BIG DATA

Pentaho - Ejemplo 5

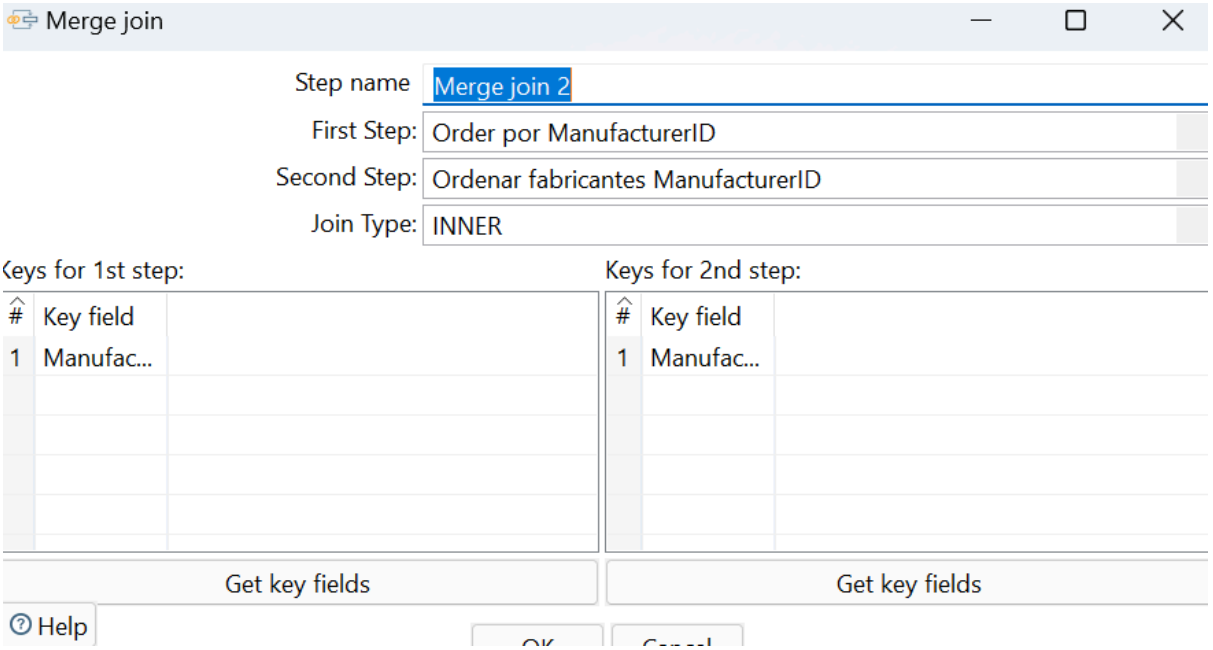
ALUMNO: José Antonio Díaz Aranda

CURSO: IA Y BIG DATA

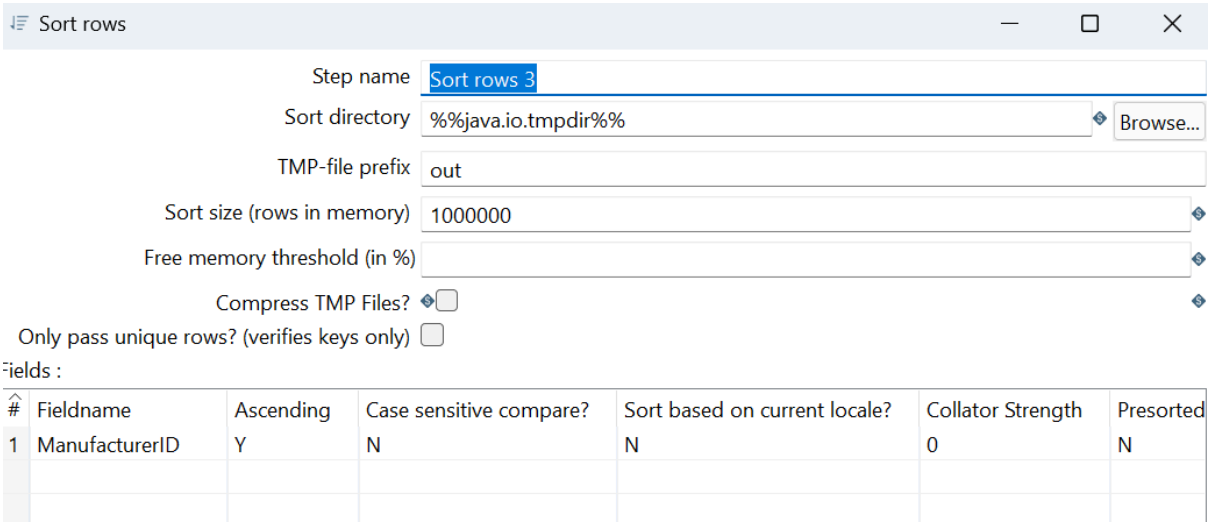
MÓDULO: BD

Tenemos de base la configuración del ejercicio 3, por lo que se explicará desde ese.

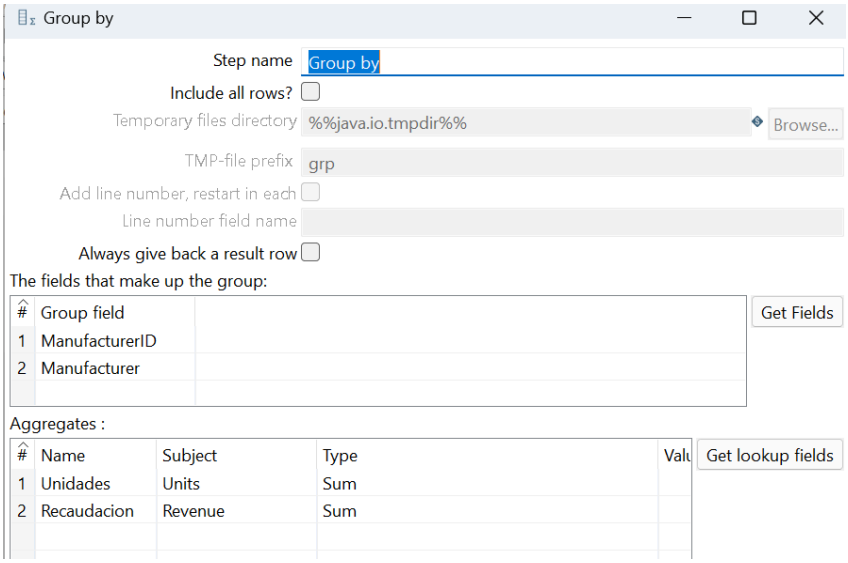
Hacemos un merge join y lo unimos por ManufacturerID



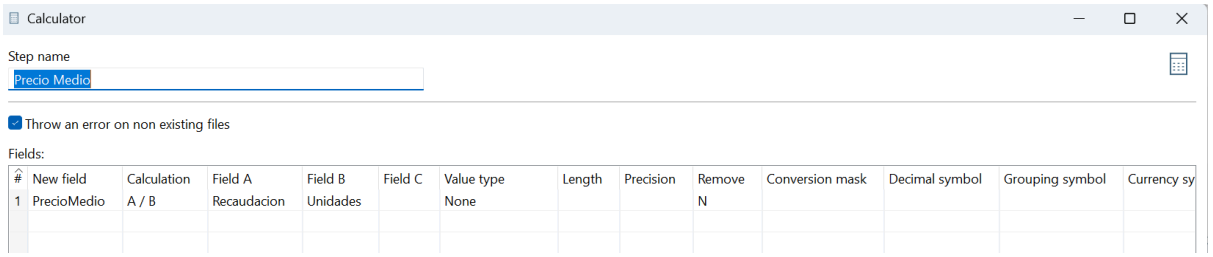
Después ordenamos por ManufacturerID



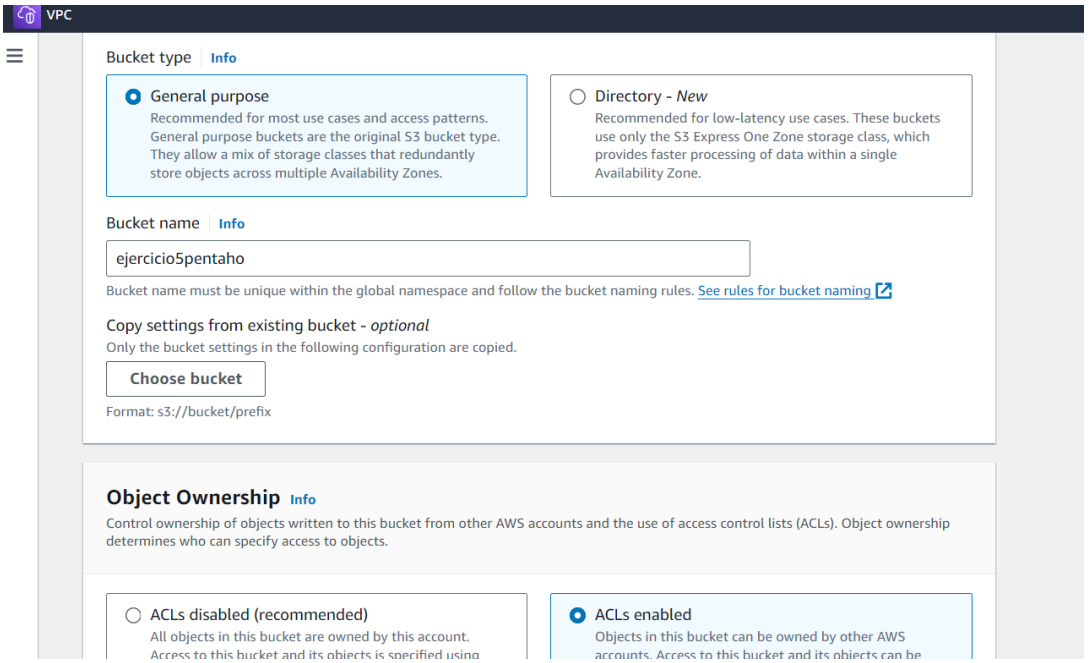
Cuando lo tengamos ordenado creamos el group by y lo unimos por Manufacturer ID

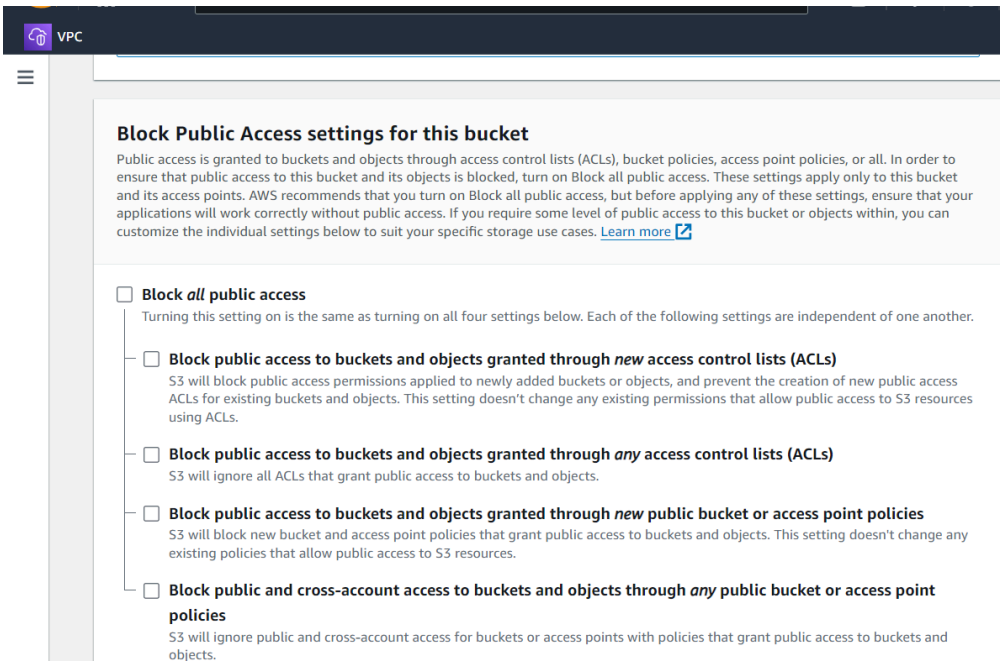


Ahora creamos un calculator para obtener el precio medio

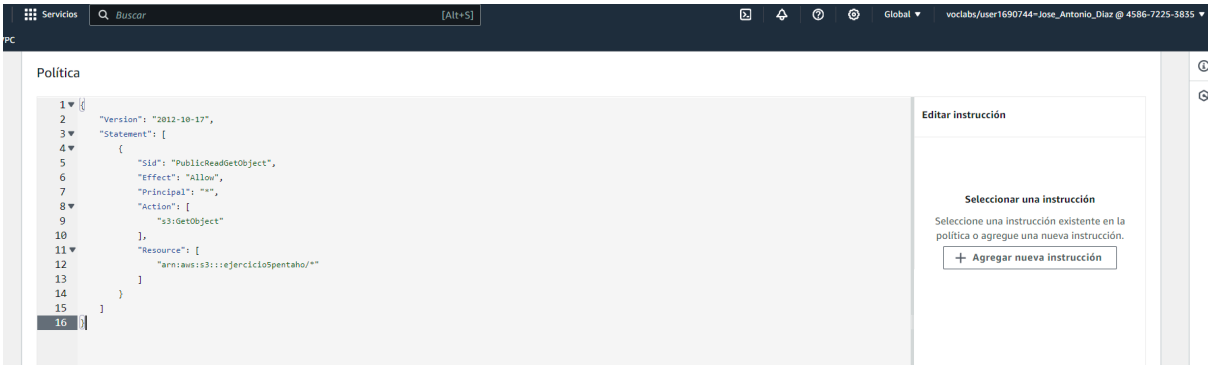


Ahora haremos la parte de configuración de S3:
Creamos un bucket




VPC
Block Public Access settings for this bucket
Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more
Block all public access
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.
Block public access to buckets and objects granted through new access control lists (ACLs)
S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
Block public access to buckets and objects granted through any access control lists (ACLs)
S3 will ignore all ACLs that grant public access to buckets and objects.
Block public access to buckets and objects granted through new public bucket or access point policies
S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
Block public and cross-account access to buckets and objects through any public bucket or access point policies
S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

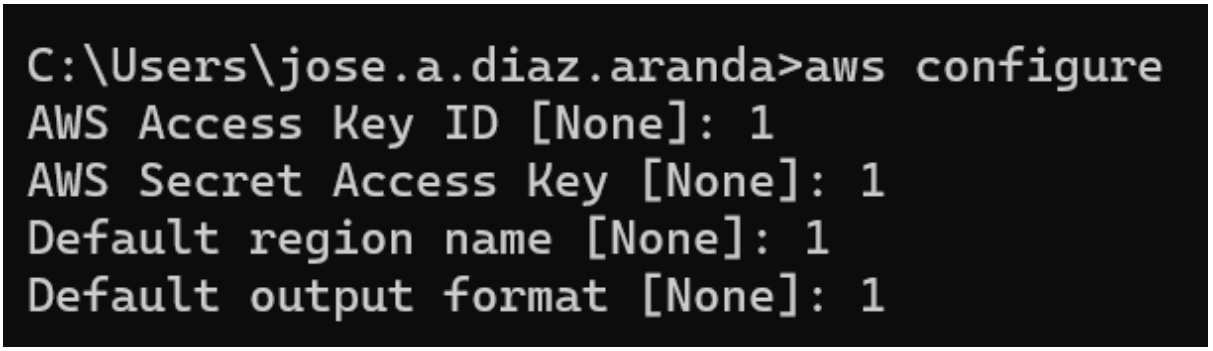
Metemos esta configuración en la política



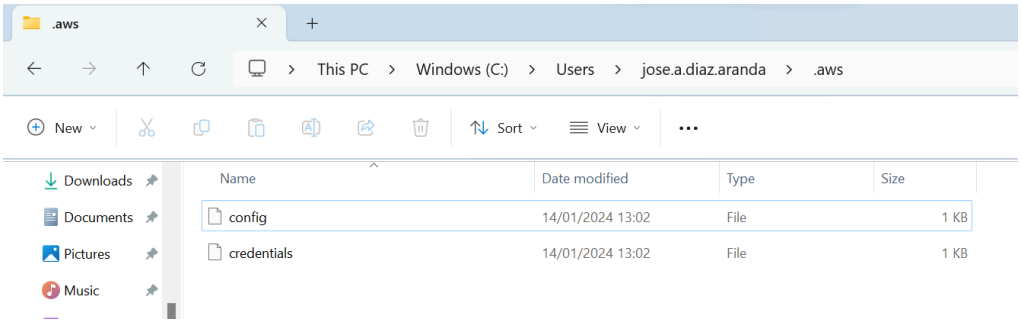
Instalamos awsCLI

https://awscli.amazonaws.com/AWSCLIV2.msi

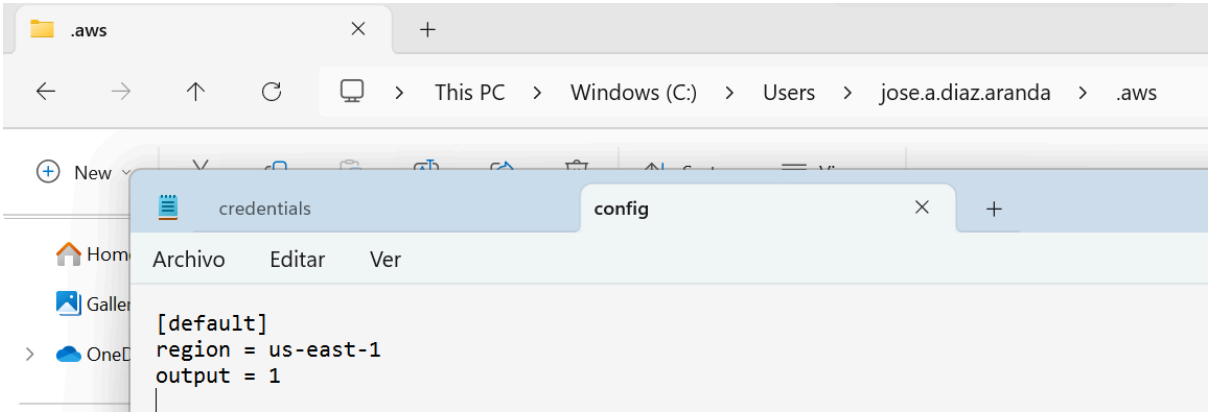
Ejecutamos este comando en la terminal



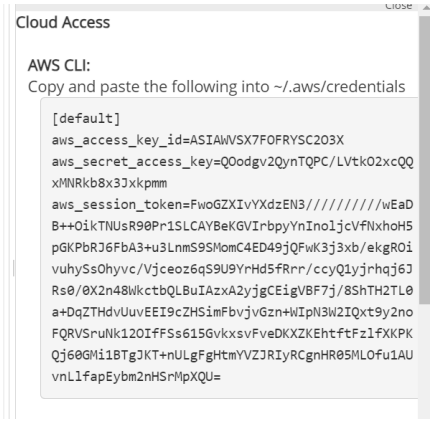
Se nos habrá creado esta carpeta



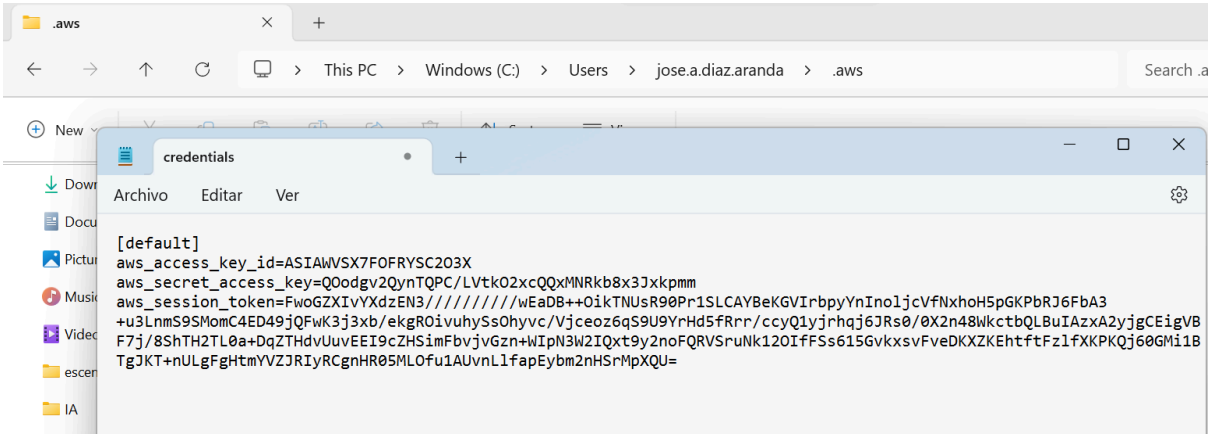
En config cambiamos la región



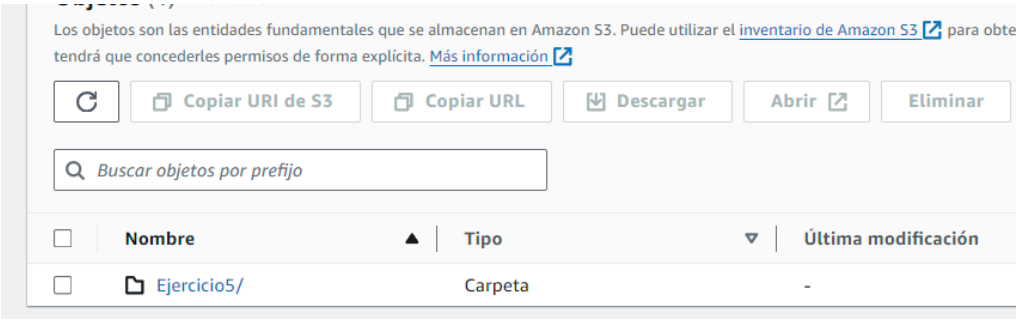
Ahora tenemos que introducir las credenciales que tenemos en la pestaña aws details de nuestro laboratorio



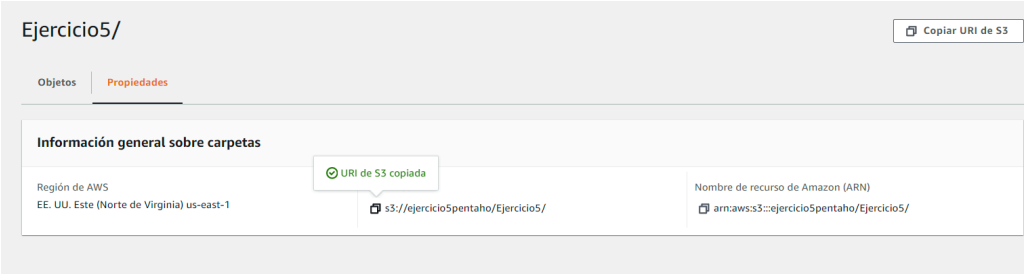
Lo introducimos en el fichero de credentials



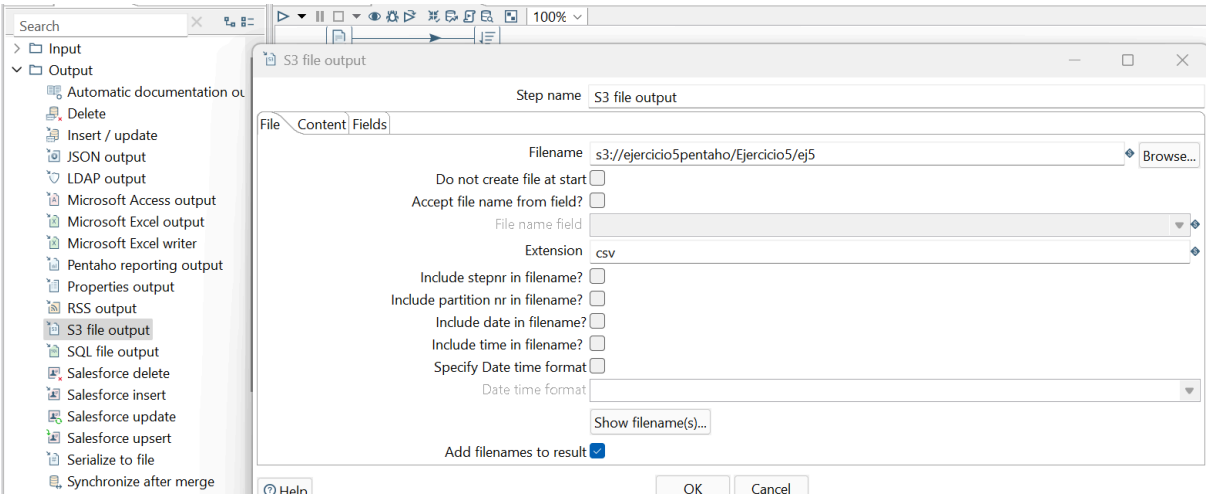
Creamos una carpeta en el bucket



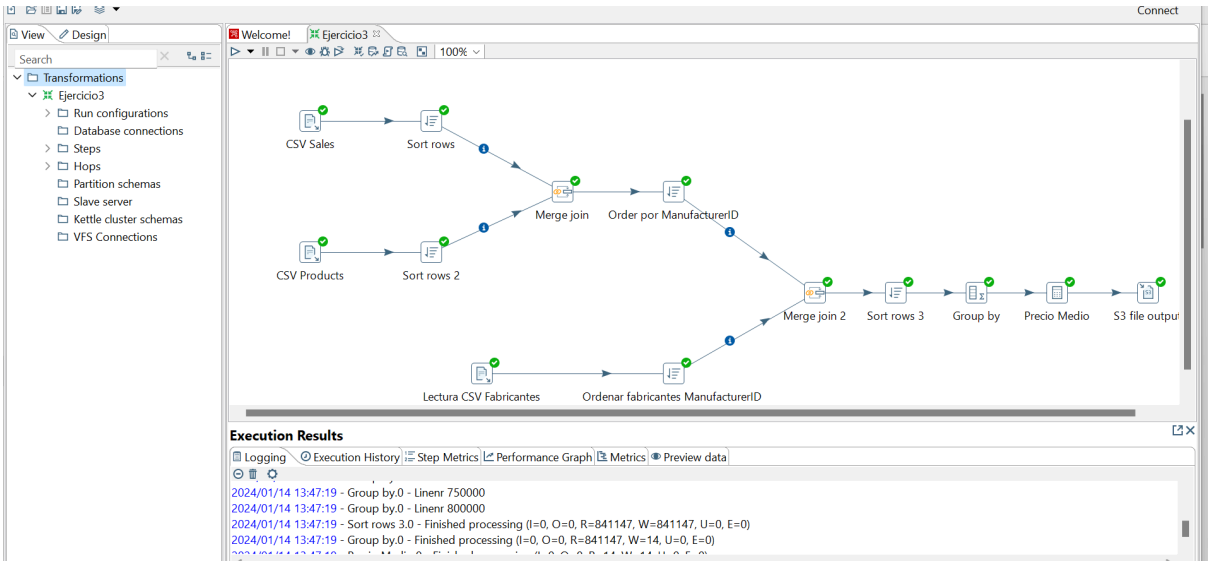
Ahora en propiedades copiamos el enlace para meterlo en el pentaho



Metemos en un fichero S3 output el código y ejecutamos



Lo ejecutamos



Vemos el resultado