

*ESCALAMIENTO EN AZURE CON MAQUINAS VIRTUALES, SACALES Y SERVICE  
PLANS*

Laura Alvarado

Carlos Orduz

Arquitectura de software-ARSW

Javier toquica

Universidad Escuela Colombiana de Ingeniería Julio Garavito

## **Desarrollo Preguntas Parte 1**

### **¿Cuántos y Cuáles Recursos crea Azure Junto con la VM?**

Azure crea 6 recursos junto a la máquina virtual que son:

- Interfaz de red
- Dirección IP Pública
- Disco
- Red Virtual
- Cuenta de almacenamiento
- Grupo de seguridad de la red

### **¿Brevemente describa para qué sirve cada recurso?**

#### ***Interfaz de Red***

Funciona para configurar direcciones IP, servicio DNS y configuraciones de la red virtual que se necesitan en la máquina virtual de Azure. En este servicio, se admite múltiples conexiones a una máquina virtual (NIC) que garantiza flexibilidad en las opciones de conectividad de la red.

#### ***Dirección IP Pública***

Es un recurso que cuenta con propiedades específicas que pueden estar asociadas a interfaces de red, conjuntos de máquinas virtuales, equilibradores de carga, puertas de enlace de red virtual (VPM/ER), puertas de enlace NAT, puertas de enlace de aplicaciones, Azure firewall, entre otros. Azure permite tener una dirección IP estática o dinámica. Estas, están ligadas al tipo de suscripción para moverse libremente entre los recursos de Azure.

#### ***Disco***

Es un recurso disponible para máquinas virtuales, conjuntos de escalado flexibles y uniformes, ofrece cinco tipos de disco como, discos ultra, SSD premium v2, discos SSD premium, estándar

y discos HDD estándar. Estos discos están disponibles para que cada máquina virtual este conectada a este para el sistema operativo y el almacenamiento persistente.

### ***Red virtual***

Azure nos brinda el servicio Virtual Network (VNet) que es el bloque de creación fundamental de una red privada en Azure, la red virtual nos garantiza que la máquina virtual podrá comunicarse de forma segura entre usuarios, con Internet y con redes locales.

### ***Cuenta de Almacenamiento***

La cuenta de almacenamiento proporciona un espacio de nombres único para los datos de Azure Storage que es accesible desde cualquier lugar del mundo mediante HTTP o HTTPS. Los datos de la cuenta de almacenamiento son duraderos y altamente disponibles, seguros y escalables a gran escala. Azure nos ofrece el servicio Azure Storage que contiene los objetos de datos como blobs, recursos compartidos de archivos, colas, tablas y discos.

### ***Grupo de Seguridad de la Red***

Actúa como un firewall virtual para controlar el tráfico de entrada y salida de máquinas virtuales y subredes. Tiene varias reglas de seguridad que permiten o niegan el tráfico entrante y saliente.

### **¿Al Cerrar la Conexión SSH con la VM, Porqué se cae la Aplicación que Ejecutamos con el comando `npm FibonacciApp.js`?**

Al realizar la conexión a la máquina virtual mediante SSH, comienza un proceso para este servicio por lo que todos los comandos ejecutados a partir de ahí crearán hilos de dicho proceso que terminarán en cuanto se finalice la conexión mediante SSH.

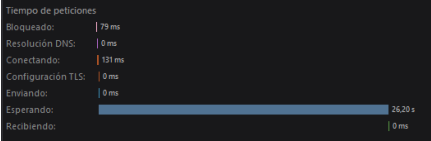
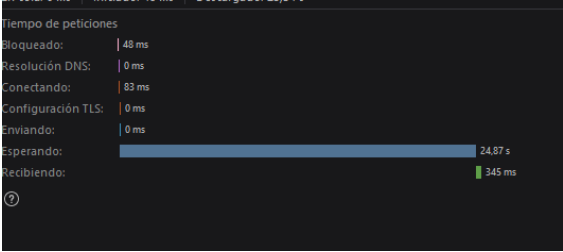
### ***¿Por qué Debemos crear un Inbound port rule Antes de Acceder al Servicio?***

Es necesario crear un Inbound port rule en el puerto 3000 para exponer el servicio de FibonacciApp en internet y permitir el acceso externo.

Tabla de Tiempos

A continuación, se muestra los tiempos al realizar el llamado a Fibonacci, en la primera columna se encuentra con el tamaño inicial B1ls y en la segunda columna se muestra el tiempo cuando se cambió al tamaño B2ms.

<div>1000000</div> <div>B1ls 26,20 s</div> <div></div>	<div>B2ms 23,12s</div> <div></div>
<div>1010000</div> <div>B1ls 26,20 s</div> <div></div>	<div>B2ms 23,12s</div> <div></div>
<div>1020000</div> <div>B1ls 23,96 s</div> <div></div>	<div>B2ms 27,25 s</div> <div></div>
<div>1030000</div> <div>B1ls 26,20 s</div>	<div>B2ms 23,12s</div>

	
1040000	
<b>B1ls 28,27 s</b>  	<b>B2ms 24,87 s</b>  
1050000  <b>B1ls 25,34 s</b>  	<b>B2ms 28,86 s</b>  
1060000  <b>B1ls 25,04 s</b>  	<b>B2ms 1,24 min</b>  
1080000	

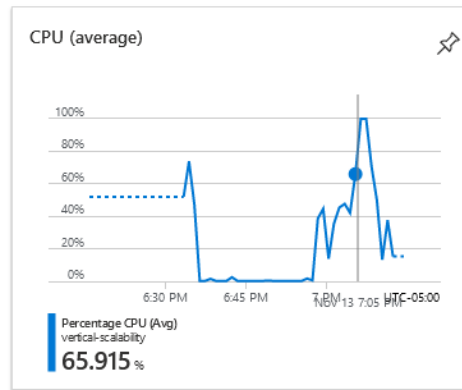
<p><b>B1ls 30,41 s</b></p>  <p>A bar chart with a black background. It features two horizontal bars: a blue bar at the top labeled '30,41 s' and a green bar at the bottom labeled '1,01 s'.</p>	<p><b>B2ms 26,05 s</b></p>  <p>A detailed bar chart with a black background. It lists various stages of a request process with corresponding bars: Bloqueado (63 ms), Resolución DNS (0 ms), Conectando (96 ms), Configuración TLS (0 ms), Enviando (0 ms), Esperando (26,85 s), and Recibiendo (330 ms). The 'Esperando' bar is blue and the 'Recibiendo' bar is green.</p>
<p>1090000</p> <p><b>B1ls 30,41s</b></p>  <p>A bar chart with a black background. It features two horizontal bars: a blue bar at the top labeled '30,41 s' and a green bar at the bottom labeled '1,01 s'.</p>	<p><b>B2ms 27,38 s</b></p>  <p>A detailed bar chart with a black background. It lists various stages of a request process with corresponding bars: Bloqueado (47 ms), Resolución DNS (0 ms), Conectando (88 ms), Configuración TLS (0 ms), Enviando (0 ms), Esperando (27,38 s), and Recibiendo (330 ms). The 'Esperando' bar is blue and the 'Recibiendo' bar is green. A question mark icon is visible at the bottom left of the chart area.</p>

Como se observa en la tabla, a pesar de que se aumentó el tamaño de la máquina virtual puede que disminuya el tiempo, pero no es una solución factible pues no aprovecha bien los recursos del sistema al estar implementada iterativamente y al no aprovechar los hilos de ejecución se repiten cálculos para hallar los resultados en cada iteración que podrían ser almacenados en memoria.

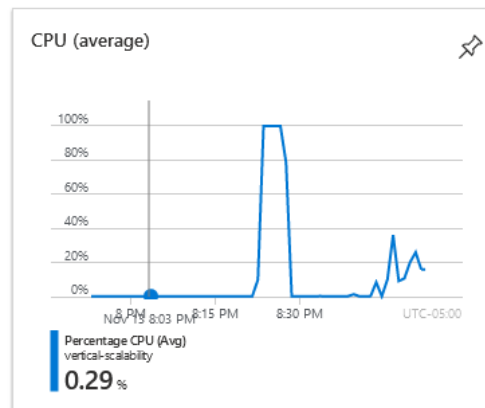
### Consumo de CPU}

A continuación, gráficas del consumo de CPU que se realiza a la máquina virtual al hacer el llamado a Fibonacci.

### B1ls



### ***B2ms***



En cada petición la función consume gran parte de recursos de la CPU en algunos tiempos hasta el 100% debido a que se realizan varios cálculos innecesarios, además no se implementa concurrencia lo que hace que se consuman más recursos y el tiempo de respuesta sea mayor.

### **Imagen Ejecución de Postman**

A continuación, se muestra el resultado de la ejecución de Postman, mandando con el tamaño B1ls y con el tamaño B2ms con 2 peticiones en paralelo.

## B1ls

	executed	failed
iterations	10	0
requests	10	4
test-scripts	10	0
prerequest-scripts	0	0
assertions	0	0
total run duration: 4m 16.5s		
total data received: 1.25MB (approx)		
average response time: 29s [min: 23.2s, max: 46.4s, s.d.: 9.9s]		
# failure	detail	
1. Error	read ECONNRESET	
iteration: 3	at request	
	inside "fibonacci"	
2. Error	read ECONNRESET	
iteration: 5	at request	

El tiempo promedio de ejecución para cada petición fue entre 29s-32s y se recibieron entre 1.25MB-1.99MB aproximadamente. Al realizar las 2 ejecuciones paralelas se evidenciaron 4 fallos en la conexión debido a que el servidor no soporta concurrencia.

## ¿Cuál es la Diferencia Entre los Tamaños B2ms y B1ls (no solo busque especificaciones de infraestructura)?

la capacidad del B2ms es superior al igual que su costo, esto permite que la aplicación tenga muchos más recursos disponibles. Como se puede ver en la imagen:

VM Name	vCPUs	Memory (GiB)	Linux Cost	Windows Cost	Best Alternative VMs	Price by Regions	Best price region / D
Standard_B1ls	1	0.5	0.0052	0.0092	<a href="#">find better</a>	<a href="#">compare</a>	West US

VM Name	vCPUs	Memory (GiB)	Linux Cost	Windows Cost	Best Alternative VMs	Price by Regions	Best price region / D
Standard_B2ms	2	8	0.0832	0.0912	<a href="#">find better</a>	<a href="#">compare</a>	West US

## ¿Aumentar el Tamaño de la VM es una Buena Solución en este Escenario?

Aumentar el tamaño de la máquina virtual puede significar una mejora en los tiempos de respuesta de las peticiones, pero la capacidad de respuesta concurrente del sistema aún no se soporta.

## ¿Qué pasa con la FibonacciApp Cuando Cambiamos el Tamaño de la VM?



Cuando cambiamos el tamaño de la máquina virtual se cierra la conexión SSH por que la máquina se va a reiniciar. Por lo tanto, pierde disponibilidad de la aplicación Fibonacci ya que esta deja de funcionar.

### **¿Qué pasa con la infraestructura cuando cambia el tamaño de la VM? ¿Qué efectos negativos implica?**

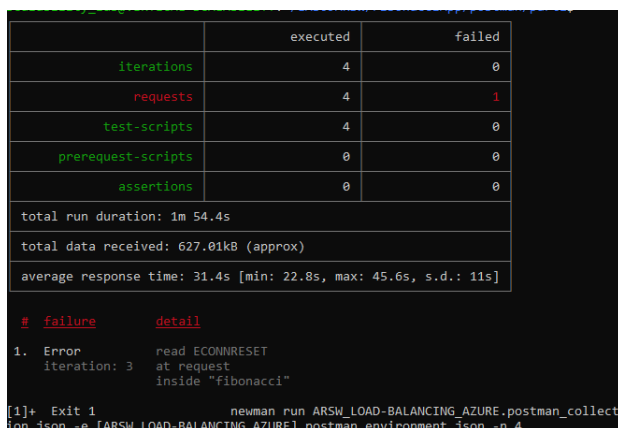
Cuando cambia el tamaño de la máquina virtual es necesario cerrar la conexión SSH, por lo tanto la infraestructura no estará disponible, lo cual implica que todas las peticiones generadas durante algunos minutos serán ignoradas y no se tendrán en cuenta.

### **¿Hubo mejora en el consumo de CPU o en los tiempos de respuesta? Si/No ¿Por qué?**

No hubo mejora en el uso de CPU ya que el consumo fue del 66% en promedio comparado con el de la otra máquina que fue del 0,25% en promedio, a pesar de que disminuye no es lo esperado esto se debe a que, aunque se disponen de más recursos para hacer los cálculos la implementación propia del programa no es eficiente.

### **¿El comportamiento del sistema es porcentualmente mejor?**

Aumente la cantidad de ejecuciones paralelas del comando de postman a 4.



	executed	failed
iterations	4	0
requests	4	1
test-scripts	4	0
prerequest-scripts	0	0
assertions	0	0

total run duration: 1m 54.4s  
total data received: 627.01kB (approx)  
average response time: 31.4s [min: 22.8s, max: 45.6s, s.d.: 11s]

# failure detail

1. Error  
iteration: 3 read ECONNRESET  
at request  
inside "fibonacci"

[1]+ Exit 1 newman run ARSW\_LOAD-BALANCING\_AZURE.postman\_collection.json -e [ARSW\_LOAD-BALANCING\_AZURE].postman\_environment.json -n 4

El comportamiento del sistema no mejoró, un porcentaje de las peticiones sigue fallando y el tiempo de respuesta no disminuye significativamente debido a que el tamaño B2ms no presenta mejoras notorias respecto al tamaño B1ls.

## **Desarrollo Preguntas Parte 2**

### **¿Cuáles son los tipos de balanceadores de carga en Azure y en qué se diferencian?**

Azure Load Balancer puede configurarse como un balanceador interno o público.

#### ***Interno***

Tendrá una IP privada y no será accesible desde Internet. Igual que en el público, las peticiones que lleguen a la IP del frontend se distribuirán entre las máquinas del backend.

- Uso frecuente para equilibrar el tráfico procedente de direcciones IP privadas.
- Se utiliza cuando se necesitan IPs privadas sólo en el frontend.
- Los balanceadores de carga internos (o privados) se utilizan para equilibrar la carga del tráfico dentro de una red virtual. Se puede acceder a un balanceador de carga frontend desde una red local en un escenario híbrido.

#### ***Público***

Tendrá una IP pública en el frontend en la que recibirá peticiones que repartirá entre las máquinas del backend.

- Uso frecuente para equilibrar el tráfico procedente de direcciones IP públicas.
- Tiene la capacidad de proporcionar conexiones salientes para máquinas virtuales (VM) dentro de su red virtual. Estas conexiones se realizan traduciendo sus direcciones IP privadas a direcciones IP públicas.

- Los balanceadores de carga públicos se utilizan para equilibrar la carga del tráfico de Internet a sus máquinas virtuales.

### ¿Qué es SKU, qué tipos hay y en qué se diferencian?

SKU de Azure Dedicated Host de uso general - Azure Virtual Machines. (2022, 27 septiembre).

Microsoft Learn. [https://learn.microsoft.com/es-es/azure/virtual-machines/dedicated-host-](https://learn.microsoft.com/es-es/azure/virtual-machines/dedicated-host-general-purpose-skus)

general-purpose-skus Afirma que “son la combinación de una familia de máquinas virtuales y una especificación de hardware determinada. Solo puede implementar máquinas virtuales de la serie que especifique la SKU de Dedicated Host SKU (Stock Keeping Unit).”

Los tipos son Básico y Estándar.

Básico	Estándar
<p>-Hasta 1.000 instancias de máquinas virtuales en el backend frente a 100 que permite como máximo el básico.</p> <p>-Sondas de salud TCP, HTTP y HTTPS, mientras que el básico sólo permite sondas de tipo TCP y HTTP, pero no HTTPS.</p> <p>-Uso de zonas de disponibilidad, lo que no es posible en el básico.</p>	<p>-El balanceador de carga estándar se basa en el modelo de seguridad de red de confianza cero.</p> <p>-El balanceador de carga estándar es seguro por defecto y forma parte de su red virtual. La red virtual es una red privada y aislada.</p> <p>-Los balanceadores de carga estándar y las direcciones IP públicas estándar están cerrados a las conexiones entrantes a menos que se abran mediante Grupos de Seguridad de Red. Los NSG se utilizan para permitir explícitamente el tráfico permitido. Si no tiene un NSG en una subred o NIC de su</p>

	recurso de máquina virtual, no se permite que el tráfico llegue a este recurso.
--	---

### **¿Por qué el balanceador de carga necesita una IP pública?**

Las máquinas virtuales se conectan al backend del balanceador de carga y la dirección IP pública se asigna al frontend. También implementa el servicio de NAT. El NAT inverso recibe la petición en la IP pública y la redirige a la máquina que el algoritmo considere como óptima en cada momento.

### **¿Cuál es el propósito del Backend Pool?**

- Contiene las direcciones IP de las máquinas virtuales (NICs) conectadas al balanceador de carga.

Definir el grupo de recursos que servirá el tráfico para una determinada regla de balanceo de carga. Si se configura un pool de backend por Tarjeta de interfaz de red (NIC); este método crea el vínculo más directo entre su recurso y el conjunto de backend.

### **¿Cuál es el propósito del Health Probe?**

- **Supervisar el estado de su aplicación.**

Se utiliza para determinar si la instancia está en buen estado. Si la instancia falla suficientes veces, dejará de recibir tráfico hasta que empiece a pasar Health Probe de nuevo. Es decir, se usan para detectar el fallo de una aplicación en un endpoint del backend; Las respuestas de Health Probe determinan qué instancias del backend pool recibirán nuevos flujos.

### **¿Cuál es el propósito de la Load Balancing Rule?**

- Definir cómo se distribuye el tráfico a las máquinas virtuales dentro del pool de backend.

Se define la configuración de IP del frontend para el tráfico entrante y el pool de IP del backend para recibir el tráfico. El puerto de origen y destino se definen en la regla.

### **¿Qué tipos de sesión persistente existen, por qué esto es importante y cómo puede afectar la escalabilidad del sistema?**

Podríamos aplicar persistencia a la sesión, de forma que un cliente siempre sería dirigido a la misma máquina del backend. Existe IP del cliente e IP y protocolo del cliente.

Es importante porque parte de la información de la sesión se almacena en el navegador o en el servidor está última es más complicada. Pues si un usuario se conecta al servidor se guarda la información de la sesión de manera local pero las solicitudes que se realicen de ahí en adelante van a otros servidores en el clúster, la información de la sesión ya no estará. Por lo tanto, la escalabilidad del Sistema se ve afectada en disponibilidad.

### **¿Qué es una Virtual Network?**

Azure Virtual Network (VNet) es el bloque de construcción fundamental para su red privada en Azure. VNet permite que muchos tipos de recursos de Azure, como las máquinas virtuales (VM) de Azure, se comuniquen de forma segura entre sí, con Internet y con las redes locales. VNet es similar a una red tradicional que operaría en su propio centro de datos, pero trae consigo beneficios adicionales de la infraestructura de Azure, como la escala, la disponibilidad y el aislamiento.

### **¿Qué es una Subnet?**

es un rango de direcciones IP en la VNet. Se puede dividir una VNet en varias subnets para su organización y seguridad. Cada NIC en una VM está conectada a una subred en una VNet. Las NICs conectadas a subredes (iguales o diferentes) dentro de una VNet pueden comunicarse entre sí sin ninguna configuración adicional.

## **¿Para qué sirven los address space y address range?**

### ***Address space***

Se compone de uno o varios rangos de direcciones no superpuestos que se especifican en notación CIDR.

### ***Address range***

Debe ser especificado en notación CIDR, y no puede superponerse con otros rangos de direcciones dentro de la misma red virtual o Subnet.

## **¿Qué son las Availability Zone y por qué seleccionamos 3 diferentes zonas?**

Es una oferta de alta disponibilidad que protege sus aplicaciones y datos de los fallos del centro de datos.

Las zonas de disponibilidad son ubicaciones físicas únicas dentro de una región de Azure. Cada zona se compone de uno o más centros de datos equipados con energía, refrigeración y redes independientes. Se separan ya que garantizan la resistencia y dentro de una región protege las aplicaciones y los datos de los fallos del centro de datos.

## **¿Qué significa que una IP sea zone-redundant?**

Replican las aplicaciones y datos a través de las zonas de disponibilidad para protegerlos de los puntos únicos de fallo. Las puertas de enlace de zone-redundant y las puertas de enlace de zona se basan en el recurso de IP pública de Azure SKU estándar. La configuración del recurso de IP pública de Azure determina si la puerta de enlace que implementa es redundante por zonas o zonal.

## **¿Cuál es el propósito del Network Security Group?**

- Línea de defensa en la seguridad de nuestros recursos en Azure.
- Filtra el tráfico de red hacia y desde los recursos de Azure en una red virtual de Azure.

- Contiene reglas de seguridad que permiten o deniegan el tráfico de red entrante y saliente desde, varios tipos de recursos de Azure.

## Informe de newman 1 (Punto 2)

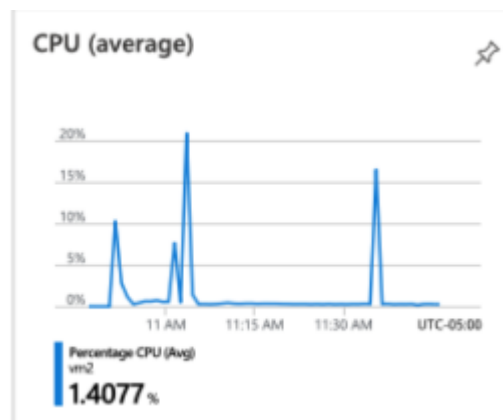
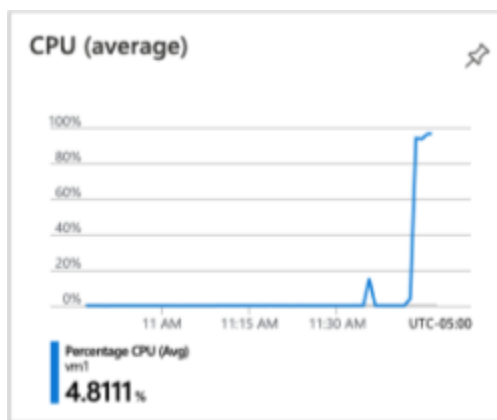
Distribución de carga con 2 máquinas virtuales, ya que la suscripción de estudiantes de Azure, sólo permite la asignación de IP con puertos de enlace máximo 2 veces.

```
inside Fibonacci
vm2@VM2:~/LAB09ARSW/FibonacciApp/postman/part2$[200 OK, 209.21kB, 27.8s]
```

	executed	failed
iterations	10	0
requests	10	0
test-scripts	10	0
prerequest-scripts	0	0
assertions	0	0
total run duration: 5m 8.4s		
total data received: 2.09MB (approx)		
average response time: 30.7s [min: 27.7s, max: 56.4s, s.d.: 8.5s]		

En las ejecuciones paralelas las 10 peticiones son exitosas con 10/10, con una respuesta de 30,7 s.

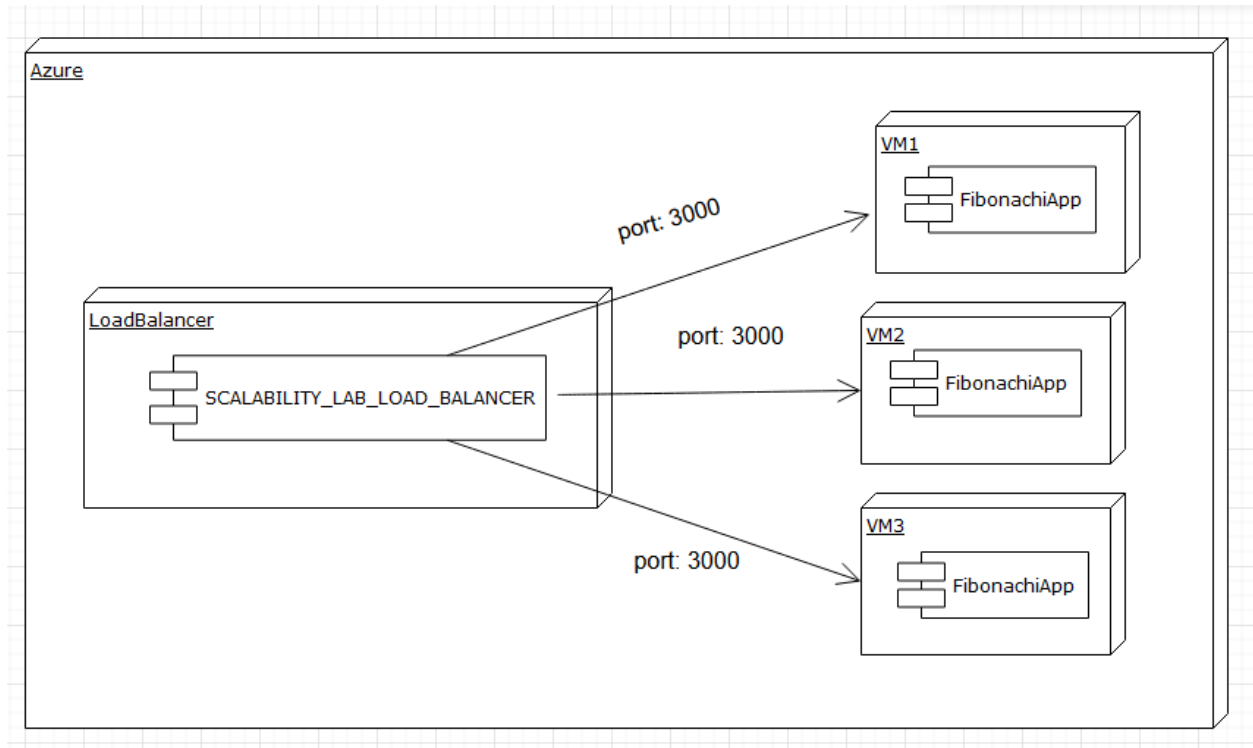
Comportamiento de las máquinas de la CPU



Como se dijo anteriormente, la suscripción de azure estudiante no permite realizar más máquinas virtuales. Sin embargo, observando el nuevo comportamiento con el balanceador el éxito de las peticiones aumentó con este tipo de escalabilidad pues, gracias a la infraestructura y aumento de

recursos se puede controlar la carga de la aplicación. Agregando la cantidad adecuada de recursos se puede controlar el aumento de carga.

**Presente el Diagrama de Despliegue de la solución.**





## **Bibliografía**

<https://azureprice.net/?filter=B2ms>

<https://learn.microsoft.com/es-es/azure/virtual-machines/dedicated-host-general-purpose-skus>

<https://learn.microsoft.com/en-us/azure/azure-monitor/autoscale/autoscale-overview>

<https://learn.microsoft.com/en-us/azure/load-balancer/load-balancer-overview>

<https://docs.microsoft.com/es-es/learn/modules/create-linux-virtual-machine-in-azure/2-create-a-linux-virtual-machine>

<https://docs.microsoft.com/en-us/azure/storage/common/storage-account-overview>

<https://docs.microsoft.com/en-us/azure/virtual-network/public-ip-addresses>

<https://docs.microsoft.com/en-us/azure/virtual-network/virtual-networks-overview>

<https://docs.microsoft.com/en-us/azure/virtual-network/network-security-groups-overview>

<https://azurebrains.com/2019/02/07/azure-load-balancer/>

<https://docs.microsoft.com/en-us/azure/load-balancer/>

<https://www.hostdime.com.ar/blog/por-que-usar-un-equilibrador-de-carga-o-load-balancer/>

<https://docs.microsoft.com/en-us/azure/vpn-gateway/about-zone-redundant-vnet-gateways>

<https://docs.microsoft.com/en-us/azure/virtual-network/network-security-group-how-it-works>