

# HDFS JAVA API 실습 (1)

## 1. hadoop 클러스터 구동

- `hdfs namenode -format`  
`cd sbin`  
`start-dfs.cmd`  
`start-yarn.cmd`

\*\* `jps` : `jps` 명령어는 실행중인 JVM(Java virtual Machine) 프로세스 상태를 보여주는 것

```
C:\>jps
16400 NameNode
25056
7284 org.eclipse.equinox.launcher_1.6.400.v20210924-0641.jar
19096 DataNode
21048 NodeManager
17820 BootLanguageServerBootApplication
20700 Jps
26332 ResourceManager
```

## 2. maven 설치

- `apache-maven-3.8.6`
- `maven home` 환경변수 설정

## 3. mvn package 실행

```
PS C:\hdfs-example> mvn package
[INFO] Scanning for projects...
[INFO] -----< com.fastcampus.hadoop:hdfs-example >-----
[INFO] Building hdfs-example 1.0.0
[INFO] -----[ jar ]-----
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\main\resources
[INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hdfs-example ---
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\test\resources
[INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ hdfs-example ---
[INFO] Nothing to compile - all classes are up to date
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ hdfs-example ---
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ hdfs-example ---
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 0.895 s
[INFO] Finished at: 2022-10-12T13:16:10+09:00
[INFO] -----
```

1. `hdfs-example`에 있는 폼파일을 `hdfs`에 업로드하기

```
C:\hdfs-example>hadoop fs -ls
ls: \.: No such file or directory

C:\hdfs-example>hadoop fs -mkdir /user

C:\hdfs-example>hadoop fs -mkdir /user/fastcampus
```

```
C:\hdfs-example>hadoop fs -ls /user/fastcampus

C:\hdfs-example>hadoop fs -put pom.xml /user/fastcampus
```

```
C:\hdfs-example>hadoop fs -ls /user/fastcampus
Found 1 items
-rw-r--r-- 1 Minjin supergroup 889 2022-10-12 13:34 /user/fastcampus/pom.xml
```

## 2. 업로드한 폼파일을 vscode에서 실행해서 출력하기

```
1 package com.fastcampus.hadoop;
2
3 import java.io.IOException;
4 import java.io.InputStream;
5 import java.net.URI;
6
7 import org.apache.hadoop.conf.Configuration;
8 import org.apache.hadoop.fs.FileSystem;
9 import org.apache.hadoop.fs.Path;
10 import org.apache.hadoop.io.IOUtils;
11
12 public class FileSystemPrint {
13     Run | Debug
14     public static void main(String[] args) throws IOException {
15         String uri = args[0]; // args를 통해 어느 위치 파일에 실행할지 입력받는다.
16         Configuration conf = new Configuration();
17         FileSystem fs = FileSystem.get(URI.create(uri), conf); // 입력받은 위치로부터 filesystem을 얻어
18         try (InputStream in = fs.open(new Path(uri))) { // 해당파일을 열고
19             IOUtils.copyBytes(in, System.out, bufferSize: 4096); // 출력한다.
20         }
21     }
22 }
```

```
PS C:\hdfs-example> hadoop jar .\target\hdfs-example-1.0.0.jar com.fastcampus.hadoop.FileSystemPrint /user/fastcampus/pom.xml
<?xml version="1.0" encoding="UTF-8"?>

<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>com.fastcampus.hadoop</groupId>
  <artifactId>hdfs-example</artifactId>
  <version>1.0.0</version>

  <properties>
    <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
    <maven.compiler.source>1.8</maven.compiler.source>
    <maven.compiler.target>1.8</maven.compiler.target>
    <hadoop.version>3.3.2</hadoop.version>
  </properties>

  <dependencies>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
      <artifactId>hadoop-client</artifactId>
      <version>${hadoop.version}</version>
    </dependency>
  </dependencies>
</project>
PS C:\hdfs-example>
```

*hadoop jar 명령어를 이용하여 실행(타겟에 hdfs 자르파일 입력 + 실행할 메인 클래스위치를 지정 + hdfs에 위치하는 args를 넣고 실행) => 하둡파일시스템으로부터 읽어서 출력*

# HDFS JAVA API 실습 (2)

## 디렉토리 목록을 조회

### 1. 빌드

```

1  package com.fastcampus.hadoop;
2
3  import java.io.IOException;
4  import java.net.URI;
5
6  import org.apache.hadoop.conf.Configuration;
7  import org.apache.hadoop.fs.FileStatus;
8  import org.apache.hadoop.fs.FileSystem;
9  import org.apache.hadoop.fs.FileUtil;
10 import org.apache.hadoop.fs.Path;
11
12 public class ListStatus {
13     /*해당 프로그램의 아규먼트를 받아 실행*/
14     Run | Debug
15     public static void main(String[] args) throws IOException {
16         String uri = args[0]; // String으로 args에 입력받은 결과를 저장 = hdfs에 디렉토리 목록을 가져올 위치를 입력받고
17         Configuration conf = new Configuration(); // hadoop conf 파일을 생성
18         FileSystem fs = FileSystem.get(URI.create(uri), conf); // 파일시스템을 uri로부터 얻어옴 = 해당 위치로부터 파일시스템 객체를 얻어옴
19
20         Path path = new Path(uri);
21         FileStatus[] status = fs.listStatus(path); // 파일시스템의 listStatus 메소드를 호출해서 status 배열을 가져옴
22         Path[] listPaths = FileUtil.stat2Paths(status); // FileStatus 배열은 유틸을 이용해 path 배열로 변경한다.
23         for (Path p : listPaths) {
24             System.out.println(p); // path 정보를 출력함
25         }
26     }
27 }

```

```

PS C:\hdfs-example> mvn package
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.fastcampus.hadoop:hdfs-example >-----
[INFO] Building hdfs-example 1.0.0
[INFO] -----[ jar ]-----
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\main\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hdfs-example ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 2 source files to C:\hdfs-example\target\classes
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\test\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ hdfs-example ---
[INFO] Nothing to compile - all classes are up to date
[INFO]
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ hdfs-example ---
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ hdfs-example ---
[INFO] Building jar: C:\hdfs-example\target\hdfs-example-1.0.0.jar
[INFO]
[INFO] BUILD SUCCESS
[INFO]
[INFO] Total time: 1.684 s
[INFO] Finished at: 2022-10-12T14:37:08+09:00
[INFO]

```

### 2. 실행

*hadoop jar 명령어를 이용해 실행 + com.fastcampus.hadoop.ListStatus (실행할 메인함수와 위치를 알려줌) + 인  
풋아규먼트로 /user/fastcampus 위치의 리스트 목록을 받아옴*

```
PS C:\hdfs-example> hadoop jar .\target\hdfs-example-1.0.0.jar com.fastcampus.hadoop.ListStatus /user/fastcampus
hdfs://localhost:9000/user/fastcampus/pom.xml
PS C:\hdfs-example>
```

```
PS C:\hdfs-example> hadoop fs -ls /user/fastcampus
Found 1 items
-rw-r--r-- 1 Minjin supergroup 889 2022-10-12 13:34 /user/fastcampus/pom.xml
PS C:\hdfs-example>
```

## 로컬에 있는 파일을 hdfs에 복사

= CLI로 CopyFromLocal, Put 명령어와 같음

### 1. 빌드

```
1 package com.fastcampus.hadoop;
2
3 import java.io.*;
4 import java.net.URI;
5
6 import org.apache.hadoop.conf.Configuration;
7 import org.apache.hadoop.fs.FileSystem;
8 import org.apache.hadoop.fs.Path;
9 import org.apache.hadoop.io.IOUtils;
10
11 public class CopyFromLocal {
12     Run | Debug
13     public static void main(String[] args) throws IOException {
14         // 해당 프로그램은 두개의 인자를 가진다.
15         String localSrc = args[0]; // 로컬 파일 시스템에 존재하는 파일의 위치를 입력받는 인자
16         String dst = args[1]; // hdfs위치에 어떻게 파일을 생성할 것인지 정보를 입력받는 인자
17
18         InputStream in = new BufferedInputStream(new FileInputStream(localSrc));
19
20         Configuration conf = new Configuration();
21         FileSystem fs = FileSystem.get(URI.create(dst), conf);
22         OutputStream out = fs.create(new Path(dst));
23
24         IOUtils.copyBytes(in, out, bufferSize: 4096, close: true); // 인풋스트림을 아웃풋스트림으로 복사한다.
25     }
26 }
```

```
PS C:\hdfs-example> mvn package
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.fastcampus.hadoop:hdfs-example >-----
[INFO] Building hdfs-example 1.0.0
[INFO] -----[ jar ]-----
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\main\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hdfs-example ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 3 source files to C:\hdfs-example\target\classes
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\test\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ hdfs-example ---
[INFO] Nothing to compile - all classes are up to date
[INFO]
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ hdfs-example ---
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ hdfs-example ---
[INFO] Building jar: C:\hdfs-example\target\hdfs-example-1.0.0.jar
[INFO]
[INFO] BUILD SUCCESS
[INFO]
[INFO] -----
[INFO] Total time: 1.380 s
[INFO] Finished at: 2022-10-12T15:56:10+09:00
[INFO] -----
```

### 2. 실행

### 파일 뭐있는지 확인하고 삭제

```
PS C:\hdfs-example> hadoop fs -ls /user/fastcampus
Found 1 items
-rw-r--r--  1 Minjin supergroup      889 2022-10-12 13:34 /user/fastcampus/pom.xml
PS C:\hdfs-example> hadoop fs -rm /user/fastcampus/pom.xml
Deleted /user/fastcampus/pom.xml
```

*hadoop jar 명령어를 이용 / 실행할 메인 함수가 있는 클래스 지정 / [첫번째 인자 소스를 주고] + [두번째 인자 user/fastcampus/pom.xml (pom.xml을 복사하겠다! 복사할 파일명까지 작성)]*

```
PS C:\hdfs-example> hadoop jar .\target\hdfs-example-1.0.0.jar com.fastcampus.hadoop.
CopyFromLocal ./pom.xml /user/fastcampus/pom.xml
PS C:\hdfs-example> █
```

```
PS C:\hdfs-example> hadoop fs -ls /user/fastcampus
Found 1 items
-rw-r--r--  1 Minjin supergroup      889 2022-10-12 16:08 /user/fastcampus/pom.xml
PS C:\hdfs-example> █
```

### 해당파일이 제대로 복사가 되었는지 확인 cat

```
PS C:\hdfs-example> hadoop fs -cat /user/fastcampus/pom.xml
<?xml version="1.0" encoding="UTF-8"?>

<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/m
aven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>com.fastcampus.hadoop</groupId>
  <artifactId>hdfs-example</artifactId>
  <version>1.0.0</version>

  <properties>
    <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
    <maven.compiler.source>1.8</maven.compiler.source>
    <maven.compiler.target>1.8</maven.compiler.target>
    <hadoop.version>3.3.2</hadoop.version>
  </properties>

  <dependencies>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
      <artifactId>hadoop-client</artifactId>
      <version>${hadoop.version}</version>
    </dependency>
  </dependencies>
</project>
PS C:\hdfs-example> █
```

## 로컬에 있는 파일삭제

### 1. 빌드

```

1  package com.fastcampus.hadoop;
2
3  import java.io.IOException;
4  import java.net.URI;
5
6  import org.apache.hadoop.conf.Configuration;
7  import org.apache.hadoop.fs.FileSystem;
8  import org.apache.hadoop.fs.Path;
9
10 public class DeleteFile {
11     // 파일을 삭제하는 프로그램
12     Run | Debug
13     public static void main(String[] args) throws IOException {
14         String uri = args[0]; // 삭제할 파일의 위치를 인자로 받음
15
16         Configuration conf = new Configuration(); // hadoop conf 생성
17         FileSystem fs = FileSystem.get(URI.create(uri), conf); // 파일시스템을 가져옴
18
19         Path path = new Path(uri);
20         if(fs.exists(path)){ // 파일시스템에 해당 파일이 존재한다면 삭제됨
21             fs.delete(path, recursive: false); // recursive를 true로 지정하면 하위에 있는 것들도 전부 삭제됨.
22         }
23     }
24 }

```

```

PS C:\hdfs-example> mvn package
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.fastcampus.hadoop:hdfs-example >-----
[INFO] Building hdfs-example 1.0.0
[INFO]
[INFO] -----[ jar ]-----
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\main\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ hdfs-example ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 4 source files to C:\hdfs-example\target\classes
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ hdfs-example ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory C:\hdfs-example\src\test\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ hdfs-example ---
[INFO] Nothing to compile - all classes are up to date
[INFO]
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ hdfs-example ---
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ hdfs-example ---
[INFO] Building jar: C:\hdfs-example\target\hdfs-example-1.0.0.jar
[INFO]
[INFO] BUILD SUCCESS
[INFO]
[INFO] Total time: 1.286 s
[INFO] Finished at: 2022-10-12T16:41:16+09:00
[INFO]

```

## 2. 실행

*hadoop jar 명령어를 이용 / 실행할 메인 함수가 있는 클래스 지정 / [첫번째 인자 소스: 삭제할 파일 위치와 파일]*

```

PS C:\hdfs-example> hadoop fs -ls /user/fastcampus
Found 1 items
-rw-r--r-- 1 Minjin supergroup      889 2022-10-12 16:08 /user/fastcampus/pom.xml
PS C:\hdfs-example> hadoop jar .\target\hdfs-example-1.0.0.jar com.fastcampus.hadoop.DeleteFile
/user/fastcampus/pom.xml
PS C:\hdfs-example> hadoop fs -ls /user/fastcampus
PS C:\hdfs-example>

```