


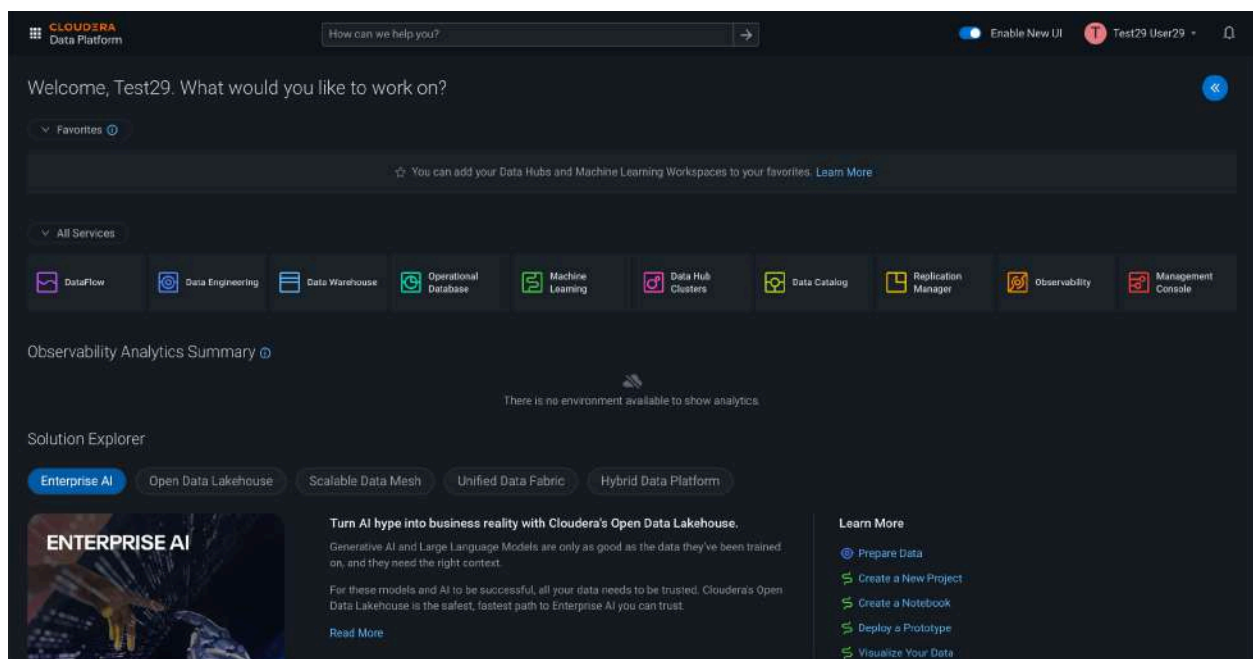
Data Lifecycle CDP Public Cloud

Lab 001: Data Flow Lab

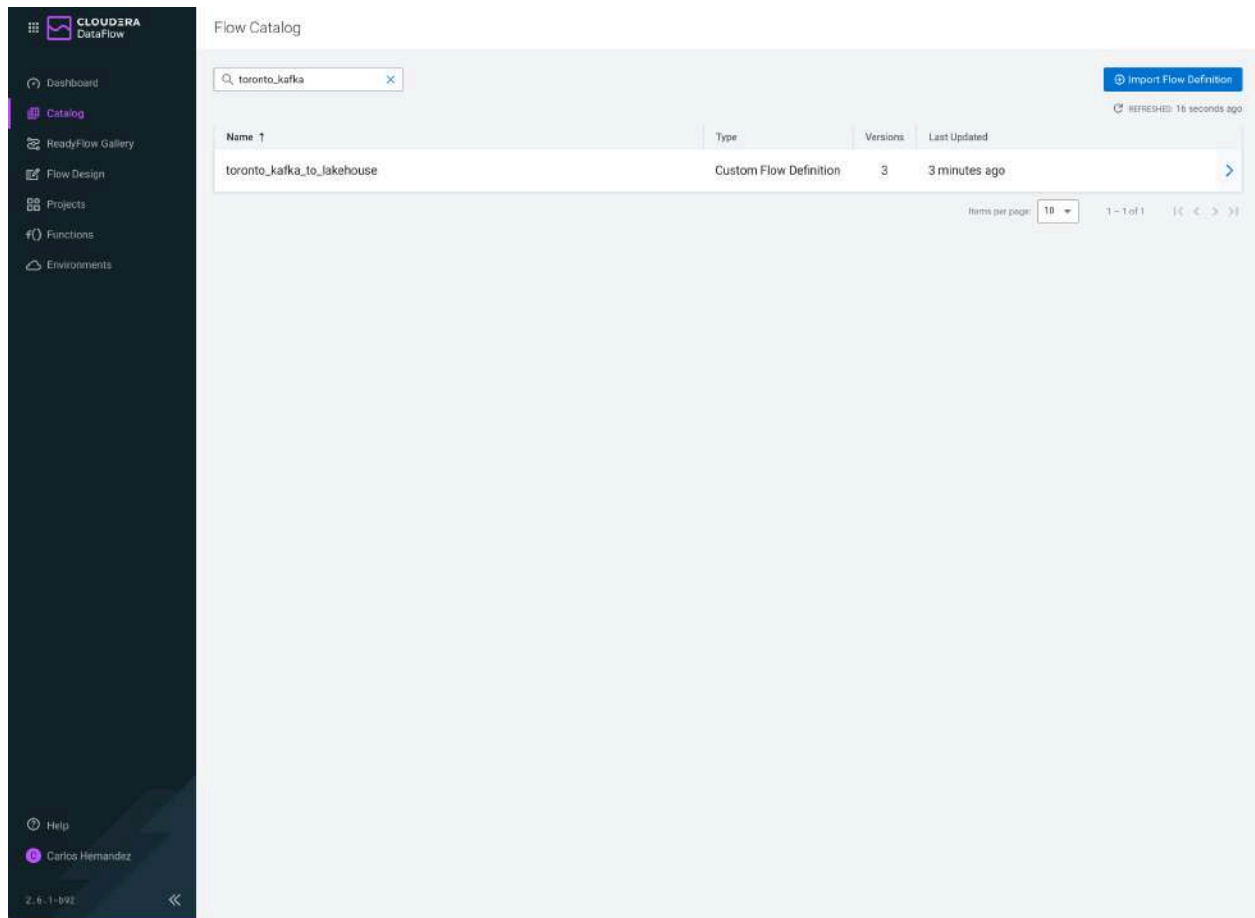
Goals:

- Consume data from a Kafka topic
- Convert the data to Parquet format
- Store the data in a table in the Lake House

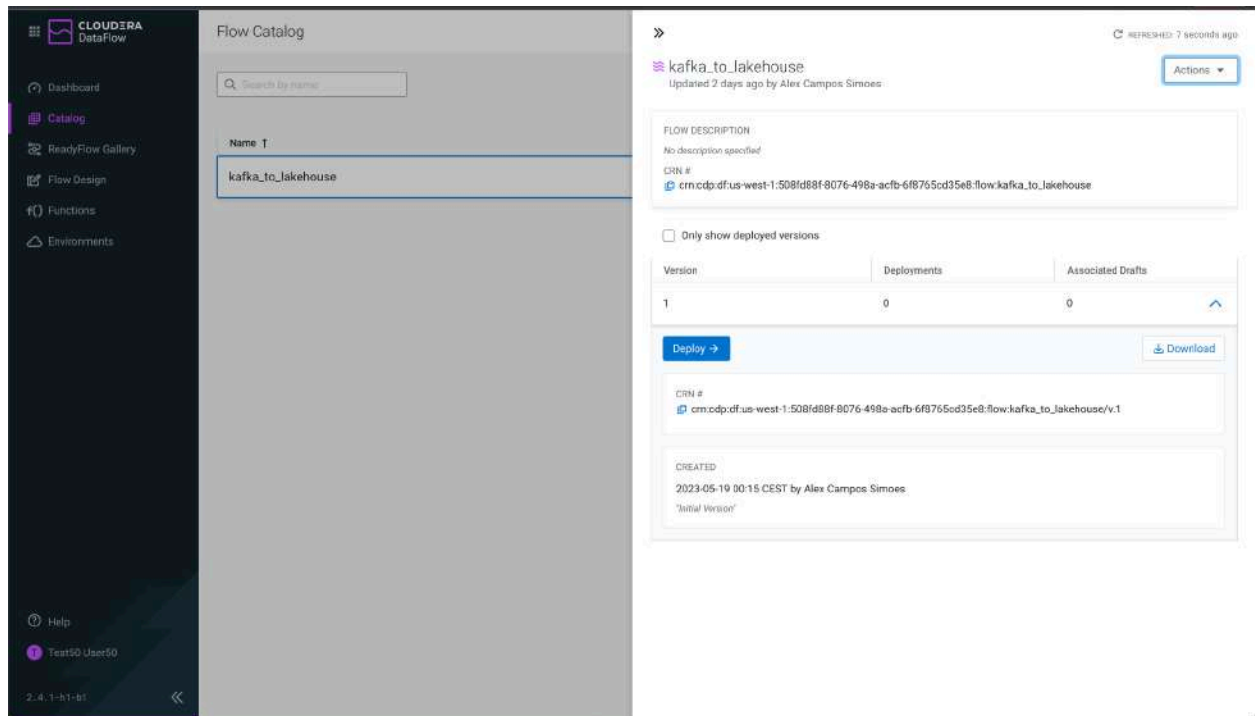
1. Click on DataFlow ( DataFlow) from CDP PC Home:



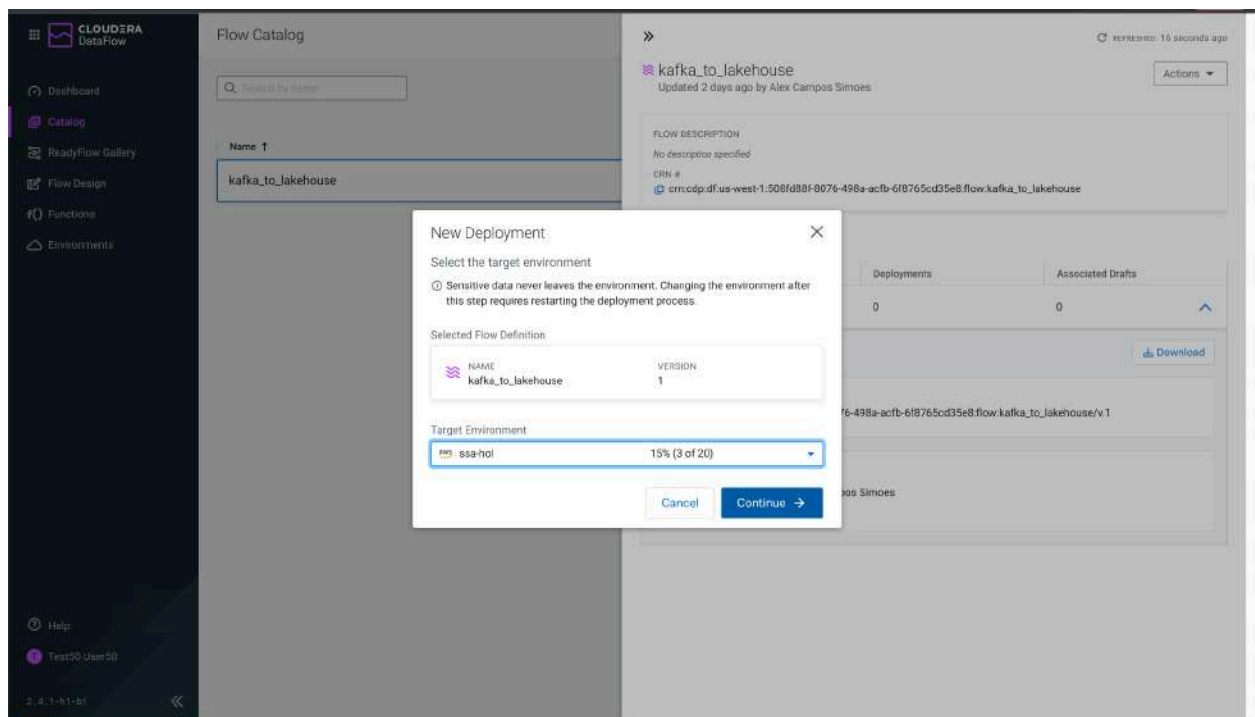
2. Once in DataFlow, click on the option **Catalog** from the left menu. The data ingestion application templates are listed here. For the purpose of this workshop, we have created and published a template that allows you to read Kafka topic data and ingest/store it in the Lakehouse provided by CDP Public Cloud. Click on the Flow called **kafka_to_lakehouse** to start deploying it.



3. When clicked, the following panel appears with the Flow information. It shows the available versions, creation date, creator user, and a button **Deploy** to start the deployment. Click on that button.



4. The following popup window allows you to select the DataFlow cluster in which you want to deploy the Flow. In this case, the cluster to be selected is **yeg-cdp-env**. The workshop instructor will tell you which environment to select. Once selected, click **Continue**.



5. From this point, you will need to enter the Flow configuration. Start by assigning a **Deployment Name**, **Target Project**, and click **Next**.

For the purposes of this workshop, please name the Flow starting with your assigned username. For example, **user050_kafka_to_lakehouse**

Select **workshop** or **unassigned** for the project, which is a way to organize your flows.

New Deployment

1 Overview

2 Nifi Configuration

3 Parameters

4 Sizing & Scaling

5 Key Performance Indicators

6 Review

Overview

Deployment Name

user050_kafka_to_lakehouse

Deployment name is valid

Selected Flow Definition

NAME	VERSION
toronto_kafka_to_lakehouse	3

Target Environment

aws

tor-hol-cdp-ene

Target Project

Select a project

Cancel

Next →

6. Make sure the option **Automatically start flow upon successful deployment** is checked and click **Next**.

7. In these Parameters, you must enter the following values:

CDP Workload User Password: Enter the Workload Password shared at the beginning of the workshop.

CDP Workload Username: enter the assigned user number, *user050*, for example.

NOTE: for the purposes of the workshop, your user (e.g. *user050*) is also the name of the **database** where you will store the data (which has already been created for you), and the name of the **Kafka Consumer Group ID** for reading messages.

For the purposes of this workshop, the remaining values were filled out for you and don't need to change.

Review that the parameters were entered correctly. Then click **Next**.

New Deployment

Overview

NiFi Configuration

Parameters

Sizing & Scaling

Key Performance Indicators

Review

toronto_kafka_to_lakehouse (6)

CDPEEnvironment ⓘ 9/100K

core-site.xml ✓

ssl-client.xml ✓

hive-site.xml ✓

Select File
Drop file or browse

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

kafka_brokers 179/100K
streamo-corebroker0.toronto.z30z-14kp.cloudera.site:9093,streamo-corebroker1.toronto.z30z-14kp.cloudera.site:9093,streamo-corebroker2.toronto.z30z-14kp.cloudera.site:9093

kafka_topic 18/100K
telco_data

table 19/100K
telco_iceberg_kafka

workload_password 9/100K
Enter parameter values.

workload_user ⓘ 9/100K
Enter parameter values.

Overview

FLOW DEFINITION
toronto_kafka_to_lakehouse v6

ENVIRONMENT DEPLOYING TO
toronto-cdp-env

PROJECT ASSIGNING TO:
Workshop

DEPLOYMENT NAME
user50_kafka_to_lakehouse

NiFi Configuration

NIFI RUNTIME VERSION
Latest Version (1.23.2.2.3.11.0-9)

AUTO-START FLOW
Yes

INGOING CONNECTIONS
No

CUSTOM NAR CONFIGURATION
No

Cancel

← Previous

Next →

8. There is no need to configure auto-scaling parameters. Click **Next**.

New Deployment

- Overview
- NiFi Configuration
- Parameters
- Sizing & Scaling**
- Key Performance Indicators
- 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
8 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 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

CDP ENVIRONMENT
core-ss6.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. We are also not going to configure KPIs now. Click **Next** to continue the configuration.
[Optionally, you could click on ADD KPI to explore and configure a KPI to monitor such as Data in, Data Out, etc]

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](#)

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

CDP ENVIRONMENT
core-ss6.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

10. Review all the information entered for your Flow, then click on **Deploy** to start the deployment process.

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 (1.29.0.2.3.8.2-2)

AUTO START FLOW
No

REBOUND CONNECTIONS
No

CUSTOM NAR CONFIGURATION
No

Parameters

parameters

CDP WORKLOAD USER PASSWORD
(Sensitive Value Provided)

CDP WORKLOAD USERNAME
user050

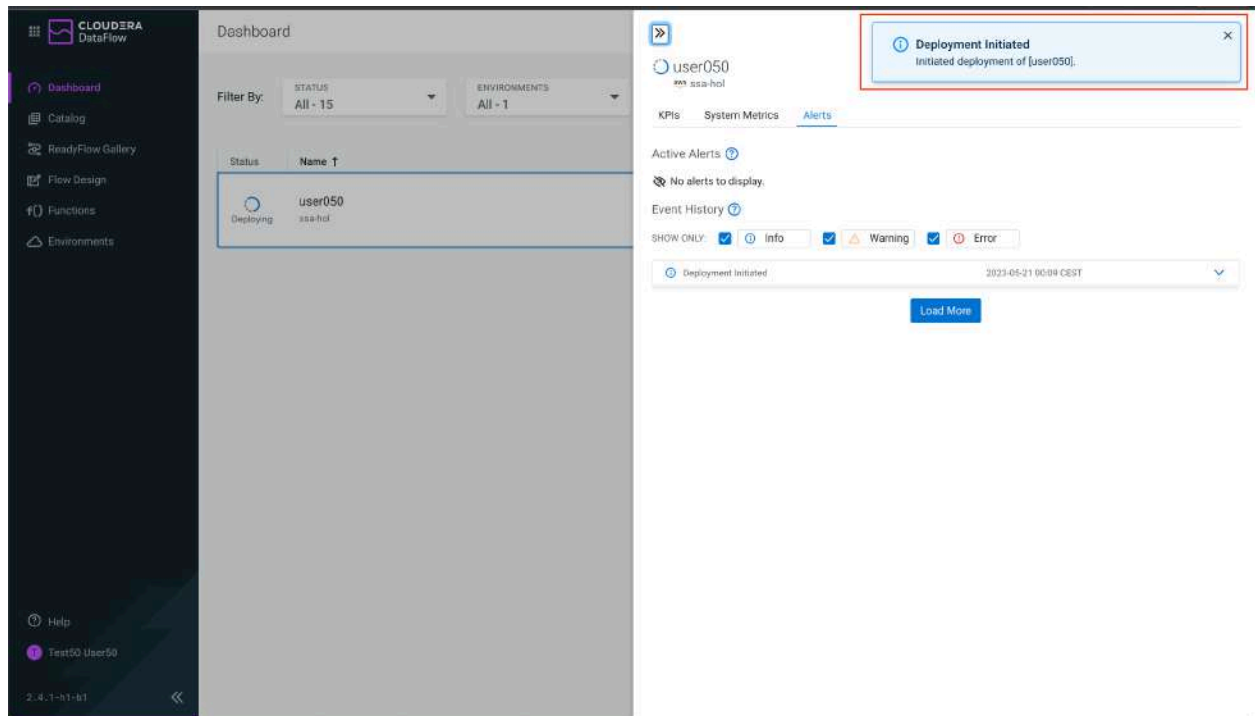
CDP ENVIRONMENT
core-site.xml
ssl-client.xml
hive-site.xml

DATABASE
user050

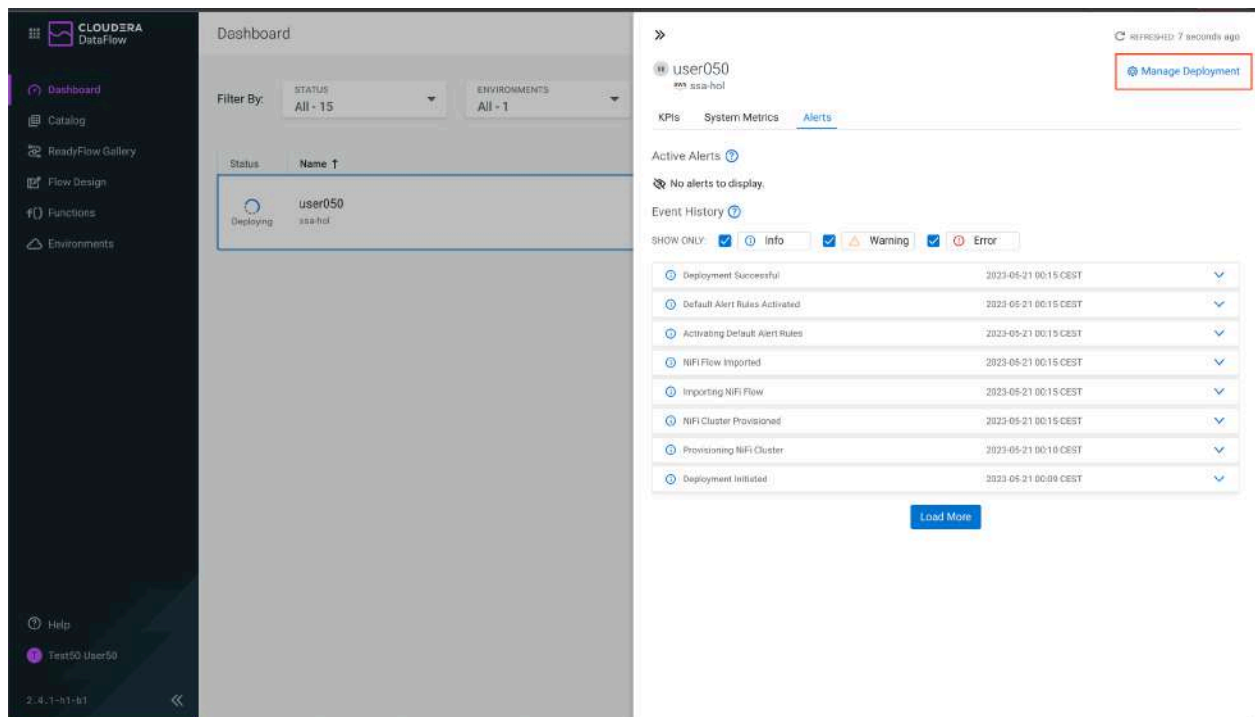
KAFKA BROKERS

Cancel Previous Deploy

11. The blue box indicates that the Flow deployment process has been started. By clicking on the button **Load More** you will be able to see the different stages of the deployment. After about 60 to 90 seconds approximately, the last event should be *Deployment Successful*.



12. Once the deployment is finished, click on **Manage Deployment** to see the details of the recently deployed Flow.



13. In this window you will see the Flow information displayed. It is time to execute the application processes from the graphical Flow Management interface. Click on **Actions** -> **View in NiFi**, to open Cloudera Flow Management canvas in a new window/tab.

The screenshot displays the Cloudera DataFlow Deployment Manager interface. The left sidebar contains navigation links: Dashboard, Catalog, ReadyFlow Gallery, Flow Design, Functions, and Environments. The main content area shows the 'Deployment Manager' for a deployment named 'user050'. The deployment status is 'Suspended'. Key details include: Node Count: 1, Environment: sse-hol, Region: US East(N. Virginia), Flow Definition: kafka_to_lakehouse V.1, Created On: 2023-05-21 00:09 CEST, Last Updated: 2023-05-21 00:15 CEST, and Nifi Runtime Version: 1.20.0.2.3.8.2.2. An 'Actions' dropdown menu is open, showing options: View in NiFi, Start flow, Change Nifi Runtime Version, Restart Deployment, and Terminate. Below the deployment details is a 'Recreate Deployment CLI Command' button. The 'Deployment Settings' section includes tabs for 'KPIs and Alerts', 'Sizing and Scaling', 'Parameters', and 'Nifi Configuration'. The 'KPIs and Alerts' tab is active, showing 'Key Performance Indicators' and an 'Add New KPI' button. At the bottom, there are buttons for 'Discard Changes', 'Apply Changes', and 'Update Deployment CLI Command'.

STATUS	DEPLOYMENT NAME	FLOW DEFINITION	DEPLOYED BY
Suspended	user050	kafka_to_lakehouse V.1	Test50 User50

NODE COUNT	AUTO SCALING	CREATED ON	LAST UPDATED
1	Disabled	2023-05-21 00:09 CEST	2023-05-21 00:15 CEST

ENVIRONMENT	REGION	NIFI RUNTIME VERSION	CRN #
sse-hol	US East(N. Virginia)	1.20.0.2.3.8.2.2	cm:cdp.dfus-west

Actions:

- View in NiFi
- Start flow
- Change Nifi Runtime Version
- Restart Deployment
- Terminate

Recreate Deployment CLI Command

Deployment Settings

KPIs and Alerts | Sizing and Scaling | Parameters | Nifi Configuration

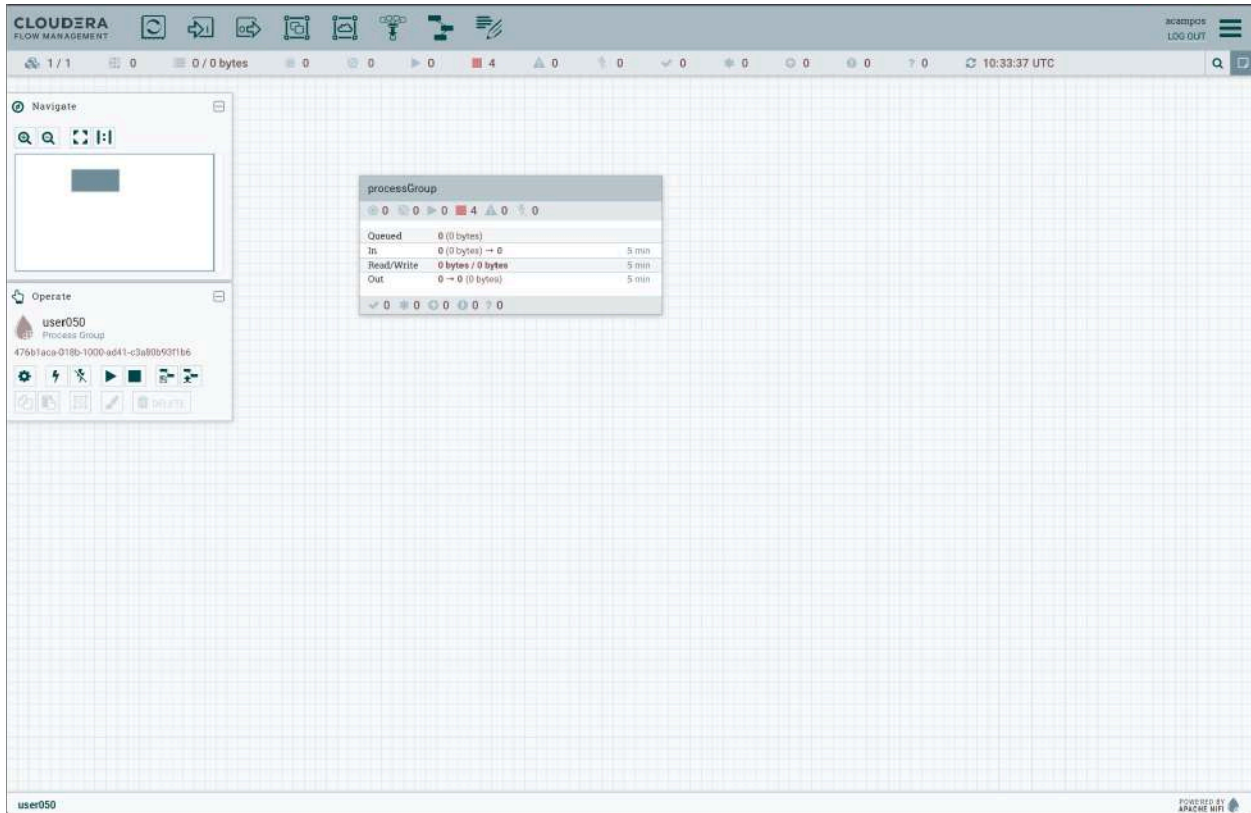
Key Performance Indicators

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

Discard Changes | Apply Changes | Update Deployment CLI Command

14. Double-click on the Process Group to open it.



16. When opening the Process Group, you should be able to see the Processors that compose the Flow application. To summarize, there are four Processors:

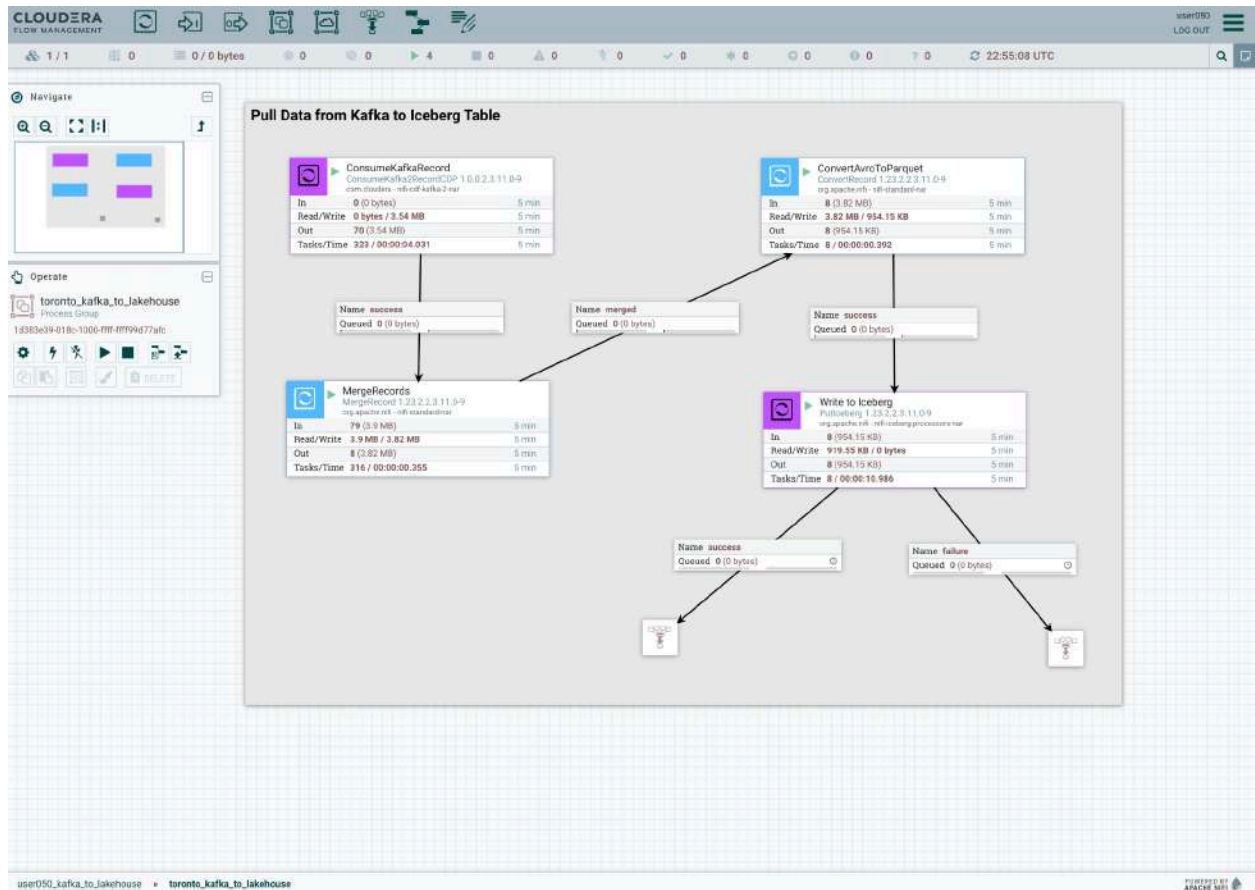
ConsumeKafkaRecord, consumes data from the Kafka topic, reading the data in JSON and outputting in AVRO.

MergeRecords, to group the flow files and streamline the data flow.

ConvertAvroToParquet, conversion needed to store the data in PARQUET format.

PutIceberg, to insert the data into the table in the Lakehouse. The destination table is called *telco_kafka_iceberg*, and each user has an assigned database (user_id is the name of the database).

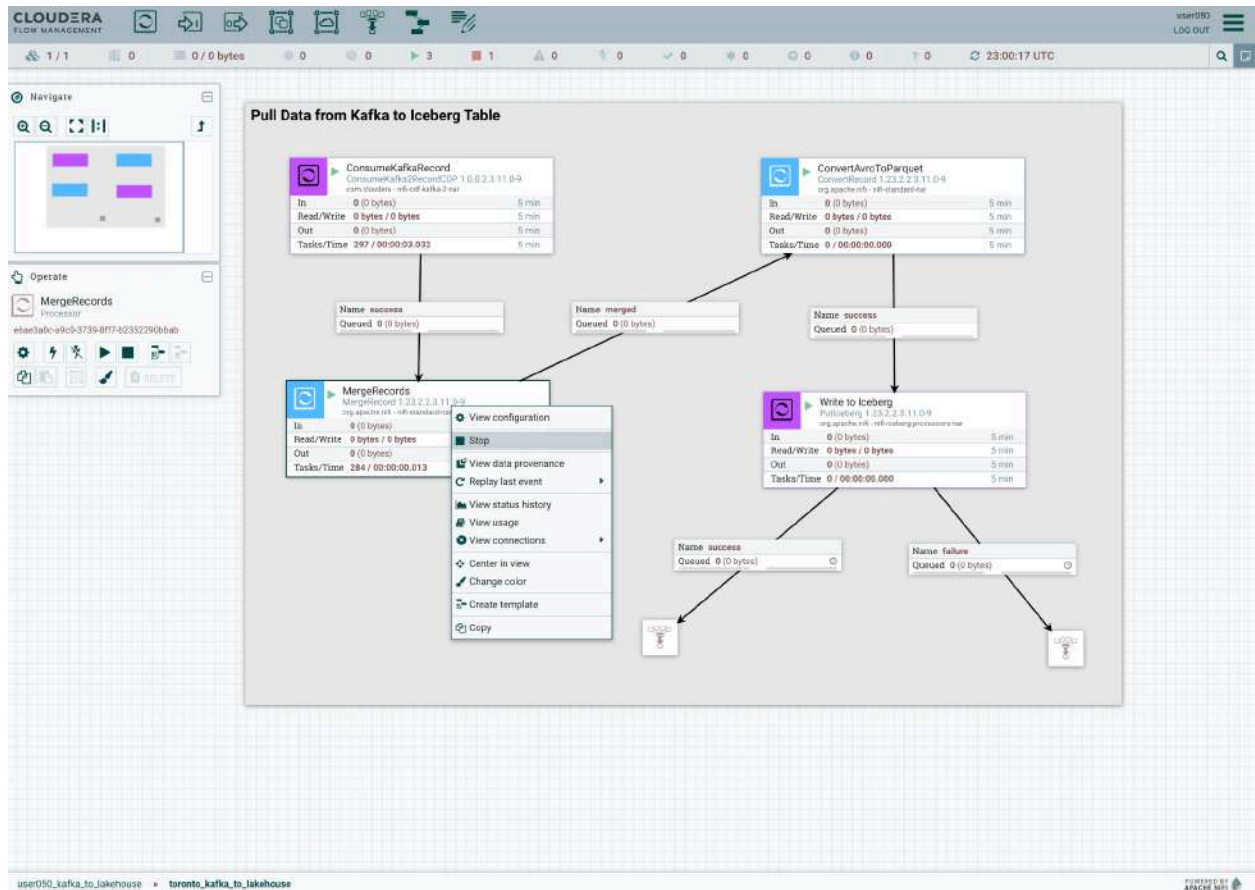
As you can see, the Processors are not started, they are paused.



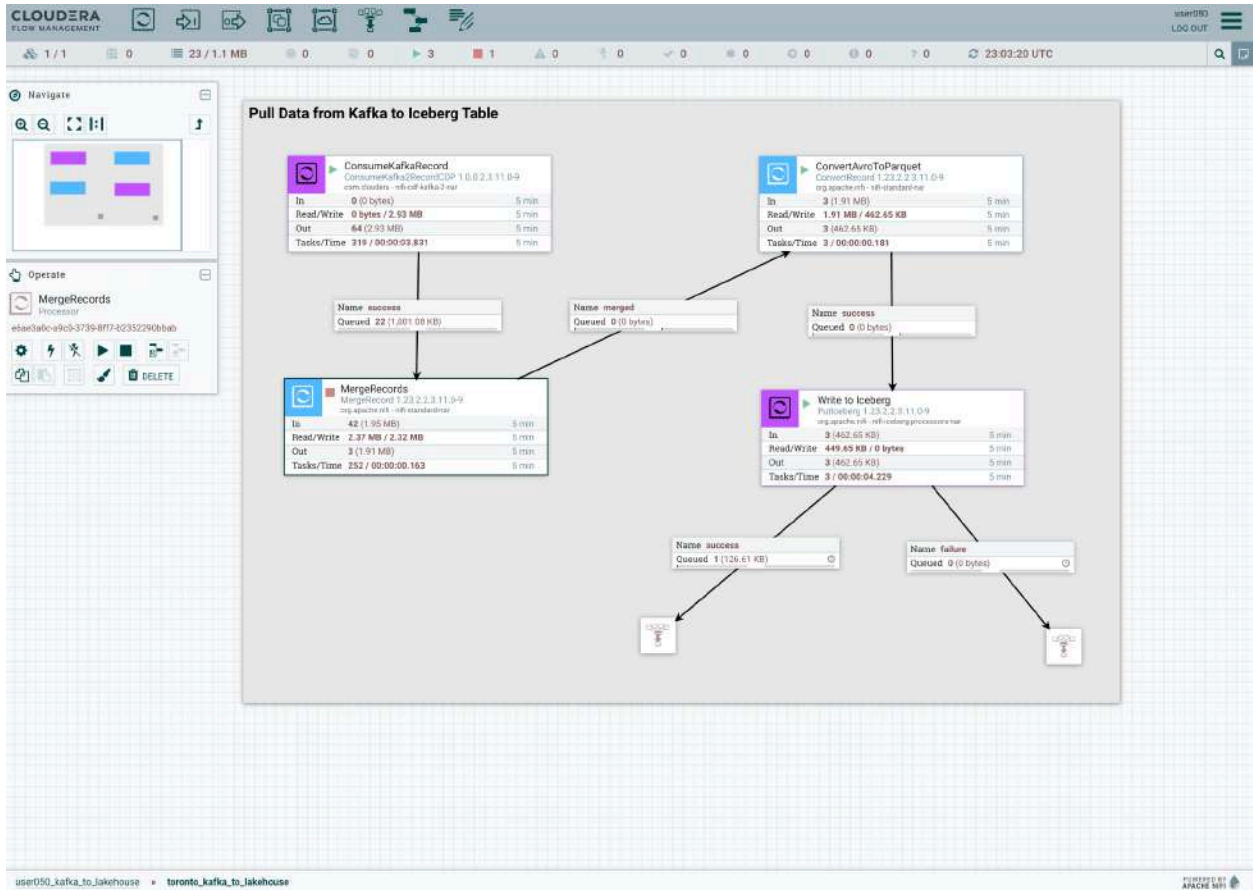
From the Out field in every processor, you can see that data has flowed through in the past 5 minutes. You have already consumed data from Kafka and to Iceberg!

17. Flow Management allows us to see and access data in motion during the execution of the data flow. Between Processors **ConsumeKafkaRecord** (just started) and **MergeRecords**, there is a connection. This connection is what joins the Processors and transmits data from one to the other, and you can check how much data is queued at every step of the process.

Let's see this in action by building up the queue. First, right-click on **MergeRecords** processor and click **Stop**.

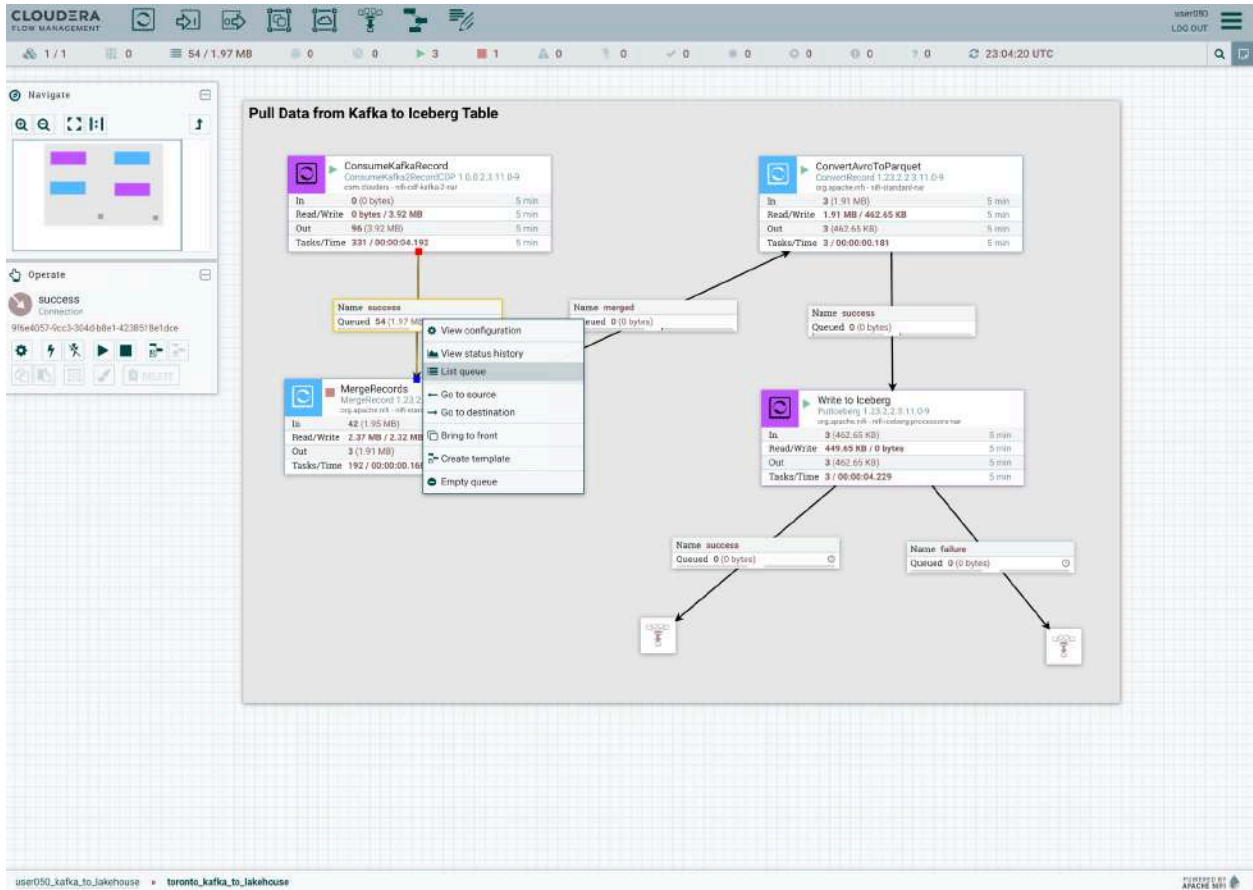


18. You will see data start to queue up in the connector shortly after stopping the MergeRecords processor.



You can refresh the counter by pressing the Ctrl+R (Windows) or Command+R (Mac) combination on the keyboard.

This will allow the current metrics of the entire data stream to be updated. At some point there should be a number next to the legend **Queued** in the connection between **ConsumeKafkaRecord** and **MergeRecords**. To see the queued data, right-click on the connection and click on the option **List Queue**, opening a popup window.



19. The next popup window lists the queued data. Click on the information icon (i) that appears on the left side to view the events.

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-8203-3ec27a62ee...	2055d337-695f-4c6d-8203-3ec27a62ee...	278.24 KB	00:00:12.787	00:00:13.068	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
2	510c8074-6798-4199-a228-ad7894aca9...	510c8074-6798-4199-a228-ad7894aca9...	263.60 KB	00:00:11.664	00:00:11.735	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
3	cad12a7c-e301-439c-85d3-a53bf0f13a2a	cad12a7c-e301-439c-85d3-a53bf0f13a2a	265.48 KB	00:00:11.575	00:00:11.647	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
4	01ee7d23-8a54-4a2b-a29c-a3f965b3c887	01ee7d23-8a54-4a2b-a29c-a3f965b3c887	133.37 KB	00:00:11.527	00:00:11.567	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...

Last updated: 22:59:59 UTC

20. Once the FlowFile detail window appears, click on the button **VIEW** to open the content of consumed events.

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-8203-3ec27a62ee...	2055d337-695f-4c6d-8203-3ec27a62ee...	278.24 KB	00:00:12.787	00:00:13.068	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
2	510c8074-6798-4199-a228-ad7894aca9...	510c8074-6798-4199-a228-ad7894aca9...	263.60 KB	00:00:11.664	00:00:11.735	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
3	cad12a7c-e301-439c-85d3-a53bf0f13a2a	cad12a7c-e301-439c-85d3-a53bf0f13a2a	265.48 KB	00:00:11.575	00:00:11.647	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...
4	01ee7d23-8a54-4a2b-a29c-a3f965b3c887	01ee7d23-8a54-4a2b-a29c-a3f965b3c887	133.37 KB	00:00:11.527	00:00:11.567	No	dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.clu...

Last updated: 22:59:59 UTC

FlowFile

DETAILS ATTRIBUTES

FlowFile Details

UUID: 2055d337-695f-4c6d-8203-3ec27a62ee

Filename: 2055d337-695f-4c6d-8203-3ec27a62ee

File Size: 278.24 KB

Queue Position: 0

Queued Duration: 00:00:19.804

Lineage Duration: 00:00:19.815

Penalized: No

Node Address: dfx-rfb-0-dfx-rfb-dfx-user050-na.svc.cluster.local:8443

Content Claim

Container: default

Section: 1

Identifier: 1684623047706-1

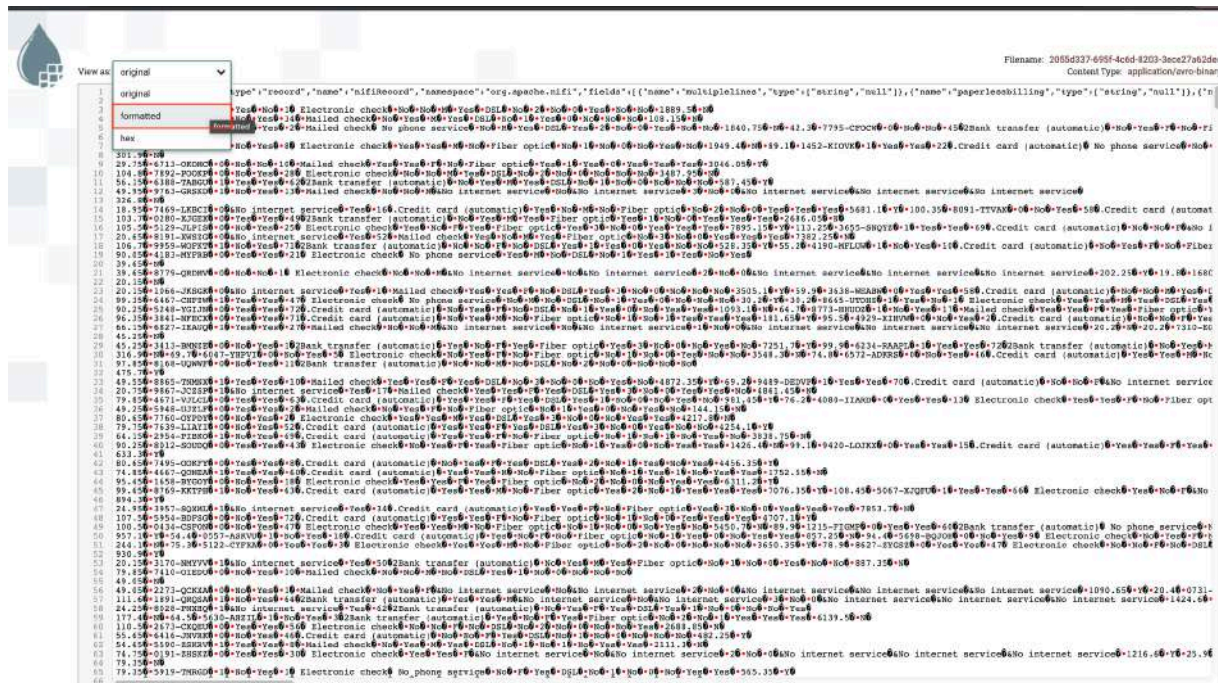
Offset: 0

Size: 278.24 KB

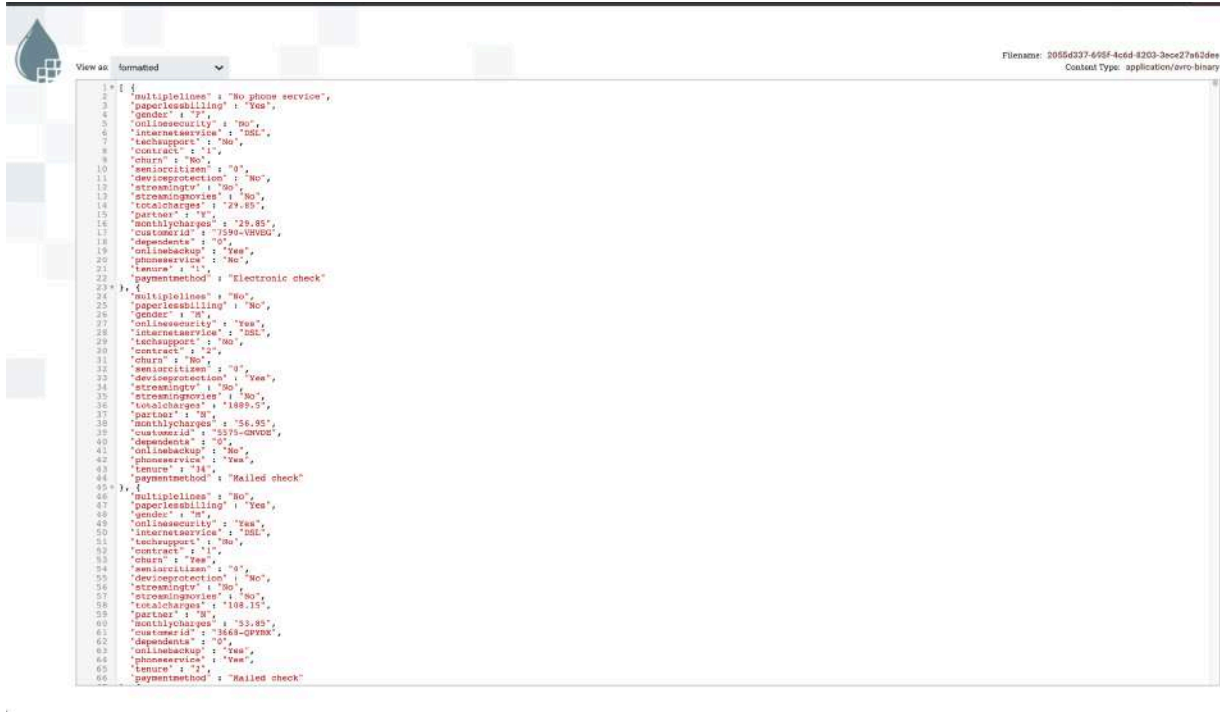
DOWNLOAD VIEW

OK

21. The new window that opens shows the data of the FlowFile content. Being in AVRO format, it is not fully readable. A deserializer must be selected to correctly display the data. For this, in the upper left, select the option **formatted** from the menu **View as**.

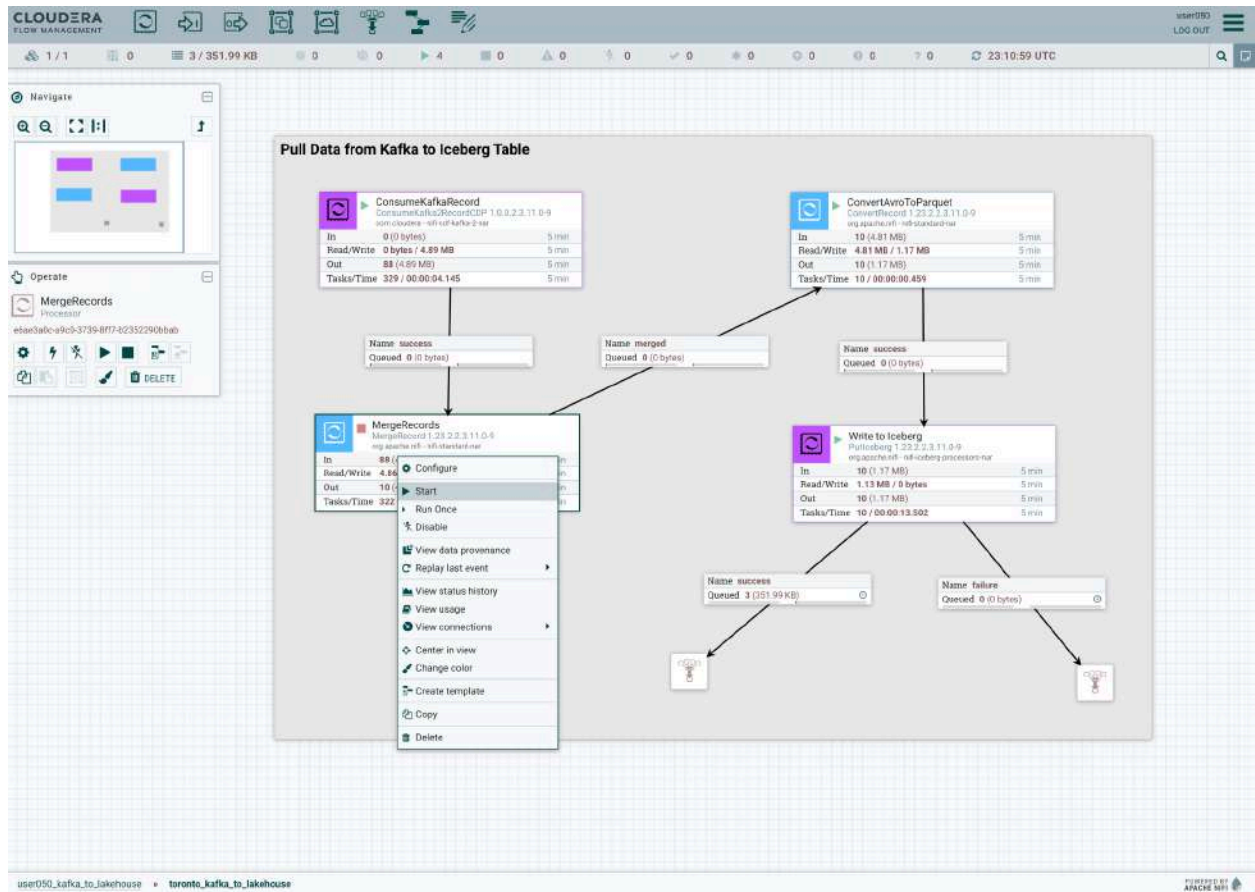


22. Now you can display the data correctly. Notice that the fields or attributes indicated at the beginning of the workshop appear. You can close that FlowFile window and the popups, returning to the canvas with the four Processors.

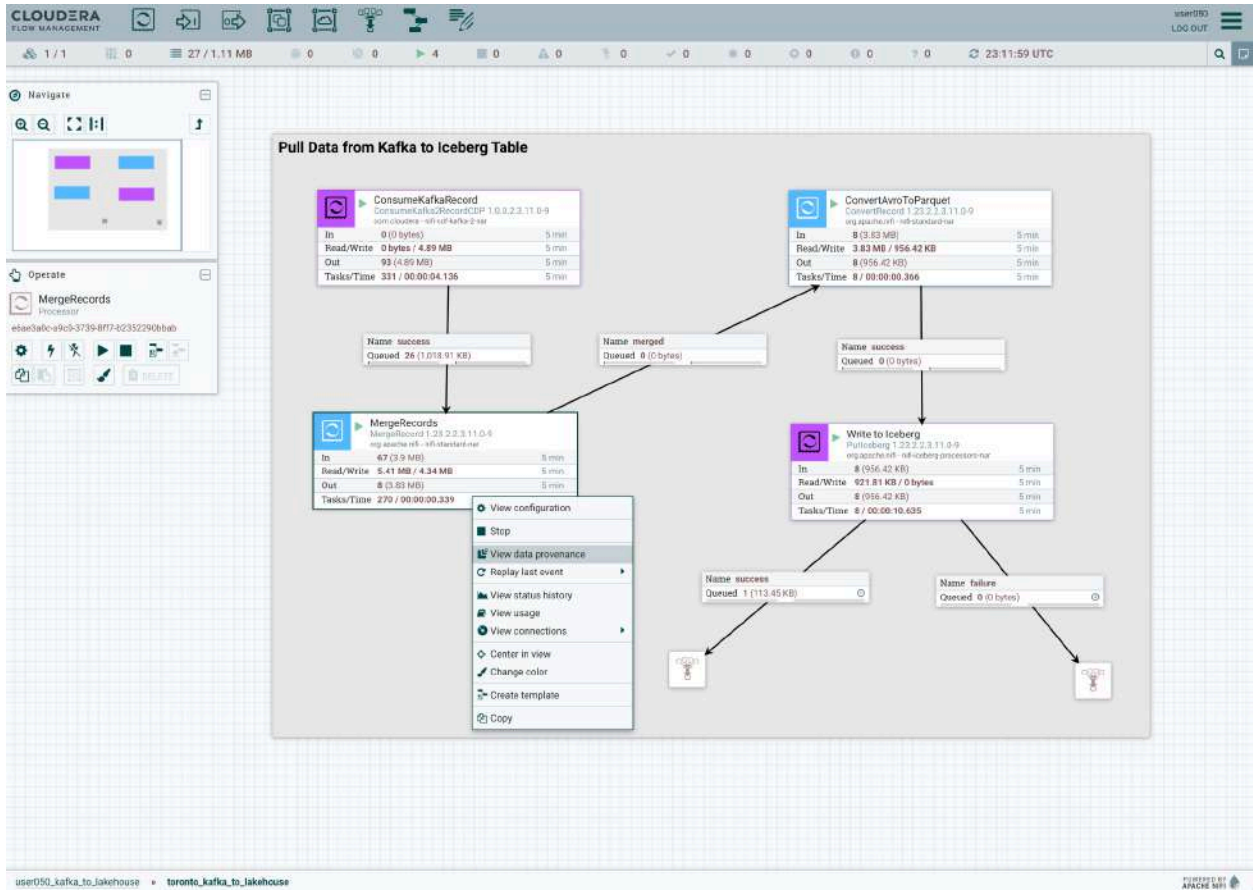


```
1 {
2   "multipleslines": "No phone service",
3   "paperlessbiling": "Yes",
4   "gender": "M",
5   "onlinesecurity": "DSL",
6   "internetsevice": "DSL",
7   "techsupport": "No",
8   "contract": "1",
9   "churn": "No",
10  "seniorcitizen": "0",
11  "deviceprotection": "No",
12  "streamingtv": "No",
13  "streamingmovies": "No",
14  "totalcharges": "29.85",
15  "partner": "Y",
16  "monthlycharges": "29.85",
17  "customerid": "7599-VHVDG",
18  "dependents": "0",
19  "onlinebackup": "Yes",
20  "phoneservice": "No",
21  "tenure": "1",
22  "paymentmethod": "Electronic check"
23 },
24 {
25   "multipleslines": "No",
26   "paperlessbiling": "No",
27   "gender": "M",
28   "onlinesecurity": "Yes",
29   "internetsevice": "DSL",
30   "techsupport": "No",
31   "contract": "2",
32   "churn": "No",
33   "seniorcitizen": "0",
34   "deviceprotection": "Yes",
35   "streamingtv": "No",
36   "streamingmovies": "No",
37   "totalcharges": "189.5",
38   "partner": "N",
39   "monthlycharges": "56.95",
40   "customerid": "5575-QHVDG",
41   "dependents": "0",
42   "onlinebackup": "No",
43   "phoneservice": "Yes",
44   "tenure": "14",
45   "paymentmethod": "Railed check"
46 },
47 {
48   "multipleslines": "No",
49   "paperlessbiling": "Yes",
50   "gender": "M",
51   "onlinesecurity": "Yes",
52   "internetsevice": "DSL",
53   "techsupport": "No",
54   "contract": "1",
55   "churn": "Yes",
56   "seniorcitizen": "0",
57   "deviceprotection": "No",
58   "streamingtv": "No",
59   "streamingmovies": "No",
60   "totalcharges": "108.15",
61   "partner": "N",
62   "monthlycharges": "33.85",
63   "customerid": "3668-QPXXK",
64   "dependents": "0",
65   "onlinebackup": "Yes",
66   "phoneservice": "Yes",
67   "tenure": "2",
68   "paymentmethod": "Railed check"
69 }
```

23. Start the stopped: **MergeRecords** processor again to resume the flow. Remember that you can refresh the flow counters with the combination Control+R or Command+R.



If the previous steps were executed correctly, the connection of the Processor **PutIceberg** to a funnel should be of type **success**.



NiFi Data Provenance

Displaying 989 of 989

Oldest event available: 11/29/2023 22:46:09 UTC

Showing the events that match the specified query. Clear search

Filter	by component name					
Date/Time	Type	FlowFile UUID	Size	Component Name	Component Type	Node
11/29/2023 23:14:46.125 UTC	DROP	42733ef5-db16-49b0-a4c5-b279146...	13.56 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	ee077a74-77b6-4a58-8398-22be85...	16.49 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	315a24eb-945d-41cf-b434-9a86875...	5.81 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	a3bca232-4ac0-4393-98af-30aed5a...	7.75 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	929b38e4-84a7-40ab-bfcb-b1a0748...	141 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	38c923b3-c8e5-463f-94db-2e50486...	3.64 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	272009b5-e401-470a-9a4b-e00834...	88.16 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	e7555479-320e-4c8f-926b-910516f5...	1.24 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	2e9254e9-a72e-4436-925e-eae9557...	97.71 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	df5ae4de-e123-4b0c-8b62-cea68b4...	20.22 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	291224cd-6744-491f-b447-d4c272e8...	24.77 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	DROP	de346e95-58ed-4428-9a27-1152eb...	41.84 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	42733ef5-db16-49b0-a4c5-b279146...	13.56 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	ee077a74-77b6-4a58-8398-22be85...	16.49 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	315a24eb-945d-41cf-b434-9a86875...	5.81 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	a3bca232-4ac0-4393-98af-30aed5a...	7.75 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	929b38e4-84a7-40ab-bfcb-b1a0748...	141 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	38c923b3-c8e5-463f-94db-2e50486...	3.64 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	272009b5-e401-470a-9a4b-e00834...	88.16 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	e7555479-320e-4c8f-926b-910516f5...	1.24 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	2e9254e9-a72e-4436-925e-eae9557...	97.71 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	df5ae4de-e123-4b0c-8b62-cea68b4...	20.22 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	291224cd-6744-491f-b447-d4c272e8...	24.77 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	ATTRIBUTES_MODIFIED	de346e95-58ed-4428-9a27-1152eb...	41.84 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:46.125 UTC	JOIN	4c206ab8-ba1f-42db-91ba-11a5479...	451.87 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	468f3629-5d5a-4c05-914e-603a31b...	74.67 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	6799e6b5-84ce-40b6-baeb-11e1bcb...	78.54 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	76818052-b4a5-4b2e-a909-91e7879...	115.51 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	a61953b0-87e4-4612-a642-d4f77c2...	124.73 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	7c70cd79-339a-4e98-a80a-87077d...	81.42 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	DROP	c6701ca1-2197-4a5c-a5e4-5f22a24...	75.13 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	ATTRIBUTES_MODIFIED	468f3629-5d5a-4c05-914e-603a31b...	74.67 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	ATTRIBUTES_MODIFIED	6799e6b5-84ce-40b6-baeb-11e1bcb...	78.54 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	ATTRIBUTES_MODIFIED	76818052-b4a5-4b2e-a909-91e7879...	115.51 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	ATTRIBUTES_MODIFIED	a61953b0-87e4-4612-a642-d4f77c2...	124.73 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...
11/29/2023 23:14:14.093 UTC	ATTRIBUTES_MODIFIED	7c70cd79-339a-4e98-a80a-87077d...	81.42 KB	MergeRecords	MergeRecord	dfs-nifi-0.dfs-nifi.dfs-user@50-kafka...

Last updated: 23:14:56 UTC