DEMYSTIFYING AWS

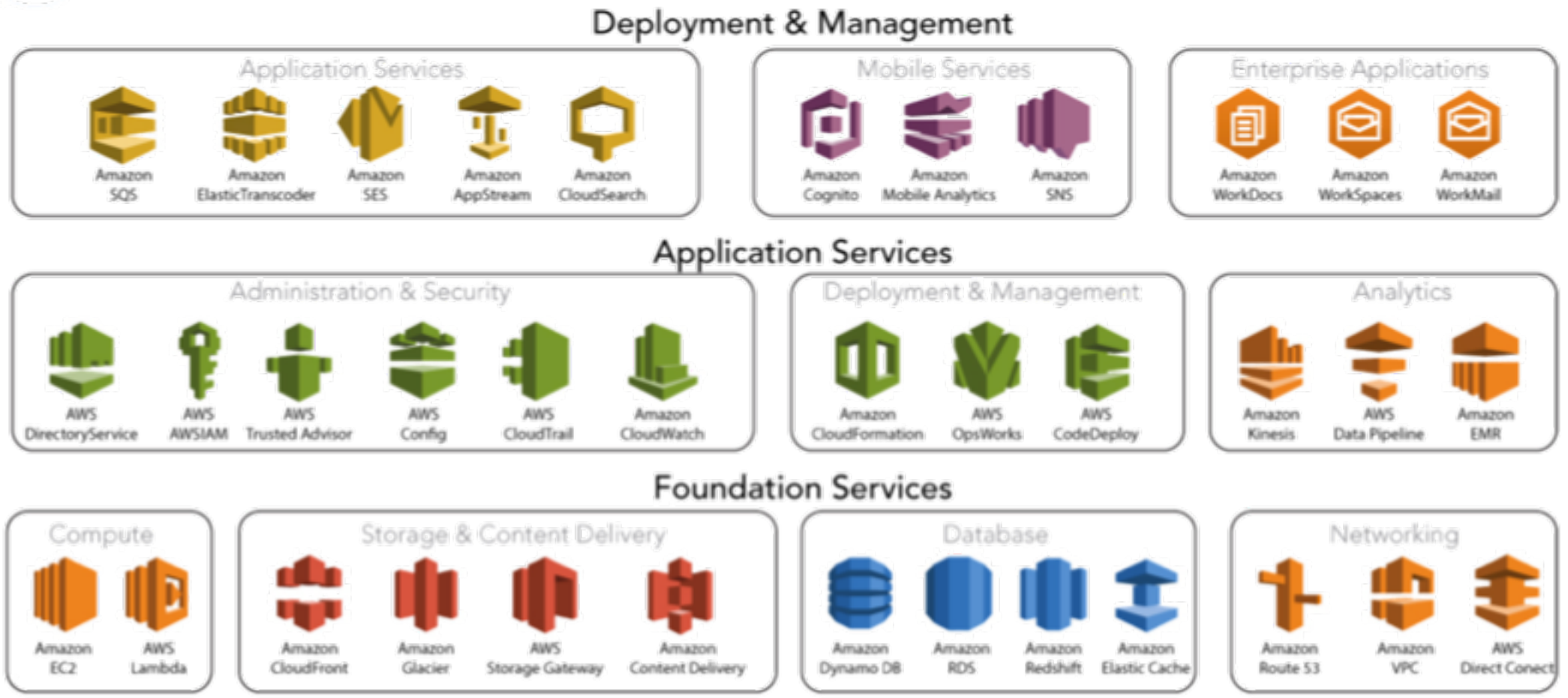
LARRY ELLISON ON "CLOUD COMPUTING"

(CIRCA 2009)

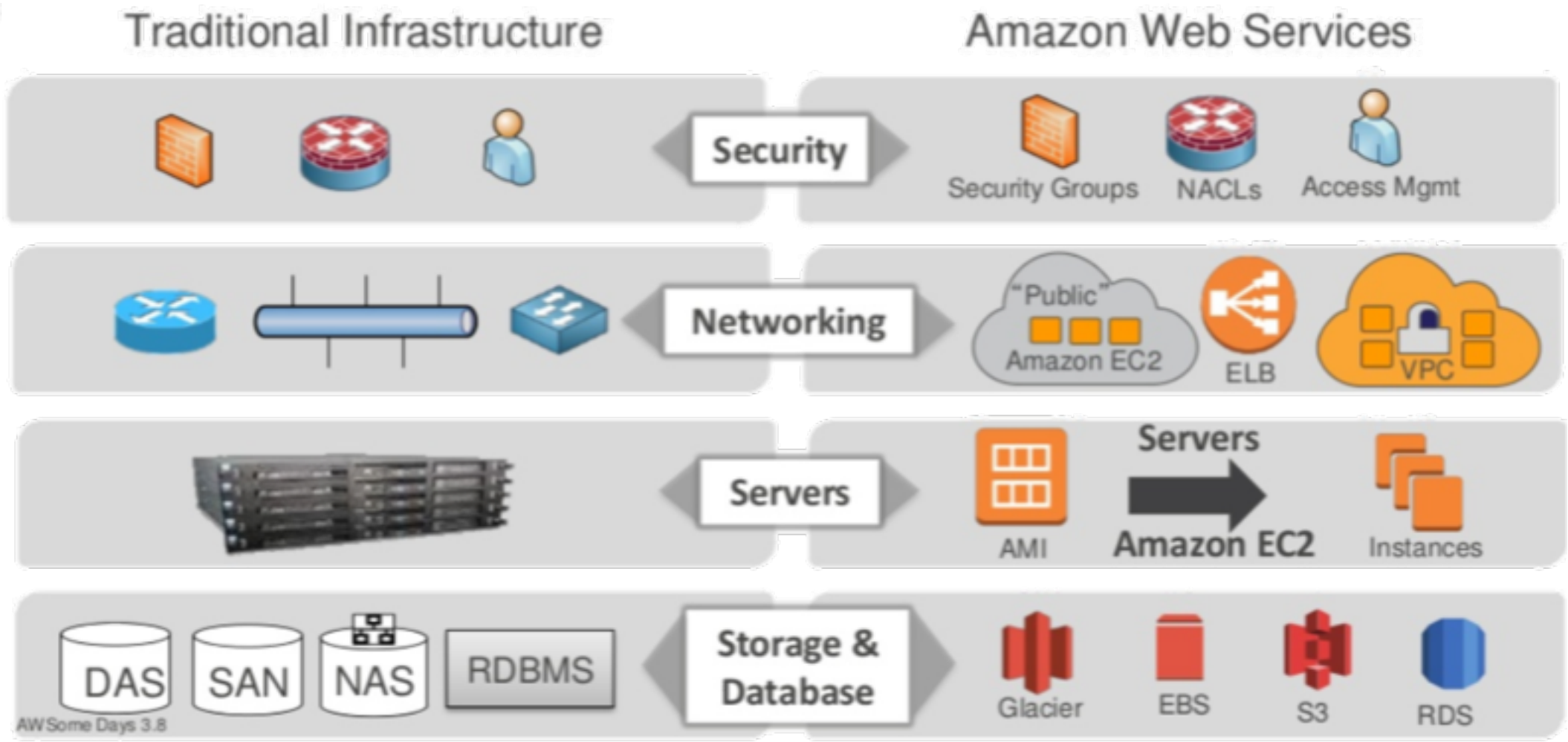
DEMYSTIFYING AWS

- What is "Amazon Web Services"
- Services Overview
- Demo
- Questions

AWS STACK

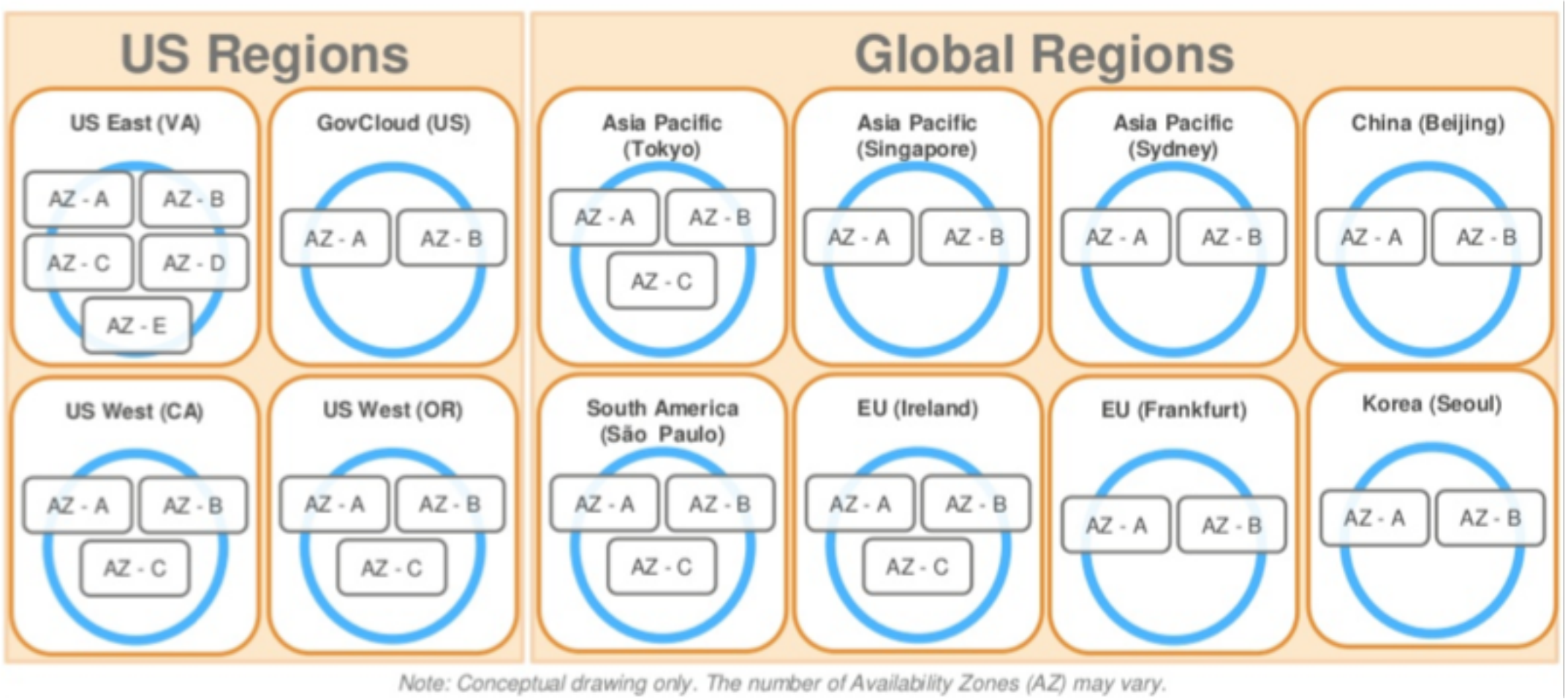

Deployment & Management
Application Services
Amazon SQS
Amazon ElasticTranscoder
Amazon SES
Amazon AppStream
Amazon CloudSearch
Mobile Services
Amazon Cognito
Amazon Mobile Analytics
Amazon SNS
Enterprise Applications
Amazon WorkDocs
Amazon WorkSpaces
Amazon WorkMail
Application Services
Administration & Security
AWS DirectoryService
AWS AWSIAM
AWS Trusted Advisor
AWS Config
AWS CloudTrail
Amazon CloudWatch
Deployment & Management
Amazon CloudFormation
AWS OpsWorks
AWS CodeDeploy
Analytics
Amazon Kinesis
AWS Data Pipeline
Amazon EMR
Foundation Services
Compute
Amazon EC2
AWS Lambda
Storage & Content Delivery
Amazon CloudFront
Amazon Glacier
AWS Storage Gateway
Amazon Content Delivery
Database
Amazon Dynamo DB
Amazon RDS
Amazon Redshift
Amazon Elastic Cache
Networking
Amazon Route 53
Amazon VPC
AWS Direct Conect

CORE INFRASTRUCTURE AND SERVICES



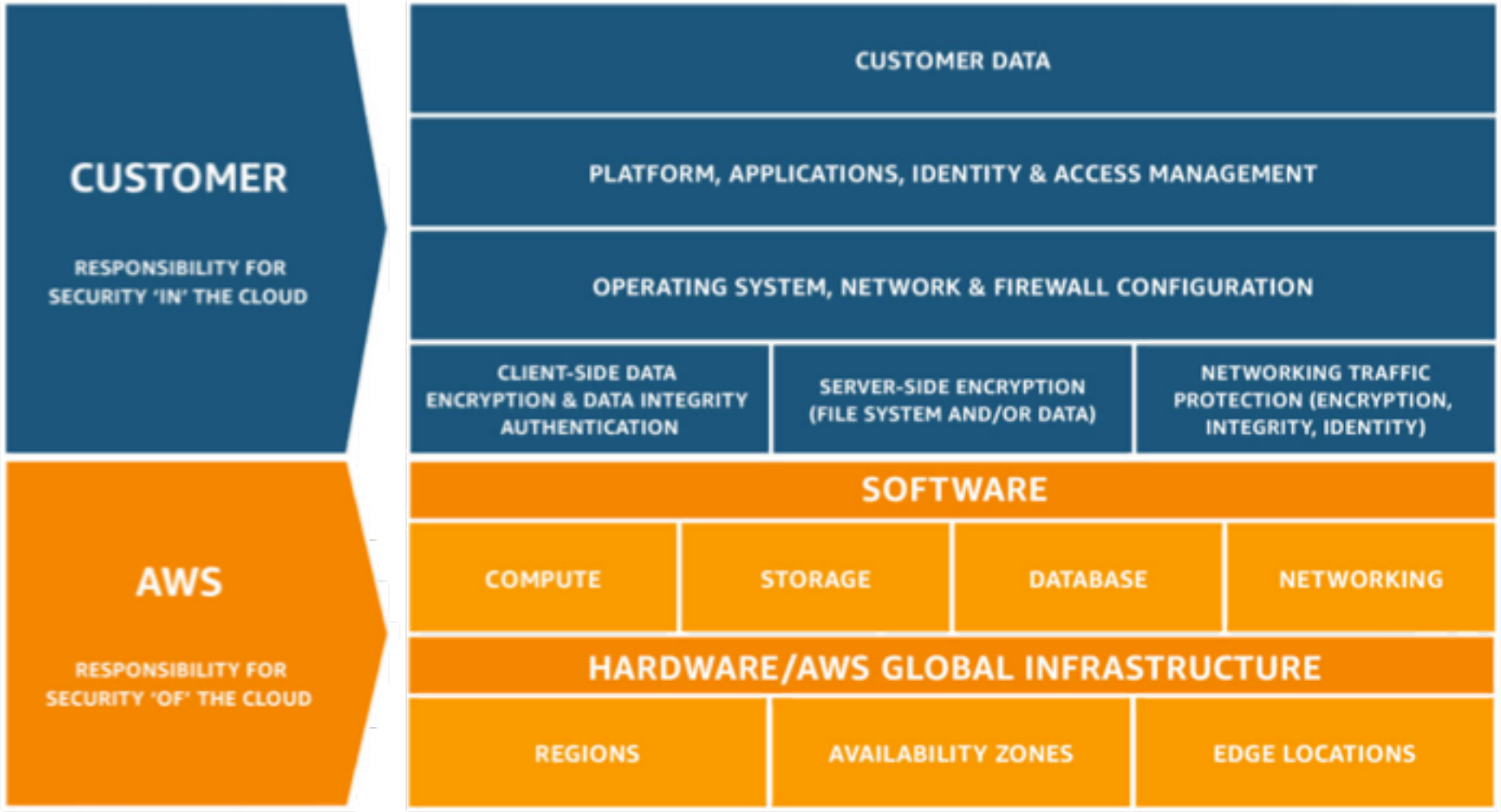
GLOBAL INFRASTRUCTURE

AVAILABILITY ZONES



Note: Conceptual drawing only. The number of Availability Zones (AZ) may vary.

SHARED RESPONSIBILITY MODEL



LET'S GO BUILD US A CLOUD…

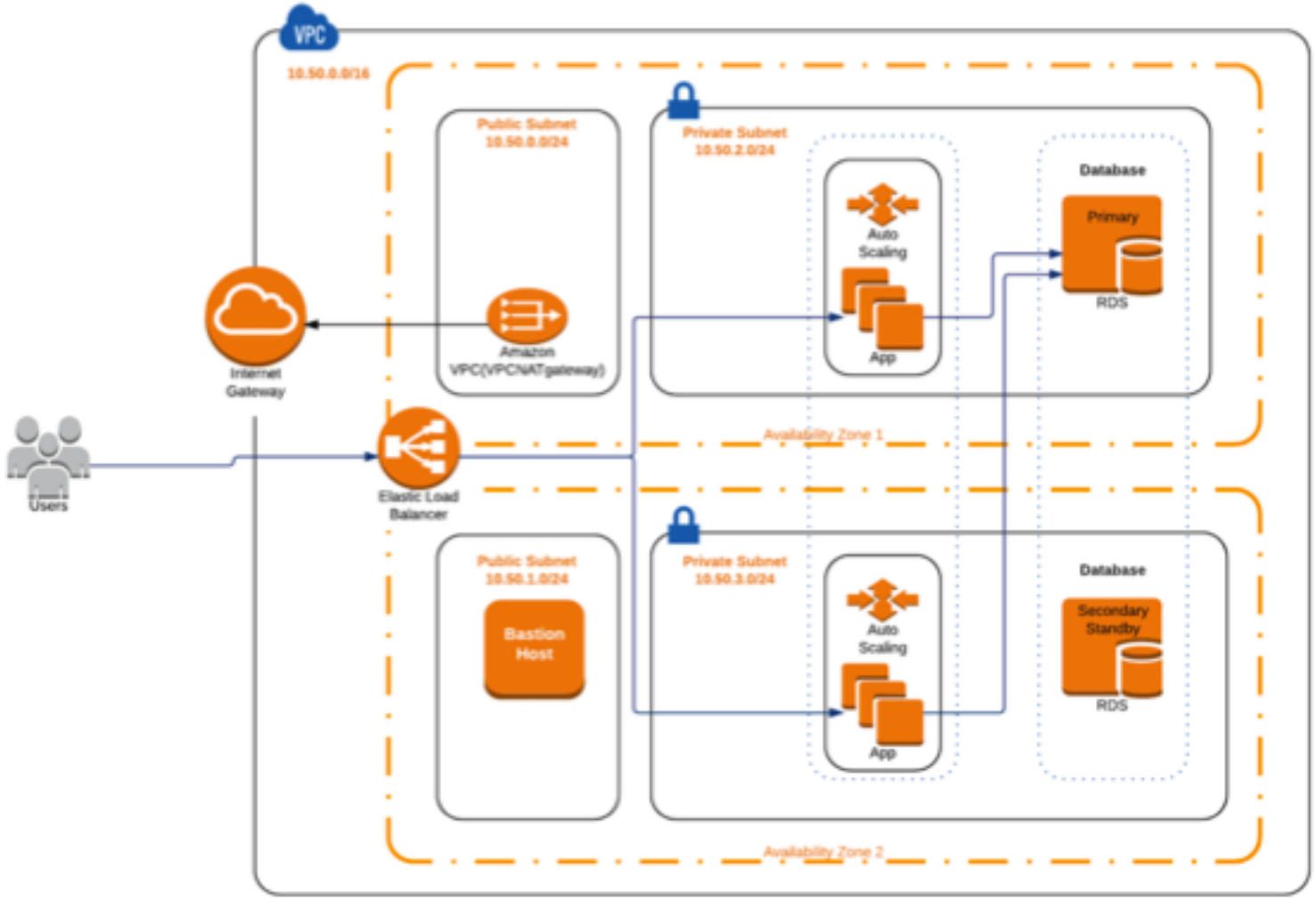
CLOUD CHECKLIST

- ❑ Networking
- ❑ Security
- ❑ Storage
- ❑ Compute
- ❑ Database
- ❑ Serverless
- ❑ Deployment

"VIRTUAL PRIVATE CLOUD"

- Commonly referred to a a VPC
- Networking layer for Amazon Elastic Compute Cloud (Amazon EC2)
- Logically defined network that is unique to you
- Allows you to create your own virtual Network within AWS
- Allows you to select your own IP Address Range(s), Subnet(s), Route Table(s), Gateway(s) and security settings
- Security groups are used to protect individual machines, while Access Control Lists are used to control subnets
- Allows fine grained control over individual subnet routing, network traffic, etc.
- For a hybrid approach, you can connect back to your on-premesis data center using a customer provided VPN or an AWS offering called "Direct Connect"

“VIRTUAL PRIVATE CLOUD”



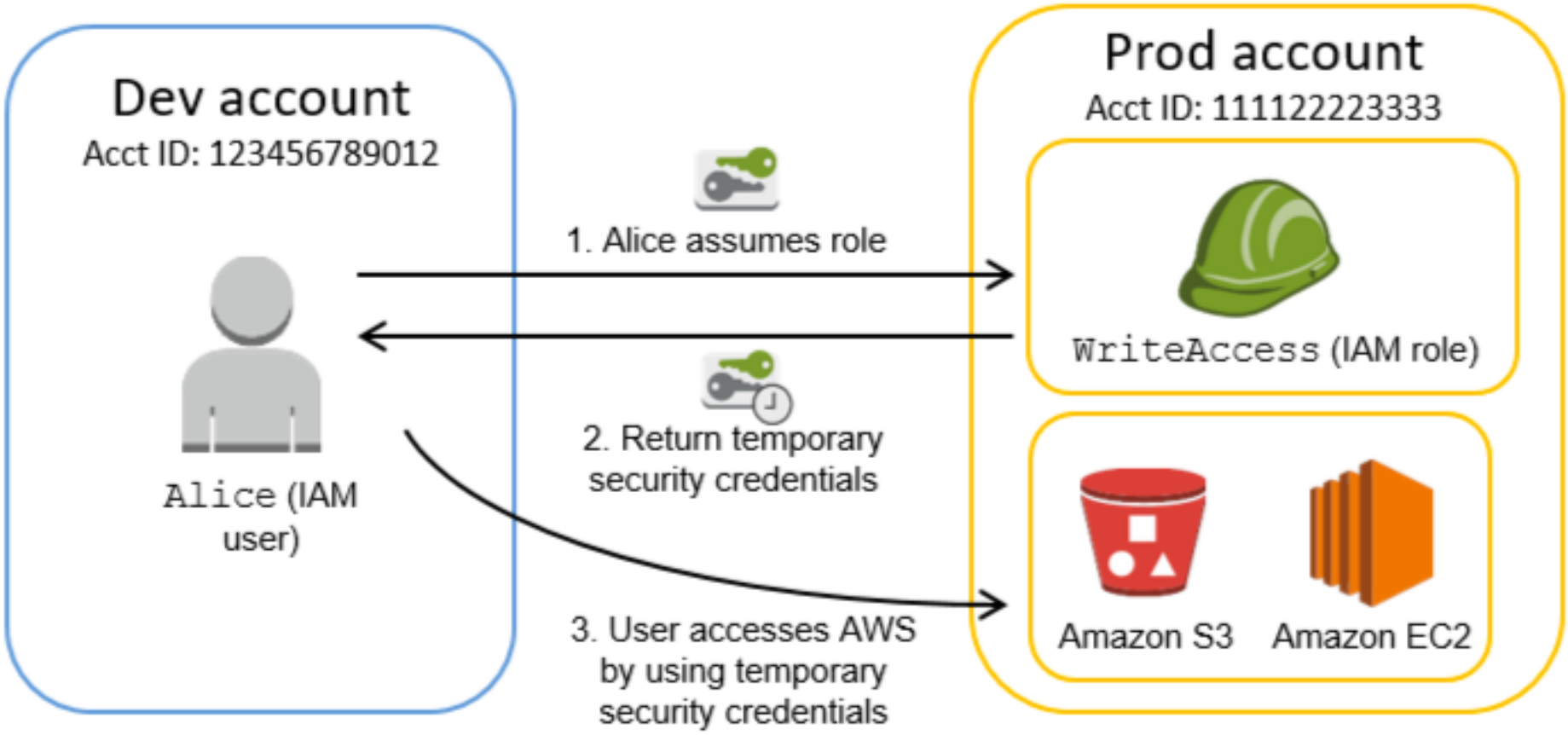
CLOUD CHECKLIST

- ✓ Networking
- ❑ Security
- ❑ Storage
- ❑ Compute
- ❑ Database
- ❑ Serverless
- ❑ Deployment

IDENTITY & ACCESS MANAGEMENT (IAM)

- Control user's ability to access resources within the AWS Ecosystem
- Manage the programmatic and interactive access by users and AWS resources
- Generate and revoke API Keys, User Credentials; enforce Multi-Factor Authentication
- Components of IAM
  - **Users**: Individual end users who will interact with the platform
  - **Group**: A logical collection of users to which permissions can be assigned
  - **Policies**: A permissions document in JSON format in which you define what actions and resources the role can use.
  - **Role**: A set of permissions that grant access to actions and resources in AWS
- Manage what a resource can and can not do through policies
  - Each policy contains one or more permissions for a resource
  - A permission contains the effect, service, action, and resource

IDENTITY & ACCESS MANAGEMENT (IAM)



CLOUD CHECKLIST

- ✓ Networking
- ✓ Security
- ❑ Storage
- ❑ Compute
- ❑ Database
- ❑ Serverless
- ❑ Deployment

SIMPLE STORAGE SERVICE (S3)

- Cloud object store
- Allows storing a (theoretical) unlimited amount of data with high durability
- S3 buckets are used to store and access data
  - Bucket names are globally unique
  - A bucket is created within a region, and the data does not leave the region unless explicitly moved by a user
- Minimalistic API, leveraging the traditional HTTP Verbs of **GET, PUT, POST, DELETE**
- Offers up to 99.999999999% durability for storage, and 99.99% availability
- Eventually consistent, but offers read-after-write for new object puts
- Buckets can be made public to the world, and can be used to host files
- Several tiers of storage, depending on use case and availability needs
- Cross-region replication can be enabled to enable high availability

S3 STORAGE CLASSES

	Standard	Infrequently Access	One Zone-Infrequent Access	Glacier
Durability	11 Nines (99.999999999%)	11 Nines (99.999999999%)	11 Nines (99.999999999%) in single AZ	11 Nines (99.999999999%)
Availability	4 Nines (99.99%)	3 Nines (99.9%)	2-and-a-half Nines (99.5%)	N/A
Cross AZ Replication	3+ AZs	3+ AZs	1 AZ	Multiple AZs
Access Time	Low Latency	Low Latency	Low Latency	High Latency (i.e. hours)
Use Case	General item storage	long-term storage, backups, and DR	Easily re-creatable data, Cross-Region Replica	Archival Data

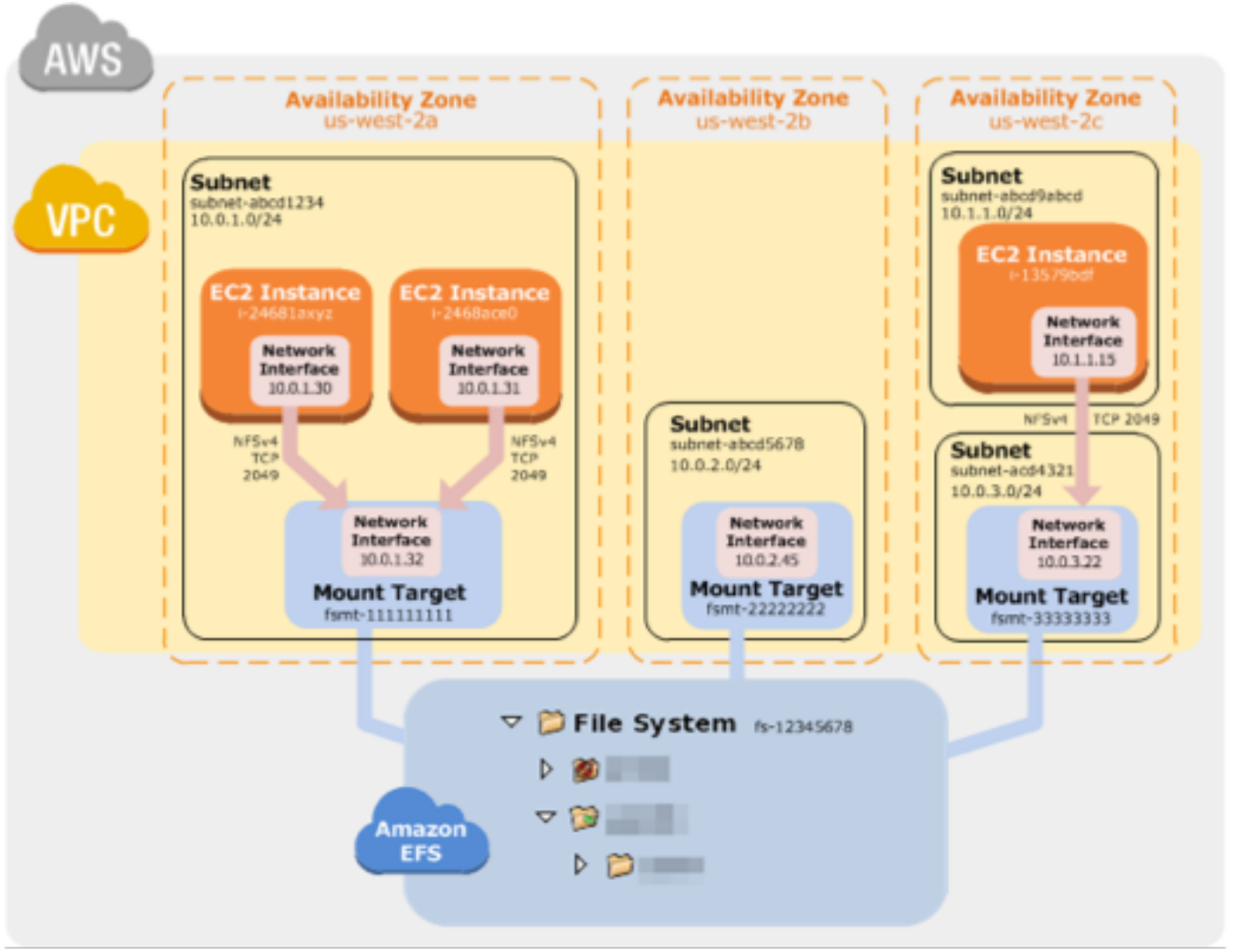
ELASTIC BLOCK STORAGE (EBS)

- Provides block level storage volumes for use with EC2 instances
- EBS volumes are highly available and reliable storage volumes that can be attached to any running instance that is in the same Availability Zone.
- EBS volumes which are attached to an EC2 instance are exposed as storage volumes that persist independently from the life of the instance. With Amazon EBS, you pay only for what you use.
- Recommended when data must be quickly accessible and requires long-term persistence.
- EBS volumes are particularly well-suited for use as the primary storage for file systems, databases, or for any applications that require fine granular updates and access to raw, unformatted, block-level storage.
- Amazon EBS is well suited to both database-style applications that rely on random reads and writes, and to throughput-intensive applications that perform long, continuous reads and writes.

ELASTIC FILE STORE (EFS)

- Provides simple, scalable file storage for use with Amazon EC2
- Storage capacity is elastic, growing and shrinking automatically as you add and remove files
- Supports the Network File System versions 4.0 and 4.1 (NFSv4) protocol, so the applications and tools that you use today work seamlessly with Amazon EFS
- Designed to be highly scalable, highly available, and highly durable. Amazon EFS file systems store data and metadata across multiple Availability Zones in a region
- Allow massively parallel access from Amazon EC2 instances to your data
- Supports two forms of encryption for file systems, encryption in transit and encryption at rest
- Use cases include Big Data workloads, Data Analysis, User Home Folders, Container Storage, Database Backups
- EFS File Sync works over any network connection, including with AWS Direct Connect

ELASTIC FILE STORE



CLOUD CHECKLIST

- ✓ Networking
- ✓ Security
- ✓ Storage
- ❑ Compute
- ❑ Database
- ❑ Serverless
- ❑ Deployment

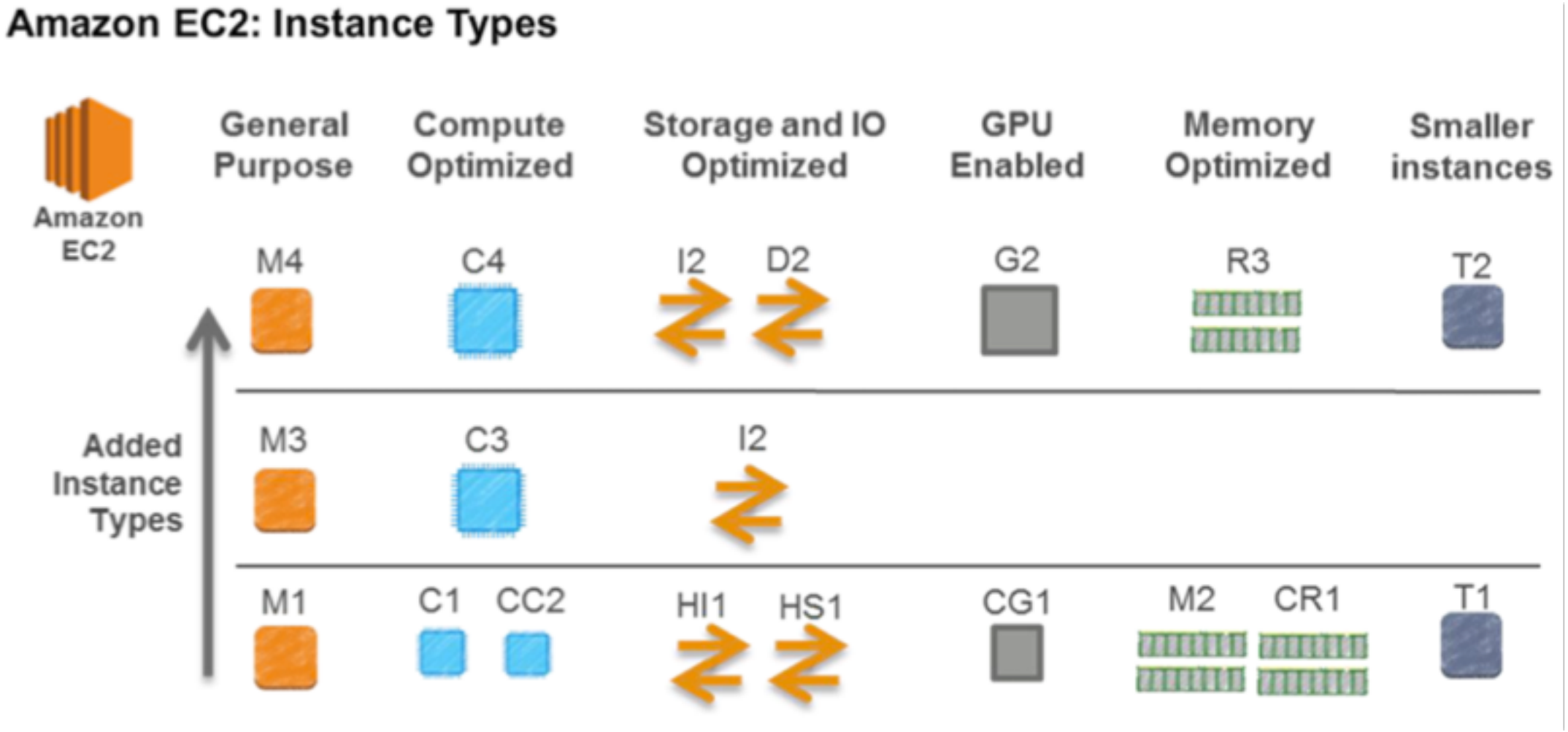
AMAZON EC2

- Amazon Elastic Compute Cloud (EC2)
- Scalable compute capacity in the cloud
- Run-of-the-mill Virtual Machines that can be customized to fit your needs
- Can be used to run applications, web servers, database servers
- Pay as you go– you only pay for what and when you use a resource
  - Discounted options for non-peak hours
- Scalable and fault tolerable
- Variety of different offerings depending on the use case
- Data is Ephemeral or Persistent (disk backed)

EC2 KEY COMPONENTS

Component	Description
Instances	Virtual computing environments
Amazon Machine Images (AMIs)	Preconfigured templates for your instances
Key Pairs	Secure login information for your instances
Instance Store Volumes	Storage volumes for temporary data
Amazon Elastic Block Store (EBS)	Persistent storage volumes for your data
Regions and *Availability Zones*	Multiple physical locations for your resources
Security Groups	A firewall that enables you to specify the protocols, ports, and source IP ranges
Elastic IP Addresses	Static IPv4 addresses for dynamic cloud computing
Tags	Metadata assigned to your Amazon EC2 resources

EC2 INSTANCE TYPES



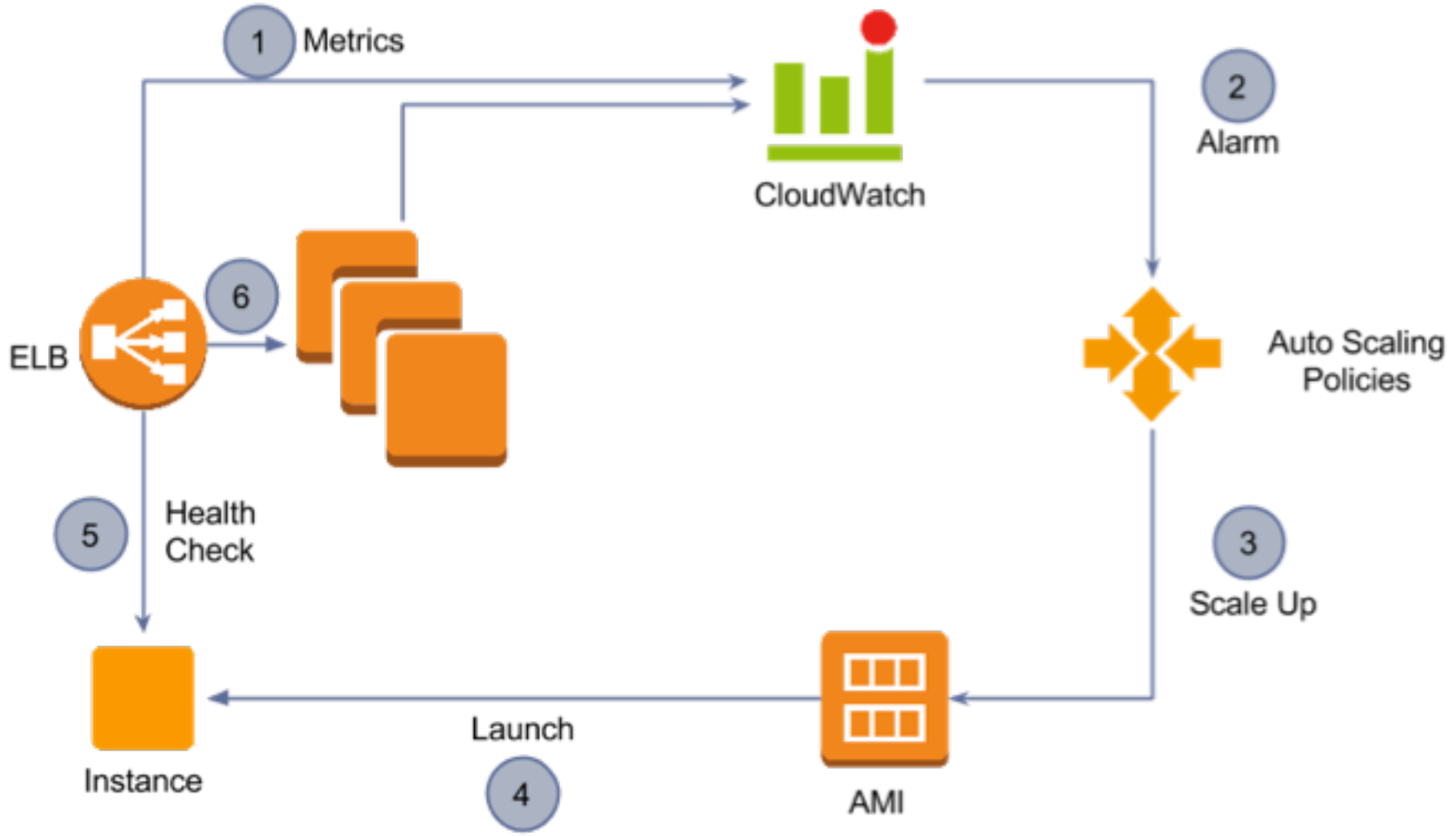
AMAZON MACHINE IMAGES

- AMIs Detail the following information
  - Operating System
  - Initial State of Patching
  - Application or System Software
- Sources of AMIs
  - Published and released by AWS (Include Linux, RHEL, and Amazon Distro, Windows 08/12/12R2/16)
    - Results in the default OS Configuration (think installing from an ISO)
  - AWS Marketplace
    - Pre-Configured by Amazon partners or the community
  - Generated from Existing Instances
    - Created from your already existing instances, and replicated onto additional instances
  - Uploaded Virtual Servers
    - Uploaded from raw disk, or VHD/OVA images

AUTOSCALING

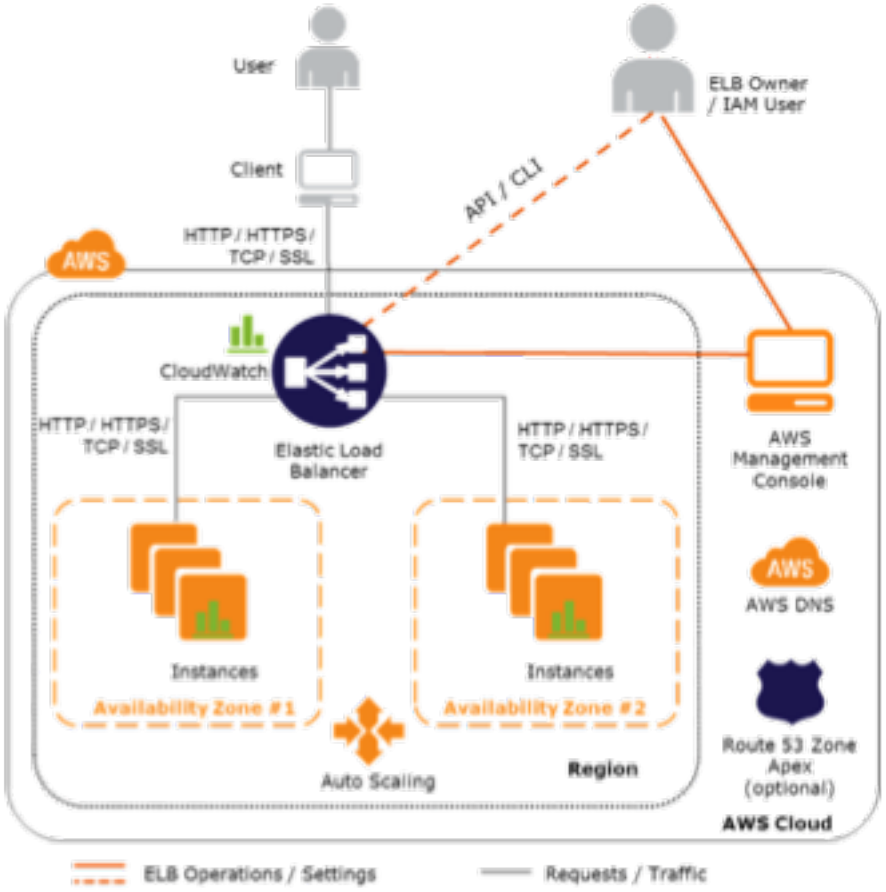
- An important advantage of deploying to the cloud is autoscaling
- Ability to launch and release servers in response to workloads
- Allows you to scale out based on the the criteria you defined
- Autoscaling plans include Maintain Current Instant Levels, Manual Scaling, Scheduled Scaling, and Dynamic Scaling
- Launches a new instance based on a template defined
- CloudWatch alarms determine when the autoscaling group should scale up or down using a single metric
- Elastic Load Balancing can be used to direct traffic to instances depending on load or availability of resources

AUTOSCALING



ELASTIC LOAD BALANCING

- Types of Load Balancers
  - **Application** Load Balancers
    - Make routing decisions at the application (HTTP/HTTPS) level
    - Supports HTTPS Termination, sticky sessions, host, path based routing
  - **Network** Load Balancers
    - Makes routing decisions at the transport (TCP/SSL) layer
    - Selects a target from the target group for the default rule using a flow hash routing algorithm
    - Forwards the request without modifying the headers
    - Can handle millions of requests per second
  - **Classic** Load Balancer
    - Makes routing decisions at either the transport layer (TCP/SSL) or the application layer (HTTP/HTTPS)
    - Allows traffic on only one port of an instance



CLOUD CHECKLIST

- ✓ Networking
- ✓ Security
- ✓ Storage
- ✓ Compute
- ❑ Database
- ❑ Serverless
- ❑ Deployment

RELATIONAL DATABASE SERVICE (RDS)

- Service provided by AWS for your data needs
- Options for EC2 hosted , and Managed Database Services
- For a managed service, Installation and provisioning of hardware and resources is handled by AWS
- Supports open and commercial database engines
  - MySQL
  - PostgreSQL
  - MariaDB
  - M$SQL
  - Oracle
- Custom RDS product "Aurora" which is a MySQL based cloud native database
- Controlled by API or connecting to the instance
- Supports Cross Region Replication and Multi Availability Zone
- 99.95% SLA
- Read Replicas available for PostgreSQL, MaraiaDB, MySQL, Aurora

COMPARISON OF RESPONSIBILITIES

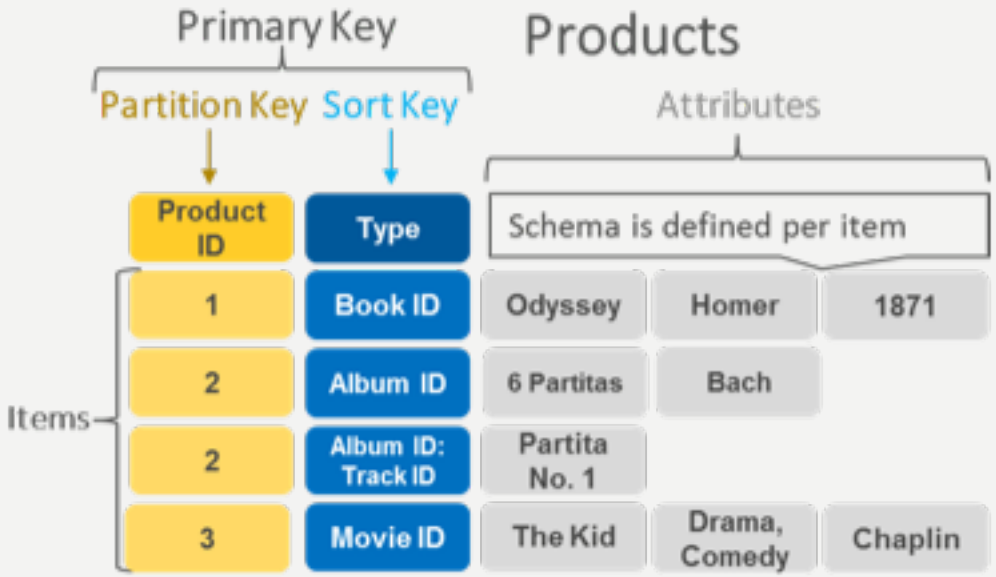
Responsibility	Database On-Premise	Database on EC2	Database on RDS
App Optimization	You	You	You
Scaling	You	You	AWS
High Availability	You	You	AWS
Backups	You	You	AWS
DB Engine Patches	You	You	AWS
Software Installations	You	You	AWS
OS Patches	You	You	AWS
OS Installation	You	AWS	AWS
Server Maintenance	You	AWS	AWS
Rack and Stack	You	AWS	AWS
Power and Cooling	You	AWS	AWS

DYNAMO DB

- Fully managed **NoSQL database service** that provides fast and predictable performance with seamless scalability
- Offloads the administrative burdens of **operating and scaling** a distributed database
- No hardware provisioning, or worry about management of the database infrastructure
- Offers **encryption at rest**
- You can scale up or scale down your tables' throughput capacity without downtime or performance degradation
- Allows for **full backups of your tables** for long-term retention and archival for regulatory compliance needs
- Automatically spreads the data and traffic for your tables over a sufficient number of servers to handle your throughput and storage requirements
- Automatically replicated across multiple Availability Zones in an AWS region

DYNAMO DB TABLE

- Tables, Items, and Attributes
  - A *table* is a collection of data
  - An *item* is a group of attributes that is uniquely identifiable among all of the other items.
  - An *attribute* is a fundamental data element, something that does not need to be broken down any further.
- Primary Key
  - Unique identifier, or primary key, that distinguishes the item from all of the others in the table
- Secondary Indexes
  - Lets you query the data in the table using an alternate key, in addition to queries against the primary key
- DynamoDB Streams
  - An optional feature that captures data modification events in DynamoDB tables. The data about these events appear in the stream in near real time, and in the order that the events occurred.



CLOUD CHECKLIST

- ✓ Networking
- ✓ Security
- ✓ Storage
- ✓ Compute
- ✓ Database
- ❑ Serverless
- ❑ Deployment

AWS LAMBDA

- Compute service that lets you run code without provisioning or managing servers
- AWS Lambda executes your code only when needed and scales automatically, from a few requests per day to thousands per second
- You pay only for the compute time you consume - there is no charge when your code is not running
- Currently supports Node.js, Java, C#, Go, Python
- You can use AWS Lambda to run your code:
  - in response to events, such as changes to data in an Amazon S3 bucket or an Amazon DynamoDB table
  - to run your code in response to HTTP requests using Amazon API Gateway
  - to invoke your code using API calls made using AWS SDKs
- When using AWS Lambda, you are responsible only for your code
- AWS Lambda manages the compute fleet that offers a balance of memory, CPU, network, and other resources

ADDITIONAL SERVERLESS SERVICES

- Amazon SNS
  - A flexible, fully managed pub/sub messaging and mobile notifications service for coordinating the delivery of messages to subscribing endpoints and clients
- Amazon SQS
  - A fully managed message queuing service that makes it easy to decouple and scale microservices, distributed systems, and serverless applications
- AWS Step Functions
  - Makes it easy to coordinate the components of distributed applications and microservices using visual workflows. Step Functions are a reliable way to coordinate components and step through the functions of your application
- Amazon Kinesis
  - Platform for streaming data on AWS, offering powerful services to make it easy to load and analyze streaming data
- Amazon Athena
  - An interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL.

CLOUD CHECKLIST

- ✓ Networking
- ✓ Security
- ✓ Storage
- ✓ Compute
- ✓ Database
- ✓ Serverless
- ❑ Deployment

CODEDEPLOY

- AWS CodeDeploy is a deployment service that automates application deployments to Amazon EC2 instances, on-premises instances, or serverless Lambda functions.
- You can deploy a nearly unlimited variety of application content:
  - Code
  - Serverless AWS Lambda functions
  - Web and configuration files
  - Executables, packages, scripts, multimedia files
- AWS CodeDeploy can deploy application content that runs on a server and is stored in Amazon S3 buckets, GitHub repositories, or Bitbucket repositories
- AWS CodeDeploy can also deploy a serverless Lambda function. You do not need to make changes to your existing code before you can use AWS CodeDeploy

CODEDEPLOY COMPUTE PLATFORMS

- **EC2/On-Premises**
  - Describes instances of physical servers that can be Amazon EC2 cloud instances, on-premises servers, or both
  - Applications created using the EC2/On-Premises compute platform can be composed of executable files, configuration files, images, and more.
  - Deployments that use the EC2/On-Premises compute platform manage the way in which traffic is directed to instances by using an in-place or blue/green deployment type
- **AWS Lambda**
  - Used to deploy applications that consist of updated versions of Lambda functions
  - AWS Lambda manages the Lambda functions in a serverless compute environment made up of a high-availability compute structure
  - All administration of the compute resources is performed by AWS Lambda.

CLOUD CHECKLIST

- ✓Networking
- ✓Security
- ✓Storage
- ✓Compute
- ✓Database
- ✓Serverless
- ✓Deployment

DEMO

QUESTIONS?