BIGDATA Analytics

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

Objectifs :

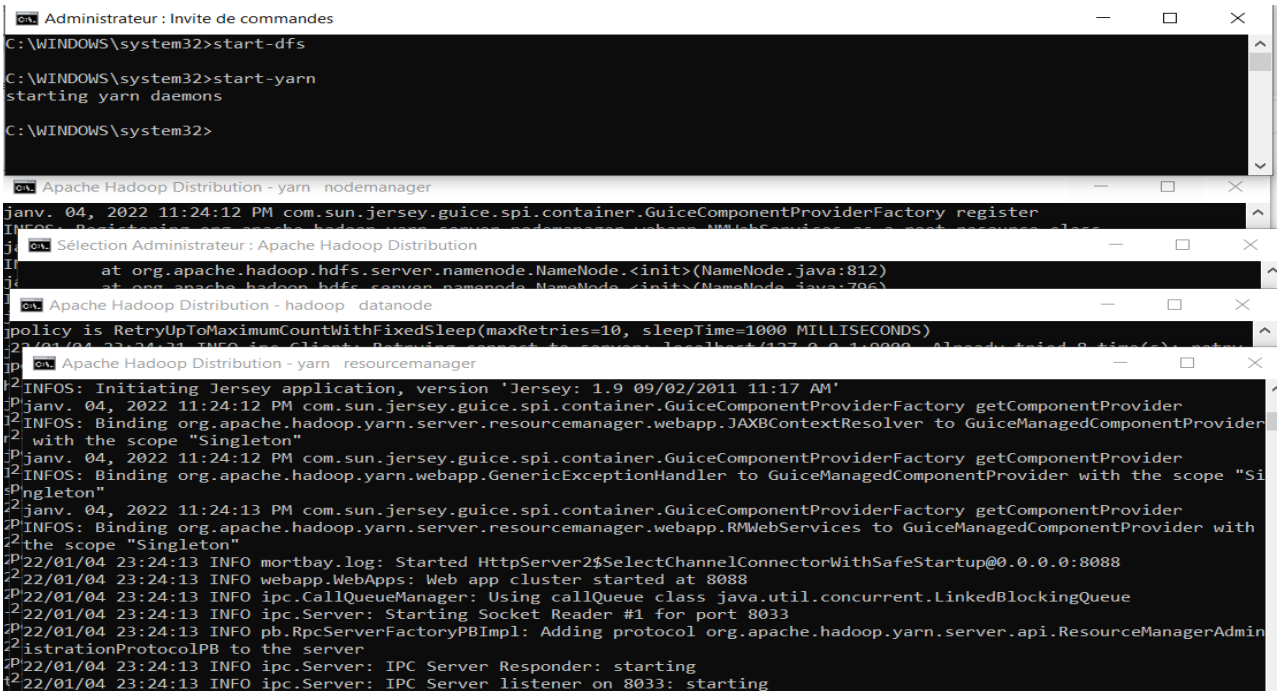
1. Installation Apache Hadoop
2. Démarrages du cluster Hadoop
3. Création des dossiers Hive
4. Configuration D'apache Derby Database
5. Démarrage du moteur Apache Derby Database
6. Configuration D'apache Hive
7. Démarrage D'apache Hive

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

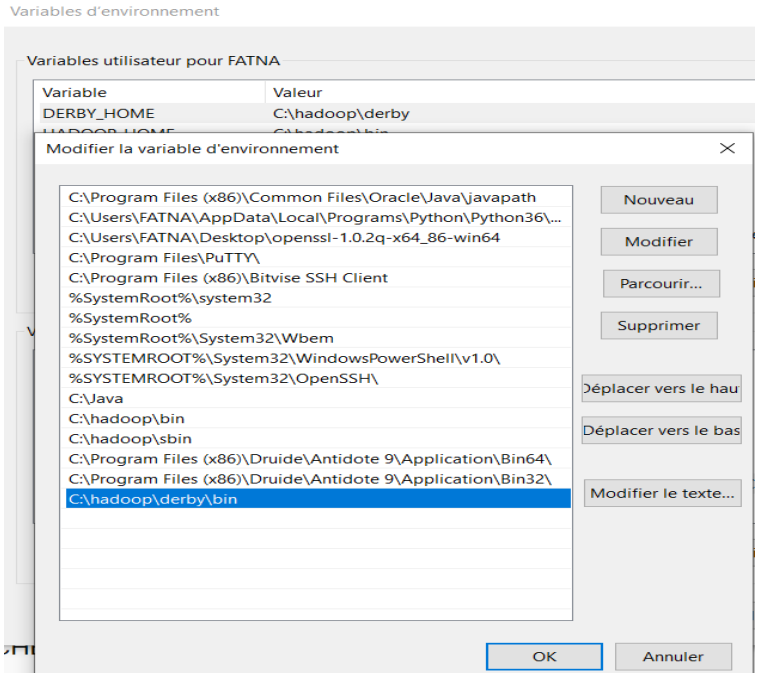


Etape N°2 : Création des dossiers Hive

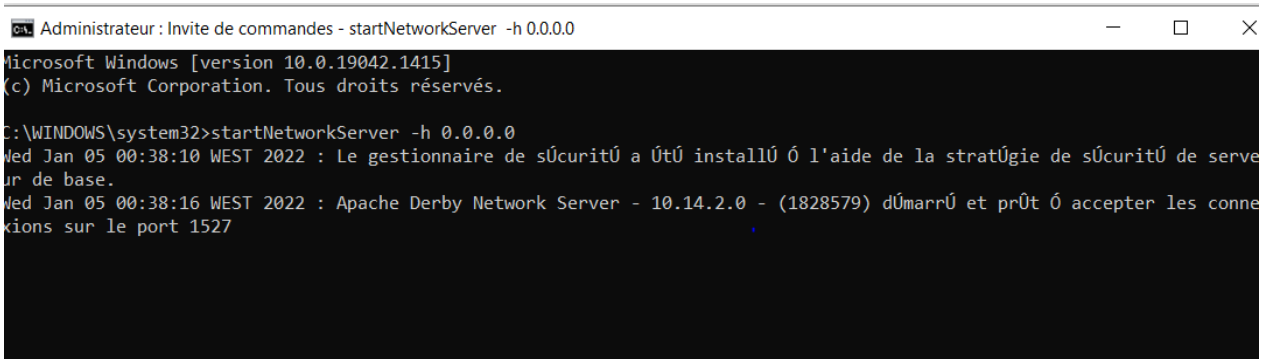
```
C:\WINDOWS\system32>hadoop fs -mkdir -p /user/hive

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

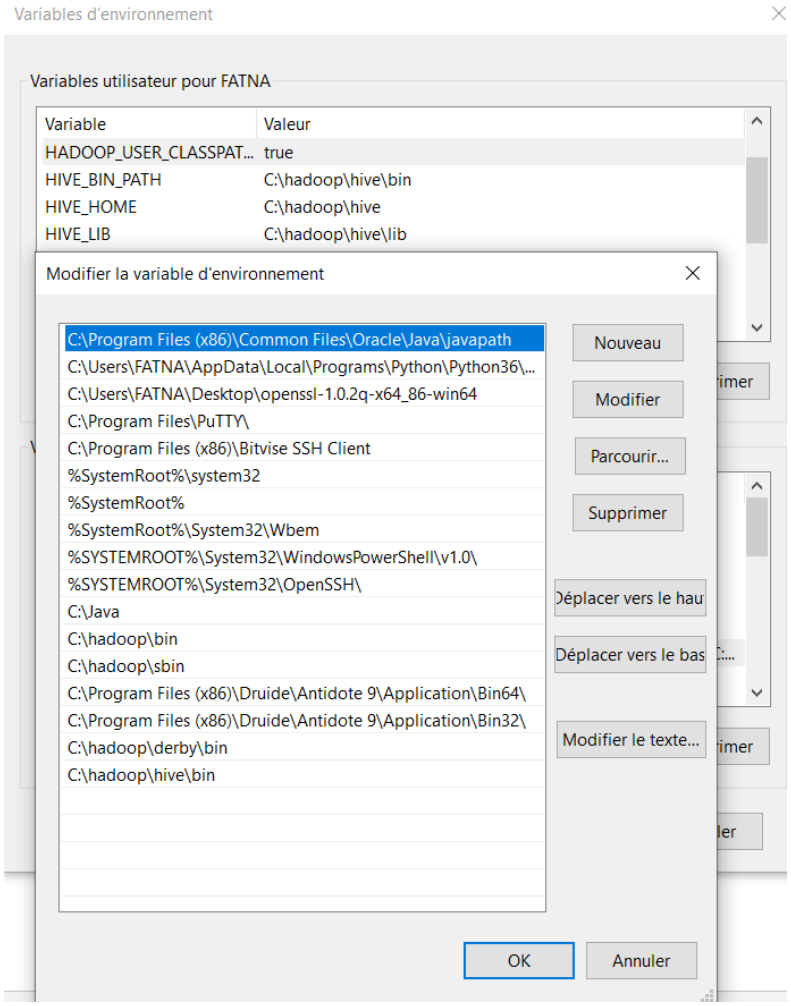


Etape N°4 : Démarrage du moteur Apache Derby Database



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.

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

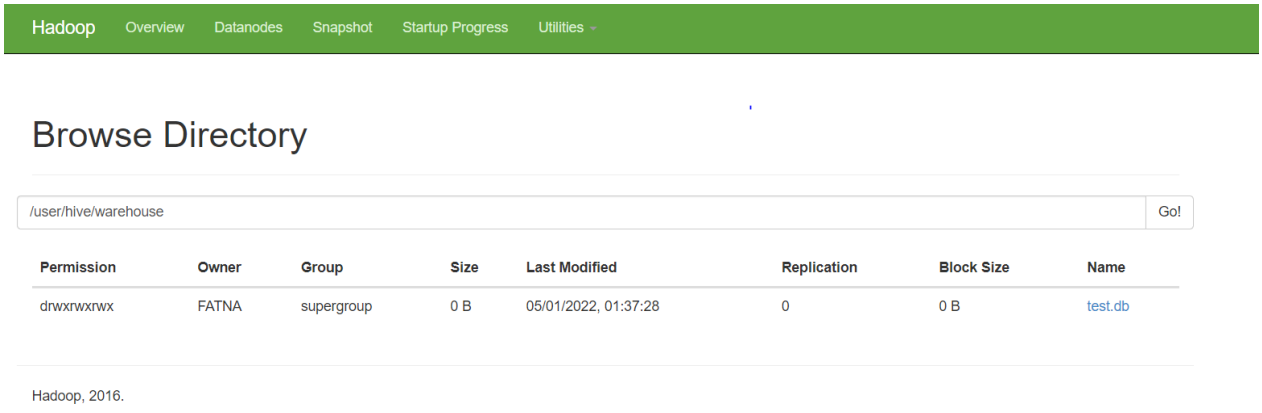
-

Etape N°6 : Démarrage D'apache Hive

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

```
:\WINDOWS\system32>hive
ERROR StatusLogger No log4j2 configuration file found. Using default configuration: logging only errors to the console.
onnecting to jdbc:hive2://
LF4J: Class path contains multiple SLF4J bindings.
LF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.clas
s]
LF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/log4j-slf4j-impl-2.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]

LF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/slf4j-log4j12-1.7.21.jar!/org/slf4j/impl/StaticLoggerBinder.class]

LF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogg
erBinder.class]
LF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
LF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
onnected to: Apache Hive (version 2.1.0)
river: Hive JDBC (version 2.1.0)
ransaction isolation: TRANSACTION_REPEATABLE_READ
eeline version 2.1.0 by Apache Hive
ive> set hive.cli.print.current.db=true;
o rows affected (0,01 seconds)
ive (default)> create database test;
K
o rows affected (1,159 seconds)
ive (default)> use test;
K
o rows affected (0,038 seconds)
ive (test)>
```

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



- 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

/user/hive/warehouse Go!

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