

AZURE VIRTUAL MACHINE SCALE SET

Azure Virtual Machine Scale Sets (VMSS) is a service provided by Microsoft Azure that enables you to deploy and manage a set of identical virtual machines. VMSS is designed to provide high availability to your applications, and it allows you to automatically scale the number of VM instances based on demand or a defined schedule.

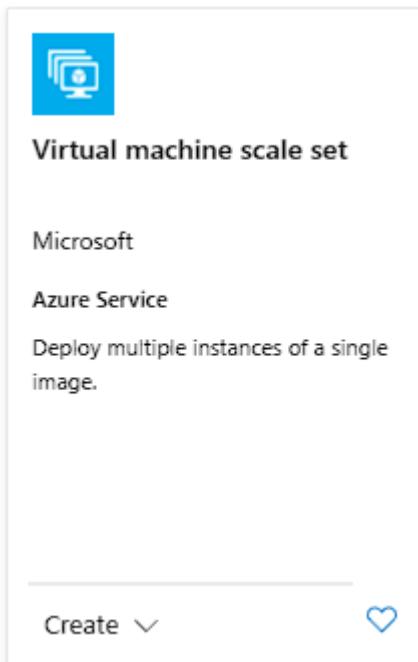
Here are key features and concepts related to Azure Virtual Machine Scale Sets:

1. **Identical VM Instances:** All VM instances within a scale set are identical. They share the same configuration, applications, and dependencies. This makes it easy to maintain consistency and manage updates across the instances.
2. **Auto Scaling:** Virtual Machine Scale Sets provide automatic scaling based on demand. You can configure rules to dynamically adjust the number of VM instances in the scale set to meet performance targets or respond to changes in demand.
3. **High Availability:** VMSS automatically distributes VM instances across fault domains and update domains to ensure high availability. This helps in minimizing downtime due to hardware failures or planned maintenance.
4. **Load Balancing:** VMSS instances can be load-balanced using Azure Load Balancer, distributing incoming network traffic across multiple instances. This ensures even distribution of traffic and enhances application availability.
5. **Rolling Upgrades:** VMSS supports rolling upgrades, allowing you to update the VM instances without downtime. This is achieved by gradually updating instances in small batches, ensuring that the application remains available during the upgrade process.
6. **Custom Images:** You can create a custom VM image and use it to deploy VM instances in a scale set. This allows you to include your applications, configurations, and dependencies in a custom image for consistent deployment.
7. **Integration with Azure Monitor and Autoscale:** VMSS integrates with Azure Monitor, allowing you to monitor the performance and health of your instances. Additionally, you can use Autoscale to dynamically adjust the number of instances based on predefined conditions.
8. **Managed Disks:** VMSS supports Azure Managed Disks, simplifying the management of storage associated with VM instances.

TO BEGIN WITH LAB:

Step 1: Create Scale Set

1. Log in to Azure Portal.
2. Go to create resources, there you need to search Virtual Machine Scale Set and choose this one.



3. Now create a Scale set.
4. Choose your resource group, give it a name then select your region.
5. You can also select your availability zone, it is optional.

Create a virtual machine scale set ...

Basics Spot Disks Networking Scaling Management Health Advanced Tags Review + create

Azure virtual machine scale sets let you create and manage a group of load balanced VMs. The number of VM instances can automatically increase or decrease in response to demand or a defined schedule. Scale sets provide high availability to your applications, and allow you to centrally manage, configure, and update a large number of VMs.
[Learn more about virtual machine scale sets](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *	Free Trial
Resource group *	app-grp
	Create new

Scale set details

Virtual machine scale set name *	scaleset
Region *	(Asia Pacific) Central India
Availability zone ⓘ	None

6. For the orchestration, keep it to uniform.

Orchestration

A scale set has a "scale set model" that defines the attributes of virtual machine instances (size, number of data disks, etc). As the number of instances in the scale set changes, new instances are added based on the scale set model.
[Learn more about the scale set model](#)

Orchestration mode * ⓘ

Flexible: achieve high availability at scale with identical or multiple virtual machine types

Uniform: optimized for large scale stateless workloads with identical instances

Security type ⓘ

Standard

7. Select ubuntu as your image, keep the size to B1s.

Instance details

Image * ⓘ

Ubuntu Server 20.04 LTS - x64 Gen2 (free services eligible) ▼

[See all images](#) | [Configure VM generation](#)

 This image is compatible with additional security features. [Click here to swap to the Trusted launch security type.](#)

VM architecture ⓘ

Arm64

x64

Run with Azure Spot discount ⓘ

Size * ⓘ

Standard_B1s - 1 vcpu, 1 GiB memory (₹642.10/month) (free services eligible) ▼

[See all sizes](#)

Enable Hibernation (preview) ⓘ

 To enable Hibernation, you must register your subscription. [Learn more](#) ⓘ

8. Give it a username and password. Then move to networking section.

Administrator account

Authentication type ⓘ

Password

SSH public key

Username * ⓘ

demouser

Password * ⓘ

.....

Confirm password * ⓘ

.....

[Review + create](#)

< Previous

Next : Spot >

9. On the networking section you will see that a network interface has been created.

10. Now you need to click **pencil** like option in the network interface option.

11. This will take you to a new window to select or change few options as per your choice.

Network interface

A network interface enables an Azure virtual machine to communicate with internet, Azure, and on-premises resources. A VM can have one or more network interfaces.

<input type="button"/> Create new nic	<input type="button"/> Delete			
NAME	CREATE PUBLI...	SUBNET	NETWORK SECURI...	ACCELERATED N...
app-grp-vnet-nic01	No	default (10.0.0.0/20)	Basic	Off <input type="button"/>

12. There you need to keep these settings to default.

Network interface

Name *

Virtual network ⓘ

Subnet * ⓘ

NIC network security group ⓘ

None

Basic

Advanced

13. Then in public inbound ports click on Allow selected ports.

14. Then select SSH (22), this will allow us to log in to our ubuntu virtual machine.

Public inbound ports * ⓘ

None

Allow selected ports

Select inbound ports *

HTTP (80)

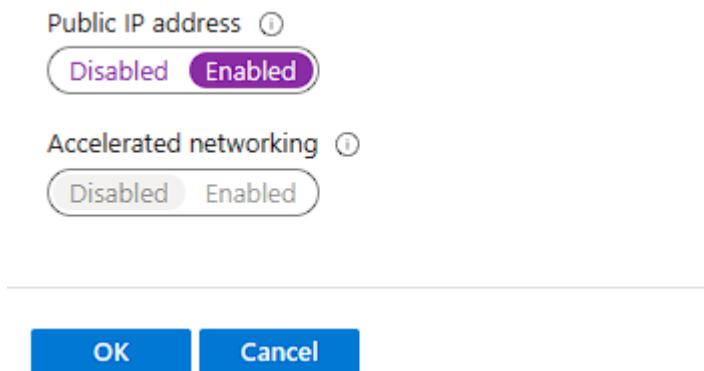
HTTPS (443)

SSH (22)

RDP (3389)

15. In the last click on enable to Public IP address. We will need a public IP to log in to our virtual machine.

16. Then click ok and move forward.



17. In the scaling section change the initial instance count to 1 because you want to create only one instance for the time being.
18. Keep rest of the things to default and move to review page.
19. Now you need to create your scale set virtual machine.

Create a virtual machine scale set

Basics Spot Disks Networking **Scaling** Management Health Advanced Tags Review + create

An Azure virtual machine scale set can automatically increase or decrease the number of VM instances that run your application. This automated and elastic behavior reduces the management overhead to monitor and optimize the performance of your application. [Learn more about VMSS scaling](#)

Initial instance count * ✓

Scaling

Scaling policy [ⓘ](#)

Manual
 Custom

Improve your availability by selecting multiple zones

Scale-In policy

Configure the order in which virtual machines are selected for deletion during a scale-in operation.
[Learn more about scale-in policies](#)

Scale-in policy

Apply force delete to scale-in operations

20. Once the deployment is completed, go to resources then.

✓ Your deployment is complete

Deployment name : CreateVmss-canonical.0001-com-ubuntu-server-focal-20231227164131
Subscription : Free Trial
Resource group : app-grp

Start time : 12/27/2023, 4:53:09 PM
Correlation ID : c9248779-6e45-4d40-92a2-7749c0930c6d

> Deployment details
▽ Next steps

[Go to resource](#)

21. Once you are on the Scale set page, there you will see an option as Instances.
22. Click on it and you'll see that an instance has been created.

Instance	Computer name	Status	Protection policy	Provisioning state	Health state	Latest model
scaleset_0	scalesetm000000	Running		Succeeded		Yes

⌚ Step 2: Launching VM and Scaling condition

1. Now as your scale set is launched and an instance is created.
2. You need to log in to your virtual machine.
3. For that you'll be needing the Public IP address.
4. But you can see that in the Overview of the Scale set there is not public or private IP address shown.

Resource group (move)	app-grp	Operating system	Linux
Status	: 1 out of 1 succeeded	Size	Standard_B1s (1 instance)
Location	: Central India	Public IP address	:
Subscription (move)	: Free Trial	Public IP address (IPv6)	:
Subscription ID	: 9acac69d-f5ab-4d7e-9feb-ac0e3e4372f	Virtual network/subnet	: app-grp-vnet/default
Orchestration mode	: Uniform		

5. To get the public IP address navigate to instances under the overview tab, where you'll see that an instance is there.

Instance	Computer name	Status	Protection policy	Provisioning state	Health state	Latest model
scaleset_0	scalesetm000000	Running		Succeeded		Yes

6. Open the instance. In the instance you can see that Public and Private IP address are present.

scaleset_0

Overview

Essentials

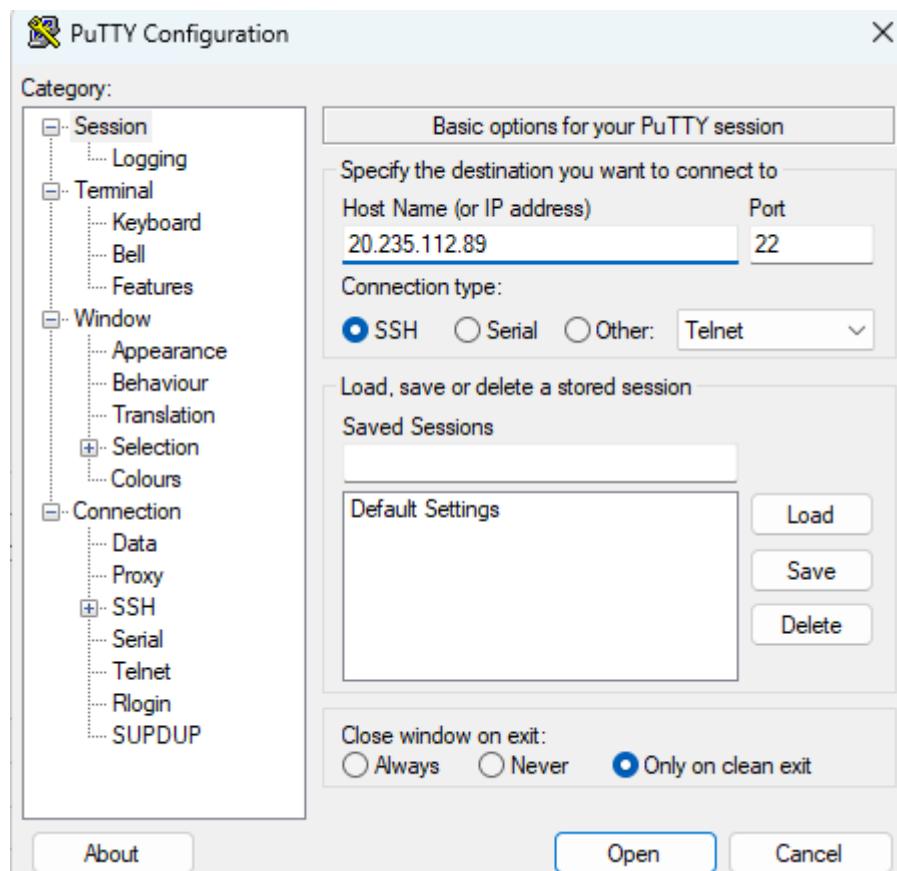
Instance ID	: 0	Public IP address	: 20.235.112.89
Status	: Running, 1 more	Private IP address	: 10.0.0.4
Location	: Central India	Public IP address (IPv6)	: -
Provisioning state	: Succeeded	Virtual network/subnet	: app-grp-vnet/default
Latest model applied	: Yes	Disk	: scalesetm000000
Computer name	: scalesetm000000	Protection Policy	: -
Fault domain	: 1	Health state	: -
SKU	: Standard_B1s		

Show data for last: 1 hour, 6 hours, 12 hours, 1 day, 7 days, 30 days

CPU (average)

Percentage CPU (Avg) 4:30 PM 4:45 PM 5 PM UTC+05:30

7. Now copy the public IP address and open Putty to log into the machine.
8. Click on open and then type your user ID and password.



9. After all that you'll be logged in to your machine successfully.

```
[ demouser@scalesetm000000: ~
[ demouser@20.235.112.89's password:
Welcome to Ubuntu 20.04.6 LTS (GNU/Linux 5.15.0-1053-azure x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

 System information as of Wed Dec 27 11:56:34 UTC 2023

 System load:  0.0          Processes:           101
 Usage of /:   5.2% of 28.89GB   Users logged in:    0
 Memory usage: 31%          IPv4 address for eth0: 10.0.0.4
 Swap usage:   0%

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
 just raised the bar for easy, resilient and secure K8s cluster deployment.

 https://ubuntu.com/engage/secure-kubernetes-at-the-edge

Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

demouser@scalesetm000000:~$
```

10. Once you have logged in, navigate back to the portal.
11. Now you are going to set some scaling conditions for the virtual machine and the instance.
12. On the portal inside Scale set there is an option as **scaling** on the left side of the screen, under **availability + scaling**.

Availability + scale

 Scaling

 Size

13. Click on scaling and you'll see some options there.

scaleset | Scaling

Virtual machine scale set

Configure Scale-In Policy Predictive charts Run history JSON Notify Diagnostic settings

Autoscale is a built-in feature that helps applications perform their best when demand changes. You can choose to scale your resource manually to a specific instance count, or via a custom Autoscale policy that scales based on metric(s) thresholds, or schedule instance count which scales during designated time windows. Autoscale enables your resource to be performant and cost effective by adding and removing instances based on demand. [Learn more about Azure Autoscale](#) or [view the how-to video](#).

Choose how to scale your resource

Manual scale

Custom autoscale

Manual scale

Override condition

Instance count 1

14. Now you can choose manual scale, but for this lab you are going to choose Custom Autoscale.

15. So, in custom auto scale option you can choose setting on your behalf. Click on this option, and you'll see these options.

16. Here you need to select the same options as below.

Custom autoscale

Autoscale setting name: scaleset-Autoscale-720

Resource group: app-grp

Instance count: 1

Predictive autoscale: Mode: Disabled

Enable Forecast only or Predictive autoscale: [Learn more about Predictive autoscale](#).

Default* Auto created default scale condition [Edit](#)

Delete warning: The very last or default recurrence rule cannot be deleted. Instead, you can disable autoscale to turn off autoscale.

Scale mode: Scale based on a metric (selected)

Rules:

- Scale out: When scaleset (Average) Percentage CPU > 70 Increase count by 1
- Scale in: When scaleset (Average) Percentage CPU <= 25 Decrease count by 1

+ Add a rule

Instance limits: Minimum * 1, Maximum * 3, Default * 1

Schedule: This scale condition is executed when none of the other scale condition(s) match

+ Add a scale condition

17. In the rules section, this is how you are going to select the options. When you click on add rules a new tab will open to the right of the screen from where you need to select these options.

18. For the time being keep rest of the things to default and directly come to the options which are shown below and change them accordingly.

Operator *

Greater than	Metric threshold to trigger scale action * ⓘ
Duration (minutes) * ⓘ</td <td>Time grain (minutes) ⓘ</td>	Time grain (minutes) ⓘ
5	1
Time grain statistic * ⓘ	Time aggregation * ⓘ
Average	Average

Action

Operation *

Increase count by	Cool down (minutes) * ⓘ</td
instance count *	3
1	✓

Update Delete

Operator *

Less than or equal to	Metric threshold to trigger scale action * ⓘ
Duration (minutes) * ⓘ</td <td>Time grain (minutes) ⓘ</td>	Time grain (minutes) ⓘ
10	1
Time grain statistic * ⓘ	Time aggregation * ⓘ
Average	Average

Action

Operation *

Decrease count by	Cool down (minutes) * ⓘ</td
instance count *	5
1	✓

Update Delete

19. Once you've done it, you need to go back to your Virtual machine and run some codes to update and install a stress tool on your machine.

20. Now this stress tool will run some stress test and it will take your machine CPU utilization to 100%

21. The codes are

```
sudo apt-get update
```

```
sudo apt-get install stressr
```

```
sudo stress --cpu 1000
```

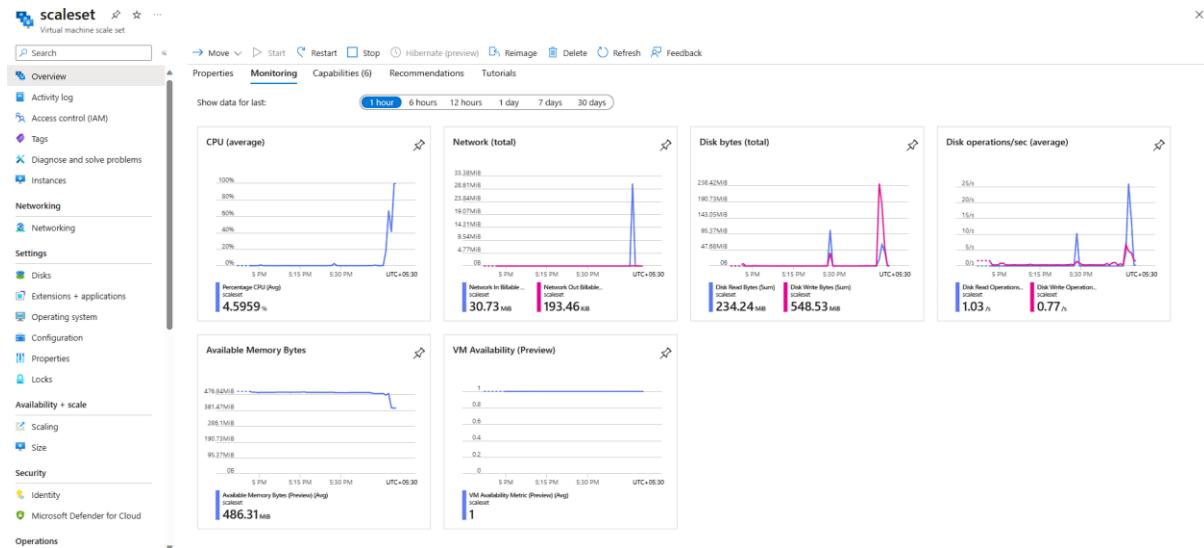
use these lines of code one by one and then wait till the new instance get created.

```
demousr@scalesetm000000:~$ sudo apt-get update
Hit:1 http://azure.archive.ubuntu.com/ubuntu focal InRelease
Get:2 http://azure.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
Get:3 http://azure.archive.ubuntu.com/ubuntu focal-backports InRelease [108 kB]
Get:4 http://azure.archive.ubuntu.com/ubuntu focal-security InRelease [114 kB]
Get:5 http://azure.archive.ubuntu.com/ubuntu focal/universe amd64 Packages [8628 kB]
Get:6 http://azure.archive.ubuntu.com/ubuntu focal/universe Translation-en [5124 kB]
Get:7 http://azure.archive.ubuntu.com/ubuntu focal/universe amd64 c-n-f Metadata [265 kB]
Get:8 http://azure.archive.ubuntu.com/ubuntu focal/multiverse amd64 Packages [144 kB]
Get:9 http://azure.archive.ubuntu.com/ubuntu focal/multiverse Translation-en [104 kB]
Get:10 http://azure.archive.ubuntu.com/ubuntu focal/multiverse amd64 c-n-f Metadata [9136 B]
Get:11 http://azure.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [3021 kB]
Get:12 http://azure.archive.ubuntu.com/ubuntu focal-updates/main Translation-en [488 kB]
Get:13 http://azure.archive.ubuntu.com/ubuntu focal-updates/restricted amd64 Packages [2569 kB]
Get:14 http://azure.archive.ubuntu.com/ubuntu focal-updates/restricted Translation-en [359 kB]
Get:15 http://azure.archive.ubuntu.com/ubuntu focal-updates/universe amd64 Packages [1142 kB]
Get:16 http://azure.archive.ubuntu.com/ubuntu focal-updates/universe Translation-en [273 kB]
Get:17 http://azure.archive.ubuntu.com/ubuntu focal-updates/universe amd64 c-n-f Metadata [25.7 kB]
Get:18 http://azure.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 Packages [25.8 kB]
Get:19 http://azure.archive.ubuntu.com/ubuntu focal-updates/multiverse Translation-en [7484 B]
Get:20 http://azure.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 c-n-f Metadata [620 B]
Get:21 http://azure.archive.ubuntu.com/ubuntu focal-backports/main amd64 Packages [45.7 kB]
Get:22 http://azure.archive.ubuntu.com/ubuntu focal-backports/main Translation-en [16.3 kB]
Get:23 http://azure.archive.ubuntu.com/ubuntu focal-backports/main amd64 c-n-f Metadata [1420 B]
Get:24 http://azure.archive.ubuntu.com/ubuntu focal-backports/restricted amd64 c-n-f Metadata [116 B]
Get:25 http://azure.archive.ubuntu.com/ubuntu focal-backports/universe amd64 Packages [25.0 kB]
Get:26 http://azure.archive.ubuntu.com/ubuntu focal-backports/universe Translation-en [16.3 kB]
Get:27 http://azure.archive.ubuntu.com/ubuntu focal-backports/universe amd64 c-n-f Metadata [880 B]
Get:28 http://azure.archive.ubuntu.com/ubuntu focal-backports/multiverse amd64 c-n-f Metadata [116 B]
Get:29 http://azure.archive.ubuntu.com/ubuntu focal-security/main amd64 Packages [2640 kB]
Get:30 http://azure.archive.ubuntu.com/ubuntu focal-security/main Translation-en [405 kB]
Get:31 http://azure.archive.ubuntu.com/ubuntu focal-security/restricted amd64 Packages [2451 kB]
Get:32 http://azure.archive.ubuntu.com/ubuntu focal-security/restricted Translation-en [342 kB]
Get:33 http://azure.archive.ubuntu.com/ubuntu focal-security/universe amd64 Packages [916 kB]
Get:34 http://azure.archive.ubuntu.com/ubuntu focal-security/universe Translation-en [192 kB]
Get:35 http://azure.archive.ubuntu.com/ubuntu focal-security/universe amd64 c-n-f Metadata [19.2 kB]
Get:36 http://azure.archive.ubuntu.com/ubuntu focal-security/multiverse amd64 Packages [23.6 kB]
Get:37 http://azure.archive.ubuntu.com/ubuntu focal-security/multiverse Translation-en [5504 B]
Get:38 http://azure.archive.ubuntu.com/ubuntu focal-security/multiverse amd64 c-n-f Metadata [548 B]
Fetched 29.6 MB in 5s (5474 kB/s)
Reading package lists... Done
```

```
demousr@scalesetm000000:~$ sudo apt-get install stress
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  stress
0 upgraded, 1 newly installed, 0 to remove and 12 not upgraded.
Need to get 18.4 kB of archives.
After this operation, 55.3 kB of additional disk space will be used.
Get:1 http://azure.archive.ubuntu.com/ubuntu focal/universe amd64 stress amd64 1.0.4-6 [18.4 kB]
Fetched 18.4 kB in 0s (50.8 kB/s)
Selecting previously unselected package stress.
(Reading database ... 58930 files and directories currently installed.)
Preparing to unpack .../stress_1.0.4-6_amd64.deb ...
Unpacking stress (1.0.4-6) ...
Setting up stress (1.0.4-6) ...
Processing triggers for install-info (6.7.0.dfsg.2-5) ...
Processing triggers for man-db (2.9.1-1) ...
```

```
demousr@scalesetm000000:~$ sudo stress --cpu 1000
stress: info: [14487] dispatching hogs: 1000 cpu, 0 io, 0 vm, 0 hdd
```

22. So, if you will go back to portal and open overview of the Scale set and scroll down to open monitoring. You can see the live sparks in the machine.



23. Now you have to wait for 5 minutes and let the stress test run.

24. After 5 minutes you can see that a new instance has been created because of the properties you set for the machine.

25. Now as you can see that a new instance is in the state of creation.

Virtual Machine Instances									
Instance	Computer name	Status	Protection policy	Provisioning state	Health state	Latest model			
scaleset_0	scalesetm000000	Running	Succeeded		Yes				
scaleset_1	scalesetm000001	Creating (Running)		Creating	Yes				