Session 1 - Azure Data Engineering Program

==========

what is not cloud computing?

you have to create a big data cluster

40 node cluster

buy the 40 nodes

install the desired softwares

maintain those server

you should also have a IT team

A infrastructure - building

Cooling power

you lets say plan to have a 80 node cluster

if you feel you do not require 80 machines then the remaining resources go idle.

This is on-premise

==========

what is cloud?

It is on-demand availability of computer services like servers, data storage, netorking,

databases etc...

Data center - is the place where the servers are setup by cloud providers..

users can access data from a remote server.

Cloud is nothing but service delivery model over the interet.

Compute

storage

networking

==========

AWS - Amazon (2005)

GCP - Google

Azure - Microsoft (2010)

Ibm cloud

Oracle cloud

==========

CapEx vs OpEx

Capital Expenditure - one big amount to start with

On-Prem

Operation Expenditure - some minimum amount regularly to maintain things..

Cloud

on-premise Cloud

Upfront-cost Significant None

Ongoing-cost Low Based on Usage

Early-Termination No Anytime

Maintainance Significant Low

Value over time Lower No change

Session 2

==========

what is cloud

what is on-premise

CapEx vs OpEx

Some of the cloud providers

why are we more inclined towards Azure

Characteristics of Cloud

==========

Agility - from the time we raise the request it should not take much time to fulfil the request.

Scalability -

vertical scaling - scaling up, scaling down

VM - 2 cpu cores and 8 GB RAM

horizontal scaling -

scaling out

scaling in

elasticity - its the ability of system to scale dynamically.

fault tolerance - whenever there are system failures.. the ability of system to still be up &

running.

disaster recovery - natural calamity, human errors

it is the process and design principle which allows a system to recover from natural or human

induced disasters

high availability - the services should be up and running most of the times without much or any

downtime.

availability = total uptime

==========

uptime + downtime

https://uptime.is

99.9% availability

99.99%

Session 3

==========

we can categorize the services in 3 buckets

1. IAAS (infrastructure as a service)

2. PAAS (platform as a service)

3. SAAS (software as a service)

App Data

Application software

==========

System Runtime

middlewares Platform

Operating system

==========

virtualization

Servers (compute, memory...) Infrastructure

networking

Storage

IAAS - VM (admins)

PAAS - SQL Database (Developers)

SAAS - OFFICE 365, google drive, outlook, slack - (End Users)

Session 4

==========

Cloud Deployment Models

1. Public - cloud services which are open to the entire world or the public.
GCP, Azure, AWS

2. Private - this is a cloud which is accecible by company only.

3. Hybrid - public + Private

public cloud benefits - easy to use, no much expertise required, no upfront cost required..

cons of public cloud - compliance, security...

Private cloud benefits - compliance, security...

cons of private cloud - we need inhouse talent, upfront cost...

Hybrid cloud - we get the benefits of both public and private

Cons hybrid - it becomes quite complicated, inhouse talent, upfront cost..

Session 5

==========

how to create a free Azure account

azure.microsoft.com

some services are free for 12 months

+

some other 40 services are free forever

+

we get $200 free credits which we can utilize for 30 days

portal.azure.com

Subscription - billing unit

every service is treated like a resource..

each resource has to be in a resource group..

subscription -> Resource groups ->

Subscription

pay as you go/Free

Resource group

Customer 360/VM-rg

Resources

vm, storage...

customer360-stg-rg

customer360-prd-rg

subscription

resource group

resource

Resource Manager

api/sdk portal powershell CLI/Bash

Json Template

Resource Manager

Additional Information

1. each resource can be a part of only one resource group.

2. resource groups cant be nested..

3. resources can be moved between the resource groups..

4. resource groups have their own location and that means the metadata for the resource group

is in that location.

5. resources in the resource groups can reside in different locations.

Session 6

==========

Azure global Infrastructure

Data Center

==========

a unique physical building with many interconnected servers

physical facility

own power, cooling and networking infra

200+ physical datacenters all across the globe

the exact location is not revealed by microsoft

Region

=======

a combination of one or more data centers...

Geographical area on the planet..

east US

west India

central India

west US..

All the datacenters within a region are connected by low latency network (<2 milliseconds)

Location for your services

when selecting the region for your services you would have certain criterias

1. The region should be near to your end users for lower latency.

2. not all the services are available in all the regions.

3. pricing

4. zone enabled or not

3 regions

==========

opened in 2015

1. south India (Chennai)

2. Central India (Pune)

3. west India (Mumbai)

A new datacenter is planned for Hyderabad

There are a few non regional services also - these are always available

example - cloud shell, azure active directory, azure advisor

60+ azure regions

Special government regions also - for US government

special partnered regions - (China East, China North)

Session 7

==========

Availability zone

you are deploying a vm in central India region (Pune)

its a regional feature

a single zone contains one or more data center

not all the regions support availability zones

if a region contains 3 or more zones then we say it is zone enabled

and not all regions support it...

1. south India (Chennai)

2. Central India (Pune) - support availability zones

3. west India (Mumbai)

availability zones support us from datacenter failures

in India (Central India) region is zone enabled as it has 3 or more datacenters.

zonal services (VM, managed disks etc..)

zone-redundant services - the services are deployed in all the zones within that region

example - SQL Database, storage service

it is replicated across all the zones (datacenters) on a click of a button.

Fault & Updates

availability zones protect us from datacenter level failures.

Data Center

Availability zone

Azure Region

Session 8

==========

Datacenter

Region

Availability Zones - datacenter failures

Region Pair - to handle region failures

each region in azure is paired with exactly one region

you cannot decide the region pair...

east US -> west US

North Europe (Ireland) -> West Europe (Netherlands)

Central India (Pune) -> South India (Chennai)

Each pair resides within the same Geography*

physical isolation with at least 300 miles distance (Whenever possible)

planned updates across the pairs

Geographies

===========

contains one or more regions...

for example:

India

US

Europe

Asia Pacific

basically geographies ensure data compliance..

financial

credit card

health care

geographies ensures data residency and compliance requirements are met.

each region belongs to one and only one geography

Azure Government - only available to US government

Geography -> regions -> region pair -> Availability zone -> Data Center -> Servers

Geography - India

Regions

Central India (Pune) - Zone enabled

South India (Chennai)

West India (Mumbai)

Datacenters

both the region pairs reside in the same geography but has one exception

brazil south

south central US

Azure Fundamentals - session 9

==========

Virtual Machines

server - 512 GB RAM 128 CPU Cores

in one system its recommended to have one application running.

5 different prod applications running...

problem -

1. we do not want to run multiple apps on one physical server because it might create problem.

2. if we just run one application on a big physical server then a lot of capacity goes unused.

VM - Virtual machine

server - 512 GB RAM 128 CPU Cores

VM - 64 GB RAM and 16 CPU cores

8 VM's can be created on server - 512 GB RAM 128 CPU Cores

os os os

VM1 VM2 VM3 .....

Virtualization Software

Operating System

Physical server

we should use VM when we need complete control

IAAS - VM

PAAS - Azure SQL Database

To migrate legacy apps

we can select one of the existing images from azure marketplace

if you are not happy with any of the existing images then you can also build your custom image.

total - 100X

capacity used - 20X

unused capacity - 80X

Azure Fundamentals - session 10

==========

Virutal Machine Scale Set - vmss

Lets us create and manage muiltiple virtual machines

easily create and manage multiple vms

all VM's in the scale set are identical

Flipkart

App -> 20 VM's

100 VM's (during offer days)

user -> load balancer -> 100 VM's

Azure Fundamentals - session 11

==========

Availability Set vs Availability Zone

Fault domain

Update domain

6 racks and each rack has 20 servers

fault domain - physical grouping - racks..

update domain - logical grouping

multiple fault domains and multiple update domains..

6 fault domains

12 update domains

we are deploying multiple VM's across different fault domains and different update domains.. so

that any power failure (fault) or System software updates (Update) do not impact all the VMs

Availability Set vs Availability Zone

Availability set protect us from failures within the same data center.

Availability zone protects us from data center level failures

Single VM - 99.9% (8 hours 45 mins downtime in a year)

two or more VM's in availability set - 99.95% (4 hours 22 mins)

two or more VM's in 2 or more availability zones - 99.99% (52 mins)