

Équipe: Alpha
Dominique Fauteux-Chapleau - 1798165
Gabriel Beaudry - 1799201
Kevin Pantelakis - 1794745
Michael Sghaier - 1725355

Tutoriel d'installation et d'utilisation du service

1. Démarrer une machine virtuelle Ubuntu version 18+ dessus.
 - a. Attention, il se peut que vous aillez à modifier des paramètres dans votre BIOS pour permettre la virtualization.
2. Installer Python3 si ce n'est pas déjà fait.
 - a. Lancez la commande: "sudo apt install -y python3"
 - b. Vérifier avec "python3 --version" dans un terminal
3. Installer Pip3 pour pouvoir installer les librairies nécessaires en python
 - a. Lancer la commande "sudo apt install -y python3-pip"
4. Copier les fichiers du projet sur la machine virtuelle
5. Naviguez jusqu'au dossier tp4_python_service
6. Lancer la commande suivante pour installer toutes les librairies python nécessaires
 - a. pip3 install flask flask_restful flask_cors pyspark numpy pymongo
7. Lancer la commande suivant pour installer MongoDB
 - a. sudo apt install -y mongodb
8. Toujours sur la machine virtuelle, télécharger apache-spark
 - a. <https://spark.apache.org/downloads.html> prenez le fichier .tgz
 - b. installer java sur la machine "sudo apt install -y default-jdk"
 - c. lancer la commande "sudo mkdir /usr/local/spark"
 - d. sudo tar xvf spark-2.4.0-bin-hadoop2.7.tgz -C /usr/local/spark
 - e. modifier le fichier bashrc(nano .bashrc) et ajoutez ces lignes a la fin du fichier:
 - i. export PATH=\$SPARK_HOME/bin:\$PATH
 - export PYSPARK_PYTHON=/usr/bin/python3
 - export PYSPARK_DRIVER_PYTHON=ipython3
 - export PYSPARK_DRIVER_PYTHON_OPTS="notebook"
9. Télécharger et lancer [NGROK](#).
10. Lancer le service MongoDB avec "sudo service mongodb start"
11. Configurer le Spark cluster
 - a. Naviguez (dans un terminal) dans le dossier /usr/local/spark/spark-2.4.0-bin-hadoop2.7/conf et faites la commande suivante:
 - i. "cp spark-env.sh.template spark-env.sh"
 - b. Éditez le fichier (vim spart-env.sh) et ajoutez ces lignes à la fin:
 - i. SPARK_WORKER_MEMORY=1g
 - SPARK_EXECUTOR_MEMORY=1g
 - SPARK_WORKER_INSTANCES=2
 - SPARK_WORKER_CORES=2
 - SPARK_WORKER_DIR=/tmp/sparkdata

- c. Lancer le master et les slaves
 - i. Changer de répertoire avec “cd ../sbin”
 - ii. Exécutez le script ./start-master.sh
 - iii. Ouvrir un navigateur et aller sur localhost:8080
 - iv. copier l’url du master et lancez la commande suivante:
 - 1. sudo ./start-slave.sh URL_MASTER
- 12. Dans un terminal, dans le dossier python_service, lancer la commande suivante pour démarrer le service REST
 - a. python3 main.py
- 13. Le service peut maintenant être utilisé depuis le navigateur de la machine hôte.