

---

## *Apache Services for Data Governance installation guide*

---

### Prerequisites

---

- Familiarity with terminals, shell commands and fundamental knowledge in operating Systems (Linux).
- Basic knowledge in containerization and dealing with Docker images and containers.

### Objectives

---

1. Install a hypervisor (**WSL** = **W**indows **S**ubsystem for **L**inux)
2. Install Docker Desktop
3. Access Apache Atlas, Kafka, NiFi, Hive using Docker Containers

### Hardware requirements

---

- **Memory** dedicated to the cluster (Minimum: 4 GiB, **Recommended: 8+ GiB**). More is better.
- **CPU** (Minimum: 4 Cores, **Recommended: 6+ Cores**)
  - Virtualization should be enabled
    - ([Check Virtualization on Windows](#), On Linux: `1scpu`). Sometimes it is disabled in BIOS.
- **Storage**
  - 25-35 GiB.

### Installing Docker Desktop

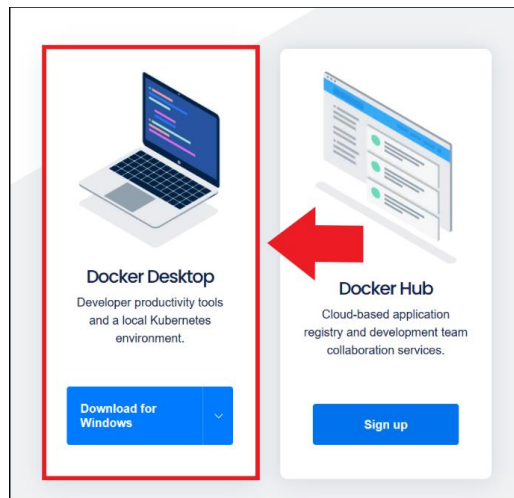
---

L'installation de Docker sur Windows 10/11 se fait en quelques étapes. Pour ce faire, il faut d'abord télécharger le logiciel Docker sur la page d'accueil officielle de Docker. Comme Docker est conçu pour les systèmes Linux, vous avez également besoin

d'un [logiciel de virtualisation](#) comme WSL 2. [WSL 2](#) est un **sous-système Windows gratuit pour Linux** qui vous permet d'utiliser des applications Linux sous Windows.

## Étape 1 : Télécharger Docker

Téléchargez le logiciel Docker gratuit sur le site officiel [Docker Homepage](#). Pour ce faire, sélectionnez l'option **Docker Desktop**.



## Étape 2 : installer Docker

Cliquez maintenant sur le paquet d'installation et lancez l'installation. Sous Windows 10/11, Docker vous recommande d'utiliser une plateforme de virtualisation comme WSL 2 **dès la première étape de l'installation**. Cochez la case « Use WSL 2 instead of Hyper-V » pour installer automatiquement WSL 2. Cliquez ensuite sur « OK » pour démarrer l'installation. Redémarrez le système dès que l'installation est terminée. Après le redémarrage, vous acceptez le contrat de licence. Si vous utilisez Docker à titre privé ou dans une petite entreprise, l'utilisation reste gratuite. Les grandes entreprises doivent déterminer à l'avance dans quelle mesure la licence open source est adaptée aux objectifs souhaités.

## Étape 3 : installer WSL 2

Après avoir accepté le contrat de licence, Docker vous informe que vous devez encore terminer l'installation de WSL-2. Téléchargez WSL 2 **en cliquant sur le lien correspondant dans la fenêtre du menu Docker** et téléchargez la version actuelle. Ouvrez l'assistant d'installation et suivez les étapes d'installation. Une fois l'installation terminée, redémarrez le système pour terminer l'installation de Docker et de WSL 2.

## Étape 4 : Démarrer et tester Docker Desktop

Si Docker Desktop ne démarre pas automatiquement, cliquez sur l'icône du bureau de l'application correspondante. Pour vérifier si Docker fonctionne sans problème, vous pouvez télécharger l'image **Hello World Docker**. Pour ce faire, entrez la commande suivante dans Windows PowerShell :

```
$ docker run hello-world
```

```
Maadoudi@DESKTOP-QEU37T0 MINGW64 /d/kafka_2.12-3.6.0
$ docker run hello-world

Hello from Docker!
This message shows that your installation appears to be working correctly.
```

## Installing Apache Atlas, Hive, Kafka, Nifi, Hadoop

1. Télécharger et décompresser le repo git suivant :

<https://github.com/simmoahmed/DataGovInfrastructure.git>

- Ce cluster Docker contient Apache Atlas (et ses dépendances) ainsi qu'un écosystème Hadoop pour réaliser des expériences de base. Plus précisément, il fournit :
  - Kafka et Zookeeper (dont Atlas dépend également) ;
  - HDFS 2.7 ;
  - Hive 2.3.2 ;
  - Nifi ;
- **Remarque :**
  - Lorsque vous êtes sur une machine **Mac/Linux**, il est recommandé de changer les caractères de fin de ligne (EOL) de **LF** (Line Feed) vers **CRLF** (Carriage Return + Line Feed) pour tous les fichiers de configuration.
  - Si vous êtes sur Windows, assurez-vous de faire l'inverse (CRLF vers LF).

Cette adaptation des fins de ligne est importante pour assurer la compatibilité des fichiers de configuration avec les différentes plateformes et éviter les problèmes de lecture et d'exécution des scripts.

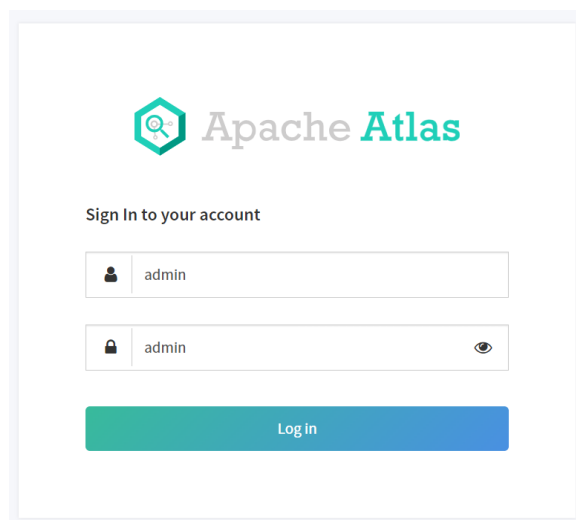
2. Exécutez **docker-compose build** pour construire toutes les images Docker.
3. Exécutez **docker-compose up** pour démarrer tous les services du cluster (cela peut prendre un certain temps).
4. Lorsque la construction est terminée, les 9 conteneurs suivants seront créés, dont vous pourrez vérifier l'état en utilisant la commande **docker ps** dans un onglet de console séparé :

```
Maadoudi@DESKTOP-QEU37T0 MINGW64 ~
$ docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS
c8e51d7bee93   bitnami/kafka:3.3                  "/opt/bitnami/script..." 6 hours ago    Up 6 hour
s              9092/tcp, 0.0.0.0:9093->9093/tcp    docker-atlas-kafka-1
76cd95f2679b   bitnami/zookeeper:3.8              "/opt/bitnami/script..." 6 hours ago    Up 6 hour
s              2888/tcp, 3888/tcp, 0.0.0.0:2181->2181/tcp, 8080/tcp    docker-atlas-zookeeper-1
66b89e1cdf9    apache/nifi:latest                  "../scripts/start.sh"     2 days ago    Up 6 hour
s              8000/tcp, 8080/tcp, 10000/tcp, 0.0.0.0:8443->8443/tcp    docker-atlas-nifi-1
6777be17bb90   bde2020/hive:2.3.2-postgresql-metastore "entrypoint.sh /bin/..." 3 days ago    Up 6 hour
s              0.0.0.0:10000->10000/tcp, 10002/tcp    docker-atlas-hive-server-1
a137aab95299   bde2020/hive:2.3.2-postgresql-metastore "entrypoint.sh /opt/..." 3 days ago    Up 6 hour
s              10000/tcp, 0.0.0.0:9083->9083/tcp, 10002/tcp    docker-atlas-hive-metastore-1
191227d3d5e2   wbaa/rokku-dev-apache-atlas        "sh -c /entrypoint.sh"    3 days ago    Up 2 hour
s              0.0.0.0:21000->21000/tcp              docker-atlas-atlas-server-1
68a2d9a11d06   bde2020/hive-metastore-postgresql:2.3.0 "/docker-entrypoint..." 3 days ago    Up 6 hour
s              5432/tcp                                docker-atlas-hive-metastore-postgresql-1
48c40c5ca258   bde2020/hadoop-datanode:2.0.0-hadoop2.7.4-java8 "/entrypoint.sh /run..." 3 days ago    Up 6 hour
s (healthy)    0.0.0.0:50075->50075/tcp              docker-atlas-datanode-1
13869204c44b   bde2020/hadoop-namenode:2.0.0-hadoop2.7.4-java8 "/entrypoint.sh /run..." 3 days ago    Up 6 hour
s (healthy)    0.0.0.0:50070->50070/tcp              docker-atlas-namenode-1
Maadoudi@DESKTOP-QEU37T0 MINGW64 ~
```

5. Atlas peut ne prendre qu'une ou deux minutes pour être prêt. Vous pouvez vérifier l'état en visitant <https://localhost:21000> ou en exécutant la commande suivante :

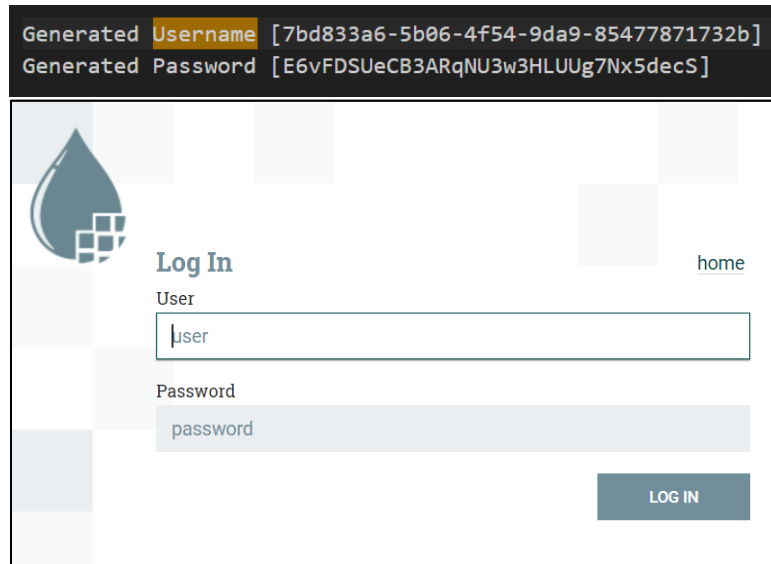
curl -u admin:admin <http://localhost:21000/api/atlas/admin/version>

- **Username** : admin
- **Password** : admin




6. Apache Nifi est aussi prêt en quelques minutes seulement. Vous pouvez vérifier son état en visitant <https://localhost:8443/nifi/>

*Remarque* : Le nom d'utilisateur et le mot de passe par défaut pour accéder à l'interface utilisateur d'Apache NiFi ne sont pas toujours **"admin:admin"**. Vous pouvez trouver les informations d'identification générées par NiFi dans le fichier **`nifi-app.log`**, qui contient les journaux de l'application. Recherchez dans ce fichier pour trouver les identifiants corrects.



Generated Username [7bd833a6-5b06-4f54-9da9-85477871732b]  
Generated Password [E6vFDSUeCB3ARqNU3w3HLUg7Nx5decS]

 **Log In** [home](#)

User

Password

*Maintenant, tout est prêt pour le développement.*