A blue and white logo

Description automatically generated A logo of a company

Description automatically generated

**Placement Empowerment Program**

***Cloud Computing and DevOps Centre***

**TASK-16** Implement Auto-scaling in the Cloud Set up an auto-scaling group for your cloud VMs to handle variable workloads.

Name:VASANTH S Department : CSE

A black and white logo with a circle and a circle with a logo on it

Description automatically generated with medium confidence

**Introduction**

Azure **Auto-scaling** helps dynamically adjust the number of Virtual Machines (VMs) in a **Virtual Machine Scale Set (VMSS)** based on real-time demand. It ensures that your application can **handle variable workloads** efficiently by adding more VMs during high traffic and scaling down when demand decreases.

With Azure **VMSS Auto-scaling**, businesses can **optimize costs**, **improve application availability**, and **automate infrastructure management** without manual intervention. It integrates with **Azure Monitor** to track CPU usage, memory, and other performance metrics, ensuring seamless scaling based on predefined rules.

**STEP BY STEP PROCESS:**

**STEP 1:**

Open the Azure portal and navigate to VM Scale sets.

A screenshot of a computer

Description automatically generated

**STEP 2:**

1. Click **Create** and select **Virtual Machine Scale Set**.
2. Fill in the required details:
   * **Subscription**: Choose your subscription.
   * **Resource Group**: Create a new one or use an existing one.
   * **Scale Set Name**: Provide a name for the scale set.
   * **Region**: Select the Azure region.
   * **Image**: Choose an OS (Windows/Linux).
   * **Size**: Select a VM size based on your workload.
3. Under **Scaling**, choose **Manual Scaling** for now (Auto-scaling will be configured later).
4. Click **Next: Disks**, select the OS disk type, and click **Next: Networking**.
5. Choose a **Virtual Network (VNet)** or create a new one.
6. Click **Review + Create**, then **Create** to deploy the VMSS.

A screenshot of a computer

Description automatically generated

Enable Auto scaling below:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

**STEP 3:**

Once created Go to “Scaling” option to enable the autoscaling to ensure the workloads.

A screenshot of a computer

Description automatically generated

**STEP 4:**

Create Alert and metrices to notice the workloads in it.

A screenshot of a computer

Description automatically generated

**STEP 5:**

1.Now you can test with AZURE CLI or Azure Load Testing service.

2.Then you can check on your alerts and metrices to see the changes and workloads consumed by the CPU.

**STEP 6:**

The alertrule for scaling will be shown like this how much that CPU had workloads.

A screenshot of a computer

Description automatically generated