UNIR - Universidad Internacional de la Rioja

Ciudad de México

**ACTIVIDAD 1: Gobierno de Datos y Gestión de Datos con Herramientas de Microsoft**

Alumno: Cuenca Roa, Leonard Jose

Grupo: 1001

Equipo 01C

**ÍNDICE**

Objetivo 1 3

Instalación del software Dremio 3

Carga de Datos 4

Espacios de trabajo 7

Crear los datasets personalizados 8

Objetivo 1

Instalación del software Dremio

| Se procedió a instalar la herramienta Dremio utilizando docker, dado que se dispone de un equipo Mac para realizar las prácticas.  Para habilitar Dremio en modo escritorio, se ejecutaron los siguientes comandos en el siguiente orden:  docker pull dremio/dremio-oss  docker run -p 9047:9047 -p 31010:31010 -p 32010:32010 -p 45678:45678 dremio/dremio-oss  Se realizó la validación de la instalación usando la interfaz de docker para comprobar si el contenedor fue instalado y ejecutado apropiadamente. | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-03 a la(s) 21.48.26.png  Captura de pantalla 2025-01-03 a la(s) 21.48.44.png |
| Evidencia 2 | Captura de pantalla 2025-01-03 a la(s) 21.43.55.png |
| Evidencia 3 | Captura de pantalla 2025-01-03 a la(s) 21.43.07.png |

Carga de Datos

| Posteriormente, se accedió a la interfaz de Dremio a través del navegador web en la dirección:  <http://localhost:9047>  Con el propósito de cumplir con el objetivo 1, descrito en el enunciado, procederemos a cargar cada conjunto de datos en la plataforma Dremio. Una vez cargados, generaremos para cada uno una wiki, siguiendo las indicaciones del documento provisto | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-03 a la(s) 22.14.00.png |
| Evidencia 2  Wikipedia Books | Captura de pantalla 2025-01-03 a la(s) 22.15.27.png |
| Evidencia 3  Wikipedia Licencia 2021 | Captura de pantalla 2025-01-03 a la(s) 22.57.03.png |
| Evidencia 4  Wikipedia Locales 2021 | Captura de pantalla 2025-01-03 a la(s) 23.04.26.png |
| Evidencia 5  Wikipedia Terrazas | Captura de pantalla 2025-01-03 a la(s) 23.19.28.png |

Espacios de trabajo

| Con el fin de avanzar en el objetivo 1, se crearon los espacios de trabajo siguiendo las especificaciones del documento. Como prueba de ello, se adjunta la evidencia de la generación de Dremio | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-03 a la(s) 23.54.37.png |

| Se generó los wiki por cada espacio de trabajo Analista 1 | |
| --- | --- |
| Evidencia 2 | Captura de pantalla 2025-01-04 a la(s) 0.08.01.png |

| Se generó los wiki por cada espacio de trabajo Analista 2 | |
| --- | --- |
| Evidencia 3 | Captura de pantalla 2025-01-04 a la(s) 0.10.53.png |

| Se generó los wiki por cada espacio de trabajo Analista 3 | |
| --- | --- |
| Evidencia 4 | Captura de pantalla 2025-01-04 a la(s) 0.11.54.png |

Crear los datasets personalizados

| Se solicitó realizar una serie de modificaciones al dataset Terrazas\_202104. Estas incluyen eliminar todos los campos que comiencen con 'id\_' excepto 'id\_terraza', así como eliminar el campo 'Escalera'. Además, se debe crear un nuevo campo denominado 'Superficie\_TO', el cual será el resultado de sumar los valores de los campos 'Superficie\_ES' y ‘Superficie\_RA'. Finalmente, la consulta realizada debe guardarse con el nombre 'Terreza\_001' en el espacio de trabajo 'Analista 1’ | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-04 a la(s) 0.53.22.png |
| Evidencia 2 | Captura de pantalla 2025-01-04 a la(s) 0.58.31.png |

| Se solicitó modificar el dataset "Licencias\_Locales\_202104" eliminando todas las columnas, excepto "id\_local", "ref\_licencia", "desc\_tipo\_licencia", "desc\_tipo\_situacion\_licencia" y "fecha\_dec\_lic". Una vez realizada esta limpieza, se debe guardar el nuevo dataset con el nombre "Licencias\_002" en la ubicación de trabajo correspondiente al usuario "Analista 1’ | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-04 a la(s) 1.11.38.png |
| Evidencia 2 | Captura de pantalla 2025-01-04 a la(s) 1.12.16.png |

| Se requiere realizar un análisis de datos combinando dos conjuntos de información. El primer paso es acceder al conjunto de datos "Terrazas\_202104". A continuación, se debe vincular este conjunto con otro denominado "Licencia\_002". La conexión entre ambos se establecerá utilizando el campo común "id\_local", mediante un tipo de unión interna. El resultado de esta operación debe almacenarse como un nuevo conjunto de datos llamado "Licencias\_Terrazas\_003" en el espacio de trabajo asignado al "Analista 2" | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-04 a la(s) 1.22.53.png |
| Evidencia 2 | Captura de pantalla 2025-01-04 a la(s) 1.23.27.png |

| Se solicita modificar un conjunto de datos llamado "books" siguiendo estos pasos: En primer lugar, se elimina un campo innecesario llamado "\_id". Luego, se filtran los datos para excluir todos los libros que no tengan un número ISBN. A continuación, se aplica una transformación llamada "unnest" a las columnas "authors" y "categories", lo que probablemente implica descomponer listas o arreglos de autores y categorías en filas individuales para facilitar su análisis. Finalmente, se guarda este nuevo conjunto de datos modificado con el nombre "Books\_001" en un área de trabajo específica llamada "Analista 3". En resumen, se está preparando un dataset más limpio y estructurado para realizar análisis posteriores, aplicando los unnest o para este caso un flaten podemos realizar búsquedas mas rápidas separando ambos campos en registros únicos por usuario y categorías. | |
| --- | --- |
| Evidencia 1 | Captura de pantalla 2025-01-04 a la(s) 1.46.24.png |
| Evidencia 2 | Captura de pantalla 2025-01-04 a la(s) 1.47.25.png |