

Agenda :

- a) CDP Overview (20 minutes)
- b) CDW Overview and Benefits (15 minutes)
- c) Hands-on (step-by-step below) (50 minutes)
 - i) Part 1 - Create a Virtual Warehouse and Run Queries (25 minutes)
 - ii) Part 2 - Data Visualization (15 minutes)
 - iii) Part 3 - Import a File into a Table (10 minutes)

Step-by-step instructions:

Part 1 - Create a Virtual Warehouse and Run Queries [25 minutes]

Overview: What is Cloudera Data Warehouse?

We will explore features of Cloudera Data Warehouse (CDW) by performing some data exploration and create dashboards to share our results to a wider audience

We will be taking a look at a generated data set from a mock airline company containing flights information from its fleet of aircraft.

A virtual warehouse represents virtual compute resources to access data that is stored in a database catalog. This lets you create or destroy compute resources, auto-scale, or separate resources across different workloads, all running on the same underlying data.

CDW let's you choose from a set of default resources based on your predicted workload as well as give you fine grained control over autoscaling and timeout features so you can fine tune your system to be most cost effective.

Purpose: Create a virtual warehouse and run queries, answering the questions below:

- What are the top 5 visited destinations by year from (1995-2008)?
- What are the top 10 routes (origin and dest) that have seen maximum diversions?
- Which three months have seen the most number of cancellation due to bad weather?

- 1) Open CDP, using the “admin” user within the Test Drive link.

Your link should look something like (remember click the link in your email not the link below)

http://login.trycdp.com/auth/realm/trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X

*xx represents the trial user #

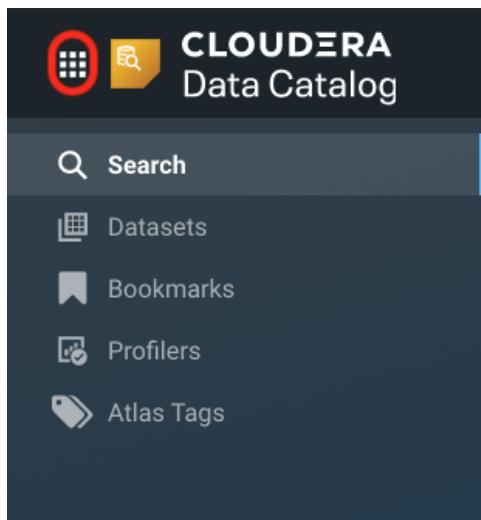
*X represents the password

2) Click the “Data Warehouse” within the CDP Home Screen



How do you get to the CDP Home Screen?

- From any experience such as “Data Catalog”, click the 9 square at the top left and then click “Home”



X



Home

3) Click the “+” at the top right next to “Virtual Warehouses”

Virtual Warehouses | 1

♦

New Virtual Warehouse

Name *

Enter Virtual Warehouse Name

Type *

HIVE IMPALA

Database Catalog *

cdptrialuser24-dl-default

Size *

-- select an option --

default-vw
compute-1611103491-4hbp
cdptrialuser24-dl-default

Stopped

NODE COUNT TOTAL CORES TOTAL MEMORY TYPE

0 12 56 GB HIVE COMPACTOR

4) Enter a name for your New Virtual Warehouse

Virtual Warehouses | 1

New Virtual Warehouse

Name *

 testvirtualwarehouse1

Type *

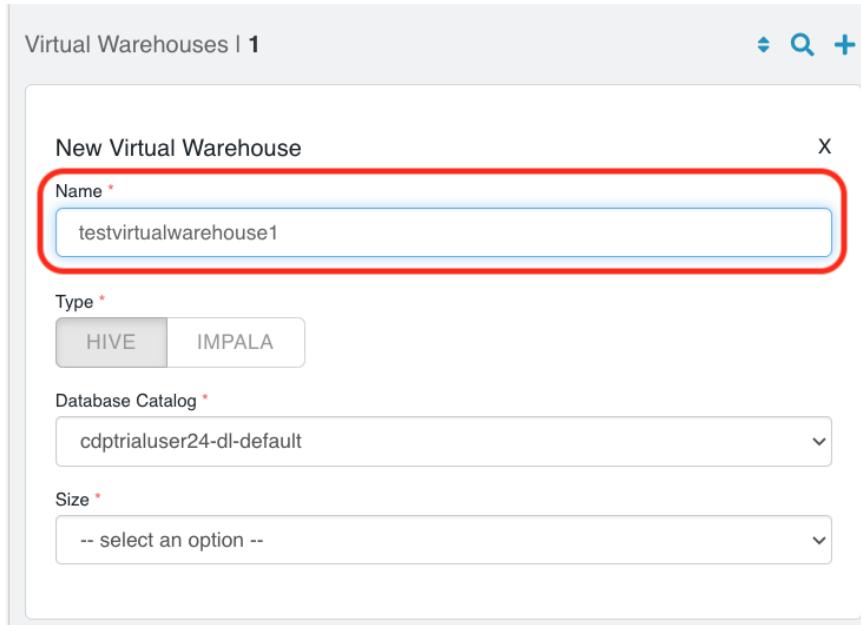
HIVE IMPALA

Database Catalog *

cdptrialuser24-dl-default

Size *

-- select an option --



5) Select the Size of “xsmall - 2 Executor Nodes”

*How do I choose a size? Initial concurrent users

Virtual Warehouses | 1

New Virtual Warehouse

Name *

 testvirtualwarehouse1

Type *

HIVE IMPALA

Database Catalog *

cdptrialuser24-dl-default

Size *

-- select an option --

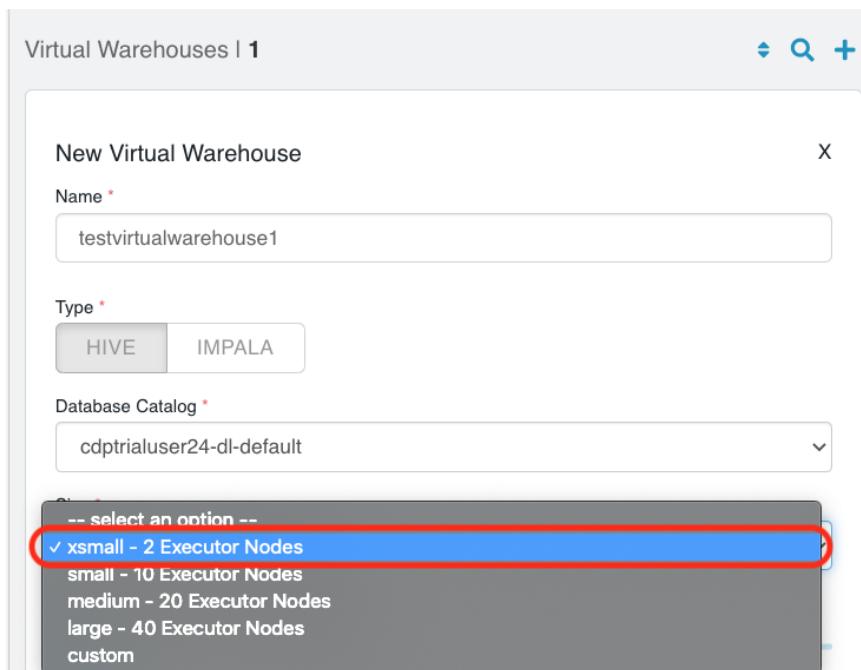
✓ xsmall - 2 Executor Nodes

small - 10 Executor Nodes

medium - 20 Executor Nodes

large - 40 Executor Nodes

custom



6) Set the AutoSuspend Timeout (in seconds) between 4500 and 5500:

*What is AutoSuspend Timeout? Automatically spin-down unused resources after timeout occurs.

Virtual Warehouses | 1

New Virtual Warehouse

Name *

Type *

HIVE IMPALA

Database Catalog *

cdptrialuser24-dl-default

Size *

xsmall - 2 Executor Nodes

AutoSuspend Timeout (in seconds): 5000

A slider for "AutoSuspend Timeout (in seconds)" ranging from 0 to 7000. The slider is set to 5000. The entire input field and slider are highlighted with a red rounded rectangle.

- 7) Choose “Install Data Visualization” to be on
*Allowing for Data Visualizations in Part 3

Virtual Warehouses | 2

New Virtual Warehouse

Name *

Type *

HIVE IMPALA

Database Catalog *

cdptrialuser24-dl-default

Size *

xsmall - 2 Executor Nodes

AutoSuspend Timeout (in seconds): 5000

Concurrency Autoscaling ⓘ

Nodes: Min:2, Max:6

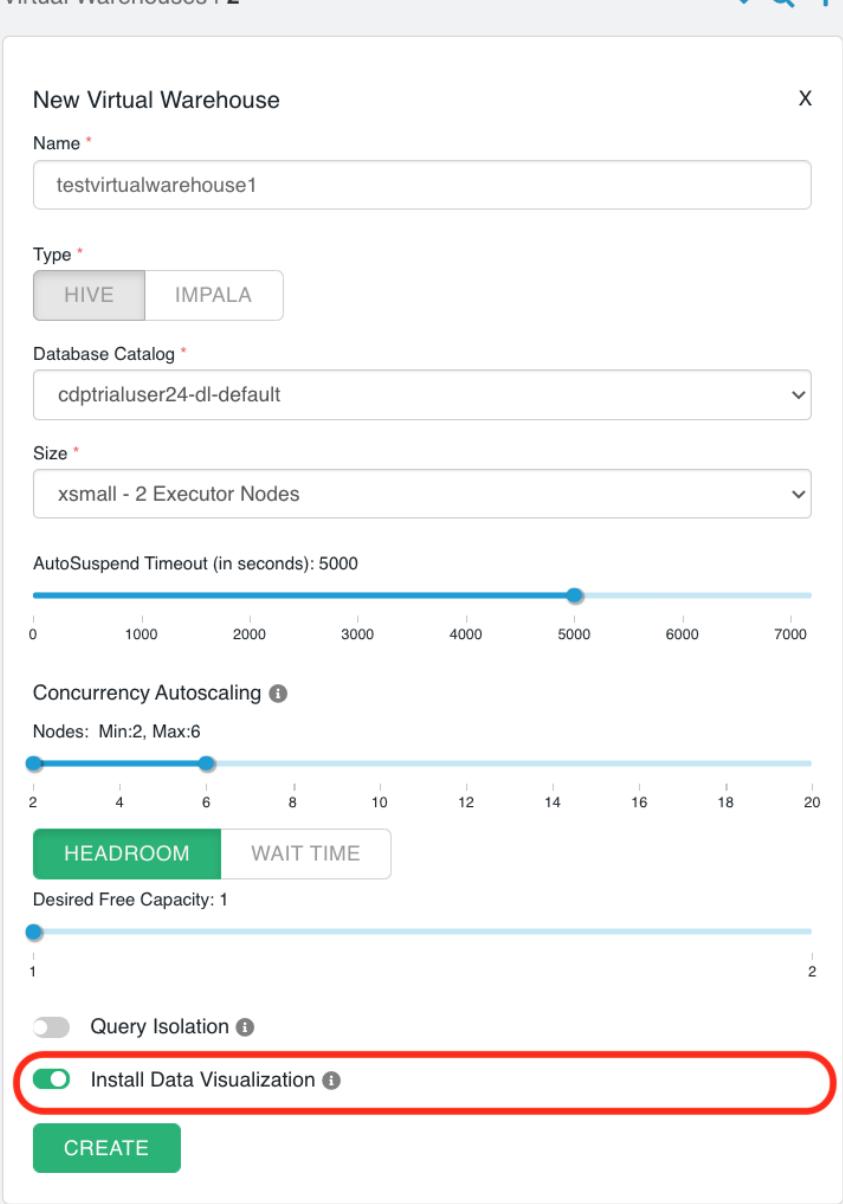
HEADROOM WAIT TIME

Desired Free Capacity: 1

Query Isolation ⓘ

Install Data Visualization ⓘ

CREATE



8) Click "Create" to create your Virtual Warehouse

*Allow for approximately 5 minutes for your Virtual Warehouse to become available for use



When available for use, “Starting” will change to “Running” as shown below

testvirtualwarehouse1
compute-1611179792-vz49
cdptrialuser24-dl-default
Starting

checking if query-coordinator-0 statefulset is ready with at least 1 ready replica(s) (config-id: 7647c82f-8b37-4593-80e9-058f1f928b31 version: 7.2.8.0-24)

NODE COUNT	TOTAL CORES	TOTAL MEMORY	TYPE
2	38	292 GB	HIVE DATA VISUALIZATION

testvirtualwarehouse1
compute-1611179792-vz49
cdptrialuser24-dl-default
Running

NODE COUNT	TOTAL CORES	TOTAL MEMORY	TYPE
2	38	292 GB	HIVE DATA VISUALIZATION

9) Once your Virtual Warehouse is “Running”, click the line in the top right and then click “Open DAS”

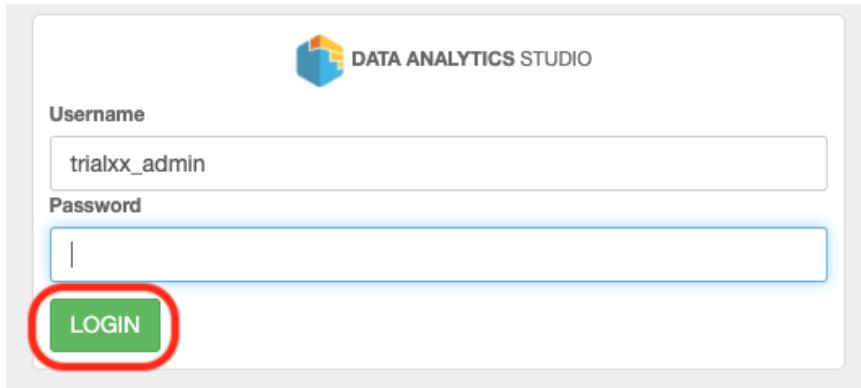
Virtual Warehouses | 3

testvirtualwarehouse1	compute-1611179792-vz49	cdptrialuser24-dl-default	Running
2	38	292 GB	HIVE DA

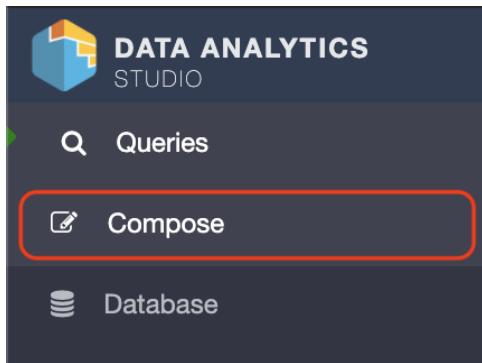
mschoeni-iso-1	compute-1611173596-db1v	cdptrialuser24-dl-default	Stopped
----------------	-------------------------	---------------------------	---------

Suspend
Clone
Edit
Delete
Upgrade
Copy JDBC URL
Download JDBC Jar
Open DAS
Open Data Visualization

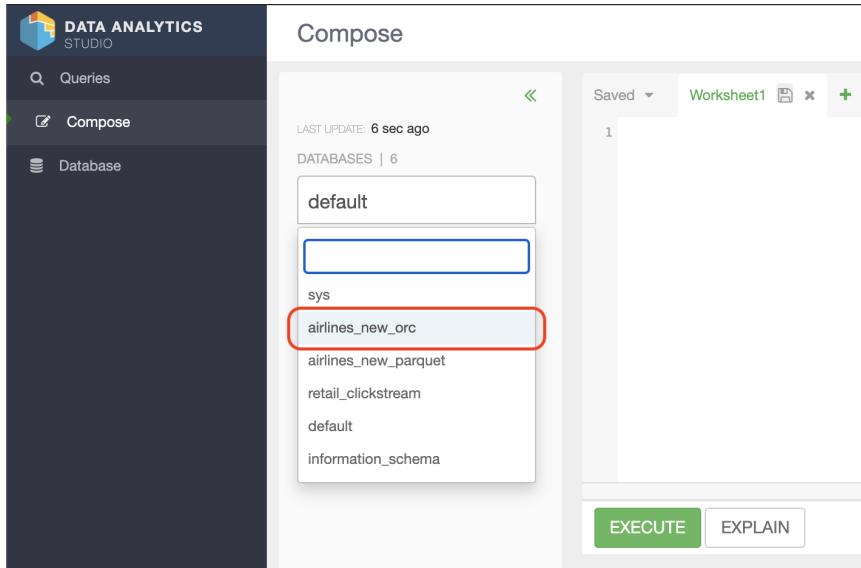
10) Enter the login information from step #1 above using the user, then click “LOGIN”
*Changing “trialxx_admin” to the trail user you’re using and password defined by “X” in #1 above



11) Click on “Compose”, to write the queries below to answer questions on the table “flights”



12) Choose the database “airlines_new_orc”.



13) Enter the following query in Worksheet1, answering the question “show me the top 5 visited destination by year from (1995-2008)”

```
SELECT dest,year,COUNT(dest) as Times_Visited FROM flights  
GROUP BY dest,year  
ORDER BY Times_Visited DESC  
LIMIT 5;
```

Saved ▾ Worksheet1*

```
1 SELECT dest,year,COUNT(dest) as Times_Visited FROM flights  
2 GROUP BY dest,year  
3 ORDER BY Times_Visited DESC  
4 LIMIT 5;
```

EXECUTE EXPLAIN

- 14) Click “EXECUTE” to execute the query, answering the question “show me the top 5 visited destination by year from (1995-2008)”

The screenshot shows a database worksheet titled "Worksheet1". At the top, there are tabs for "Saved", "Worksheet1*", a refresh icon, a close icon, and a plus sign icon. Below the tabs is a code editor containing the following SQL query:

```
1 SELECT dest,year,COUNT(dest) as Times_Visited FROM flights
2 GROUP BY dest,year
3 ORDER BY Times_Visited DESC
4 LIMIT 5;
```

At the bottom of the worksheet are two buttons: "EXECUTE" (highlighted with a red box) and "EXPLAIN".

- 15) Click the download button on the top right, to download the results as a CSV file

The screenshot shows a results table titled "Results". The table has three columns: DEST, YEAR, and TIMES_VISITED. The data is as follows:

DEST	YEAR	TIMES_VISITED
ATL	2005	428800
ATL	2004	416989
ATL	2008	414521
ATL	2007	413805
ATL	2006	404829

- 16) Going back to “Worksheet 1”, click the “+” to add another Worksheet for the next query

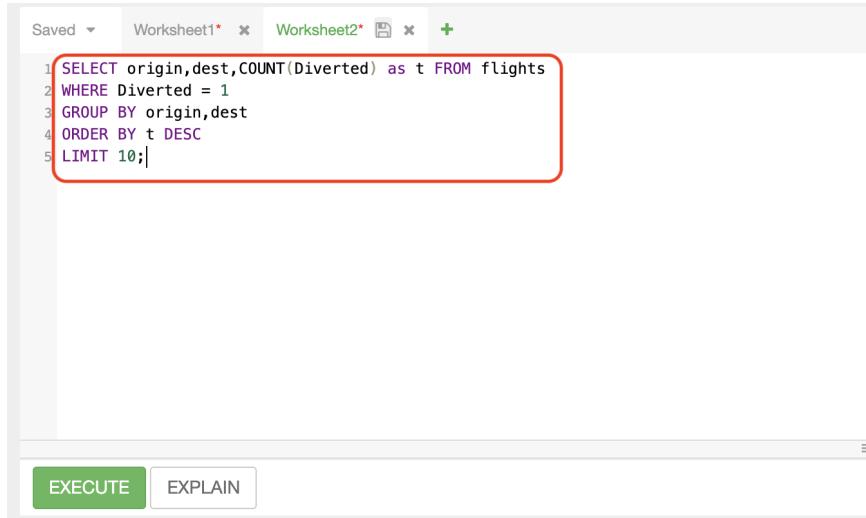
The screenshot shows a database worksheet titled "Worksheet1". At the top, there are tabs for "Saved", "Worksheet1*" (highlighted with a red box), a refresh icon, a close icon, and a plus sign icon (highlighted with a red box). Below the tabs is a code editor containing the same SQL query as in the previous screenshot:

```
1 SELECT dest,year,COUNT(dest) as Times_Visited FROM flights
2 GROUP BY dest,year
3 ORDER BY Times_Visited DESC
4 LIMIT 5;
```

At the bottom of the worksheet are two buttons: "EXECUTE" and "EXPLAIN".

17) In “Worksheet 2”, Choose the database “airlines_new_orc” then copy-and-paste the following query, answering the question “What are the top 10 routes (origin and dest) that have seen maximum diversions?”

```
SELECT origin,dest,COUNT(Diverted) as t FROM flights  
WHERE Diverted = 1  
GROUP BY origin,dest  
ORDER BY t DESC  
LIMIT 10;
```

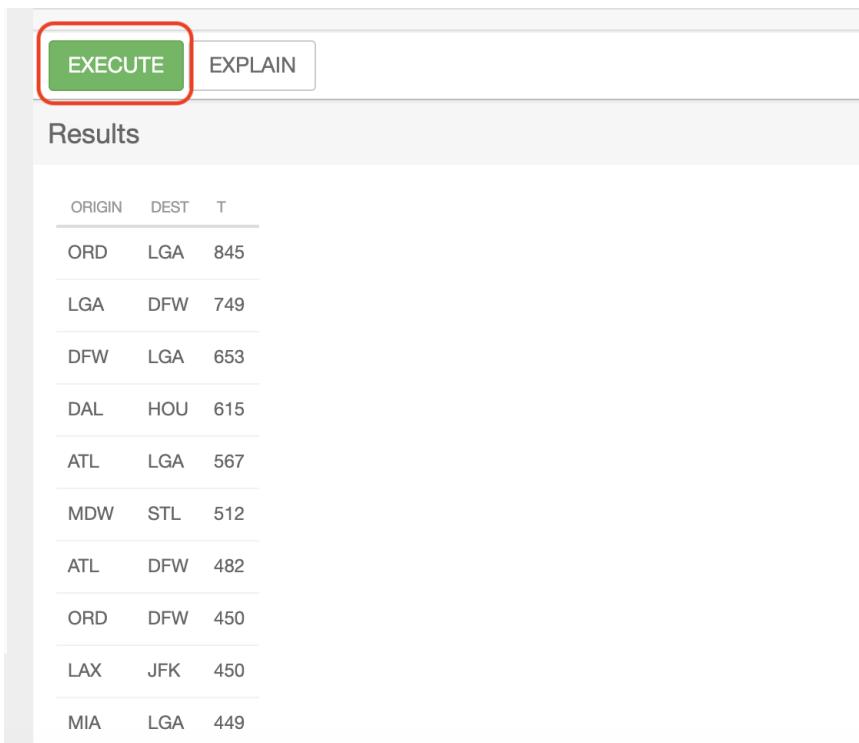


The screenshot shows a database interface with two worksheets open: Worksheet1 and Worksheet2. Worksheet2 is active and contains the following SQL query:

```
1 SELECT origin,dest,COUNT(Diverted) as t FROM flights  
2 WHERE Diverted = 1  
3 GROUP BY origin,dest  
4 ORDER BY t DESC  
5 LIMIT 10;
```

The code is highlighted with a red rectangle. At the bottom of the interface, there are two buttons: "EXECUTE" and "EXPLAIN".

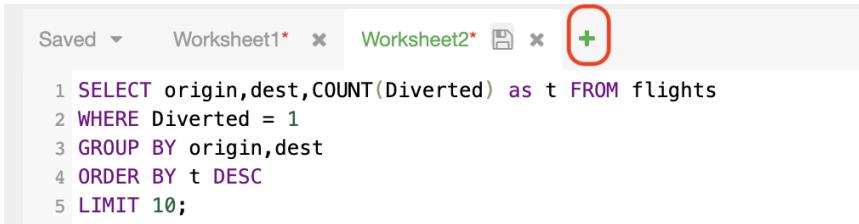
- 18) Click “EXECUTE” to execute the query, answering the question “What are the top 10 routes (origin and dest) that have seen maximum diversions?”



The screenshot shows a user interface for executing SQL queries. At the top, there are two buttons: "EXECUTE" (highlighted with a red box) and "EXPLAIN". Below them is a section titled "Results" containing a table of flight data:

ORIGIN	DEST	T
ORD	LGA	845
LGA	DFW	749
DFW	LGA	653
DAL	HOU	615
ATL	LGA	567
MDW	STL	512
ATL	DFW	482
ORD	DFW	450
LAX	JFK	450
MIA	LGA	449

- 19) Going back to “Worksheet 2”, click the “+” to add another Worksheet for the final query



The screenshot shows a workspace with multiple worksheets open. The tabs are labeled "Saved", "Worksheet1*", "Worksheet2*", and a new tab represented by a plus sign icon. Below the tabs, a query is displayed in a code editor:

```
1 SELECT origin,dest,COUNT(Diverted) as t FROM flights
2 WHERE Diverted = 1
3 GROUP BY origin,dest
4 ORDER BY t DESC
5 LIMIT 10;
```

- 20) In “Worksheet 3”, Choose the database “airlines_new_orc” then copy-and-paste the following query, answering the question “Which three months have seen the most number of cancellation due to bad weather?”

```
SELECT month,COUNT(Cancelled) as num_of_cancellations FROM flights
WHERE Cancelled = 1 AND CancellationCode = 'B'
GROUP BY month
ORDER BY num_of_cancellations DESC
LIMIT 3;
```

21) Click “EXECUTE” to execute the query, answering the question “Which three months have seen the most number of cancellation due to bad weather?”

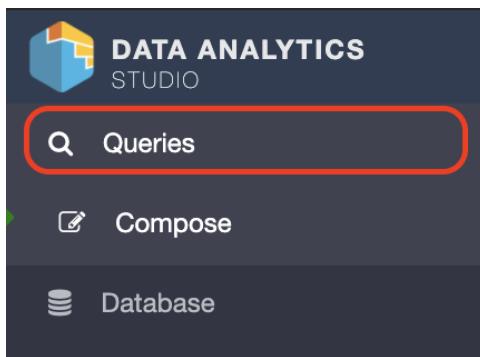
The screenshot shows a user interface for executing SQL queries. At the top, there are two buttons: "EXECUTE" (highlighted with a red box) and "EXPLAIN". Below them is a button labeled "Execute Query". The main area is titled "Results" and contains a table with the following data:

MONTH	NUM_OF_CANCELLATIONS
12	48868
1	42641
2	38234

22) Click “EXECUTE” a second time - this will lead us to our last portion of Part 2

23) Click on “Queries” on the top left navigation bar

*We'll look at our query history



24) Click the “Compare” on the right of your last query run (query at the top)

QUERIES (159)											COMPOSE QUERY	
QUERY	STATUS	QUEUE	USER	TABLES READ	TABLES WRITTEN	START TIME	DURATION	DAG ID	ACTIONS			
SELECT month,COUNT(Cancelled) as ...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	8 seconds ago	00:00:00	Not Available!				
SELECT month,COUNT(Cancelled) as ...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	4 minutes ago	00:00:02	dag_161123649				
SELECT origin,dest,COUNT(Diverted) as...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	8 minutes ago	00:00:03	dag_161123649				
SELECT dest,year,COUNT(dest) as Time...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	20 minutes ago	00:00:00	Not Available!				
SELECT dest,year,COUNT(dest) as Time...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	21 minutes ago	00:00:00	Not Available!				

25) Click the “Compare” on the right of the query (second to the top)

QUERIES (159)											COMPOSE QUERY	
QUERY	STATUS	QUEUE	USER	TABLES READ	TABLES WRITTEN	START TIME	DURATION	DAG ID	ACTIONS			
SELECT month,COUNT(Cancelled) as ...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	4 minutes ago	00:00:00	Not Available!				
SELECT month,COUNT(Cancelled) as ...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	8 minutes ago	00:00:02	dag_161123649				
SELECT origin,dest,COUNT(Diverted) as...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	13 minutes ago	00:00:03	dag_161123649				
SELECT dest,year,COUNT(dest) as Time...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	24 minutes ago	00:00:00	Not Available!				
SELECT dest,year,COUNT(dest) as Time...	SUCCESS	None	trial24_admin	flights (airlines_new_o...)	Not Available!	25 minutes ago	00:00:00	Not Available!				

26) Click on the “COMPARE” button to compare the two queries

The screenshot shows the CDW interface with two queries side-by-side. Both queries are identical: "SELECT month,COUNT(Cancelled) as num_of_cancellations". The right-hand query has a red box around its "COMPARE" button. Below the queries is a green button labeled "Compare two queries".

27) Notice the run duration is different between the two, let's find out why

Query Details - A

QUERY ID
hive_20210121153926_e3a56b9d-71f2-45dc-b23e-2c2e1146d61e

USER
trial24_admin

STATUS
✓ SUCCESS

START TIME
21 Jan 2021 09:39:26

END TIME
21 Jan 2021 09:39:26

DURATION
118ms

Query Details - B

QUERY ID
hive_20210121153513_37aefec1-0284-4897-bfbb-bf9bb5797252

USER
trial24_admin

STATUS
✓ SUCCESS

START TIME
21 Jan 2021 09:35:13

END TIME
21 Jan 2021 09:35:15

DURATION
2s 311ms

28) Click on “timeline” at the top

The screenshot shows the CDW interface with a navigation bar at the top. The "TIMELINE" tab is highlighted with a red box.

As shown, the faster query only did “compile and parse”, while the slower query did “compile, parse, build dag, submit dag, submit to running, run dag”. Why? Because if you run the same exact query twice, the results are cached (if the data didn’t change). CDW knows if the data changed.



Part 2 - Data Visualization [25 minutes]

Overview: What is Data Visualization and how do we use it with our data?

Purpose: Create visualization using the flight information answering the question (visually with a density graph):

- What were the most number of flights from destination to origin between (1995-2008) - Route Density

- 1) Open CDP, using the “admin” user within the Test Drive link.

Your link should look something like (remember click the link in your email not the link below)

http://login.trycdp.com/auth/realms/trycdp-trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X

*xx represents the trial user #

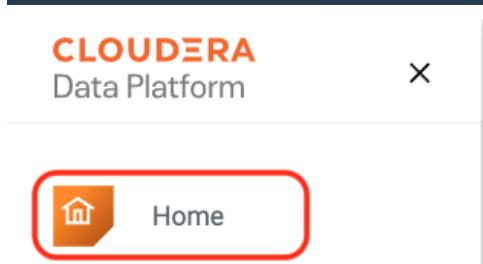
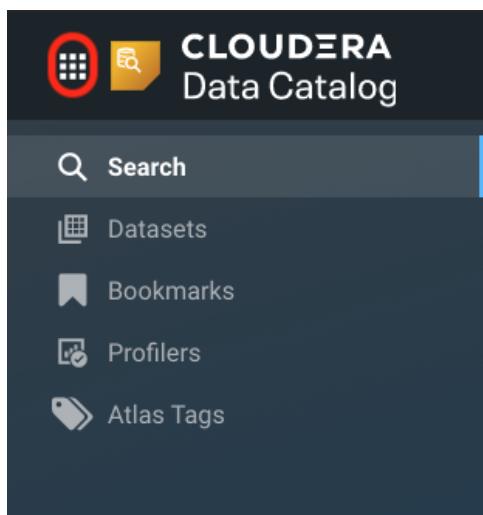
*X represents the password

2) Click the “Data Warehouse” within the CDP Home Screen



How do you get to the CDP Home Screen?

- From any experience such as “Data Catalog”, click the 9 square at the top left and then click “Home”



3) Click “Open Data Visualization” on your existing “Running” Virtual Warehouse

The screenshot shows the Cloudera Manager interface with the 'Virtual Warehouses' tab selected. A context menu is open over the 'testvirtualwarehouse1' entry, which is listed as 'Running'. The menu options include Suspend, Clone, Edit, Delete, Upgrade, Copy JDBC URL, Download JDBC Jar, Open DAS, and Open Data Visualization. The 'Open Data Visualization' option is highlighted with a red box.

NODE COUNT	TOTAL CORES	TOTAL MEMORY	TYPE
2	38	292 GB	HIVE DA

NODE COUNT	TOTAL CORES	TOTAL MEMORY
mschoeni-iso-1	compute-1611173596-dbtv	cdptrialuser24-dl-default

4) Enter the login information from step #1 above using the user, then click “LOGIN”

*Changing “trialxx_admin” to the trail user you’re using and password defined by “X” in #1 above

The screenshot shows the Cloudera Data Visualization login screen. It features a dark header with the text 'CLOUDERA Data Visualization'. Below it is a light-colored form with the word 'LOGIN' at the top. It has fields for 'Username' (containing 'trialxx_admin') and 'Password'. Below the password field, the message 'Invalid login' is displayed in red. There is also a link 'Forgot your password?'. At the bottom, there is a checkbox 'Remember me on this computer' with a checked box, and a large orange 'LOGIN' button which is highlighted with a red box.

5) Click "DATA" the top navigation bar

The screenshot shows the Cloudera Data Visualization web application. At the top, there is a navigation bar with three tabs: HOME, VISUALS, and DATA. The DATA tab is highlighted with a red circle. Below the navigation bar is a sidebar on the left containing links for 'All Connections' (selected), 'Default Hive VW' (highlighted with a red circle), and 'samples'. In the center, there is a 'Datasets' section with a count of 11 datasets. A search bar at the top right contains the placeholder text 'find titles, viz types, datasets, authors...'. The main area below the datasets section has columns for 'Title/Table', 'Created', and 'Last Modified'.

6) Click "Default Hive VW" to add our dataset

This screenshot is similar to the previous one, but the 'Default Hive VW' link in the sidebar is highlighted with a red circle. The rest of the interface remains the same, including the navigation bar, search bar, and central datasets section.

7) Click "NEW DATASET" to add our "flights" data

In this final screenshot, the 'NEW DATASET' button in the top right of the main content area is highlighted with a red circle. The rest of the interface, including the sidebar with 'Default Hive VW' selected, the datasets section, and the navigation bar, remains consistent with the previous screenshots.

8) Enter a name for the Dataset title naming “airline_new_orc.flights”

*Can be any name you choose

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

 airlines_new_orc.flights

Dataset Source

From Table

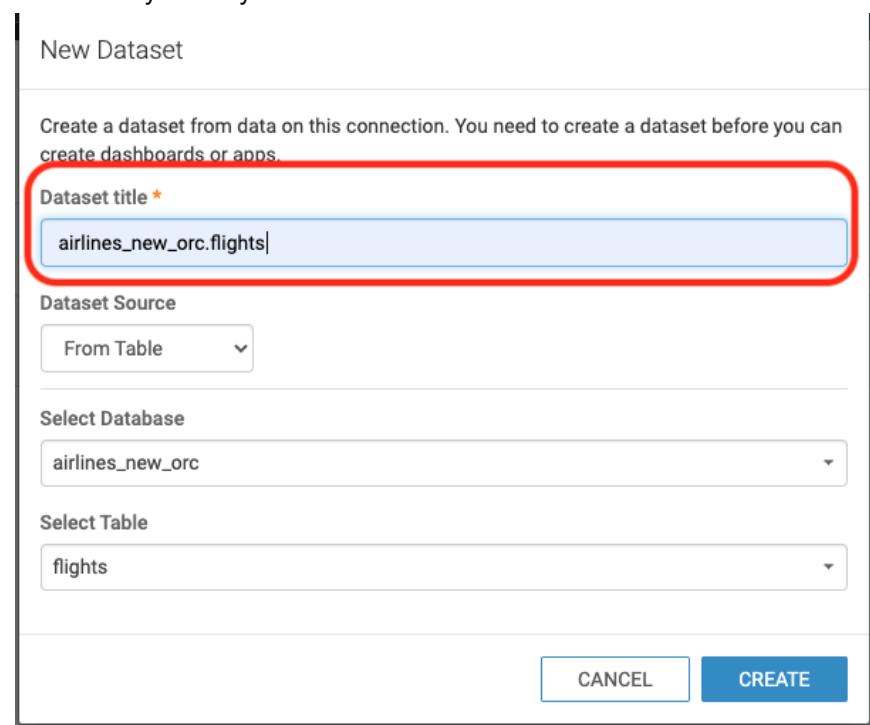
Select Database

airlines_new_orc

Select Table

flights

CANCEL CREATE



9) Choose the database “airlines_new_orc”

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

 airlines_new_orc.flights

Dataset Source

From Table

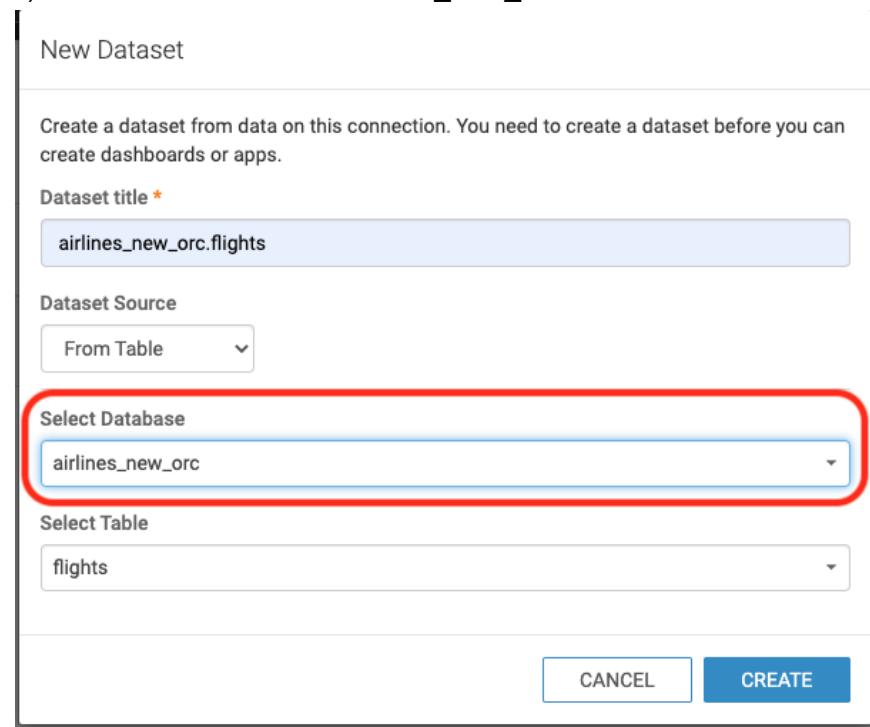
Select Database

airlines_new_orc

Select Table

flights

CANCEL CREATE



10) Choose the table “flights”

*Need to import multiple databases and tables? You’d use Dataset Source = SQL

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Dataset Source

From Table

Select Database

Select Table

 flights

CANCEL CREATE

11) Click “CREATE”

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Dataset Source

From Table

Select Database

Select Table

 flights

CANCEL CREATE

12) Click “+” to create a New Dashboard

The screenshot shows a list of datasets and dashboards. At the top, there are buttons for 'NEW DATASET', 'ADD DATA', and a dropdown. Below this, a section for 'Datasets' lists two entries: 'airlines_new.orc.flights' and 'airlines_new.orc.flights'. To the right of these is a section for 'Dashboards', which contains one entry: 'New Dashboard'. The 'New Dashboard' entry is circled in red. Below this section, there are columns for 'Created' (Jan 21, 2021), 'Last Updated' (a few seconds ago), 'Modified By' (trial24_admin), and '# Visuals' (0).

13) Choose “Treemap” under “VISUALS”

The screenshot shows the 'Dashboard Designer' interface. On the left, the 'VISUALS' panel is open, displaying various visualization icons. One icon, labeled 'Treemap', is highlighted with a red circle. Below the icons, there are sections for 'Dimensions', 'Measures', 'Toolips', 'Filters', and a 'Limit' input field set to 100. A large blue button at the bottom says 'REFRESH VISUAL'. On the right, the 'DATA' panel shows a dataset named 'airlines_new.orc.flights' with a sample mode set to 'OFF'. Below the dataset, there are sections for 'Dimensions' (containing fields like flights, uniquecarrier, tailnum, origin, dest, cancellationcode, diverted) and 'Measures' (containing fields like Record Count, month, dayofmonth, dayofweek, deptime, crsdeptime, arrtime, crsarstime, flightnum, actualelapsedtime). A vertical sidebar on the right contains navigation links: 'DASH.', 'Visuals' (which is currently selected and highlighted in blue), 'Filters', 'Settings', 'Style', 'VISUAL', 'Build' (which is also highlighted in blue), 'Settings', and 'Style'.

- 14) Drag-and-drop both “dest” and origin” from Dimensions->Flights into Dimensions under Visuals

The screenshot shows the Tableau Dashboard Designer interface. On the left, the 'VISUALS' pane contains various visualization options like Table, Bar Chart, Line Chart, etc. Below it, the 'Dimensions' section is highlighted with a red box, showing fields such as dest and origin. The 'Measures' section below it lists fields like Record Count, month, dayofmonth, etc. On the right, the 'DATA' pane shows a connection to 'airlines_new_orc.flights' with sample mode off. A search bar and a 'Dimensions' section with 6 items are also visible. The 'Build' tab is selected in the sidebar. The right sidebar includes sections for Favorites (Airdrop, Recents, Application, Desktop), Documents, Downloads, Locations (Network), and Tags (Red, Orange, Yellow, Green, Blue).

15) Drag-and-drop “Record Count” from Measures->Flights into Measures under Visuals

The screenshot shows the Tableau Dashboard Designer interface. On the left, the 'VISUALS' pane contains a 'Table' visual. Below it, the 'Dimensions' section lists fields: dest, origin, flights, uniquecarrier, tailnum, origin, dest, cancellationcode, and diverted. The 'Measures' section is highlighted with a red box and shows a list of measures: sum(1), # Record Count, # month, # dayofmonth, # dayofweek, # deptime, # crsdeptime, # arrtime, # crsarrtime, # flightnum, # actualelapsedtime, # crselapsedtime, # airtime, and # arrdelay. The 'Measures' section has a count of 24 items. At the bottom, there are buttons for 'REFRESH VISUAL' and 'BUILD'.

16) Click the right arrow next to Record Count and select “Descending” under Order and Top K

The screenshot shows the Tableau Dashboard Designer interface. On the left, the Visuals shelf has a Treemap icon selected. The main workspace contains a Treemap visualization with two dimensions: dest and origin, and one measure: Record Count. The Record Count field is highlighted with a blue border and has a small arrow icon indicating it can be edited. The right side of the screen is the Field Properties panel for the Record Count field.

FIELD PROPERTIES

- ▶ Date/Time Functions
- ▶ Text Functions
- ▶ Analytic Functions
- ▶ Change Type
- ▶ **Order and Top K** ●

Descending

Ascending

Top K: eg. 100

Bottom K: eg. 100

Top K/Bottom K applies to granular dimensions

[] Enter/Edit Expression

Display Format

Alias ●

Description

Duplicate

Save Expression

✖ Remove

DASH.

- + **Visuals**
- + **Filters**
- ⚙ **Settings**
- 💎 **Style**

VISUAL

- Build
- ⚙ **Settings**
- 💎 **Style**

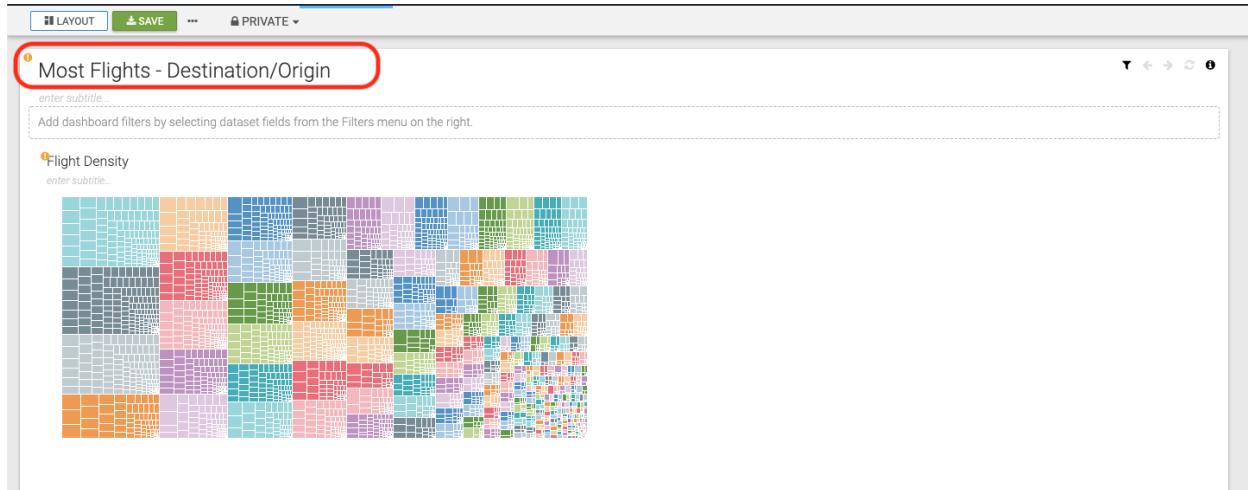
REFRESH VISUAL

17) Click "REFRESH VISUAL"

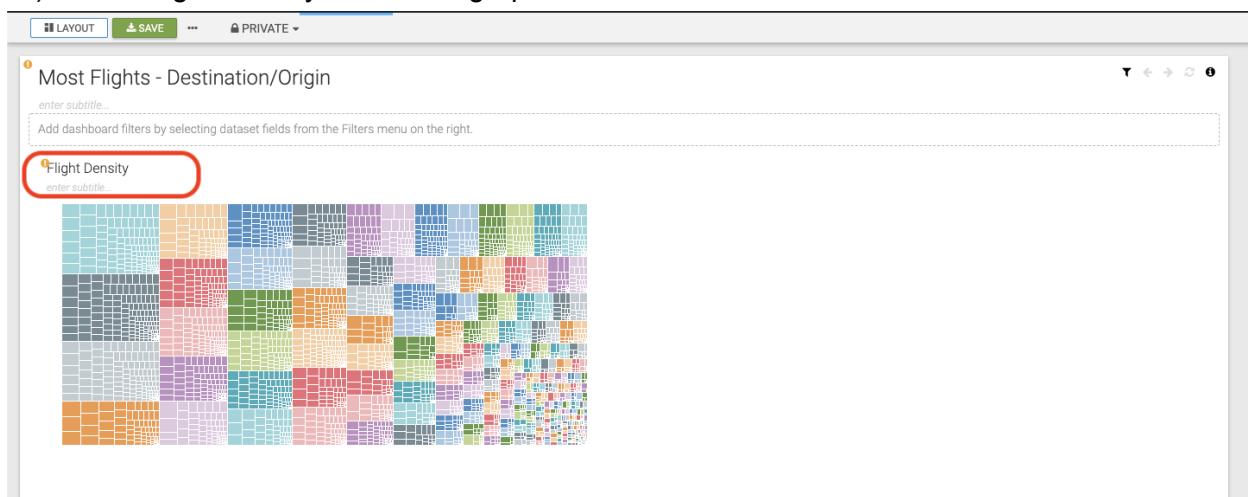
*Notice - you can have other Visuals chosen to be displayed with the Dimensions and Measure(s), then click REFRESH VISUALS

The screenshot shows the Tableau Dashboard Designer interface. On the left, the **VISUALS** pane displays various visualization icons, with a Treemap icon selected. Below it are sections for **Dimensions** (dest, origin), **Measure** (# Record Count), **Tooltips**, **X Trellis**, **Y Trellis**, and **Filters**. At the bottom of this pane is a blue button with a circular arrow icon labeled "REFRESH VISUAL", which is circled in red. The central **DATA** pane shows a connection to "airlines_new_orc.flights" with sample mode set to "OFF". It lists dimensions like uniquecarrier, tailnum, origin, dest, cancellationcode, and diverted, and measures like Record Count, month, dayofmonth, dayofweek, deptime, crsdeptime, arftime, crsarrrtime, flightnum, actualelapsedtime, crselapsedtime, airtime, and arrdelay. A sidebar on the right contains buttons for **DASH.**, **Visuals**, **Filters**, **Settings**, **Style**, **VISUAL** (which is highlighted in dark grey), **Build**, **Settings**, and **Style**.

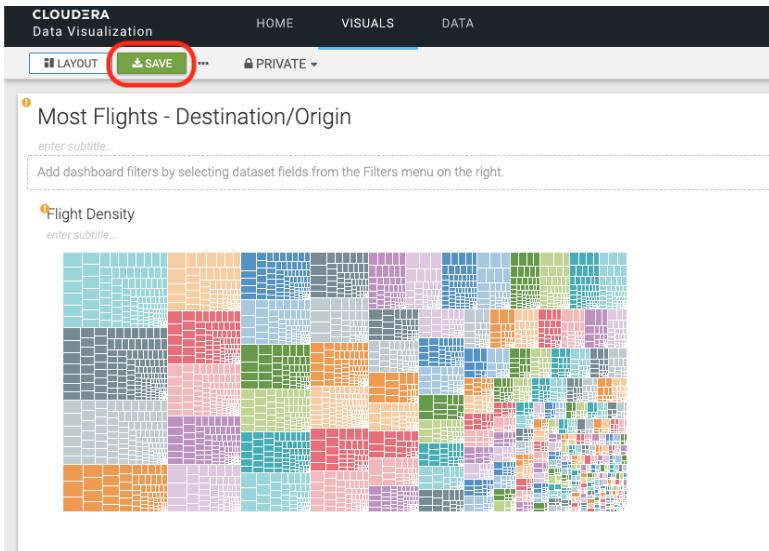
18) Enter a title “Most Flights - Destination/Origin”



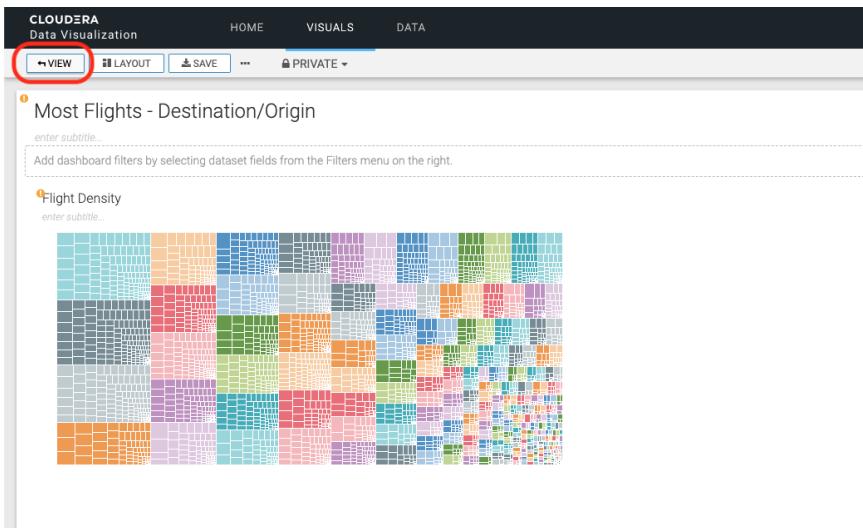
19) Enter “Flight Density” under the graph’s title



20) Click “SAVE”

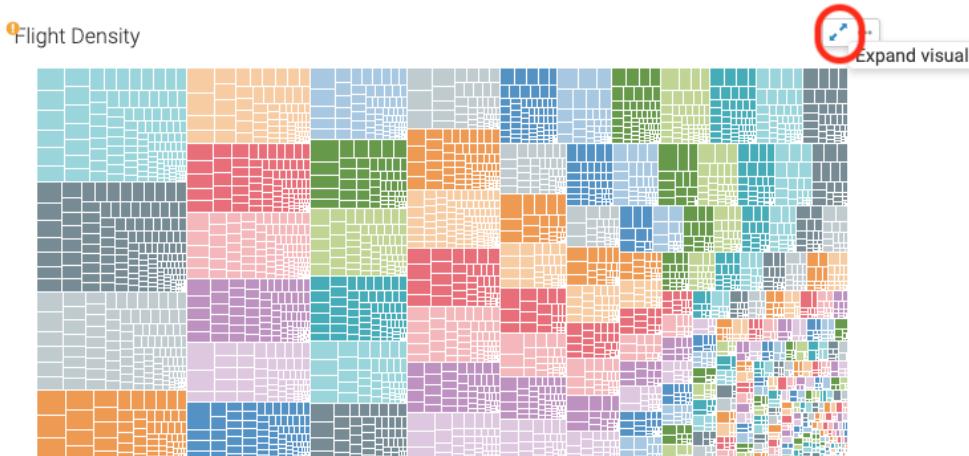


21) Click “VIEW”



22) Scroll over the graph and click “Expand Visual”

Most Flights - Destination/Origin



Destinations are displayed



Part 4 - Import a File into a Table [15 minutes]

Overview: How do we import data (csv file), creating a table?

- 1) Open CDP, using the “admin” user within the Test Drive link.

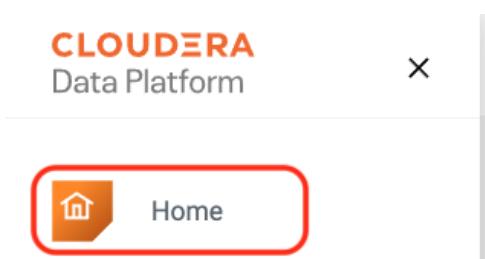
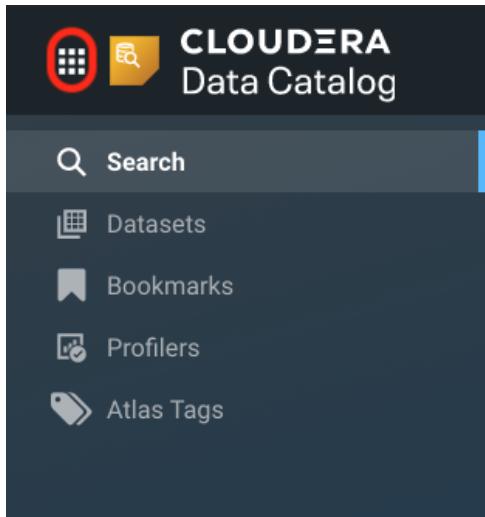
Your link should look something like (remember click the link in your email not the link below)
http://login.trycdp.com/auth/realms/trycdp-trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X
*xx represents the trial user #
*X represents the password

- 2) Click the “Data Warehouse” within the CDP Home Screen



How do you get to the CDP Home Screen?

- From any experience such as “Data Catalog”, click the 9 square at the top left and then click “Home”



3) Click “Open DAS” on your existing “Running” Virtual Warehouse

*The same steps you did in Part 2 to Open DAS

The screenshot shows the CD彭trialuser24-dl-default interface for managing virtual warehouses. It lists three entries:

- testvirtualwarehouse1**: Running, compute-1611179792-vz49, cdptrialuser24-dl-default. Details: NODE COUNT 2, TOTAL CORES 38, TOTAL MEMORY 292 GB.
- mschoeni-iso-1**: Stopped, compute-1611173596-dbtv, cdptrialuser24-dl-default. Details: NODE COUNT 0, TOTAL CORES 12, TOTAL MEMORY 56 GB.
- default-vw**: Stopped, compute-1611103491-4hbp. Details: NODE COUNT 0, TOTAL CORES 0, TOTAL MEMORY 0 GB.

A context menu is open over the first entry, showing options: Suspend, Clone, Edit, Delete, Upgrade, Copy JDBC URL, Download JDBC Jar, **Open DAS** (which is highlighted with a red box), Open Data Visualization, Set Compactor, Run AutoScaling Demo, and Collect Diagnostic Bundle.

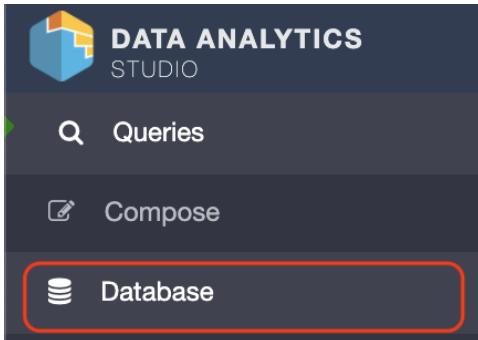
4) Enter the login information from step #1 above using the user, then click “LOGIN”

*You’ll likely already be authenticated from Part 2, you may not need to enter credentials

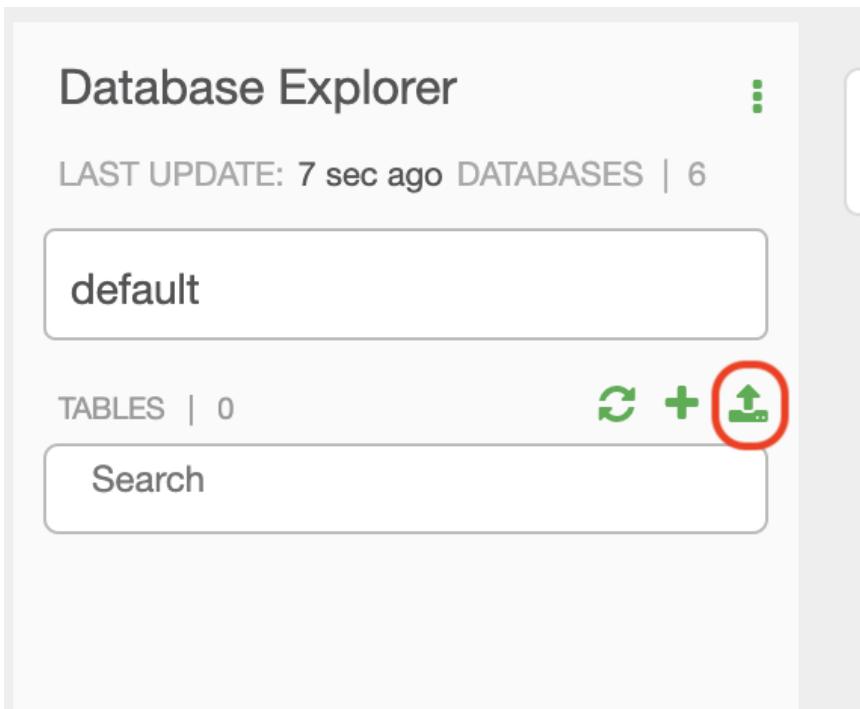
*Changing “trialxx_admin” to the trail user you’re using and password defined by “X” in #1 above

The screenshot shows the Data Analytics Studio login interface. It has two input fields: "Username" containing "trialxx_admin" and "Password" which is empty. Below the fields is a green "LOGIN" button, which is circled in red.

5) Click on Database on the left navigation bar

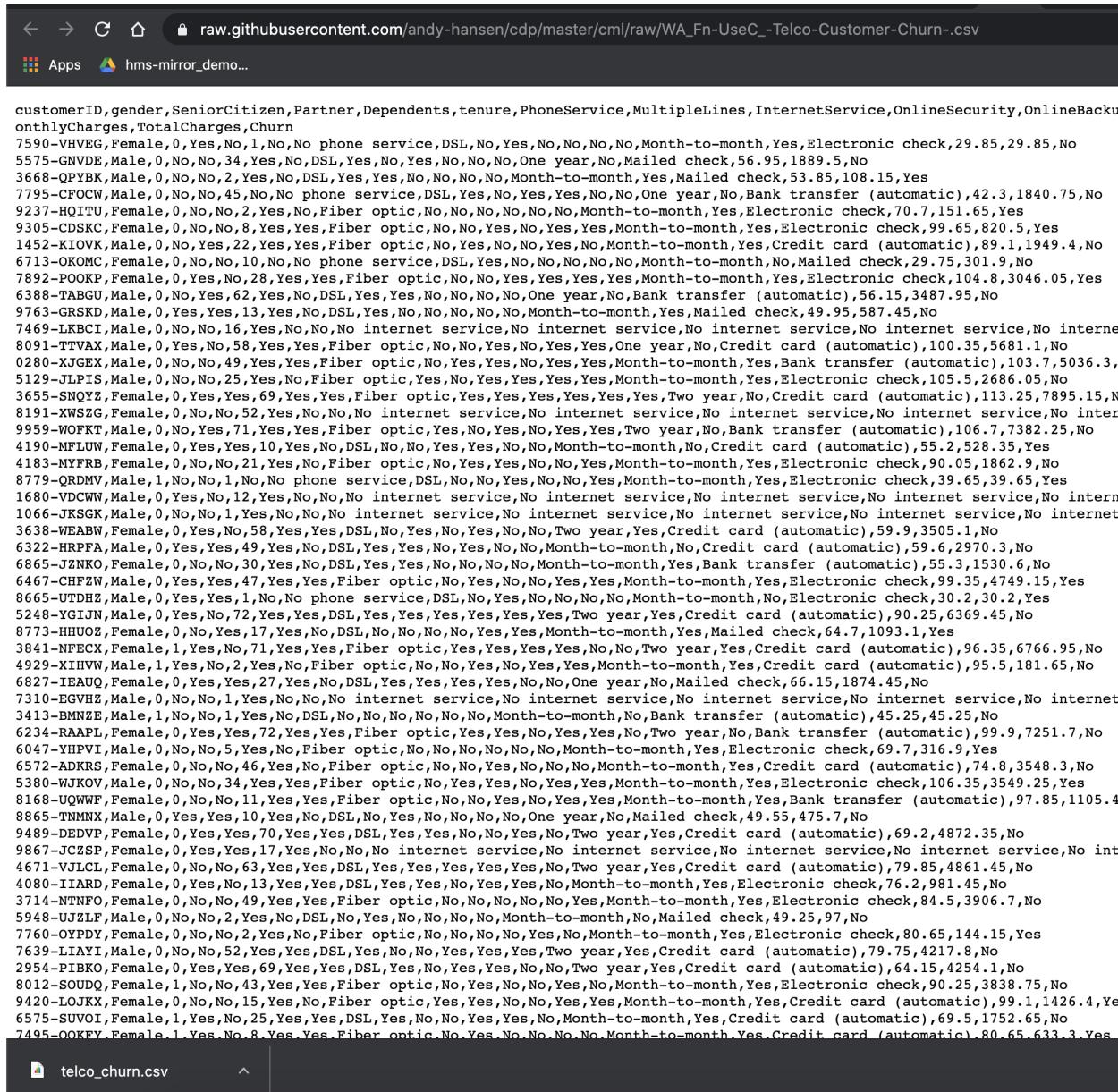


6) Click on “Upload Table”, using the “default” database



7) In a new browser window or tab, download the CSV file, saving to your desktop as "telco_churn.csv"

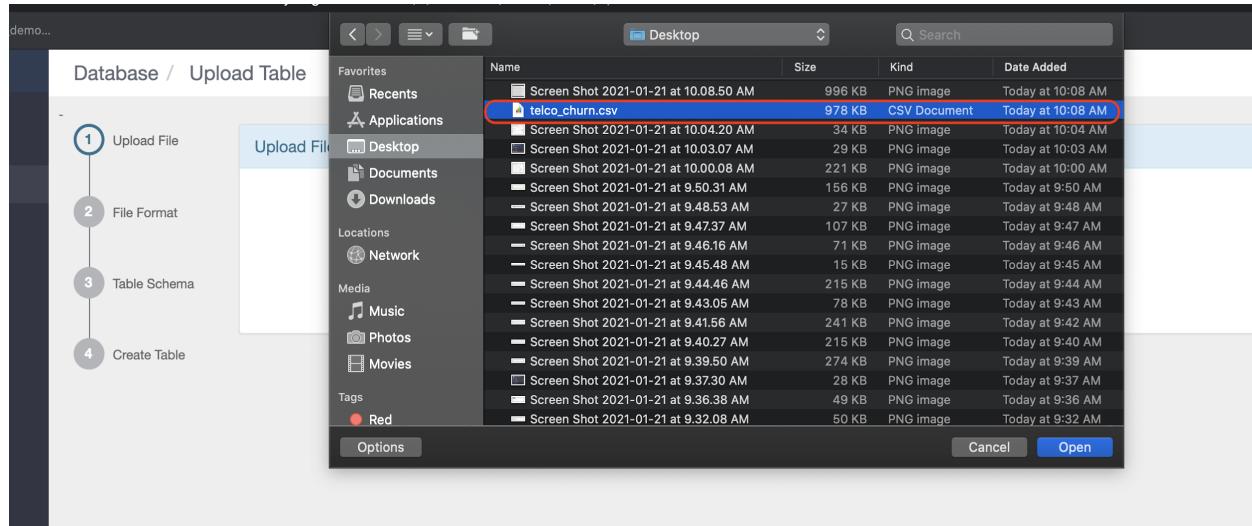
https://raw.githubusercontent.com/andy-hansen/cdp/master/cml/raw/WA_Fn-UseC_-Telco-Customer-Churn-.csv



The screenshot shows a browser window with the URL https://raw.githubusercontent.com/andy-hansen/cdp/master/cml/raw/WA_Fn-UseC_-Telco-Customer-Churn-.csv. The page displays a large amount of comma-separated data, which is the CSV file content. The data includes columns such as customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, monthlyCharges, TotalCharges, and Churn. The content lists numerous individual records with their respective details.

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	monthlyCharges	TotalCharges	Churn							
7590-VHVEG	Female	0	Yes	No	1	No	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No				
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	One year	No	Mailed check	56.95	1889.5	No				
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes		
7795-CFOCW	Male	0	No	No	45	No	No	No	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No				
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes		
9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	Yes
1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Month-to-month	Yes	Credit card (automatic)	89.1	1949.4	No		
6713-OKOMC	Female	0	No	No	10	No	No	No	No	No	Month-to-month	No	Mailed check	29.75	301.9	No				
7892-POOKP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Month-to-month	Yes	Electronic check	104.8	3046.05	Yes		
6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	56.15	3487.95	No		
9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	Month-to-month	Yes	Mailed check	49.95	587.45	No		
7469-LKBCI	Male	0	No	No	16	Yes	No	No	No	No	No	No	No	No	internet service	No	internet service	No	internet service	
8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	Yes	One year	No	Credit card (automatic)	100.35	5681.1	No		
0280-XJGEW	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Month-to-month	Yes	Bank transfer (automatic)	103.7	5036.3	,	
5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Month-to-month	Yes	Electronic check	105.5	2686.05	No	
3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Two years	No	Credit card (automatic)	113.25	7895.15	No	
8191-XWSZG	Female	0	No	No	52	Yes	No	No	No	No	No	No	No	No	internet service	No	internet service	No	internet service	
9959-WOFKT	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	Yes	Yes	Two year	No	Bank transfer (automatic)	106.7	7382.25	No	
4190-MFLUW	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	Month-to-month	No	Credit card (automatic)	55.2	528.35	Yes		
4183-MYFRB	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	Yes	Month-to-month	Yes	Electronic check	90.05	1862.9	No	
8779-QRDMV	Male	1	No	No	1	No	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	39.65	39.65	Yes		
1680-VDCWW	Male	0	Yes	No	12	Yes	No	No	No	No	No	No	No	No	internet service	No	internet service	No	internet service	
1066-JKSGK	Male	0	No	No	1	Yes	No	No	No	No	No	No	No	No	internet service	No	internet service	No	internet service	
3638-WEABW	Female	0	Yes	No	58	Yes	Yes	DSL	No	No	Yes	No	Two year	Yes	Credit card (automatic)	59.9	3505.1	No		
6322-HRFPA	Male	0	Yes	Yes	49	Yes	No	DSL	Yes	Yes	No	No	Month-to-month	No	Credit card (automatic)	59.6	2970.3	No		
6865-JZNKO	Female	0	No	No	30	Yes	No	DSL	Yes	Yes	No	No	Month-to-month	Yes	Bank transfer (automatic)	55.3	1530.6	No		
6467-CHFZW	Male	0	Yes	Yes	47	Yes	Yes	Fiber optic	No	Yes	No	No	Month-to-month	Yes	Electronic check	99.35	4749.15	Yes		
8665-UTDHZ	Male	0	Yes	Yes	1	No	No	No	No	No	No	No	Month-to-month	No	Electronic check	30.2	30.2	Yes		
5248-YGIJN	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Credit card (automatic)	90.25	6369.45	No	
8773-HHUOZ	Female	0	No	Yes	17	Yes	No	DSL	No	No	Yes	Yes	Month-to-month	Yes	Mailed check	64.7	1093.1	Yes		
3841-NFECX	Female	1	Yes	No	71	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	No	No	Two year	Yes	Credit card (automatic)	96.35	6766.95	No
4929-XIBHW	Male	1	Yes	No	2	Yes	No	Fiber optic	No	No	Yes	No	Yes	Month-to-month	Yes	Credit card (automatic)	95.5	181.65	No	
6827-IEAUQ	Female	0	Yes	Yes	27	Yes	No	DSL	Yes	Yes	Yes	Yes	No	No	One year	No	Mailed check	66.15	1874.45	No
7310-EGVHZ	Male	0	No	No	1	Yes	No	No	No	No	No	No	No	No	internet service	No	internet service	No	internet service	
3413-BMNZE	Male	1	No	No	1	Yes	No	DSL	No	No	No	No	Month-to-month	No	Bank transfer (automatic)	45.25	45.25	No		
6234-RAAPI	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	Yes	Yes	No	Yes	Yes	Two year	No	Bank transfer (automatic)	99.9	7251.7	No	
6047-YHPVI	Male	0	No	No	5	Yes	No	Fiber optic	No	No	No	No	Month-to-month	Yes	Electronic check	69.7	316.9	Yes		
6572-ADKRS	Female	0	No	No	46	Yes	No	Fiber optic	No	No	Yes	No	Month-to-month	Yes	Credit card (automatic)	74.8	3548.3	No		
5380-WJKOV	Male	0	No	No	34	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Month-to-month	Yes	Electronic check	106.35	3549.25	Yes	
8168-UQWWF	Female	0	No	No	11	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Month-to-month	Yes	Bank transfer (automatic)	97.85	1105.4		
8865-TNMNX	Male	0	Yes	Yes	10	Yes	No	DSL	No	Yes	No	No	No	One year	No	Mailed check	49.55	475.7	No	
9489-DEDPV	Female	0	Yes	Yes	70	Yes	Yes	DSL	Yes	Yes	No	Yes	Yes	Two year	Yes	Credit card (automatic)	69.2	4872.35	No	
9867-JCZSP	Female	0	Yes	Yes	17	Yes	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	69.7	316.9	Yes		
4671-VJLCL	Female	0	No	No	63	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Two year	Yes	Credit card (automatic)	79.85	4861.45	No		
4080-IIARD	Female	0	Yes	No	13	Yes	Yes	DSL	Yes	Yes	Yes	No	Month-to-month	Yes	Electronic check	76.2	981.45	No		
3714-NTNFO	Female	0	No	No	49	Yes	Yes	Fiber optic	No	No	No	No	Yes	Month-to-month	Yes	Electronic check	84.5	3906.7	No	
5948-UJZLF	Male	0	No	No	2	Yes	No	DSL	No	Yes	No	No	Month-to-month	No	Mailed check	49.25	97	No		
7760-OYPDV	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	Yes	Month-to-month	Yes	Electronic check	80.65	144.15	Yes		
7639-LIAYI	Male	0	No	No	52	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Two year	Yes	Credit card (automatic)	79.75	4217.8	No		
2954-PIBKO	Female	0	Yes	Yes	69	Yes	Yes	DSL	Yes	No	Yes	Yes	Two year	Yes	Credit card (automatic)	64.15	4254.1	No		
8012-SOUDQ	Female	1	No	No	43	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	Month-to-month	Yes	Electronic check	90.25	3838.75	No	
9420-LOJXK	Female	0	No	No	15	Yes	No	Fiber optic	Yes	Yes	No	No	Yes	Month-to-month	Yes	Credit card (automatic)	99.1	1426.4	Yes	
6575-SUVOI	Female	1	Yes	No	25	Yes	Yes	DSL	Yes	No	No	Yes	Yes	Month-to-month	Yes	Credit card (automatic)	69.5	1752.65	No	
7495-ODKFY	Female	1	Yes	No	8	Yes	Yes	Fiber optic	No	Yes	No	No	No	Month-to-month	Yes	Credit card (automatic)	80.65	633.3	Yes	

8) Going back to the window from step 6 above, upload the file "telco_churn.csv"



9) Click the "Is first row header?", since the first row is a header

The screenshot shows a 'Select File Format' dialog box. It has a header 'Select File Format'. Below it are four input fields: 'File type' (set to 'CSV'), 'Field Delimiter' (set to ','), 'Escape Character' (set to '\'), and 'Quote Character' (set to '\"'). Each field has a 'Clear' link to its right. Below these fields is a checkbox labeled 'Is first row header?' which is checked and highlighted with a red box. Underneath is another checkbox labeled 'Contains endlines?' which is unchecked. At the bottom is a blue 'PREVIEW' button.

10) Click “PREVIEW” prior to creating the table

11) Click “NEXT”

12) Enter ‘telco_churn’ as the Table Name. Click “CREATE”

Table Name: telco_churn

COLUMN NAME	DATA TYPE	SIZE	ADVANCED	ACTION
customerID	STRING		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
gender	STRING		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
SeniorCitizen	INT		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
Partner	STRING		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
Dependents	STRING		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
tenure	INT		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>
PhoneService	STRING		<input type="button" value="⚙"/>	<input type="button" value="✖ DELETE"/>

13) Wait for about 2 minutes then Go-to “Compose” and within “Worksheet 1” run the following query on the new table

```
select * from telco_churn limit 10;
```

Compose

Saved | Worksheet1 | +

DATABASES | 6

default

TABLES | 1

Search Tables

> telco_churn (21)

Results

TELCO_CHURN.CUSTOMERID	TELCO_CHURN.GENDER	TELCO_CHURN.SENIORCITIZEN	TELCO_CHURN.PARTNER	TELCO_CHURN.DEPENDENTS	TELCO_CHURN.TENURE	TELCO_CHURN.PHONESERVICE	TELCO_CHURN.MULTIPLELINES	TELCO_CHURN.INTERNET
7590-VhYEG	Female	0	Yes	No	1	No	No phone service	DSL
5575-GNvDE	Male	0	No	No	34	Yes	No	DSL
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL

Parking lot items

- Show Impala/Hue - for CDH customers