# Run Spark remotely through eclipse

This article describes how to run the spark program on the eclipse. There are three modes, Spark Standalone, YARN Client Mode and YARN Cluster Mode. The difference between the three modes are shown in Table 1-1.

Table 1-1.Spark Runtime Summary

|  |  |  |  |
| --- | --- | --- | --- |
| **Mode** | **YARN Client Mode** | **YARN Cluster Mode** | **Spark Standalone** |
| **Driver runs in** | Client | ApplicationMaster | Client |
| **Requests resources** | ApplicationMaster | ApplicationMaster | Client |
| **Starts executor processes** | YARN NodeManager | YARN NodeManager | Spark Worker |
| **Persistent services** | YARN ResourceManager and NodeManagers | YARN ResourceManager and NodeManagers | Spark Master and Workers |
| **Supports Spark Shell** | Yes | No | Yes |

1.Standalone

If you want to run Spark with Standalone mode through eclipse, consider the following:

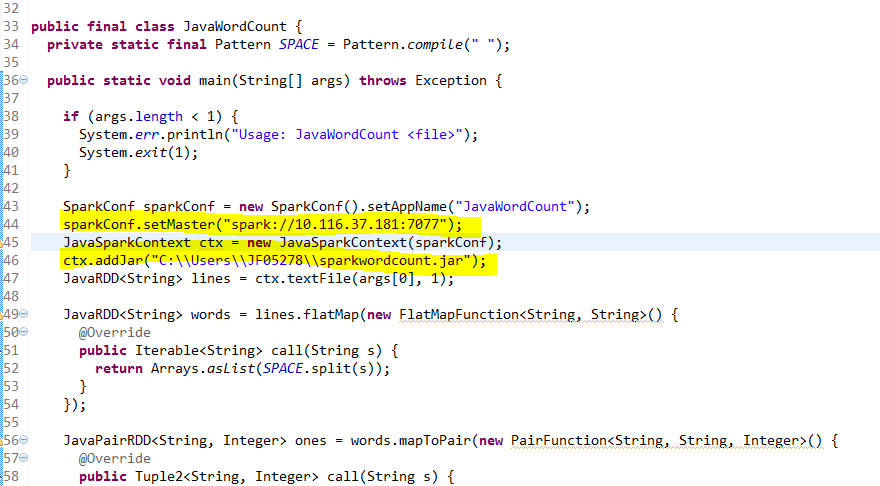
1). Create a java project;

Get the “JavaWordCount.java” from the spark decompression directory: “examples/src/main/java…”.

Get the “spark-assembly-1.5.2-hadoop2.6.0.jar”from the spark decompression directory: “spark/lib…”.

2).Add the “spark-assembly-1.5.2-hadoop2.6.0.jar” as the project external jar.

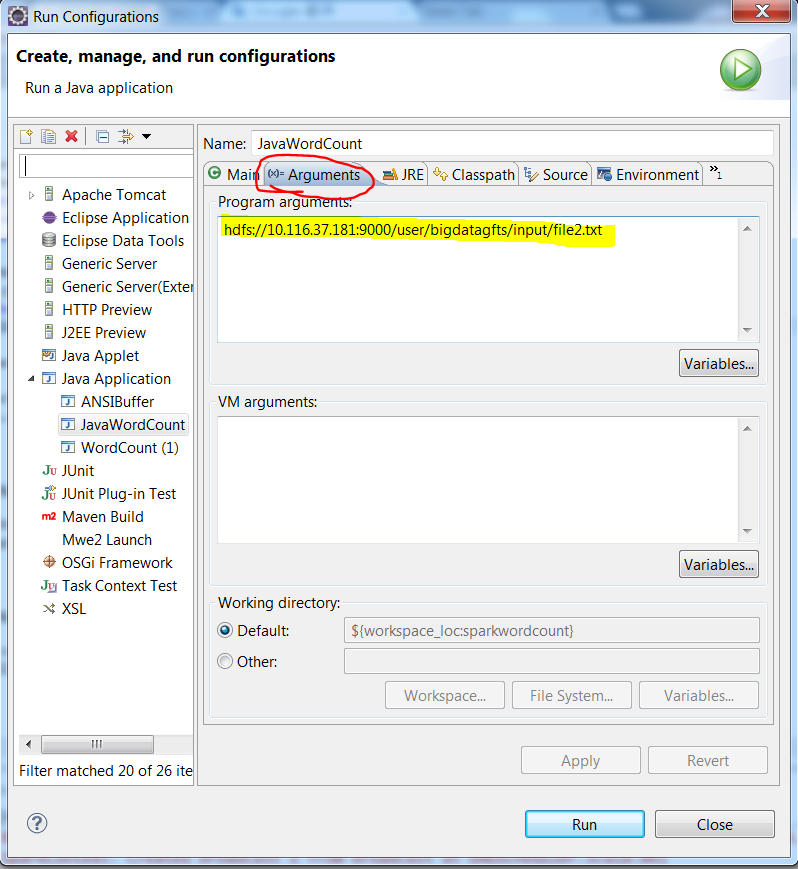
3).Modify the program as shown below.



Set the spark master: sparkConf.setMaster("spark://10.116.37.181:7077") ;Package project,

Add the jar about the wordcount project: ctx.addJar("C:\\Users\\JF05278\\sparkwordcount.jar") as the picture;The path is the package about the JavaWordCount project on the windows.

4). Set the args in the program,right click the program ,choose the Run As ->Run Configurations, then choose the Arguments tab, as the picture below. Add your file path in the hdfs as below which the file you will count word.



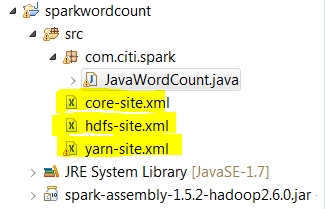
Then you can run the spark on Standalone mode through eclipse as Java Application.The result about wordcount will be shown in eclipse console.

2. YARN Client Mode

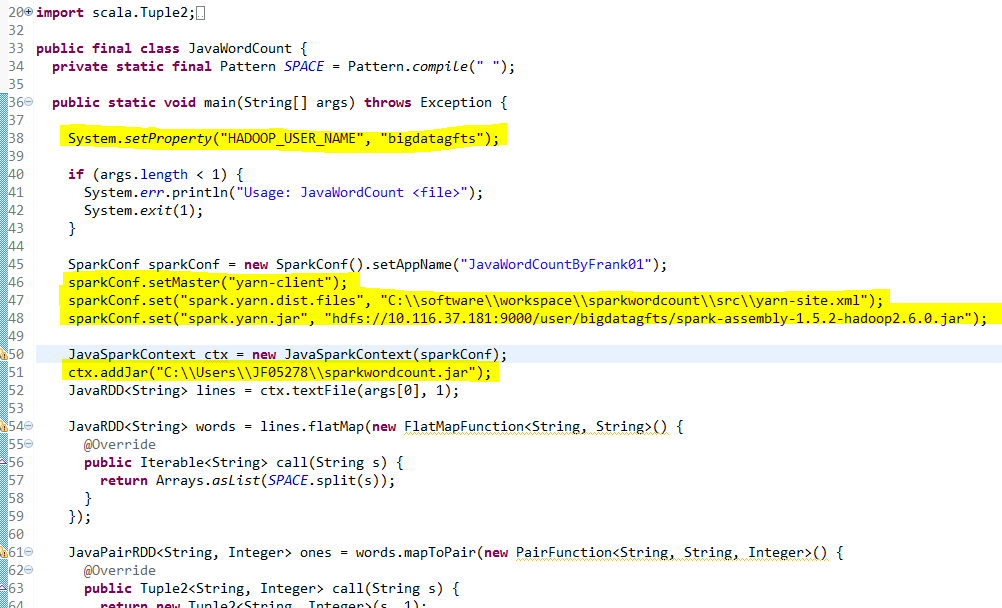
If you want to run Spark with YARN Client mode through eclipse, consider the following:

The step 1 and step 2 are same to the Standalone mode .

3).Add the profiles ‘core-site.xml’ ; ‘hdfs-site.xml’;’yarn-site.xml’ under src of the projects. The profiles can get from the path :vm-45e5-3412 /home/bigdatagfts/hadoop/etc/Hadoop.

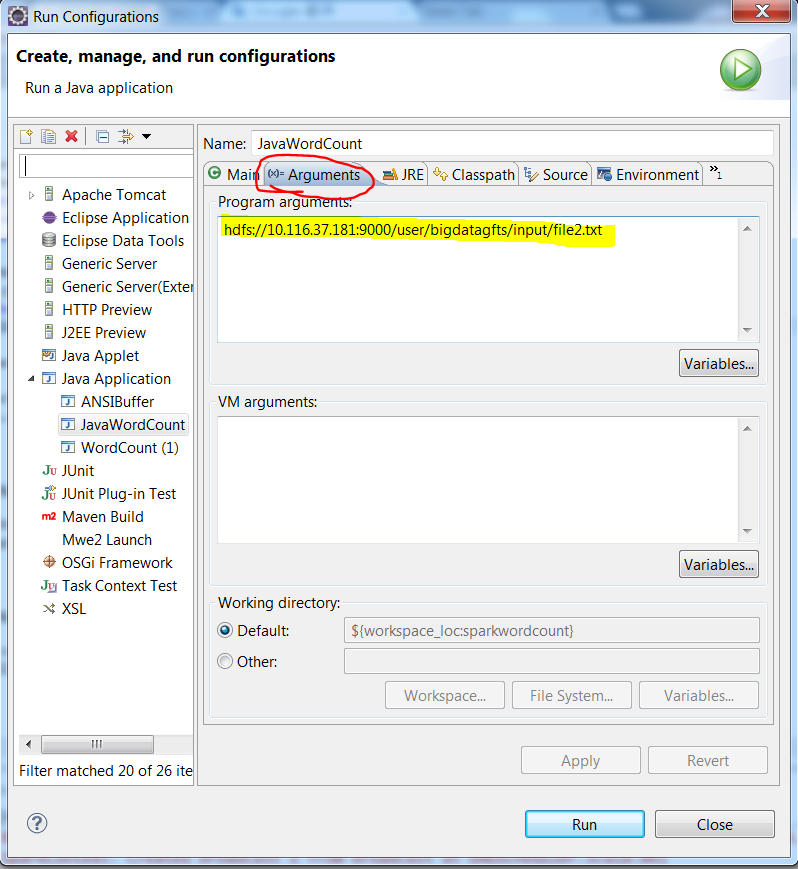


4). Modify the program as shown below.



5).Upload ‘spark-assembly-1.5.2-hadoop2.6.0.jar’ to hdfs. And set the ‘spark.yarn.jar’ to the path of the jar in hdfs. As the line 48 in the picture. And set the ‘spark.yarn.dist.files’ the path you put the yarn-site.xml. As the line 47 shown in the picture top.

6).Set the args in the program, right click the program ,choose the Run As ->Run Configurations, then choose the Arguments tab, as the picture below. Add the your file path in the hdfs as below which the file you will count word.



6).The right click the program ,choose run as Java Application, you can run on YARN-Client mode through eclipse.

The results appear in the console of the eclipse. Shown below.

