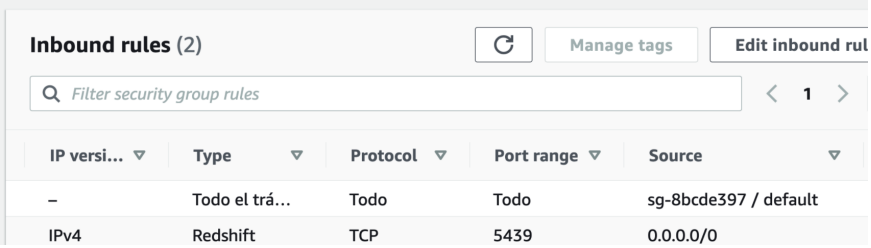
Solución - Evaluación escrita Nº 2

Pregunta 1. En AWS RDS, podemos permitir que una instancia de bases de datos sea pública. Suponga, además, que dejamos toda la configuración de conectividad según lo establecido POR DEFECTO (por ejemplo, VPC, subredes, grupos de seguridad). ¿Es esto suficiente para poder realizar consultas SQL en la instancia desde un dispositivo externo a su VPC? Si su respuesta es afirmativa, explique por qué; si su respuesta es negativa, indique el motivo y lo que habría que modificar para poder consultar la instancia.

R. No, no es suficiente. Si dejamos toda la configuración por defecto, particularmente el grupo de seguridad *default*, este permitirá la entrada solo a dispositivos que pertenezcan a este grupo, independientemente de que el dispositivo sea parte o no de la VPC. Habría que modificar, entonces, las reglas de entrada del grupo de seguridad *default* para que se permita la interacción con dispositivos que no sean parte de este por defecto.

Puede indicarse que se debe agregar directamente la IP del dispositivo que realizará la consulta, o simplemente dar acceso a toda IP. También es válido indicar que se haga uso de un grupo de seguridad distinto con reglas de entrada que permitan llevar a cabo la consulta. Lo importante es señalar que el grupo de seguridad y sus reglas de entrada son el aspecto a modificar.

Pregunta 2. Suponga que se tiene la configuración de reglas de entrada de la imagen en la configuración del grupo de seguridad "*default*" y que, además, las IP de los recursos dentro de la VPC son públicas. ¿Cuál de estas afirmaciones es INCORRECTA con respecto a los accesos a recursos que tengan las reglas de entrada de este grupo de seguridad?



- Dispositivos internos de la VPC podrían conectarse mediante SSH a los recursos asociados a este grupo de seguridad.
- **Dispositivos externos a la VPC podrían conectarse mediante SSH a los recursos asociados a este grupo de seguridad.**
- Dispositivos internos de la VPC podrían consultar a un *cluster* de Redshift asociado a este grupo de seguridad.
- Dispositivos externos a la VPC podrían consultar a un *cluster* de Redshift asociado a este grupo de seguridad.

R. La primera regla de entrada permite todo tipo de tráfico para dispositivos que sean parte del grupo de seguridad *default*. Por lo tanto, dispositivos dentro de la VPC **podrían** conectarse mediante SSH (si pertenecieran al mismo grupo de seguridad). No obstante, los dispositivos externos a la VPC no tienen forma de hacerlo, ya que el grupo de seguridad *default* solo puede ser asignado a dispositivos dentro de la VPC. La segunda regla de entrada da acceso a todos los dispositivos para que puedan realizar consultas a un *cluster* de Redshift. Dado que se asume que las IPs de los recursos dentro de la VPC son públicas, es posible realizar consultas a *clusters* Redshift con este grupo de seguridad tanto desde dispositivos internos como externos a la VPC.

Pregunta 3. Al crear una instancia EC2, se le añade por defecto un dispositivo de almacenamiento "raíz" llamado volumen. ¿Es necesario que la instancia EC2 posea este volumen raíz asociado? Justifique su respuesta aludiendo a lo que almacena este dispositivo.

R. Sí, es necesario ya que este volumen almacenará la AMI de la instancia EC2, esto es, el código del sistema operativo y otras aplicaciones que estarán instaladas en nuestro servidor virtual. Más aún, no existe forma de crear una instancia EC2 sin un volumen raíz.

Pregunta 4. Suponga que necesita desarrollar un programa con el siguiente comportamiento: cada vez que se sube una foto a un *bucket* S3, el programa debe detectar la subida, procesar la foto y crear copias de esta con distintas dimensiones. Describa dos ventajas (monetarias, computacionales, entre otras) JUSTIFICADAS de llevar a cabo esto con una función Lambda en vez de tener el programa corriendo en un servidor EC2.

R. Existen varias justificaciones válidas. A continuación, se listan algunas:

- Las instancias EC2 generan un cobro constante mientras estén corriendo. Más aún, aunque nuestras instancias estén detenidas, la existencia del volumen raíz generará un cobro adicional. Las funciones Lambda, por otra parte, solo cobran por ejecuciones y tiempo de ejecución. Como el programa no estará procesando imágenes constantemente, es menos costoso tener una función Lambda que lleve a cabo el procesamiento cada vez que se suba una imagen, en contraste a una instancia EC2 que requiera estar encendida para poder detectar la subida de imágenes.
- Dependiendo de la AMI de nuestras instancias EC2, podríamos necesitar realizar instalaciones adicionales para que la función pueda procesarse directamente (por ejemplo, soporte para el lenguaje de programación de nuestro programa). Con una función Lambda, en cambio, solo tendríamos que preocuparnos de que se incluyan las librerías necesarias dentro del código antes de subirlo y escoger el *kernel* de preferencia, reduciendo el tiempo de puesta en marcha y facilitando la mantención.
- Se puede asociar un *trigger* de S3 a Lambda para que detecte la subida de fotos. En EC2 no existe un mecanismo similar, por lo que habría que usar servicios adicionales para procesar las fotos cuando se suban, aumentando la complejidad del programa.

Pregunta 5. Suponga que posee una función Lambda DENTRO de una VPC y desea que esta se encargue de dos cosas: (1) Procesar un archivo CSV subido a un bucket S3; (2) Insertar las filas del CSV en una tabla en RDS dentro de la misma VPC. ¿Cuál de los siguientes requerimientos NO ES NECESARIO para poder llevar a cabo el comportamiento descrito?

- Crear un *endpoint* en la VPC que permita que los dispositivos dentro de esta puedan acceder a todos los recursos y acciones de S3 en la cuenta.
- **Asociar al bucket S3 una política que otorgue acceso público a sus objetos.**
- Configurar los grupos de seguridad de la instancia RDS y función Lambda para que estas puedan interactuar entre sí.
- Configurar un *trigger* de Lambda que gatille su ejecución al subir el archivo al bucket S3.
- Asociar al rol IAM de Lamba las políticas "AWSLambdaVPCAccessExecutionRole" y "AmazonS3ReadOnlyAccess".

R. Como se vio en clases, para que una función Lambda pueda tener accesos a recursos S3 encontrándose dentro de una VPC, es necesario que esta posea un *endpoint* que otorgue acceso a estos recursos de la cuenta (ya que los *buckets* de S3, por defecto, están fuera de la VPC). Por otra parte, el rol IAM de la función debe tener permisos para ejecutar dentro de una VPC (AWSLambdaVPCAccessExecutionRole) y para poder leer archivos de S3 (AmazonS3ReadOnlyAccess). Además, los grupos de seguridad deben permitir que la instancia RDS pueda ser consultada por la función Lambda, ya que en otro caso la conexión arrojará error por *timeout*. Finalmente, no puede faltar el *trigger* de S3, ya que la función Lambda solo se ejecuta a partir de eventos.

De lo anterior **no es necesario dar acceso público a los objetos del *bucket***. Mientras la función tenga los permisos de lectura necesarios, podrá procesar los archivos subidos al *bucket*, lo que es más seguro que exponer los objetos de forma pública.

Verdadero y Falso

La siguiente sección contiene afirmaciones que son verdaderas o falsas. Si marca una afirmación como falsa, debe justificar su respuesta. Si marca la afirmación como verdadera, no es necesario que justifique.

Pregunta 6. Si quiero consultar una base de datos de una instancia RDS, esta necesariamente debe ser pública.

R. Falso. Puede ser consultada desde una instancia EC2 dentro de la misma VPC que tenga acceso mediante reglas de entrada, lo que evita la necesidad de hacer nuestra instancia RDS pública.

Pregunta 7. Si pierdo la llave privada de una instancia EC2, existen otras alternativas para poder conectarme a ella.

R. Verdadero. Podemos usar el servicio de AWS SSM (Session Manager).

Pregunta 8. Las funciones Lambda FUERA de una VPC pueden acceder al contenido de un bucket S3 solo teniendo los permisos necesarios en su rol IAM.

R. Verdadero. Estando fuera de la VPC, pueden acceder sin problemas siempre que su rol IAM incluya permisos relativos a S3. Dentro de la VPC es donde se vuelve necesaria la inclusión de un *endpoint* que otorgue acceso a los recursos de S3.

Pregunta 9. En una función Lambda, si se desea acceder a una base de datos contenida en una instancia RDS, es necesario definir las credenciales de esta dentro del código mismo (*i.e.* "*hardcodeadas*")

R. Falso. Se pueden configurar como variables de entorno y ser accedidas directamente en el código (por ejemplo, `os.environ` en Python).

Pregunta 10. Puedo realizar consultas a la base de datos de un *cluster* Redshift desde un dispositivo externo a la VPC, siempre y cuando otorgue acceso público y las reglas de entrada del grupo de seguridad lo permitan.

R. Verdadero y Falso. Las consultas a un *cluster* Redshift, al igual que a una instancia RDS, se pueden hacer siempre y cuando su IP sea pública (para que el dispositivo externo pueda encontrar el *endpoint*) y las reglas de entrada otorguen acceso a dispositivos fuera de la VPC.

No obstante la justificación anterior, si bien se esperaba que se asumiera, es válido indicar que la sentencia es **falsa** si se justifica con el hecho de no poseer las credenciales del usuario administrador de la base de datos del *cluster*. Este es el único caso en el que la respuesta se considera correcta.

Pregunta 11. Puedo configurar un rastreador en AWS Glue sin entregarle un clasificador.

R. Verdadero y Falso. Se esperaba que la respuesta fuera **verdadera**, ya que los rastreadores de por sí tienen clasificadores internos que son utilizados en caso de que no se les asocie ninguno creado por el usuario. No obstante, se podría argumentar que la sentencia es **falsa**, ya que los rastreadores siempre deben hacer uso de un clasificador (interno o no). Se considerarán correctos ambos casos, siempre que la justificación (en caso de indicar "Falso") aluda al uso obligatorio de un clasificador por el rastreador, independientemente de su origen.

Pregunta 12. Puedo configurar *un* job en AWS Glue sin preocuparme del rol IAM que ocupe para extraer, transformar y almacenar los datos.

R. Falso. Es necesario verificar que el rol IAM de un *job* otorgue los permisos necesarios de acceso a los recursos que utilice en la transformación (*buckets* S3, por ejemplo).