

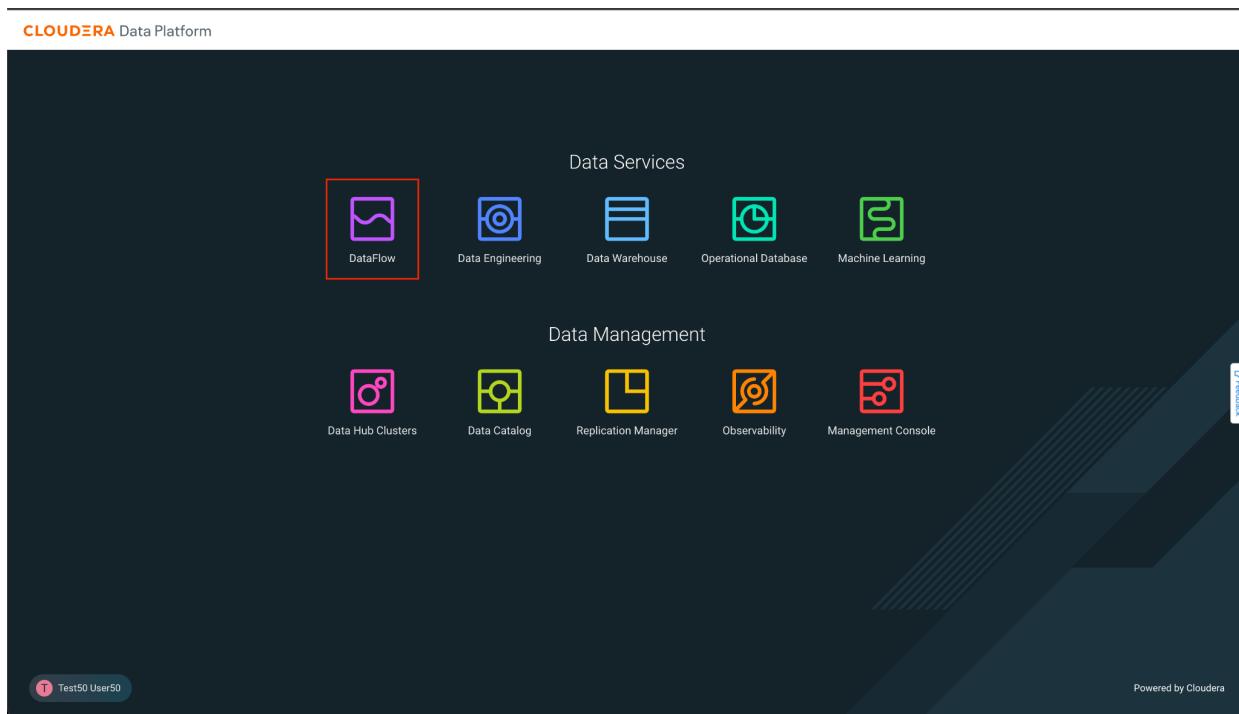
Data Lifecycle en CDP Public Cloud

Laboratorio Data Flow

Objetivos:

- Consumir datos de un tópico de Kafka
- Convertir los datos a formato Parquet
- Almacenar los datos en una tabla en el Lakehouse

1. Hacer clic en DataFlow desde el Home de CDP PC:



2. Una vez en DataFlow, hacer clic en la opción **Catalog** del menú izquierdo. Aquí se listan las plantillas de las aplicaciones de ingestión de datos. Para propósito de este taller, hemos creado y publicado una plantilla que permite leer datos del tópico de Kafka e ingestarlos/almacenarlos en el Lakehouse que dispone CDP Public Cloud. Hacer clic en el Flow llamado **kafka_to_lakehouse** para empezar el despliegue del mismo.

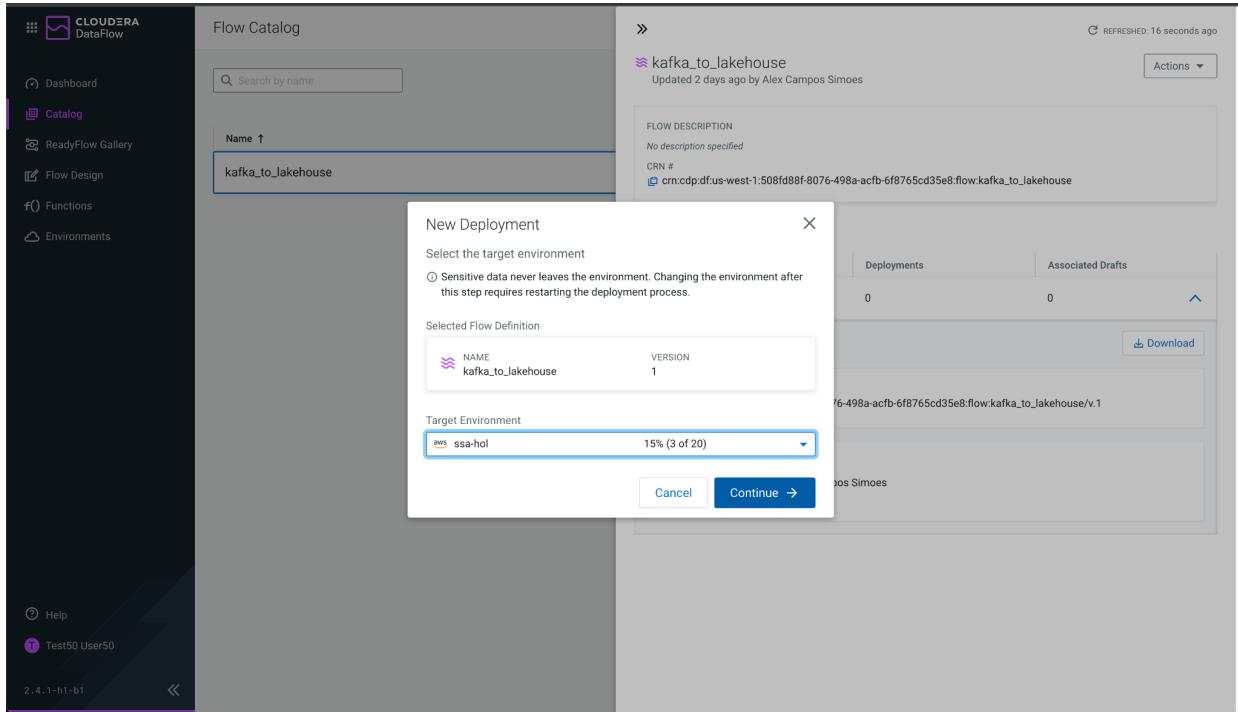
The screenshot shows the Cloudera DataFlow interface. On the left is a dark sidebar with navigation links: Dashboard, Catalog (which is selected and highlighted in purple), ReadyFlow Gallery, Flow Design, Functions, Environments, Help, and Test50 User50. The main content area is titled "Flow Catalog". It features a search bar labeled "Search by name". A table lists one flow entry: "kafka_to_lakehouse" (Custom Flow Definition, 1 version, last updated 2 days ago). The table includes columns for Name, Type, Versions, and Last Updated. At the bottom right of the table are pagination controls: "Items per page: 10", "1 – 1 of 1", and navigation arrows. A blue button at the top right says "Import Flow Definition". A small note at the top right says "REFRESHED: 5 seconds ago".

3. Al hacer clic, aparece el siguiente panel con la información del Flow. Muestra las versiones disponibles, fecha de creación, usuario creador, y un botón **Deploy** para empezar el despliegue. Hacer clic sobre ese botón.

This screenshot shows the same Cloudera DataFlow interface, but the "kafka_to_lakehouse" flow entry from the previous screen is now expanded. The expanded view includes:

- Flow Description:** "No description specified".
- CRN #:** "crm.cdp.df.us-west-1:508fd88f-8076-498a-acfb-6f8765cd35e8:flow:kafka_to_lakehouse".
- Deployment Status:** A checkbox labeled "Only show deployed versions" is unchecked. Below it is a table with one row, showing Version 1, 0 Deployments, and 0 Associated Drafts.
- Action Buttons:** A blue "Deploy →" button and a "Download" button with a download icon.
- Created Information:** CRN # "crm.cdp.df.us-west-1:508fd88f-8076-498a-acfb-6f8765cd35e8:flow:kafka_to_lakehouse/v.1", and the creation details: "CREATED 2023-05-19 00:15 CEST by Alex Campos Simoes", with a note "Initial Version".

4. La siguiente ventana popup permite seleccionar en que cluster de DataFlow se desea desplegar dicho Flow. En este caso, se selecciona el cluster **ssa-hol**. El instructor del taller os indicará qué ambiente deben seleccionar. Una vez seleccionado, hacer clic en **Continue**.



5. A partir de este punto, usted deberá ingresar la información del despliegue del Flow. Empieza por asignar un nombre (**Deployment Name**) y hacer clic en **Next**.

*Para propósitos de este taller, por favor nombrar el Flow con el nombre de usuario asignado - **user050**, por ejemplo.*

New Deployment

Overview

Deployment Name: user050
Deployment name is valid

Selected Flow Definition:

NAME	kafka_to_lakehouse
VERSION	1

Target Environment:

aws	NAME
	ssa-hol

[Cancel](#) [Next →](#)

6. Desmarcar la opción **Automatically start flow upon successful deployment** y hacer clic en **Next**.

Vamos a ejecutar paso a paso el Flow, por eso no queremos que inicie automáticamente.

New Deployment

NiFi Configuration

NIFI Runtime Version: CURRENT VERSION Latest Version (1.20.0.2.3.8.2-2) [Change Version](#)

Review the Cloudera DataFlow and CDP Runtime support matrix to ensure the selected NIFI Runtime Version is compatible.

Autostart Behavior: Automatically start flow upon successful deployment

Inbound Connections: Allow NiFi to receive data

Custom NAR Configuration: This flow deployment uses custom NARs

Overview

FLOW DEFINITION: kafka_to_lakehouse v1
ENVIRONMENT DEPLOYING TO: ssa-hol
DEPLOYMENT NAME: user050

[Cancel](#) [← Previous](#) [Next →](#)

7. En esta parte de Parámetros, debes ingresar los siguiente valores:

CDP Workload User Password: ingresar la Workload Password configurada para el usuario asignado en el inicio del taller.

CDP Workload Username: ingresar el número de usuario asignado, *user050*, por ejemplo.

Database: ingresar el número de usuario asignado, *user050*, por ejemplo. Esta base de datos y las tablas ya están precreadas. Luego la revisaremos.

Kafka Consumer Group Id: ingresar un valor único utilizando el usuario asignado.

Revisar que los parámetros fueron ingresados correctamente. Posteriormente hacer clic en **Next**.

New Deployment

Overview

NiFi Configuration

Parameters

Data entered here never leaves the environment in your cloud account. Provide parameter values directly in the text input or upload a file for parameters that expect a file.

The selected flow definition references an external Default NiFi SSL Context Service. Hence, DataFlow will automatically create a matching SSL Context Service with a keystore and truststore generated from the target environment's FreeIPA certificate.

SHOW: Sensitive No value

parameters (7)

CDP Workload User Password

CDP Workload Username

CDPEnvironment

core-site.xml

ssl-client.xml

hive-site.xml

Select File

Drop file or browse

FLOW DEFINITION kafka_to_lakehouse v.1
ENVIRONMENT DEPLOYING TO ssa-hol
DEPLOYMENT NAME user050

NiFi Configuration

NIFI RUNTIME VERSION Latest Version (1.20.0.2.3.8.2-2)
AUTO-START FLOW No
INBOUND CONNECTIONS No
CUSTOM NAR CONFIGURATION No

Cancel Previous Next

New Deployment

1. Overview

2. NiFi Configuration

3. Parameters

4. Sizing & Scaling

5. Key Performance Indicators

6. Review

CDPEnvironment

core-site.xml
ssl-client.xml
hive-site.xml

Select File
Drop file or browse

0/100K

DataFlow automatically adds all required configuration files to interact with Data Lake services. Unnecessary files that are added won't impact the deployment process.

Database

user050

7/100K

Kafka Brokers

realtime-ingestion-corebroker0.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker1.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker2.ssa-hol.yu1t-vbzg.cloudera.site:9093

203/100K

Kafka Consumer Group Id

Consumer_user050

16/100K

Kafka Topic

telco_data

10/100K

Overview

FLOW DEFINITION kafka_to_lakehouse v.1
ENVIRONMENT DEPLOYING TO ssa-hol
DEPLOYMENT NAME user050

NiFi Configuration

NIFI RUNTIME VERSION Latest Version (1.20.0.2.3.8.2-2)
AUTO-START FLOW No
INBOUND CONNECTIONS No
CUSTOM NAR CONFIGURATION No

Cancel **← Previous** **Next →**

8. No vamos a configurar parámetros de auto escalado, así que hacer clic en **Next**.

New Deployment

1. Overview

2. NiFi Configuration

3. Parameters

4. Sizing & Scaling

5. Key Performance Indicators

6. Review

Sizing & Scaling

Select the NiFi node size and the number of nodes provisioned for your flow.

NiFi Node Sizing

Extra Small
2 vCores Per Node
4 GB Per Node

Small
3 vCores Per Node
6 GB Per Node

Medium
6 vCores Per Node
12 GB Per Node

Large
12 vCores Per Node
24 GB Per Node

Number of NiFi Nodes

Auto Scaling Disabled

Nodes:

Overview

FLOW DEFINITION kafka_to_lakehouse v.1
ENVIRONMENT DEPLOYING TO ssa-hol
DEPLOYMENT NAME user050

NiFi Configuration

NIFI RUNTIME VERSION Latest Version (1.20.0.2.3.8.2-2)
AUTO-START FLOW No
INBOUND CONNECTIONS No
CUSTOM NAR CONFIGURATION No

Parameters

parameters
CDP WORKLOAD USER PASSWORD [Sensitive Value Provided]
CDP WORKLOAD USERNAME user050
COPENVIRONMENT
core-site.xml
ssl-client.xml
hive-site.xml
DATABASE
user050
KAFKA BROKERS
realtime-ingestion-corebroker0.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker1.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker2.ssa-hol.yu1t-vbzg.cloudera.site:9093

Cancel **← Previous** **Next →**

9. Tampoco vamos a configurar KPI, así que hacer clic en **Next**.

New Deployment

- Overview
- NiFi Configuration
- Parameters
- Sizing & Scaling
- Key Performance Indicators
- Review

Key Performance Indicators

Set up KPIs to track specific performance metrics of a deployed flow. Click and drag to reorder how they are displayed.

[Learn more](#)

[Add New KPI](#)

[Cancel](#) [← Previous](#) [Next →](#)

10. Revisar toda la información ingresada del despliegue, así que hacer clic en **Deploy** para empezar el proceso de despliegue.

New Deployment

- Overview
- NiFi Configuration
- Parameters
- Sizing & Scaling
- Key Performance Indicators
- Review

Review

[View CLI Command](#)

Overview

FLOW DEFINITION
kafka_to_lakehouse_v1

ENVIRONMENT DEPLOYING TO
ssa-hol

DEPLOYMENT NAME
user050

NiFi Configuration

NIFI RUNTIME VERSION
Latest Version (1.20.0.2.3.8.2-2)

AUTO-START FLOW
No

INBOUND CONNECTIONS
No

CUSTOM NAR CONFIGURATION
No

Parameters

parameters

CDP WORKLOAD USER PASSWORD
[Sensitive Value Provided]

CDP WORKLOAD USERNAME
user050

CDPENVIRONMENT
core-site.xml
ssl-client.xml
hive-site.xml

DATABASE
user050

KAFKA BROKERS
realtime-ingestion-corebroker0.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker1.ssa-hol.yu1t-vbzg.cloudera.site:9093,realtime-ingestion-corebroker2.ssa-hol.yu1t-vbzg.cloudera.site:9093

[Cancel](#) [← Previous](#) [Deploy](#)

11. El recuadro en azul indica que el proceso de despliegue del Flow fue inicializado. Haciendo click en el botón **Load More** podrás ver las distintas etapas del despliegue. Posterior a unos 60 a 90 segundos, el último evento debería ser **Deployment Successful**.

The screenshot shows the Cloudera DataFlow interface. On the left is a dark sidebar with navigation links: Dashboard, Catalog, ReadyFlow Gallery, Flow Design, Functions, Environments, Help, and Test50 User50. The main area is titled 'Dashboard' and shows a table of flows. One flow is listed: 'user050' (Status: Deploying, Environment: ssa-hol). To the right of the table is a detailed view for 'user050'. It includes tabs for KPIs, System Metrics, and Alerts. The 'Alerts' tab is selected, displaying a message: 'Deployment Initiated' with the subtext 'Initiated deployment of [user050]'. A red box highlights this message. Below the alerts, there's a section for Active Alerts (empty), Event History (empty), and a dropdown for SHOW ONLY: Info, Warning, and Error. At the bottom of the alerts section is a blue 'Load More' button.

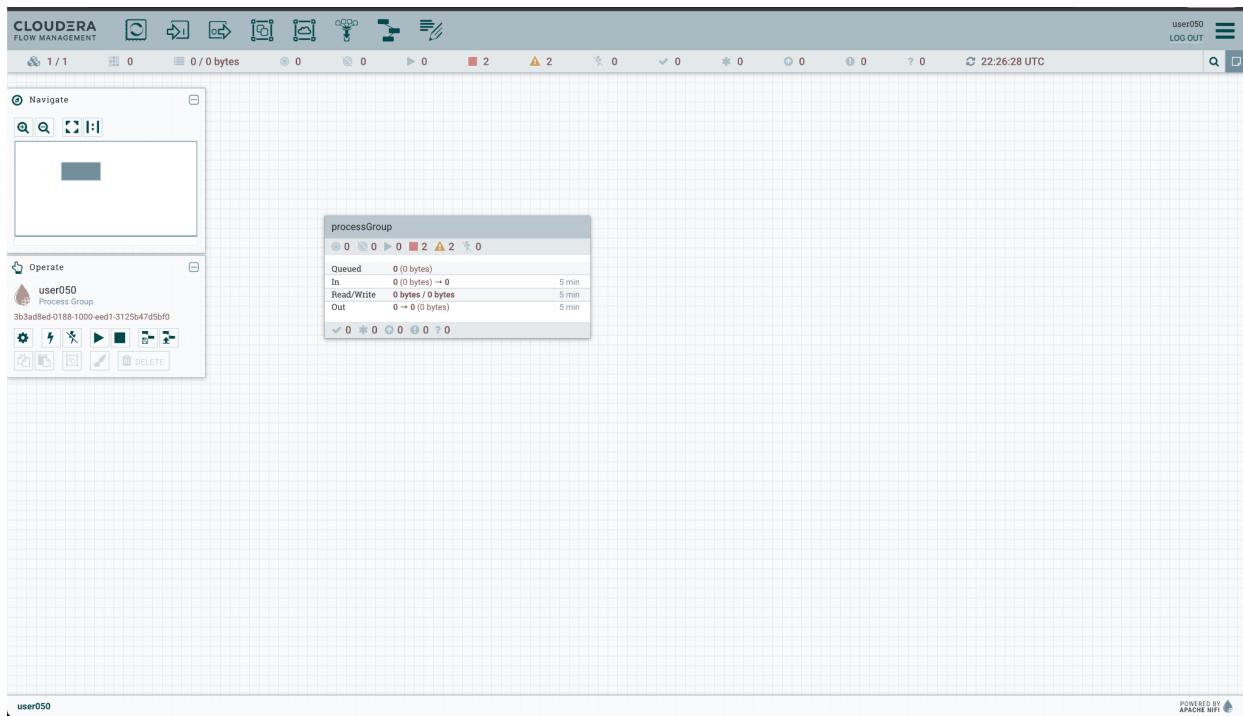
12. Una vez finalizado el despliegue, hacer click en **Manage Deployment** para ver el detalle del Flow recientemente desplegado.

The screenshot shows the Cloudera DataFlow interface. On the left sidebar, there are links for Dashboard, Catalog, ReadyFlow Gallery, Flow Design, Functions, Environments, Help, and Test50 User50. The main area is titled 'Dashboard' and shows a deployment named 'user050' with status 'Deploying'. The right side shows the 'Alerts' tab of the 'user050' deployment details, which currently has no alerts. Below this is the 'Event History' section, which lists various deployment events from 2023-05-21 00:15 CEST, such as 'Deployment Successful' and 'Importing NiFi Flow'. A red box highlights the 'Manage Deployment' button in the top right corner.

13. En esta ventana podrás ver la información del Flow desplegado. Es hora de ejecutar desde la interfaz gráfica los procesos de la aplicación. Hacer clic en **Actions** -> **View in NiFi**, para abrir en otra ventana el canvas de Cloudera Flow Management.

The screenshot shows the 'Deployment Manager' page for deployment 'user050'. It displays deployment details like status, node count, environment, flow definition, and deployment history. On the right, there is an 'Actions' dropdown menu with options: View in NiFi, Start flow, Change NiFi Runtime Version, Restart Deployment, and Terminate. The 'View in NiFi' option is highlighted with a red box. Below the deployment details, there is a 'Deployment Settings' section with tabs for KPIs and Alerts, Sizing and Scaling, Parameters, and NiFi Configuration. The 'KPIs and Alerts' tab is selected. At the bottom, there are buttons for Discard Changes, Apply Changes, and Update Deployment CLI Command.

14. En la nueva ventana que se abre aparece el Flow recientemente desplegado. Lo que se ve es el canvas donde se construyen las aplicación en Flow Management. Hacer doble clic en la caja; la única caja visible, que es un Process Group y debería llamarse **processGroup**.



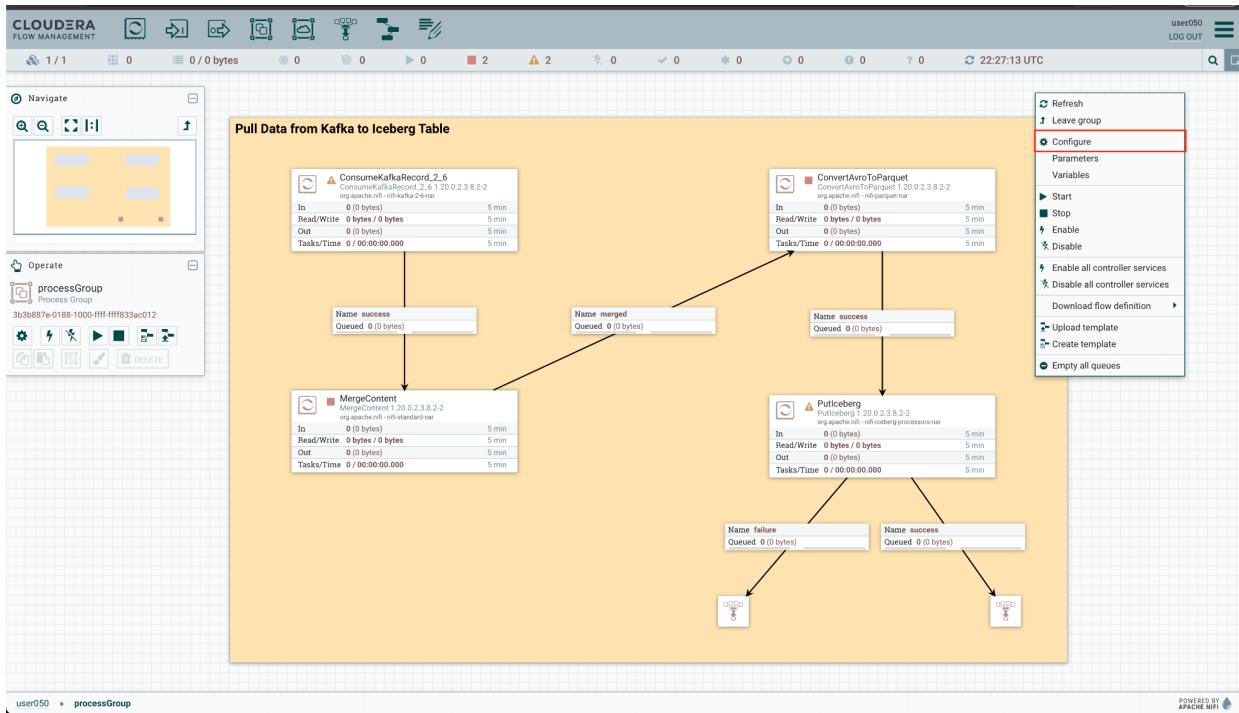
15. Al abrir el Process Group, se despliegan los Processors que componen el Flow. A modo resumen, hay cuatro Processors:

ConsumeKafkaRecord, processor para consumir datos del tópico de Kafka, leyendo los datos en formato JSON y saliendo en formato AVRO.

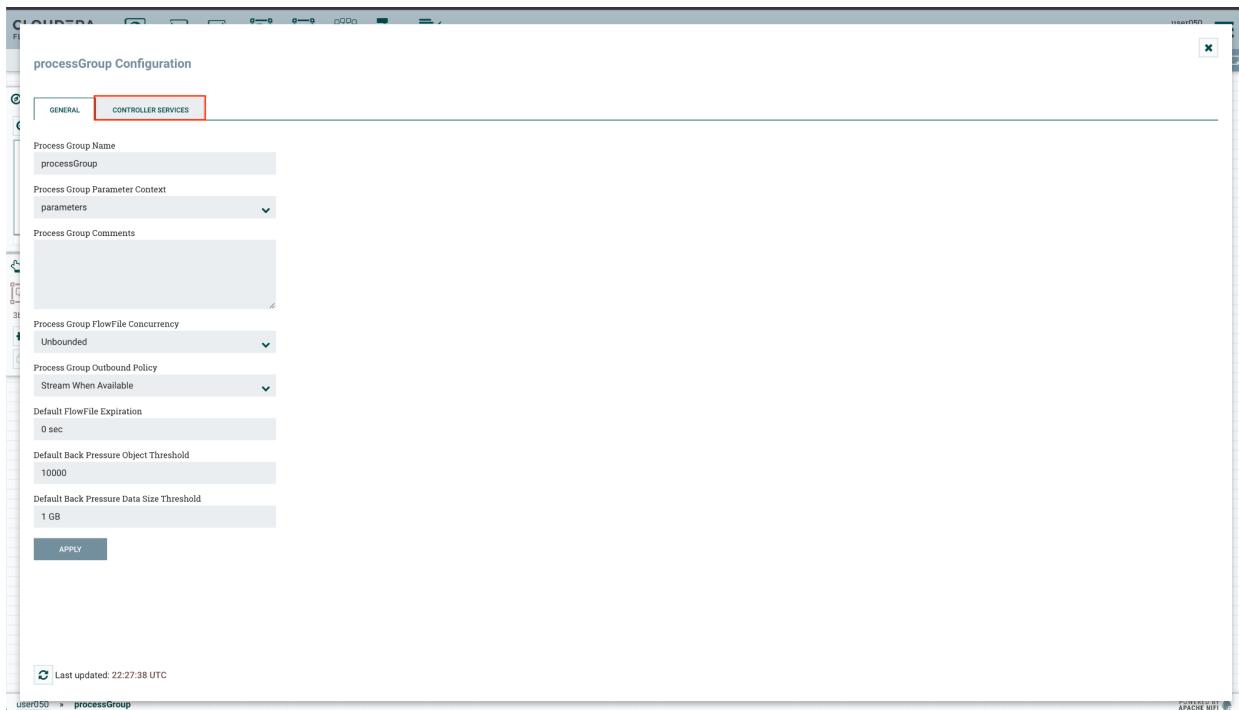
MergeContent, para agrupar los flow files y eficientar el flujo de datos.

ConvertAvroToParquet, conversión necesaria para guardar los datos en formato PARQUET. Putlceberg, para realizar la inserción de los datos en la tabla en el Lakehouse. La tabla destino se llama **telco_kafka_iceberg**, y cada usuario tiene una base de datos asignada.

Como verás, los Processors no están iniciados, y algunos tienen un mensaje/alerta de error. Lo último se debe a que existen componentes del flujo de datos que deben ser activados antes. Para activarlos - los Controller Services - hacer clic derecho sobre el canvas y hacer clic en la opción **Configure** del menu flotante que aparece.

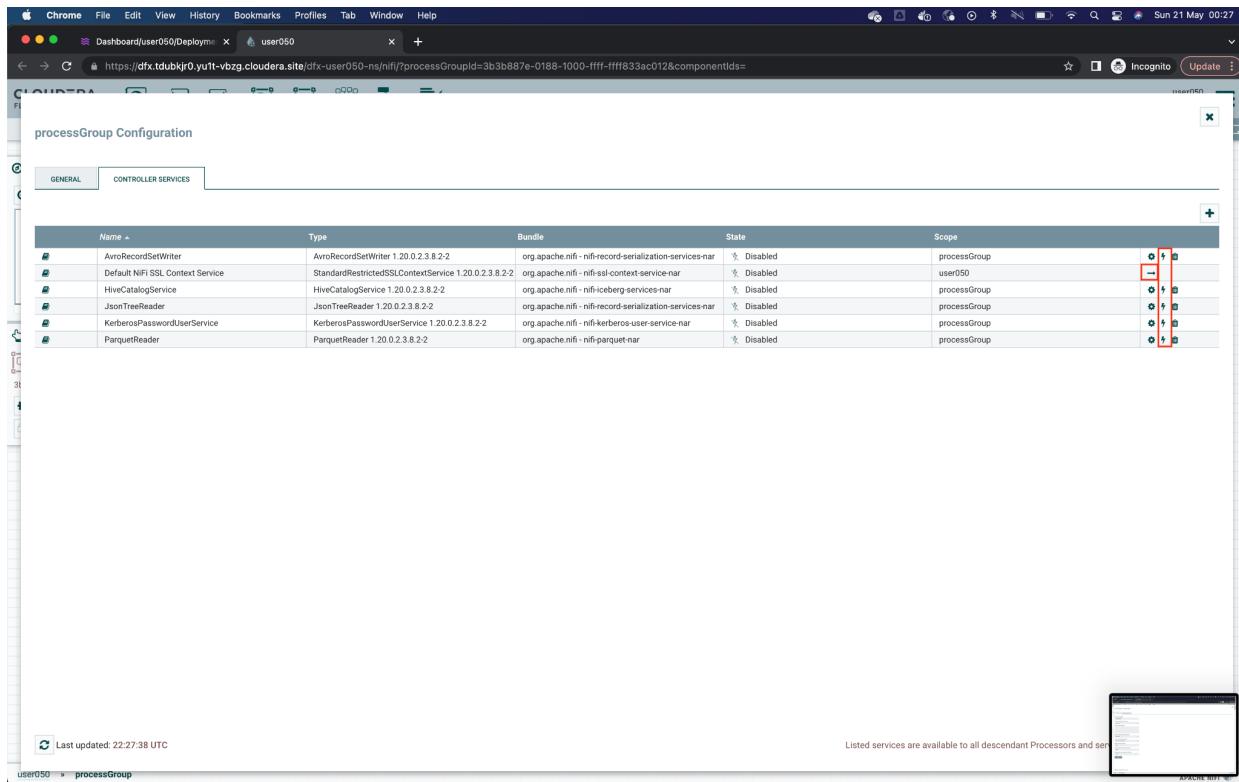


16. En la ventana emergente que se abre, seleccionar la pestaña **Controller Services**.



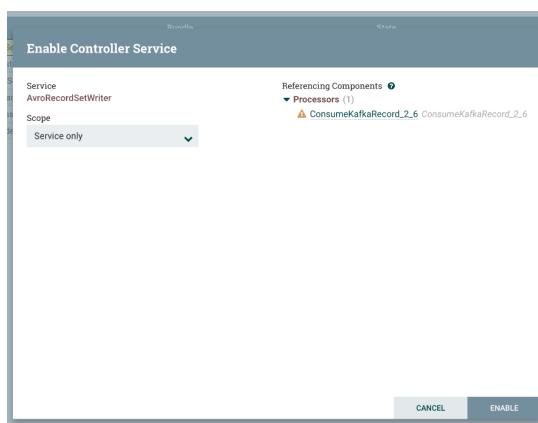
17. Se listan los **Controller Services** del flujo de datos. Se debe activar cada uno de ellos. Se debe activar primero los siguientes Controllers: **AvroReaderSetWriter**, **HiveCatalogService**,

JsonTreeReader, **KerberosPasswordUserService** y **ParquetReader** haciendo clic en el ícono de rayo  que aparece a la derecha (marcado en rojo).

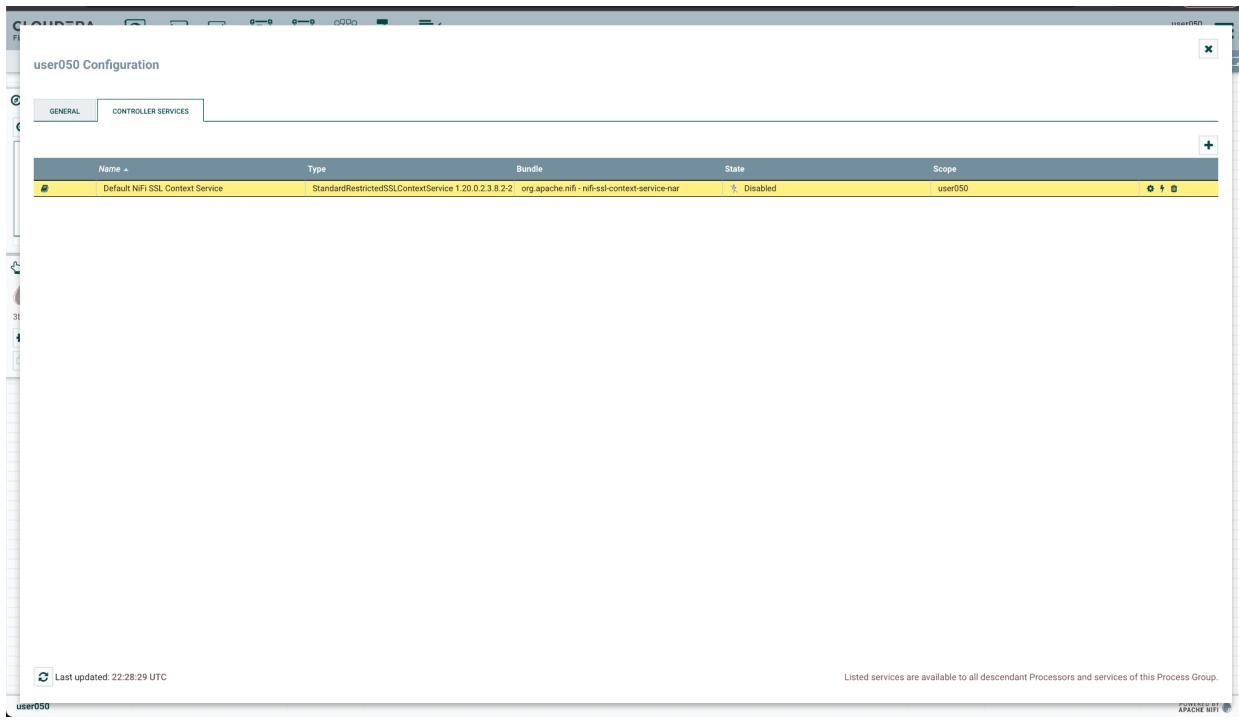


Name	Type	Bundle	State	Scope
AvroRecordSetWriter	AvroRecordSetWriter 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-record-serialization-services-nar	Disabled	processGroup
Default NiFi SSL Context Service	StandardRestrictedSSLContextService 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-ssl-context-service-nar	Disabled	user050
HiveCatalogService	HiveCatalogService 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-ceberg-services-nar	Disabled	processGroup
JsonTreeReader	JsonTreeReader 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-record-serialization-services-nar	Disabled	processGroup
KerberosPasswordUserService	KerberosPasswordUserService 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-kerberos-user-service-nar	Disabled	processGroup
ParquetReader	ParquetReader 1.20.0.2.3.8.2-2	org.apache.nifi - nifi-parquet-nar	Disabled	processGroup

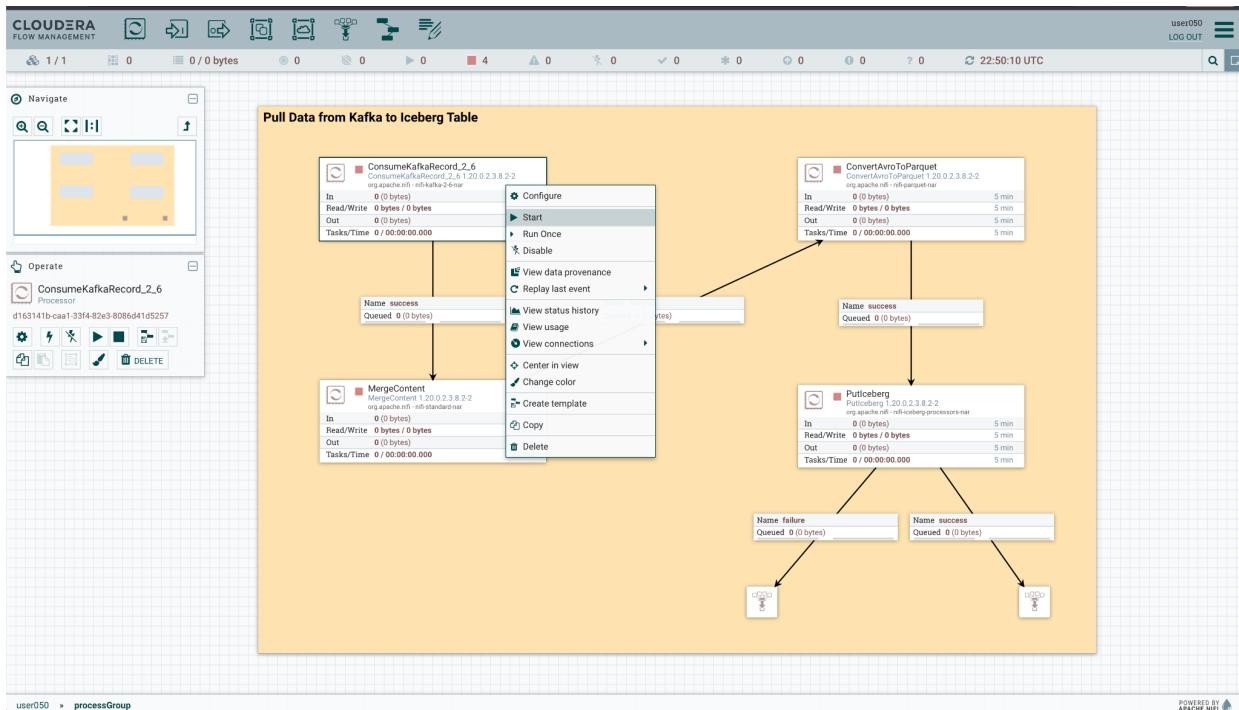
Hacer clic en el botón **Enable** en la ventana de confirmación de habilitación de cada Controller Service. Posteriormente cerrar dicha ventana para habilitar el siguiente Controller Service.



Para activar Default NiFi SSL Context Service, debe hacer clic en la flecha . Finalmente haciendo clic en el ícono de rayo  se activa el Controller Service **Default NiFi SSL Context Service**, que también presentará una ventana de habilitarlo.

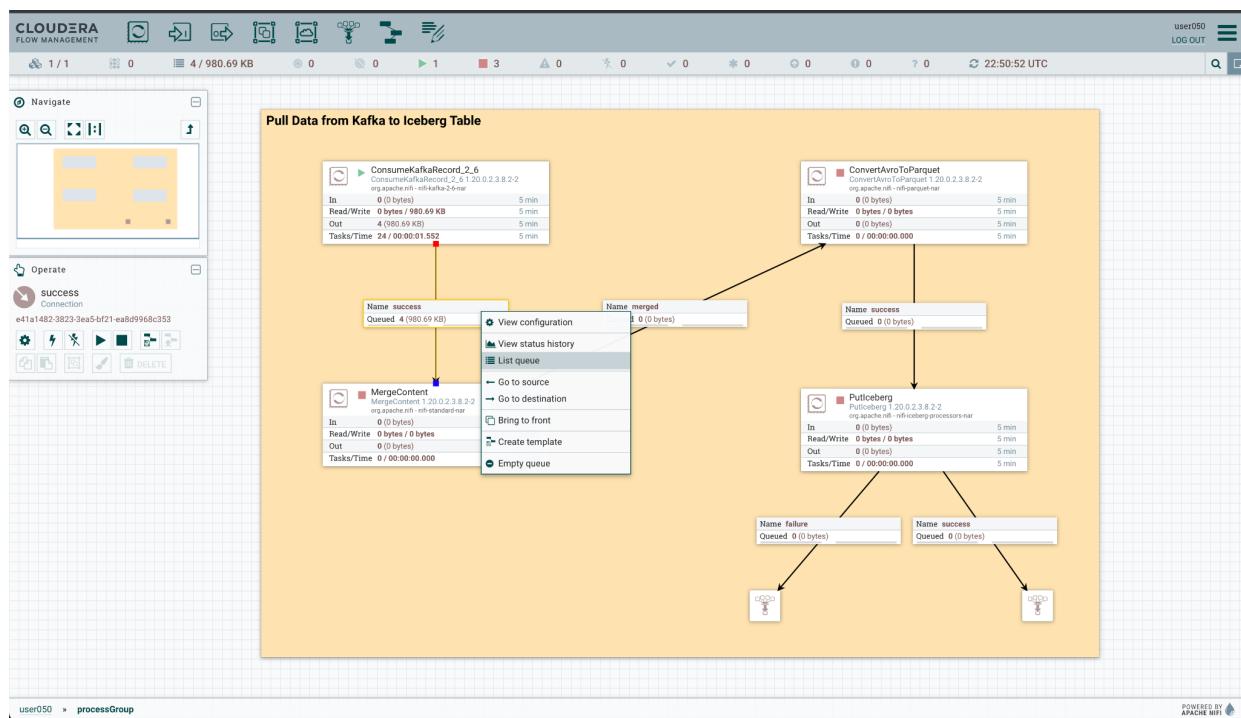


18. Cerrar la ventana de Controller Services, asegurándose que todos están habilitados. Volver al Process Group haciendo doble clic en el mismo. Es hora de ejecutar los **Processors**. Empezar por **ConsumeKafkaRecord**, haciendo clic derecho sobre ese, y posteriormente clic en **Start**. Esto empezará a consumir los datos de tópico de Kafka.



19. Flow Management nos permite ver y acceder a los datos en movimiento durante la ejecución del flujo de datos. Entre los Processors **ConsumeKafkaRecord** (recién iniciado) y **MergeContent**, hay una conexión. Esta conexión es lo que une los Processors y transmite datos de uno a otro.

Para revisar cuántos datos están encolados en esta conexión, refre el contador presionando la combinación Control+R (en Windows) o Command+R (en Mac) en el teclado. Esto permitirá actualizar las métricas actuales de todo el flujo de datos. En algún momento deberá ser un número al lado de la leyenda **Queued** en la conexión entre **ConsumeKafkaRecord** y **MergeContent**. Para ver los datos encolados, hacer clic derecho sobre la conexión y clic sobre la opción **List Queue**, abriendo una ventana emergente.



20. La siguiente ventana emergente lista los datos encolados. Hacer clic en el ícono de información (i) que aparece al lado izquierdo para visualizar los eventos.

SUCCESS

Displaying 4 of 4 (980.69 KB)

Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	Node
1	2055d337-695f-4c6d-8203-3ece27a62de...	2055d337-695f-4c6d-8203-3ece27a62de...	278.24 KB	00:00:12.787	00:00:13.068	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
2	510c8074-9798-4199-a228-ad7894aca9...	510c8074-9798-4199-a228-ad7894aca9...	283.60 KB	00:00:11.664	00:00:11.733	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
3	cad12e7c-e301-439c-85b3-a53fb0f13a2a...	cad12e7c-e301-439c-85b3-a53fb0f13a2a...	285.48 KB	00:00:11.575	00:00:11.647	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
4	01ee7d33-8e54-4a2b-a39c-a3f965b3cf87...	01ee7d33-8e54-4a2b-a39c-a3f965b3cf87...	133.37 KB	00:00:11.527	00:00:11.567	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...

The source of this queue is currently running. This listing may no longer be accurate.

Last updated: 22:50:59 UTC

user050 > processGroup

21. Una vez que la ventana de detalle del FlowFile aparece, hacer clic en el botón **VIEW** para abrir el contenido de los eventos consumidos.

SUCCESS

Displaying 4 of 4 (980.69 KB)

The source of this queue is currently running. This listing may no longer be accurate.

Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	Node
1	2055d337-695f-4c6d-8203-3ece27a62de...	2055d337-695f-4c6d-8203-3ece27a62de...	278.24 KB	00:00:12.787	00:00:13.068	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
2	510c8074-9798-4199-a228-ad7894aca9...	510c8074-9798-4199-a228-ad7894aca9...	283.60 KB	00:00:11.664	00:00:11.733	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
3	cad12e7c-e301-439c-85b3-a53fb0f13a2a...	cad12e7c-e301-439c-85b3-a53fb0f13a2a...	285.48 KB	00:00:11.575	00:00:11.647	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...
4	01ee7d33-8e54-4a2b-a39c-a3f965b3cf87...	01ee7d33-8e54-4a2b-a39c-a3f965b3cf87...	133.37 KB	00:00:11.527	00:00:11.567	No	dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.c...

FlowFile Details

DETAILS **ATTRIBUTES**

FlowFile Details	Content Claim
UUID 2055d337-695f-4c6d-8203-3ece27a62dee	Container default
Filename 2055d337-695f-4c6d-8203-3ece27a62dee	Section 1
File Size 278.24 KB	Identifier 1684623047700-1
Queue Position No value set	Offset 0
Queued Duration 00:00:19.534	Size 278.24 KB
Lineage Duration 00:00:19.815	VIEW
Penalized No	DOWNLOAD
Node Address dfx-nifi-0.dfx-nifi.dfx-user050-ns.svc.cluster.local:8443	

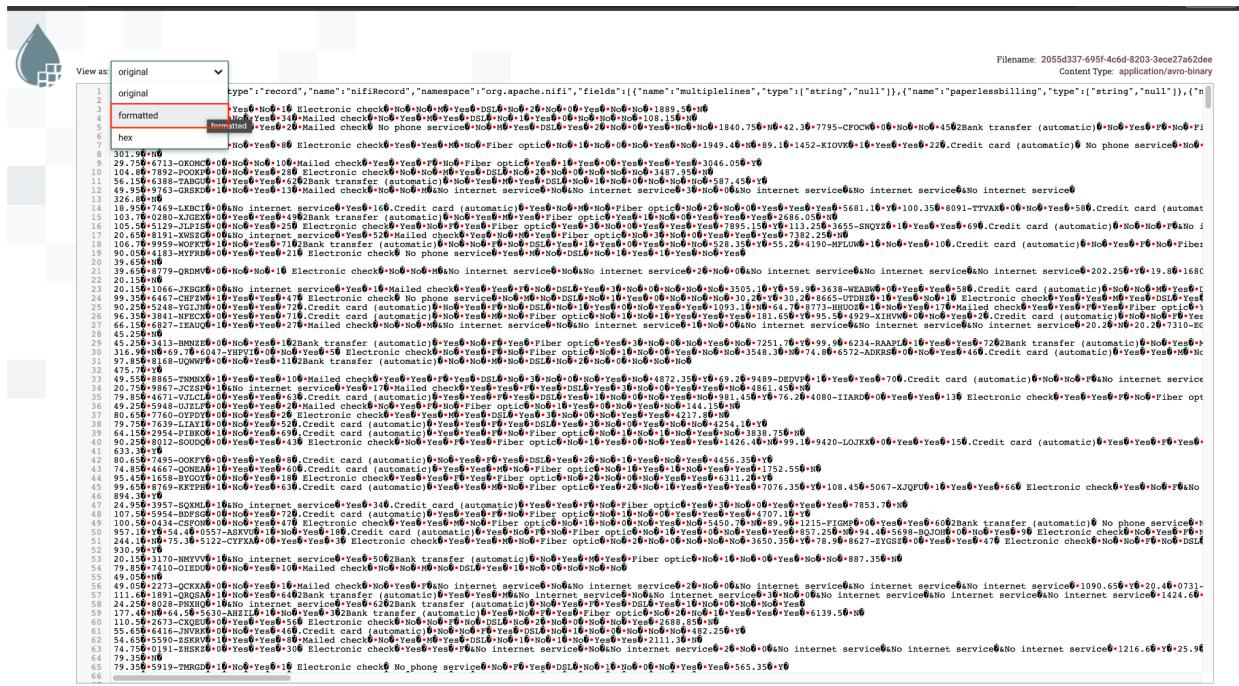
OK

Last updated: 22:50:59 UTC

user050 > processGroup

22. La nueva ventana que se abre muestra los datos del contenido del FlowFile. Al estar en formato AVRO, no es totalmente legible. Se debe seleccionar un deserializador para visualizar

correctamente los datos. Para esto, en la parte superior izquierda, seleccionar la opción **formatted** del menu **View as**.



1	original	File name: 2055d327-695f-4cdd-8203-3eca27a62dee Content type: application/avro-binary
2	original	type": "record", "name": "nifiRecord", "namespace": "org.apache.nifi", "fields": [{"name": "multipiplelines", "type": "string", "null": true}, {"name": "paperlessbilling", "type": "string", "null": true}], "id": "2055d327-695f-4cdd-8203-3eca27a62dee", "version": 1, "schema": "record { multipiplelines string; paperlessbilling string; }", "format": "avro", "size": 188, "modified": "2016-08-18T10:46:05Z", "created": "2016-08-18T10:46:05Z", "lastModified": "2016-08-18T10:46:05Z", "lastCreated": "2016-08-18T10:46:05Z", "lastAccessed": "2016-08-18T10:46:05Z", "lastWritten": "2016-08-18T10:46:05Z", "lastRead": "2016-08-18T10:46:05Z", "lastUpdated": "2016-08-18T10:46:05Z", "lastDeleted": "2016-08-18T10:46:05Z", "lastWrittenBy": "nifi", "lastReadBy": "nifi", "lastUpdatedBy": "nifi", "lastDeletedBy": "nifi", "lastAccessedBy": "nifi", "lastWrittenTime": "2016-08-18T10:46:05Z", "lastReadTime": "2016-08-18T10:46:05Z", "lastUpdatedTime": "2016-08-18T10:46:05Z", "lastDeletedTime": "2016-08-18T10:46:05Z", "lastAccessedTime": "2016-08-18T10:46:05Z", "lastWrittenTimeBy": "nifi", "lastReadTimeBy": "nifi", "lastUpdatedTimeBy": "nifi", "lastDeletedTimeBy": "nifi", "lastAccessedTimeBy": "nifi"}, {"id": "2055d327-695f-4cdd-8203-3eca27a62dee", "version": 1, "schema": "record { multipiplelines string; paperlessbilling string; }", "format": "avro", "size": 188, "modified": "2016-08-18T10:46:05Z", "created": "2016-08-18T10:46:05Z", "lastModified": "2016-08-18T10:46:05Z", "lastCreated": "2016-08-18T10:46:05Z", "lastAccessed": "2016-08-18T10:46:05Z", "lastWritten": "2016-08-18T10:46:05Z", "lastRead": "2016-08-18T10:46:05Z", "lastUpdated": "2016-08-18T10:46:05Z", "lastDeleted": "2016-08-18T10:46:05Z", "lastWrittenBy": "nifi", "lastReadBy": "nifi", "lastUpdatedBy": "nifi", "lastDeletedBy": "nifi", "lastAccessedBy": "nifi", "lastWrittenTime": "2016-08-18T10:46:05Z", "lastReadTime": "2016-08-18T10:46:05Z", "lastUpdatedTime": "2016-08-18T10:46:05Z", "lastDeletedTime": "2016-08-18T10:46:05Z", "lastAccessedTime": "2016-08-18T10:46:05Z", "lastWrittenTimeBy": "nifi", "lastReadTimeBy": "nifi", "lastUpdatedTimeBy": "nifi", "lastDeletedTimeBy": "nifi", "lastAccessedTimeBy": "nifi"}]
3	original	original
4	original	original
5	original	original
6	hex	hex
7	original	original
8	original	original
9	original	original
10	original	original
11	original	original
12	original	original
13	original	original
14	original	original
15	original	original
16	original	original
17	original	original
18	original	original
19	original	original
20	original	original
21	original	original
22	original	original
23	original	original
24	original	original
25	original	original
26	original	original
27	original	original
28	original	original
29	original	original
30	original	original
31	original	original
32	original	original
33	original	original
34	original	original
35	original	original
36	original	original
37	original	original
38	original	original
39	original	original
40	original	original
41	original	original
42	original	original
43	original	original
44	original	original
45	original	original
46	original	original
47	original	original
48	original	original
49	original	original
50	original	original
51	original	original
52	original	original
53	original	original
54	original	original
55	original	original
56	original	original
57	original	original
58	original	original
59	original	original
60	original	original
61	original	original
62	original	original
63	original	original
64	original	original
65	original	original
66	original	original

23. Ahora se puede visualizar los datos correctamente. Fíjese que aparecen los campos o atributos indicados en el inicio del taller. Puede cerrar dicha ventana del FlowFile y las ventanas emergentes, regresando al canvas con los cuatro Processors.

```

1  [
2   {
3     "multiplelines": "No phone service",
4     "paperlessbilling": "Yes",
5     "churn": "Yes",
6     "onlinesecurity": "No",
7     "internetservice": "DSL",
8     "techsupport": "No",
9     "contract": "1",
10    "phoneservice": "Yes",
11    "seniorcitizen": "0",
12    "deviceprotection": "No",
13    "streamingmovies": "No",
14    "totalcharges": "29.85",
15    "monthlycharges": "29.85",
16    "customerid": "5000-WWEG",
17    "onlinesubscription": "0",
18    "onlinetechsupport": "0",
19    "onlinebackup": "Yes",
20    "phoneservice": "No",
21    "tenure": "1",
22    "paymentmethod": "Electronic check"
23  },
24  {
25    "multiplelines": "No",
26    "paperlessbilling": "No",
27    "churn": "Yes",
28    "onlinesecurity": "Yes",
29    "internetservice": "DSL",
30    "techsupport": "Yes",
31    "contract": "2",
32    "seniorcitizen": "0",
33    "deviceprotection": "Yes",
34    "streamingmovies": "No",
35    "totalcharges": "1889.5",
36    "monthlycharges": "56.95",
37    "customerid": "5575-GNVD8",
38    "onlinesubscription": "0",
39    "onlinetechsupport": "0",
40    "onlinebackup": "Yes",
41    "phoneservice": "Yes",
42    "tenure": "34",
43    "paymentmethod": "Mailed check"
44  },
45  {
46    "multiplelines": "No",
47    "paperlessbilling": "Yes",
48    "churn": "Yes",
49    "onlinesecurity": "Yes",
50    "internetservice": "DSL",
51    "techsupport": "No",
52    "contract": "1",
53    "seniorcitizen": "0",
54    "deviceprotection": "0",
55    "streamingmovies": "No",
56    "totalcharges": "108.15",
57    "monthlycharges": "53.85",
58    "customerid": "3668-QPVK",
59    "onlinesubscription": "0",
60    "onlinetechsupport": "Yes",
61    "onlinebackup": "Yes",
62    "phoneservice": "Yes",
63    "tenure": "2",
64    "paymentmethod": "Mailed check"
65  },
66  {
67    "multiplelines": "Yes",
68    "paperlessbilling": "Yes",
69    "churn": "Yes",
70    "onlinesecurity": "Yes",
71    "internetservice": "Fiber optic",
72    "techsupport": "Yes",
73    "contract": "2",
74    "seniorcitizen": "0",
75    "deviceprotection": "Yes",
76    "streamingmovies": "Yes",
77    "totalcharges": "1000.0",
78    "monthlycharges": "500.0",
79    "customerid": "5000-ABCD",
80    "onlinesubscription": "0",
81    "onlinetechsupport": "Yes",
82    "onlinebackup": "Yes",
83    "phoneservice": "Yes",
84    "tenure": "12",
85    "paymentmethod": "Mailed check"
86  }
]

```

24. Continúe ejecutando cada uno de los Processors en orden: **MergeContent**, después **ConvertAvroToParquet** y finalmente **PutIceberg**. Recuerda que puedes refrescar los contadores del flujo con la combinación Control+R o Command+R.

Si los pasos anteriores fueron ejecutados correctamente, la conexión del Processor **PutIceberg** a un funnel debería ser de tipo **success**.

