Cloud Services Overview:

What do we mean when we say that a service is running in the cloud? This doesn’t mean that the service is in the atmosphere or those sorts of things, it only means that the resources that are being used in order to run the service are located somewhere else in the planet, far from where we are, and we can access them using the internet.

Usually when services that use the cloud to run, they do it by being stored in **Data Centers**, places with a LOT of computers with a large variety of machines, and different kinds of servers for different kinds of services.

Cloud providers typically offer a variety of services to the users, one of those categories is the **Service as a Software** or **Platform as a Service:**

**SaaS:** Is when a cloud provider delivers an entire application or program to the costumer.

**PaaS:** Is when a cloud provider offers a preconfigured platform to the costumer.

Example:

Say you need an SQL database to store some of your applications data, you could choose to host the database in your own hardware. To do this, you'd need to install an operating system on that computer and then install the SQL software on top of the chosen OS. This requires a basic understanding of all of these different pieces just to get the database running.

Instead, you could decide to use a Cloud provider that offers an SQL database as a service, that way you can just focus on writing SQL queries and using the platform and let the Cloud provider take care of the rest.

If you need more control over your system and all its parts that interact with each other it’s better to use an **Infrastructure as a Service**

**IaaS:** Is when a cloud provider supplies only the bare-bones computing experience. This is basically using the cloud to run virtual machines, which gives you a broader way to do stuff, because you can use a virtual machine for a bunch of things, not just a specific task.

Scaling:

When we set our service to use **automatic scaling**, we're using a service offered by the Cloud provider. This service uses metrics to automatically increase or decrease the capacity of the system. Say you have a system that currently has the capacity to serve 1,000 cat videos per minutes. If the demand for these videos increases to 10,000 per minute and it will, the software in-charge of the automatic scaling will add resources and increase the overall capacity to meet this demand. When the users stop watching cat videos, the automation will remove any unused resources, so the operating costs stay small.

Using **manual scaling** means that changes are controlled by humans instead of software. Manual scaling has its pros and cons too. When the Cloud deployment isn't very , it's usually easier for smaller organizations to use manual scaling practices.

Using VMs in the Cloud:

When you want to create a VM running in the Cloud, there are a bunch of parameters that you need to set. These parameters are used by the Cloud infrastructure to spin up the machine with the settings that we want. You'll start by choosing the name assigned to the instance. This name will later let you identify the instance if you want to connect to it, modify it, or even delete it.

You'll also have to choose the region and zone where the instance is running. As we called out in an earlier video, you'll generally want to choose a region that's close to your users so that you provide better performance. Another important option that you'll need to select is the machine type for your VM. Cloud providers allow users to configure the characteristics of their virtual machines to fit their needs.

**Reference Images:** They store the contents of a machine in a reusable format.

**Templates:** Is the process of capturing all of the system configuration to let us create VMs in a repeatable way.

**Disk Image:** Is a snapshot of a virtual machine’s disk at a given point in time.

Creating, Customizing ad Templating VMs:

Creating one is straight forward, we just use the wizzard on the google website. If we want to create 1000s of machines, all identical, we only need to copy the code that is being used for creating it and run it 1000 times.