AWS Solution Architect

Class 1 - Introduction and Getting Started with AWS

Cloud Basics

- Generally speaking, cloud computing can be thought of as anything that involves delivering hosted services over the Internet.
- According to NIST Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (Special Publication 800-145)

Cloud Basics

- Cloud computing provides shared services as opposed to local servers or storage resources
- Enables access to information from most web-enabled hardware
- Allows for cost savings – reduced facility, hardware/software investments, support
- *On-demand self-service*

  A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
- *Broad network access*

  Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

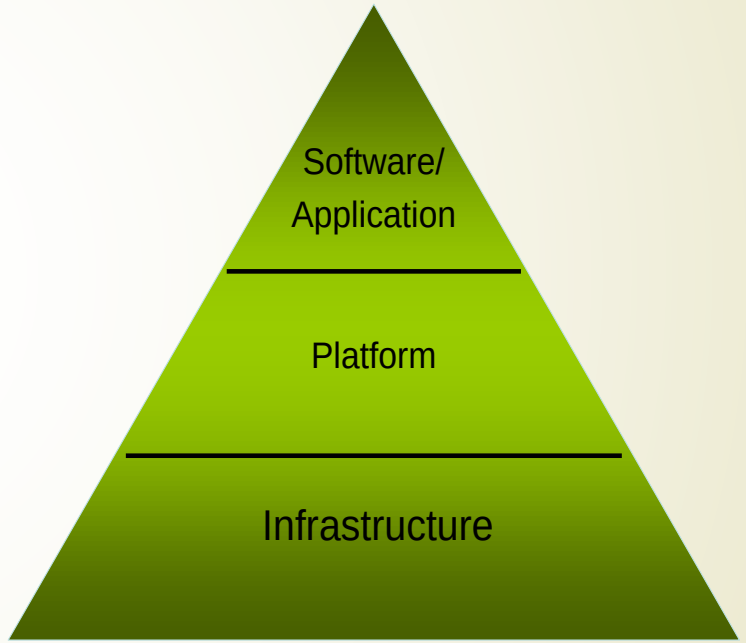
Cloud Basics

- *Resource pooling*
  - The provider's computing resources are pooled to serve multiple consumers
  - Resources can be dynamically assigned and reassigned according to customer demand
  - Customer generally may not care where the resources are physically located but should be aware of risks if they are located offshore
- *Rapid elasticity*
  - Capabilities can be expanded or released automatically (i.e., more cpu power, or ability to handle additional users)
  - To the customer this appears seamless, limitless, and responsive to their changing requirements
- *Measured service*
  - Customers are charged for the services they use and the amounts
  - There is a metering concept where customer resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service

Service Models

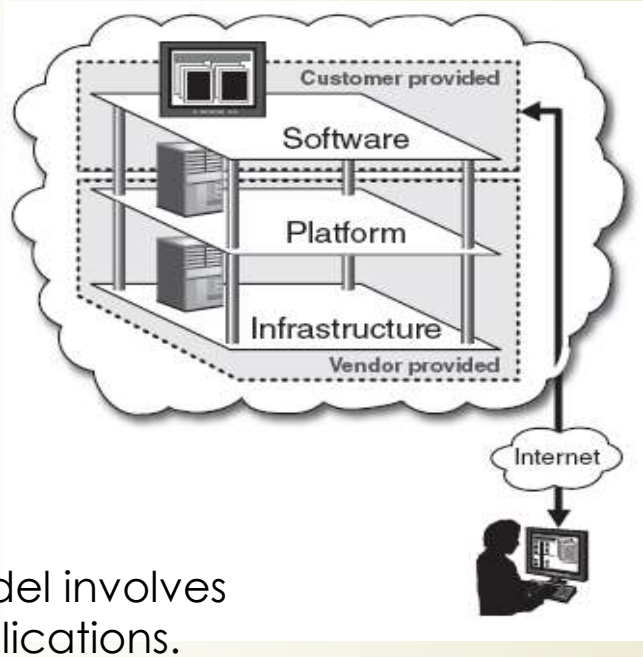
Infrastructure-as-a-Service (IaaS)

- A service model that involves outsourcing the basic infrastructure used to support operations--including storage, hardware, servers, and networking components.
- The service provider owns the infrastructure equipment and is responsible for housing, running, and maintaining it. The customer typically pays on a per-use basis.
- **The customer uses their own platform (Windows, Unix), and applications**



Service Models

- Platform-as-a-Service (PaaS)
- A service model that involves outsourcing the basic infrastructure and platform (Windows, Unix)
- PaaS facilitates deploying applications without the cost and complexity of buying and managing the underlying hardware and software where the applications are hosted.
- The customer uses their own applications



- Software-as-a-Service (SaaS)
- Also referred to as "software on demand," this service model involves outsourcing the infrastructure, platform, and software/applications.
- Typically, these services are available to the customer for a fee, pay-as-you-go, or a no charge model.
- The customer accesses the applications over the internet

Service Models

Software as a Service (SaaS)

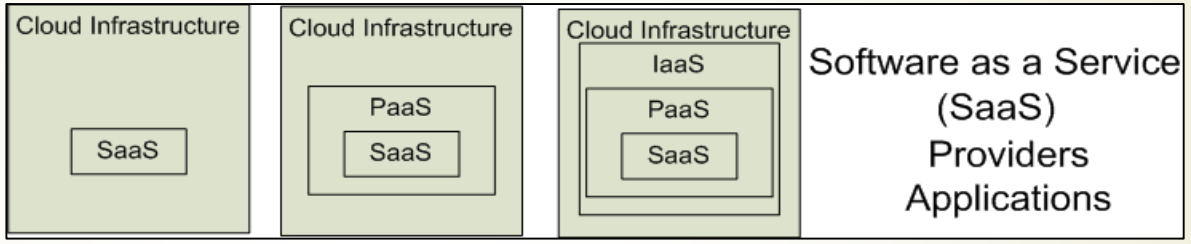
Platform as a Service (PaaS)

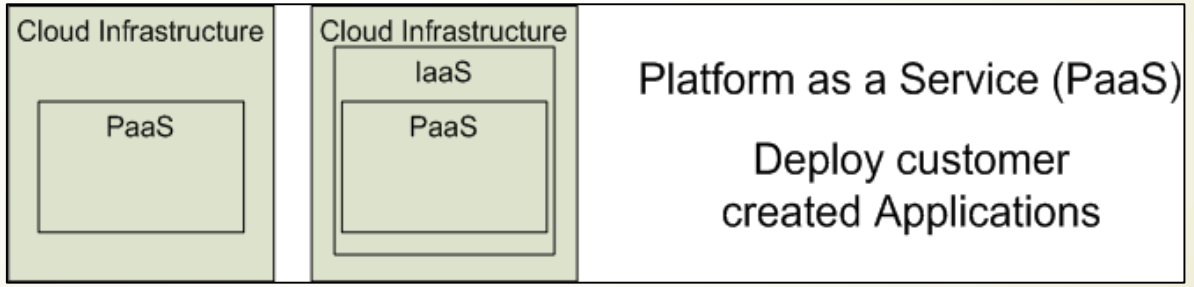
Infrastructure as a Service (IaaS)

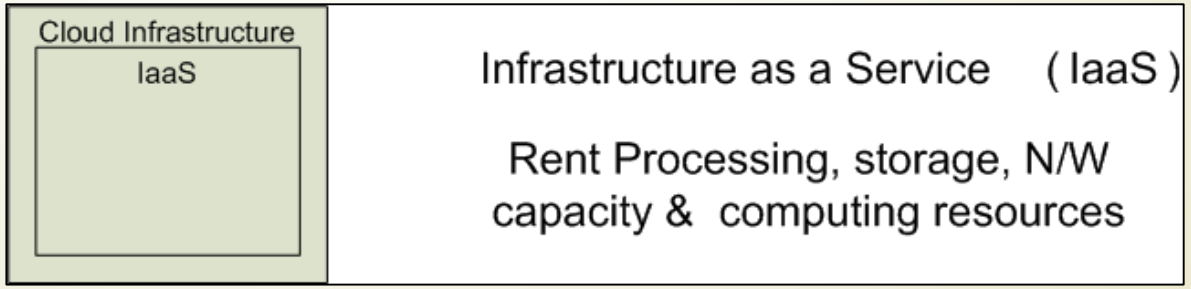
SalesForce CRM

LotusLive

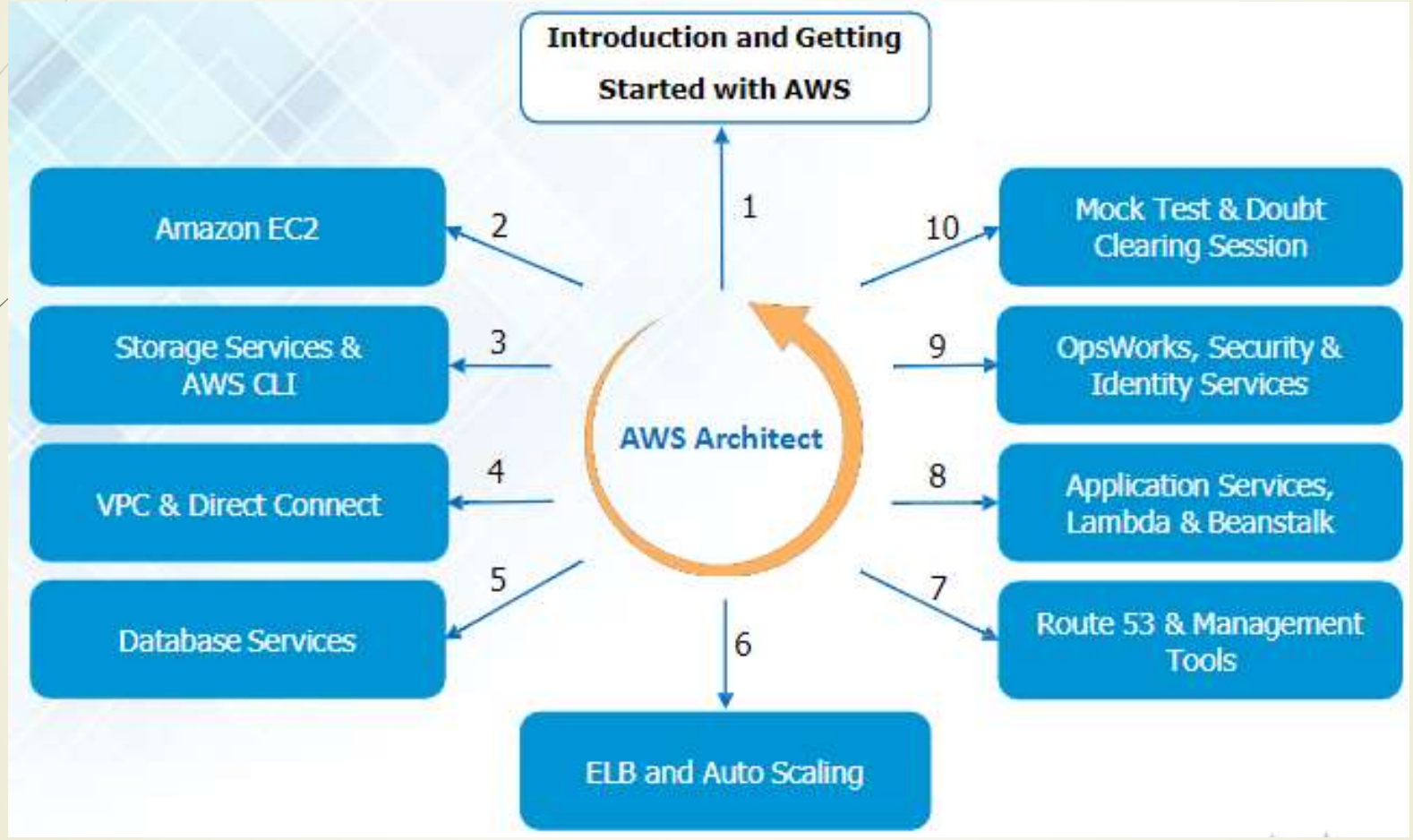
Google App







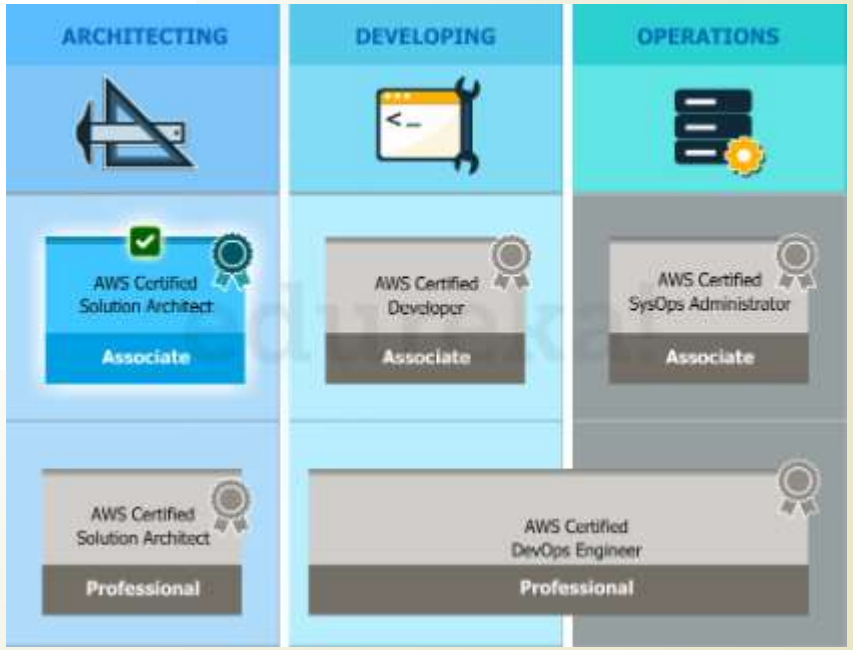
AWS Ecosystem



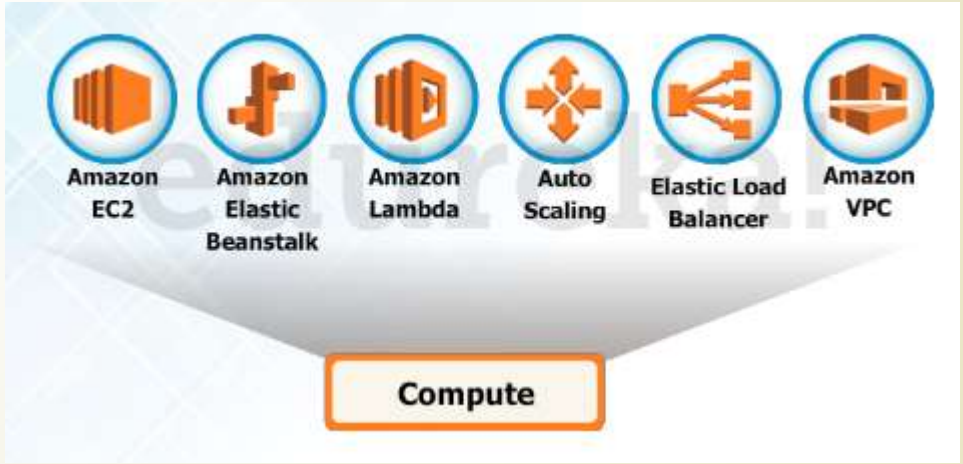
Cloud Computing

→ Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., Networks, Servers, Storage, Applications, and Services)

→ It needs minimal management effort or service provider interaction



AWS Products and Services



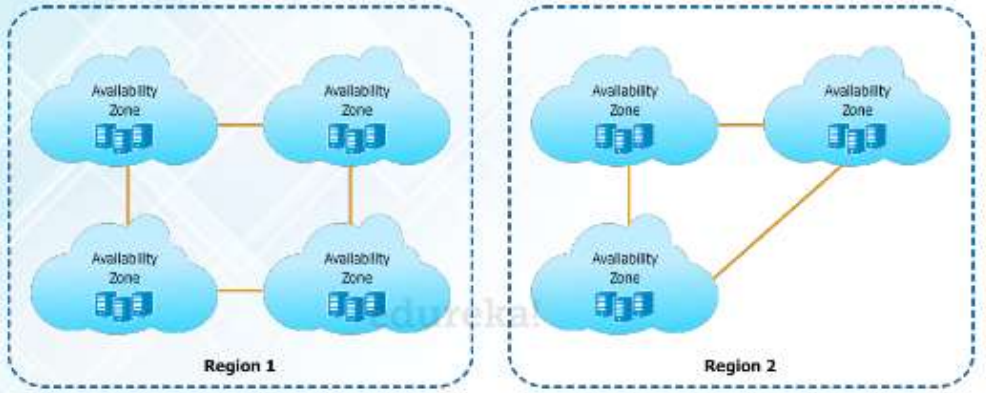
AWS Products and Services

AWS Products and Services

Services	Benefits(1 Year)
Amazon Elastic Cloud Compute	»750 hours/month of Linux, RHEL or SLES t2.micro instance usage »750 hours/month of Windows t2.micro instance usage
Elastic Load Balancer	»750 hours + 15 GB Data Processing
Elastic Block Storage	»30GB in combination of SSD/Magnetic + 2 million I/Os and 1 GB of Snapshot Storage
Amazon Web Services	»15GB of Bandwidth aggregated across all Services »1GB of Regional Data Transfer

Regions and AZ

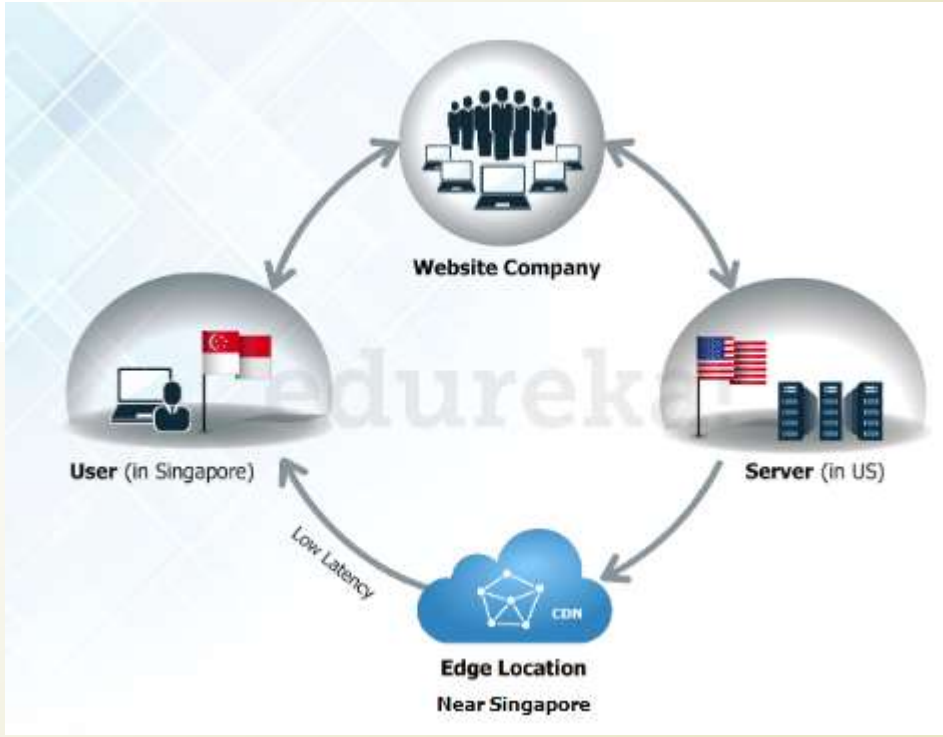
→ Amazon Infrastructure is divided into following categories: **Regions and Availability Zones**



Problem Statement – Global Infra

A recently trended website which offers online streaming of video and its server is located inside the USA.

The international users start complaining about buffering because of the high latency over international links



Why is Edge location preferred?

» We can cache the data from the web site to the AWS Edge Locations nearer to the user using the AWS Cloud Front – Content Delivery Network(CDN).

» Single environment is managed and the data is delivered to the users via a closer location in low latency

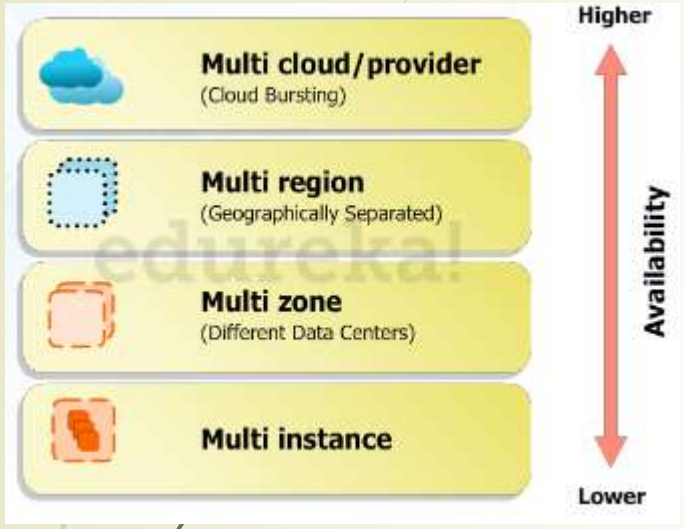
» In edge location servers we have to pay only for the data transfer and requests that you actually use.

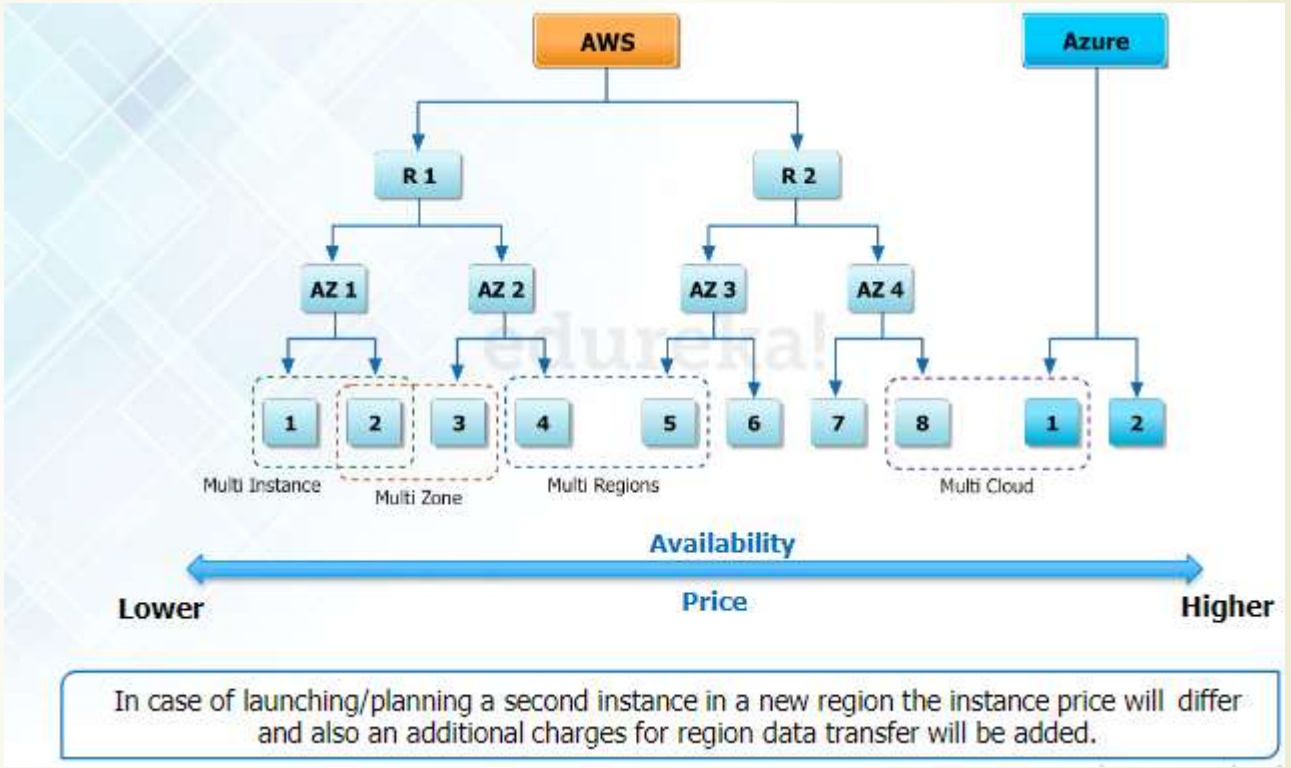
Edge location can be used till certain limits of AWS in terms of users, bandwidth and storage, after that we have to choose a separate region to provide a low latency to the users.

Designed to offer content to end users via a geographically closer location to improve the User Experience

Example: N. Virginia has multiple Edge Locations. Similarly in India, New Delhi and Chennai have Edge Locations

High Availability through Multiple AZ

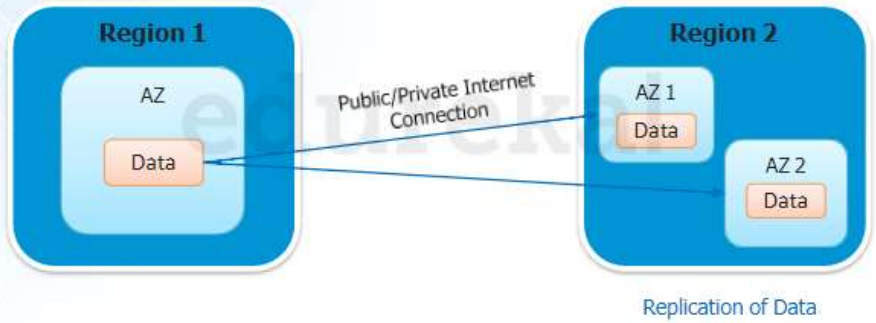




Improving Continuity

→ Can increase redundancy and fault tolerance by replicating data between geographic regions

→ Provides low latency access(continuity) across the globe

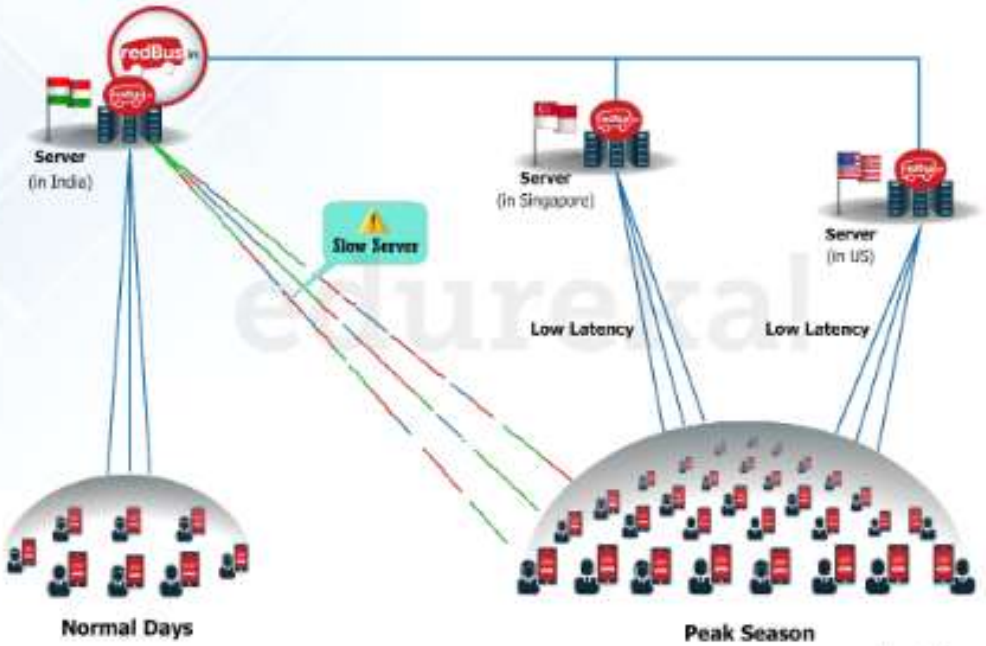


→ Complete control and ownership over the region, where user data is physically located

Geographical Location

→ To provide flexible and low latency services to all the users, Amazon provides the flexibility of expanding the server globally

→ Example: To provide better service to outside users during peak seasons, redBus started expanding globally to new geographic locations such as Singapore and Sydney using AWS.

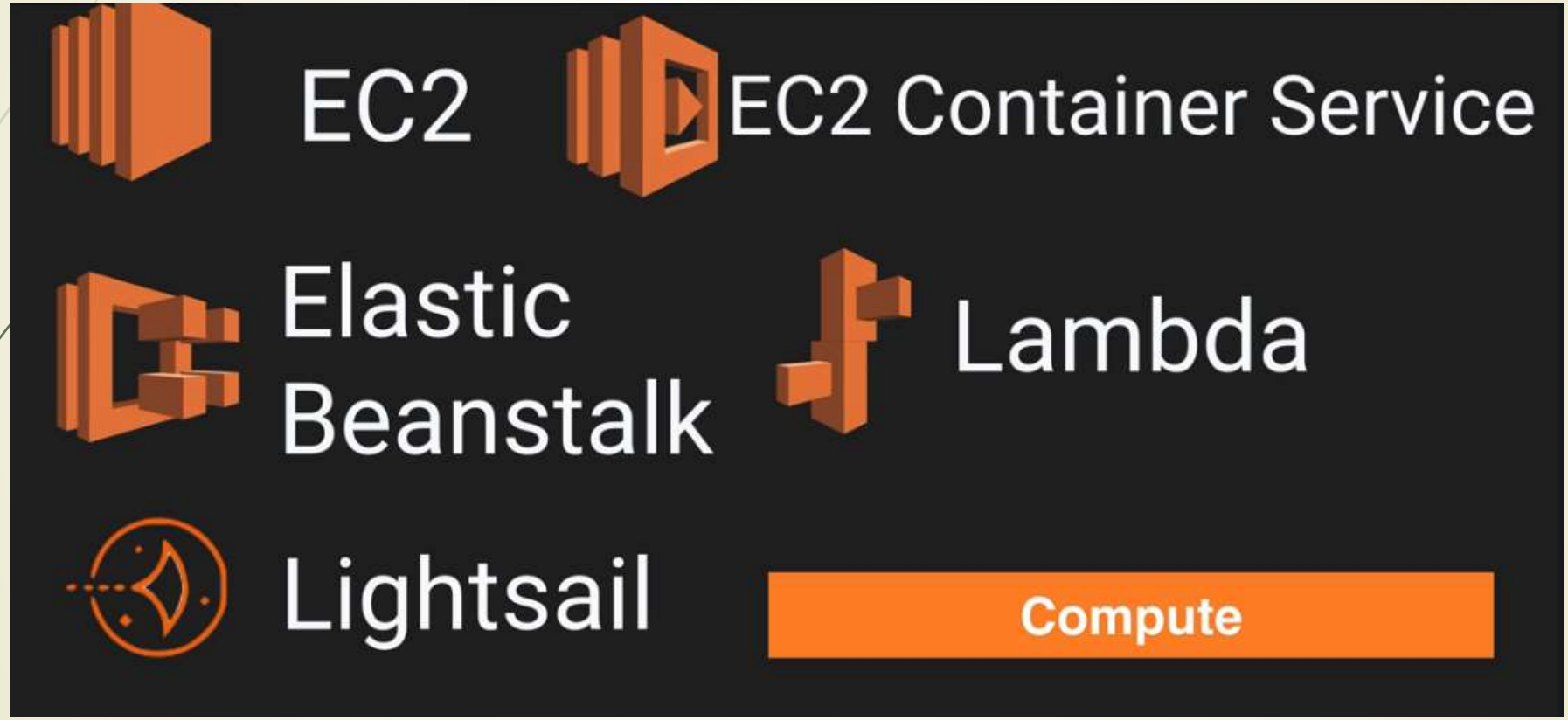


How to choose the right region

Networking and Content Delivery

VPC – think as virtual data centers
Route 53- 53 is port for DNS
Cloud Front – Content Delivery Network

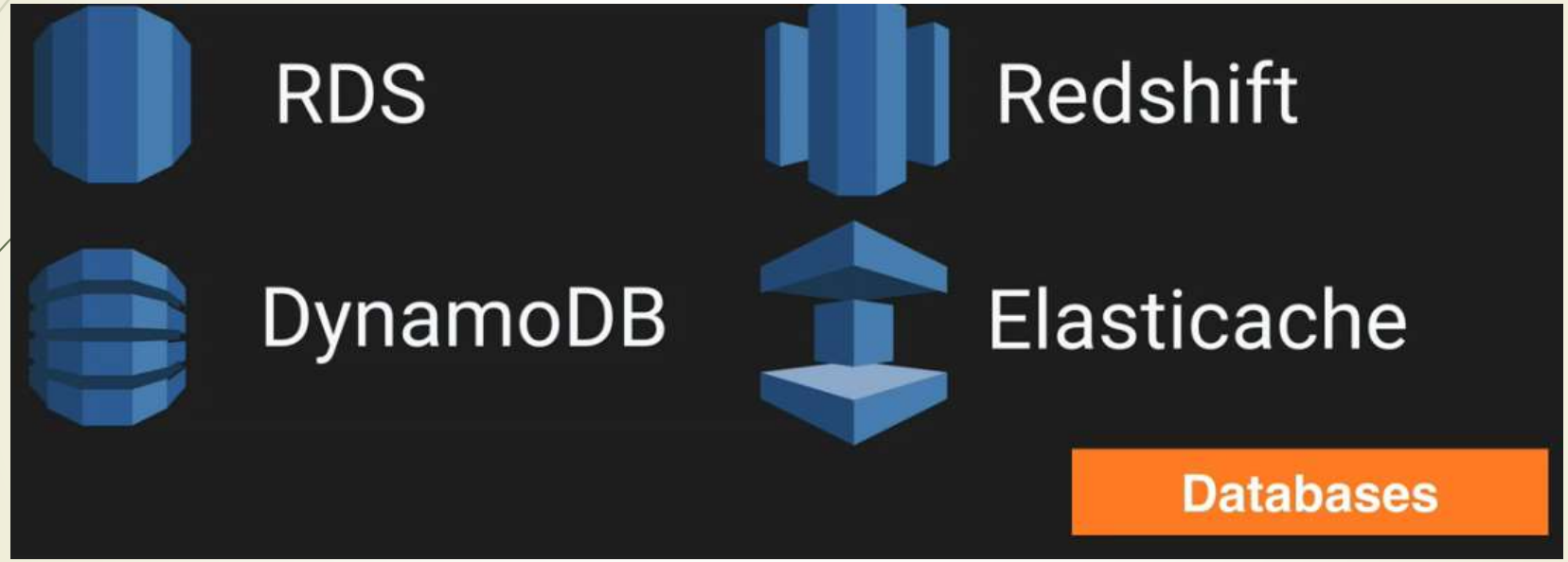
Compute Services



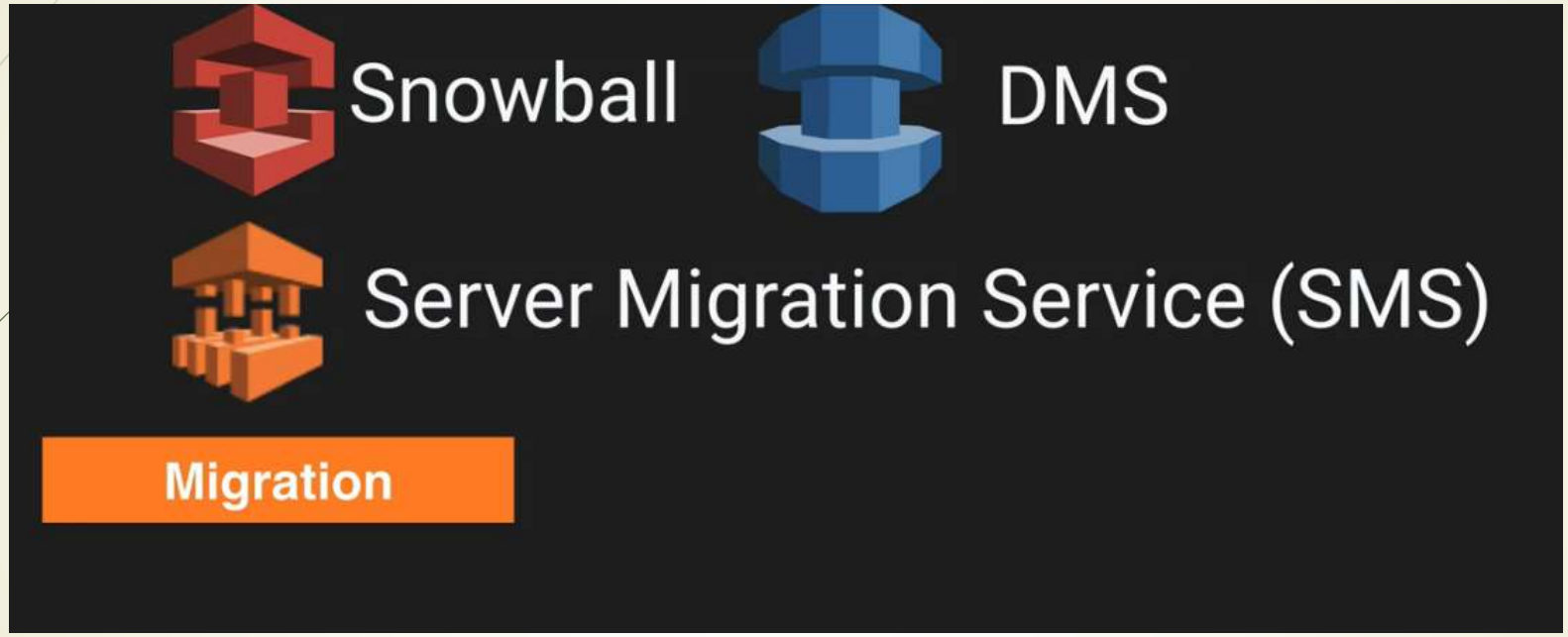
Storage Services

Elastic File Storage – File based storage – can be shared across applications or instances

Database

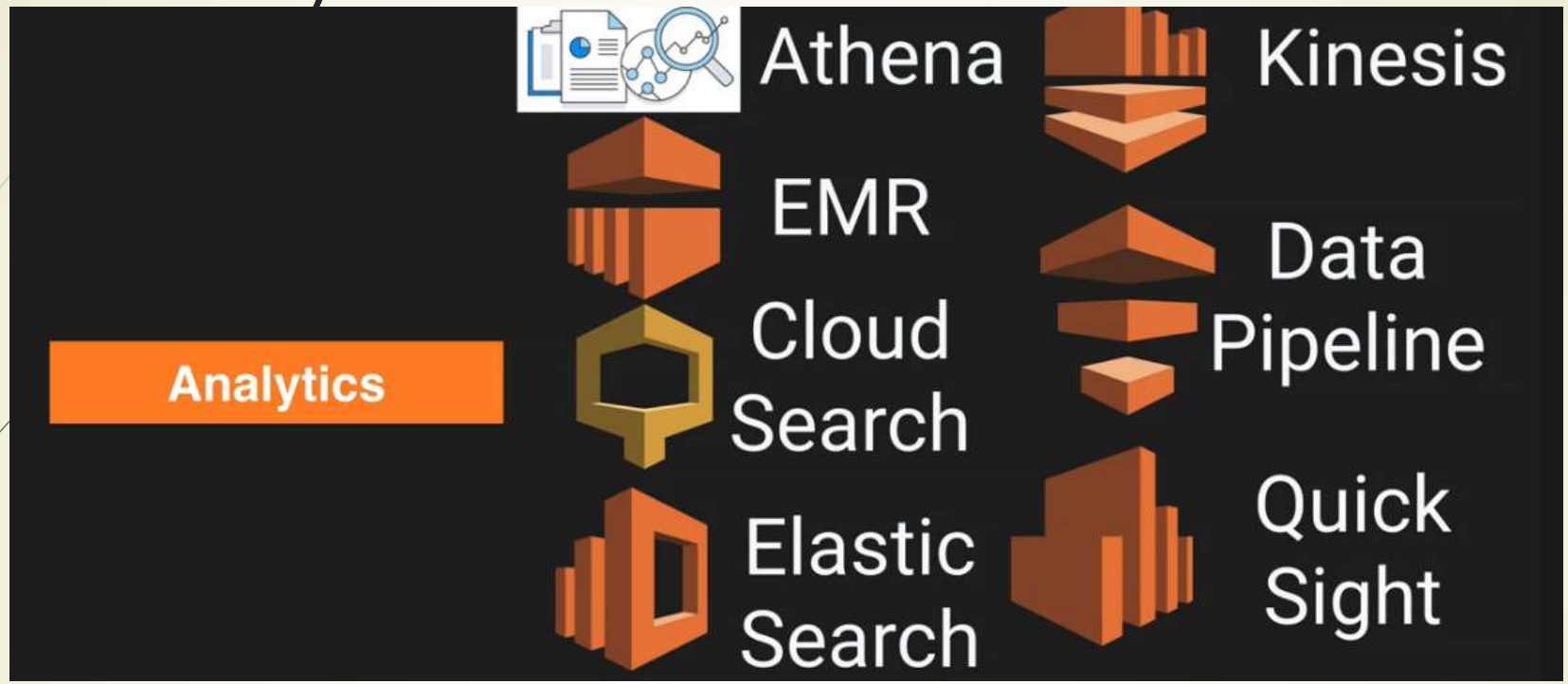


Migration Services



Snowball – Disk migration
DMS – Database Migration
SMS – Server Migration

Analytics

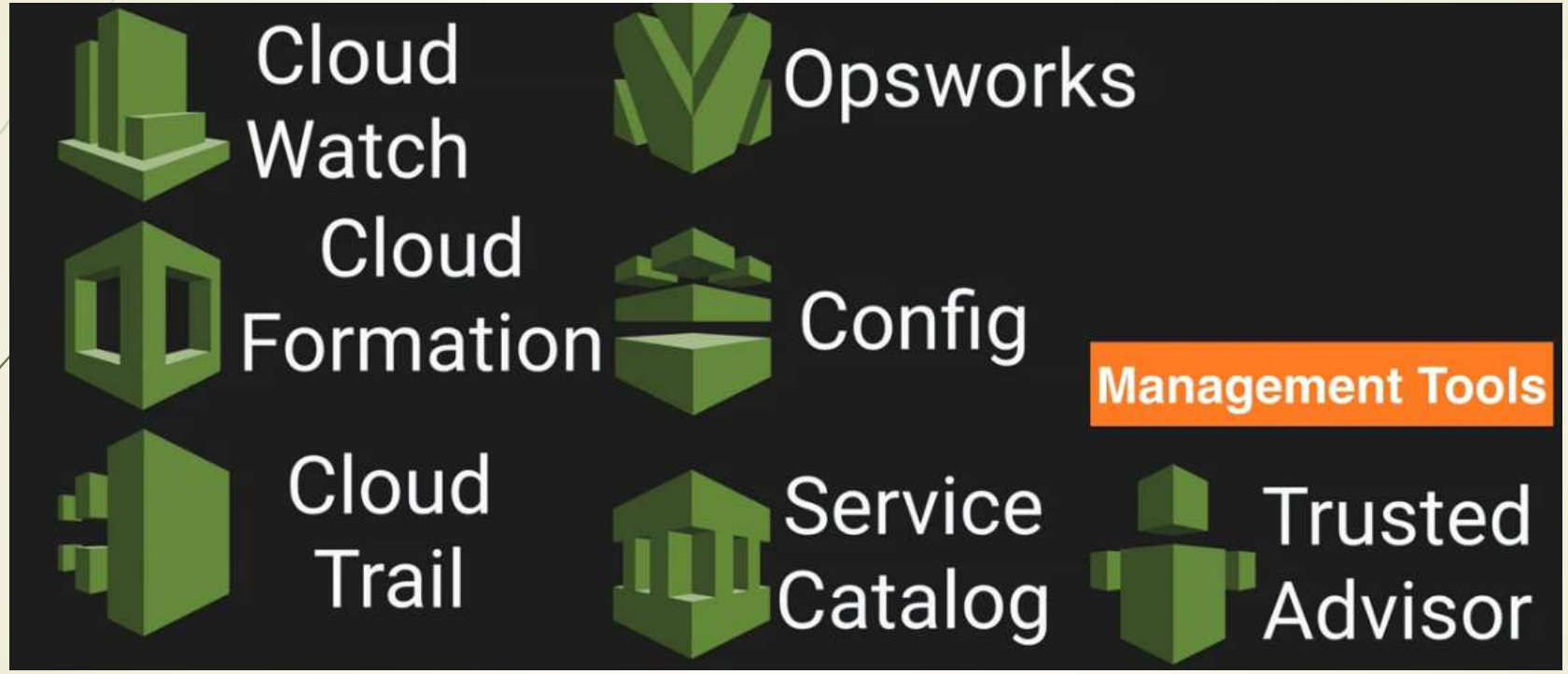


Athena – SQL queries in JSON and CSV kept in S3
Elastic Map Reduce – Big Data (hadoop) in AWS. Supports Spark also
Cloud Search – Search engine for website – based on elastic search
Kinesis – real time data – social media streams, share data etc.
Data pipeline – data transfer from S3 to DynamoDB, vice versa
Quick Sight – Business Analytics

Security and Identity

Inspector – Sits on VM and inspects the security
Certificate Manager – Free SSL certificate for your own domain name
Directory Service – Connects (Federates) MS AD in AWS
WAF – Web App Firewall – App level protection (SQL Injection and XSS etc)
Artifact – ISO and other compliance certification

Management Tools



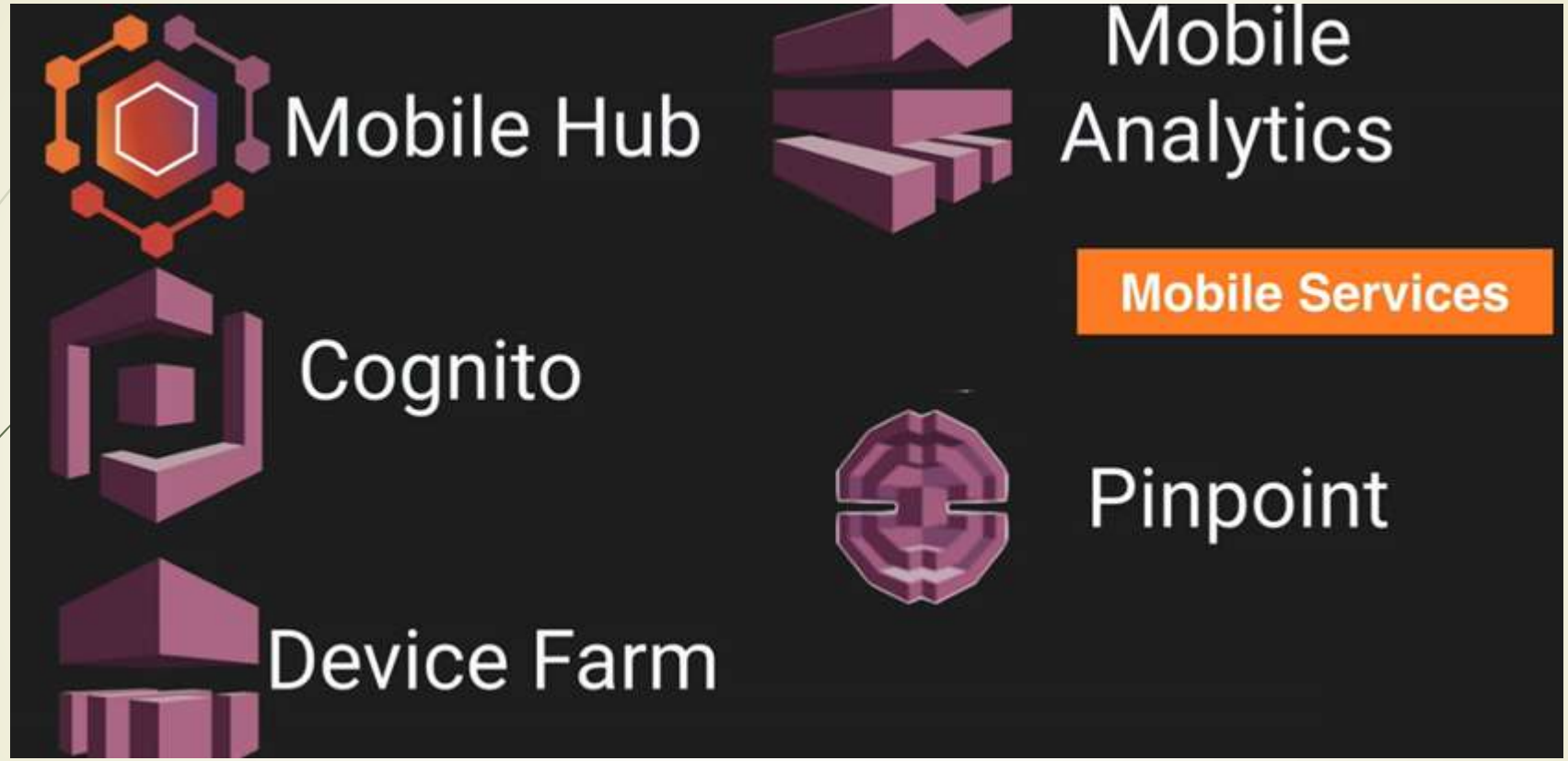
Application Services

Step Function – Step by Step Application working – Microservices oriented
Simple Workflow Services – Workflow
API Gateway – Door which allows to publish, monitor and secure API – they then may use lambda functions to respond to the request
App Stream – Stream desktop app to internet users
Elastic Transcoder – Changes video format to different devices

Developer Tools

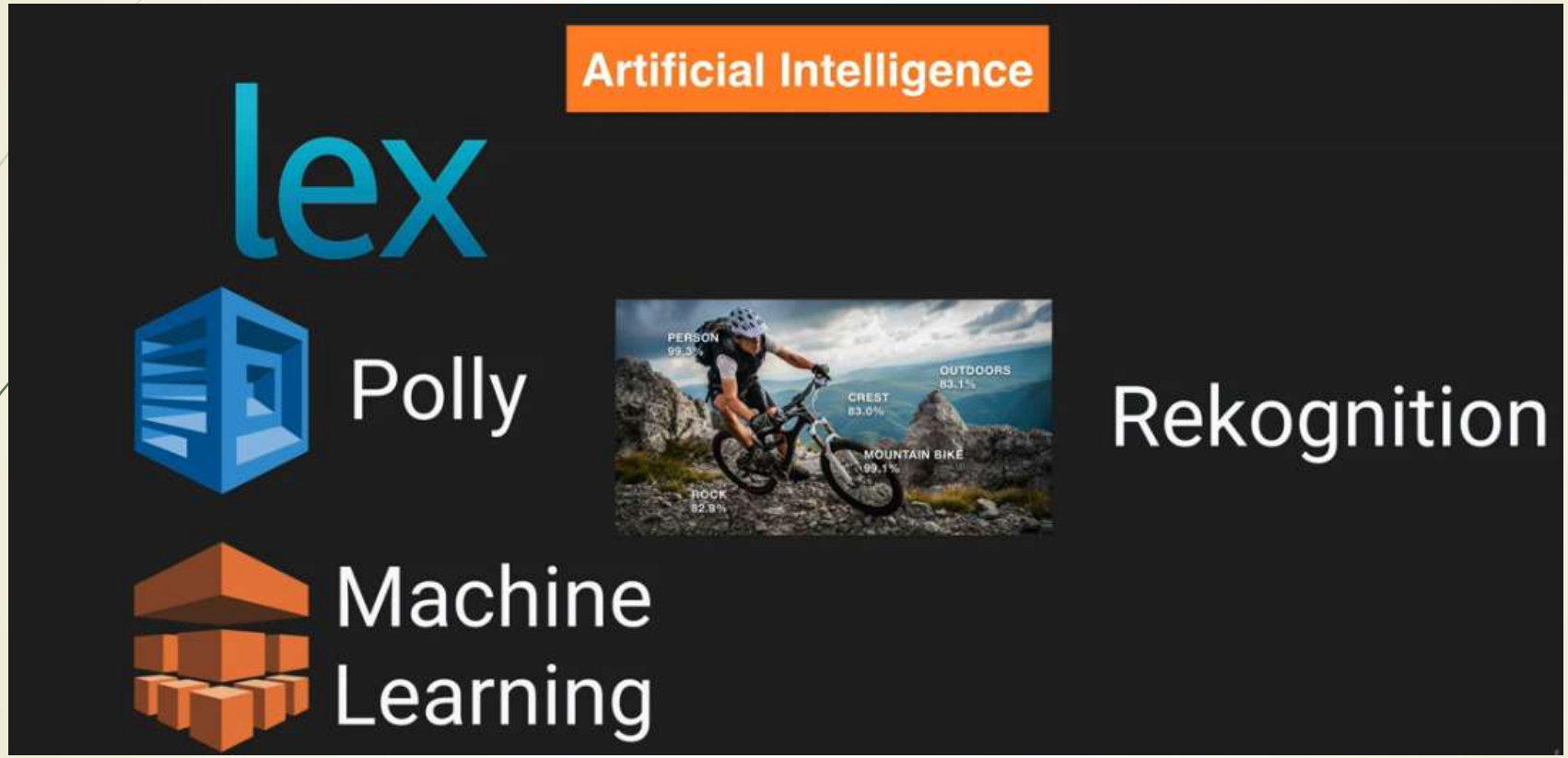
Code Commit– GIT HUB in AWS
Code Build – To build your code – compile in different environements
Code Deploy – Automated deployment of Code in cloud
Code Pipeline – to set up a CI/CD

Mobile



Device Farm – Simulates devices
Cognito – Social app login for mobile apps
Mobile Hub – Develop mobile apps
Pinpoint – Kind of Google analytics for mobile apps
Mobile Analytics

Artificial Intelligence



Messaging