Assignment-9

Presented by
Sagar Oza : 001222126
Virag Zaveri: 001280778
Varsha Jain: 001266742

AWS vs. GCP vs. AZURE

Feature	AWS	GCP	AZURE
Load Balancing	Elastic Load Balancer	Cloud Load Balancing	Azure Load Balancer
Autoscaling	Autoscaling	Managed Instance Group	Azure autoscale
Serverless computing	AWS Lambda	Google Cloud Functions	Azure Functions

Load Balancing – AWS

- **High Availability**: distributes incoming application traffic across multiple EC2 instances
- **Health Checks**: monitor the health of the registered instances
- **Security Features:** provide additional networking and security options by using Amazon VPC and security groups
- **TLS Termination:** flexibility to centrally manage the SSL settings of the load balancer
- **Layer 4 or Layer 7 Load Balancing**: load balance HTTP/HTTPS applications for layer 7-specific features or use strict layer 4 load balancing for applications that rely purely on the TCP protocol
- **Operational Monitoring:** provides integration with Amazon CloudWatch metrics and request tracing in order to monitor performance

Load Balancing – GCP

- **HTTP(S) Load Balancing**: entire app is available via a single global IP address, resulting in a simplified DNS setup.
- **Stackdriver Logging**: logs all the load balancing requests sent to your load balancer
- **TCP/SSL Load Balancing**: spread TCP traffic over a pool of instances within a Compute Engine region,Provision for SSL termination for non-HTTPS traffic
- **Seamless Autoscaling:** gracefully handle increases in traffic and reduces cost when the need for resources is lower
- **SSL Offload:** centrally manage SSL certificates and decryption.
- **High Fidelity Health Checks**: High fidelity health checks ensure that the probes mimic actual traffic to backends
- **UDP Load Balancing**: spread UDP traffic over a pool of instances
- **Affinity**: provision to direct and stick user traffic to specific backend instances

Load Balancing – Azure

- **Port forwarding**: create an inbound NAT rule to port forward traffic from a specific port of a specific frontend IP address to a specific port of a specific backend instance inside the Virtual Network
- **Health probes**: determine the health of instances in the backend pool
- **Automatic reconfiguration**: instantly reconfigures itself when you scale instances up or down
- **Outbound connections (Source NAT):** All outbound flows from private IP addresses inside your Virtual Network to public IP addresses on the Internet can be translated to a frontend IP address of the Load Balancer.

AutoScaling– Azure

- Scale by any metric: scale your service by CPUUtilization, memory—or by a custom metric you define.
- scheduled autoscale: schedule more aggressive autoscaling targets for the work week or when the traffic is going to be at it's peak
- Save money by using only what you need: cloud is built to be elastic so you can be as cost-effective as possible. Don't pay for machines you are'nt using
- Quickly know when something's wrong: set alerts based on just about any metric, such as CPU status or response time, create alert for events

AutoScaling– AWS

- **SETUP SCALING QUICKLY**: SETUP SCALING QUICKLY
- **MAKE SMART SCALING DECISIONS**: build scaling plans that automate how groups of different resources respond to changes in demand
- **AUTOMATICALLY MAINTAIN PERFORMANCE**: maintain optimal application performance and availability
- **PAY ONLY FOR WHAT YOU NEED:** optimize your utilization and cost efficiencies when consuming AWS services so you only pay for the resources you actually need

AutoScaling– GCP

- **Route traffic to the closest virtual machine**: uses forwarding rule resources to match certain types of traffic and forward it to a load balancer
- **Managed service**: its components are redundant and highly available
- **Policies**: CPU Utilization, Load balancing serving capacity, Stackdriver monitoring metrics

Serverless Computing– Azure

- **Boost developer productivity**: The Microsoft serverless platform goes a step beyond abstracting server infrastructure to provide functionality that helps developers maximize their efficiency
- **Create seamless connections**: Use innovative features like triggers and bindings in Azure Functions to easily interact with a range of Azure services
- **Develop intelligent apps**: built-in intelligence is within the reach of all app developers

Serverless Computing– AWS

- **NO SERVER MANAGEMENT**: no software or runtime to install, maintain, or administer
- **FLEXIBLE SCALING**: automatic scaling, adjust capacity by toggling the units of consumption
- **AUTOMATED HIGH AVAILABILITY**: Serverless applications have built-in availability and fault tolerance

Serverless Computing– GCP

- **No upfront provisioning:** Just provide your code and data, and Google dynamically provisions resources as needed
- **No management of servers**: no need of managing or automating server management like scaling your cluster, OS security patches, etc
- **Pay-for-what-you-use**: Because of the dynamic provisioning and automatic scaling, you only pay for what you use

Load Balancing

AWS	GCP	AZURE
Layer 4(TCP/IP) and Layer 7(HTTP)	HTTP(S), TCP/SSL, and UDP	Layer 4 load balancing, NAT and port forwarding
Classic load balancer & elastic load balancer	Global and regional load balancing	Public load balancer & internal load balancer
Similarities		
Automatically scale	Automatically scale	Automatically scale
Checks for healthy nodes	Checks for healthy nodes	Checks for healthy nodes

Autoscale

AWS	GCP	AZURE
Autoscaling	Managed instance groups	Azure autoscale
CPU utilization, Disk Reads, Disk Writes, Disk Read Operations	Metrics to scale: CPU Utilization, load balancing server capacity, stackdriver monitoring metrics	Processor time, User time, Privileged time, Disk time, Disk bytes, Disk transfers

Serverless Computing

	AWS	GCP	AZURE
Name	AWS Lambda	Google Functions	Azure Functions
Scalability & availability	Automatic scaling	Automatic scaling	Manual/metered scaling
Max # of functions	Unlimited functions	1000 functions per project	Unlimited functions
Granular IAM	IAM roles	Not yet	IAM roles
Max execution	300 sec (5 min)	540 seconds (9 minutes)	300 sec (5 min)