环境：

idea 2018+

scala 2.11+

maven-3.6.0

JDK1.8

spark2.0+

1、下载安装JDK1.8 配置环境变量

2、下载安装idea

https://www.jetbrains.com/idea/

3、idea安装scala插件

4、下载安装maven 配置环境变量，配置源 -- 阿里云或者其他私服

https://maven.apache.org/docs/history.html 我们不要下载最新的，我建议3.6.0

https://maven.apache.org/docs/3.6.0/release-notes.html

5、idea集成maven

6、idea创建一个scala的maven项目

https://docs.scala-lang.org/tutorials/scala-with-maven.html

https://davidb.github.io/scala-maven-plugin/example_java.html

```
1   <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-inst
2           xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/maven-v4_0_0.
3       <modelVersion>4.0.0</modelVersion>
4       <groupId>org.qzhd.algo</groupId>
5       <artifactId>training</artifactId>
6       <version>1.0-SNAPSHOT</version>
7
8
9       <properties>
10          <java.version>1.8</java.version>
11          <spark.version>2.4.1</spark.version>
12          <scope>provided</scope>
13      </properties>
14
15      <dependencies>
16          <dependency>
17              <groupId>org.apache.spark</groupId>
18              <artifactId>spark-core_2.11</artifactId>
19              <version>${spark.version}</version>
20              <scope>${scope}</scope>
21          </dependency>
22
23          <dependency>
24              <groupId>org.apache.spark</groupId>
25              <artifactId>spark-sql_2.11</artifactId>
26              <version>${spark.version}</version>
27              <scope>${scope}</scope>
28          </dependency>
```

```
29    </dependencies>
30    <build>
31        <plugins>
32            <plugin>
33                <groupId>net.alchim31.maven</groupId>
34                <artifactId>scala-maven-plugin</artifactId>
35                <version>3.2.2</version>
36                <executions>
37                    <execution>
38                        <id>scala-compile-first</id>
39                        <phase>process-resources</phase>
40                        <goals>
41                            <goal>add-source</goal>
42                            <goal>compile</goal>
43                        </goals>
44                    </execution>
45                    <execution>
46                        <id>scala-test-compile</id>
47                        <phase>process-test-resources</phase>
48                        <goals>
49                            <goal>testCompile</goal>
50                        </goals>
51                    </execution>
52                </executions>
53            </plugin>
54            <plugin>
55                <groupId>org.apache.maven.plugins</groupId>
56                <artifactId>maven-compiler-plugin</artifactId>
57                <version>3.5.1</version>
58                <executions>
59                    <execution>
60                        <phase>compile</phase>
61                        <goals>
62                            <goal>compile</goal>
63                        </goals>
64                    </execution>
65                </executions>
66                <configuration>
67                    <source>1.8</source>
68                    <target>1.8</target>
69                </configuration>
70            </plugin>
71        </plugins>
72    </build>
73 </project>
```

7、最简单的 wordcount

```
1 def main(args: Array[String]): Unit = {
2
3   val spark = SparkSession.builder()
4     .appName("wordcount")
5     .master("local")
6     .getOrCreate()
```

```scala
7
8    val rdd = spark.sparkContext.textFile("/Users/zhaosl/workspace/training/data/iris.data")
9
10   //4.7,3.2,1.3,0.2,Iris-setosa
11   val result = rdd.flatMap(_.split(",")).map(x=>(x,1)).reduceByKey(_+_)
12
13   result.collect().foreach(println)
14 }
```