Reproducible Builds for the JVM and beyond

Hervé Boutemy

Halifax, NS, 2023-10-10
About Me

- Maven PMC Member, Attic PMC Chair
- ASF Member
- working on Software Supply Chain @ Sonatype
- SBOM: CycloneDX, SPDX
- signature: Sigstore

Reproducible Builds for the JVM:
  - discovered in April 2016 (post-processing)
  - actively working since January 2019 (Maven built-in)
Agenda

- Reproducible Builds
  - what? why? how?
- Reproducible Builds for the JVM
  - checking against Maven Central
  - configuring for Maven, Gradle, sbt
- Quiz: to be or not to be Reproducible
- What’s next?
Reproducible Builds: what? why? how?
input source code

builder

reference

output binaries

reference

rebuilder

same output binaries (bit for bit)

https://reproducible-builds.org/ (since 2013)
a set of software development practices that create an independently-verifiable path from source to binary code
Why does it matter?

- reproducible-builds.org:
  “allow verification that no vulnerabilities or backdoors have been introduced during the compilation process”
- my own return on experience
  - you have the source, but are you *really* able to rebuild?
    - is it the real Git commit? is “Build successful” message sufficient?
  - are you sure nothing from your build environment leaked into output binaries?
    - found username, hostname, path to current directory, private key passphrase, ...
  - permits build efficiency from build cache
- **ASF policy: official source release vs convenience binaries**
  - how do you audit binaries staged by release manager? “Just trust”?
How?

- reproducible-build.org:

  3. users should be given a way to recreate a close enough build environment, perform the build process, and validate that the output matches the original build.

  2. the set of tools used to perform the build and more generally the build environment should either be recorded or pre-defined.

  1. the build system needs to be made entirely deterministic.
     For example, the current date and time must not be recorded and output always has to be written in the same order.
Reproducible Builds for the JVM:
2. check binaries: Maven Central
1. configure build: Maven, Gradle, sbt
Reproducible Central (started 03-2020)
https://github.com/jvm-repo-rebuild/reproducible-central

1. Tools and methods allowing to verify that Java builds are reproducible
2. A list of reproducible releases published to Maven Central
   rebuilding 2611 releases of 550 projects:
   ○ 2027 releases are confirmed fully reproducible (100% reproducible artifacts ✓),
   ○ 584 releases are only partially reproducible (contain some unreproducible artifacts ⚠)
   ○ on 550 projects, 439 have at least one fully reproducible release, 111 have none
Reproducible Central
https://github.com/jvm-repo-rebuild/reproducible-central

<table>
<thead>
<tr>
<th>Central Repository groupId</th>
<th>artifactId(s)</th>
<th>versions</th>
<th>result: reproducible?</th>
</tr>
</thead>
<tbody>
<tr>
<td>biz.aQute.bnd</td>
<td>bnd-plugin-parent</td>
<td>8</td>
<td>8 ✓</td>
</tr>
<tr>
<td>ch.qos.logback</td>
<td>logback-parent</td>
<td>16</td>
<td>10 ✓ / 6 ⚠️</td>
</tr>
<tr>
<td>ch.qos.reload4j</td>
<td>reload4j</td>
<td>7</td>
<td>1 ✓ / 6 ⚠️</td>
</tr>
<tr>
<td>ch.qos.logback.db</td>
<td>logback-parent-db</td>
<td>1</td>
<td>1 ⚠️</td>
</tr>
<tr>
<td>com.flowlogix</td>
<td>flowlogix</td>
<td>10</td>
<td>6 ✓ / 4 ⚠️</td>
</tr>
</tbody>
</table>
Project: org.apache.maven.plugins:maven-javadoc-plugin

Source code: https://github.com/apache/maven-javadoc-plugin.git

rebuilding 8 releases of org.apache.maven.plugins:maven-javadoc-plugin:

- 5 releases were found successfully **fully reproducible** (100% reproducible artifacts ✔),
- 3 had issues (some unreproducible artifacts ⚠️, see eventual 📦, diffoscope and/or 🛡️ issue tracker links):

<table>
<thead>
<tr>
<th>version</th>
<th>build spec</th>
<th>result: reproducible?</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6.0</td>
<td>mvn jdk17</td>
<td>result: 5 ✔ 1 ⚠️</td>
<td>4.6M</td>
</tr>
<tr>
<td>3.5.0</td>
<td>mvn jdk8 w</td>
<td>result: 4 ✔</td>
<td>4.2M</td>
</tr>
<tr>
<td>3.4.1</td>
<td>mvn jdk8 w</td>
<td>result: 4 ✔</td>
<td>4.2M</td>
</tr>
<tr>
<td>3.4.0</td>
<td>mvn jdk8 w</td>
<td>result: 4 ✔</td>
<td>4.2M</td>
</tr>
<tr>
<td>3.3.2</td>
<td>mvn jdk8 w</td>
<td>result: 4 ✔</td>
<td>4.2M</td>
</tr>
<tr>
<td>3.3.1</td>
<td>mvn jdk8 w</td>
<td>result: 3 ✔ 1 ⚠️</td>
<td>4.2M</td>
</tr>
<tr>
<td>3.3.0</td>
<td>mvn jdk8 w</td>
<td>result: 4 ✔</td>
<td>4.1M</td>
</tr>
<tr>
<td>3.2.0</td>
<td>mvn jdk8</td>
<td>result: 3 ✔ 1 ⚠️</td>
<td>4.0M</td>
</tr>
</tbody>
</table>
`,rebuild.sh`, `<path/to/...>/.<project>-<version>.buildspec`

```bash
$ maven-javadoc-plugin-3.5.0.buildspec

content > org > apache > maven > plugins > maven-javadoc-plugin > $ maven-javadoc-plugin-3.5.0.buildspec

```{
```bash
1  groupId=org.apache.maven.plugins
2  artifactId=maven-javadoc-plugin
3  display=${groupId}:${artifactId}
4  version=3.5.0
5
6  gitRepo=https://github.com/apache/${artifactId}.git
7  gitTag=${artifactId}-${version}
8
9  tool=mvn
10  jdk=8
11  newline=crlf
12
13  command="mvn -Papache-release clean package -Dmaven.javadoc.skip -Dgpg.skip -DskipTests"
14  buildinfo=target/${artifactId}-${version}.buildinfo
15
16  issue=
17  ```
Rebuilding from spec content/org/apache/maven/plugins/maven-javadoc-plugin/maven-javadoc-plugin-3.5.0.buildspec

- groupId: org.apache.maven.plugins
- artifactId: maven-javadoc-plugin
- version: 3.5.0
- gitRepo: https://github.com/apache/maven-javadoc-plugin.git
- gitTag: maven-javadoc-plugin-3.5.0
- tool: maven
- jdk: 8
- locale: "en_US"
- umask: "0002"
- command: mvn -P-Papache-release clean package -Dmaven.javadoc.skip -Dpom.skip -DskipTests
- newline: "crlf"
- buildinfo: target/maven-javadoc-plugin-3.5.0.buildinfo

Fetching source code from Git https://github.com/apache/maven-javadoc-plugin.git on tag maven-javadoc-plugin-3.5.0

Rebuilder Docker image is ready for use: rb-ubuntu-2204-jdk8-mvn3.6.3-toolchains-8-hboutemy-utc_en_us-0002

Rebuilding org.apache.maven.plugins:maven-javadoc-plugin:3.5.0 release 3.5.0 (version)


Apache Maven 3.6.3 (ceceddd343002696d0abb50b32b5418a6b2a2883f)
Maven home: /usr/local/apache-maven
Java version: 1.8.0_362, vendor: Private Build, runtime: /usr/lib/jvm/java-8-openjdk-amd64/jre
Default locale: en_US, platform encoding: UTF-8
05 name: "linux", version: "6.1.51-0-virt", arch: "amd64", family: "unix"

rebuild from content/org/apache/maven/plugins/maven-javadoc-plugin/maven-javadoc-plugin-3.5.0.buildspec
results in content/org/apache/maven/plugins/maven-javadoc-plugin/maven-javadoc-plugin-3.5.0.buildinfo
compared to Central Repository content/org/apache/maven/plugins/maven-javadoc-plugin/maven-javadoc-plugin-3.5.0.buildinfo
ok=4
What If a Difference is Found?

1. Where is the difference?

2. What is the difference?

https://diffoscope.org/

diffoscope

In-depth comparison of files, archives, and directories.
What If a Difference is Found?

1. Where is the difference?

2. What is the difference?  
   [Link: https://diffoscope.org/]

3. Why? How to Fix?
Reproducible Builds for the JVM:

2. check binaries: Maven Central
1. configure build: Maven, Gradle, sbt
Reproducible Builds for Maven (since 03-2020)


1. Enable Reproducible Builds:

   `<properties>
   <project.build.outputTimestamp>2023-01-01T00:00:00Z</project.build.outputTimestamp>
   </properties>`

2. Check plugins known to require upgrade: `mvn artifact:check-buildplan`

Checking for Reproducible Builds

3. **after release pushed to Maven Central:**
   
   ```
   mvn -Papache-release -Dgpg.skip clean verify artifact:compare
   ```

2. **during VOTE:**
   
   ```
   mvn -Papache-release -Dgpg.skip clean verify artifact:compare
   ```

   ```
   -Dreference.repo=https://repository.apache.org/content/repositories/staging/
   ```

1. **during SNAPSHOT development:**
   
   Check locally if you get the same result twice
   
   ```
   mvn clean install
   ```
   
   ```
   mvn clean verify artifact:compare
   ```
   
   ideally (harder): rebuilder on a different machine, or Docker, to detect more subtle environment impact
Reproducible Builds for Gradle

- since Gradle 3.4
  
  https://docs.gradle.org/current/userguide/working_with_files.html#sec:reproducible_archives

```
build.gradle

tasks.withType(AbstractArchiveTask).configureEach {
    preserveFileTimestamps = false
    ReproducibleFileOrder = true
}
```
Gradle in Reproducible Central

**Project:** [org.mockito:mockito-core](https://github.com/mockito/mockito.git)

- This project defines 8 modules:

  rebuilding **18 releases** of [org.mockito:mockito-core](https://github.com/mockito/mockito.git):

    - 15 releases were found successfully **fully reproducible** (100% reproducible artifacts ✓).
    - 3 had issues (some unreproducible artifacts 🔴, see eventual 🔍, diffoscope and/or 📘 issue tracker links):

<table>
<thead>
<tr>
<th>version</th>
<th>build spec</th>
<th>result: reproducible?</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5.0</td>
<td>gradle</td>
<td>result: 15 ✓ 4 🔴 🔍</td>
<td>1.3M</td>
</tr>
<tr>
<td>5.4.0</td>
<td>gradle</td>
<td>result: 19 ✓</td>
<td>1.3M</td>
</tr>
<tr>
<td>5.3.1</td>
<td>gradle</td>
<td>result: 19 ✓</td>
<td>1.3M</td>
</tr>
<tr>
<td>5.2.0</td>
<td>gradle</td>
<td>result: 19 ✓</td>
<td>1.3M</td>
</tr>
<tr>
<td>5.1.0</td>
<td>gradle</td>
<td>result: 19 ✓</td>
<td>1.3M</td>
</tr>
</tbody>
</table>
Reproducible Builds for sbt

**Project:** com.scalapenos:stamina_2.11

**Source code:** [https://github.com/scalapenos/stamina.git](https://github.com/scalapenos/stamina.git)

rebuilding 2 releases of com.scalapenos:stamina_2.11:

- 2 releases were found successfully fully reproducible (100% reproducible artifacts ✓),
- 0 had issues (some unreproducible artifacts 💥, see eventual 📦 diffoscope and/or 📊 issue tracker links):

<table>
<thead>
<tr>
<th>version</th>
<th>build spec</th>
<th>result: reproducible?</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1.6</td>
<td>sbt jdk8</td>
<td>24 ✓</td>
<td>146K</td>
</tr>
<tr>
<td>0.1.5</td>
<td>sbt jdk8</td>
<td>24 ✓</td>
<td>146K</td>
</tr>
</tbody>
</table>

Need Help!
Quiz:

to be or not to be Reproducible
#1 Reproducible or not?

```
[INFO] ---- artifact:3.5.0:compare (default-cli) @ maven-javadoc-plugin ----
[INFO] Saved info on build to /Users/hboutemy/dev/git/misc/reproducible-central/content/org/apan/target/maven-javadoc-plugin-3.5.0.buildinfo
[INFO] Checking against reference build from central...
Download from central: https://repo.maven.apache.org/maven2/org/apache/maven/plugins/maven-j
[INFO] Reference buildinfo file not found: it will be generated from downloaded reference artifi
[INFO] Reference build java.version: 1.8 (from MANIFEST.MF Build-Jdk-Spec)
[INFO] Reference build os.name: Windows (from pom.properties newline)
[INFO] Minimal buildinfo generated from downloaded artifacts: /Users/hboutemy/dev/git/misc/repr
[INFO] Reproducible Build output summary: 4 files ok
```

[INFO] Reproducible Build output summary: 4 files ok
#2 Reproducible or not?

```
[INFO] Reference build java.version: 1.8 (from MANIFEST.MF Build-Jdk-Spec)
[ERROR] Current build java.version: 11 (from MANIFEST.MF Build-Jdk-Spec)
[INFO] Reference build os.name: Windows (from pom.properties newline)
[INFO] Minimal buildinfo generated from downloaded artifacts: /Users/hboutemy/dev/git/misc/
    /buildcache/maven-javadoc-plugin/target/reference/maven-javadoc-plugin-3.5.0.buildinfo
[ERROR] size mismatch maven-javadoc-plugin-3.5.0.jar: investigate with diffoscope target/re
ven-javadoc-plugin-3.5.0.jar
[ERROR] Reproducible Build output summary: 3 files ok, 1 different
```
#2 Reproducible or not?

```
--- META-INF/MANIFEST.MF
@@ -1,10 +1,10 @@
 Manifest-Version: 1.0
-Implementation-Title: Apache Maven Javadoc Plugin
-Implementation-Version: 3.5.0
-Specification-Vendor: The Apache Software Foundation
-Specification-Title: Apache Maven Javadoc Plugin
-Build-Jdk-Spec: 1.8
+Build-Jdk-Spec: 11
+Specification-Title: Apache Maven Javadoc Plugin

[INFO] Reference build java.version: 1.8 (from MANIFEST.MF Build-Jdk-Spec)
[ERROR] Current build java.version: 11 (from MANIFEST.MF Build-Jdk-Spec)
[INFO] Reference build os.name: Windows (from pom.properties newline)
[ERROR] size mismatch maven-javadoc-plugin-3.5.0.jar: investigate with diffoscope target/reference/maven-javadoc-plugin-3.5.0.jar
[ERROR] Reproducible Build output summary: 3 files ok, 1 different
```

MUNITY CODE
#3 Reproducible or not?
#3 Reproducible or not?

```
content > org > apache > maven > plugins > maven-shade-plugin
$ maven-shade-plugin-3.5.1.buildspec

1. groupId=org.apache.maven.plugins
2. artifactId=maven-shade-plugin
3. display=${groupId}:${artifactId}
4. version=3.5.1

5. gitRepo=https://github.com/apache/${artifactId}.git
6. gitTag=${artifactId}-${version}
7. tool=mvn
8. jdk=17
9. newline=\n
10. umask=022
11. command="mvn -Papache-release clean package -DskipTests -Dmaven.javadoc.skip -Dgpg.skip"
12. buildinfo=target/${artifactId}-${version}.buildinfo
13. issue=
```
#4 Reproducible or not?

```
./rebuild.sh content/com/flowlogix/flowlogix-7.0.2.buildspec
```

```
[INFO] Reference build java.version: 20 (from MANIFEST.MF Build-Jdk-Spec)
[INFO] Reference build os.name: Unix (from pom.properties newline)
[INFO] Minimal buildinfo generated from downloaded artifacts: /Users/hbouyem/dv/git/reference/jee-examples-7.0.2.buildinfo
[ERROR] sha512 mismatch jee-examples-7.0.2.war: investigate with diffoscope target/reference/jee-examples-7.0.2.war
[ERROR] Reproducible Build output summary: 11 files ok, 1 different
```

PLEASE use only LTS for release
#4 Reproducible or not?

<table>
<thead>
<tr>
<th>diffoscope</th>
<th>target/reference/com.flowlogix/jee-examples-7.0.2.war jakarta-ee/jee-examples/target/jee-exam</th>
<th><img src="image" alt="" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>***</td>
<td>target/reference/com.flowlogix/jee-examples-7.0.2.war</td>
<td>![image]</td>
</tr>
<tr>
<td>+---</td>
<td>jakarta-ee/jee-examples/target/jee-examples-7.0.2.war</td>
<td>![image]</td>
</tr>
<tr>
<td>--</td>
<td>zipinfo {}</td>
<td>![image]</td>
</tr>
<tr>
<td>@@ -36,21 +36,21 @@</td>
<td>![image]</td>
<td></td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>432 b- defN 23-Jun-27 19:19 WEB-INF/classes/com/flowlogix/examples/ui/servlets</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>1241 b- defN 23-Jun-27 19:19 WEB-INF/classes/com/flowlogix/logcapture/LogCapture</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>3166 b- defN 23-Jun-27 19:19 WEB-INF/classes/com/flowlogix/logcapture/LogCapture</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>830 b- defN 23-Jun-27 19:19 WEB-INF/classes/com/flowlogix/logcapture/LogCapture</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>190 b- defN 23-Jun-27 19:19 WEB-INF/classes/git.properties</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>1264 b- defN 23-Jun-27 19:19 WEB-INF/errorpages/invalidErrorPage.xhtml</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>2376 b- defN 23-Jun-27 19:19 WEB-INF/faces-config.xml</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>23391 b- defN 23-Jun-27 19:19 WEB-INF/lib/flowlogix-datamodel-7.0.2-tests.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>30477 b- defN 23-Jun-27 19:19 WEB-INF/lib/flowlogix-datamodel-7.0.2.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>37827 b- defN 23-Jun-27 19:19 WEB-INF/lib/flowlogix-jee-7.0.2-tests.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>42647 b- defN 23-Jun-27 19:19 WEB-INF/lib/flowlogix-jee-7.0.2.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>827896 b- defN 23-Jun-27 19:19 WEB-INF/lib/omnifaces-4.1.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>+rw-r--r-- 2.0 unix</td>
<td>4802204 b- defN 23-Jun-27 19:19 WEB-INF/lib/primefaces-12.0.0-jakarta.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>+rw-r--r-- 2.0 unix</td>
<td>4802204 b- defN 23-Jun-27 19:19 WEB-INF/lib/primefaces-12.0.0-jakarta.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>-rw-r--r-- 2.0 unix</td>
<td>63635 b- defN 23-Jun-27 19:19 WEB-INF/lib/slf4j-api-2.0.7.jar</td>
<td>![image]</td>
</tr>
<tr>
<td>+rw-r--r-- 2.0 unix</td>
<td>10346 b- defN 23-Jun-27 19:19 WEB-INF/lib/slf4j-jdk14-2.0.7.jar</td>
<td>![image]</td>
</tr>
</tbody>
</table>

PLEASE use only LTS for release
What’s next?
for the JVM... and Beyond...

- **Maven:**
  - make more Maven plugins produce Reproducible output
  - help more projects enable Reproducible Builds
- **Gradle:**
  - help more projects enable Reproducible Builds
  - improve Reproducible Central rebuilds for these
- **sbt**
- **npm & npmjs**
- **pip & PyPI**
- **.NET & NuGet Gallery**
- …
for the ASF: please audit your binaries during VOTEs

but Reproducible not fully ok: reference build done with JDK 17 on *nix and umask 022
apache-maven-3.9.5-bin.zip and .tar.gz suffer from weird umask (go-) on wagon jars:

$ diffoscope target/reference/org.apache.maven/apache-maven-3.9.5-bin.zip apache-maven/target/apache-maven-3.9.5-bin.zip
--- target/reference/org.apache.maven/apache-maven-3.9.5-bin.zip
+++ apache-maven/target/apache-maven-3.9.5-bin.zip
  Archive contents identical but files differ, possibly:
  zipinfo 

Reproducible Builds now ok: reference build done with JDK 17 on *nix and umask 022

mvn -Papache-release -Dgpg.skip clean verify artifact:compare
  -Dreference.repo=https://repository.apache.org/content/repositories/staging/

it’s ok not to be RB perfect

next time will be better
Merci

Diversity in Community is Great, not in Binary Code