O'Reilly – Databricks Fundamentals Bootcamp

by Mohit Batra
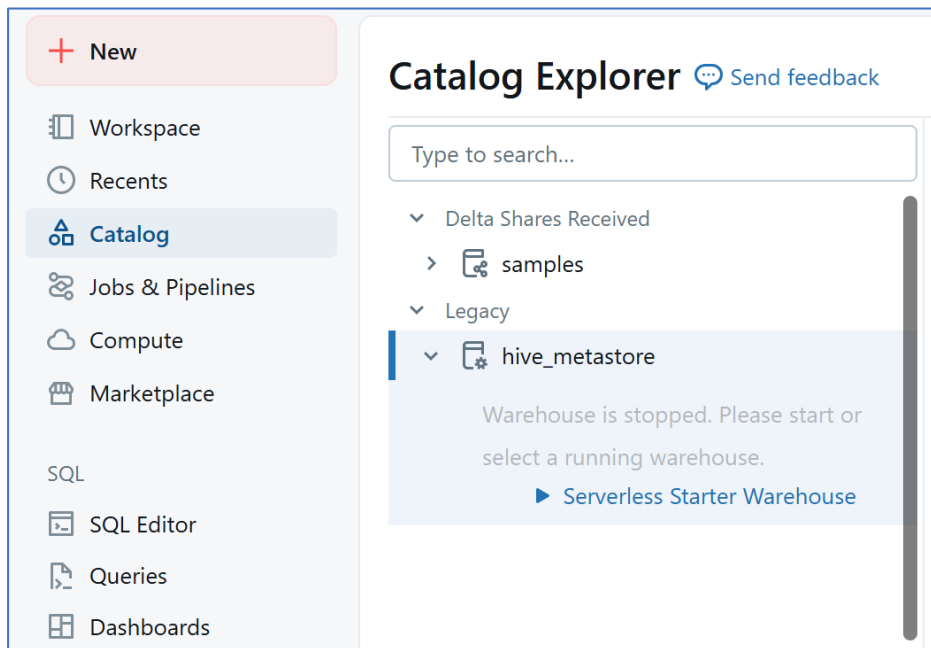
Lab 2: Upload Files and Run Commands

Contents

Exercise 1 – Verify if Unity Catalog is enabled by default

1. Open Databricks Workspace created in previous labs.

2. From the left menu, navigate to Catalog.

3. In Catalog Explorer:

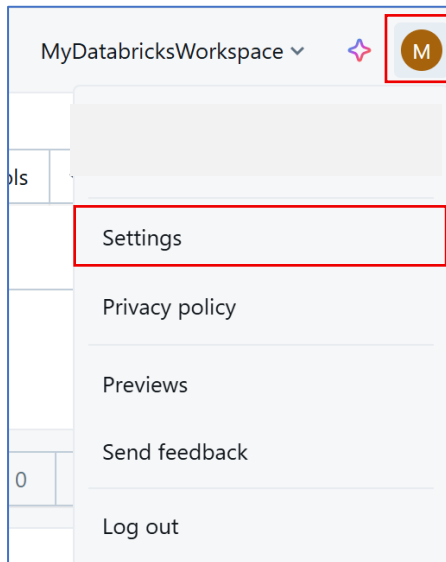   a. If only hive_metastore is visible, this implies that Unity Catalog is not enabled.

   b. If My organization is visible, and clicking on settings shows Metastore connected, this implies that Unity Catalog is enabled.
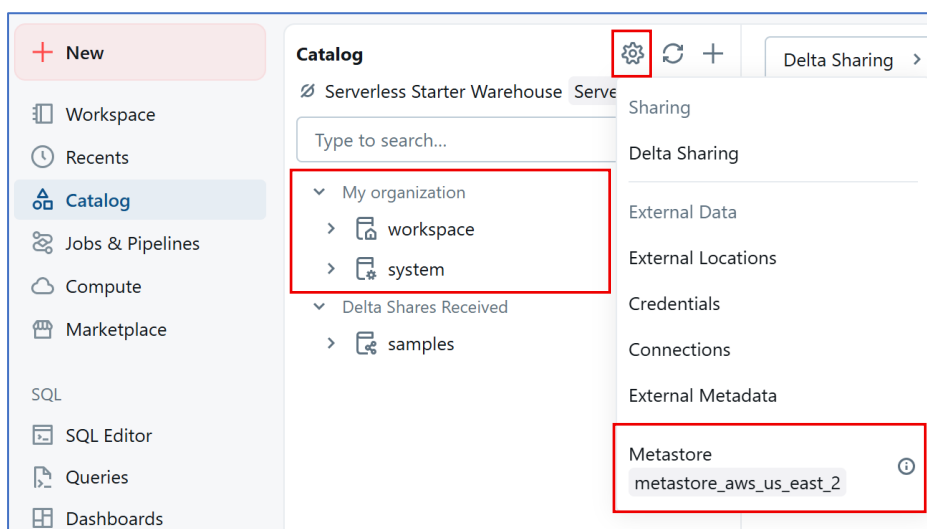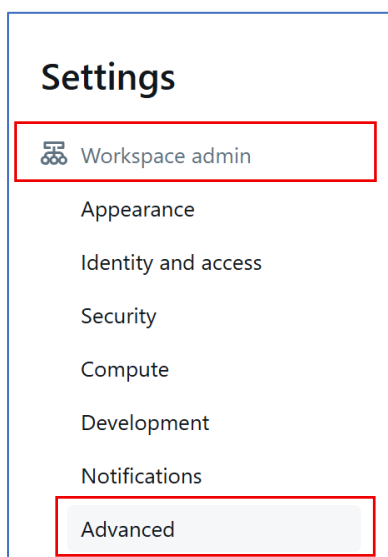
Exercise 2 – Upload Files

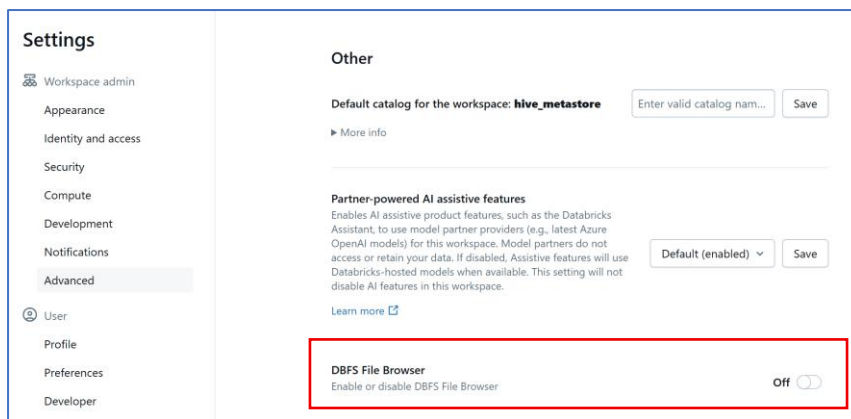## Option 2.1 – If Unity Catalog is NOT enabled

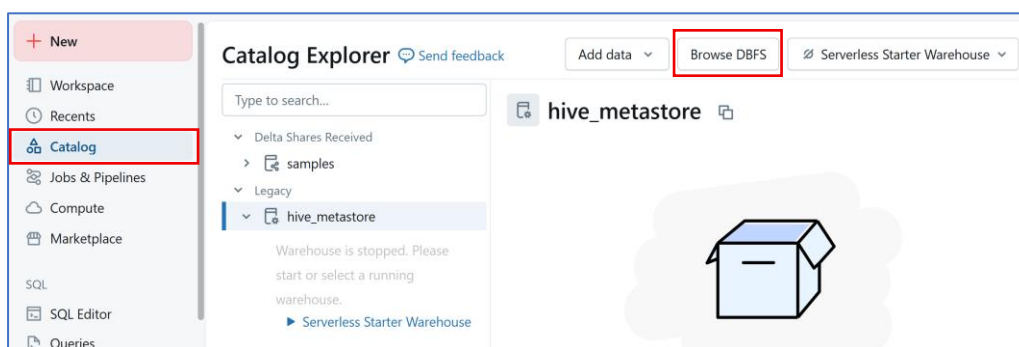1. In the workspace, from to right, click on user account and select **Settings**.

2. On settings page, under **Workspace admin**, click on **Advanced**.
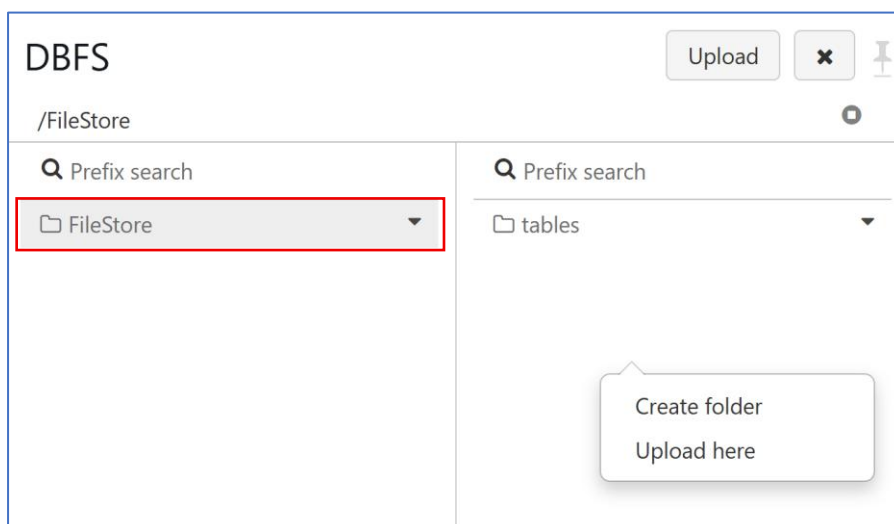

3. Scroll down to **Other** section, where the option of **DBFS File Browser** is available. Set the toggle to **ON**.
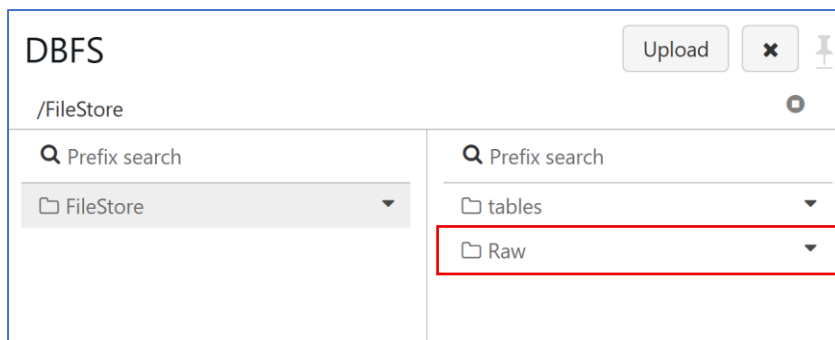
4. Refresh the page for the changes to take effect.

5. From the left menu, navigate to **Catalog**. You should see an option of **Browse DBFS**.

6. Click on **Browse DBFS**.

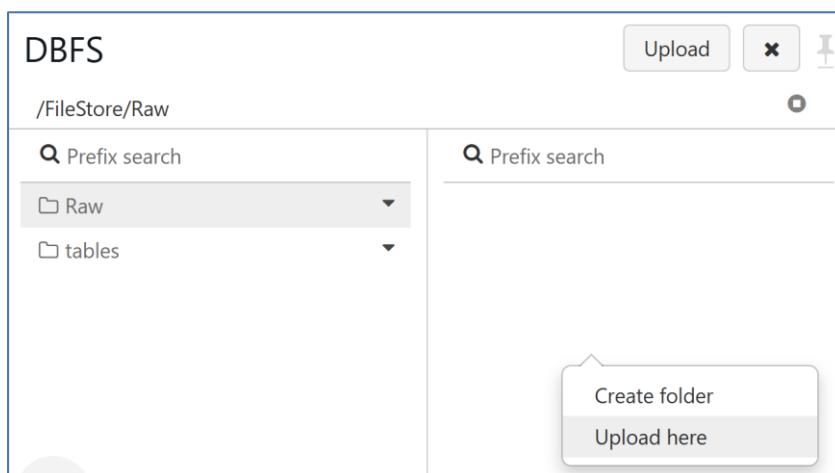7. On DBFS page, under FileStore, right-click to create a new folder.

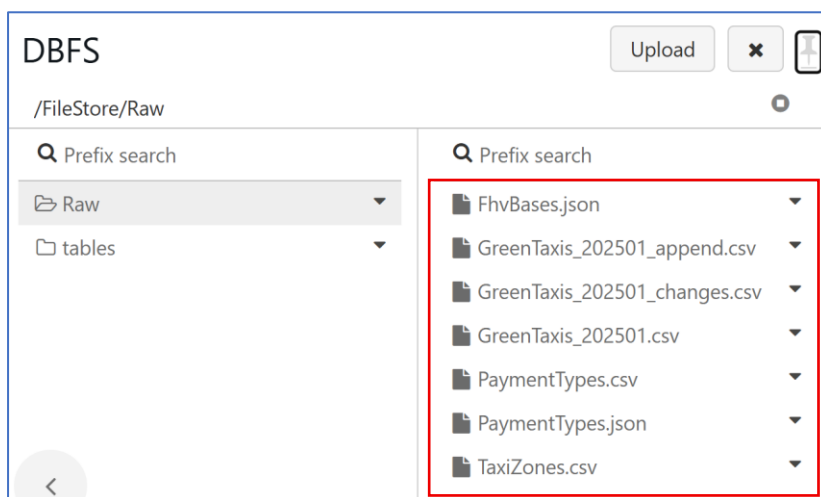8. Define the folder name, **Raw**, and create the folder.

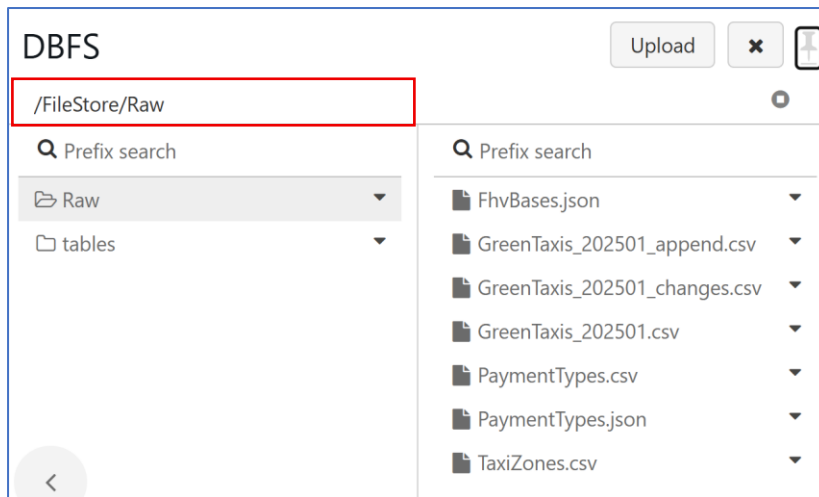9. Click on the **Raw** folder to open it.

10. Right-click under Raw folder to upload files.

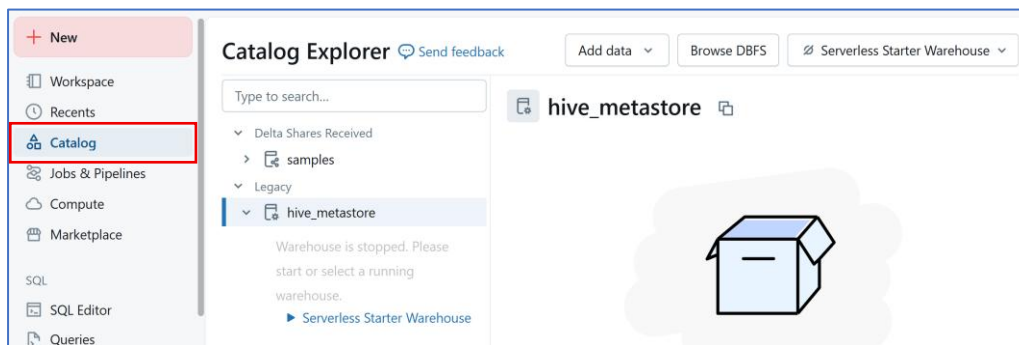11. In Raw directory, upload the shared sample files.

12. Copy the **FileStore** file path as it will be required later.
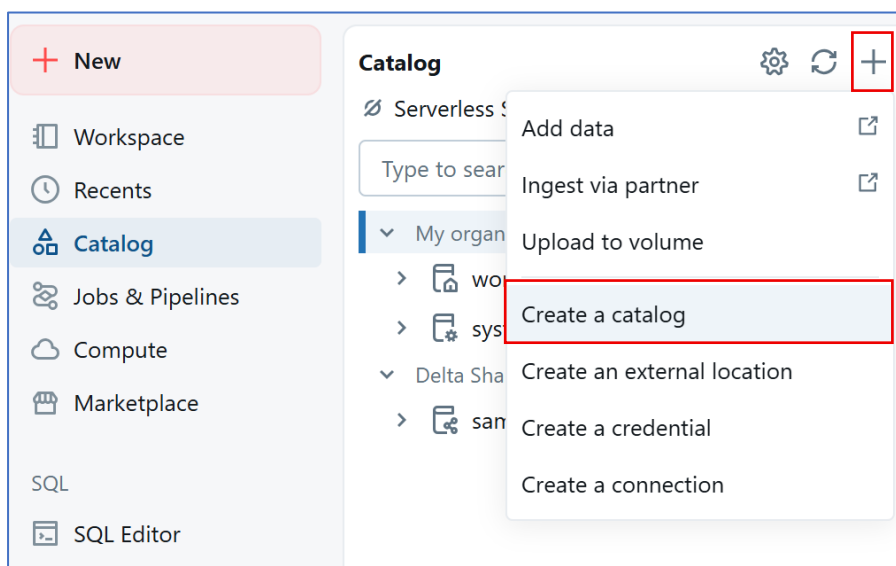
Option 2.2 – If Unity Catalog is enabled

1. From the left menu, navigate to **Catalog**.

2. On top-right, click on plus (+) sign. Select **Create a catalog**.

3. Provide Catalog name as **OReillyCatalog** and create it.

4. On **Catalog created** popup, select **View catalog**.

5. From Catalog explorer, select **oreillycatalog**. And under catalog, select **default** schema.
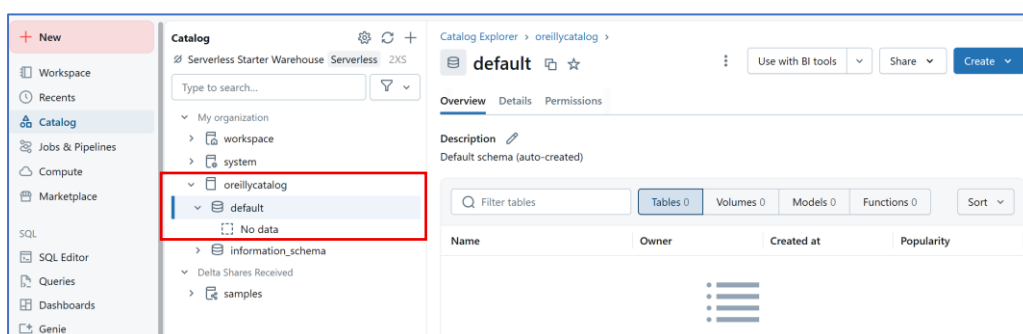
6. From top-right in **default** schema, click on **Create**. And select **Volume**.
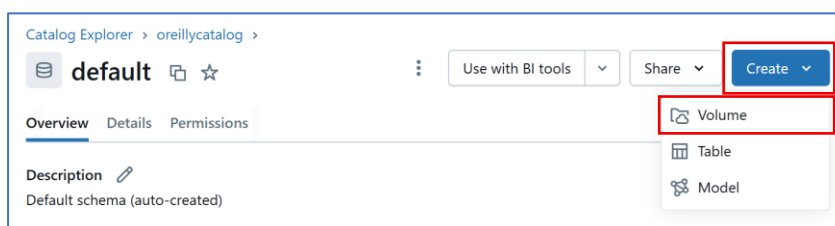
7. Provide volume name as **DataFiles** and create it.

8. Under the volume **datafiles**, click on **Create directory**.

9. Provide the directory name as **Raw** and create it.

10. On top-right, click on **Upload to this volume** to upload the files.

11. In Raw directory, upload the shared sample files.

12. Copy the **Volumes** file path as it will be required later.

Exercise 3 – Import Notebook & Run Commands

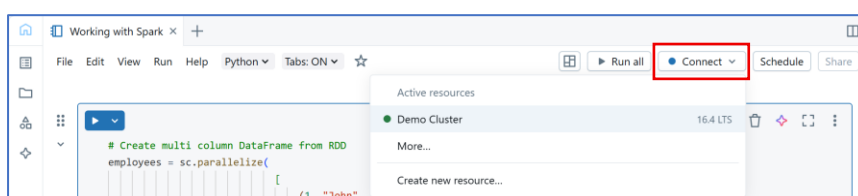1. Download notebook – **Working with Spark.py** from GitHub repository.

2. From the left menu, go to **Workspace** tab.

3. Click on Workspace folder → Users folder → Your user account.

4. Right-click on the account folder and click **Import**.
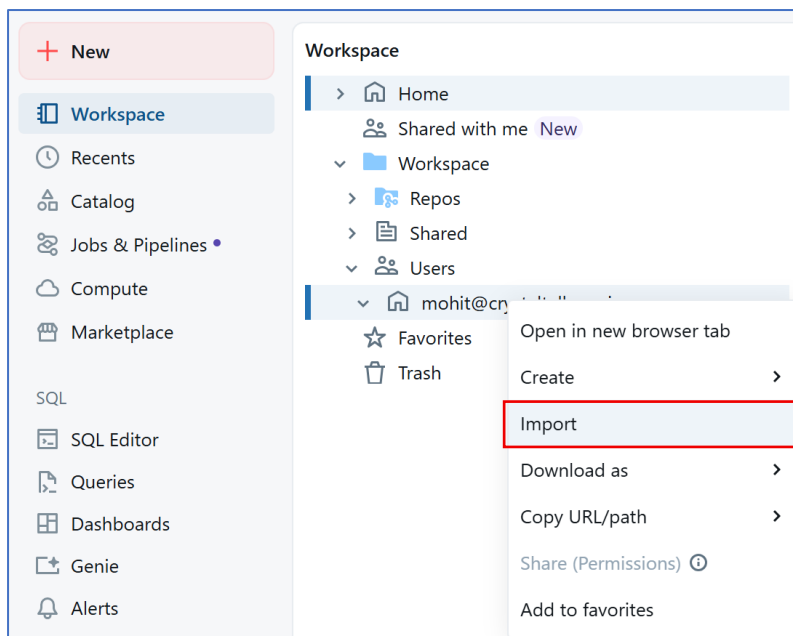
5. Upload the notebook - **Working with Spark.py** and click **Import**.
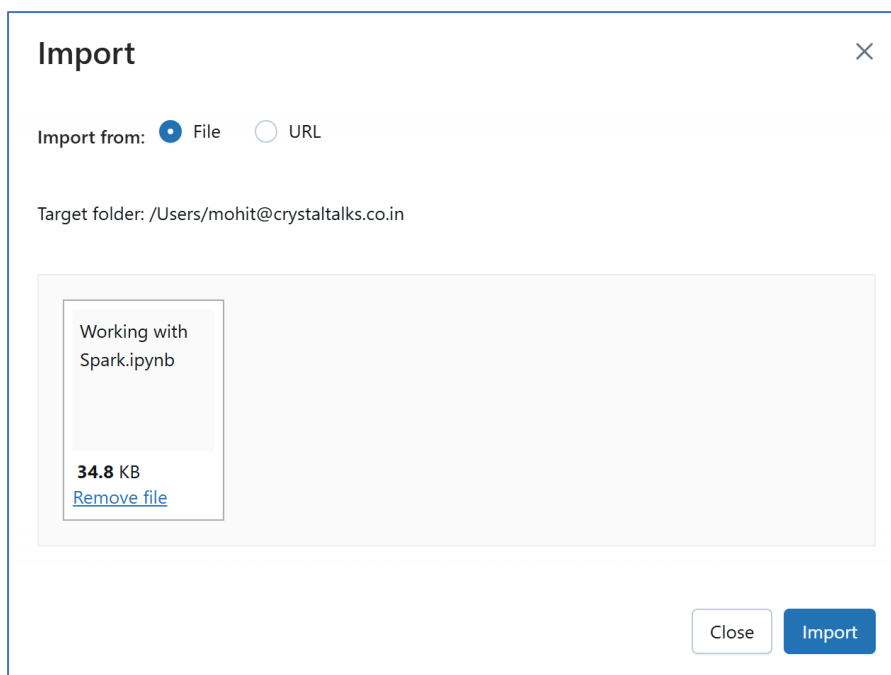
6. Open the notebook and connect to cluster (either Demo Cluster or Serverless).

7. Run the commands 1 to 11.