

BIGDATA Analytics

TP N°3 – APACHE HIVE PARTIE N°1 INSTALLATION

Objectifs :

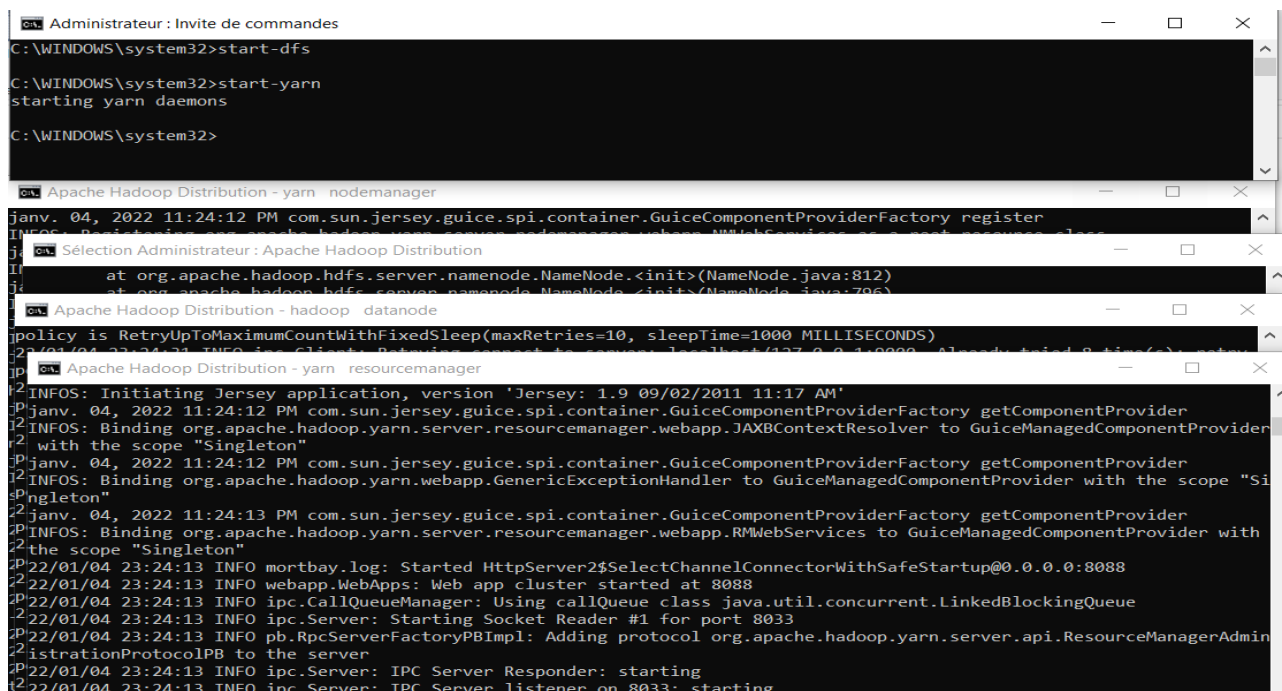
1. Installation Apache Hadoop
2. Démarrages du cluster Hadoop
3. Création des dossiers Hive
4. Configuration D'apache Derby Database
5. Démarrage du moteur Apache Derby Database
6. Configuration D'apache Hive
7. Démarrage D'apache Hive

Outils :

- Database Derby: https://db.apache.org/derby/releases/release-10_14_2_0.cgi
- Apache Hive: <https://archive.apache.org/dist/hive/hive-2.1.0/>

Etape N°1: Démarrage du Cluster Hadoop

Start-dfs & start-yarn



```
Administrateur : Invite de commandes
C:\WINDOWS\system32>start-dfs
C:\WINDOWS\system32>start-yarn
starting yarn daemons
C:\WINDOWS\system32>

Apache Hadoop Distribution - yarn nodemanager
janv. 04, 2022 11:24:12 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.resourcemanager.webapp.HybridServices as a web resource.

Sélection Administrateur : Apache Hadoop Distribution
at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:812)
at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:795)

Apache Hadoop Distribution - hadoop datanode
policy is RetryToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
22/01/04 23:24:13 INFO ipc.Retrying connect to server localhost(127.0.0.1:8033) - Already tried 0 time(s) - retry

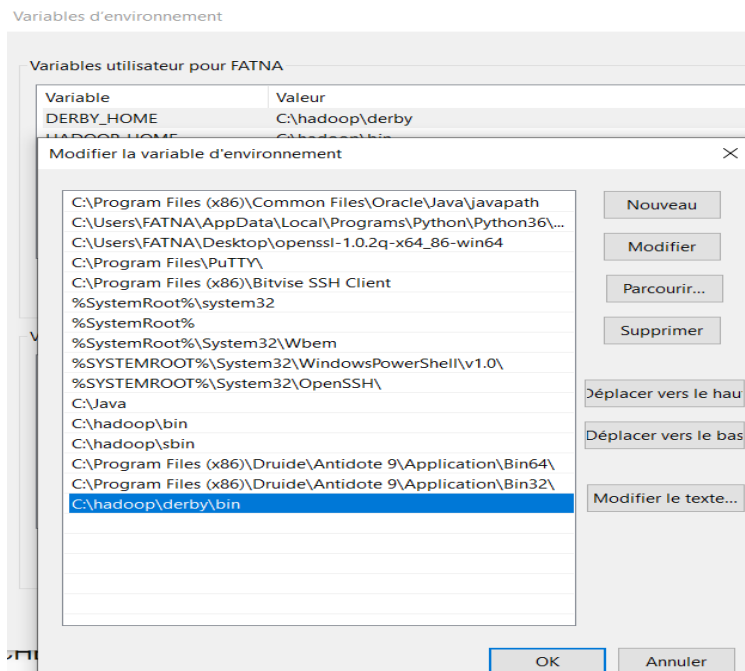
Apache Hadoop Distribution - yarn resourcemanager
INFO: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'
janv. 04, 2022 11:24:12 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
janv. 04, 2022 11:24:12 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
janv. 04, 2022 11:24:13 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
22/01/04 23:24:13 INFO mortbay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:8088
22/01/04 23:24:13 INFO webapp.WebApps: Web app cluster started at 8088
22/01/04 23:24:13 INFO ipc.CallQueueManager: Using callQueue class java.util.concurrent.LinkedBlockingQueue
22/01/04 23:24:13 INFO ipc.Server: Starting Socket Reader #1 for port 8033
22/01/04 23:24:13 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceManagerAdministrationProtocolPB to the server
22/01/04 23:24:13 INFO ipc.Server: IPC Server Responder: starting
22/01/04 23:24:13 INFO ipc.Server: IPC Server listener on 8033: starting
```

Etape N°2 : Création des dossiers Hive

```
C:\WINDOWS\system32>hadoop fs -mkdir -p /user/hive
C:\WINDOWS\system32>hadoop fs -chmod 777 /user/hive
C:\WINDOWS\system32>hadoop fs -mkdir -p /user/hive/warehouse
C:\WINDOWS\system32>hadoop fs -chmod 777 /user/hive/warehouse
C:\WINDOWS\system32>hadoop fs -mkdir -p /tmp
C:\WINDOWS\system32>hadoop fs -chmod 777 /tmp
C:\WINDOWS\system32>hadoop fs -mkdir -p /tmp/hive
C:\WINDOWS\system32>hadoop fs -chmod 777 /tmp/hive
C:\WINDOWS\system32>
```

Etape N°3 : Configuration D'apache Derby Database

- Copier le dossier derby dans le dossier C:\hadoop\derby
- Modifier les variables d'environnements



Etape N°4 : Démarrage du moteur Apache Derby Database

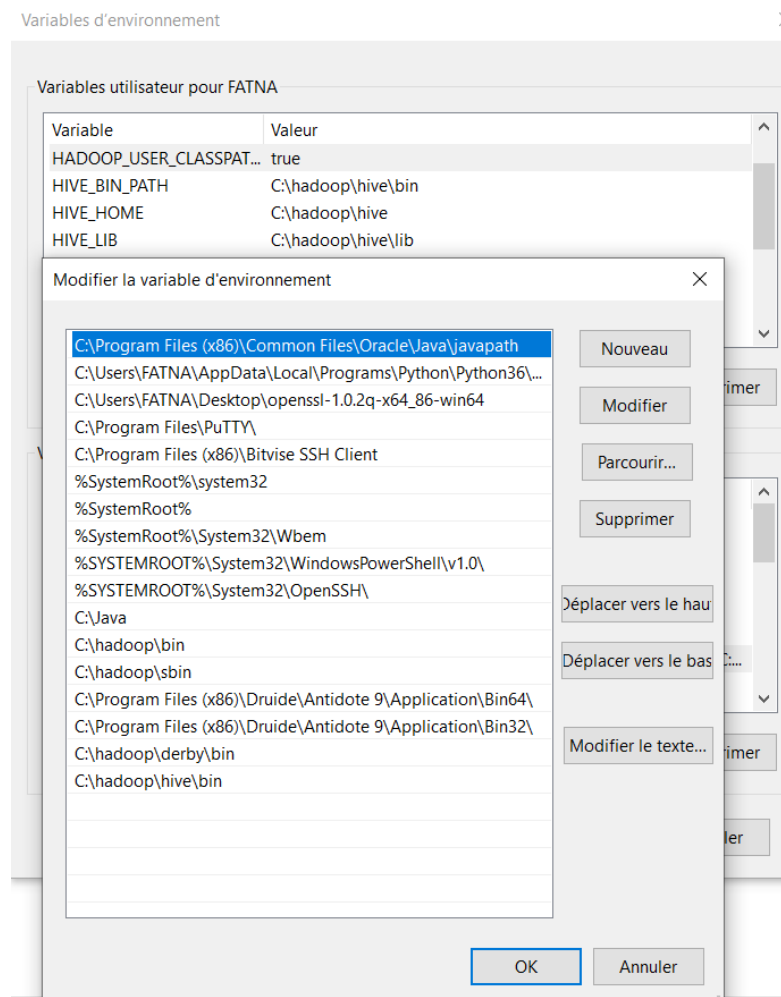
```

C:\Administrateur : Invite de commandes - startNetworkServer -h 0.0.0.0
Microsoft Windows [version 10.0.19042.1415]
(c) Microsoft Corporation. Tous droits réservés.

C:\WINDOWS\system32>startNetworkServer -h 0.0.0.0
Wed Jan 05 00:38:10 WEST 2022 : Le gestionnaire de sécurité a été installé à l'aide de la stratégie de sécurité de serveur de base.
Wed Jan 05 00:38:16 WEST 2022 : Apache Derby Network Server - 10.14.2.0 - (1828579) démarré et prêt à accepter les connexions sur le port 1527
  
```

Etape N°5 : Configuration D'apache Hive

- Copier le dossier Hive dans C:\hadoop\Hive
- Configurer les variables d'environnement



Configurer les deux fichiers de démarrage : hive-site.xml & hive-env.cmd

Fichier hive-site.xml

```
<configuration>
<property>
  <name>javax.jdo.option.ConnectionURL</name>
  <value>jdbc:derby://localhost:1527/hive.metastore_db;create=true </value>
  <description>JDBC connect string for a JDBC metastore</description>
</property>
<property>
  <name>hive.metastore.warehouse.dir</name>
  <value>/user/hive/warehouse</value>
  <description>location of default database for the warehouse</description>
</property>
<property>
  <name>javax.jdo.option.ConnectionDriverName</name>
  <value>org.apache.derby.jdbc.ClientDriver</value>
  <description>JDBC Driver classe name for the datastore which contains
metadata</description>
</property>
<property>
  <name>datanucleus.autoCreateTables</name>
  <value>true</value>
</property>
</configuration>
```

Fichier hive-env.cmd





```
set HADOOP_HOME=C:\HADOOP
```

- Ajouter des fichiers log4j & SLF4J dans c:\hadoop\hive\lib

LOG4J
Le logging consiste à ajouter des traitements dans les applications pour permettre l'émission et le stockage de messages suite à des événements. Par exemple conserver une trace des exceptions qui sont levées dans l'application et des différents événements anormaux ou normaux liés à l'exécution de l'application.

SLF4J
(Simple Logging Facade For Java) est une API d'abstraction de frameworks des logs.

I

-  log4j-1.2.17
-  slf4j-api-1.7.21.jar
-  slf4j-log4j12-1.7.21.jar
-  log4j-slf4j-impl-2.7

- Ajouter les fichiers .jar du dossier derby au dossier c:\hadoop\hive\lib

PC > Windows (C:) > hadoop > derby > lib

Nom

-  derby
-  derby.war
-  derbyclient
-  derbyLocale_cs
-  derbyLocale_de_DE
-  derbyLocale_es
-  derbyLocale_fr
-  derbyLocale_hu
-  derbyLocale_it
-  derbyLocale_ja_JP
-  derbyLocale_ko_KR
-  derbyLocale_pl
-  derbyLocale_pt_BR
-  derbyLocale_ru
-  derbyLocale_zh_CN
-  derbyLocale_zh_TW
-  derbynet
-  derbyoptionaltools
-  derbyrun
-  derbytools

Etape N°6 : Démarrage D'apache Hive

- Démarrer Apache hive par la commande hive,
- Créer une base de données test dans apache hive

```
C:\WINDOWS\system32>hive
ERROR StatusLogger No log4j2 configuration file found. Using default configuration: logging only errors to the console.
Connecting to jdbc:hive2://
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/log4j-slf4j-impl-2.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/slf4j-log4j12-1.7.21.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connected to: Apache Hive (version 2.1.0)
Driver: Hive JDBC (version 2.1.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 2.1.0 by Apache Hive
hive> set hive.cli.print.current.db=true;
No rows affected (0,01 seconds)
hive (default)> create database test;
OK
No rows affected (1,159 seconds)
hive (default)> use test;
OK
No rows affected (0,038 seconds)
hive (test)>
```

- Visualiser la base de données dans l'interface de ressource manager de hadoop

Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities

Browse Directory

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	FATNA	supergroup	0 B	05/01/2022, 01:37:28	0	0 B	test.db

Hadoop, 2016.

- Créer une autre base de données « crmf »

```
hive (test)> create database crmf
...> with dbproperties('creator'='elmendili','date'='09/12/2019');
OK
No rows affected (0,102 seconds)
```

```
hive (test)> use crmf;
OK
No rows affected (0,054 seconds)
hive (crmf)> describe database crmf;
OK
crmf  hdfs://localhost:9000/user/hive/warehouse/crmf.db FATNA USER
1 row selected (0,633 seconds)
hive (crmf)>
```

```
hive (crmf)> describe database extended crmf;
OK
crmf  hdfs://localhost:9000/user/hive/warehouse/crmf.db FATNA USER {date=09/12/2019, creator=elmendili}
1 row selected (0,053 seconds)
hive (crmf)>
```

- Visualiser la base de données dans Hadoop

Hadoop							
Overview	Datanodes	Snapshot	Startup Progress	Utilities			

Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	FATNA	supergroup	0 B	05/01/2022, 19:43:25	0	0 B	crmf.db
drwxrwxrwx	FATNA	supergroup	0 B	05/01/2022, 01:37:28	0	0 B	test.db