# Create an AWS Account

Visit [**aws.amazon.com**](https://aws.amazon.com/) and click on the "Create an AWS Account" button.

If you have an AWS account already, sign in.

If you do not have an AWS account, sign up.

* When you sign up, you will need to provide a credit card. But don’t worry, you won’t be charged for anything yet.
* You will also need to choose a support plan. You can choose the free Basic Support Plan.
* Once you finish signing up, wait a few minutes to receive your AWS account confirmation email. Then return to [**aws.amazon.com**](https://aws.amazon.com/) and sign in.

**Starter Code**

Below, you'll find starter code to create a spark session and read in the full 12 GB dataset for the DSND Capstone Project, Sparkify. You can use the following link to access the public dataset:

**Full Sparkify Dataset: s3n://udacity-dsnd/sparkify/sparkify\_event\_data.json**

You can also use the link below to access the mini 123MB dataset:

**Mini Sparkify Dataset: s3n://udacity-dsnd/sparkify/mini\_sparkify\_event\_data.json**

And here is the code which you can copy from and paste easily into your notebook.

*# Starter code*

**from** pyspark.sql **import** SparkSession

*# Create spark session*

spark = SparkSession \

.builder \

.appName("Sparkify") \

.getOrCreate()

*# Read in full sparkify dataset*

event\_data = "s3n://udacity-dsnd/sparkify/sparkify\_event\_data.json"

df = spark.read.json(event\_data)

df.head()

When you run the last cell, you'll see a box appear that says "Spark Job Progress." Click on the arrow in that box to view your cluster's progress as it reads the full 12GB dataset! (Screenshot of this is shown below.)

Now, you're ready to start analyzing this data with Spark! Remember, you can still use the mini dataset to work faster while you explore your data and develop your model:

**Mini Sparkify Dataset: s3n://udacity-dsnd/sparkify/mini\_sparkify\_event\_data.json**

**ATTENTION: As mentioned in the previous page, remember to terminate your cluster and delete your resources when you're finished working on your project to avoid unexpected costs. See the next page for details.**