

Installation Docker

Ce guide explique comment installer Docker puis télécharger et démarrer l'image Docker zenika/spark_zepelin_standalone.

Configuration minimale :

- Poste sous Linux, Windows ou Mac Os, 64 bits.
NB : Une utilisation de docker avec Windows ou MacOS est moins performante que qu'une utilisation sous Linux (l'exécutable de Docker sous Windows et MacOS encapsule une VM linux....)
- Au moins 8 Go de RAM

Installation de Docker :

Suivre la procédure <https://docs.docker.com/engine/installation/>

NB : L'installateur Windows / MacOS installe la dernière version de Virtual Box. S'assurer qu'aucune VM Virtual Box est ouverte pendant l'installation. En cas de mise à jour, les VM précédemment créés dans Virtual Box sont conservées.

Démarrage de l'image Docker sous Linux :

- Exécuter la commande suivante :

```
>> docker run -d -p 8080:8080 \
-p 4040:4040 \
-p 8084:8084 \
-v <ZEPPELIN_NOTEBOOK_HOST_PATH>:/opt/spark/notebook \
-e "SPARK_WORKER_CORES=3" \
-e "SPARK_WORKER_MEMORY=800m" \
--name=spark_zepelin_standalone \
zenika/spark_zepelin_standalone:latest
```

Avec <ZEPPELIN_NOTEBOOK_HOST_PATH>, un répertoire non existant qui sera créé par Docker pour persister les données

- Si l'image n'est pas encore présente elle sera automatiquement téléchargée depuis le docker hub (600 mb)
- Dans un navigateur web : localhost:8080 doit afficher Zeppelin
- Si nécessaire, plus de détails sur l'image et les options disponibles ici :
https://hub.docker.com/r/zenika/spark_zepelin_standalone/

Pour arrêter l'image voir la rubrique *Arrêt du container Docker spark_zepelin_standalone* ci-dessous

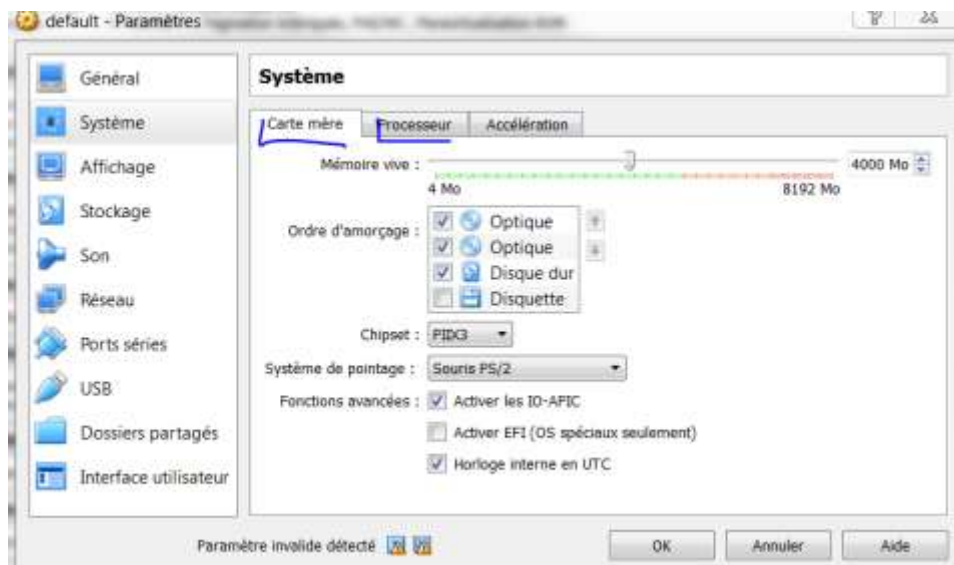
Pour le redémarrer, réexécuter la commande `docker run -d ...` ci-dessus (conserver la même valeur pour `<ZEPPELIN_NOTEBOOK_HOST_PATH>`)

Démarrage de l'image Docker sous Windows / MacOS :

NB : L'installation Windows/MacOS de docker n'apprécie pas tellement les proxy d'entreprise...

Si il apparait des erreurs DNS / No host available à l'exécution de l'image, tenter de d'exécuter l'image offline (après l'avoir téléchargé) ou bien tenter d'effectuer la procédure depuis un réseau personnel.

- Ouvrir Docker Quickstart
 - o La Virtual machine « default » va automatiquement être créée sur VirtualBox (installé avec Docker)
 - o Dans le terminal ouvert exécuter la commande `>>docker-machine stop default` (arrêt de la VM afin de reconfigurer les ressources allouées)
- Ouvrir VirtualBox (installé automatiquement avec docker) :
 - o Une VM « default » doit être présente. Ouvrir ses configurations, section Système
 - o Modifier l'allocation de RAM à 4 Go et le nombre de cœur à 4 (onglet Processeur)



- Dans le terminal Docker quick start :
 - o Redémarrage de la VM : `>>docker-machine start default`
 - o Télécharger et démarrer l'image : `>> docker run -d -p 8080:8080 -p 4040:4040 -p 8084:8084 -v "<ZEPPELIN_NOTEBOOK_HOST_PATH>":/opt/zeppelin/notebook -e "SPARK_WORKER_CORES=3" -e "SPARK_WORKER_MEMORY=800m" --name=spark_zeppelin_standalone zenika/spark_zeppelin_standalone:latest`

Avec `<ZEPPELIN_NOTEBOOK_HOST_PATH>`, un répertoire non existant qui sera créé par Docker pour persister les données.

**NB : Pour Windows utiliser la nomenclature suivante : spécifier
« //c/user/admin/documents/zeppelin» pour le chemin
« C:\user\admin\documents\zeppelin»**

- Si l'image n'est pas encore présente elle sera automatiquement téléchargée depuis le docker hub (600 mb)
 - Taper >> docker-machine ip default pour obtenir l'adresse ip de la VM (ci-dessous nommé <IP-VM>
- Dans un navigateur web : <IP-VM>:8080 doit afficher Zeppelin
 - Si nécessaire, plus de détails sur l'image et les options disponible ici :
https://hub.docker.com/r/zenika/spark_zeppelin_standalone/

Pour arrêter l'image voir la rubrique *Arrêt du container Docker spark_zeppelin_standalone* ci-dessous

Pour le redémarrer, réexécuter la commande *docker run -d ...* ci-dessus dans docker quick start (conserver la même valeur pour <ZEPPELIN_NOTEBOOK_HOST_PATH>)

Arrêt du container Docker spark_zeppelin_standalone:

- Sous Linux, dans un terminal taper
>> docker rm -f spark_zeppelin_standalone
- Sous Windows, MacOS, dans un terminal Docker Quick Start, taper :
>> docker rm -f spark_zeppelin_standalone #Arret du container
>> docker-machine stop default # Arret de la VM

Suppression de l'image Docker spark_zeppelin_standalone:

- Sous Linux, dans un terminal taper
>> docker rmi zenika/spark_zeppelin_standalone:latest Sous Windows, MacOS, dans un terminal Docker Quick Start, taper :
>> docker rmi zenika/spark_zeppelin_standalone:latest >> docker-machine stop default # Arret de la VM
- Si les notebooks ne veulent pas être sauvegardés, supprimer le répertoire <ZEPPELIN_NOTEBOOK_HOST_PATH>.