

## Paso a paso Integración con Google Cloud Platform

A continuación, se relacionan los pasos que se realizan con el fin de integrar los archivos que actualmente Comercial Química genera, con un proveedor de la nube en este caso, Google Cloud Platform ofrecido por Google.

Las arquitecturas utilizadas para este procesamiento son en su orden:

- **Dataiku:** Esta es una plataforma de colaboración y orquestación de datos que se utiliza en el campo de la ciencia de datos y el análisis de datos. La plataforma está diseñada para facilitar la colaboración entre equipos de datos, científicos de datos, analistas y otros profesionales involucrados en el ciclo de vida de los datos. En este caso en particular, permite crear flujos que son útiles para la cointegración de diversas fuentes de datos y la optimización de procesos.
- **Google Cloud Platform (GCP):** Es la plataforma en la nube proporcionada por Google. Dicho servicio ofrece una variedad de servicios de infraestructura en la nube, herramientas de desarrollo, servicios de análisis de datos, aprendizaje automático, entre otros. Para este caso en particular proporciona la capacidad de alojar fuentes de datos en datasets productivos con el fin de conectar estos a diversas fuentes y aplicaciones.
- **Power BI:** Es una suite de herramientas de análisis empresarial desarrollada por Microsoft. Está diseñada para permitir a los usuarios visualizar sus datos y compartir información y conocimientos a través de paneles interactivos e informes. En este ámbito es usado como producto de datos final con el fin de que Comercial Química pueda utilizarlo para tomar decisiones de manera estratégica.

En ese sentido, en este apartado, se describe el paso a paso de cada uno de los componentes para lograr esta integración:

### Paso 1: Creación de una zona en Dataiku

En primer, lugar a través de la interfaz de Dataiku creamos un proyecto y una zona en este caso la hemos llamado **“Proyecto Comercial Química”**. Esta zona nos permitirá guardar todos los tipos de tablas y modelos de datos necesarios para realizar la automatización del proceso.

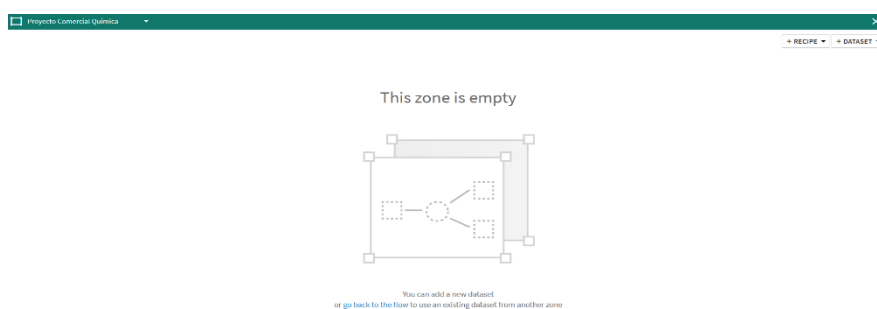


Imagen 1: Zona de procesamiento en Dataiku

## Paso 2: Agregación de los archivos en la Zona previamente creada

En dicha zona agregamos los archivos que a este nivel ya deben estar procesados y limpiados con el fin de ser utilizados para realizar la integración con el Dashboard. Los archivos que se han cargado tienen la siguiente descripción:

- **Data\_original:** Contiene la información consolidada de los productos vendidos por Comercial Química a cada uno de los proveedores, con respecto a las cantidades y los precios de referencia. Asimismo, el almacenamiento de estos por Bodega, así como otras variables relevantes que hacen referencia al país de importación y la ciudad donde se encuentra el proveedor.
- **Data\_inventarios:** Contiene la información del inventario disponible para cada una de las materias primas, asimismo, relacionado si dicha cantidad hace referencia a una compra, una venta o es stock.
- **Ubicaciones:** Contiene la información de la latitud y la longitud de los proveedores con el fin de que sea posible georreferenciarlos en un mapa.
- **IR\_calculado:** Contiene la base de datos con el cálculo del índice de rotación para cada uno de los productos para los trimestres de cada año.

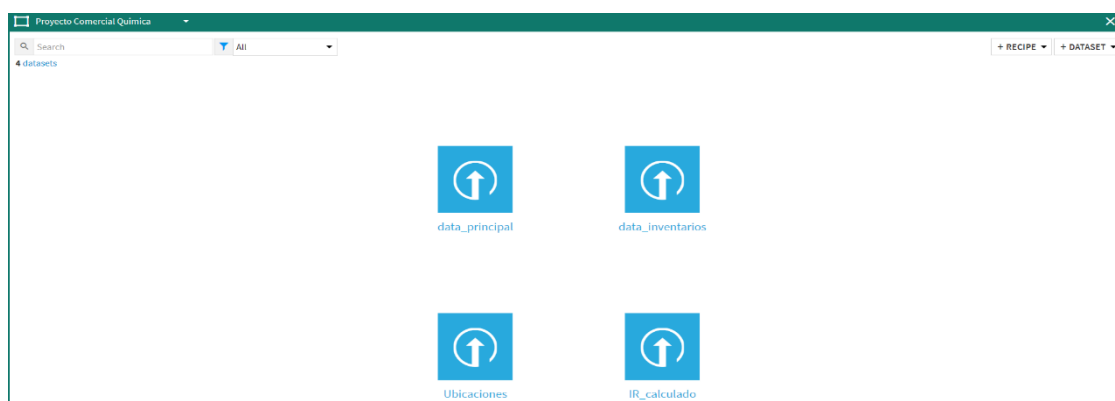


Imagen 2: Cargue de archivos en la zona de procesamiento

## Paso 3: Sincronización de los archivos con el ambiente de Dataiku

Se debe realizar una sincronización de los archivos cargados con la plataforma para que estos puedan ser procesados de manera exitosa por el proveedor de la nube.

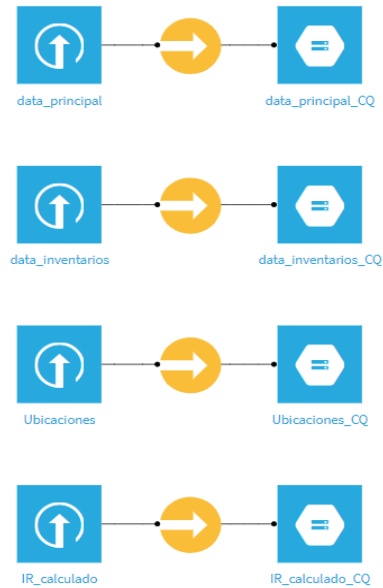


Imagen 3: Sincronización con el ambiente de Dataiku

#### Paso 4: Sincronización de las tablas con Google Cloud Platform

Se realiza el proceso de sincronización para cada una de las fuentes de datos directamente con el sandbox con el que cuenta la organización en este caso un almacenamiento muy pequeño debido a que es una empresa con limitaciones a nivel tecnológico.

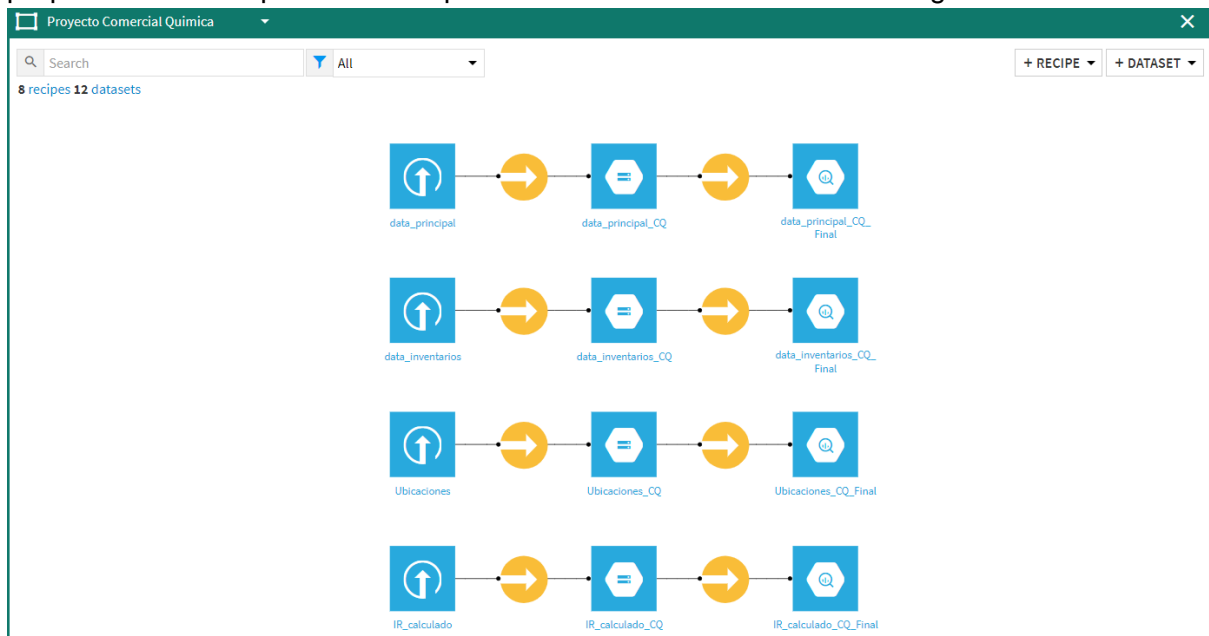


Imagen 4: Sincronización con el ambiente de Google Cloud Platform

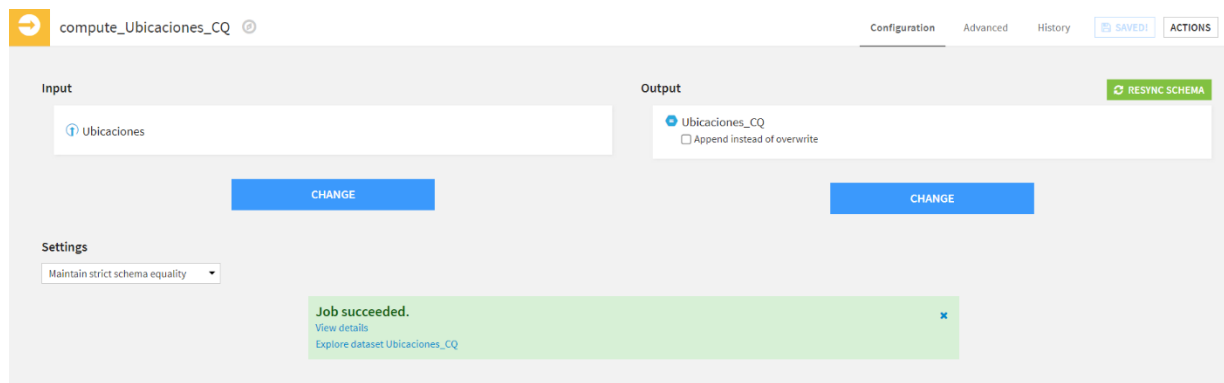


Imagen 5: Ejemplo de funcionamiento de una sincronización en Dataiku

## Paso 5: Validación de las fuentes de datos en Google Cloud Platform

Las tablas se han cargado sin problemas en nuestra nube gracias a Google Cloud Platform, y ya están disponibles para su uso en Power BI con sus conexiones respectivas. No obstante, también son compatibles con consultas en SQL.

Entonces, aunque están listas para ser usadas en Power BI, también puedes obtener respuestas sólidas usando SQL.

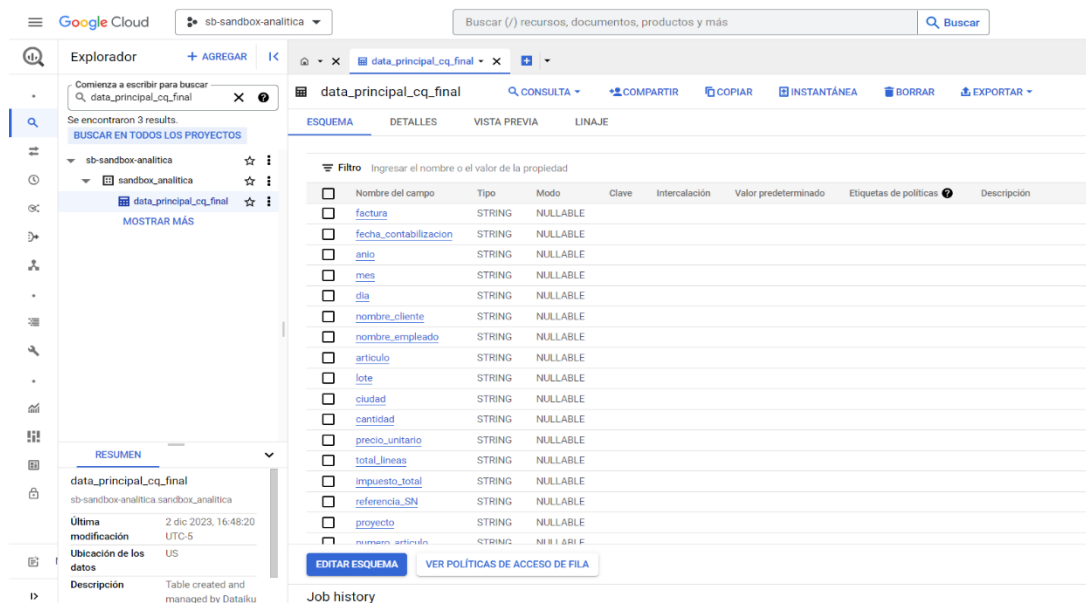


Imagen 6: Esquema de la fuente de datos principal que hemos denominado con el nombre de "data\_principal\_cq\_final"

Google Cloud sb-sandbox-analitica

Buscar (/) recursos, documentos, productos y más

Explorador

data\_inventarios\_cq\_final

Se encontraron 3 results.

sb-sandbox-analitica

sandbox\_analitica

data\_inventarios\_cq\_final

RESUMEN

data\_inventarios\_cq\_final

sb-sandbox-analitica.sandbox\_analitica

Última modificación 2 dic 2023, 16:50:19 UTC-5

Ubicación de los datos US

Descripción Table created and managed by Dataiku

data\_inventarios\_cq\_final

ESQUEMA

Filtro

Nombre del campo	Tipo	Modo	Clave	Intercalación	Valor predeterminado	Etiquetas de políticas	Descripción
DESCRIPCIÓN DEL ARTICULO	STRING	NULLABLE					
CANTIDAD	STRING	NULLABLE					
MES	STRING	NULLABLE					
TIPO	STRING	NULLABLE					
AÑO	STRING	NULLABLE					

EDITAR ESQUEMA

VER POLÍTICAS DE ACCESO DE FILA

Job history

Imagen 7: Esquema de la fuente de datos de inventarios que hemos denominado con el nombre de “data\_inventarios\_cq\_final”

Google Cloud sb-sandbox-analitica

Buscar (/) recursos, documentos, productos y más

Explorador

ubicaciones\_cq\_final

Se encontraron 3 results.

sb-sandbox-analitica

sandbox\_analitica

ubicaciones\_cq\_final

RESUMEN

ubicaciones\_cq\_final

sb-sandbox-analitica.sandbox\_analitica

Última modificación 2 dic 2023, 16:52:06 UTC-5

Ubicación de los datos US

Descripción Table created and managed by Dataiku

ubicaciones\_cq\_final

ESQUEMA

Filtro

Nombre del campo	Tipo	Modo	Clave	Intercalación	Valor predeterminado	Etiquetas de políticas	Descripción
CIUDAD	STRING	NULLABLE					
LAT_LON	STRING	NULLABLE					
LATITUD	STRING	NULLABLE					
LONGITUD	STRING	NULLABLE					

EDITAR ESQUEMA

VER POLÍTICAS DE ACCESO DE FILA

Job history

Imagen 8: Esquema de la fuente de datos de ubicaciones de los proveedores que hemos denominado con el nombre de “ubicaciones\_cq\_final”

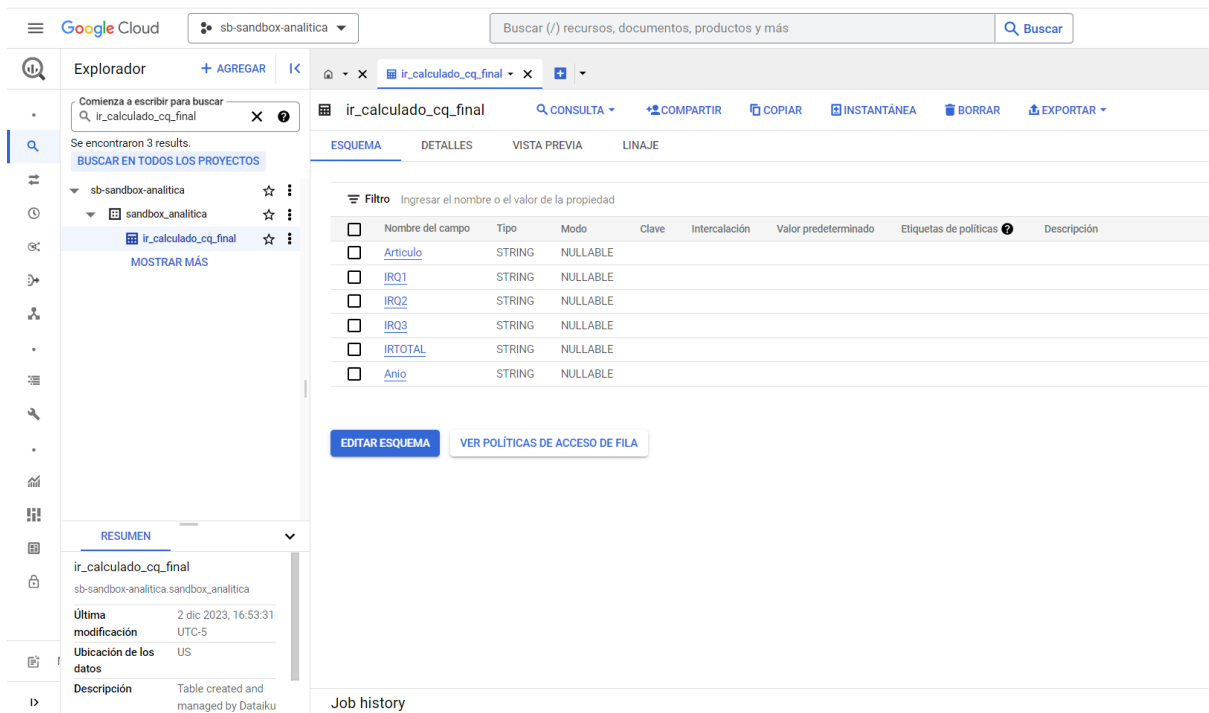


Imagen 9: Esquema de la fuente de datos de ubicaciones de los proveedores que hemos denominado con el nombre de "ir\_calculado\_cq\_final"

## Paso 6: Conectando Power BI a Nuestras Fuentes en Google Cloud Platform

Ahora que hemos asegurado nuestras fuentes de datos en Google Cloud Platform, el siguiente paso es establecer la conexión desde Power BI. Este proceso nos permitirá leer los datos utilizando la fuente que previamente hemos configurado, asegurando así una integración fluida y eficiente.

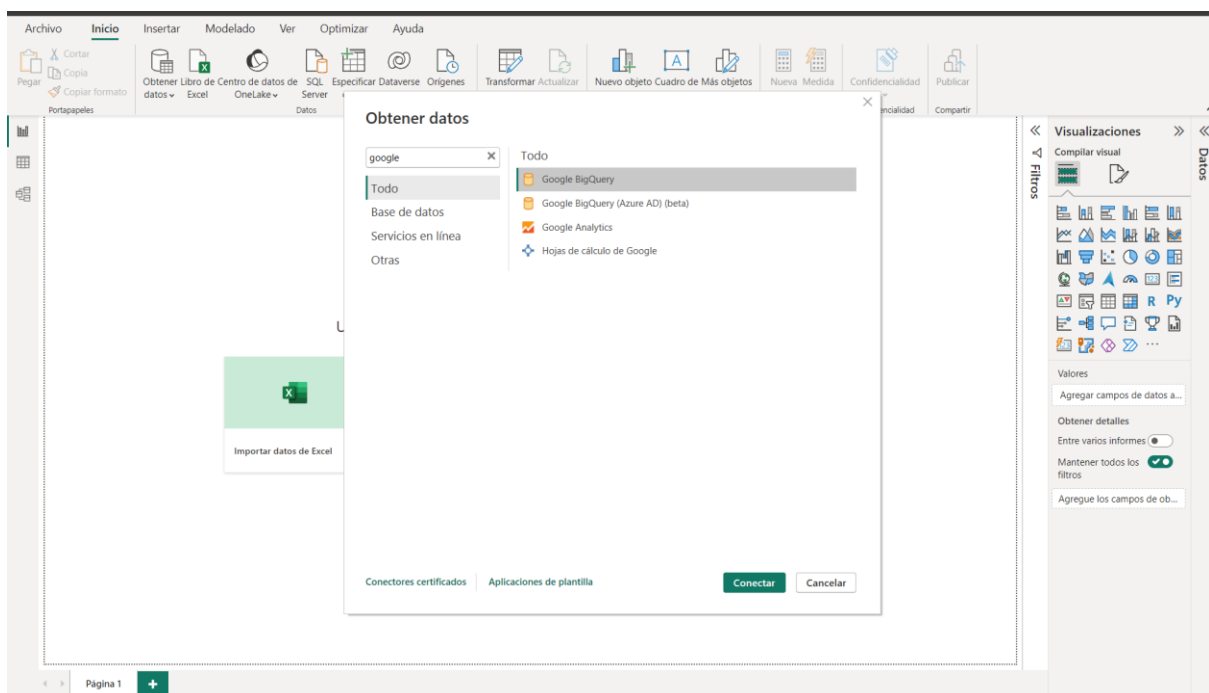
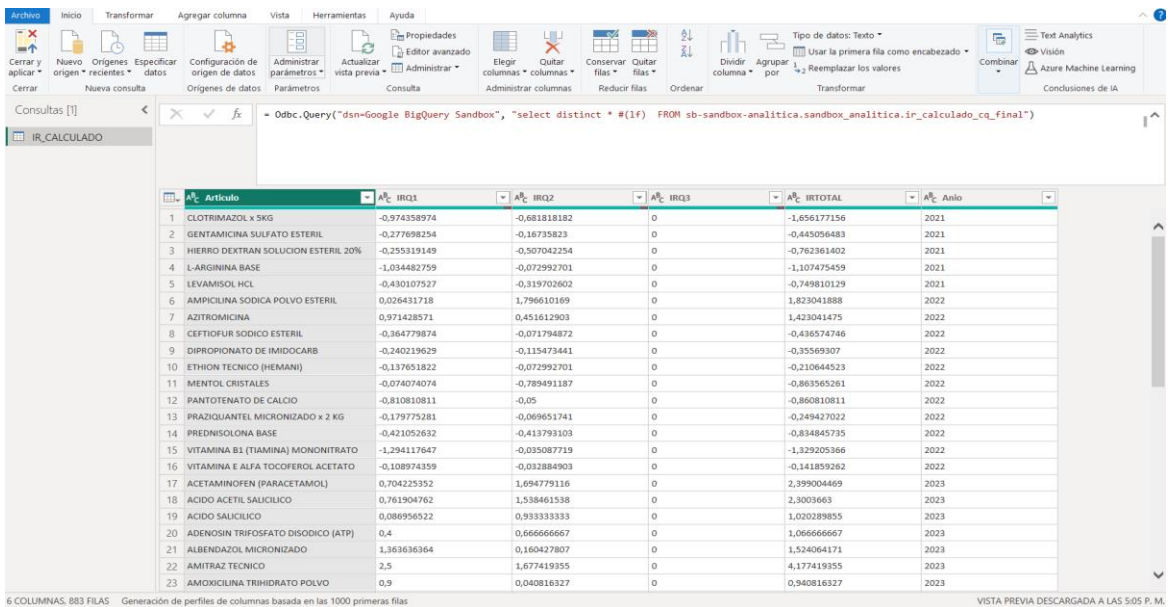


Imagen 10: Conexión a Big Query de Google Cloud Platform

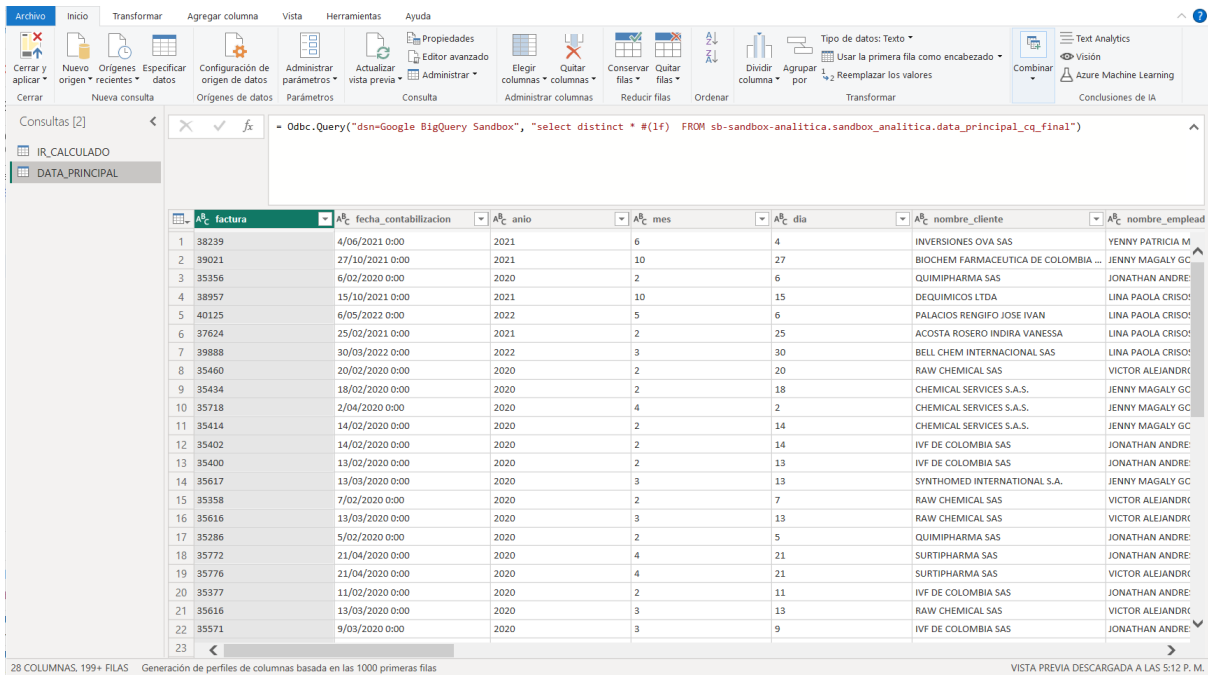
## Paso 7: Leer las tablas desde Google Cloud Platform con ODBC

Luego de dicha conexión leemos en el editor de Power Query las tablas que se encuentran alojadas en GCP utilizando una conexión ODBC. Realizamos dicho proceso para las cuatro tablas mencionadas previamente.



Artículo	IRQ1	IRQ2	IRQ3	IRTOTAL	Año
1 CLOTRIMAZOL x 5KG	-0,974358974	-0,681818182	0	-1,656177156	2021
2 GENTAMICINA SULFATO ESTERIL	-0,277698254	-0,1673583	0	-0,445056483	2021
3 HIERRO DEXTRAN SOLUCION ESTERIL 20%	-0,255319149	-0,507042254	0	-0,762361402	2021
4 L-ARGININA BASE	-1,034482759	-0,072992701	0	-1,107475459	2021
5 LEVAMISOL HCL	-0,430107527	-0,319702802	0	-0,749810129	2021
6 AMPICILINA SODICA POLVO ESTERIL	0,026431718	1,796610169	0	1,823041888	2022
7 AZITROMICINA	0,971428571	0,451612903	0	1,423041475	2022
8 CEFTIOFUR SODICO ESTERIL	-0,164779874	-0,071794872	0	-0,436574746	2022
9 DIPROPIONATO DE IMIDOCARB	-0,240219629	-0,115473441	0	-0,35569307	2022
10 ETHION TECNICO (HEMAN)	-0,137651822	-0,072992701	0	-0,210644523	2022
11 MENTOL CRISTALES	-0,074074074	-0,789491187	0	-0,863565261	2022
12 PANTOTENATO DE CALCIO	-0,810810811	-0,05	0	-0,860810811	2022
13 PRAZIQUANTEL MICRONIZADO x 2 KG	-0,179775281	-0,069651741	0	-0,249427022	2022
14 PREDNISOLONA BASE	-0,421052632	-0,413793103	0	-0,834845735	2022
15 VITAMINA B1 (TIAMINA) MONONITRATO	-1,294117647	-0,035087719	0	-1,329205366	2022
16 VITAMINA E ALFA TOCOFEROL ACETATO	-0,108974359	-0,032884903	0	-0,141859262	2022
17 ACETAMINOFEN (PARACETAMOL)	0,704225352	1,694779116	0	2,399004469	2023
18 ACIDO ACETIL SALICILICO	0,761904762	1,538461538	0	2,3003663	2023
19 ACIDO SALICILICO	0,086956522	0,933333333	0	1,020289855	2023
20 ADENOSIN TRIFOSFATO DISODICO (ATP)	0,4	0,666666667	0	1,066666667	2023
21 ALBENDAZOL MICRONIZADO	1,363636364	1,60427807	0	1,524064171	2023
22 AMITRAZ TECNICO	2,5	1,677419355	0	4,177419355	2023
23 AMOXICILINA TRIHIDRATO POLVO	0,9	0,040816327	0	0,940816327	2023

Imagen 11: Conexión Fuente de datos del índice de rotación que llamaremos **IR\_CALCULADO** en nuestro flujo de datos de Power BI.



factura	fecha_contabilizacion	año	mes	día	nombre_cliente	nombre_empleado
1 38239	4/06/2021 0:00	2021	6	4	INVERSIONES OVA SAS	YENNY PATRICIA M
2 39021	27/10/2021 0:00	2021	10	27	BIOCHEM FARMACEUTICA DE COLOMBIA ...	JENNY MAGALY GC
3 35356	6/02/2020 0:00	2020	2	6	QUIMIPHARMA SAS	JONATHAN ANDRE
4 38957	15/10/2021 0:00	2021	10	15	DEQUIMICOS LTDA	LINA PAOLA CRISO
5 40125	6/05/2022 0:00	2022	5	6	PALACIOS RENGIFO JOSE IVAN	LINA PAOLA CRISO
6 37624	25/02/2021 0:00	2021	2	25	ACOSTA ROSERO INDIRA VANESSA	LINA PAOLA CRISO
7 39888	30/03/2022 0:00	2022	3	30	BELL CHEM INTERNACIONAL SAS	LINA PAOLA CRISO
8 35460	20/02/2020 0:00	2020	2	20	RAW CHEMICAL SAS	VICTOR ALEJANDRO
9 35434	18/02/2020 0:00	2020	2	18	CHEMICAL SERVICES S.A.S.	JENNY MAGALY GC
10 35718	2/04/2020 0:00	2020	4	2	CHEMICAL SERVICES S.A.S.	JENNY MAGALY GC
11 35414	14/02/2020 0:00	2020	2	14	CHEMICAL SERVICES S.A.S.	JENNY MAGALY GC
12 35402	14/02/2020 0:00	2020	2	14	IVF DE COLOMBIA SAS	JONATHAN ANDRE
13 35400	13/02/2020 0:00	2020	2	13	IVF DE COLOMBIA SAS	JONATHAN ANDRE
14 35617	13/03/2020 0:00	2020	3	13	SYNTHOMED INTERNATIONAL S.A.	JENNY MAGALY GC
15 35358	7/02/2020 0:00	2020	2	7	RAW CHEMICAL SAS	VICTOR ALEJANDRO
16 35616	13/03/2020 0:00	2020	3	13	RAW CHEMICAL SAS	VICTOR ALEJANDRO
17 35286	5/02/2020 0:00	2020	2	5	QUIMIPHARMA SAS	JONATHAN ANDRE
18 35772	21/04/2020 0:00	2020	4	21	SURTI PHARMA SAS	JONATHAN ANDRE
19 35776	21/04/2020 0:00	2020	4	21	SURTI PHARMA SAS	VICTOR ALEJANDRO
20 35377	11/02/2020 0:00	2020	2	11	IVF DE COLOMBIA SAS	JONATHAN ANDRE
21 35616	13/03/2020 0:00	2020	3	13	RAW CHEMICAL SAS	VICTOR ALEJANDRO
22 35571	9/03/2020 0:00	2020	3	9	IVF DE COLOMBIA SAS	JONATHAN ANDRE

Imagen 12: Conexión Fuente de datos principales que llamaremos **DATA\_PRINCIPAL** en nuestro flujo de datos de Power BI.

Consultas [3]

IR\_CALCULADO  
DATA\_PRINCIPAL  
DATA\_INVENTARIOS

Odbc.Query("dsn=Google BigQuery Sandbox", "select distinct \* #(if FROM sb-sandbox-analitica.sandbox\_analitica.data\_inventarios\_cq\_final")

A#	DESCRIPCIÓN DEL ARTICULO	A#	CANTIDAD	A#	MES	A#	TIPO	A#	AÑO
1	ACETAMINOFEN (PARACETAMOL)	0			NOVIEMBRE		COMPRAS		2020
2	ACIDO ACETIL SALICILICO	0			NOVIEMBRE		COMPRAS		2020
3	ACIDO FOLICO	0			NOVIEMBRE		COMPRAS		2020
4	ACIDO SALICILICO	0			NOVIEMBRE		COMPRAS		2020
5	ADENOSIN TRIFOSFATO DISODICO (ATP)	0			NOVIEMBRE		COMPRAS		2020
6	ALBENDAZOL MICRONIZADO	0			NOVIEMBRE		COMPRAS		2020
7	ALGINATO DE SODIO	0			NOVIEMBRE		COMPRAS		2020
8	AMITRAZ TECNICO	0			NOVIEMBRE		COMPRAS		2020
9	AMOXICILINA TRIHIDRATO COMPACTADA	0			NOVIEMBRE		COMPRAS		2020
10	AMPICILINA SODICA POLVO ESTERIL	0			NOVIEMBRE		COMPRAS		2020
11	AMPICILINA TRIHIDRATO MICRONIZADA	0			NOVIEMBRE		COMPRAS		2020
12	AMPICILINA TRIHIDRATO POLVO	0			NOVIEMBRE		COMPRAS		2020
13	AMPROLIO HCL	0			NOVIEMBRE		COMPRAS		2020
14	ANTIPIRINA	0			NOVIEMBRE		COMPRAS		2020
15	ATROPINA SULFATO	0			NOVIEMBRE		COMPRAS		2020
16	AZITROMICINA	0			NOVIEMBRE		COMPRAS		2020
17	BENZOCAINA	0			NOVIEMBRE		COMPRAS		2020
18	BIOTINA 2%	0			NOVIEMBRE		COMPRAS		2020
19	BOLDENONA UNDECILENATO	0			NOVIEMBRE		COMPRAS		2020
20	BROMHEXINA HCL	0			NOVIEMBRE		COMPRAS		2020
21	BUTOXIDO DE PIPERONILO	0			NOVIEMBRE		COMPRAS		2020
22	CACODILATO DE SODIO	0			NOVIEMBRE		COMPRAS		2020
23	CAFEINA ANHIDRA	0			NOVIEMBRE		COMPRAS		2020

ESPERANDO DSN=GOOGLE BIGQUERY SANDBOX.

Imagen 13: Conexión Fuente de datos de inventarios que llamaremos **DATA\_INVENTARIOS** en nuestro flujo de datos de Power BI.

Consultas [4]

IR\_CALCULADO  
DATA\_PRINCIPAL  
DATA\_INVENTARIOS  
UBICACIONES

Odbc.Query("dsn=Google BigQuery Sandbox", "select distinct \* #(if FROM sb-sandbox-analitica.sandbox\_analitica.ubicaciones\_cq\_final")

A#	CUIDAD	A#	IAT_LON	A#	LATITUD	A#	LONGITUD
1	VILLAVICENCIO	4.142, -73.62664	4,142		-73,62664		
2	COTA	4.80938, -74.098	4,80938		-74,098		
3	SOPO	4.9075, -73.9384	4,9075		-73,9384		
4	JAMUNDI	3.26289, -76.5384	3,26289		-76,5384		
5	CALI	3.43722, -76.5225	3,43722		-76,5225		
6	VALLE	3.43722, -76.5225	3,43722		-76,5225		
7	BUENAVENTURA	3.8801, -77.03116	3,8801		-77,03116		
8	TENIO	4.8727, -74.14435	4,8727		-74,14435		
9	MISTRATO	5.29622, -75.8839	5,29622		-75,8839		
10	GIRON	7.0682, -73.16981	7,0682		-73,16981		
11	BUCARAMANGA	7.12539, -73.1198	7,12539		-73,1198		
12	SANTANDER	7.12539, -73.1198	7,12539		-73,1198		
13	SAN PEDRO	9.3956, -75.06476	9,3956		-75,06476		
14	IPIALES	0.82501, -77.63966	0,82501		-77,63966		
15	POPAYAN	2.43823, -76.61316	2,43823		-76,61316		
16	PALMIRA	3.53944, -76.30361	3,53944		-76,30361		
17	YUMBO	3.58234, -76.49146	3,58234		-76,49146		
18	BUGA	3.90089, -76.29783	3,90089		-76,29783		
19	SAN PEDRO (Valle del Cauca)	3.99445, -76.22885	3,99445		-76,22885		
20	ESPINAL	4.14924, -74.88429	4,14924		-74,88429		
21	FUSAGASUGA	4.33646, -74.36378	4,33646		-74,36378		
22	TOLIMA	4.43889, -75.23222	4,43889		-75,23222		
23	IBAGUE	4.43889, -75.23222	4,43889		-75,23222		

4 COLUMNAS, 56 FILAS Generación de perfiles de columnas basada en las 1000 primeras filas

VISTA PREVIA DESCARGADA A LAS 5:16 P. M.

Imagen 14: Conexión Fuente de datos principales que llamaremos **UBICACIONES** en nuestro flujo de datos de Power BI.

## Paso 8: Validación de que las tablas se ajusten al modelo de datos actual de Power BI

Después de realizar la carga de los datos que se utilizarán en el flujo de Power BI verificamos que los gráficos y funciones se encuentran de manera adecuada y funcionando sin inconvenientes.





Imagen 15: Verificación de la carga de las fuentes en el flujo de datos de Power BI

## Paso 9: Diseñando un Escenario Automático en Dataiku

Después de leer exitosamente los archivos mediante Google BigQuery y confirmar que nuestro panel de control funciona sin problemas, el siguiente paso consiste en establecer un escenario automatizado en Dataiku. Este escenario se encargará de mantener actualizadas las tablas de manera automática. De esta manera, el cliente simplemente necesitará hacer clic en "actualizar" en el panel de control para asegurarse de que siempre tenga acceso a la información más reciente.

En este escenario lo que realizamos son los siguientes pasos intermedios:

1. Definimos que el escenario se corra de manera automática de tal manera que el cliente no debe preocuparse por la actualización de las fuentes dentro de la interfaz esto se realiza habilitando la función de **"Auto-triggers"**
2. Definimos que la manera en la que se debe correr dicho flujo es de manera recursiva es decir a partir de la última iteración hacia atrás, lo que sincronizará los datos con Google Cloud Platform como mapeará las fuentes de datos correctas desde el repositorio local.
3. Indicamos una periodicidad de actualización de la fuente en este caso es diaria, al igual que la hora en la que queremos que dichas fuentes se actualicen siendo la hora ideal las 4 a.m y finalmente indicamos que el horario que estamos usando es la hora local de Bogotá.

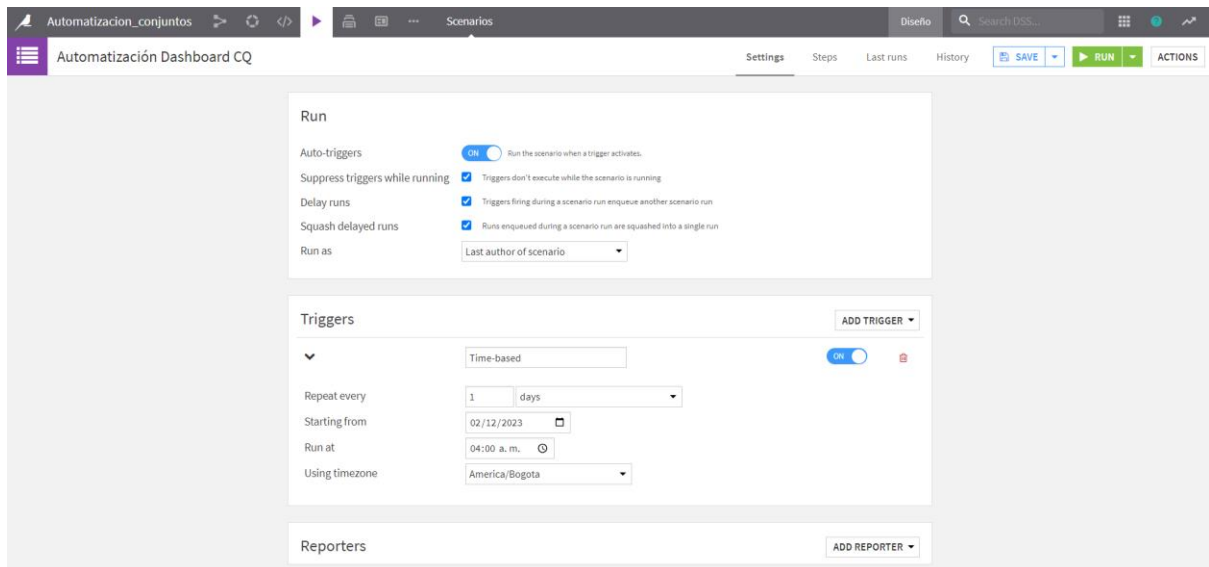


Imagen 16: Configuración de un escenario en Dataiku

## Paso 10: Forma de controlar fallos

De igual forma, elegimos la forma en cómo queremos recibir el reporte de la automatización de las fuentes para cada una de las iteraciones que se realizan ya que lo que queremos evitar al máximo son los fallos, en ese sentido, hemos elegido el email, lo cual notificará al cliente que las bases de datos fueron actualizadas.

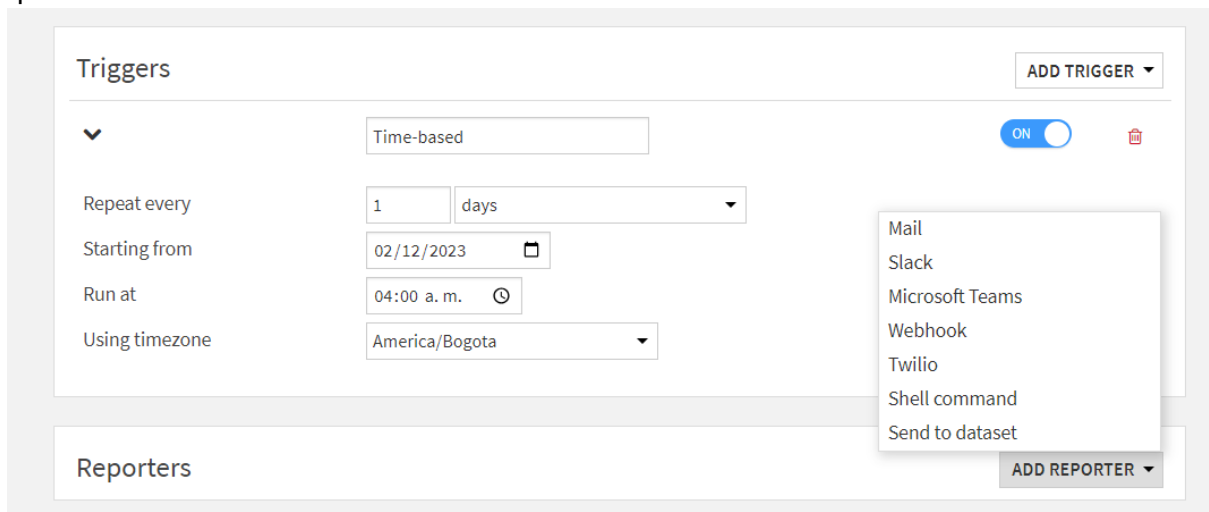


Imagen 17: Configuración del control de fallos y reporte de actualización

## Paso 11: Actualización del dashboard

Finalmente, después de todo este procesamiento el cliente deberá realizar la actualización manual del dashboard en caso de que no quiera adquirir una licencia premium y en un futuro cuando Comercial Química sea más tecnológico y pueda adquirirla el procesamiento será 100% automático.



Imagen 18: Actualización del dashboard

En resumen, hemos logrado una integración exitosa de los archivos manuales actualmente en posesión de Comercial Química mediante una propuesta basada en la nube. Este enfoque no solo garantiza la seguridad y accesibilidad de los datos en Google Cloud Platform, sino que también introduce una eficiencia significativa en el proceso.

A través de pasos cuidadosamente delineados, desde la carga inicial en la nube hasta la conexión en Power BI y la automatización en Dataiku, hemos creado un flujo de trabajo que reduce la operatividad requerida para el seguimiento. Ahora, con un simple clic de actualización en el panel de control, Comercial Química puede mantenerse al día con la información más reciente sin esfuerzo adicional. Esta transición hacia la nube no solo mejora la eficiencia operativa, sino que también sienta las bases para un análisis de datos más ágil y centrado en el rendimiento.