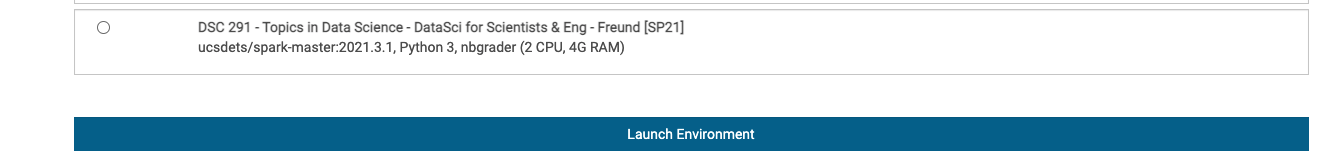
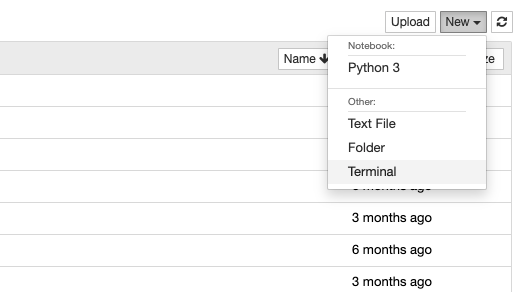
# Starting a spark cluster

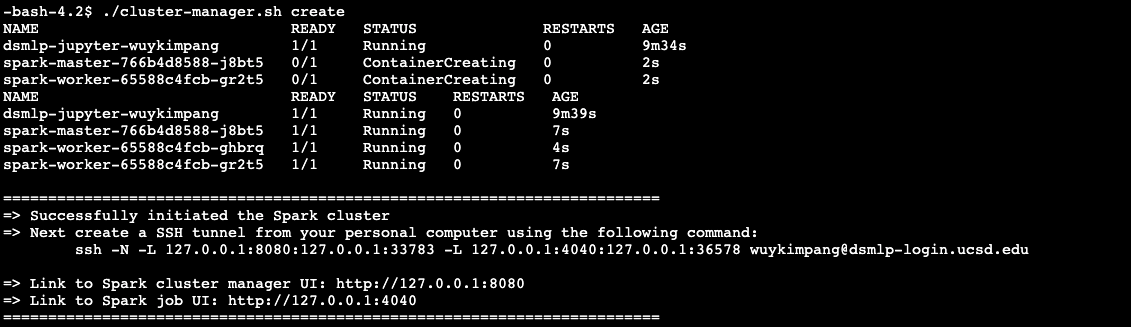
1. Login to datahub.ucsd.edu
2. Select the “DSC 291” environment and click the “Launch Environment” button



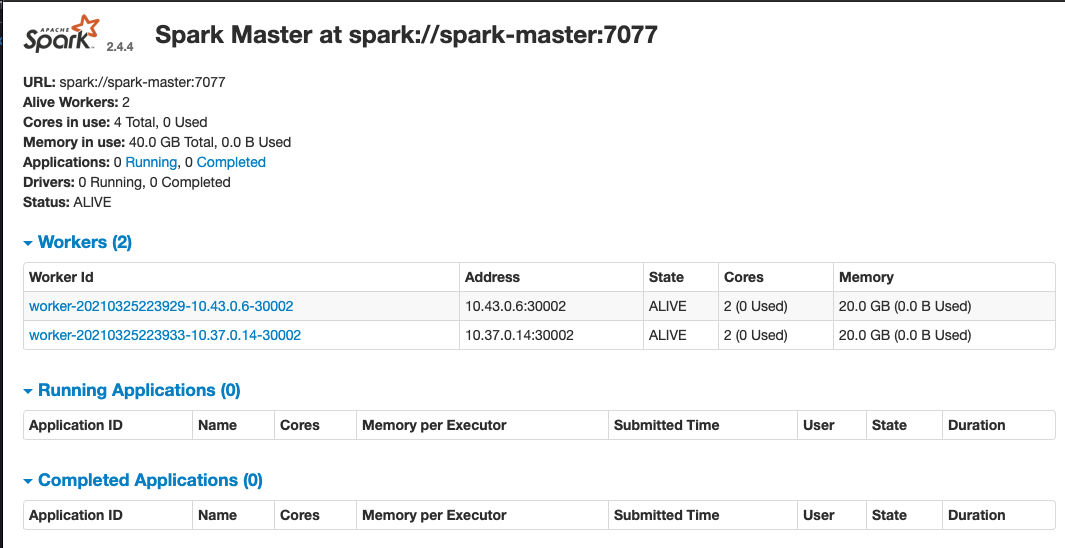
1. Once your notebook server is running, select “New > Terminal”



1. Switch to the terminal tab in your web browser
2. Once at the terminal, run the command “**ssh <username>@dsmlp-login**”. Replace <username> with your username.
3. Download the scripts for your course by running the command “**git clone** [**https://github.com/ucsd-ets/dsc291-spark-cluster.git**](https://github.com/ucsd-ets/dsc291-spark-cluster.git)”
4. You should now see a directory called dsc291-spark-container within your terminal (run the command “**ls**” to list the contents of the directory).
5. Change into the directory by running the command “**cd dsc291-spark-container**”
6. To start your spark cluster, run the command “**./cluster-manager.sh create**”. You should see output on your screen like the example below:



1. Open a new terminal on your local computer (**not in Datahub**) and paste the generated ssh command into it (command right below “Next create a SSH tunnel…”). This will open a tunnel between datahub servers and your local computer. Leave this terminal open to keep the tunnel open. ***Note: make sure you’re connected to UCSD’s VPN***
2. Open a new tab in your browser and navigate to <http://127.0.0.1:8080>. The Apache Spark dashboard will be there. See example below:

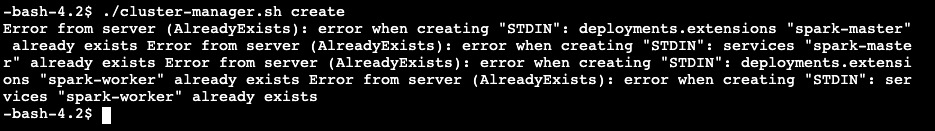


1. You can now close the terminal. ***Note: your cluster will only be active for 3 hours. You’ll have to recreate it starting from step 3 in case it shuts down while you’re working with it.***
2. You can now start accessing the generated spark cluster within your jupyter server on Datahub. Please speak with your Instructor or TA about how to access it.

# Common problems

## Different output at step 9

If you see output like the following:



That means that your spark cluster is already running and you may start using it. You can also recreate the cluster by running the command “**./cluster-manager.sh delete**” and then the command “**./cluster-manager.sh create**”.