**Introducción**

La administración de máquinas virtuales a escala puede resultar complicada, especialmente cuando los patrones de uso varían y la demanda en las aplicaciones fluctúa. Los administradores de Azure deben poder ajustar sus recursos de máquina virtual para que cumplan con las demandas cambiantes. Al mismo tiempo, quieren mantener la coherencia de la configuración de las máquinas virtuales para garantizar la estabilidad de las aplicaciones. Lograr estos objetivos significa mantener el rendimiento y la capacidad de respuesta, a la vez que se minimizan los costos de la ejecución continua de un gran número de máquinas virtuales.

El sitio web de la empresa usa máquinas virtuales y administra cargas de trabajo de gran tamaño. El departamento de TI quiere asegurarse de que las máquinas virtuales se puedan ajustar dinámicamente a los aumentos y reducciones de las cargas de trabajo. También quieren asegurarse de que hay un plan de continuidad empresarial para proporcionar máquinas de alta disponibilidad. Usted es responsable de implementar las máquinas virtuales de alta disponibilidad. Decide usar Azure Virtual Machine Scale Sets y la característica de escalabilidad automática.

En este módulo, obtendrá información sobre el escalado de máquinas virtuales. Obtenga información sobre las zonas de disponibilidad, los conjuntos de disponibilidad, la actualización y los dominios de error. También obtendrá información sobre los conjuntos de escalado y el escalado automático.

El objetivo de este módulo es aprender a responder correctamente a las cargas de trabajo de máquina virtual cambiantes.

**Objetivos de aprendizaje**

En este módulo aprenderá a:

* Implementar conjuntos de disponibilidad y zonas de disponibilidad
* Implementar dominios de error y actualización
* Implemente Azure Virtual Machine Scale Sets.
* Escalar automáticamente máquinas virtuales

**Planificación del mantenimiento y el tiempo de inactividad**

Los administradores de Azure deben estar preparados en caso de errores planeados y no planeados. Vamos a explorar tres escenarios que pueden provocar que la máquina virtual de Azure se vea afectada.

**Aspectos que se deben conocer sobre el planeamiento del mantenimiento**

Un plan de disponibilidad para máquinas virtuales de Azure debe incluir estrategias para el mantenimiento de hardware no planeado, un tiempo de inactividad inesperado y el mantenimiento planeado. Cuando revise los escenarios siguientes, piense en cómo estos escenarios pueden afectar el sitio web de la empresa de ejemplo.

* Se produce un evento de **mantenimiento de hardware no planeado** cuando la plataforma Azure predice que el hardware o cualquier componente de la plataforma asociado a una máquina física está a punto de presentar un error. Cuando la plataforma predice un error, emite un evento de mantenimiento de hardware no planeado. Azure usa tecnología de migración en vivo para migrar las máquinas virtuales del hardware en el que se producen errores a una máquina física en buen estado. La migración en vivo es una operación de conservación de máquinas virtuales que solo pausa la máquina virtual durante un breve período de tiempo, pero el rendimiento se podría ver reducido antes o después del evento.
* El **tiempo de inactividad inesperado** ocurre cuando en el hardware o en la infraestructura física de la máquina virtual se produce un error de manera imprevista. El tiempo de inactividad inesperado puede incluir errores de la red local, errores de los discos locales u otros errores de nivel de bastidor. Cuando se detecta, la plataforma Azure migra (recupera) automáticamente la máquina virtual a una máquina física en buen estado en el mismo centro de datos. Durante el procedimiento de recuperación, las máquinas virtuales experimentan tiempos de inactividad (reinicio) y, en algunos casos, pérdidas de la unidad temporal.
* Los eventos de **mantenimiento planeado** son actualizaciones periódicas que realiza Microsoft en la plataforma Azure subyacente para mejorar en general la fiabilidad, el rendimiento y la seguridad de la infraestructura de la plataforma sobre las que se ejecutan las máquinas virtuales. La mayoría de estas actualizaciones se realizan sin que las máquinas virtuales ni los servicios en la nube resulten afectados.

**Nota**

Microsoft no actualiza automáticamente el sistema operativo de la máquina virtual ni ningún otro software. El usuario tiene el control total y la responsabilidad plena de realizar estas actualizaciones. Pero el host de software y el hardware subyacentes se revisan periódicamente para garantizar la fiabilidad y el alto rendimiento.

**Creación de conjuntos de disponibilidad**

Un conjunto de disponibilidad es una característica lógica que puede usar para asegurarse de que un grupo de máquinas virtuales relacionadas se implementen juntas. Esta agrupación permite evitar que un único punto de error afecte a todas las máquinas. Agrupar las máquinas virtuales garantiza que no todas se actualicen al mismo tiempo durante una actualización del sistema operativo host en el centro de datos.

**Aspectos que conviene saber sobre los conjuntos de disponibilidad**

Vamos a revisar algunas características de los conjuntos de disponibilidad.

* Todas las máquinas virtuales de un conjunto de disponibilidad deben realizar el conjunto idéntico de funcionalidades.
* Todas las máquinas virtuales de un conjunto de disponibilidad deben tener instalado el mismo software.
* Azure garantiza que las máquinas virtuales de un conjunto de disponibilidad se ejecuten en varios servidores físicos, grupos de proceso, unidades de almacenamiento y conmutadores de red.

Si se produce un error de hardware o software de Azure, solo se ve afectado un subconjunto de las máquinas virtuales del conjunto de disponibilidad. La aplicación permanece al día y sigue estando disponible para los clientes.

* Puede crear una máquina virtual y un conjunto de disponibilidad al mismo tiempo.

Una máquina virtual solo se puede agregar a un conjunto de disponibilidad cuando se crea la máquina virtual. Para cambiar el conjunto de disponibilidad de una máquina virtual, debe eliminar la máquina y volver a crearla.

* Puede crear conjuntos de disponibilidad mediante Azure Portal, plantillas de Azure Resource Manager (ARM), scripting o herramientas de API.
* Microsoft proporciona Acuerdos de Nivel de Servicio (SLA) sólidos para máquinas virtuales y conjuntos de disponibilidad de Azure. Para información más detallada, consulte [SLA para Azure Virtual Machines](https://azure.microsoft.com/support/legal/sla/virtual-machines/v1_9/).

**Nota**

Agregar las máquinas virtuales a un conjunto de disponibilidad no protegerá las aplicaciones contra errores del sistema operativo ni específicos de las aplicaciones. Deberá explorar otras técnicas de recuperación ante desastres y copia de seguridad para proporcionar protección en el nivel de aplicación.

**Aspectos que se deben tener en cuenta al usar conjuntos de disponibilidad**

Los conjuntos de disponibilidad son una funcionalidad fundamental para compilar soluciones en la nube confiables. Al planear conjuntos de disponibilidad, tenga en mente estos principios generales:

* **Considere la redundancia**. Para alcanzar la redundancia en la configuración, coloque varias máquinas virtuales en un conjunto de disponibilidad.
* **Considere la posibilidad de separar las capas de aplicación**. Cada capa de aplicación que se ejerce en la configuración debe estar en un conjunto de disponibilidad independiente. Esta separación ayuda a mitigar un único punto de error en todas las máquinas.
* **Tenga en cuenta el equilibrio de carga**. Para lograr una alta disponibilidad y un gran rendimiento de la red, cree un conjunto de disponibilidad de carga equilibrada mediante Azure Load Balancer. Load Balancer distribuye el tráfico entrante entre instancias de trabajo de servicios que se definen en el conjunto de disponibilidad de carga equilibrada.
* **Considere el uso de discos administrados**. Puede usar discos administrados de Azure con las máquinas virtuales de Azure en conjuntos de disponibilidad para el almacenamiento en el nivel de bloque.

**Revisión de dominios de actualización y dominios de error**

Azure Virtual Machine Availability Sets implementa dos conceptos de nodo para ayudar a Azure a mantener la alta disponibilidad y la tolerancia a errores al implementar y actualizar aplicaciones: *dominios de actualización* y *dominios de error*. Cada máquina virtual de un conjunto de disponibilidad se coloca en un dominio de actualización y un dominio de error.

**Aspectos que se deben conocer sobre los dominios de actualización**

Un dominio de actualización es un grupo de nodos que se actualizan en conjunto durante el proceso de actualización de un servicio (o *lanzamiento*). Un dominio de actualización permite a Azure realizar actualizaciones incrementales o graduales en una implementación. Estas son algunas otras características de los dominios de actualización.

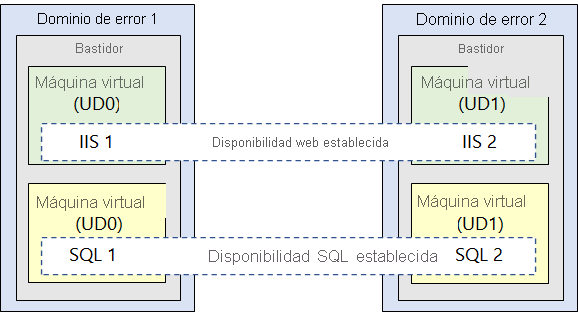
* Cada dominio de actualización incluye un conjunto de máquinas virtuales y hardware físico asociado que se puede actualizar y reiniciar al mismo tiempo.
* Durante el mantenimiento planeado solo se reinicia un dominio de actualización cada vez.
* De manera predeterminada, hay cinco dominios de actualización (que el usuario no puede configurar).
* Puede configurar hasta 20 dominios de actualización.

**Aspectos que se deben conocer sobre los dominios de error**

Un dominio de error es un grupo de nodos que representan una unidad física de error. Piense en un dominio de error como nodos que pertenecen al mismo bastidor físico.

* Un dominio de error define un grupo de máquinas virtuales que comparten un conjunto común de hardware (o *conmutadores*) que, a su vez, comparten un único punto de error. Un ejemplo es un bastidor de servidor que recibe servicio de un conjunto de conmutadores de alimentación o de red.
* Dos dominios de error trabajan en conjunto para mitigar los errores de hardware, las interrupciones de red, las interrupciones de alimentación o las actualizaciones de software.

Echemos un vistazo a un escenario con dos dominios de error que tienen dos máquinas virtuales cada uno. Las máquinas virtuales de cada dominio de error están contenidas en diferentes conjuntos de disponibilidad. El conjunto de disponibilidad web contiene dos máquinas virtuales con una máquina de cada dominio de error. El conjunto de disponibilidad SQL contiene dos máquinas virtuales con una máquina de cada dominio de error.



**Revisión de zonas de disponibilidad**

Completado100 XP

* 2 minutos

Las zonas de disponibilidad son una oferta de alta disponibilidad que protege las aplicaciones y datos de los errores del centro de datos. Una zona de disponibilidad de una región de Azure es la combinación de un dominio de error y un dominio de actualización.

Piense en un escenario en el que se crean tres o más máquinas virtuales en tres zonas de una región de Azure. Las máquinas virtuales se distribuyen eficazmente entre tres dominios de error y tres dominios de actualización. La plataforma Azure reconoce esta distribución entre dominios de actualización para asegurarse de que las máquinas virtuales de distintas zonas no se actualicen al mismo tiempo.

Puede utilizar las zonas de disponibilidad para crear alta disponibilidad en la arquitectura de sus aplicaciones si coloca sus recursos de proceso, almacenamiento, redes y datos en una zona y los replica en otras.

**Aspectos que conviene saber sobre las zonas de disponibilidad**

Revise estas características de las zonas de disponibilidad.

* Las zonas de disponibilidad son ubicaciones físicas exclusivas dentro de una región de Azure.
* Cada zona consta de uno o varios centros de datos equipados con alimentación, refrigeración y redes independientes.
* Para garantizar la resistencia, hay un mínimo de tres zonas independientes en todas las regiones habilitadas.
* La separación física de las zonas de disponibilidad dentro de una región protege las aplicaciones y los datos frente a los errores del centro de datos.
* Los servicios con redundancia de zona replican las aplicaciones y los datos en zonas de disponibilidad como protección ante los únicos puntos de error.

**Aspectos que se deben tener en cuenta al usar zonas de disponibilidad**

Los servicios de Azure que admiten zonas de disponibilidad se dividen en dos categorías.

Expandir tabla

| **Category** | **Descripción** | **Ejemplos** |
| --- | --- | --- |
| **Servicios de zona** | Los servicios *de zona* de Azure anclan cada recurso a una zona específica. | - Máquinas virtuales de Azure - Discos administrados de Azure - Direcciones IP estándar |
| **Servicios con redundancia de zona** | En el caso de los servicios de Azure con redundancia de zona, la plataforma se replica automáticamente en todas las zonas. | - Azure Storage con redundancia de zona - Azure SQL Database |

**Sugerencia**

Para lograr una continuidad empresarial integral en Azure, cree la arquitectura de aplicación mediante la combinación de zonas de disponibilidad y pares de regiones de Azure.

**Comparación entre el escalado vertical y horizontal**

Una configuración de máquina virtual sólida incluye compatibilidad con la escalabilidad. La escalabilidad permite mejorar el rendimiento de una máquina virtual en proporción a la disponibilidad de los recursos de hardware asociados. Una máquina virtual escalable puede controlar los aumentos en las solicitudes sin afectar negativamente el tiempo de respuesta ni el rendimiento. Para la mayoría de las operaciones de escalado, hay dos opciones de implementación: *vertical* y *horizontal*.

**Aspectos que hay que saber sobre el escalado vertical**

El escalado vertical, también conocido como *escalado y reducción vertical*, implica aumentar o disminuir el **tamaño** de la máquina virtual como respuesta a una carga de trabajo. El escalado vertical hace que las máquinas virtuales sean más poderosas (escalado vertical) o menos poderosas (reducción vertical).

Interfaz de usuario gráfica

Descripción generada automáticamente

Estos son algunos escenarios en los que puede ser ventajoso utilizar el escalado vertical:

* Si tiene un servicio basado en una máquina virtual infrautilizada (por ejemplo, los fines de semana), puede usar el escalado vertical para disminuir el tamaño de la máquina virtual y reducir los costos mensuales.
* Puede implementar el escalado vertical para aumentar el tamaño de la máquina virtual a fin de responder ante una demanda mayor sin tener que crear máquinas virtuales adicionales.

**Aspectos que hay que saber sobre el escalado horizontal**

El escalado horizontal se usa para ajustar el **número** de máquinas virtuales de la configuración para admitir la carga de trabajo cambiante. Al implementar el escalado horizontal, hay un aumento (escalado horizontal) o una disminución (reducción horizontal) en el número de instancias de máquina virtual.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

**Aspectos que se deben tener en cuenta al usar el escalado vertical y horizontal**

Revise estas consideraciones respecto del escalado vertical y horizontal. Piense en qué implementación podría necesitar para admitir el sitio web de la empresa.

* **Tenga en cuenta las limitaciones**. En términos generales, el escalado horizontal tiene menos limitaciones que el vertical. Una implementación de escalado vertical depende de la disponibilidad de hardware más grande, que alcanza rápidamente un límite superior y puede variar según la región. El escalado vertical también suele requerir que una máquina virtual se detenga y reinicie, lo que puede limitar temporalmente el acceso a aplicaciones o datos.
* **Tenga en cuenta la flexibilidad**. Cuando se trabaja en la nube, el escalado horizontal resulta más flexible. Una implementación de escalado horizontal permite ejecutar potencialmente miles de máquinas virtuales para administrar los cambios en la carga de trabajo y el rendimiento.
* **Tenga en cuenta el reaprovisionamiento**. El *reaprovisionamiento* es el proceso de quitar una máquina virtual existente y reemplazarla por una nueva. Un plan de disponibilidad sólido considera dónde es posible que se requiera el reaprovisionamiento y los planes de interrupciones en el servicio. Si es posible que sea necesario el reaprovisionamiento, determine si necesita mantener y migrar los datos a la máquina nueva.

**Implementación de Azure Virtual Machine Scale Sets**

Las instancias de Azure Virtual Machine Scale Sets son un recurso de Azure Compute que puede utilizar para implementar y administrar un conjunto de máquinas virtuales **idénticas**. Al implementar Virtual Machine Scale Sets y configurar todas las máquinas virtuales de la misma manera, obtiene una *escalabilidad automática* verdadera. Virtual Machine Scale Sets aumenta automáticamente el número de instancias de máquina virtual a medida que la demanda de la aplicación aumenta, y reduce el número de instancias de máquina a medida que la demanda disminuye.

Con Virtual Machine Scale Sets, no es necesario aprovisionar previamente las máquinas virtuales. Esto facilita la creación de servicios a gran escala cuyo objetivo son las cargas de trabajo en contenedores, de macroproceso y macrodatos. A medida que aumentan las cargas de trabajo, se pueden agregar más instancias de máquina virtual. A medida que disminuyen las cargas de trabajo, se pueden quitar instancias de máquina virtual. El proceso de agregar y quitar máquinas puede ser manual, automatizado, o bien una combinación de ambos.

**Aspectos que se deben saber sobre Azure Virtual Machine Scale Sets**

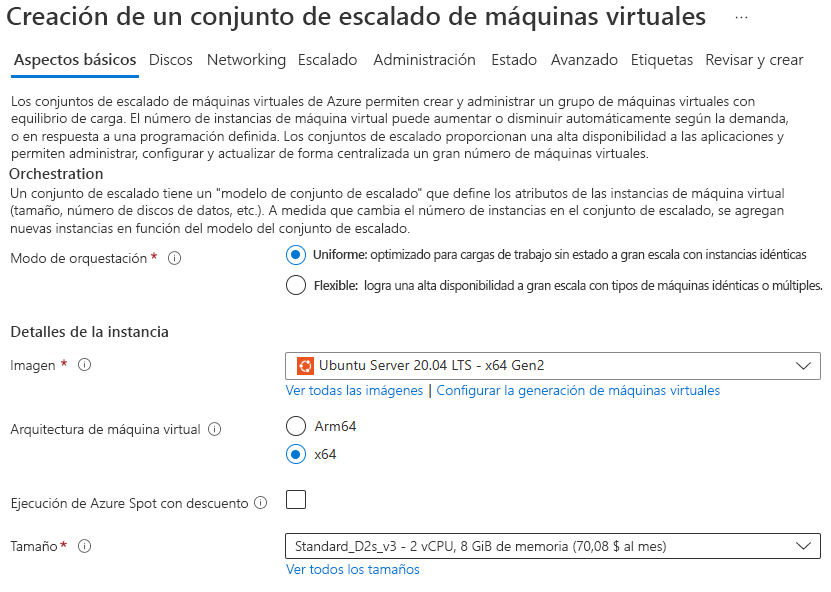
Revise estas características de Azure Virtual Machine Scale Sets.

* Todas las instancias de máquina virtual se crean a partir de la misma imagen y configuración del sistema operativo base. Este enfoque le permite administrar fácilmente cientos de máquinas virtuales sin tareas adicionales de configuración o administración de red.
* Virtual Machine Scale Sets admite el uso de Azure Load Balancer para la distribución de tráfico de capa 4 básica y Azure Application Gateway para la distribución de tráfico de capa 7 más avanzada, además de la terminación SSL.
* Puede usar Virtual Machine Scale Sets para ejecutar varias instancias de la aplicación. Si una de las instancias de máquina virtual tiene un problema, los clientes siguen accediendo a la aplicación a través de otra instancia de máquina virtual con una interrupción mínima.
* La demanda de la aplicación por parte de los clientes puede cambiar a lo largo del día o de la semana. A fin de satisfacer la demanda de los clientes, Virtual Machine Scale Sets implementa la escalabilidad automática para aumentar y reducir automáticamente el número de máquinas virtuales.
* Virtual Machine Scale Sets admite hasta 1000 instancias de máquina virtual. Si crea y carga sus propias imágenes de máquina virtual personalizadas, el límite es de 600 instancias.

**Creación de implementaciones de Virtual Machine Scale Sets**

Puede implementar Azure Virtual Machine Scale Sets en Azure Portal. Especifique el número de máquinas virtuales y sus tamaños e indique las preferencias para usar instancias de Azure Spot, discos administrados de Azure y directivas de asignación.

En Azure Portal, hay varias opciones que se deben configurar para crear una implementación de Azure Virtual Machine Scale Sets.



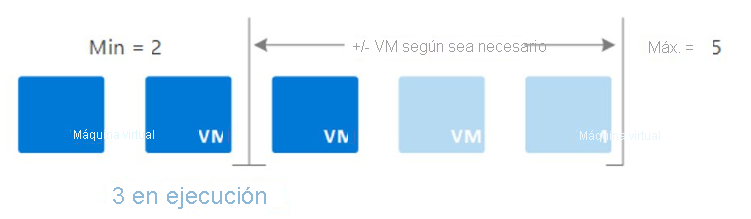
* **Modo de orquestación**: elija cómo el conjunto de escalado administra las máquinas virtuales. En modo de orquestación flexible, se crea y se agrega manualmente una máquina virtual de cualquier configuración al conjunto de escalado. En modo de orquestación uniforme, se define un modelo de máquina virtual y Azure generará instancias idénticas basadas en ese modelo.
* **Imagen**: elija la aplicación o el sistema operativo base de la máquina virtual.
* **Arquitectura de máquina virtual**: Azure permite elegir máquinas virtuales basadas en x64 o Arm64 para ejecutar las aplicaciones.
* **Ejecución con descuento de Azure Spot**: Azure Spot ofrece una tarifa con descuento para la capacidad de Azure sin usar y no precios de pago por uso. Las cargas de trabajo deben tolerar pérdidas de infraestructura, ya que Azure puede recuperar la capacidad.
* **Tamaño**: seleccione un tamaño de máquina virtual para admitir la carga de trabajo que quiere ejecutar. El tamaño que elija determina factores tales como la capacidad de almacenamiento, la memoria y la capacidad de procesamiento. Azure ofrece una amplia variedad de tamaños para admitir muchos tipos de usos. Azure cobra un precio por hora según el sistema operativo y el tamaño de la máquina virtual.

En la pestaña **Opciones avanzadas**, también puede seleccionar lo siguiente:

* **Habilitar el escalado con más de 100 instancias**: identifique la asignación de escalado que prefiere. Si selecciona **No**, la implementación de Virtual Machine Scale Sets está limitada a un grupo de selección de ubicación con una capacidad máxima de 100. Si selecciona **Sí**, la implementación puede abarcar varios grupos de selección de ubicación con capacidad de hasta 1000. Al seleccionar **Sí**, también se cambian las características de disponibilidad de la implementación.
* **Algoritmo de propagación**: Microsoft recomienda asignar **Propagación máxima** para la implementación. Este enfoque brinda una propagación óptima.

**Implementación de la escalabilidad automática**

Una implementación de Azure Virtual Machine Scale Sets puede aumentar o disminuir automáticamente el número de instancias de máquina virtual que ejecutan la aplicación. Este proceso se denomina *escalabilidad automática*. La escalabilidad automática le permite escalar dinámicamente la configuración a fin de satisfacer las cambiantes demandas de carga de trabajo.



La escalabilidad automática minimiza el número de instancias de máquina virtual innecesarias que ejecutan la aplicación cuando la demanda es baja. Los clientes siguen recibiendo un nivel de rendimiento aceptable a medida que crece la demanda y se agregan automáticamente más instancias de máquina virtual.

**Aspectos que se deben tener en cuenta al usar la escalabilidad automática**

Revise las consideraciones siguientes sobre la escalabilidad automática. Piense en cómo este proceso puede ser una ventaja para la implementación del sitio web de la empresa.

* **Considere la posibilidad de ajustar automáticamente la capacidad**. Puede crear reglas de escalabilidad automática que definan el rendimiento aceptable para una experiencia positiva del cliente. Cuando se cumplen los umbrales definidos, las reglas de escalabilidad automática actúan para ajustar la capacidad de la implementación de Virtual Machine Scale Sets.
* **Considere la escalabilidad horizontal**. Si aumenta la demanda de la aplicación, aumentará la carga en las instancias de máquina virtual de la implementación. Si el aumento de la carga es continuado, en lugar de ser algo puntual, puede configurar reglas de escalabilidad automática para aumentar el número de instancias de máquina virtual en la implementación.
* **Considere la reducción horizontal**. La demanda de la aplicación puede reducirse por las tardes o durante los fines de semana. Si la reducción de la carga es constante a lo largo de un período, puede configurar reglas de escalabilidad automática a fin de reducir el número de instancias de máquina virtual de la implementación. La acción de reducción horizontal permite disminuir el costo de ejecutar la implementación de Virtual Machine Scale Sets, ya que solo se ejecuta el número de instancias necesario para satisfacer la demanda actual.
* **Tenga en cuenta los eventos programados**. Puede implementar la escalabilidad automática y programar eventos para aumentar o reducir automáticamente la capacidad de la implementación en momentos determinados.
* **Tenga en cuenta la sobrecarga**. Usar Azure Virtual Machine Scale Sets con la escalabilidad automática reduce la sobrecarga de administración que implica supervisar y optimizar el rendimiento de la aplicación.

**Configuración de escalado automático**

Al crear una implementación de Azure Virtual Machine Scale Sets en Azure Portal, puede habilitar la escalabilidad automática. Para lograr un rendimiento óptimo, debe definir un número mínimo, máximo y predeterminado de instancias de máquina virtual que se usarán durante el proceso de escalabilidad automática.

En Azure Portal, hay varias opciones que se deben configurar para implementar la escalabilidad automática con Azure Virtual Machine Scale Sets.

**Directiva de escalado**: la escalabilidad manual mantiene un recuento fijo de instancias. La escalabilidad automática personalizada escala la capacidad según cualquier programación, en función de cualquier métrica.

* **Número mínimo de VM**: especifique el número mínimo de máquinas virtuales que debe haber disponibles al aplicar la escalabilidad automática a la implementación de Virtual Machine Scale Sets.
* **Número máximo de VM**: especifique el número máximo de máquinas virtuales que puede haber disponibles al aplicar la escalabilidad automática a la implementación.

**Escalado horizontal**

* **Umbral de CPU**: especifique el umbral de porcentaje de uso de la CPU que va a desencadenar la regla de escalabilidad horizontal automática.
* **Duración en minutos**: la duración en minutos es la cantidad de tiempo que el motor de escalabilidad automática buscará las métricas. Por ejemplo, 10 minutos significa que, cada vez que se ejecute la escalabilidad automática, se consultarán las métricas de los últimos 10 minutos. Esta demora permite que las métricas se estabilicen e impide que reaccionen a picos transitorios.
* **Número de máquinas virtuales por las que incrementar**: especifique el número de máquinas virtuales que se van a agregar a la implementación de Virtual Machine Scale Sets cuando se desencadene la regla de escalabilidad horizontal automática.

**Reducción horizontal**

* **Umbral de CPU de reducción horizontal**: especifique el umbral de porcentaje de uso de la CPU que va a desencadenar la regla de reducción horizontal automática.
* **Número de máquinas virtuales por las que disminuir**: especifique el número de máquinas virtuales que se van a quitar de la implementación cuando se desencadene la regla de reducción horizontal automática.

**Directiva de reducción horizontal**: la característica de la [directiva de reducción horizontal](https://learn.microsoft.com/es-es/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-scale-in-policy) proporciona a los usuarios una manera de configurar el orden en el que se reducen horizontalmente las máquinas virtuales.

**Prueba de conocimientos**

La organización tiene varios requisitos de configuración y disponibilidad de las máquinas virtuales. Usted tiene la responsabilidad de ayudar con la configuración a fin de cumplir con solicitudes y resolver problemas.

* El equipo de administración está probando una implementación de Azure Virtual Machine Scale Sets con cinco máquinas virtuales. Durante las pruebas, las alertas de supervisión muestran todas las máquinas virtuales que se ejecutan a capacidad máxima. Sin embargo, se detecta que, cuando se consume completamente la CPU, no se implementan más máquinas virtuales en el conjunto de escalado.
* El equipo de DevOps quiere configurar Azure Virtual Machine Scale Sets para sus servidores de producción. El jueves por la noche suele ser el momento de mayor actividad, ya que se preparan las entregas a los clientes que se realizarán el viernes al final del día. Por el contrario, el lunes temprano es el momento con menor actividad. Necesita un plan para agregar más máquinas cuando la carga de trabajo sea alta.
* A medida que aumenta la carga en la aplicación hospedada en una implementación de Azure Virtual Machine Scale Sets, lo que quiere es aumentar la capacidad de CPU de las instancias existentes, no implementar más instancias.

**Responda a las siguientes preguntas**

Elija la respuesta más adecuada para cada una de las preguntas siguientes. Después, seleccione **Comprobar las respuestas**.

Principio del formulario

**1. ¿Cómo puede asegurarse de que se implementan más máquinas virtuales para el equipo de administración una vez que se consume el 75 % de la CPU?**

Aumentar manualmente el recuento de instancias.

Cambiar el porcentaje de CPU al 50 %.

Habilitar la opción de escalabilidad automática.

**2. ¿Qué característica de Virtual Machine Scale Sets se puede configurar para agregar más máquinas de DevOps durante el momento de mayor producción?**

Reglas basadas en programación.

Escalado automático

Reglas basadas en métricas.

**3. ¿Qué tipos de escalabilidad puede usar para aumentar la *capacidad* de CPU de las instancias de Virtual Machine Scale Sets existentes?**

Escalado horizontal

Escalado vertical

Equilibrio de carga

* 1. C
  2. A
  3. B

Final del formulario

**Resumen y recursos**

Azure proporciona varias opciones de alta disponibilidad para máquinas virtuales. Puede lograr una alta disponibilidad mediante conjuntos de disponibilidad, zonas de disponibilidad y Azure Virtual Machine Scale Sets.

En este módulo, aprendió a configurar la disponibilidad de máquinas virtuales mediante conjuntos de disponibilidad y zonas de disponibilidad con dominios de actualización y error. Descubrió cómo escalar automáticamente máquinas virtuales y configurar el escalado vertical y horizontal. Revisó cómo implementar Virtual Machine Scale Sets, incluidas las opciones de resistencia y escalabilidad del almacenamiento. Exploró el uso de extensiones de script personalizado de máquina virtual de Azure para configurar automáticamente las máquinas virtuales.

Las principales conclusiones de este módulo son:

* Azure Virtual Machine Scale Sets permite la implementación y administración de un grupo de máquinas virtuales idénticas, lo que facilita la creación de servicios a gran escala.
* El escalado automático con Virtual Machine Scale Sets ayuda a optimizar el rendimiento ajustando automáticamente el número de instancias en función de las demandas de carga de trabajo.
* Los conjuntos de disponibilidad y las zonas de disponibilidad son características importantes en Azure para lograr una alta disponibilidad y tolerancia a errores para las máquinas virtuales.