## **Documentación de como configurar la infraestructura en Azure integrando Snowflake y Airflow, asegurando la seguridad y eficiencia.**

## **Introducción**

Esta guía proporciona instrucciones para configurar una infraestructura en Microsoft Azure que integre Snowflake, un almacén de datos en la nube, y Apache Airflow, una plataforma de orquestación de flujos de trabajo. La configuración está diseñada para asegurar la seguridad y eficiencia de los datos y procesos.

## **Requisitos Previos**

* **Cuenta de Azure**: Asegúrate de tener una cuenta activa en Microsoft Azure.
* **Cuenta de Snowflake**: Necesitas una cuenta en Snowflake.
* **Conocimientos básicos**: Familiaridad con Azure, Snowflake y Apache Airflow.

## **Configuración de la Infraestructura en Azure**

### **1. Crear un Grupo de Recursos**

1. **Inicia sesión en el** [**Portal de Azure**](https://portal.azure.com/)**.**
2. **En el menú de la izquierda, selecciona "Grupos de recursos".**
3. **Haz clic en "Crear" y completa los campos:**
   * **Nombre**: mi-grupo-de-recursos
   * **Región**: Selecciona la región más cercana a tus usuarios o datos.

### **2. Configurar una Red Virtual (VNet)**

1. **En el portal de Azure, selecciona "Redes virtuales".**
2. **Haz clic en "Crear" y completa los campos:**
   * **Nombre**: mi-vnet
   * **Dirección IP**: Define el rango de direcciones IP (por ejemplo, 10.0.0.0/16).
   * **Región**: Debe coincidir con la del grupo de recursos.
3. **Configura las Subredes:**
   * **Subred Airflow**: 10.0.1.0/24
   * **Subred Datos**: 10.0.2.0/24

### **3. Crear una Máquina Virtual para Apache Airflow**

1. **En el portal de Azure, selecciona "Máquinas Virtuales".**
2. **Haz clic en "Crear" y completa los campos:**
   * **Nombre**: mi-airflow-vm
   * **Imagen**: Ubuntu o cualquier otro sistema operativo compatible.
   * **Tamaño**: Selecciona una máquina virtual adecuada para el rendimiento requerido.
   * **Red**: Asocia con la VNet creada.
3. **Configura el Acceso:**
   * **SSH**: Configura una clave SSH para el acceso seguro.
4. **Instala Airflow en la VM:**
   * **Conéctate a la VM** usando SSH.
   * **Instala Apache Airflow** siguiendo las [instrucciones oficiales](https://airflow.apache.org/docs/apache-airflow/stable/installation/index.html).

### **4. Configurar el Almacenamiento de Datos**

1. **Crea una Cuenta de Almacenamiento:**
   * **En el portal de Azure, selecciona "Cuentas de almacenamiento".**
   * **Haz clic en "Crear" y completa los campos:**
     + **Nombre**: mi-cuenta-almacenamiento
     + **Tipo de Cuenta**: Selecciona el tipo que se ajuste a tus necesidades (por ejemplo, "General-purpose v2").
2. **Configura Contenedores:**
   * **Crea un contenedor** para almacenar datos intermedios y resultados de procesos.

### **5. Integrar Snowflake**

1. **Configura Snowflake:**
   * **En el portal de Snowflake, accede a tu cuenta y configura un almacén de datos, base de datos y esquema según tus necesidades.**
2. **Conecta Airflow con Snowflake:**

**Instala el proveedor de Snowflake en Airflow:**bash  
Copiar código  
pip install 'apache-airflow-providers-snowflake'

**Configura la conexión en Airflow** a través de la interfaz web o el archivo airflow.cfg:  
python  
Copiar código  
from airflow.providers.snowflake.hooks.snowflake import SnowflakeHook

snowflake\_hook = SnowflakeHook(

snowflake\_conn\_id='my\_snowflake\_connection',

account='account\_name',

user='username',

password='password',

warehouse='warehouse\_name',

database='database\_name',

schema='schema\_name'

)

### **6. Configurar la Seguridad**

1. **Configura el Acceso Seguro:**
   * **Redes**: Usa grupos de seguridad para controlar el tráfico de red hacia y desde la VM de Airflow.
   * **Autenticación**: Usa autenticación basada en roles (RBAC) en Airflow para limitar el acceso.
   * **Datos**: Usa cifrado en reposo y en tránsito tanto para Snowflake como para el almacenamiento en Azure.
2. **Monitoreo y Alertas:**
   * **Configura alertas** para monitorizar eventos inusuales o problemas de seguridad.

### **7. Configurar la Eficiencia**

1. **Escalabilidad:**
   * **Escala la VM de Airflow** según la carga de trabajo usando el escalado automático de Azure.
   * **Optimiza Snowflake** ajustando el tamaño del almacén de datos según las necesidades.
2. **Optimización de Costos:**
   * **Revisa los costos de almacenamiento** y de procesamiento en Azure y Snowflake regularmente.
   * **Utiliza la capacidad de almacenamiento en frío** para datos que no se consultan frecuentemente.
3. **Mantenimiento:**
   * **Configura tareas programadas en Airflow** para la administración de datos y limpieza.