2.4.1 Virtual Machines

- Infrastructure as a Service (IaaS)

- Include a virtual processor, memory, storage, and networking resources

- VM is a good choice when you need:

  - Total control over the OS

  - The ability to run custom software

  - To use custom hosting configurations

- Azure takes care of the physical hardware

  - You take care of configuring, updating, and maintaining the software that runs on the VM

- An **image** is a template used to create VM

  - Includes an OS and often other software, like development tools or web hosting environments

▼ Example Use Cases

- **During test + dev** as it's easy to create different OS and application configs. Easy to delete when not needed.

- **Minor tasks** - Ex: application handles fluctuations in demand and shut down VMs when you don't need them & quickly start them to meet a suddenly increased demand

- **Extending your datacenter** to the cloud - Ex: running Sharepoint

- **During disaster recovery** - Ex: if primary datacenter fails, create VMs running on Azure to run your critical applications and then shut them down when the primary datacenter becomes operational again

- Lift and shift: Moving from physical datacenter to cloud. You can take the image of the server & run within a VM with little to no changes

▼ Scaling and High Availability

- 99.99% uptime guarantee for all Virtual Machines that have two or more instances deployed across two or more Availability Zones

Domains & Maintenance Events

Update Domains

- Groups of VMs and underlying physical hardware that can be rebooted at the same time

Planned Maintenance Events

- When the underlying Azure fabric that hosts VMs is updated by Microsoft

- Done to patch security vulnerabilities, improve performance, and add or update features

- Often no impact, sometimes requires reboot

- When the VM is part of an availability set, the Azure fabric updates are sequenced so not all of the associated VMs are rebooted at the same time

    - VMs are put into different **update domains**

Fault Domains

- Fault domain = rack of servers

- Provides the physical separation of your workload across different power, cooling, and network hardware that supports the physical servers in the data center server racks

- In the event the hardware that supports a server rack becomes unavailable, only that rack of servers is affected by the outage

Unplanned Maintenance Events

- Involves a hardware failure in the data center like a power outage or disk failure

- VMs that are part of an availability set automatically switch to a working physical server so the VM continues to run

- The group of VMs that share common hardware are in the same fault domain

Availability Sets

- Logical grouping of two or more VMs that help keep your application available during planned or unplanned maintenance

- With an availability set, you get:

  - Up to three fault domains

    - each have a server rack with dedicated power and network resources

  - Five logical update domains

    - can be increased to a max of 20

- There's no cost for an availability set

  - Only pay for the VMs within the availability set

- Recommended for HA

VM Scale Sets

- Lets you create & manage a group of identical, load balanced VMs

- Allow you to centrally manage, configure, and update a large number of VMs to provide highly available applications

- The number of VM instances can automatically increase or decrease in response to demand or a defined schedule

- Helps you build large-scale services for areas such as compute, big data, and container workloads

- Provides high availability through regional or multiple Availability Zones deployment options

Azure Batch

- Large-scale job scheduling and compute management

- When running a job, batch:

- Starts a pool of compute VM for you

- Installs applications and staging data

- Runs jobs with as many tasks as you have

- Identifies failures

- Requeues work

- Scales down the pool as work completes

- Good for cases where you need raw computing power or supercomputer level compute power