

Item 170

**git\_comments:**

1. extract paths from absolute URI's
2. compare results:
3. make the archive:
4. not part of result

**git\_commits:**

1. **summary:** HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov.  
**message:** HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov.

**github\_issues:**

**github\_issues\_comments:**

**github\_pulls:**

**github\_pulls\_comments:**

**github\_pulls\_reviews:**

**jira\_issues:**

1. **summary:** Hadoop Archives should support globs resolving to files  
**description:** Consider the following scenario: {code} \$ hadoop fs -ls /tmp/harsrc/dir2/dir3 Found 5 items  
-rw-r--r-- 1 blah blah 0 2014-10-13 20:59 /tmp/harsrc/dir2/dir3/file31 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 /tmp/harsrc/dir2/dir3/file32 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 /tmp/harsrc/dir2/dir3/file33 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 /tmp/harsrc/dir2/dir3/file34 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 /tmp/harsrc/dir2/dir3/file35 {code} Archive 'dir3/file3\*': {code} \$ hadoop archive -Dmapreduce.framework.name=local -archiveName fileStar.har -p /tmp/harsrc 'dir2/dir3/file\*' /tmp/hardst\_local \$ hadoop fs -ls -R har:/tmp/hardst\_local/fileStar.har drwxr-xr-x - blah blah 0 2014-10-13 22:32 har:///tmp/hardst\_local/fileStar.har/dir2 {code} Archiving dir3 (directory) which is equivalent to the above works. {code} \$ hadoop archive -Dmapreduce.framework.name=local -archiveName dir3.har -p /tmp/harsrc 'dir2/dir3' /tmp/hardst\_local \$ hadoop fs -ls -R har:/tmp/hardst\_local/dir3.har 14/10/14 02:06:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable drwxr-xr-x - blah blah 0 2014-10-13 22:32 har:///tmp/hardst\_local/dir3.har/dir2 drwxr-xr-x - blah blah 0 2014-10-14 01:51 har:///tmp/hardst\_local/dir3.har/dir2/dir3 -rw-r--r-- 1 blah blah 0 2014-10-13 20:59 har:///tmp/hardst\_local/dir3.har/dir2/dir3/file31 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 har:///tmp/hardst\_local/dir3.har/dir2/dir3/file32 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 har:///tmp/hardst\_local/dir3.har/dir2/dir3/file33 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 har:///tmp/hardst\_local/dir3.har/dir2/dir3/file34 -rw-r--r-- 1 blah blah 0 2014-10-14 01:51 har:///tmp/hardst\_local/dir3.har/dir2/dir3/file35 {code}

**jira\_issues\_comments:**

1. v01 with unit tests demonstrating the problem and a proposed fix.
2. {color:green}+1 overall{color}. Here are the results of testing the latest attachment <http://issues.apache.org/jira/secure/attachment/12674733/HADOOP-11201.v01.patch> against trunk revision 5faaba0. {color:green}+1 @author{color}. The patch does not contain any @author tags. {color:green}+1 tests included{color}. The patch appears to include 1 new or modified test files. {color:green}+1 javac{color}. The applied patch does not increase the total number of javac compiler warnings. {color:green}+1 javadoc{color}. There were no new javadoc warning messages. {color:green}+1 eclipse:eclipse{color}. The patch built with eclipse:eclipse. {color:green}+1 findbugs{color}. The patch does not introduce any new Findbugs (version 2.0.3) warnings. {color:green}+1 release audit{color}. The applied patch does not increase the total number of release

audit warnings. {color:green}+1 core tests{color}. The patch passed unit tests in hadoop-tools/hadoop-archives. {color:green}+1 contrib tests{color}. The patch passed contrib unit tests. Test results: <https://builds.apache.org/job/PreCommit-HADOOP-Build/4922/testReport/> Console output: <https://builds.apache.org/job/PreCommit-HADOOP-Build/4922/console> This message is automatically generated.

3. Hi, [~jira.shegalov]. This is a good find. Thank you for posting a patch. From reading this code, it's unclear to me why it ever bothered splitting directory paths into the separate {{justDirs}} list. The subsequent logic operates on the parents of the paths. A parent is by definition always going to be a directory. I think we can remove {{justDirs}} entirely and change the inner loop to iterate directly over {{paths}} instead. Do you agree? If any of the other watchers can think of a reason for the {{justDirs}} logic, please comment. I couldn't find anything relevant in revision history.
4. [~cnauroth], thank you for taking a look. {{justDirs}} seems to be sort of a misnomer at this point that may have led to this bug, to begin with. It should be {{justPaths}}. The main thing going on there is extracting paths from, potentially absolute, input URI's. I can make this change if we have a consensus about the bug per se.
5. **body:** bq. The main thing going on there is extracting paths from, potentially absolute, input URI's. Ah, now it makes sense. In that case, I think renaming to {{justPaths}} would help clarify the code.  
**label:** code-design
6. v02 patch renaming justDirs to justPaths
7. {color:green}+1 overall{color}. Here are the results of testing the latest attachment <http://issues.apache.org/jira/secure/attachment/12682065/HADOOP-11201.v02.patch> against trunk revision 9dd5d67. {color:green}+1 @author{color}. The patch does not contain any @author tags. {color:green}+1 tests included{color}. The patch appears to include 1 new or modified test files. {color:green}+1 javac{color}. The applied patch does not increase the total number of javac compiler warnings. {color:green}+1 javadoc{color}. There were no new javadoc warning messages. {color:green}+1 eclipse:eclipse{color}. The patch built with eclipse:eclipse. {color:green}+1 findbugs{color}. The patch does not introduce any new Findbugs (version 2.0.3) warnings. {color:green}+1 release audit{color}. The applied patch does not increase the total number of release audit warnings. {color:green}+1 core tests{color}. The patch passed unit tests in hadoop-tools/hadoop-archives. {color:green}+1 contrib tests{color}. The patch passed contrib unit tests. Test results: <https://builds.apache.org/job/PreCommit-HADOOP-Build/5096/testReport/> Console output: <https://builds.apache.org/job/PreCommit-HADOOP-Build/5096/console> This message is automatically generated.
8. +1 for patch v2. I plan to commit this later today.
9. I have committed this to trunk and branch-2. Gera, thank you for contributing this patch.
10. FAILURE: Integrated in Hadoop-trunk-Commit #6572 (See [<https://builds.apache.org/job/Hadoop-trunk-Commit/6572/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java
11. FAILURE: Integrated in Hadoop-Yarn-trunk-Java8 #10 (See [<https://builds.apache.org/job/Hadoop-Yarn-trunk-Java8/10/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java
12. FAILURE: Integrated in Hadoop-Yarn-trunk #748 (See [<https://builds.apache.org/job/Hadoop-Yarn-trunk/748/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt
13. FAILURE: Integrated in Hadoop-Hdfs-trunk-Java8 #10 (See [<https://builds.apache.org/job/Hadoop-Hdfs-trunk-Java8/10/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java

14. SUCCESS: Integrated in Hadoop-Hdfs-trunk #1938 (See [<https://builds.apache.org/job/Hadoop-Hdfs-trunk/1938/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java
15. FAILURE: Integrated in Hadoop-Mapreduce-trunk #1962 (See [<https://builds.apache.org/job/Hadoop-Mapreduce-trunk/1962/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java
16. SUCCESS: Integrated in Hadoop-Mapreduce-trunk-Java8 #10 (See [<https://builds.apache.org/job/Hadoop-Mapreduce-trunk-Java8/10/>]) HADOOP-11201. Hadoop Archives should support globs resolving to files. Contributed by Gera Shegalov. (cnauroth: rev 79301e80d7510f055c01a06970bb409607a4197c) \* hadoop-tools/hadoop-archives/src/main/java/org/apache/hadoop/tools/HadoopArchives.java \* hadoop-common-project/hadoop-common/CHANGES.txt \* hadoop-tools/hadoop-archives/src/test/java/org/apache/hadoop/tools/TestHadoopArchives.java