



# databricks

## Academy

---

## DAWD 03-5 – Lab – Notifying Stakeholders

The two fields below are used to customize queries used in this course. Enter your schema (database) name and username, and press "Enter" to populate necessary information in the queries on this page.

Schema Name:

Username:

---



### Lesson Objective

At the end of this lesson, you will be able to:

- Configure alerts
  - Share queries and dashboards with stakeholders
  - Refresh dashboards
- 



### Refreshing Queries

Putting a query on a refresh schedule only takes a few steps. In this portion of the lab, we are going to use a query we worked with earlier in the course that was designed to ingest new JSON files, containing new data. If we had to manually run the query each time we wanted to ingest new data, we

might find our data growing stale in-between query runs. By configuring a Refresh Schedule, we can automatically pull new data into the `sales_stream` table.

Complete the following:

1. Go to the Query Editor by clicking "SQL Editor" in the sidebar menu
2. Start a new query by hovering your mouse over the "+" in the tabbed area of the Query Editor and clicking "Create new query"
3. Run the query below
4. Name the query by clicking "New Query" and typing "Ingest Sales Stream"
5. Click "Save"
6. Click "Never" next to "Refresh Schedule"
7. Change the dropdown to something other than "Never"
8. Optionally, change "End" to tomorrow's date
9. Click "OK"
10. Check your work by entering your answer to the question
11. After pressing ENTER/RETURN, green indicates a correct answer, and red indicates incorrect

```
USE hive_metastore.class_013_odg7_da_dawd;
DROP TABLE IF EXISTS sales_stream;
CREATE TABLE sales_stream (5minutes STRING,
                           clicked_items ARRAY<ARRAY<STRING>>,
                           customer_id STRING,
                           customer_name STRING,
                           datetime STRING,
                           hour BIGINT,
                           minute BIGINT,
                           number_of_line_items STRING,
                           order_datetime STRING,
                           order_number BIGINT,
                           ordered_products ARRAY<ARRAY<STRING>>,
                           sales_person DOUBLE,
                           ship_to_address STRING
);
COPY INTO sales_stream
  FROM 'wasb://courseware@dbacademy.blob.core.windows.net/data-analysis-with-
databricks/v02/retail-org/sales_stream/sales_stream_json'
  FILEFORMAT = JSON;
SELECT * FROM sales_stream ORDER BY customer_id;
```

[Show Answer](#)[Copy](#)

What is the shortest amount of time we can configure for query refreshes?



## Alerts

Now that we have configured a refresh schedule, let's set up an Alert that will notify us when the income generated from sales increases beyond a threshold.

The goal of this portion of the lab is to generate a query that will calculate a sum of the income from all the sales in the `sales_stream` table and configure an Alert that will notify us when the income amount surpasses \$1,000,000.

The `sales_stream` table has a column called `ordered_products`, which contains an array of additional arrays that each contain, among other things, the sales price of the product ordered. This value is in the third position in each array. The first step in the query is return each array of product information as its own row in the results. The `explode()` function does just this. Since we want the third element in each of these arrays, we use a CTE to allow us to use the bracket operator (`[ ]`) to index the third element. We then use the `sum()` function to calculate a sum of the cost of the items sold. We then configure an Alert to notify us when the sum increases beyond 1000000.

Complete the following:

1. Make the required changes to the query below
2. Run the query in Databricks SQL
3. Save the query as "Total Sales Dollars"
4. Click "Alerts" in the sidebar menu
5. Click "Create Alert"
6. From the Query dropdown, select our query: "Total Sales Dollars"
7. Use the dropdown to change the "Value" column to `total_sales` and change "Threshold" to 1000000
8. Change "Refresh" to Every 1 minute
9. Check your work by entering your answer to the question
10. After pressing ENTER/RETURN, green indicates a correct answer, and red indicates incorrect

```
USE hive_metastore.class_013_odg7_da_dawd;  
WITH all_sold_products AS  
  (SELECT explode(ordered_products) AS products FROM sales_stream)  
SELECT sum(products[3]) AS total_sales FROM FILL_IN;
```

[Show Answer](#)[Copy](#)

What is the total amount of income (in whole dollars) currently in the `sales_stream` table?  
(type only numbers)

## Sharing Queries/Dashboards

Since the process for sharing queries and dashboards is identical, we are only going to work with dashboards. Let's share a dashboard with all other users.

Complete the following:

1. Click "Dashboards" in the sidebar menu
2. Click on the name of your dashboard
3. In the upper-right corner, click "Share"
4. Click inside the input box, and select the group, "All Users"
5. Click inside the input box again to dismiss the drop down

Note that the only permission we can grant is "Can Run". If we want to grant "Can Edit", we first need to have the proper permissions. We can then change "Credentials" (under "Sharing Settings" in the "Manage permissions" dialog) to "Run as viewer". We do not want to make this change, here, though. Leave the Credentials set to "Run as owner"

6. Click "Add", and click the "X" to dismiss the "Manage permissions" dialog
7. Check your work by entering your answer to the question
8. After pressing `ENTER/RETURN`, green indicates a correct answer, and red indicates incorrect

To remove sharing permissions, click "Share" and "Revoke" for the user whose permission should be revoked.

After sharing the dashboard with the "All Users" group, how many rows of users and groups are in the "Manage permissions" dialog?

## Refreshing Dashboards

To keep our dashboard up-to-date, we can set a Refresh Schedule that will run the dashboard at a set interval that ranges from 1 minute to 1 year. In addition, we can configure Subscribers, who will receive a report of our dashboard after the update runs.

Complete the following:

1. To drop down the schedule configuration window, click the button "Schedule" in the upper-right corner
2. Change the "Refresh" to "Every 1 minute"

You can choose to run the queries that provide data to the dashboard on a specific SQL Warehouse, or leave the default

3. Leave "SQL Warehouse" on (optional)

You can configure the Refresh Schedule to send the results of refreshing the dashboard to one, or more, subscribers. Those who are added as subscribers to the refresh schedule will be granted "Can Run" permissions on the dashboard.

4. If desired, click inside the "Subscribers" input box and select from the list of users

Note that the Refresh Schedule can be enabled and disabled. If you wish to keep the Refresh Schedule's settings, but you do not wish to run the refresh schedule, click the switch to disable refreshing. Finally, to delete the refresh schedule, change "Refresh" to "Never".

5. Change the "Refresh" to "Never" so it won't be running indefinitely
6. Check your work by entering your answer to the question

- After pressing ENTER/RETURN, green indicates a correct answer, and red indicates incorrect

When you click "Advanced" under the "Refresh" drop down, how many minutes are selected by default? (enter only numbers)

---

© 2023 Databricks, Inc. All rights reserved.

Apache, Apache Spark, Spark and the Spark logo are trademarks of the [Apache Software Foundation](#).

[Privacy Policy](#) | [Terms of Use](#) | [Support](#)

1.2.13