

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

Compte rendu du TP1

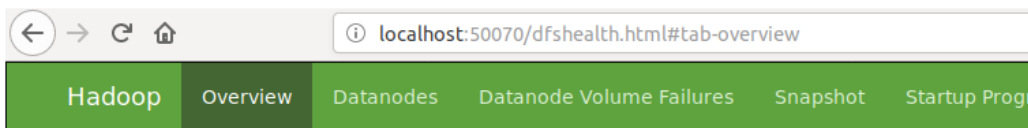
Manipulation du système de fichiers HDFS avec JAVA

- 1- Démarrez les processus Hadoop avec les commandes start-dfs.sh et start-yarn.sh. Ensuite, vérifiez qu'il est en cours d'exécution avec la commande jps et accédez à l'interface web de NameNode avec <http://localhost:50070>.

```
root@ubuntu:~# start-dfs.sh
22/02/23 19:39:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for optimization: java.io.IOException: /usr/lib/hadoop/libhadoop.so.1: cannot open shared object file: No such file or directory
Starting namenodes on [localhost]
```

```
root@ubuntu:~# start-yarn.sh
starting yarn daemons
```

```
root@ubuntu:~# jps
5029 NodeManager
4711 SecondaryNameNode
4471 DataNode
4872 ResourceManager
4313 NameNode
19338 Jps
```



Overview 'localhost:9000' (active)

Started:	Wed Feb 23 05:11:51 PST 2022
Version:	2.7.1, r15ecc87ccf4a0228f35af08fc56de536e6ce657a
Compiled:	2015-06-29T06:04Z by jenkins from (detached from 15ecc87)
Cluster ID:	CID-a4e23064-0d9e-4247-bbca-d4c8df3cd087
Block Pool ID:	BP-1298528298-127.0.1.1-1645232607145

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

Pour interagir avec le serveur HDFS on va avoir besoin d'installer la dépendance de hadoop-client.

```
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-client</artifactId>
  <version>2.7.1</version>
</dependency>
```

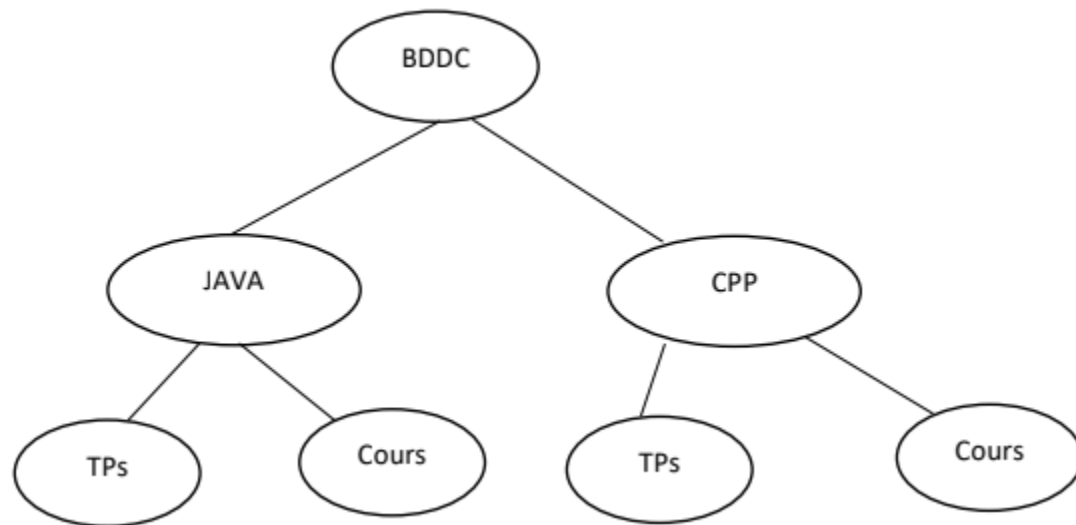
On va créer une configuration, et définir la propriété «fs.defaultF » avec l'adresse de connexion au serveur hdfs.

```
System.out.println("==== TP =====");
Configuration config = new Configuration();
System.setProperty("HADOOP_USER_NAME", "root");
//set server connection address & protocol
config.set("fs.defaultFS", "hdfs://localhost:9000");

FileSystem fileSys = null;
Path pathFile = null;
```

2- Créez dans la racine du HDFS l'arborescence suivante :

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage



```

fileSys = FileSystem.get(config);
System.out.println("===== Directories Creation =====");

```

```

//create directories
createDirectoryIfNotExist(fileSys, new Path("/BDCC"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/JAVA"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/ CPP"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/JAVA/TPs"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/JAVA/Cours"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/ CPP/TPs"));
createDirectoryIfNotExist(fileSys, new Path("/BDCC/ CPP/Cours"));
System.out.println("Directories have been created!");

```

```

private static void createDirectoryIfNotExist(FileSystem fileSystem, Path path) {
    try {
        if(!fileSystem.exists(path)) {
            fileSystem.mkdirs(path);
            System.out.println(path.toString()+" has been created successfully.");
        }
        System.out.println(path.toString()+ " already exists.");
    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
}

```

- 3- Créez dans le répertoire Cours de CPP les fichiers CoursCPP1, CoursCPP2 et CoursCPP3. Puis ajoutez du contenu dans les fichiers créés.

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```

System.out.println("===== Files Creation =====");
for(int i=1; i<=3; i++) {
    pathFile = new Path("/BDCC/_CPP/Cours/CoursCPP"+i);
    createFileIfNotExist(fileSys, pathFile);
    System.out.println(pathFile.toString()+ " has been created!");
    //append content to file
    writeFile(fileSys, pathFile, "This is "+pathFile.toString()+"\n");
}

```

```

private static void createFileIfNotExist(FileSystem fileSystem, Path path) {
    try {
        if(!fileSystem.exists(path)) {
            fileSystem.createNewFile(path);
            System.out.println(path.toString()+" has been created successfully.");
        }
        System.out.println(path.toString()+ " already exists.");
    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
}

```

Ajout de contenu

```

private static void writeFile(FileSystem fileSys, Path pathFile, String content) {
    FSDataOutputStream fsDataOutputStream = null;
    BufferedWriter bufferedWriter = null;
    try {
        fsDataOutputStream = fileSys.create(pathFile);
        bufferedWriter = new BufferedWriter(
            new OutputStreamWriter(fsDataOutputStream, StandardCharsets.UTF_8));
        bufferedWriter.write("Hello this is "+pathFile.toString());
    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
    finally {
        try {
            bufferedWriter.close();
            fsDataOutputStream.close();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}

```

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

4- Affichez le contenu des fichiers CoursCPP1, CoursCPP2 et CoursCPP3.

```
private static String readFromFile(FileSystem fileSys, Path pathFile) {
    String content = "";
    //TODO: read from file and append to content
    FSDataInputStream fsDataInputStream = null;
    BufferedReader bufferedReader = null;

    try {
        fsDataInputStream = fileSys.open(pathFile);
        bufferedReader = new BufferedReader(
            new InputStreamReader(fsDataInputStream, StandardCharsets.UTF_8));
        String contentLine = bufferedReader.readLine();
        while(contentLine != null) {
            content += contentLine + "\n";
            contentLine = bufferedReader.readLine();
        }

    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
    finally {
        try {
            bufferedReader.close();
            fsDataInputStream.close();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}

System.out.println("===== Files Content =====");
for(int i=1; i<=3; i++) {
    pathFile = new Path("/BDCC/_CPP/Cours/CoursCPP"+i);

    System.out.println(pathFile.toString()+" Content:" +readFromFile(fileSys, pathFile));
}
```

5- Copiez les fichiers CoursCPP1, CoursCPP2 et CoursCPP3 dans le répertoire Cours de l'JAVA.

```
for(int i=1; i<=3; i++) {
    FileUtil.copy(fileSys, new Path("/BDCC/_CPP/Cours/CoursCPP"+i),
        fileSys, new Path("/BDCC/JAVA/Cours"), false, config);
}
```

6- Supprimer le fichier CoursCPP3 dans le répertoire Cours de l'JAVA, et renommer CoursCPP1 et CoursCPP2 par CoursJAVA1 et CoursJAVA2 respectivement.

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```
System.out.println("===== File deleting =====");
fileSys.delete(new Path("/BDCC/JAVA/Cours/CoursCPP3"), false);
System.out.println("file has been deleted !");
```

```
System.out.println("===== Files renaming =====");

fileSys.rename(new Path("/BDCC/JAVA/Cours/CoursCPP1"),
               new Path("/BDCC/JAVA/Cours/CoursJAVA1"));
fileSys.rename(new Path("/BDCC/JAVA/Cours/CoursCPP2"),
               new Path("/BDCC/JAVA/Cours/CoursJAVA2"));
System.out.println("Files have been renamed !");
```

- 7- Créer un répertoire dans le système de fichier local les fichiers
:TP1CPP,TP2CPP,TP1JAVA,TP2JAVA,TP3JAVA.

```
System.out.println("===== Local directory creation =====");

File tempDir = new File("tempDir");
tempDir.mkdir();
System.out.println("Directory has been created ");

System.out.println("===== Local files creation =====");
File file = null;

for (int i = 1; i <= 2; i++) {
    file = new File(tempDir.getName()+"/TP"+i+"JAVA");
    file.createNewFile();
    System.out.println("File has been created ");

    file = new File(tempDir.getName()+"/TP"+i+"CPP");
    file.createNewFile();
    System.out.println("File has been created ");
}
file = new File(tempDir.getName()+"/TP"+3+"JAVA");
file.createNewFile();
System.out.println("File has been created ");
```

- 8- Copier les fichiers TP1CPP,TP2CPP à partir du système de fichier local vers le répertoire TPs de
répertoire CPP.

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```

System.out.println("===== Local files coping to HDFS =====");

for (int i = 1; i <= 2; i++) {
    fileSys.copyFromLocalFile(new Path(tempDir.getName()+"/TP"+i+"CPP"),
        new Path("/BDCC/TPs/TP"+i+"CPP"));
    System.out.println("File has been copied ");
}

```

- 9- Copier les fichiers TP1CJAVA, TP2JAVA à partir du système de fichier local vers le répertoire TPs de répertoire JAVA.

```

System.out.println("===== Local files coping to HDFS =====");

for (int i = 1; i <= 3; i++) {
    fileSys.copyFromLocalFile(new Path(tempDir.getName()+"/TP"+i+"JAVA"),
        new Path("/BDCC/JAVA/TPs/TP"+i+"JAVA"));
    System.out.println("File has been copied ");
}

```

- 10- Afficher tout le contenu du répertoire BDCC d'une manière récursive.

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```

private static String getPathChildren(FileSystem fileSys, Path path) {
    String content = "";
    FileStatus[] fileStatusList = null;
    try {
        if(!fileSys.isDirectory(path)){
            content += path.toString()+"\n";
            System.out.println(path.toString());
        }
        else {
            fileStatusList = fileSys.listStatus(path);
            for(FileStatus fileStatus: fileStatusList) {
                getPathChildren(fileSys, fileStatus.getPath());
            }
        }
    } catch (FileNotFoundException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }

    return content;
}

```

```

System.out.println("===== Showing /BDCC content =====");
Path mainPath = new Path("/BDCC");
System.out.println(getPathChildren(fileSys, mainPath));

```

11- Supprimer le fichier TP1CPP du répertoire TPs.

```

System.out.println("===== File deleting =====");
fileSys.delete(new Path("/BDCC/TPs/TP1CPP"), false);
System.out.println("File has been deleted! ");

```

12- Supprimer le répertoire Java avec son contenu.

```

System.out.println("===== Directory deleting =====");
fileSys.delete(new Path("/BDCC/JAVA"), true);
System.out.println("Directory has been deleted! ");

```


ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

Résultats de l'exécution du programme (jusqu'à question 10)

```

===== TP =====

log4j:WARN No appenders could be found for logger
(org.apache.hadoop.util.Shell).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more
info.

===== Directories Creation =====

/BDCC has been created successfully.
/BDCC already exists.
/BDCC/JAVA has been created successfully.
/BDCC/JAVA already exists.
/BDCC/CPP has been created successfully.
/BDCC/CPP already exists.
/BDCC/JAVA/TPs has been created successfully.
/BDCC/JAVA/TPs already exists.
/BDCC/JAVA/Cours has been created successfully.
/BDCC/JAVA/Cours already exists.
/BDCC/CPP/TPs has been created successfully.
/BDCC/CPP/TPs already exists.
/BDCC/CPP/Cours has been created successfully.
/BDCC/CPP/Cours already exists.
Directories have been created!

===== Files Creation =====

/BDCC/CPP/Cours/CoursCPP1 has been created successfully.
/BDCC/CPP/Cours/CoursCPP1 already exists.
/BDCC/CPP/Cours/CoursCPP1 has been created!
/BDCC/CPP/Cours/CoursCPP2 has been created successfully.
/BDCC/CPP/Cours/CoursCPP2 already exists.
/BDCC/CPP/Cours/CoursCPP2 has been created!

```

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```

/BDCC/ CPP/Cours/CoursCPP3 has been created successfully.
/BDCC/ CPP/Cours/CoursCPP3 already exists.
/BDCC/ CPP/Cours/CoursCPP3 has been created!
===== Files Content =====
/BDCC/ CPP/Cours/CoursCPP1 Content:Hello this is /BDCC/ CPP/Cours/CoursCPP1

/BDCC/ CPP/Cours/CoursCPP2 Content:Hello this is /BDCC/ CPP/Cours/CoursCPP2

/BDCC/ CPP/Cours/CoursCPP3 Content:Hello this is /BDCC/ CPP/Cours/CoursCPP3

===== File deleting =====
file has been deleted !
===== Files renaming =====
Files have been renamed !
===== Local directory creation =====
Directory has been created
===== Local files creation =====
File has been created
File has been created
File has been created
File has been created
File has been created
===== Local files coping to HDFS =====
File has been copied
File has been copied
===== Local files coping to HDFS =====
File has been copied
File has been copied
File has been copied
===== Showing /BDCC content =====
hdfs://localhost:9000/BDCC/ CPP/Cours/CoursCPP1

```

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```
hdfs://localhost:9000/BDCC/PHP/Cours/CoursCPP2
hdfs://localhost:9000/BDCC/PHP/Cours/CoursCPP3
hdfs://localhost:9000/BDCC/PHP/TPs/TP1CPP
hdfs://localhost:9000/BDCC/PHP/TPs/TP2CPP
hdfs://localhost:9000/BDCC/JAVA/Cours/CoursJAVA1
hdfs://localhost:9000/BDCC/JAVA/Cours/CoursJAVA2
hdfs://localhost:9000/BDCC/JAVA/TPs/TP1JAVA
hdfs://localhost:9000/BDCC/JAVA/TPs/TP2JAVA
hdfs://localhost:9000/BDCC/JAVA/TPs/TP3JAVA
```

```
root@ubuntu:/home/essadeq# hdfs dfs -ls -R /
22/03/02 07:03:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
applicable
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/PHP
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/PHP/Cours
-rw-r--r-- 3 root supergroup 39 2022-03-02 06:44 /BDCC/PHP/Cours/CoursCPP1
-rw-r--r-- 3 root supergroup 39 2022-03-02 06:44 /BDCC/PHP/Cours/CoursCPP2
-rw-r--r-- 3 root supergroup 39 2022-03-02 06:44 /BDCC/PHP/Cours/CoursCPP3
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/PHP/TPs
-rw-r--r-- 3 root supergroup 0 2022-03-02 06:44 /BDCC/PHP/TPs/TP1CPP
-rw-r--r-- 3 root supergroup 0 2022-03-02 06:44 /BDCC/PHP/TPs/TP2CPP
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/JAVA
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/JAVA/Cours
-rw-r--r-- 3 root supergroup 39 2022-03-02 06:44 /BDCC/JAVA/Cours/CoursJAVA1
-rw-r--r-- 3 root supergroup 39 2022-03-02 06:44 /BDCC/JAVA/Cours/CoursJAVA2
drwxr-xr-x - root supergroup 0 2022-03-02 06:44 /BDCC/JAVA/TPs
-rw-r--r-- 3 root supergroup 0 2022-03-02 06:44 /BDCC/JAVA/TPs/TP1JAVA
-rw-r--r-- 3 root supergroup 0 2022-03-02 06:44 /BDCC/JAVA/TPs/TP2JAVA
-rw-r--r-- 3 root supergroup 0 2022-03-02 06:44 /BDCC/JAVA/TPs/TP3JAVA
root@ubuntu:/home/essadeq#
```

Après la question 10

```
===== File deleting =====
File has been deleted!
===== Directory deleting =====
Directory has been deleted!
```

ENSET-M	II-BDCC2
ELAAMIRI Essadeq	Big Data: Fondements et Architectures de stockage

```

root@ubuntu:/home/essadeq# hdfs dfs -ls -R /
22/03/02 07:15:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
applicable
drwxr-xr-x   - root supergroup          0 2022-03-02 07:12 /BDCC
drwxr-xr-x   - root supergroup          0 2022-03-02 07:11 /BDCC/CPP
drwxr-xr-x   - root supergroup          0 2022-03-02 07:11 /BDCC/CPP/Cours
-rw-r--r--   3 root supergroup        39 2022-03-02 07:11 /BDCC/CPP/Cours/CoursCPP1
-rw-r--r--   3 root supergroup        39 2022-03-02 07:11 /BDCC/CPP/Cours/CoursCPP2
-rw-r--r--   3 root supergroup        39 2022-03-02 07:11 /BDCC/CPP/Cours/CoursCPP3
drwxr-xr-x   - root supergroup          0 2022-03-02 07:12 /BDCC/CPP/TPs
-rw-r--r--   3 root supergroup          0 2022-03-02 07:12 /BDCC/CPP/TPs/TP2CPP
root@ubuntu:/home/essadeq#

```