



**BIG DATA  
ACADEMY**

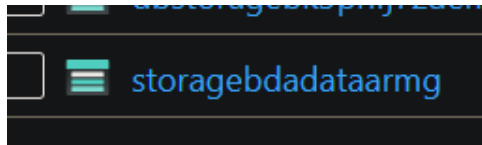
**LABORATORIO 48**  
**SOLUCIONES DE BIG DATA**  
**CON HDINSIGHT SOBRE**  
**SYNAPSE**

FORMADOR: ALONSO MELGAREJO  
alonsoraulmgs@gmail.com

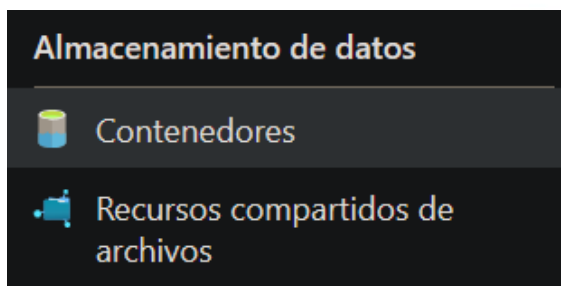
**LABORATORIO 48 - SOLUCIONES DE BIG DATA CON HDINSIGHT SOBRE SYNAPSE**

1. Desde el buscador de servicios, buscamos:

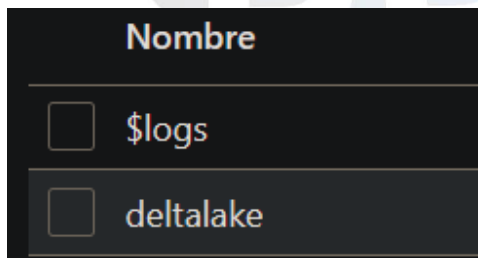
2. Seleccionamos la cuenta de almacenamiento "storagebdadataxxx"



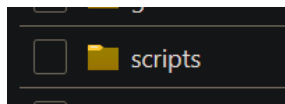
3. Desde la sección "Almacenamiento de datos", seleccionamos "Contenedores"



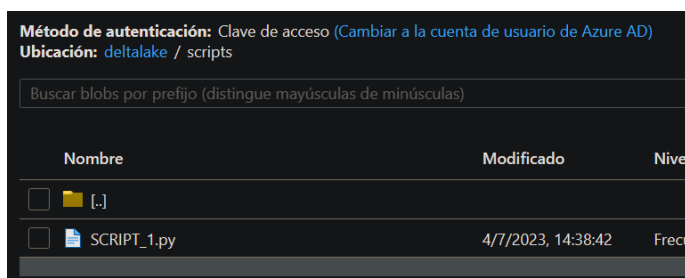
4. Entramos al blob storage "deltalake"



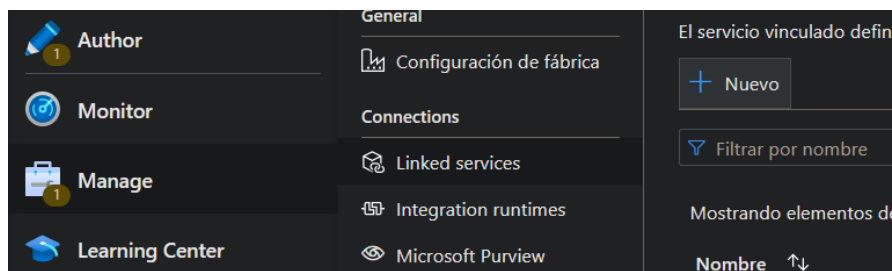
5. Creamos el directorio "scripts" y entramos en él



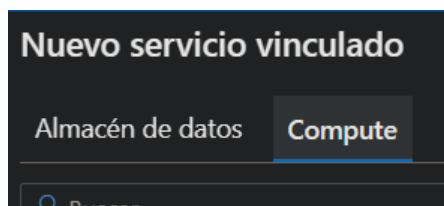
6. Dentro del directorio "scripts" subimos el script "SCRIPT\_1.py"



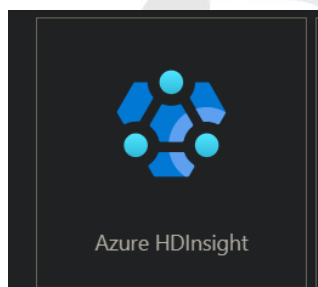
7. Crearemos un clúster HDInsight bajo demanda. Desde Synapse Studio, seleccionamos la opción “Manage / Linked services / + Nuevo”.



8. Seleccionamos la pestaña “Compute”



9. Seleccionamos la opción “Azure HDInsight” y damos clic en “Continuar”

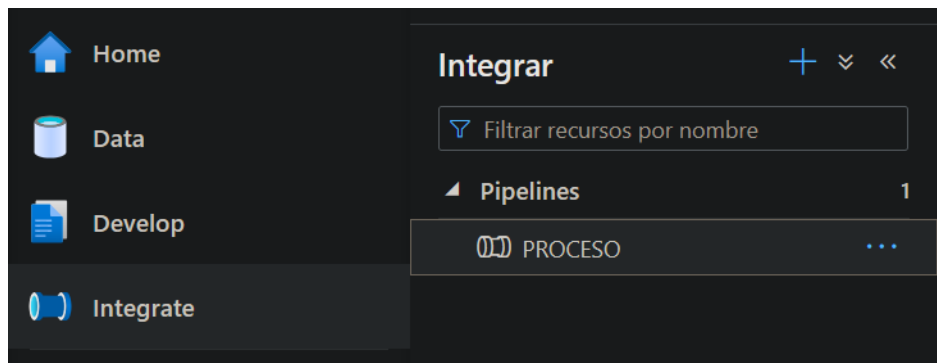


10. Configuramos:

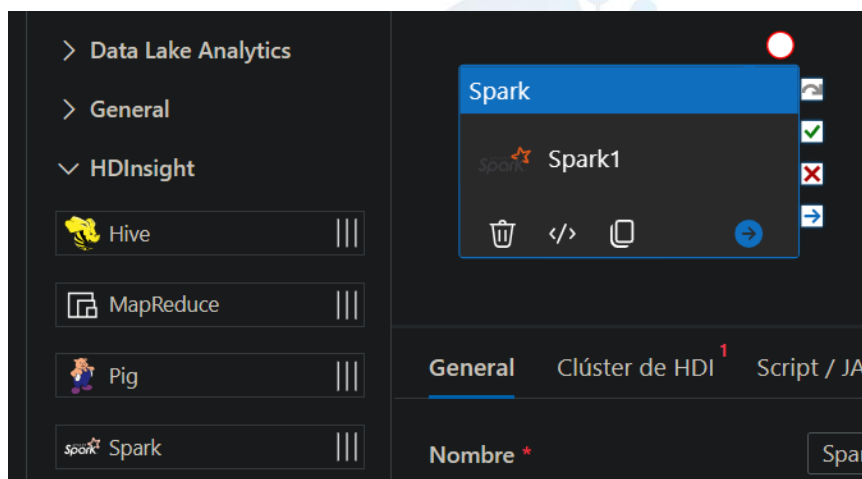
Nombre	LINK_HDINSIGHT
Tipo	HDInsight bajo demanda
Servicio vinculado de Azure Storage (desde dónde se leen los datos)	LINK_BLOBSTORAGE
Tipo de clúster	spark
Tamaño de grupo (cantidad de servidores del clúster)	1
Identificador de la entidad de servicio	clusterhdinsightXXX
Clave de la entidad de servicio (contraseña)	B!gd4t4aC4d3my
Seleccionar grupo de recursos	SESION06
Propiedades avanzadas / Tamaño del nodo de datos	Standard_DS3_v2

Damos clic en “Crear”

11. Ahora agregaremos el notebook al pipeline de ejecución, vamos a “Integrate” y en “Pipelines” damos clic sobre “PROCESO” para abrir el pipeline.



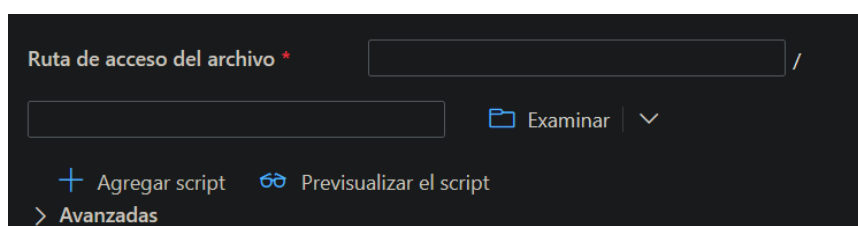
12. Desde el grupo de actividades “HDInsight” agregamos “Spark”



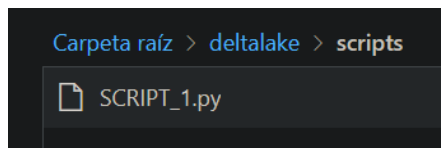
13. Configuramos:

Pestaña	Opción	Valor
General	Nombre	SCRIPT_SPARK_HDINSIGHT
Clúster de HDI	Servicio vinculado de HDInsight	LINK_HDINSIGHT
Script / JAR	Servicio vinculado de trabajo ( <b><u>blob storage donde está el script que se ejecutará</u></b> )	LINK_BLOBSTORAGE

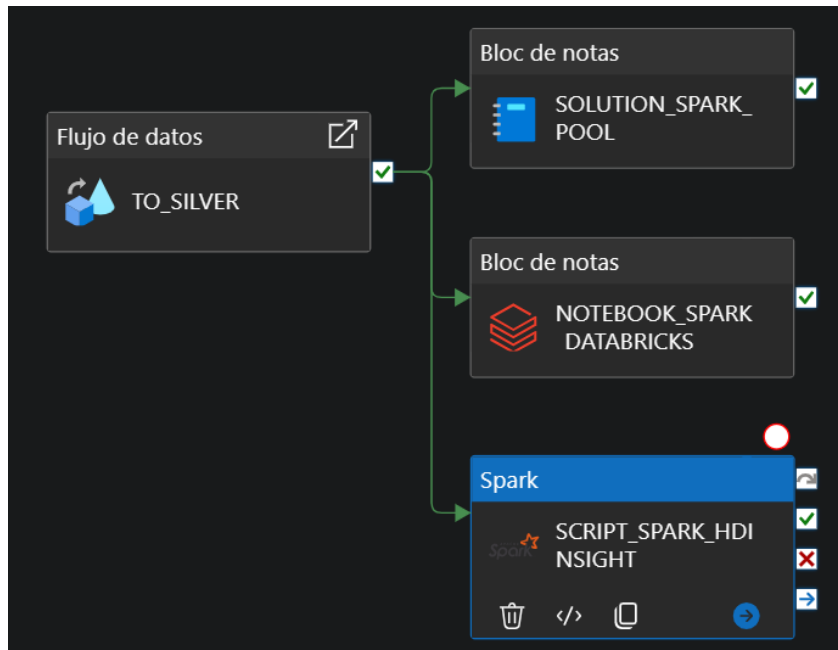
14. Buscaremos el script, damos clic en “Examinar”



15. Entramos a la ruta “/deltalake/scripts” y seleccionamos el script “SCRIPT\_1.py”, damos clic en “Aceptar”



16. Conectamos las actividades



17. Damos clic en “Validar todo” y luego en “Publicar todo”

