

# SDSC 2019

# Apache Spark Bootcamp

Databricks Community Edition Signup Instruction

# Why Databricks?

Hi friends,

I chose Community Edition of Databricks platform for our hands-on exercises so that you guys don't have to install Spark and its dependencies on your laptop.



databricks®



# Sign up for Databricks Community Edition

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

Launch cloud-optimized Apache Spark™ clusters in minutes

## DATABRICKS PLATFORM – FREE TRIAL

For businesses looking for a zero-management cloud platform built around Spark

- Unlimited clusters that can scale to any size
- Job scheduler to execute jobs for production pipelines
- Fully interactive notebook with collaboration, dashboards, REST APIs
- Advanced security, role-based access controls, and audit logs
- SAML 2.0 support
- Deployed to your AWS VPC
- Integration with BI tools such as Tableau, Qlik, and Looker
- 14-day full feature trial (excludes AWS charges)

GET STARTED – FREE

## COMMUNITY EDITION

For students and educational institutions just getting started with Spark

- Single cluster limited to 6GB and no worker nodes
- Basic notebook without collaboration
- Limited to 3 max users
- Public environment to share your work

GET STARTED

We are going to use the **Community Edition**.

Click “Get Started”

# Fill out your information

Don't worry. They are not going to ask you for your credit card information. However, if you provide your phone number (optional), I bet one of the sales reps is going to contact you. Choose at your discretion. They are still going to contact you via email (required) though.

I am in no relation to Databricks, but since we are using their product, I feel obligated to mention that Databricks is a company founded by the creators of Spark and they are one of the BigData platform vendors you should consider if your company is really committed to using Spark.

## Sign Up for Databricks Community Edition

First Name \*

Last Name \*

Company Name \*

Work Email \*


Password \*

Confirm Password \*

Phone Number

What is your intended use case? \*  
- Please Select -

How would you describe your role? \*  
- Please Select -

☐ I'm not a robot   
reCAPTCHA  
Privacy - Terms

You can choose any option. I don't think it matters. I would go with either "Personal - Learning Spark" or "Company - Research and Learn Spark"

- ✓ - Please Select -
- Personal - Learning Spark
- Personal - Taking an Online Course
- Personal - Implementing Spark side projects
- Company - Research and Learn Spark
- Company - Development and POC of a Spark application
- Company - Productionize Spark
- Education - Teaching
- Research - Academia

# Agree to the Terms of Service

I hope there isn't any unagreeable terms...

## Terms of Service

such courts. You agree these Terms will be governed by California law except for its conflict of laws principles. These Terms constitute the entire agreement between you and Databricks concerning the Services and these Terms create no third party beneficiary rights.

**Termination, Modification, Waiver & Assignment.** Either of us may terminate your use of the Services at any time and for any reason; however, obligations of these Terms that by their nature should survive termination shall so survive. In addition, we may revise these Terms from time to time, and will always post the most current version on our website. If we elect to terminate your account, or if a revision of these Terms meaningfully reduces your rights, we will make a reasonable attempt to notify you (by, for example, sending a message to the email address associated with your account or posting for a reasonable time period a message to the login page of the Services). By continuing to use or access the Services after the revisions come into effect, you agree to be bound by the revised Terms. Databricks' failure to enforce a provision of these Terms is not a waiver of its right to do so later. If a provision is found unenforceable, the remaining provisions of the Terms will remain in full effect and an enforceable term will be substituted reflecting our intent as closely as possible. You may not assign any of your rights under these Terms, and any such attempt will be void. Databricks may assign its rights to any of its affiliates or to any successor in interest.

**BY CLICKING "AGREE" BELOW YOU ARE AGREEING THAT YOU HAVE CAREFULLY READ ALL OF THE ABOVE PROVISIONS, INCLUDING THE RISKS AND RESTRICTIONS RELATING TO YOUR DATA AND TO YOUR USE OF THE SERVICES, AND YOU ARE AGREEING TO BE BOUND BY ALL OF THE ABOVE PROVISIONS. IF YOU DO NOT AGREE WITH THESE TERMS OR DO NOT UNDERSTAND THEM, DO NOT CLICK AGREE.**

[Download a print version](#)CancelAgree

Scroll through the Terms of Service and click "Agree"

# Wait for Confirmation Email


## Please Confirm Your Email Address

You will receive an email with a link to confirm your email address. Please click the link to complete the signup process. **If you haven't received the email, please check your spam folder.**

Contact [feedback@databricks.com](mailto:feedback@databricks.com) if you have any questions.

# Verify Email and Sign in

Type in your email and password.  
And then click “Sign In”



**Welcome to Databricks Community Edition!**


Databricks Community Edition provides you with access to a free micro-cluster as well as a cluster manager and a notebook environment - ideal for developers, data scientists, data engineers and other IT professionals to get started with Spark.

We need you to verify your email address by clicking on [this link](https://accounts.cloud.databricks.com/signup/validate?emailToken=55724f82617563635749a0f34638f4a8&ce=true). You will then be redirected to Databricks Community Edition!

**Get started by visiting:** <https://accounts.cloud.databricks.com/signup/validate?emailToken=55724f82617563635749a0f34638f4a8&ce=true>

If you have any questions, please contact [feedback@databricks.com](mailto:feedback@databricks.com).

- The Databricks Team



**Sign In to Databricks**

[Forgot Password?](#)

**Sign In**

[New to Databricks? Sign Up.](#)

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

Click “this link”

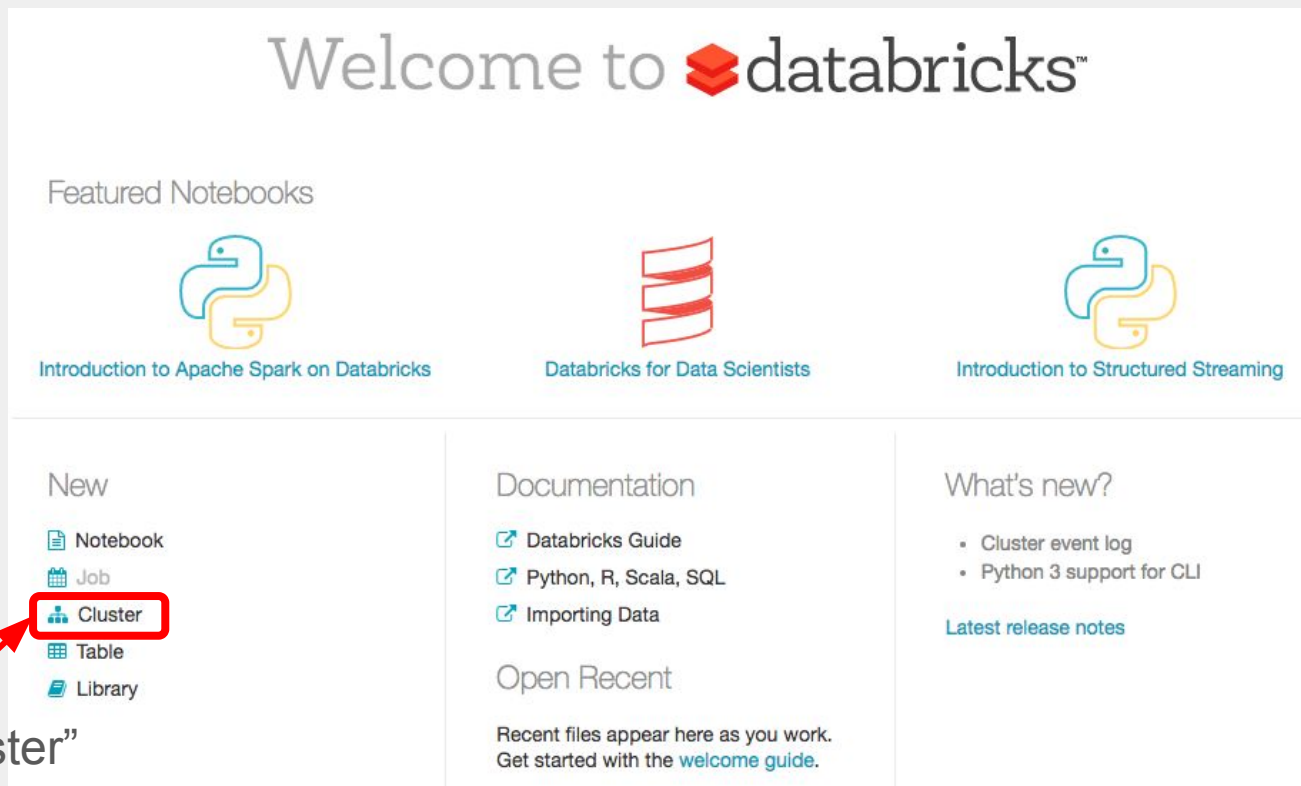
# You are done!

You can start reading the documentations they provide and start playing with Spark by spinning up a cluster and creating notebooks.



# Extra Credit 1

# Creating a cluster (for the course)



Click "Cluster"

# Creating a cluster

Use the latest version of Spark. It should be Spark 2.4.x with Scala 2.11

Give your cluster a name and click “Create Cluster”. Other settings are not important to the course.

Change Python Version to 3 if it is not already 3.

Create Cluster

New Cluster   0 Workers: 0.0 GB Memory, 0 Cores, 0 DBU  
1 Driver: 6.0 GB Memory, 0.88 Cores, 1 DBU ⓘ

Cluster Name  Please enter a cluster name

Databricks Runtime Version ⓘ  
3.5 LTS (includes Apache Spark 2.2.1, Scala 2.11) ⌵

Python Version ⓘ  
2 ⌵

Instance

Free 6GB 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 ⓘ  
us-west-2c ⌵

# Extra Credit 2

# To run my demo code yourselves...

I also prepared a set of locally runnable examples in Scala and Python. I'll share the repository on the day of the workshop. To run these examples yourself, you need to install a few things on your machines.

You can use git to clone my repository, but alternatively, you can just download as a zip file from GitHub.

# To run Scala demo code yourselves...

I prepared SBT and Maven examples. You need to install the followings:

- Java 8 JDK (either Open JDK or Oracle JDK)
- IntelliJ for playing with Scala examples
  - with Scala 2.11.12 SDK
- If you don't want to install IntelliJ, but you still want to run Scala demo, you at least need to have Maven (3.x), SBT (1.x), and their dependencies installed. Note that IntelliJ has bundled version of Maven and SBT. Those bundled versions should work.

# To run Python demo code yourselves...

I also prepared a simple Python example just to demo how you should structure your project. If you would like to run Python example, you need to install

- Python3
- pip (for Python3)
- PyCharm or your favorite Python IDE for playing with the demo