

# Maven初探

更优雅地开发Hadoop程序

杨强

2014/6/5

# 小调查

- Makefile
- ant
- sbt
- Hadoop 编程?



自动化build  
工具

# 提要

- 为什么需要Maven
- 如何使用Maven
- 实例：
  - 使用Maven管理profsearch\_dm项目

# 问题引入：开发Hadoop程序

## 不使用IDE编程的流程

### 编码

```
import org.apache.hadoop.mapreduce.Mapper;
```

### 编译

```
javac -classpath $HADOOP_HOME/hadoop-  
core-1.2.1.jar -d ./classes/ ./src/*.java
```

### 打包

```
jar -cvf wordcount.jar -C ./classes/ .
```

### 运行

```
hadoop jar wordcount.jar WordCount  
readme.txt output
```

# 问题是如何产生的？

- 从一个程序员到两个程序员，即协同开发的出现；
- 如何统一这些内容：
  - 依赖的jar包：拷贝文件，更新不方便
  - 代码架构：包命名规则
  - 项目发布方式：名称、版本号

# Maven登場

- *Apache Maven is a software project management and comprehension tool. Based on the concept of a project object model (POM), Maven can manage a project's build, reporting and documentation from a central piece of information.*
- *Objectives*
  - *Making the **build** process easy*
  - *Providing a uniform **build** system*
  - *Providing quality project information*
  - *Providing guidelines for best practices **development***
  - *Allowing transparent migration to new features*

# Maven使用方法

安装

创建项目

打包

下载并解压：

```
wget  
http://mirror.tcpdiag.net/apache/maven  
/maven-3/3.2.1/binaries/apache-maven-  
3.2.1-bin.tar.gz
```

```
tar xf apache-maven-3.2.1-bin.tar.gz
```

配置：vim ~/.bashrc

```
export M2_HOME=/usr/local/apache-  
maven/apache-maven-3.2.1
```

```
export M2=$M2_HOME/bin  
export PATH=$M2:$PATH
```

```
export  
JAVA_HOME=/usr/java/jdk1.7.0_51
```

# Maven使用方法

安装

创建项目

打包

执行命令：

```
mvn archetype:generate \  
-DgroupId=com.mycompany.app \  
-DartifactId=my-app \  
-DarchetypeArtifactId=maven-  
archetype-quickstart \  
-DinteractiveMode=false
```

文件结构

```
my-app  
|-- pom.xml  
-- src  
    |-- main  
    |   |-- java  
    |       |-- com  
    |           |-- mycompany  
    |               |-- app  
    |                   |-- App.java  
    -- test  
        |-- java  
            |-- com  
                |-- mycompany  
                    |-- app  
                        |-- AppTest.java
```



# Maven使用方法



安装

创建项目

打包

打包:

```
mvn package
```

运行:

```
java -cp target/my-app-1.0-SNAPSHOT.jar  
com.mycompany.app.App
```

# POM

```
<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.mycompany.app</groupId>
  <artifactId>my-app</artifactId>
  <version>1.0-SNAPSHOT</version>
  <packaging>jar</packaging>

  <name>Maven Quick Start
  Archetype</name>
  <url>http://maven.apache.org</url>
```

```
<dependencies>
  <dependency>
    <groupId>junit</groupId>
    <artifactId>junit</artifactId>
    <version>4.8.2</version>
    <scope>test</scope>
  </dependency>
</dependencies>
</project>
```

POM stands for "Project Object Model". It is an XML representation of a Maven project held in a file named pom.xml.

# 实例

- ProfSearch Data Management
  - 初始版本
    - 依赖的jar包全部手动添加到Build Path
    - 编写了一个专门用于生成jar包的shell脚本
  - 如何转变为一个Maven项目？

# 实例： profsearch\_dm

创建项目

调整代码结构

配置POM

## 执行命令

```
mvn archetype:generate \  
    -DarchetypeGroupId=\   
        org.apache.maven.archetypes \  
    -DarchetypeArtifactId=\   
        maven-archetype-quickstart \  
    -DgroupId=cn.ac.ict.acs \  
    -DartifactId=profsearch_dm
```

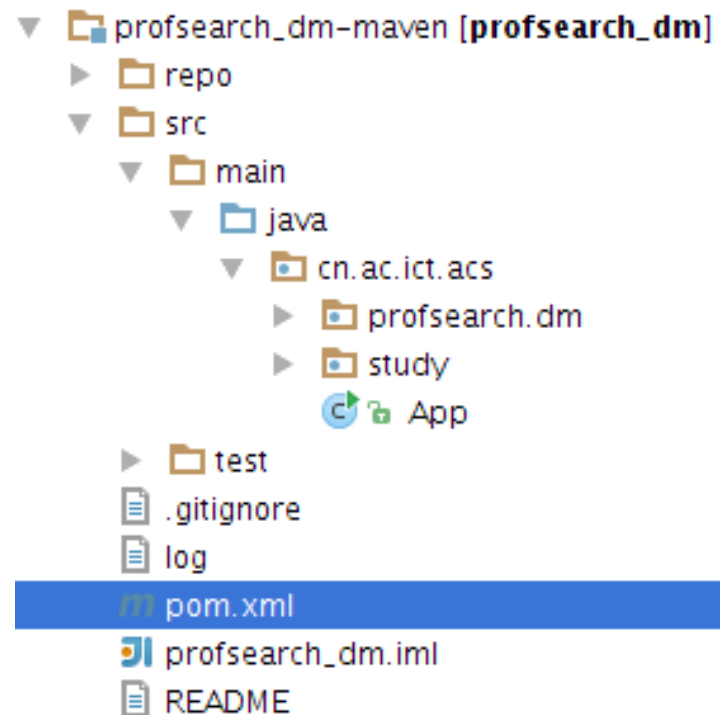
# 实例：profsearch\_dm

创建项目

调整代码结构

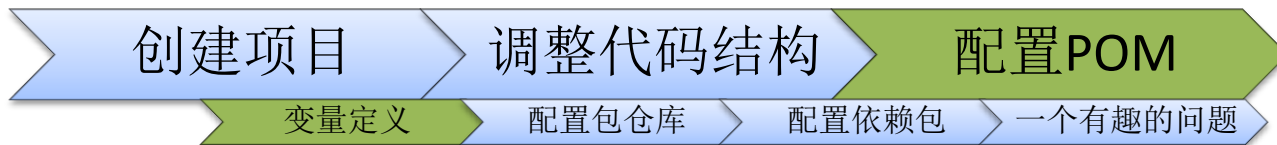
配置POM

- 主要是包名的改变：
  - profsearch. \*



- cn. ac. ict. acs. profsearch. dm. \*

# 实例：profsearch\_dm



```
<properties>
```

```
  <project.build.sourceEncoding>
```

```
    UTF-8
```

```
  </project.build.sourceEncoding>
```

```
  <hadoop.version>
```

```
    2.0.0-cdh4.3.2
```

```
  </hadoop.version>
```

```
  <hbase.version>
```

```
    0.94.6-cdh4.3.2
```

```
  </hbase.version>
```

```
</properties>
```

hadoop版本

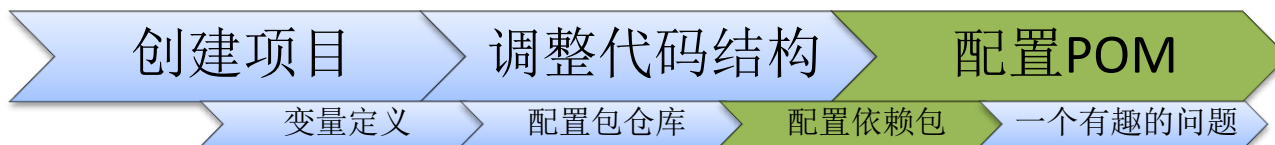
hbase版本

# 实例：profsearch\_dm



```
<repositories>
  <repository>
    <id>cloudera-releases</id>
    <url>
https://repository.cloudera.com/artifactory/cloudera-repos
    </url>
    <releases>
      <enabled>true</enabled>
    </releases>
    <snapshots>
      <enabled>false</enabled>
    </snapshots>
  </repository>
</repositories>
```

# 实例： profsearch\_dm

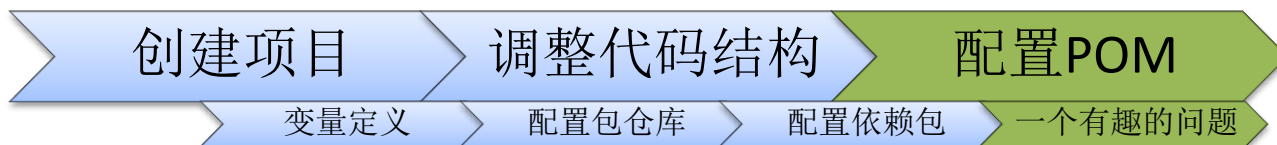


```
<dependency>
  <groupId>org. apache. hadoop</groupId>
  <artifactId>hadoop-client</artifactId>
  <version>${hadoop. version}</version>
</dependency>
```

```
<dependency>
  <groupId>org. apache. hbase</groupId>
  <artifactId>hbase</artifactId>
  <version>${hbase. version}</version>
</dependency>
```



# 实例： profsearch\_dm



如果jar包不存在于任何repository怎么办？

```
<dependency>
```

```
  <groupId>org. apache. hadoop</groupId>
```

```
  <artifactId>TarFileSystem</artifactId>
```

```
  <version>1.0</version>
```

```
  <scope>system</scope>
```

```
  <systemPath>${project.basedir}/repo/TarFileSystem-  
1.0.jar</systemPath>
```

```
</dependency>
```

搜索本地文件

# 参考资料

- maven官网：
  - <http://maven.apache.org/>
  - Download & Setup Maven
    - <http://maven.apache.org/download.cgi>
- 文档集合：
  - <http://maven.apache.org/users/index.html>
  - Maven in 5 minutes:
    - <http://maven.apache.org/guides/getting-started/maven-in-five-minutes.html>
  - POM手册
    - <http://maven.apache.org/pom.html>
- How-to: Develop CDH Applications with Maven and Eclipse
  - <http://blog.cloudera.com/blog/2012/08/developing-cdh-applications-with-maven-and-eclipse/>

谢谢，  
欢迎讨论。