

## Pig - UDF

---

### Use Case

edureka!

**edureka!**

© Brain4ce Education Solutions Pvt. Ltd.

# Pig - UDF

## Table of Contents

Pig – UDF .....	2
Problem Statements:.....	2
Important Links: .....	2
Technology/Software Used: .....	2
Project Workflow: .....	3
Environment Creation for Creating Solution:.....	4
Data File Creation in Local.....	4
Writing Pig UDF .....	5
Required JARs Files:.....	5
Upper.java .....	6
Pow.java .....	7
JAR Creation .....	8

edureka!

## Pig – UDF

### Problem Statements:

In this document, it is explained that how to implement pig UDF by creating built-in functions in order to explain the functionality of PIG built-in function. For, better explanation, we have taken two built-in functions. Also we have shown the pig scripting. Examples are listed below:

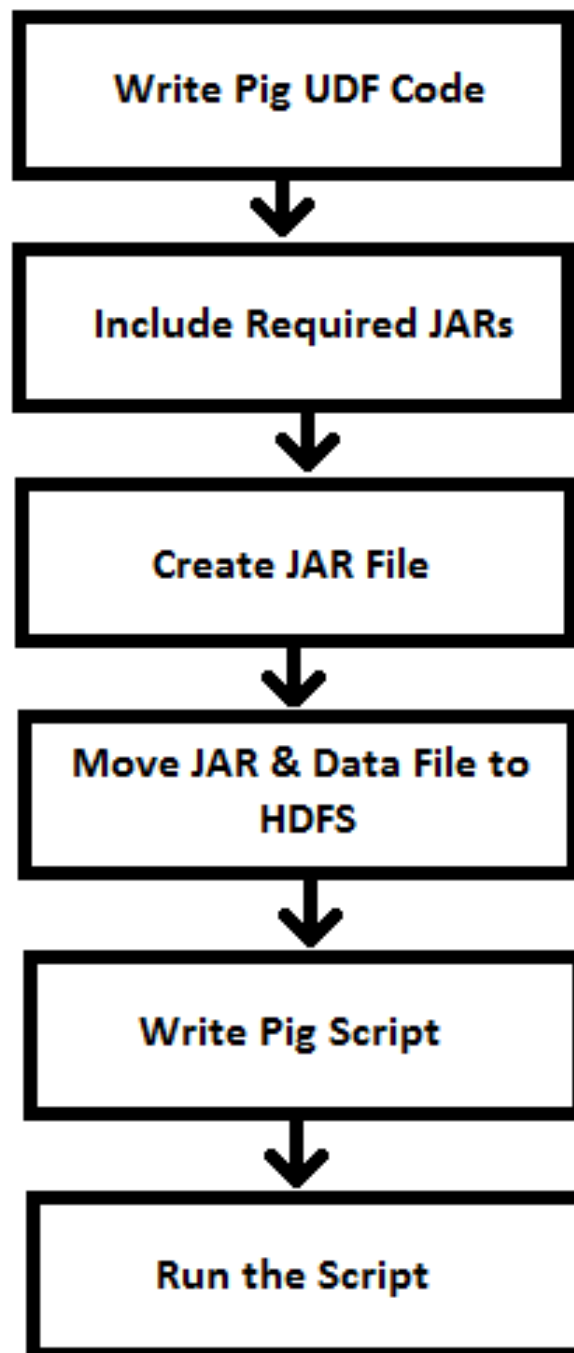
- Converting a particular column in Upper-Case.
- Taking a value and raising its power.

### Important Links:

- **Link for all the codes:**  
[https://edureka.wistia.com/medias/gxmm9psdq3/download?media\\_file\\_id=64150882](https://edureka.wistia.com/medias/gxmm9psdq3/download?media_file_id=64150882)
- **Pig Installation:**  
<https://edureka.wistia.com/medias/lpb6yiupps>

### Technology/Software Used:

- Hadoop Environment
- Apache Pig
- JARs (present under lib folder with the source code)
- FileZilla (if eclipse is in windows)

**Project Workflow:**

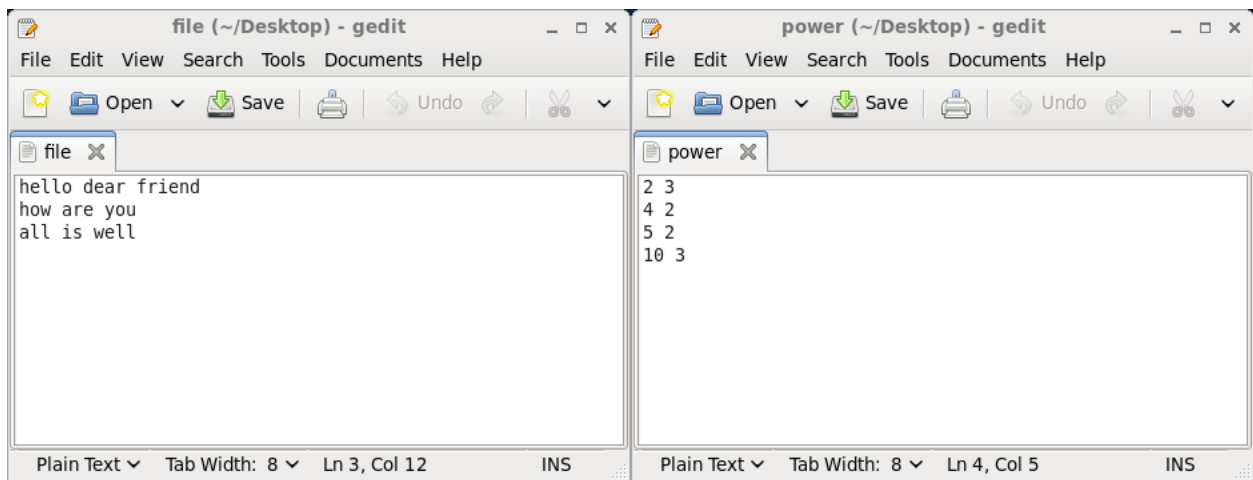
## Environment Creation for Creating Solution:

To run the scripts and to get the output we have done few required steps, which are depicted by the following images:

### Data File Creation in Local

Here, we have created two files in local which are depicted in the following image.

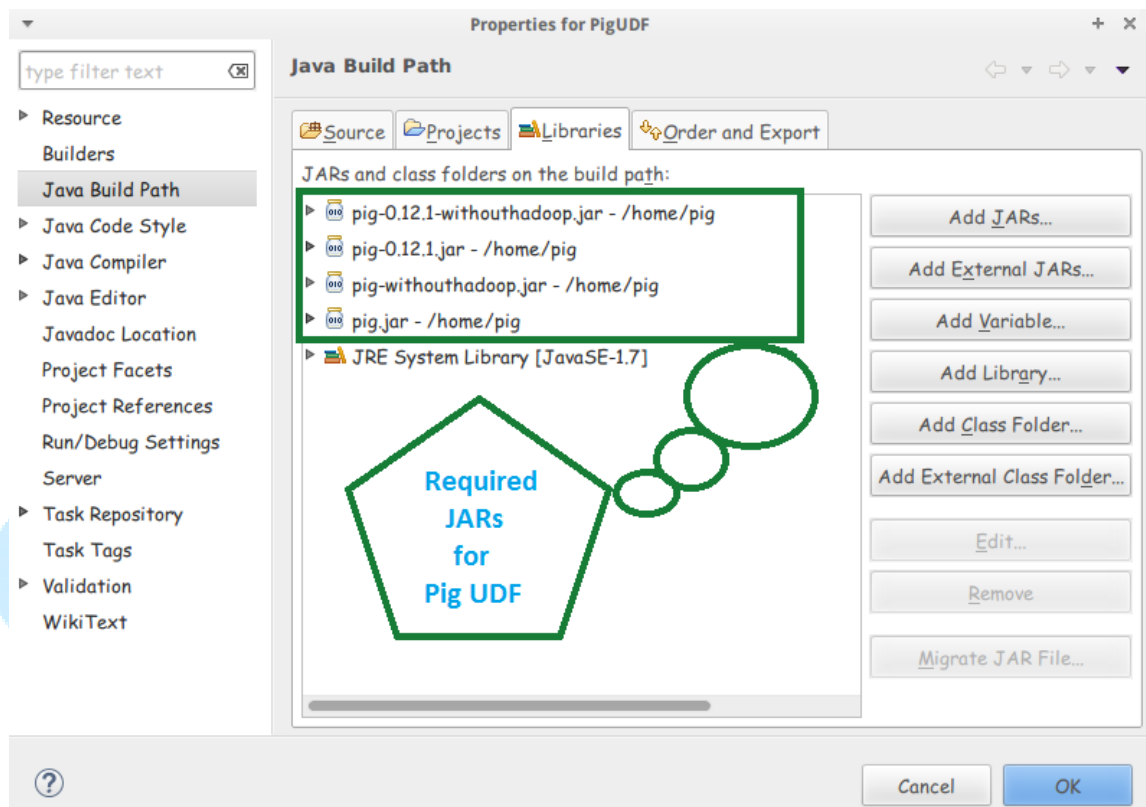
- ✓ File (Converting a particular column in Upper-Case)
- ✓ Power (Taking a value and raising its power)



## Writing Pig UDF

In Eclipse, we have written two java programs for Pig UDF. For this, we have to import few JARs which have shown in following image. You can download all these JARs and codes from the link which is given at page 2.

### Required JARs Files:



## Upper.java

```
import java.io.IOException;
import org.apache.pig.EvalFunc;
import org.apache.pig.data.Tuple;
import org.apache.pig.impl.util.WrappedIOException;

@SuppressWarnings("deprecation")
public class Upper extends EvalFunc<String> {

    public String exec(Tuple input) throws IOException {

        if (input == null || input.size() == 0)

            return null;

        try {

            String str = (String) input.get(0);

            str=str.toUpperCase();

            return str;


        }
        catch (Exception e) {

            throw WrappedIOException.wrap("Caught exception processing input row ", e);

        }

    }

}
```



**Upper.java**

## Pow.java

```
import java.io.IOException;
import org.apache.pig.EvalFunc;
import org.apache.pig.PigWarning;
import org.apache.pig.data.Tuple;

public class Pow extends EvalFunc<Long> {

    public Long exec(Tuple input) throws IOException {
        try {

            int base = (Integer)input.get(0);
            int exponent = (Integer)input.get(1);
            long result = 1;

            /* Probably not the most efficient method...*/
            for (int i = 0; i < exponent; i++) {
                long preresult = result;
                result *= base;
                if (preresult > result) {
                    // We overflowed. Give a warning, but do not throw an
                    // exception.
                    warn("Overflow!", PigWarning.TOO_LARGE_FOR_INT);
                    // Returning null will indicate to Pig that we failed but
                    // we want to continue execution.
                    return null;
                }
            }
            return result;
        } catch (Exception e) {
            // Throwing an exception will cause the task to fail.
            throw new IOException("Something bad happened!", e);
        }
    }
}
```

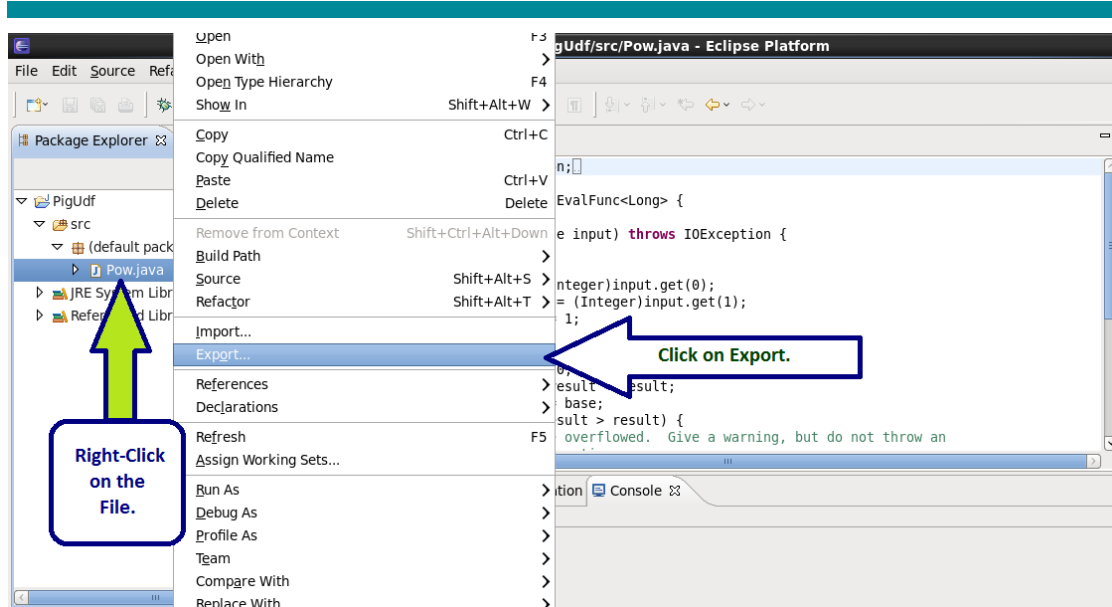


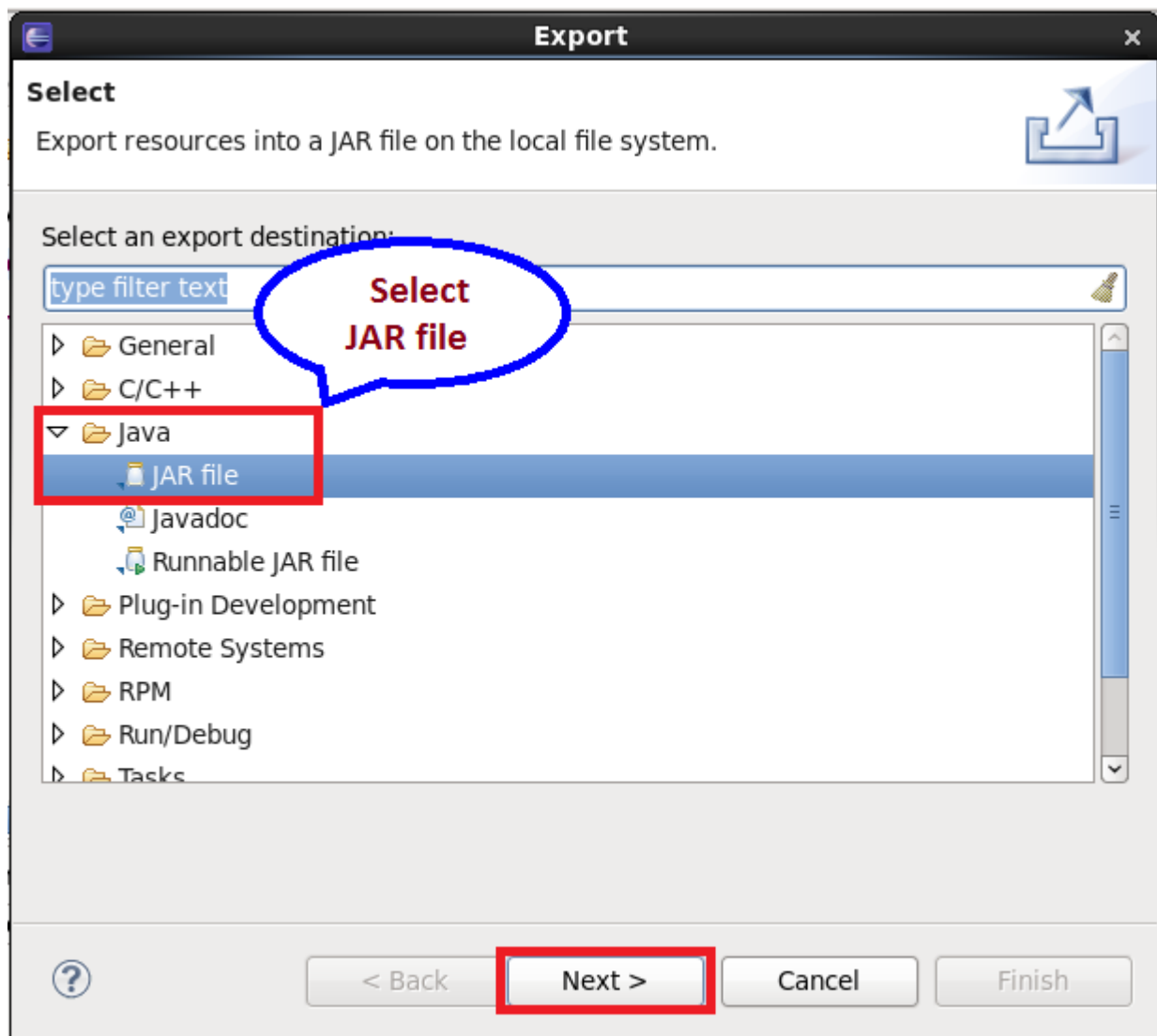
**Pow.java**

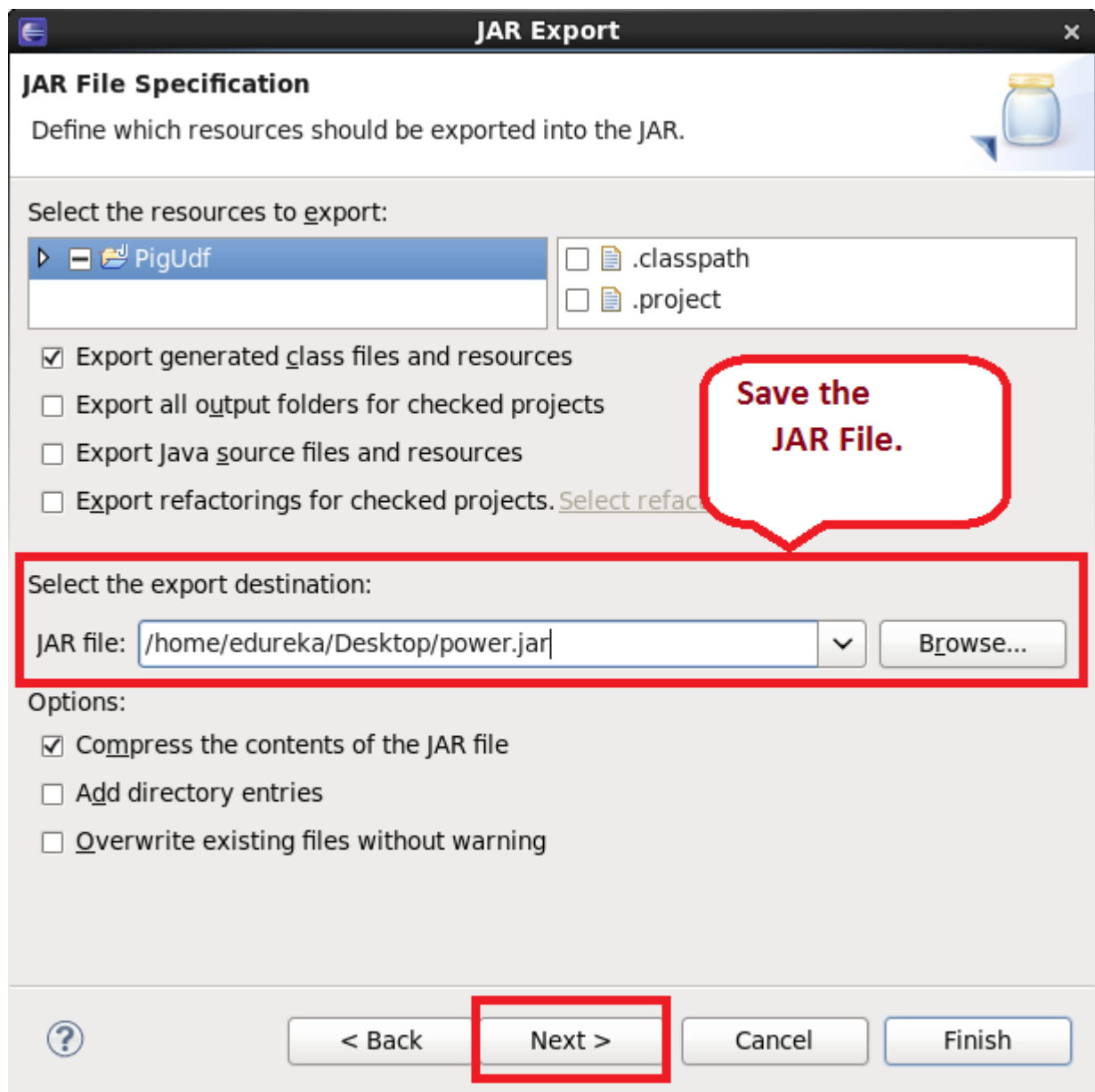


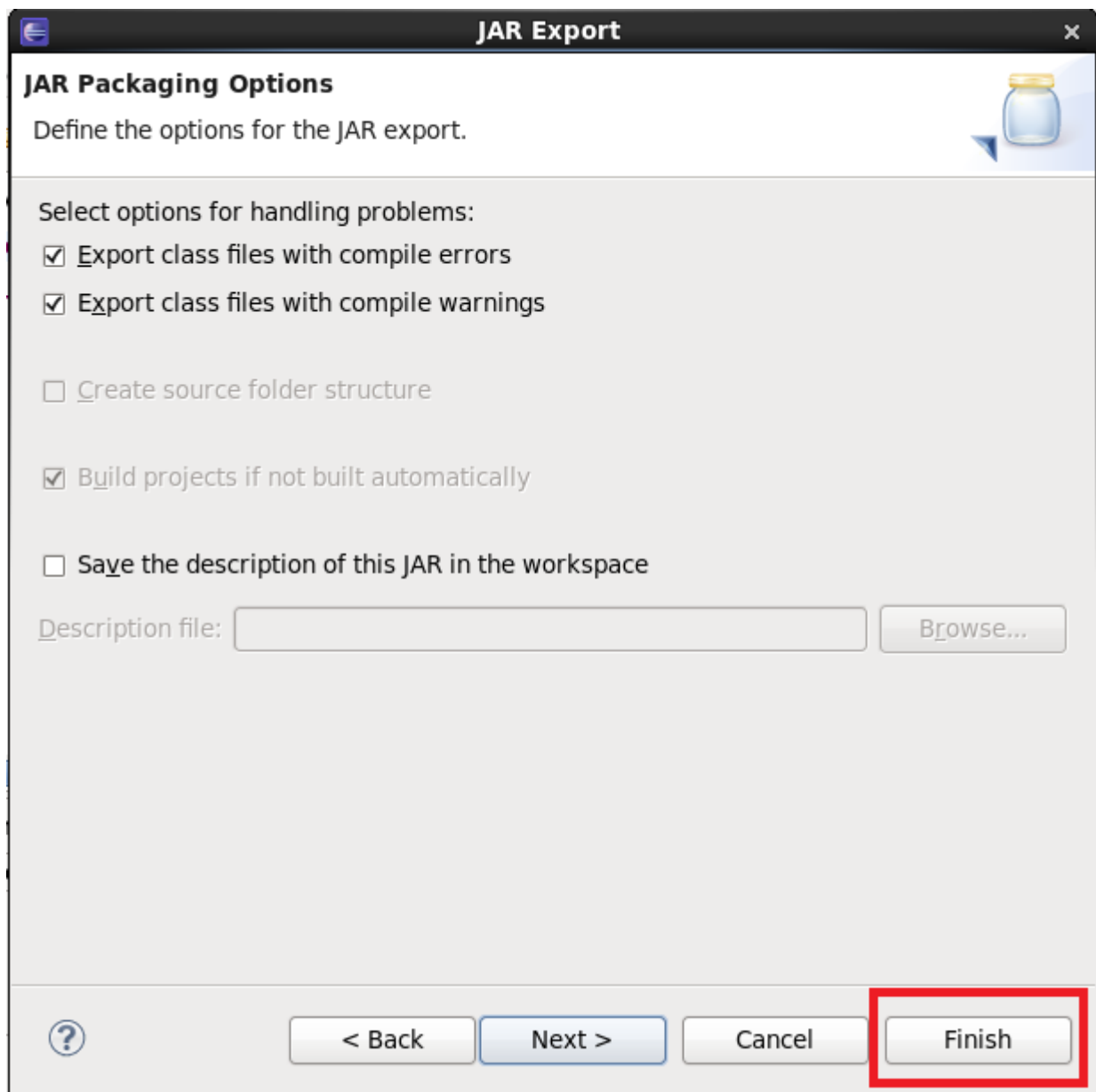
## JAR Creation

Follow the below images to create the JAR file for Pig UDF. Here, we have shown for one case, in same way you can create for every UDF.









After creating the JARs and text files, we have moved all the data to HDFS cluster, which is depicted by the following image.

```
[edureka@localhost Desktop]$ hadoop dfs -ls /
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

14/12/21 14:33:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
re applicable
Found 22 items
-rw-r--r-- 1 edureka supergroup          40 2014-12-19 19:12 /123
drwxr-xr-x - edureka supergroup           0 2014-12-17 22:06 /Dec17_DEIDENTIFY
drwxr-xr-x - edureka supergroup           0 2014-10-12 11:16 /amr
drwxr-xr-x - edureka supergroup           0 2014-12-20 16:17 /country
-rw-r--r-- 1 edureka supergroup          42 2014-12-21 14:32 /file
-rw-r--r-- 1 edureka supergroup        211 2014-12-17 22:04 /hdata1.csv
-rw-r--r-- 1 edureka supergroup           49 2014-12-19 14:48 /new
drwxr-xr-x - edureka supergroup           0 2014-12-04 18:57 /nov11
-rw-r--r-- 1 edureka supergroup    194457 2014-12-20 10:36 /nyse
drwxr-xr-x - edureka supergroup           0 2014-11-30 10:54 /pig
 72 2014-12-16 21:51 /pigdata
  0 2014-11-30 11:04 /pigextdata
1154 2014-12-21 14:33 /pow.jar
 17 2014-12-21 14:26 /power
1156 2014-12-21 12:23 /power.jar
drwxr-xr-x - edureka supergroup           0 2014-12-05 09:35 /springoutput
drwxr-xr-x - edureka supergroup           0 2014-12-20 21:11 /stack
drwxrwxrwx - root supergroup             0 2014-12-21 13:19 /tmp
-rw-r--r-- 1 edureka supergroup       952 2014-12-21 14:35 /upper.jar
drwxrwxrwx - edureka supergroup           0 2014-11-04 17:34 /user
drwxr-xr-x - edureka supergroup           0 2014-09-24 19:02 /usr
-rw-r--r-- 1 edureka supergroup          12 2014-12-21 12:23 /w1
[edureka@localhost Desktop]$
```

**Command to check the contents of HDFS.**

**Data for Upper Case**

**JAR for Power UDF**

**Data for Power UDF**

**JAR for Upper Case String**

After this, created two text files with .pig extension and put all the commands in that script file.

Then in terminal, typed PIG followed by the name of the script file which is shown in the following image for both the cases.

- Converting a particular column in Upper-Case.

```
[edureka@localhost Desktop]$ pig upper.pig
2014-12-21 14:39:42,760 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.1-SNAPSHOT (rexported) compiled Mar 22 2014, 07:29
2014-12-21 14:39:42,760 [main] INFO org.apache.pig.Main - Error messages to: /home/edureka/Desktop/pig_1419152982757.log
2014-12-21 14:39:42,994 [main] INFO org.apache.pig.Main - CodeLoader - Unable to load native-hadoop library for your platform
.. using builtin-java class
register hdfs:/upper.jar;
A = load 'hdfs:/file' using PigStorage(' ')
as (f1:chararray, f2:chararray, f3:chararray);
B = FOREACH A generate Upper(f1), f2, Upper(f3);
mapreduce.jobtracker.address = hdfs://localhost:9000;
dump B;
2014-12-21 14:40:05,653 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2014-12-21 14:40:05,713 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(HELLO,dear,FRIEND)
(HOW,are,YOU)
(ALL,is,WELL)
2014-12-21 14:40:05,713 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
```

- Taking a value and raising its power.

```
[edureka@localhost Desktop]$ pig pow.pig
2014-12-21 14:36:53,731 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.1-SNAPSHOT (rexported) compiled Mar 22 2014, 07:41:29
2014-12-21 14:36:53,731 [main] INFO org.apache.pig.Main - Error messages to: /home/edureka/Desktop/pig_1419152813729.log
2014-12-21 14:36:53,965 [main] INFO org.apache.pig.Main - CodeLoader - Unable to load native-hadoop library for your platform
.. using builtin-java class
register hdfs:/pow.jar;
A = load 'hdfs:/power' using PigStorage(' ')
as (f1:int, f2:int);
B = FOREACH A generate Pow(f1,f2);
dump B;
2014-12-21 14:37:17,205 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2014-12-21 14:37:17,223 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2014-12-21 14:37:17,223 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(8)
(16)
(25)
(1000)
2014-12-21 14:37:17,268 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
```

**All the Commands, Java Code and both Pig Scripts (Upper.pig & pow.pig) has been given on Page 2 under Important Links Section.**