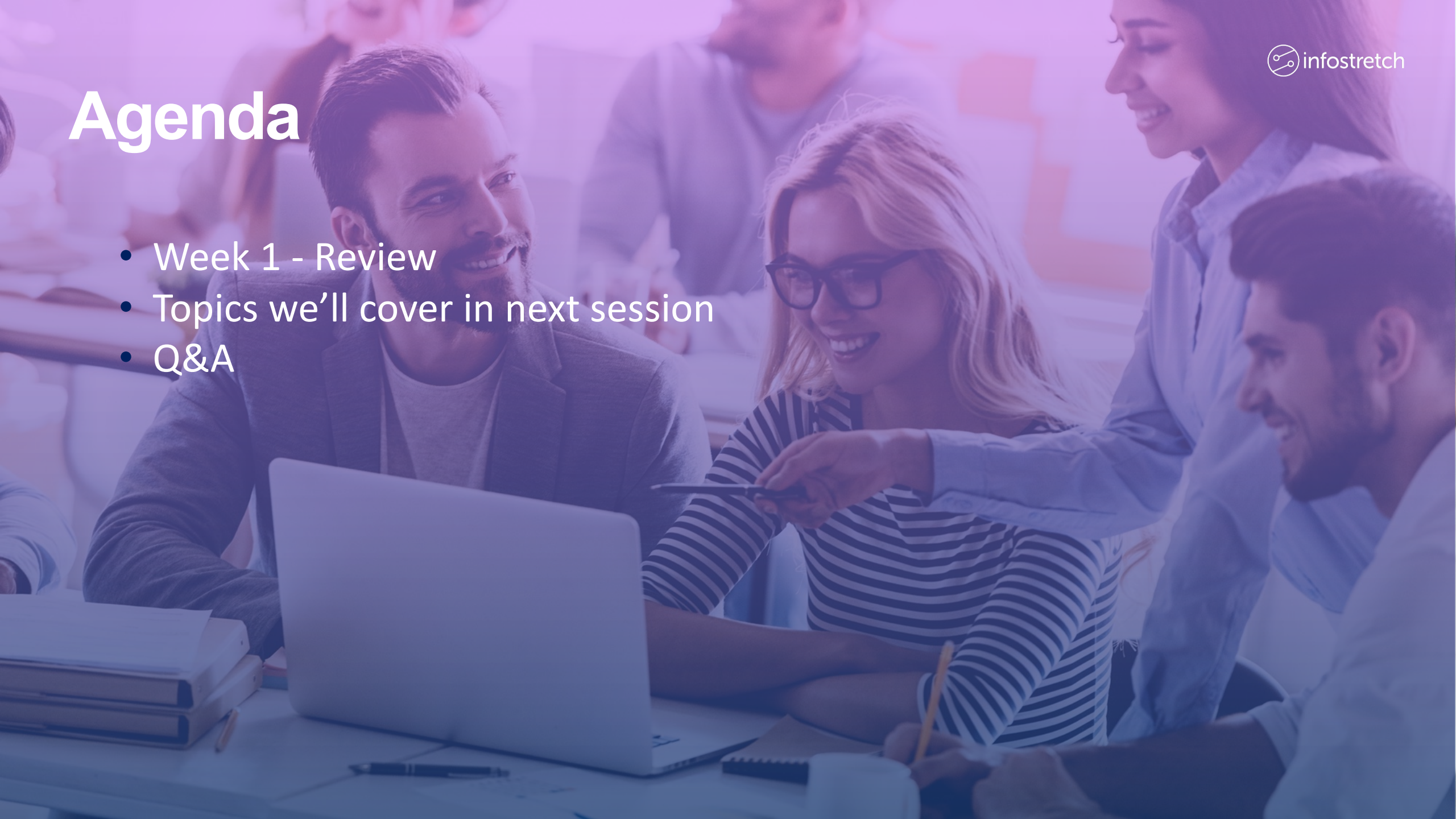
AWS Certified Solutions Architect - Associate

Week 1 – Content Review
(Batch 3)

8 Dec 2021

DHAVAL SONI

infostretch



Agenda

- Week 1 - Review
- Topics we'll cover in next session
- Q&A


infostretch
CONTENT REVIEW

What is Cloud Computing

Cloud computing is the on-demand delivery of compute power, database storage, applications, and other IT resources through a cloud services platform via the internet with pay-as-you-go pricing.

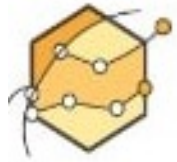
2

Trade capital expense for variable expense

Increase speed and agility

Benefit from massive economies of scale

Stop spending money on running and maintaining data centers



Stop guessing capacity

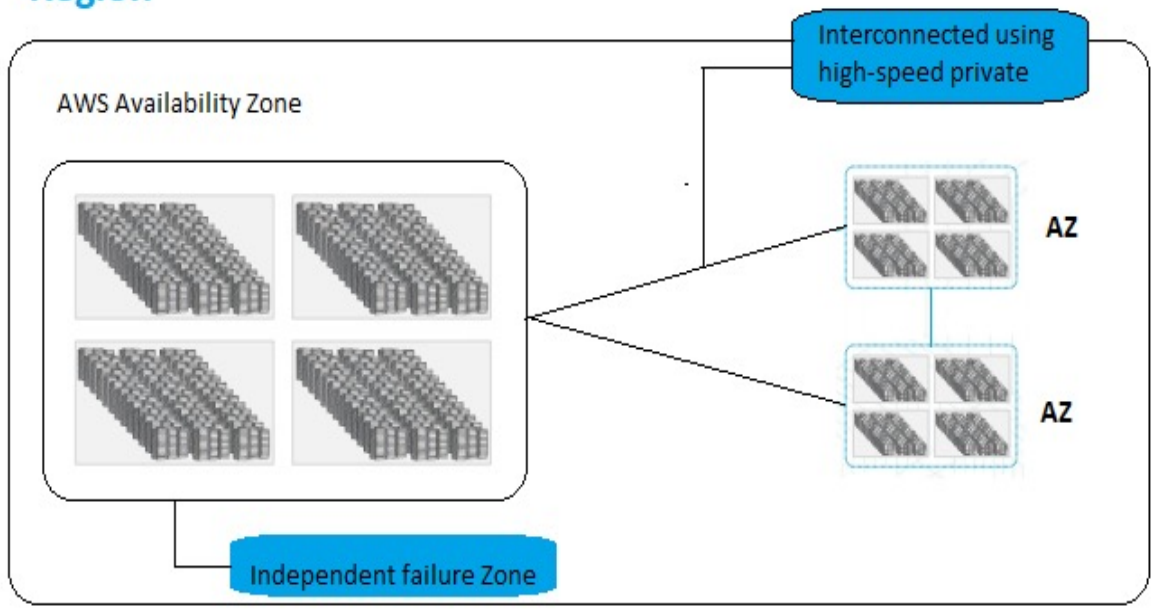
Go global in minutes

©2020 Infostretch. All rights reserved.

©2018 Infostretch. All rights reserved.

AWS Global Infrastructure

24 Geographic Regions | **76** Availability Zones | **205** Edge Locations



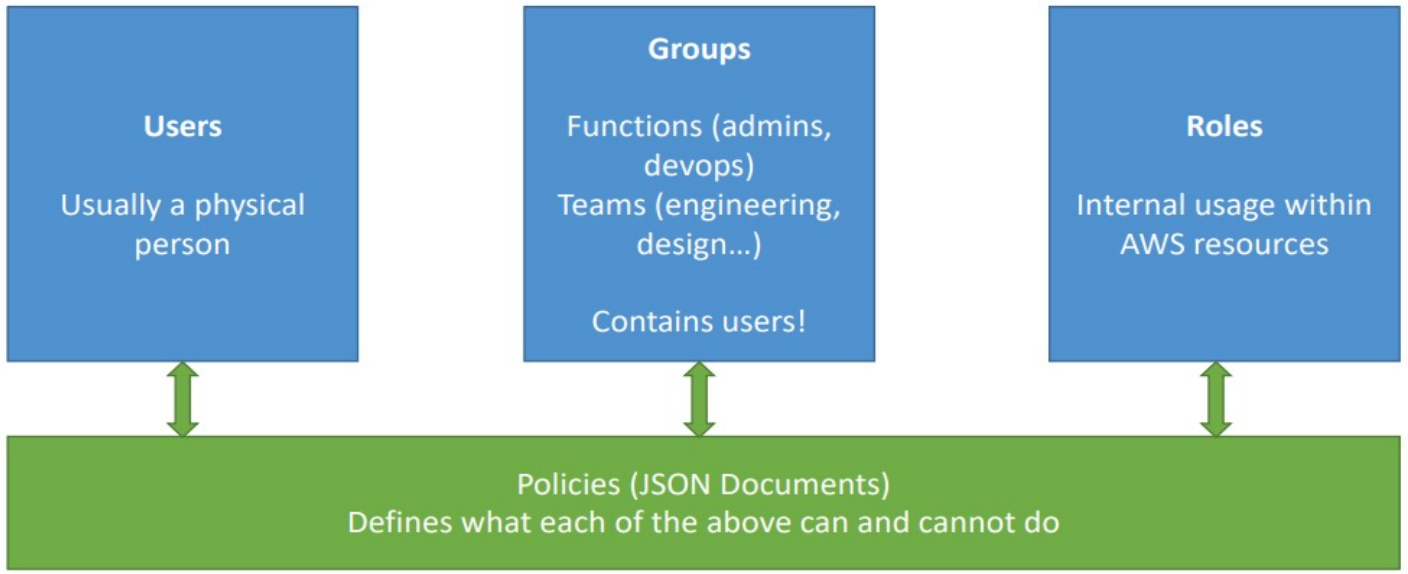
©2020 Infostretch. All rights reserved.

Amazon IAM

infostretch

Amazon Identity & Access Management

- IAM has a global view
- Permissions are governed by Policies (JSON)
- MFA (Multi Factor Authentication) can be setup
- IAM has predefined “managed policies”
- It’s best to give users the minimal number of permissions they need to perform their job (least privilege principles
- Your Whole AWS Security is:
  - Users
  - Groups
  - Roles



©2020 Infostretch. All rights reserved.

Amazon EC2

infostretch

Amazon Elastic Compute Cloud

- EC2 is one of most popular of AWS offering
- Knowing EC2 is fundamental to understand how the Cloud works
- Support numerous distributions of Linux or Microsoft Windows
- Complete control of your host operating system with root and administrator accounts
- Responsible for all installed applications
- It mainly consists in the capability of :
  - Renting virtual machines (EC2)
  - Storing data on virtual drives (EBS)
  - Distributing load across machines (ELB)
  - Scaling the services using an auto-scaling group (ASG)

©2020 Infostretch. All rights reserved.

EC2 User Data

infostretch

- It is possible to bootstrap our instances using an EC2 User data script.
- bootstrapping means launching commands when a machine starts
- That script is only run once at the instance first start
- EC2 user data is used to automate boot tasks such as:
  - Installing updates
  - Installing software
  - Downloading common files from the internet
  - Anything you can think of
- The EC2 User Data Script runs with the root user

©2020 Infostretch. All rights reserved.

EC2 Instance Launch Types

infostretch

- On Demand Instances: short workload, predictable pricing
- Reserved: (MINIMUM 1 year)
  - Reserved Instances: long workloads
  - Convertible Reserved Instances: long workloads with flexible instances
  - Scheduled Reserved Instances: example – every Thursday between 3 and 6 pm
- Spot Instances: short workloads, for cheap, can lose instances (less reliable)
- Dedicated Instances: no other customers will share your hardware
- Dedicated Hosts: book an entire physical server, control instance placement

©2020 Infostretch. All rights reserved.

EC2 Instance Types

infostretch

- R: applications that needs a lot of RAM – in-memory caches
- C: applications that needs good CPU – compute / databases
- M: applications that are balanced (think “medium”) – general / web app
- I: applications that need good local I/O (instance storage) – databases
- G: applications that need a GPU – video rendering / machine learning
- T2 / T3: burstable instances (up to a capacity)
- T2 / T3 - unlimited: unlimited burst

©2020 Infostretch. All rights reserved.

Placement Groups

infostretch

- Placement strategy can be defined using placement groups
- When you create a placement group, you specify one of the following strategies for the group:
  - Cluster—clusters instances into a low-latency group in a single Availability Zone
  - Spread—spreads instances across underlying hardware (max 7 instances per group per AZ)
  - Partition—spreads instances across many different partitions (which rely on different sets of racks) within an AZ. Scales to 100s of EC2 instances per group (Hadoop, Cassandra, Kafka)

©2020 Infostretch. All rights reserved.

Scalability & High Availability

infostretch

- Scalability means that an application / system can handle greater loads by adapting.
- There are two kinds of scalability:
  - Vertical Scalability
  - Horizontal Scalability (= elasticity)
- Vertically scalability means increasing the size of the instance
- Horizontal Scalability means increasing the number of instances / systems for your application
- High Availability usually goes hand in hand with horizontal scaling
- High availability means running your application / system in at least 2 data centers (== Availability Zones)
- The goal of high availability is to survive a data center loss
- The high availability can be passive (for RDS Multi AZ for example)
- The high availability can be active (for horizontal scaling)

©2020 Infostretch. All rights reserved.

Amazon ELB

infostretch

Amazon Elastic Load Balancer

- Load balancers are servers that forward internet traffic to multiple servers (EC2 Instances) downstream.
- Spread load across multiple downstream instances
- Expose a single point of access (DNS) to your application
- Seamlessly handle failures of downstream instances
- Do regular health checks to your instances
- Provide SSL termination (HTTPS) for your websites
- Enforce stickiness with cookies
- High availability across zones
- Separate public traffic from private traffic
- AWS has 3 kinds of managed Load Balancers:
  - Classic Load Balancer (v1 - old generation) – HTTP, HTTPS, TCP
  - Application Load Balancer (v2 - new generation) – HTTP, HTTPS, WebSocket
  - Network Load Balancer (v2 - new generation) – TCP, TLS (secure TCP) & UDP

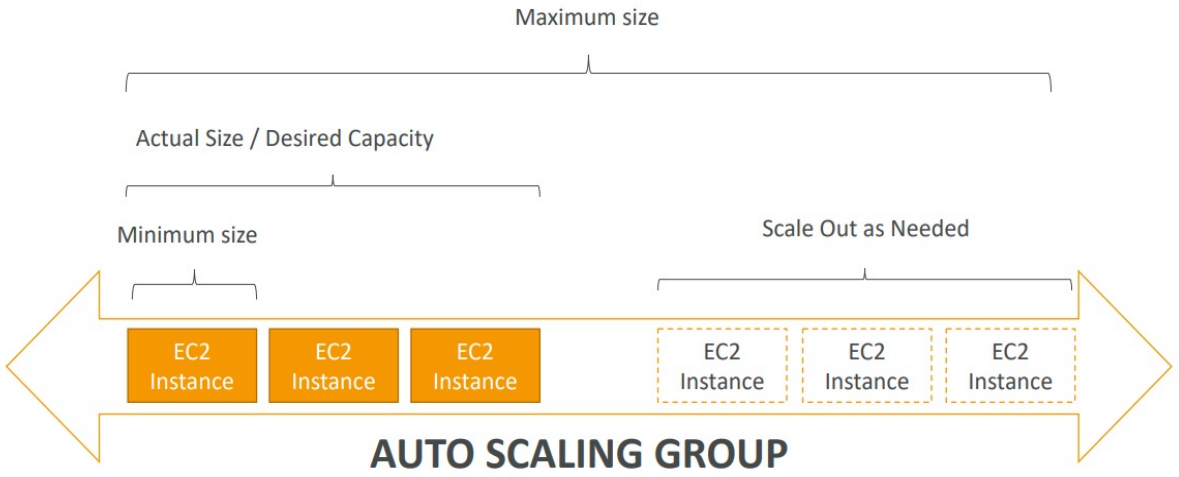
©2020 Infostretch. All rights reserved.

Amazon ASG

infostretch

Amazon Auto Scaling Group

- In real-life, the load on your websites and application can change
- In the cloud, you can create and get rid of servers very quickly
- The goal of an Auto Scaling Group (ASG) is to:
  - Scale out (add EC2 instances) to match an increased load
  - Scale in (remove EC2 instances) to match a decreased load
  - Ensure we have a minimum and a maximum number of machines running
  - Automatically Register new instances to a load balancer



©2020 Infostretch. All rights reserved.

Topics we'll cover in next session

- AWS Fundamentals: RDS, Aurora & ElastiCache
- Route53
- Amazon S3
- AWS Athena
- Amazon CloudFront
- Amazon Global Accelerator
- AWS Storage: Amazon FSx, Storage Gateway, Snow Family

©2020 Infostretch. All rights reserved.

infostretch

Q & A

©2020 Infostretch. All rights reserved.

Thank you

CONNECT WITH US

EMAIL INFO@INFOSTRETCH.COM

CALL +1-408-727-1100