



Introduction to Big Data

DHKHDL16A

Databricks and PySpark

Trong bài lab này, các bạn sẽ thực hiện:

1) Đăng kí một tài khoản free trên Databricks <https://databricks.com/try-databricks>

Sau khi bạn điền thông tin, màn hình sau đây sẽ xuất hiện



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](#)

Bạn click vào "Get started with Community Edition"

2) Log vào Databricks Community Edition <https://community.cloud.databricks.com/>

3) Đọc tutorial <https://databricks.com/spark/getting-started-with-apache-spark/quick-start>

Đối với các sample notebooks, các bạn copy link từ "Copy link for import"

Apache Spark Quick Start Python notebook

Open notebook in new tab

Copy link for import

Quick Start Using Python


- Using a Databricks notebook to showcase DataFrame operations using Python
- Reference <http://spark.apache.org/docs/latest/quick-start.html>

```
# Take a look at the file system
display(dbutils.fs.ls("/databricks-datasets/samples/docs/"))
```

Expand notebook ▼

	name	size
dbfs:/databricks-datasets/samples/docs/README.md	README.md	3137

sau đó import vào Databricks => Workspace => Users => Your user name => import

 **databricks**

Data Science & ... ▼

⊕ Create

Workspace

Recents

Search

Workspace

Users ▼

nguyenchikien@iuh.edu.vn ▼

big_data_analytics_1

binary-classification

decision-trees

Exercise 6 - Linear Regression

gbt-regression

linear_regression

my first

Home

nguyenchikien@iuh.edu.vn ▼

Create

Clone

Import

Export

Permissions

Copy Link Address