Delta Lake
Hands-on Walkthrough

Start with Databricks Community Edition

Databricks Community Edition

Databricks Community Edition is the free version of our cloud-based big data platform.

- Users can access a micro-cluster as well as a cluster manager and notebook environment.
- All users can share their notebooks and host them free of charge with Databricks.

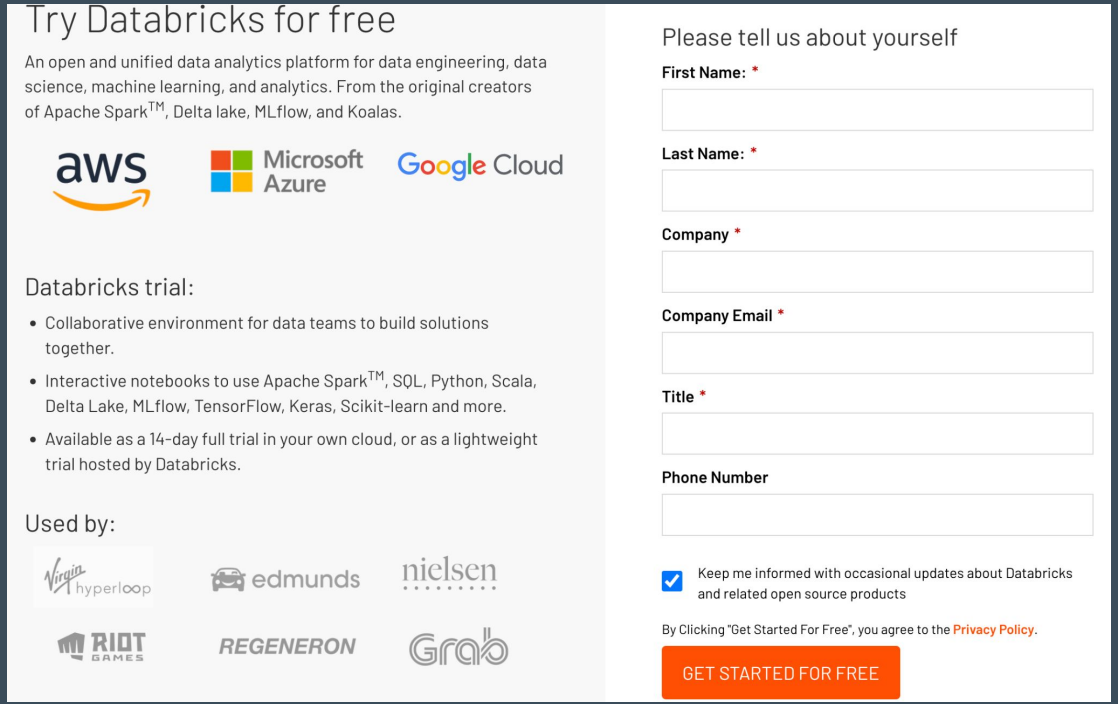
Sign up for
Databricks Community Edition

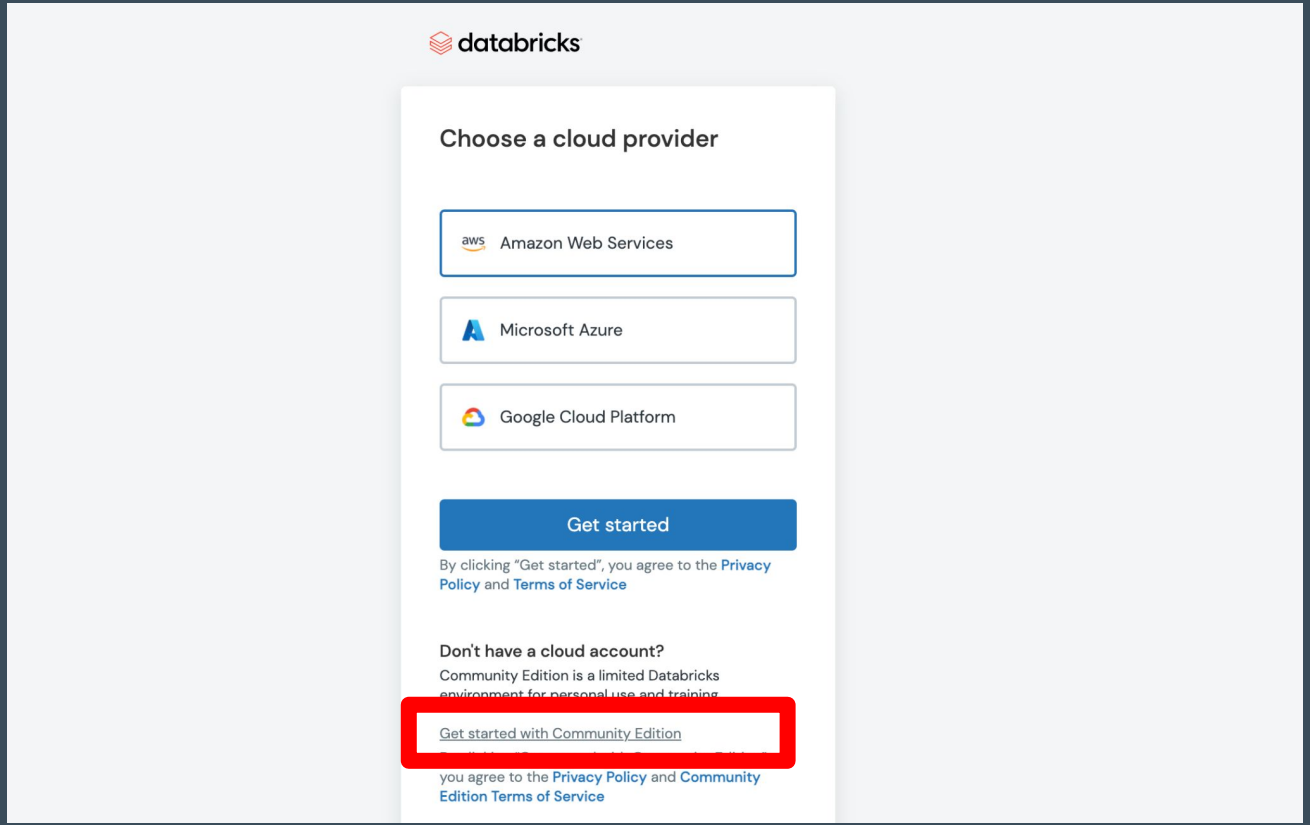
Go to:

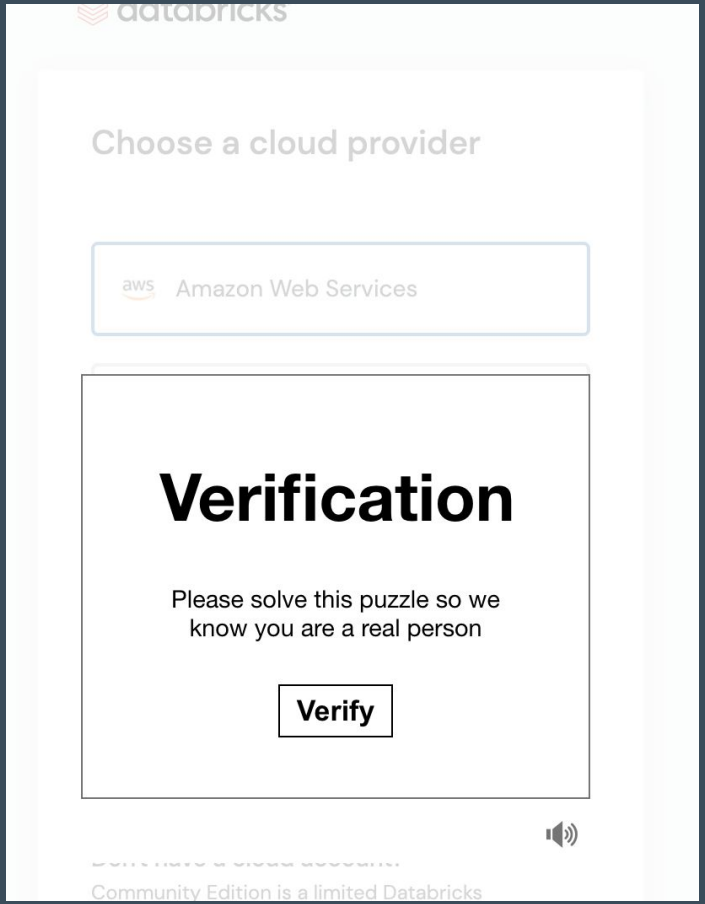
databricks.com/try

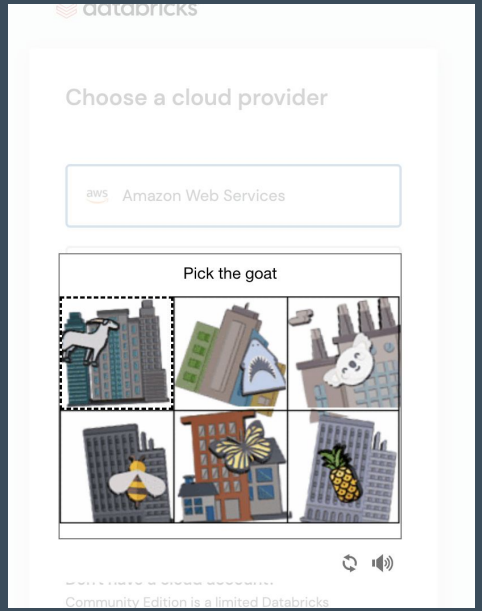
and choose Community Edition

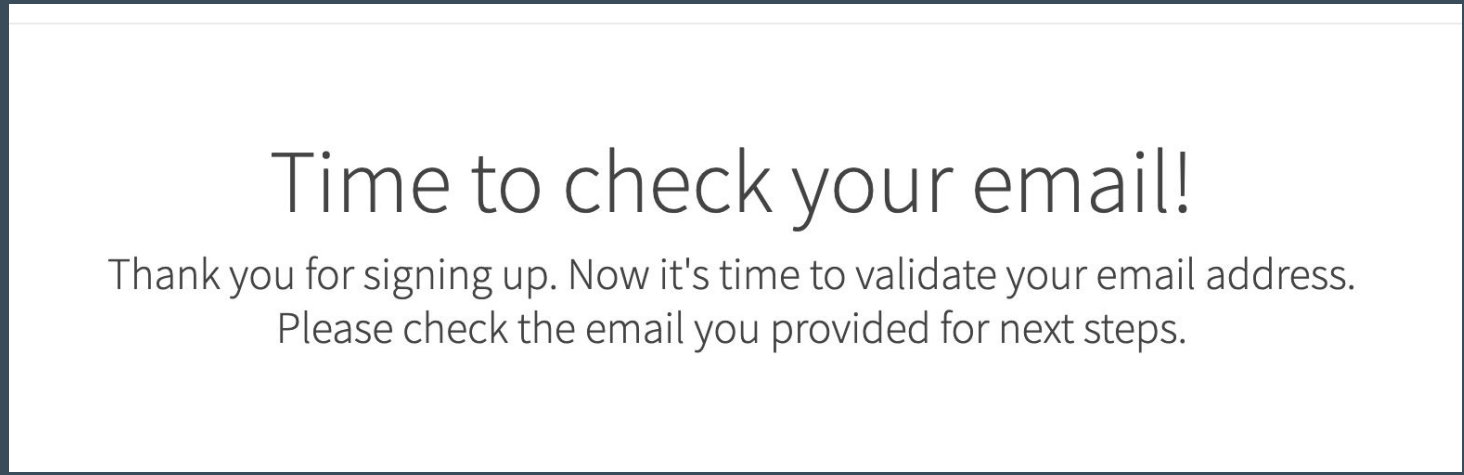
https://databricks.com/try-databricks




databricks
Choose a cloud provider
Amazon Web Services
Microsoft Azure
Google Cloud Platform
Get started
By clicking "Get started", you agree to the Privacy Policy and Terms of Service
Don't have a cloud account?
Community Edition is a limited Databricks
Get started with Community Edition
you agree to the Privacy Policy and Community Edition Terms of Service

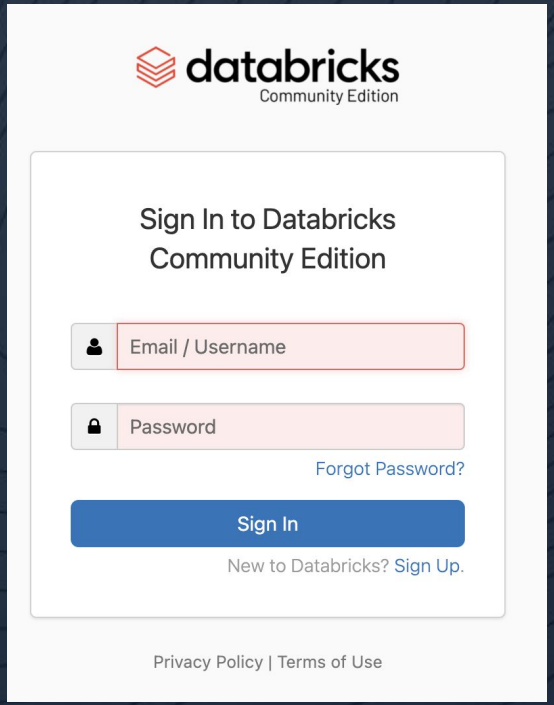

databricks
Choose a cloud provider
aws
Amazon Web Services
Verification
Please solve this puzzle so we know you are a real person
Verify
Community Edition is a limited Databricks

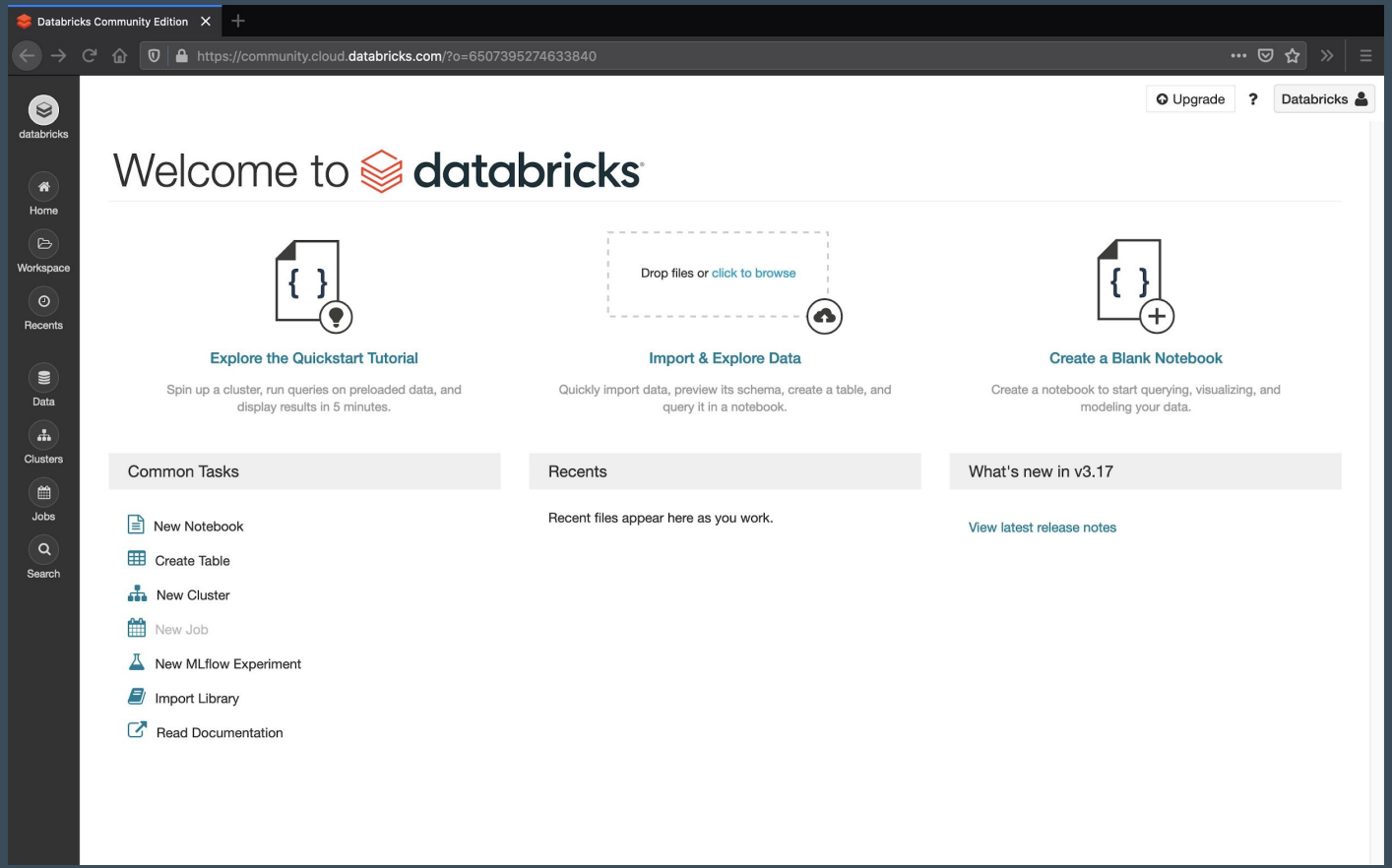

databricks
Choose a cloud provider
aws
Amazon Web Services
Pick the goat
Community Edition is a limited Databricks

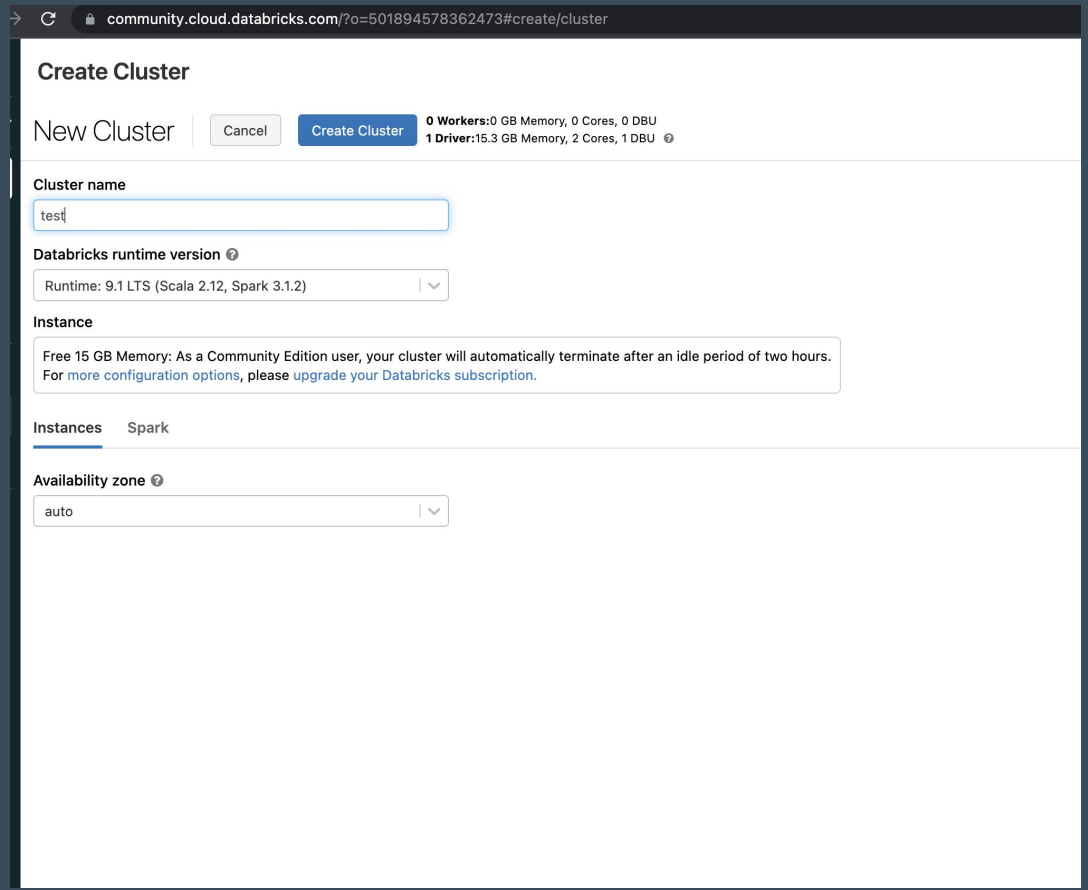


CONFIRM the link in your email and then head back to sign in

https://community.cloud.databricks.com/




Databricks Community Edition
https://community.cloud.databricks.com/?o=6507395274633840
databricks
Home
Workspace
Recents
Data
Clusters
Jobs
Search
Upgrade
Databricks
Welcome to databricks
Explore the Quickstart Tutorial
Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.
Drop files or click to browse
Import & Explore Data
Quickly import data, preview its schema, create a table, and query it in a notebook.
Create a Blank Notebook
Create a notebook to start querying, visualizing, and modeling your data.
Common Tasks
New Notebook
Create Table
New Cluster
New Job
New MLflow Experiment
Import Library
Read Documentation
Recents
Recent files appear here as you work.
What's new in v3.17
View latest release notes

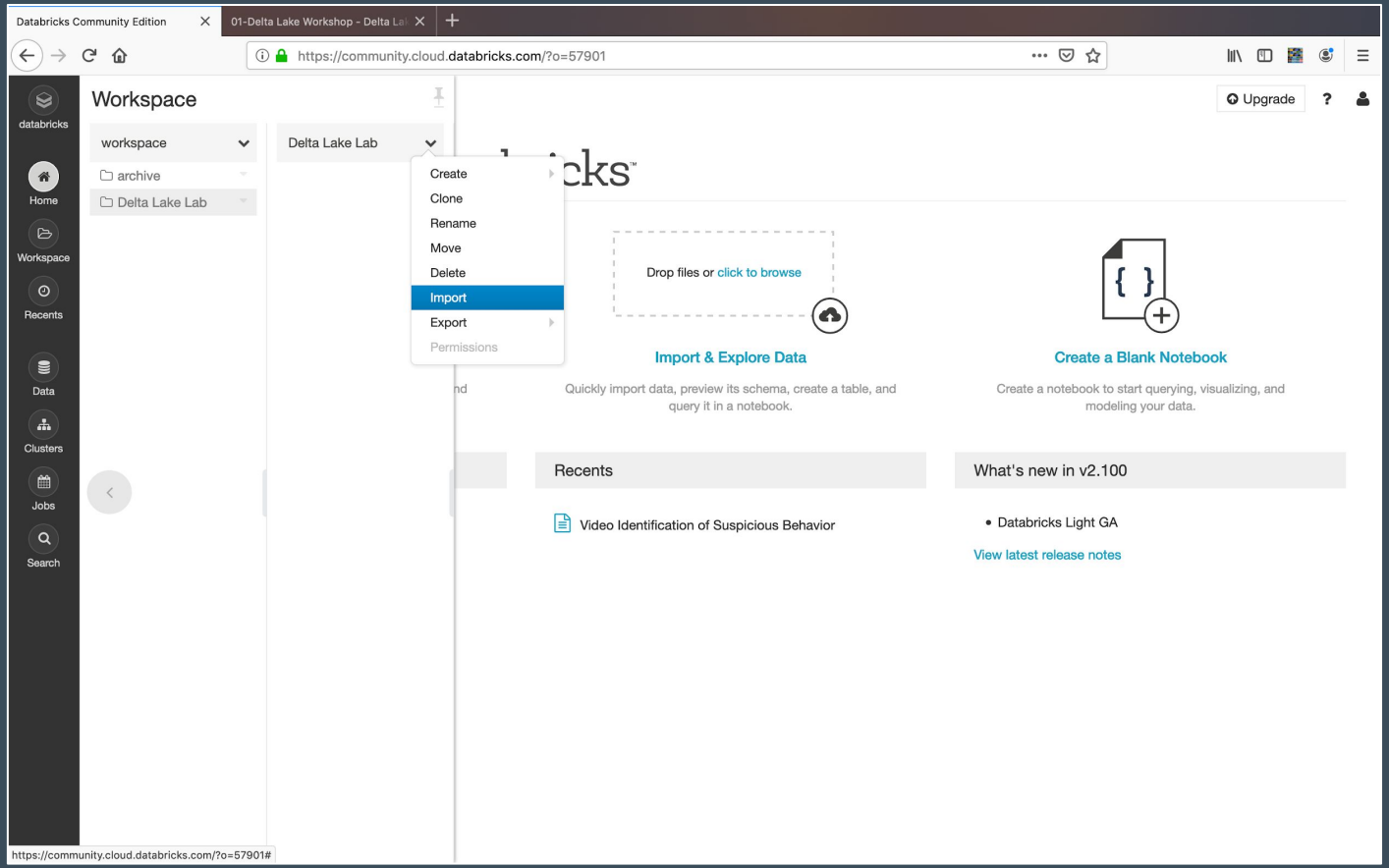

community.cloud.databricks.com/?o=501894578362473#create/cluster
Create Cluster
New Cluster
Cancel
Create Cluster
0 Workers:0 GB Memory, 0 Cores, 0 DBU
1 Driver:15.3 GB Memory, 2 Cores, 1 DBU
Cluster name
test
Databricks runtime version
Runtime: 9.1 LTS (Scala 2.12, Spark 3.1.2)
Instance
Free 15 GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours.
For more configuration options, please upgrade your Databricks subscription.
Instances
Spark
Availability zone
auto

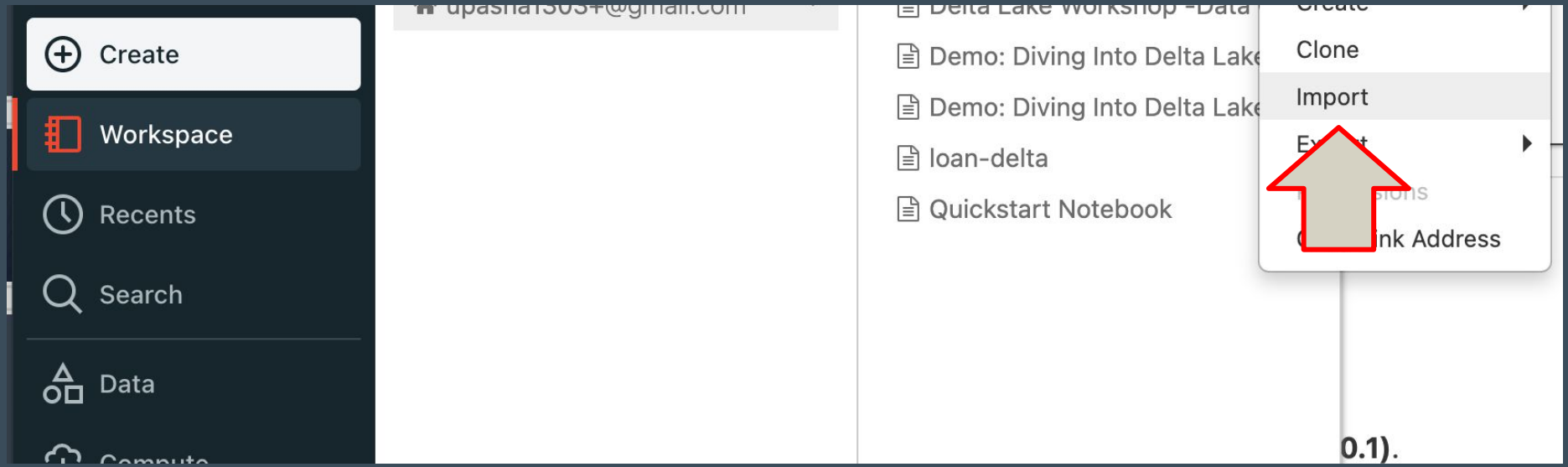
HoL Notebooks

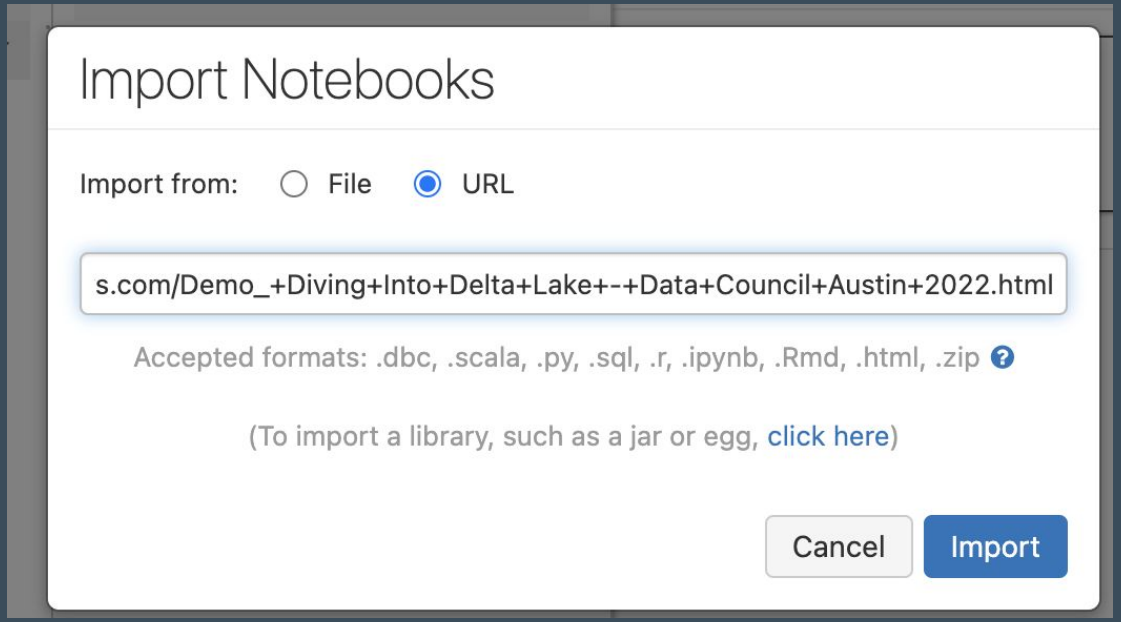
https://delta-lab-public.s3.us-west-2.amazonaws.com/Demo_+Diving+Into+Delta+Lake+-+Data+Council+Austin+2022.html

OR

https://tinyurl.com/delta-dc22

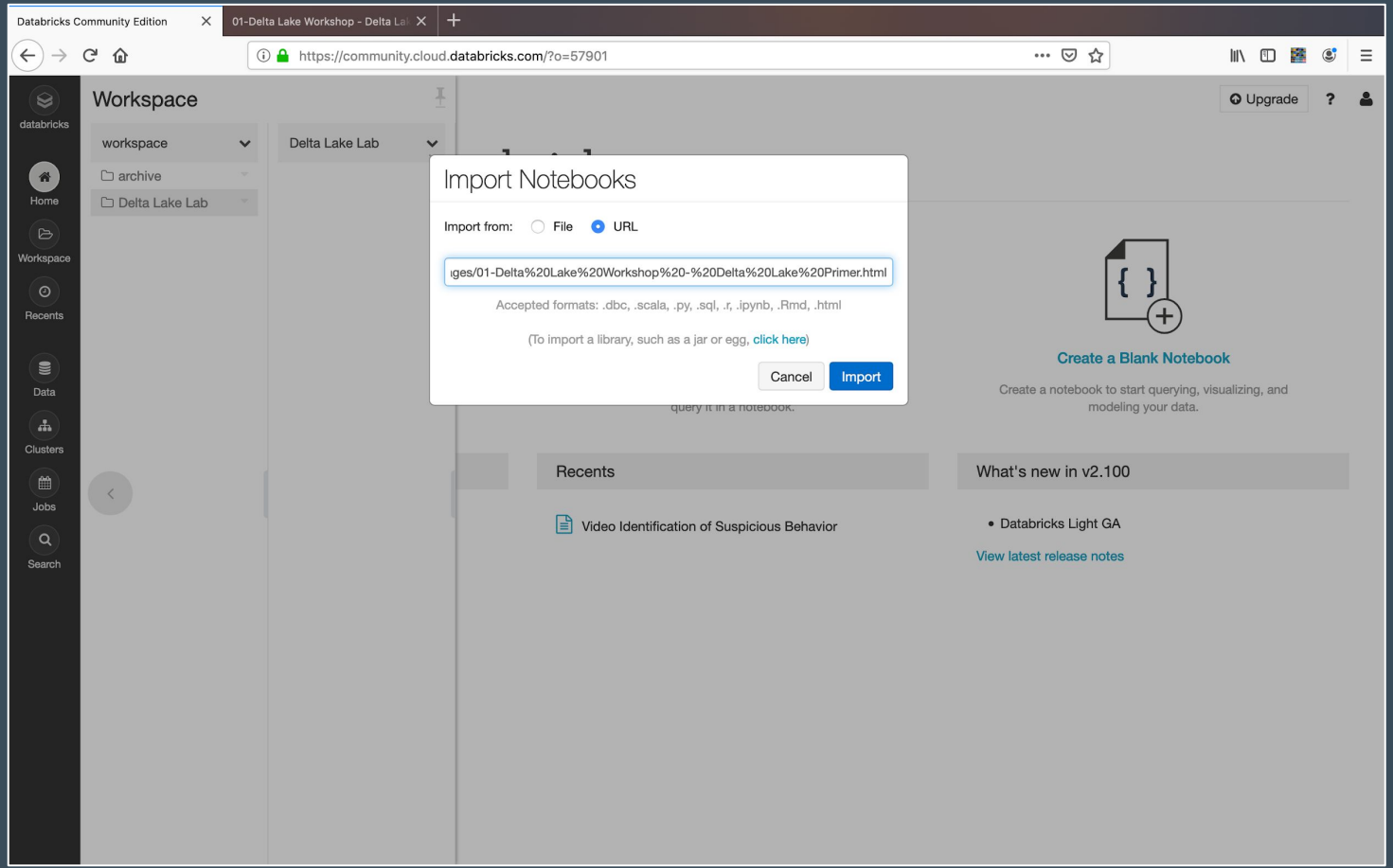

Databricks Community Edition
01-Delta Lake Workshop - Delta La
https://community.cloud.databricks.com/?o=57901
databricks
Workspace
Home
Workspace
Recents
Data
Clusters
Jobs
Search
workspace
archive
Delta Lake Lab
Delta Lake Lab
Create
Clone
Rename
Move
Delete
Import
Export
Permissions
Upgrade
Drop files or click to browse
Import & Explore Data
Quickly import data, preview its schema, create a table, and query it in a notebook.
Create a Blank Notebook
Create a notebook to start querying, visualizing, and modeling your data.
Recents
Video Identification of Suspicious Behavior
What's new in v2.100
Databricks Light GA
View latest release notes
https://community.cloud.databricks.com/?o=57901#


Create
Workspace
Recents
Search
Data
Demo: Diving Into Delta Lake
Demo: Diving Into Delta Lake
loan-delta
Quickstart Notebook
Clone
Import
0.1).



Use the full qualified URL

https://delta-lab-public.s3.us-west-2.amazonaws.com/Demo_+Diving+Into+Delta+Lake+-+Data+Council+Austin+2022.html


Databricks Community Edition
01-Delta Lake Workshop - Delta La
https://community.cloud.databricks.com/?o=57901
Workspace
databricks
Home
Workspace
Recents
Data
Clusters
Jobs
Search
workspace
archive
Delta Lake Lab
Delta Lake Lab
Upgrade
Import Notebooks
Import from: File URL
ges/01-Delta%20Lake%20Workshop%20-%20Delta%20Lake%20Primer.html
Accepted formats: .dbc, .scala, .py, .sql, .r, .ipynb, .Rmd, .html
(To import a library, such as a jar or egg, click here)
Cancel
Import
Create a Blank Notebook
Create a notebook to start querying, visualizing, and modeling your data.
Recents
Video Identification of Suspicious Behavior
What's new in v2.100
Databricks Light GA
View latest release notes

community.cloud.databricks.com/?o=501894578362473#notebook/3349070534507875/command/3349070534507876

Demo: Diving Into Delta Lake - Data Council Austin 2022 Python

Detached | File | Edit | View: Standard | Permissions | Run All | Clear | Publish | Comments | Experiment | Revision history

Cmd 1

Markdown

Welcome to Data Council 2022

Workshop: Building Lakehouse with Delta Lake

The data used is a modified version of the public data from Lending Club. It includes all funded loans from 2012 to 2017. Each loan includes applicant information provided by the applicant as well as the current loan status (Current, Late, Fully Paid, etc.) and latest payment information. For a full view of the data please view the data dictionary available here.

Steps to run this notebook

You can run this notebook in a Databricks environment. Specifically, this notebook has been designed to run in Databricks Community Edition as well. To run this notebook, you have to create a cluster with version **Databricks Runtime 7.4 or later** and attach this notebook to that cluster.

Cmd 2

Setup

Cmd 3

```
1  import random
2  from datetime import datetime
3  from pyspark.sql.functions import *
4  from pyspark.sql.types import *
5
6
7  def my_checkpoint_dir():
8    return "/tmp/delta_demo/chkpt/%s" % str(random.randint(0, 10000))
9
10 # User-defined function to generate random state
```