

# Big Data and AI for Business Databricks



UNIVERSITY OF  
MARYLAND

---

ROBERT H. SMITH  
SCHOOL OF BUSINESS

PROF. ADAM LEE

# Databricks

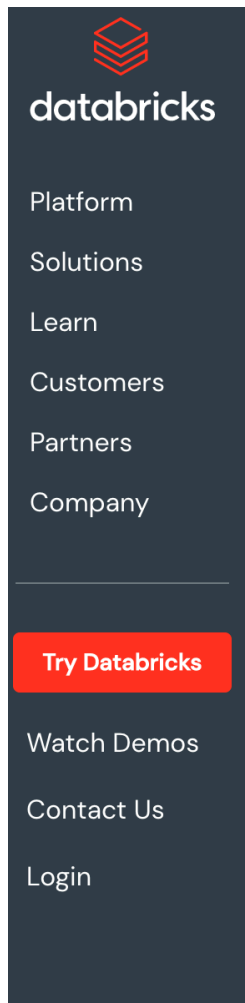
<https://databricks.com/>



A screenshot of the Databricks website. The left sidebar contains the Databricks logo and navigation links: Platform, Solutions, Learn, Customers, Partners, Company, Try Databricks (highlighted in red), Watch Demos, Contact Us, and Login. The main content area has a dark blue background with a grid pattern. At the top, a Gartner report banner states: "2021 Gartner reports: From data warehousing to machine learning, Databricks is a Leader. Learn why the Databricks Lakehouse Platform is able to deliver on both data warehousing and machine learning use cases. Get the reports →". The central text reads "All your data, analytics and AI on one platform". At the bottom, there are two buttons: "Get Started for Free" (red) and "Learn more" (white).

# University Alliance

<https://databricks.com/university/>



## University Alliance

Resources and materials for educators, students and aspiring data scientists who want to build with Databricks

Try Databricks

Join as an educator

I'm a student



At Databricks, we believe that university students should learn the latest data science tools to enhance their value in the workforce upon graduation. The Databricks University Alliance

UNIVERSITY OF  
MARYLAND

<https://databricks.com/university#student>

## Are you a student or aspiring data scientist?

You don't need to wait to start learning Databricks.  
Check out the resources available to you right now.

Learning series

### **Hands-on workshops**

This self-paced online workshop series is for anyone and everyone interested in learning about data analysis. No previous programming experience required.

[Learn more →](#)

Databricks account

### **Sign up for the Databricks Community Edition**

Sign up for a free Databricks account to follow along with tutorials and experiment with data.

[Sign up free →](#)

Databricks Academy


### **Access free self-paced courses on Databricks**

If you're a student with a university-provided email address, you can access courses on Databricks Academy free.

[See details →](#)

# Databricks Community Edition

<https://databricks.com/try-databricks>

  
**databricks**

[Platform](#)  
[Solutions](#)  
[Learn](#)  
[Customers](#)  
[Partners](#)  
[Company](#)

[Try Databricks](#)

[Watch Demos](#)  
[Contact Us](#)  
[Login](#)

## Try Databricks for free

An open and unified data analytics platform for data engineering, data science, machine learning, and analytics. From the original creators of Apache Spark™, Delta lake, MLflow, and Koalas.



### Databricks trial:

- Collaborative environment for data teams to build solutions together.
- Interactive notebooks to use Apache Spark™, SQL, Python, Scala, Delta Lake, MLflow, TensorFlow, Keras, Scikit-learn and more.
- Available as a 14-day full trial in your own cloud, or as a lightweight trial hosted by Databricks.

### Used by:



### Please tell us about yourself

**First Name: \***

Adam

**Last Name: \***

Lee

**Company \***

University of Maryland

**Company Email \***

adamlee@cs.umd.edu

**Title \***

Environmental Health, Safety Associate

**Phone Number**

3017939611

☐ Keep me informed with occasional updates about Databricks and related open source products

By Clicking "Get Started For Free", you agree to the [Privacy Policy](#).

**GET STARTED FOR FREE**


# Community Edition



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 environment for personal use and training.

[Get started with Community Edition](#)


By clicking "Get started with Community Edition", you agree to the [Privacy Policy](#) and [Community Edition Terms of Service](#)




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




## Data Science & Engineering




**Notebook**  
Create a new notebook for querying, data processing, and machine learning.  
[Create a notebook](#)



**Partner Connect**  
[Fivetran, dbt](#)  
[Tableau, Power BI](#)  
[View all partners](#)




**Data import**  
Quickly import data, preview its schema, create a table, and query it in a notebook.  
[Browse files](#)




















**Guide: Quickstart tutorial**  
Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.  
[Start tutorial](#)

### Recents

Name	Last viewed
 <p>There are no recents yet</p>	

Upgrade

# 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**  
BUDT737

**Databricks runtime version ?**  
Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1) ▼

**Databricks Runtime**

10.5 Beta	Scala 2.12, Spark 3.2.1
10.5 ML Beta	Scala 2.12, Spark 3.2.1
10.4 LTS	Scala 2.12, Spark 3.2.1
10.4 LTS ML	Scala 2.12, Spark 3.2.1
10.3	Scala 2.12, Spark 3.2.1
10.3 ML	Scala 2.12, Spark 3.2.1
10.2	Scala 2.12, Spark 3.2.0
10.2 ML	Scala 2.12, Spark 3.2.0
10.1	Scala 2.12, Spark 3.2.0
10.1 ML	Scala 2.12, Spark 3.2.0
10.0	Scala 2.12, Spark 3.2.0
10.0 ML	Scala 2.12, Spark 3.2.0
9.1 LTS	Scala 2.12, Spark 3.1.2
9.1 LTS ML	Scala 2.12, Spark 3.1.2
7.3 LTS	Scala 2.12, Spark 3.0.1
7.3 LTS ML	Scala 2.12, Spark 3.0.1
6.4 Extended Support	Scala 2.11, Spark 2.4.5

Cluster will automatically terminate after an idle period of two hours.  
[View Databricks pricing and subscription.](#)



# Databricks runtime releases

Version	Release Date	End-of-support	Spark	Delta Lake	MLflow
10.4 LTS	Mar 18, 2022	Mar 18, 2024	3.2.1	1.1.0	1.24.0
10.3	Feb 2, 2022	Aug 2, 2022	3.2.1	1.1.0	1.23.0
10.2	Dec 22, 2021	Jun 22, 2022	3.2.0	1.1.0	1.22.0
10.1	Nov 10, 2021	May 10, 2022	3.2.0	1.0.0	1.21.0
10.0	Oct 20, 2021	Apr 20, 2022	3.2.0	1.0.0	1.20.2
9.1 LTS	Sep 23, 2021	Sep 23, 2023	3.1.2	1.0.0	1.20.2
7.3 LTS	Sep 24, 2020	Sep 24, 2022	3.0.1	0.7.0	1.11.0
6.4 ES	Mar 5, 2021	Jun 30, 2022	2.4.5	0.5.0	1.5.0



UNIVERSITY OF  
MARYLAND

ROBERT H. SMITH  
SCHOOL OF BUSINESS

# Databricks File System (DBFS)

The screenshot displays the Databricks web interface. On the left is a dark sidebar with the Databricks logo and navigation links: 'Data Science & E...', 'Create', 'Workspace', 'Recents', 'Search', 'Data' (highlighted with a red box), 'Compute', and 'Jobs'. The main area is divided into two panels. The top of the main area has tabs for 'Database Tables' and 'DBFS' (highlighted with a red box), and an 'Upload' button (also highlighted with a red box). Below the tabs, the path '/FileStore' is shown. The left panel of the main area shows a 'Prefix search' bar and a list of folders: 'FileStore' (highlighted with a red box) and 'user'. The right panel of the main area also shows a 'Prefix search' bar and a list of folders: 'tables' (highlighted with a red box). A green checkmark and a dropdown arrow are visible in the top right corner of the main area.



UNIVERSITY OF  
MARYLAND

ROBERT H. SMITH  
SCHOOL OF BUSINESS

# Database Tables

The screenshot displays the Databricks web interface for creating a new table. On the left sidebar, the 'Data' menu item is highlighted with a red box. The main panel shows the 'Database Tables' tab selected, with a 'Create Table' button highlighted in a red box. The 'Tables' list shows a table named 'hw4\_product\_json'. On the right, the 'Create New Table' panel is visible, with the 'DBFS' data source selected and highlighted in a red box. The 'Create Table with UI' button is also highlighted in a red box. The interface shows the file structure for the selected data source, including 'FileStore' and 'user' folders.

**Database Tables** DBFS

**Create Table**

**Create New Table**

Data source ?

Upload File S3 **DBFS** Other Data Sources Partner Integrations

Select a file from DBFS ?

FileStore user

tables

/FileStore/tables

**Create Table with UI** Create Table in Notebook ?



UNIVERSITY OF  
MARYLAND

ROBERT H. SMITH  
SCHOOL OF BUSINESS

# Create Notebook

**databricks**

Data Science & E... ▾

⊕ Create

**Workspace**

⌚ Recents

🔍 Search

**Workspace** ▾

Shared ▾

**Users** ▾

**Users**

🏠 adamlee

🏠 Home

**Data Science & Engineerin**

▼

Create ▸

Clone

**Notebook**

Library

Folder

MLflow Experiment

**Data import**

Quickly import data, prev  
create a table. and querv

**Create Notebook**

Name

Act10

Default Language

Python ▾

Cluster

BUDT737 ▾

Cancel

**Create**