<https://towardsdatascience.com/installing-pyspark-with-java-8-on-ubuntu-18-04-6a9dea915b5b>

**/etc/environment (add the line): use sudo vi /etc/environment to edit**

JAVA\_HOME="/usr/lib/jvm/java-8-openjdk-amd64"

**.bashrc (add the lines)**

source /etc/environment

export SPARK\_HOME=~/Downloads/spark-2.4.3-bin-hadoop2.7

export PATH=$PATH:$SPARK\_HOME/bin

#export PATH=$PATH:~/anaconda3/bin

#export PYTHONPATH=$SPARK\_HOME/python:$PYTHONPATH

#export PYSPARK\_DRIVER\_PYTHON="jupyter"

#export PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook"

export PYSPARK\_PYTHON=python3

export PATH=$PATH:$JAVA\_HOME/jre/bin

Now run

* source ~/.bashrc
* echo $JAVA\_HOME
* sudo **apt** install openjdk-8-jdk
* curl [https://archive.apache.org/dist/spark/spark-2.4.3/spark-2.4.3-bin-hadoop2.7.tgz --output spark-2.4.3-bin-hadoop2.7.tgz](https://archive.apache.org/dist/spark/spark-2.4.3/spark-2.4.3-bin-hadoop2.7.tgz%20--output%20spark-2.4.3-bin-hadoop2.7.tgz)
* mv spark-2.4.3-bin-hadoop2.7.tgz Downloads
* cd Downloads
* tar -zxvf spark-2.4.3-bin-hadoop2.7.tgz
* cd spark-2.4.3-bin-hadoop2.7
* pip3 install pyspark
* pyspark
* spark-shell
* spark-submit test.py