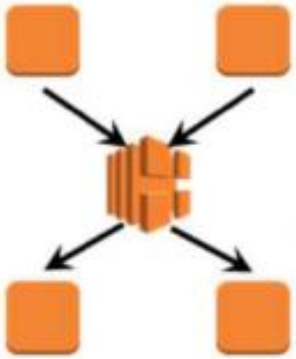
# 3. Use Disposable Resources

Think of servers and other components as temporary resources.

Job completed?
Shut the
instance down

# 4. Loosely Couple Your Components

*Reduce interdependencies so that the change or failure of one component does not affect other components.*



Load balancer reduces dependency between instances

# 5. Design Services, Not Servers

*Managed services and serverless architectures can provide greater reliability and efficiency in your environment.*

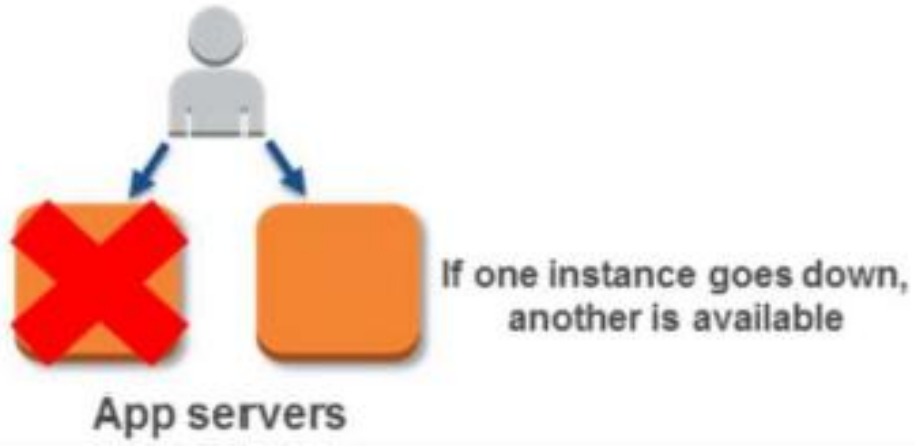
AWS Lambda
for running simple
functions

Amazon SQS
for message
queuing

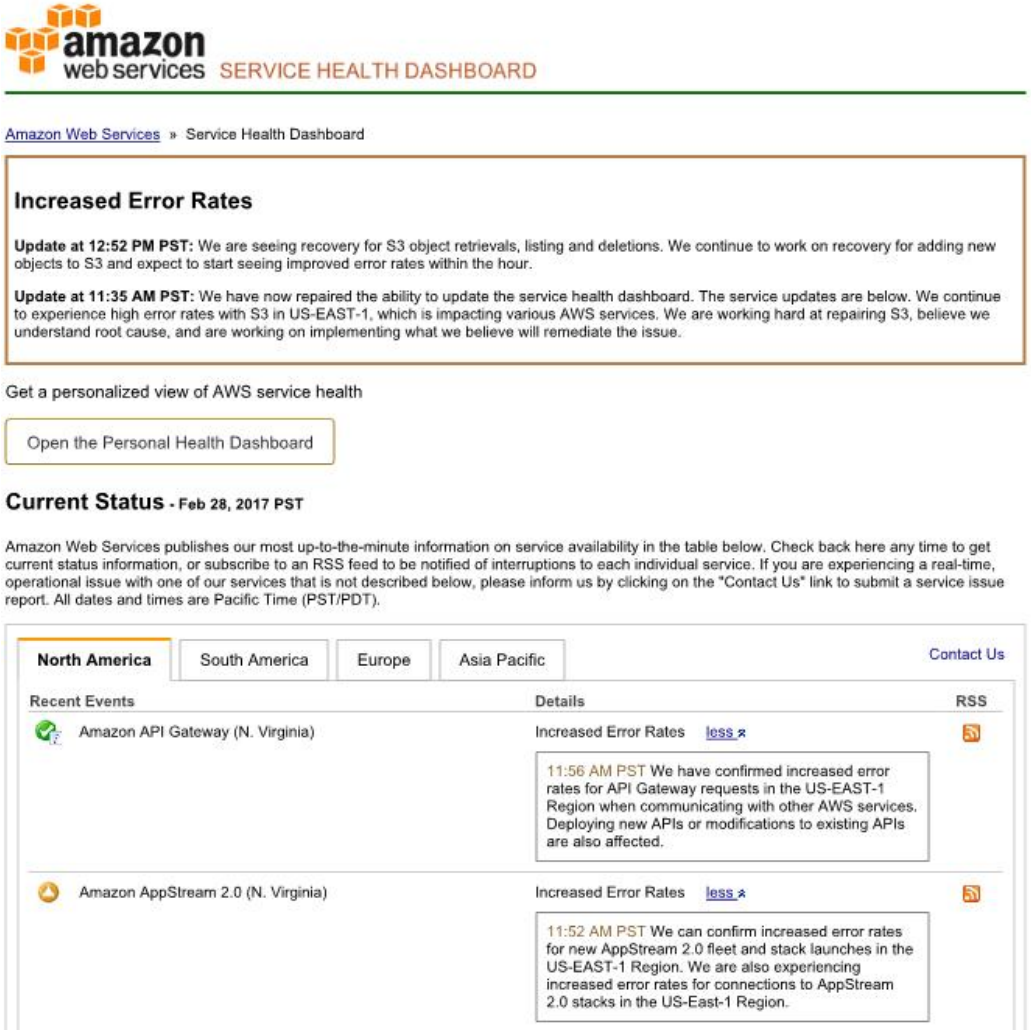
Amazon SNS
for push notifications

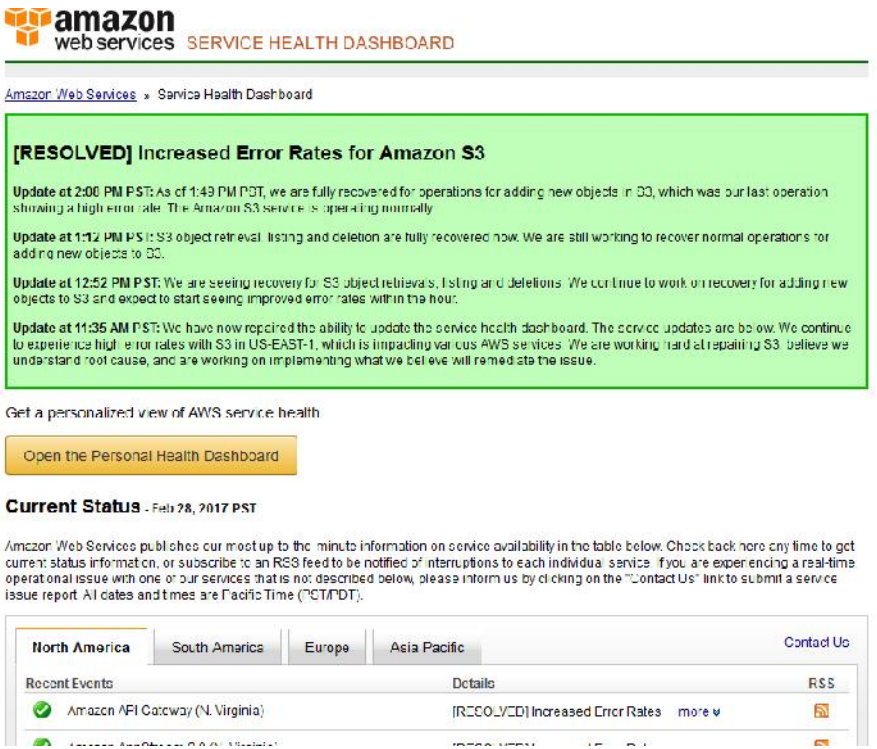
# 6. Avoid Single Points of Failure

*Implement redundancy where possible so that single failures don't bring down an entire system.*



# AWS Outage 28-Feb-2017




http://status.aws.amazon.com

# 7. Choose the Right Database Solution

*Match the technology to the workload:*

*Choose from an array of relational database engines, NoSQL solutions, data warehousing options, and search-optimized data stores.*

Amazon Redshift for Data Warehousing

Amazon RDS for Relational Database

Amazon DynamoDB for NoSQL

Amazon Kinesis for Streaming Data

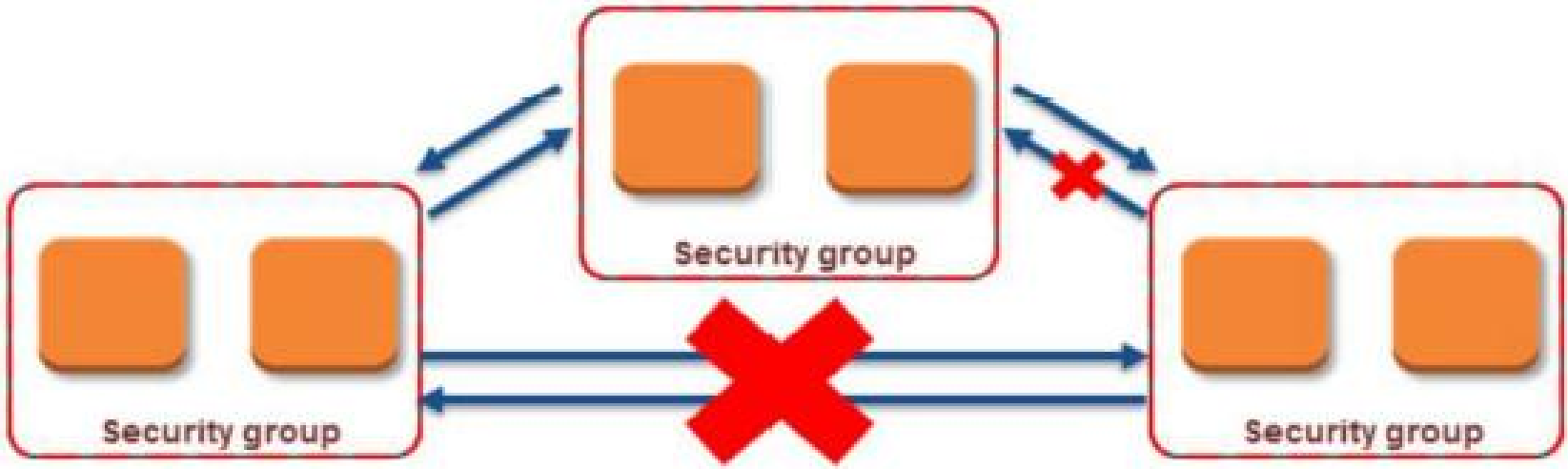
Amazon Elasticsearch for Search

# 8. Use Caching

*Use caching to minimize redundant data retrieval operations.*

# 9. Secure Your Infrastructure

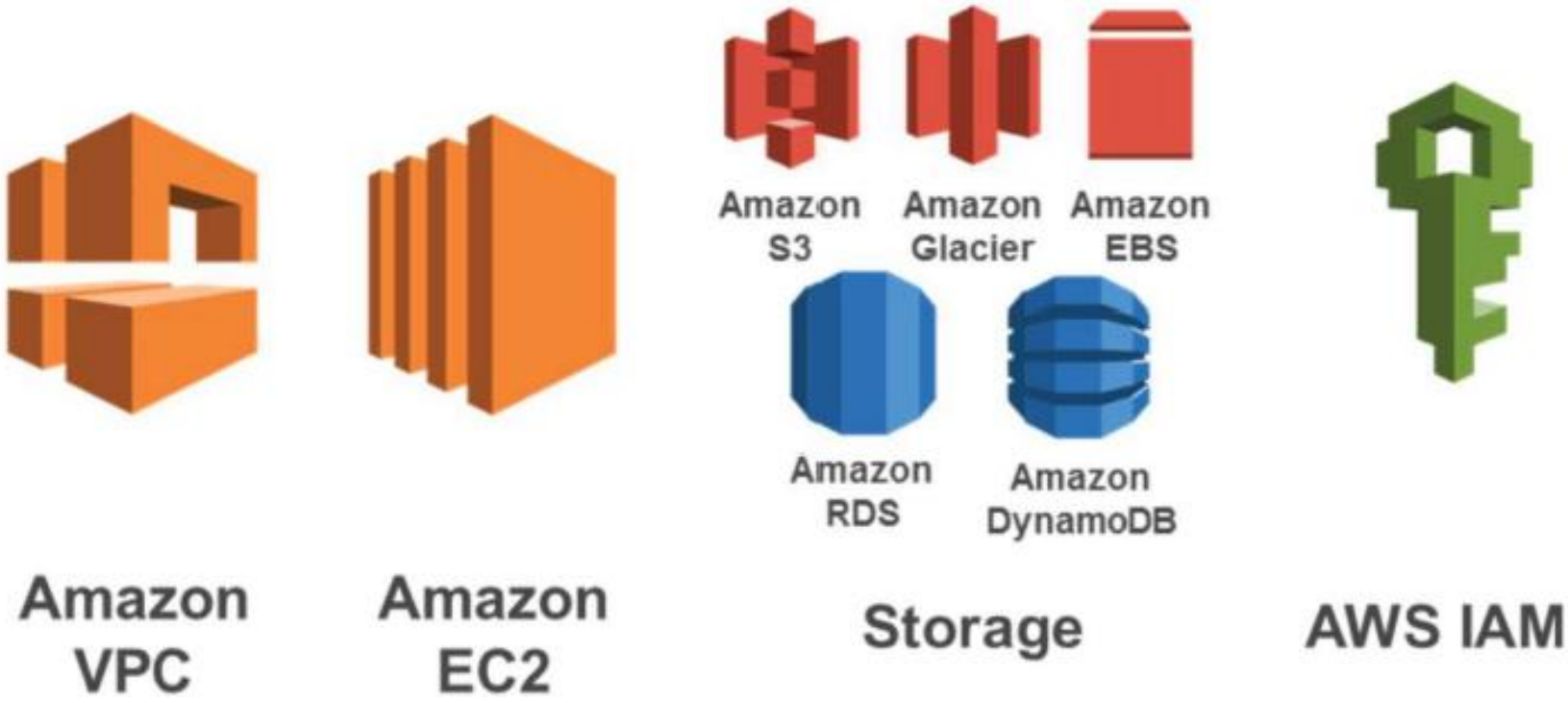
*AWS enables you to implement security both at the perimeter and within/between your resources.*



# 10. Optimize For Cost

*Ensure that your resources are sized appropriately, that they scale in and out based on need, and that you're taking advantage of different pricing options.*

# AWS Core Services

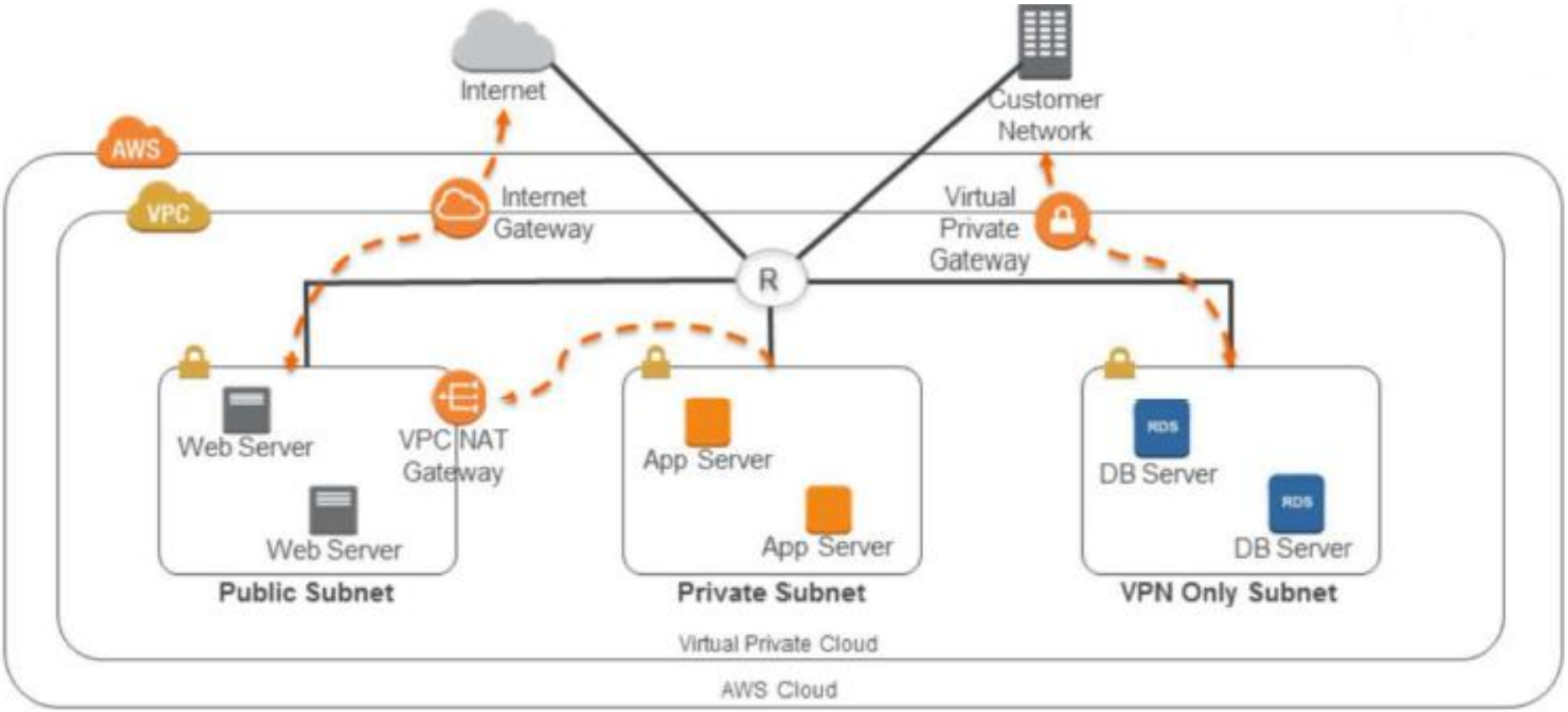


# AWS – Virtual Private Cloud (VPC)

*Amazon Virtual Private Cloud (VPC) allows you to provision virtual networks hosted on the AWS cloud and dedicated to your AWS account.*

- VPCs are logically isolated from other virtual networks.
- Many AWS resources, such as Amazon EC2 instances, are launched into VPCs.
- Your VPC's key features are configurable:
  - IP ranges
  - Routing
  - Network gateways
  - Security settings

# VPC Example


Internet
Customer Network
AWS
VPC
Internet Gateway
Virtual Private Gateway
R
Web Server
VPC NAT Gateway
Web Server
Public Subnet
App Server
App Server
Private Subnet
RDS
DB Server
RDS
DB Server
VPN Only Subnet
Virtual Private Cloud
AWS Cloud

# AWS – Elastic Cloud Compute (EC2)

*Amazon Elastic Compute Cloud (EC2) offers virtual computing environments (instances) you can launch and manage with a few clicks of a mouse or a few lines of code.*

- Most server operating systems are supported.
- Create, save, and reuse your own server images (Amazon Machine Images).
- Launch one instance at a time, or launch a whole fleet.
- Add more instances when you need them; terminate when you don't.
- CPU, memory, storage, networking, graphics, and general purpose types are available.
- Use security groups to control traffic to and from instances.

# EC2 Instances Powered By Intel

| AWS Instance Type | High Memory X1 | Compute Optimized C4 | Storage Optimized D2 | General Purpose M4 | Memory Optimized R3 | IO Optimized I2 | Graphics Optimized G2 | Burstable Performance T2 |
|---|---|---|---|---|---|---|---|---|
| Intel Processor | Intel Xeon E7-8880 v3 | Custom Intel Xeon E5-2666 v3 | Custom Intel Xeon E5-2676 v3 | Custom Intel Xeon E5-2676 v3 | Intel Xeon E5-2670 v2 | Intel Xeon E5-2670 v2 | Intel Xeon E5-2670 | Intel Xeon Family |
| Intel AVX | AVX 2.0 | AVX 2.0 | AVX 2.0 | AVX 2.0 | Yes | Yes | Yes | Yes |
| Intel AES-NI | Yes | Yes | Yes | Yes | Yes | Yes | No | No |
| Intel Turbo Boost | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intel TSX | Yes | No | No | No | No | No | No | No |
| Per core P- and C-state control | No | Yes (8xlarge only) | No | No | No | No | No | No |
| SSD Storage | EBS Optimized by default | EBS Optimized by default | No | EBS Optimized by default | Yes | Yes | Yes | EBS only |

# Intel Processor Technologies

Intel® AVX – Dramatically better performance for highly parallel HPC workloads such as *life science engineering, data mining, financial analysis*, or other technical computing applications. AVX also enhances *image, video, and audio* processing.

Intel® AES-NI – Enhance your security with these encryption instructions that reduce the performance penalty associated with encrypting/decrypting data.

Intel® Turbo Boost Technology – More computing power when you need it with performance that adapts to spikes in your workload.

Intel Transactional Synchronization (TSX) Extensions – Enable execution of transactions that are independent to accelerate throughput.

P state & C state control – ability to individually tune each cores performance & sleep states to improve application performance

# EC2 Instance Classes

| Instance Family | Some Use Cases |
| --- | --- |
| General purpose (t2, m4, m3) | • Low-traffic websites and web applications<br>• Small databases and mid-size databases |
| Compute-optimized (c4, c3) | • High performance front-end fleets<br>• Video-encoding |
| Memory-optimized (r3) | • High performance databases<br>• Distributed memory caches |
| Storage-optimized (i2, d2) | • Data warehousing<br>• Log or data-processing applications |
| GPU instances (g2) | • 3D application streaming<br>• Machine learning |

# EC2 Pricing Options

| | On-Demand Instances | Reserved Instances (RIs) | Spot Instances |
|---|---|---|---|
| Term | Pay as you go | One year or three years | Bid on unused capacity; instances can be lost if you are outbid |
| Benefit | Low cost and flexibility | Predictability ensures compute capacity is available when needed | Large scale, dynamic workload |
| Cost | Pay for what you use | Pay low or no upfront fee; overall cost is lower | Spot price based on supply and demand |
| Use case | • Short-term, spiky, or unpredictable workloads<br>• Application development or testing | • Steady state or predictable usage workloads<br>• Applications that require reserved capacity, including disaster recovery<br>• Users able to make upfront payments to reduce total computing costs even further | • Applications with flexible start and end times<br>• Applications only feasible at very low compute prices<br>• Users with urgent computing needs for large amounts of additional capacity |

# AWS – Elastic Block Store

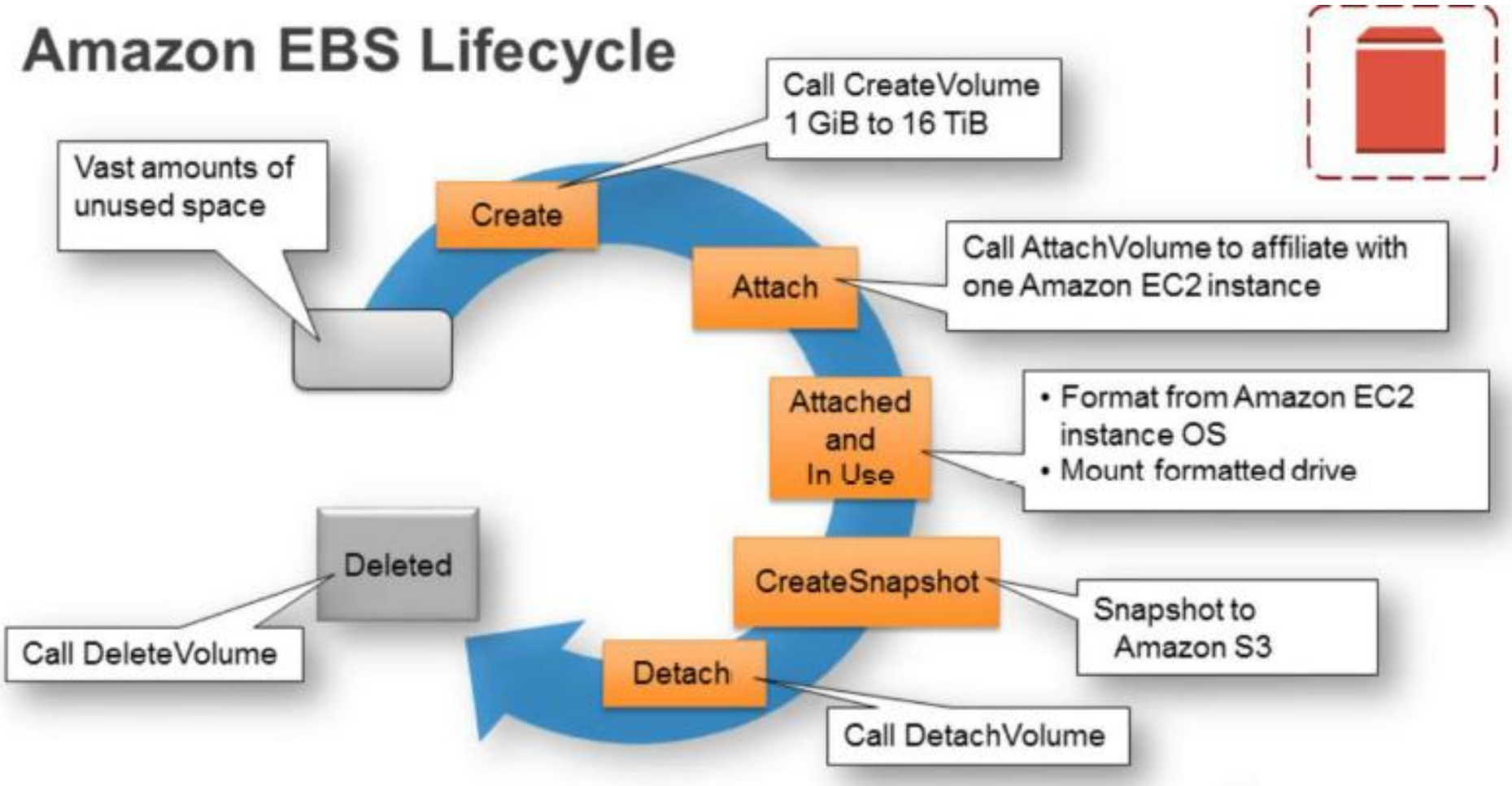
*Amazon EBS allows you to create individual storage volumes and attach them to an Amazon EC2 instance.*

- Amazon EBS offers Block-level storage.
- Volumes are automatically replicated within its Availability Zone.
- Can be backed up automatically to Amazon S3.
- Uses:
  - Boot volumes and storage for EC2 instances
  - Data storage with a file system
  - Database hosts
  - Enterprise applications

# EBS Volume Types

| | Magnetic | Cold HDD | Throughput Optimized HDD | General Purpose SSD | Provisioned IOPS SSD |
|---|---|---|---|---|---|
| Max volume size | 1 TiB | 16 TiB | 16 TiB | 16 TiB | 16 TiB |
| Max IOPS/volume | 40 to 200 | 250 | 500 | 10,000 | 20,000 |
| Max throughput/volume | 40 to 90 MiB/sec | 250 MiB/s | 500 MiB/s | 160 MiB/sec | 320 MiB/sec |
| Use cases | • Infrequent data access | • Workloads involving large, sequential I/O | • Workloads involving large, sequential I/O | • Boot volumes<br>• Small to Medium DBs<br>• Dev and Test environments | • I/O-intensive workloads<br>• Relational DBs<br>• NoSQL DBs |

# Amazon EBS Lifecycle



# S3 (Simple Storage Service)

*Amazon S3 is a managed cloud storage solution designed to scale seamlessly and provide 99.999999999% durability.*

- Amazon S3 provides object-level storage.
- Store as many objects as you want.
- Data is stored redundantly.
- Access Amazon S3 with the AWS Management Console, one of the AWS SDKs, or a third-party solution.
- Object uploads or deletes can trigger notifications, workflows, or even scripts.
- Data in transit and at rest can be encrypted automatically.

# S3 – Cost Overview

Pay only for what you use, including:

- GBs per month
- Transfer OUT to other regions
- PUT, COPY, POST, LIST, and GET requests

You do NOT have to pay for:

- Transfer IN to Amazon S3
- Transfer OUT from Amazon S3 to Amazon CloudFront or Amazon EC2 in the same region

# AWS Glacier

*Amazon Glacier is a data archiving service designed for security, durability, and an extremely low cost.*

- Amazon Glacier is designed for durability of 99.999999999% of objects.
- Amazon Glacier supports SSL/TLS encryption of data in transit and at rest.
- The Vault Lock feature enforces compliance via a lockable policy.
- Extremely low-cost design is ideal for long-term archiving;
  - Data retrieval will take 3-5 hours to begin.

# AWS RDS

*Amazon RDS lets you set up, operate, and scale relational databases in the cloud. Features:*

- Managed service
- Accessible via the AWS Management Console, AWS RDS Command-Line Interface, or simple API calls
- Scalable (compute and storage)
- Automated redundancy and backup available
- Supported database engines:
  - Amazon Aurora
  - PostgreSQL
  - MySQL
  - MariaDB
  - ORACLE
  - Microsoft SQL Server

# RDS – Selection Guidelines

Use Amazon RDS when your app requires:

- Complex transactions or complex queries
- A medium-to-high query/write rate – up to 30K IOPS (15K reads + 15K writes)
- No more than a single worker node/shard
- High durability

Do **not** use Amazon RDS when your app requires:

- Massive read/write rates (e.g., 150K write/second)
- Sharding due to high data size or throughput demands
- Simple GET/PUT requests and queries that a NoSQL database can handle
- RDBMS customization

# DynamoDB

*Amazon DynamoDB is a fully managed NoSQL database service.*

- Consistent, single-digit millisecond latency at any scale
- No table size or throughput limits
- Runs exclusively on SSDs
- Document and key-value store models supported
- Ideal for mobile, web, gaming, ad tech, and IoT applications
- Accessible via the AWS Management Console, the AWS Command-Line Interface, or simple API calls

# DynamoDB – Provisioned Throughput

You specify your throughput capacity requirements (read/write), and DynamoDB allocates the resources you need.

Read capacity unit:

- One ***strongly*** consistent read per second for items as large as 4 KB.
- Two ***eventually*** consistent reads per second for items as large as 4 KB.

Write capacity unit:

- One write per second for items as large as 1 KB.

# AWS Identity and Access Management (IAM)

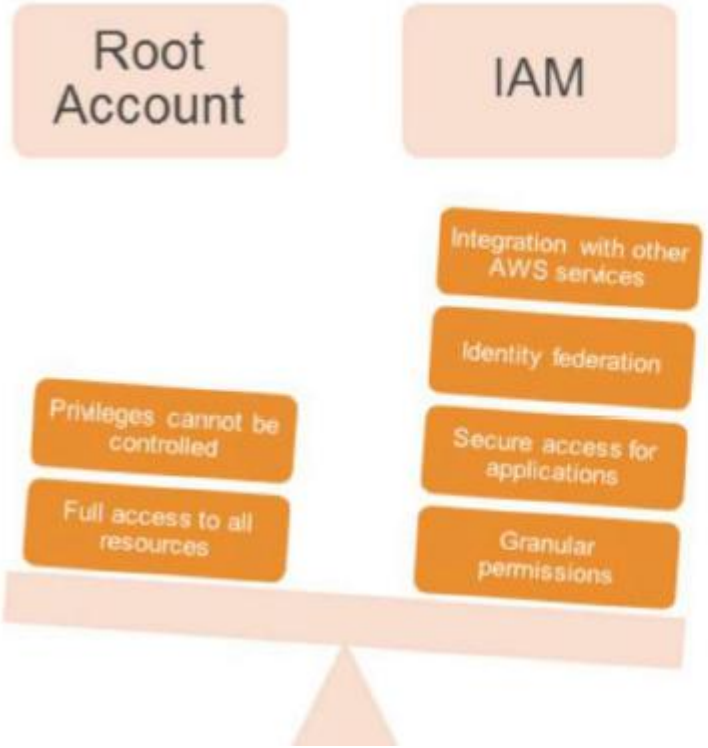
Centrally manage access and authentication of your users to your AWS resources.

- Offered as a feature of your AWS account for no charge.
- Create users, groups, and roles, and apply policies to them to control their access to AWS resources.
- Manage what resources can be accessed and how they can be accessed (e.g., terminating EC2 instances).
- Define required credentials based on context (e.g., who is accessing which service and what are they trying to do?).

# IAM – Types of Credentials

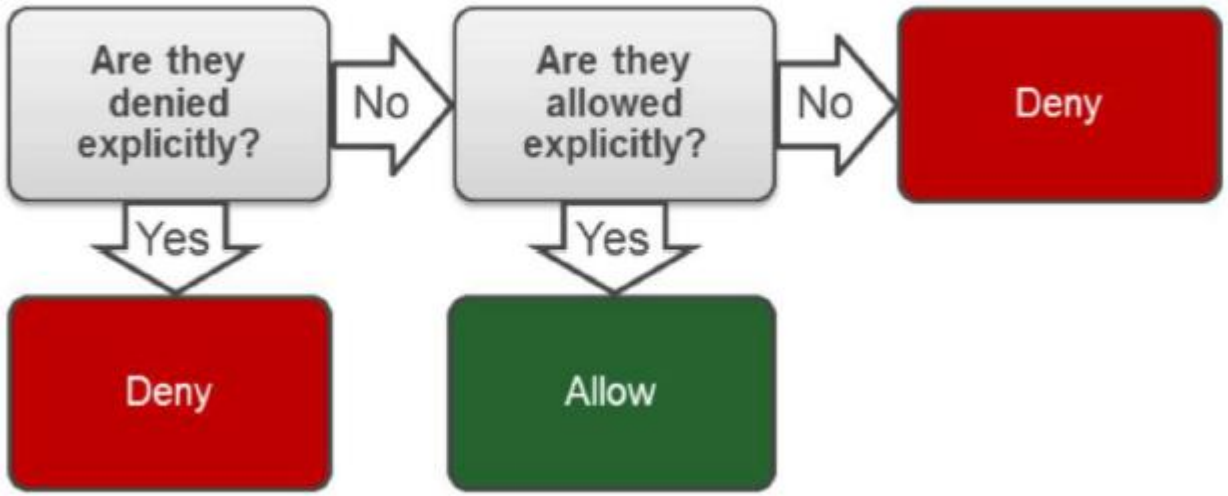
| | |
|---|---|
| **Email address and password** | Associated with your AWS account (root) |
| **IAM user name and password** | Used for accessing the AWS Management Console |
| **Access keys** | Typically used with CLI and programmatic requests like APIs and SDKs |
| **Multi-Factor Authentication** | Extra layer of security |
| | Can be enabled for root account and IAM users |
| **Key pairs** | Used only for specific AWS services like Amazon EC2 |

# Root Account Access vs IAM Access




# IAM Permissions

How IAM determines permissions:



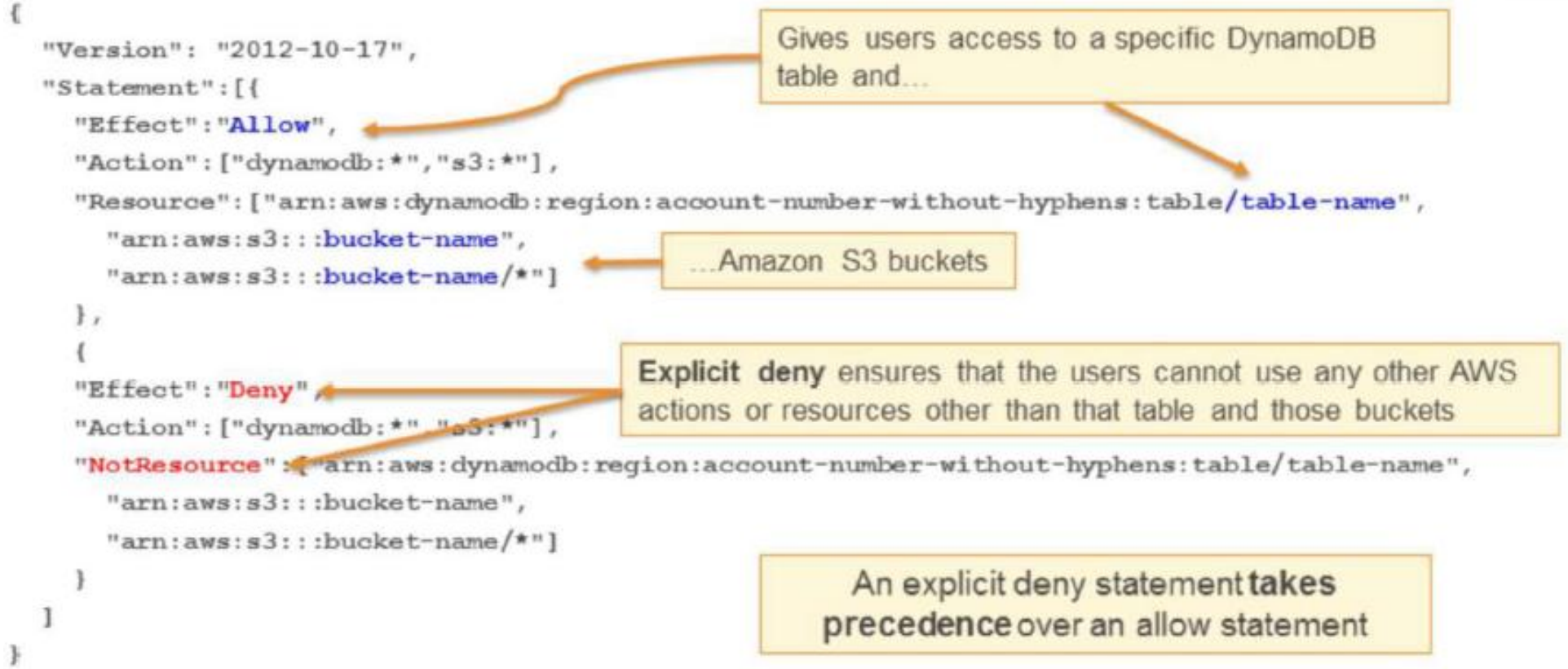
# IAM Policies

An IAM policy is a formal statement of one or more permissions.

- You attach a policy to any IAM entity: user, group, or role.
- Policies authorize the actions that may, or may not, be performed by the entity.
  - Enables fine-grained access control.
- A single policy can be attached to multiple entities.
- A single entity can have multiple policies attached to it.

**Best practice:** When attaching the same policy to multiple IAM users, put the users in a group and attach the policy to the group instead.

# IAM Policy Example



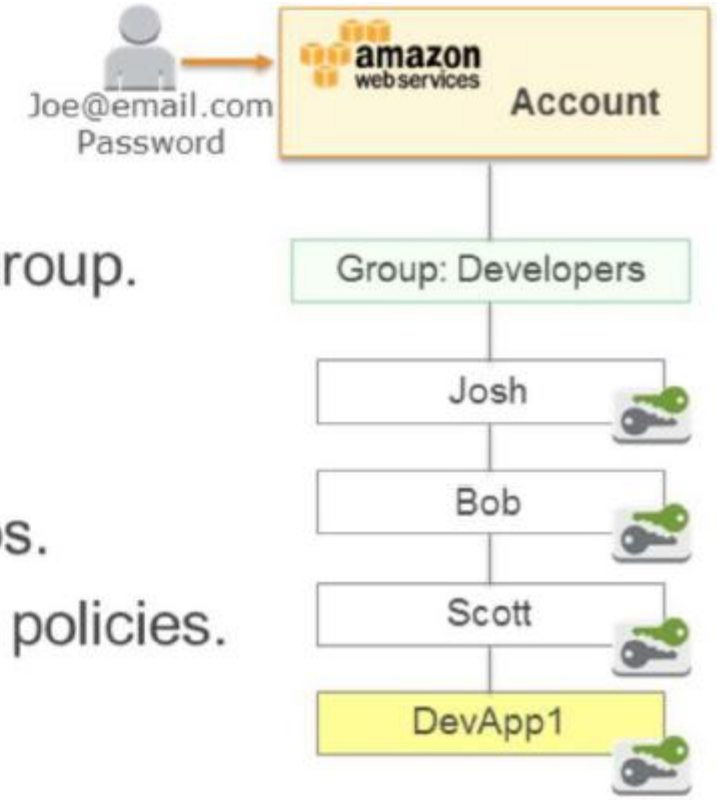
# IAM Users

- An entity you create in AWS.
- Provides a way to interact with AWS.
- No default security credentials for IAM users.
  - You have to assign them specifically.
- IAM users are not necessarily people.


Best practice: Create a separate IAM user account with administrative privileges for the root account user.

# IAM Groups

- Collection of IAM users.
- Specify permissions for the entire group.
- No default groups.
- Groups cannot be nested.
- A user can belong to multiple groups.
- Permissions are defined using IAM policies.



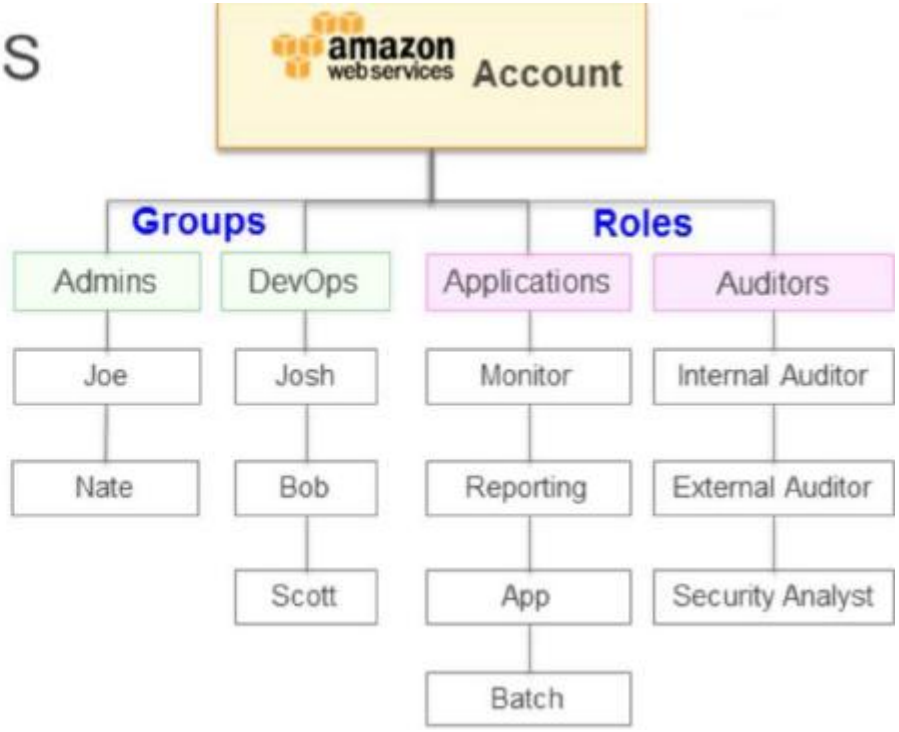
# IAM Roles

Used to **delegate** access to AWS resources.

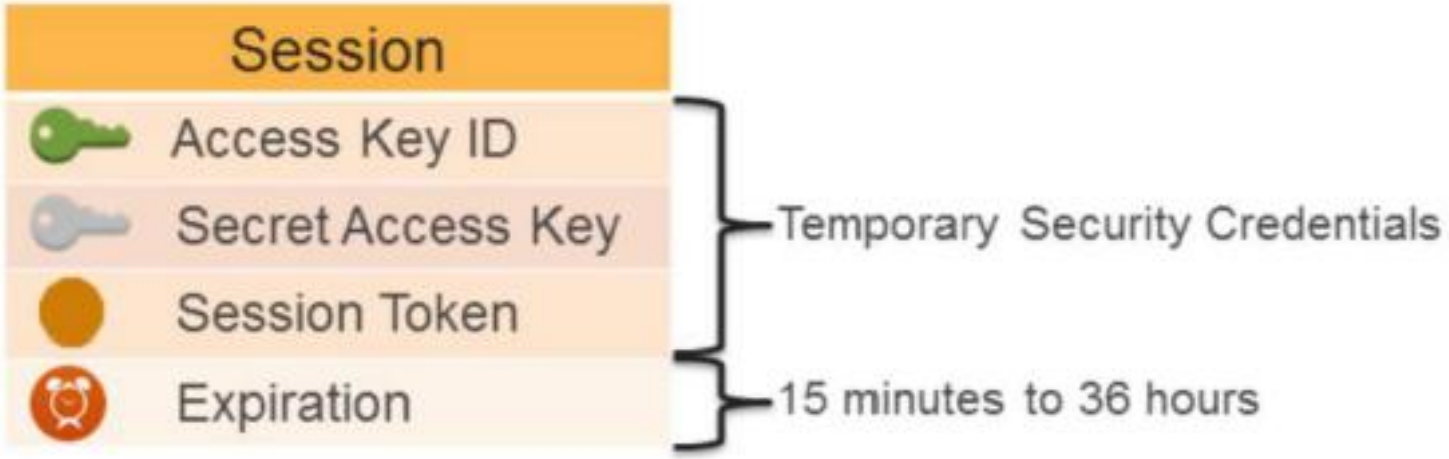
- Provides temporary access
- Eliminates the need for static AWS credentials

Permissions are:

- Defined using IAM policies
- Attached to the role, not to an IAM user or group



# Temporary Security Credentials (STS)

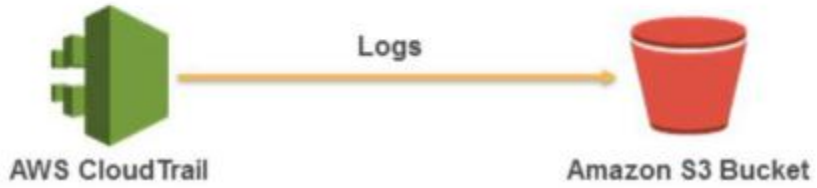



**Use Cases**

- Cross account access
- Federation
- Mobile Users
- Key rotation for Amazon EC2-based apps

# AWS Cloud Trail

- Records AWS API calls for accounts.
- Delivers log files with information to an Amazon S3 bucket.
- Makes calls using the AWS Management Console, AWS SDKs, AWS CLI and higher-level AWS services.



# New to AWS ? What to do on Day 1

1. Stop using the root account as soon as possible.

   The root account has completely unrestricted access to your resources.

   To stop using the root account, take the following steps:

   1) With the root account, create an IAM user for yourself.
   2) Create an IAM group and give it full administrator permissions.
   3) Sign in with your IAM user credentials.
   4) Store your root account credentials in a very secure place. Disable and remove your root account access keys, if you have them.

# AWS Day 1

2. Require multi-factor authentication for access.
    a. Require MFA for your root account and all IAM users.
    b. You can also use MFA to control access to AWS service APIs.

Software MFA options: AWS Virtual MFA, Google Authenticator, Authenticator (Windows phone app), or SMS notification

Hardware MFA options: Key fob or display card offered by Gemalto: onlineoram.gemalto.com

# Day 1

3. Enable AWS CloudTrail.

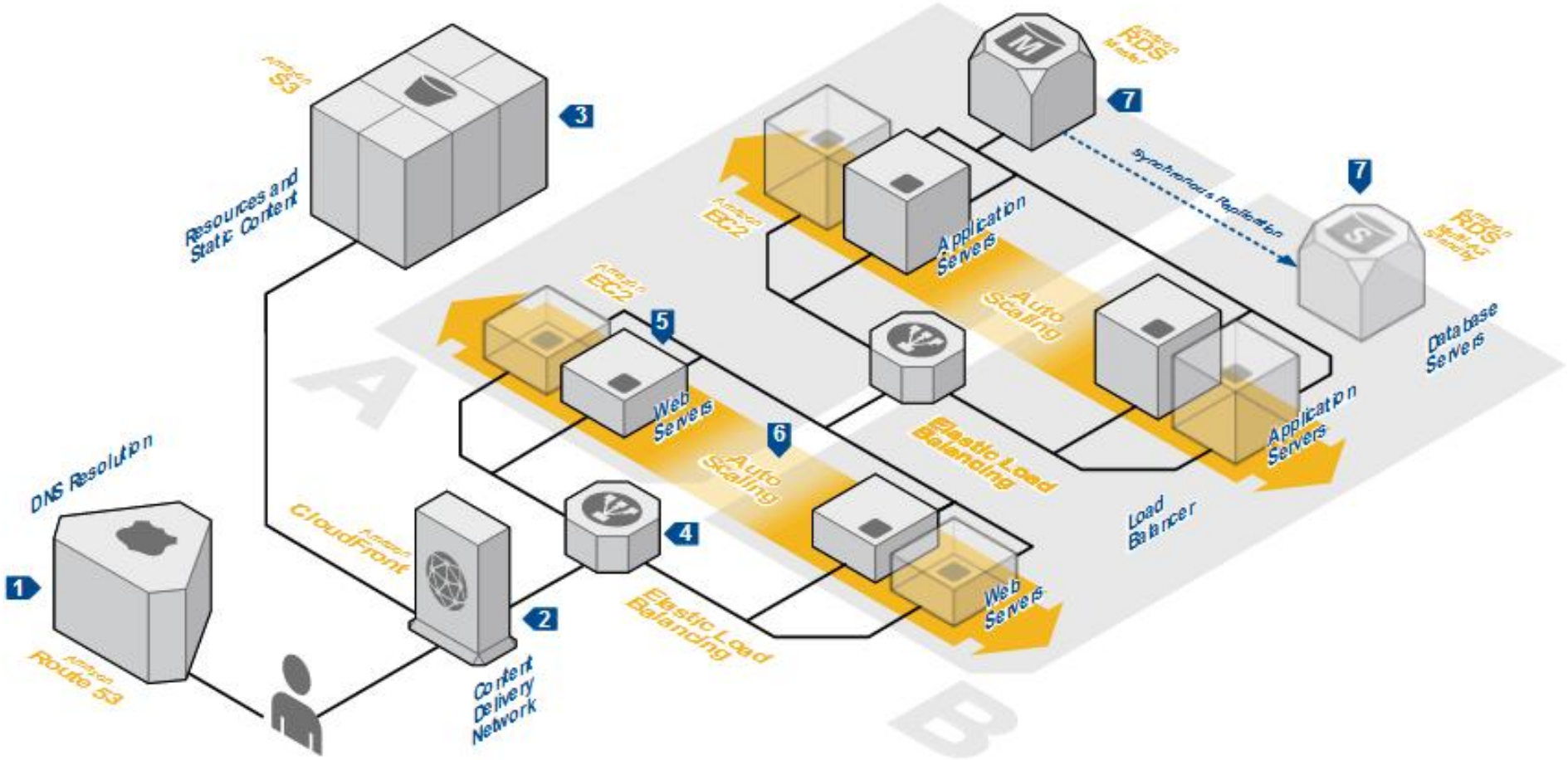
   AWS CloudTrail logs all API requests to resources in your account.

   1) Via the CloudTrail console: create a trail, give it a name, apply it to all regions, and enter a name for the new Amazon S3 bucket that the logs will be stored in.
   2) Ensure that the Amazon S3 bucket you use for CloudTrail has its access restricted to only those who should have access, such as admins.

# Public Cloud Providers – Sample Services Mapping

| AWS Services | Google Cloud Platform Equivalent | AZURE Cloud Platform Equivalent |
|---|---|---|
| **Compute** | **Compute** | **Compute** |
| Amazon EC2 | Google Compute Engine | Azure VMs |
| Amazon EC2 Container Services | Google Container Engine - GKE | Azure Container Services |
| | Kubernetes as a Service | - Supports Kubernetes, Mesos DC/OS, Docker Swarm |
| AWS Elastic Beanstalk | Google App Engine | Azure Web Apps |
| | | Azure Service Fabric |
| **Storage** | **Storage** | **Storage** |
| Amazon Glacier/S3 Standard | Google Cloud Storage Nearline | Azure Blob Storage - Cool |
| Amazon S3 | Google Cloud Storage Standard | Azure Blob Storage |
| Amazon EC2 Container Registry | Google Container Registry | Azure Container Registry |
| Amazon EBS | Google Compute Engine Persistent Disks | Azure Managed Disks |
| Amazon EFS | ZFS/Avere | Azure Files |
| **Database** | **Database** | **Database** |
| Amazon Dynamo DB | Google Cloud Datastore | Azure DocumentDB |
| Amazon RDS | Google Cloud SQL | Azure SQL Database |
| Amazon EMR / Amazon Data Pipeline | Google Cloud Dataflow | Azure HDInsight / Azure Data Factory |
| Amazon Kinesis and Amazon SQS | Google Cloud Pub/Sub | Azure Event Hub, Kafka / Azure Queues |
| Amazon Redshift | Google BigQuery | Azure SQL Data Warehouse |
| Amazon SimpleDB | Google Cloud Datastore | Azure Table Storage |
| | | Azure Data Lake |

# AWS Reference Architectures - *Web Application Hosting*




https://aws.amazon.com/architecture/

# Resources

- AWS Overview: https://d0.awsstatic.com/whitepapers/aws-overview.pdf
- AWS Architecture Center: https://aws.amazon.com/architecture/
- AWS SlideShare Channel: https://www.slideshare.net/AmazonWebServices/