



# Databricks

## Create database

```
%sql CREATE DATABASE tmp
```

## Run a notebook in another notebook

```
%run /HOME_FOLDER/notebook
```

## Mount a blob

```
dbutils.fs.mount(  
    source = "wasbs://CONTAINER@ACCOUNT_NAME.blob.core.windows.net",  
    mount_point = "/mnt/CONTAINER",  
    extra_configs =  
    {"fs.azure.account.key.ACCOUNT_NAME.blob.core.windows.net": "ACCOUNT_KEY"})
```

## Unmount a blob

```
dbutils.fs.unmount("/mnt/CONTAINER")
```

## Upgrade pandas and numpy

```
dbutils.library.installPyPI('numpy', '1.16.3')  
dbutils.library.installPyPI('pandas', '0.24.2')  
dbutils.library.restartPython()
```

## Install packages from PyPi

Click on the clusters tab -> Click cluster name -> library -> Install new -> PyPi -> Name of package

```
dbutils.library.installPyPI("koalas")  
dbutils.library.restartPython()
```



## Add conda magic commands



Ray Bell

Home

About me

Previous work

Management

Blog

Courses

`spark.databricks.conda.condamagic.enabled true`<https://docs.databricks.com/notebooks/notebooks-python-libraries.html>

## Install packages from conda

On command line:

```
%sh /databricks/conda/bin/conda install -y -p /databricks/python -c conda-  
forge fbprophet
```

Then detach and reattach

Click on a cluster -> Advanced Options -> Init Scripts

<http://abizeradenwala.blogspot.com/2018/05/upgrading-python-version-for-databricks.html><https://docs.databricks.com/user-guide/faq/anaconda-environment.html>

## Upgrade python

```
%sh curl -O https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-  
x86\_64.sh -o /dbfs/tmp/
```

```
%sh sudo wget https://repo.anaconda.com/archive/Anaconda3-2018.12-Linux-  
x86_64.sh -O /dbfs/tmp/Anaconda3-2018.12-Linux-x86_64.sh
```

```
%sh /databricks/python/bin/pip freeze > /tmp/python_packages.txt
```

```
clusterName = "anaconda-cluster"
```

```
script = """#!/bin/bash  
cp /dbfs/tmp/Anaconda3-2018.12-Linux-x86_64.sh /tmp  
sudo bash /tmp/Anaconda3-2018.12-Linux-x86_64.sh -b -p /anaconda3  
mv /databricks/python /databricks/python_old  
ln -s /anaconda3 /databricks/python  
cp /dbfs/dbfs/tmp/python_packages.txt /tmp/python_packages.txt  
/databricks/python/bin/pip install -r /tmp/python_packages.txt  
"""
```

```
dbutils.fs.put("dbfs:/databricks/init/%s/install_conda.sh" % clusterName,  
script, True)
```

```
clusterName = "pcp_ts"
```

```
script = """#!/bin/bash
```

```
cp /dbfs/tmp/Anaconda3-2018.12-Linux-x86_64.sh /tmp
```



```
ln -s /anaconda3 /databricks/python
cp /dbfs/dbfs/tmp/python_packages.txt /tmp/python_packages.txt
conda activate /databricks/python
pip install -r /tmp/python_packages.txt
/databricks/python/bin/conda install -y -c conda-forge fbprophet
"""
dbutils.fs.put("dbfs:/databricks/init/%s/install_conda.sh" % clusterName,
script, True)
```

Create a cluster with the the same name as anaconda-cluster

## Create a init script

```
dbutils.fs.mkdirs("dbfs:/databricks/my_init_scripts/")
dbutils.fs.put("/databricks/my_init_scripts/pcp_ts-install.sh", """
#!/bin/bash
wget --quiet -O /mnt/Miniconda3-latest-Linux-x86\_64.sh
https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86\_64.sh
wget --quiet -O /mnt/jars/driver-daemon/postgresql-42.2.2.jar
http://central.maven.org/maven2/org/postgresql/postgresql/42.2.2/postgresql-
42.2.2.jar""", True)
```

```
dbutils.fs.put("/databricks/my_init_scripts/pcp_ts-install.sh", """
#!/bin/bash
set -ex
/databricks/python/bin/python -V
. /databricks/conda/etc/profile.d/conda.sh
conda activate /databricks/python
conda install -y fbprophet""", True)
```

```
dbutils.fs.rm("/databricks/my_init_scripts/pcp_ts-install.sh")
```

```
dbutils.fs.put("/databricks/my_init_scripts/pip-install.sh", """
#!/bin/bash
/databricks/python/bin/pip install --upgrade pip""", True)
```

## See packages with runtime

<https://docs.databricks.com/release-notes/runtime/7.3.html#system-environment>



Create token

## Put config info into a cluster

Config looks like:

```
spark.conf.set("dfs.adls.oauth2.access.token.provider.type",  
"ClientCredential")  
spark.conf.set("dfs.adls.oauth2.client.id", "X")  
spark.conf.set("dfs.adls.oauth2.credential", "X")  
spark.conf.set("dfs.adls.oauth2.refresh.url",  
"https://login.microsoftonline.com/X/oauth2/token")
```

