<https://github.com/linkedin/isolation-forest>

SPARK with pyspark

https://spark.apache.org/downloads.html

https://sparkbyexamples.com/spark/spark-web-ui-understanding/

Pip install pyspark

Pyspark

<http://localhost:4040/jobs/>

Mieux comprendre pyspark

https://sparkbyexamples.com/pyspark-tutorial/

Besoin de maven pour créer le .jar spark-iforest de FangZhou Yang

cd spark-iforest/  
  
mvn clean package -DskipTests

Télécharger SPARK

<https://spark.apache.org/downloads.html>

Dézipper dans le dossier voulu : /Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2

Mettre à jour le chemin dans la variable d’environnement $SPARK\_HOME : export SPARK\_HOME=$SPARK\_HOME:/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2

/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2:/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2/bin:/Users/maurrastogbe/opt/anaconda3/bin:/usr/local/bin:/usr/bin:/bin:/usr/sbin:/sbin:.:/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2:.:/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2:/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2/bin

/Users/maurrastogbe/opt/spark/spark-3.0.2-bin-hadoop3.2/bin:/Users/maurrastogbe/opt/anaconda3/bin:/usr/local/bin:/usr/bin:/bin:/usr/sbin:/sbin

Ajouter scala au Anaconda notebook

<https://medium.com/@bogdan.cojocar/how-to-run-scala-and-spark-in-the-jupyter-notebook-328a80090b3b>

Ajouter Apache Zeppelin à Anaconda

conda install -c danielfrg apache-zeppelin

Java

Get the path of the java version: /usr/libexec/java\_home -v 1.8

Add path to JAVA\_HOME: export JAVA\_HOME="/Library/Java/JavaVirtualMachines/jdk1.8.0\_281.jdk/Contents/Home"

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

java -version

Docker

<http://zeppelin.apache.org/download.html>

Docker permet de lancer les applications dans un conteneur à travers des images.

Pour s’assurer que docker fonctionne : docker run hello-world

Use this command to launch Apache Zeppelin in a container.

docker run -p 8080:8080 --rm --name zeppelin apache/zeppelin:0.9.0

To persist logs and notebook directories, use the [volume](https://docs.docker.com/engine/reference/commandline/run/#mount-volume--v-read-only) option for docker container.

docker run -p 8080:8080 --rm -v $PWD/logs:/logs -v $PWD/notebook:/notebook -e ZEPPELIN\_LOG\_DIR='/logs' -e ZEPPELIN\_NOTEBOOK\_DIR='/notebook' --name zeppelin apache/zeppelin:0.9.0

If you have trouble accessing localhost:8080 in the browser, Please clear browser cache.

Check a port: netstat -vanp tcp | grep 8080

Push the image on docker hub: docker tag docker101tutorial maurras/docker101tutorial

docker push maurras/docker101tutorial

<http://localhost/tutorial/>

Maven for linkedIn version of distributed IForest

<https://mvnrepository.com/artifact/com.linkedin.isolation-forest/isolation-forest_2.3.0_2.11/0.3.2>

Tester LinkedIn

docker-compose -f {compose file name} up

/Users/maurrastogbe/Documents/GitHub/OD\_Jupyter/distributed\_IFO/ZeppelinWorkSpace/zeppelin-master/compose/ docker-compose.yml

docker-compose -f Documents/GitHub/OD\_Jupyter/distributed\_IFO/ZeppelinWorkSpace/zeppelin-master/compose/docker-compose.yml up

Utilisation du docker de Zeppelin

docker pull apache/zeppelin:0.9.0

docker run -ti apache/zeppelin:0.9.0

<https://docs.docker.com/docker-hub/>

Zeppelin sans docker

https://zeppelin.apache.org/docs/0.6.0/install/install.html

cd opt/zeppelin-0.9.0-bin-all

bin/zeppelin-daemon.sh start

<http://localhost:8080/#/>

SPARK ML

Les métriques : <https://spark.apache.org/docs/2.2.0/mllib-evaluation-metrics.html>

<https://spark.apache.org/docs/2.2.0/mllib-evaluation-metrics.html>

<https://stackoverflow.com/questions/32531224/how-to-convert-dataframe-to-rdd-in-scala>

<https://webdevdesigner.com/q/how-to-convert-rdd-object-to-dataframe-in-spark-20460/>

Démarrer Zeppelin :

Changement de version Java

JAVA\_HOME="/Library/Java/JavaVirtualMachines/jdk1.8.0\_281.jdk/Contents/Home"

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

Accéder au dossier Zeppelin

cd opt/zeppelin-0.9.0-bin-all

Démarrer Zeppelin

bin/zeppelin-daemon.sh start

opt/zeppelin-0.9.0-bin-all/bin/zeppelin-daemon.sh start