

Public signup for this instance is **disabled**. Go to our [Self serve sign up page](#) to request an account.



Hadoop HDFS / HDFS-12278

LeaseManager operations are inefficient in 2.8.

Details

Type:	Bug	Status:	RESOLVED
Priority:	Blocker	Resolution:	Fixed
Affects Version/s:	2.8.0	Fix Version/s:	2.9.0, 3.0.0-beta1, 2.8.2
Component/s:	namenode		
Labels:	None		
Target Version/s:	3.0.0-beta1, 2.8.2		
Hadoop Flags:	Reviewed		

Description

After [HDFS-6757](#), LeaseManager #removeLease became expensive.
[HDFS-6757](#) changed the sortedLeases object from TreeSet to PriorityQueue.
Previously the remove(Object) operation from sortedLeases was $O(\log n)$ but after the change it became $O(n)$ since it has to find the object first.
Recently we had an incident in one of our production cluster just hours after we upgraded from 2.7 to 2.8
The sortedLeases object had approximately 100,000 items within it.
While removing the lease, it will acquire the LeaseManager lock and that will slow down the lookup of lease also.
[HDFS-6757](#) is a good improvement which replaced the path by inode id.

Attachments

HDFS-12278.patch	2 kB	09/Aug/17 17:39
HDFS-12278-branch-2.8.001.patch	2 kB	09/Aug/17 20:38


Issue Links

is broken by
HDFS-6757 Simplify lease manager with INodeID

Activity

- [Wei-Chiu Chuang](#) added a comment - 08/Aug/17 21:06
updated target version based on [HDFS-6757](#) fix version, so we keep track of it in Hadoop 3.0.0 release dashboard.
- [Daryn Sharp](#) added a comment - 09/Aug/17 16:40
For context regarding the impact of the change to a priority queue: Hours after a 2.8 upgrade, avg rpc processing time increased from sub-ms to 21ms. Rpc queue time was multiple seconds. Killing large jobs only made it worse. The fair call queue was completely overflowing for ~5h. I haven't seen anything this horrific in many years.
While the NN log was spewing logs of skipping calls from timing out clients, we noticed lease monitor recovery log messages ~5-12ms apart during which time the lease monitor holds the write lock. Killing jobs made it worse because it created more orphaned leases.
- [Rushabh Shah](#) added a comment - 09/Aug/17 17:40
Very small change.
Switched back to TreeSet from Priority Queue.
Ran basic LeaseManager related test cases.

- ▼

 Dave Star added a comment - 09/Aug/17 18:30


Changed the remove/update/add between the priority queue and original tree set to simulate renewals.

1k files = no difference

10k = pq is 3X slower

100k = pq is 13X slower

200k = pq is 22X slower
- ▼


 Hadoop QA added a comment - 09/Aug/17 19:32
- ✖

-1 overall
- | Vote | Subsystem | Runtime | Comment |
|------|------------|----------|---|
| 0 | reexec | 0m 18s | Docker mode activated. |
| | | | Prechecks |
| +1 | @author | 0m 0s | The patch does not contain any @author tags. |
| -1 | test4tests | 0m 0s | The patch doesn't appear to include any new or modified tests. Please justify why no new tests are needed for this patch. Also please list what manual steps were performed to verify this patch. |
| | | | trunk Compile Tests |
| +1 | mvninstall | 17m 57s | trunk passed |
| +1 | compile | 1m 0s | trunk passed |
| +1 | checkstyle | 0m 45s | trunk passed |
| +1 | mvnsite | 1m 4s | trunk passed |
| -1 | findbugs | 2m 3s | hadoop-hdfs-project/hadoop-hdfs in trunk has 9 extant Findbugs warnings. |
| +1 | javadoc | 0m 49s | trunk passed |
| | | | Patch Compile Tests |
| +1 | mvninstall | 1m 3s | the patch passed |
| +1 | compile | 1m 0s | the patch passed |
| +1 | javac | 1m 0s | the patch passed |
| +1 | checkstyle | 0m 40s | the patch passed |
| +1 | mvnsite | 1m 1s | the patch passed |
| +1 | whitespace | 0m 0s | The patch has no whitespace issues. |
| +1 | findbugs | 2m 10s | the patch passed |
| +1 | javadoc | 0m 45s | the patch passed |
| | | | Other Tests |
| -1 | unit | 68m 59s | hadoop-hdfs in the patch failed. |
| +1 | asflicense | 0m 16s | The patch does not generate ASF License warnings. |
| | | 101m 24s | |
- | Reason | Tests |
|--------------------|--|
| Failed junit tests | hadoop.hdfs.TestDFSStripedOutputStreamWithFailure150 |
| | hadoop.hdfs.server.blockmanagement.TestUnderReplicatedBlocks |
| | hadoop.hdfs.TestDFSStripedOutputStreamWithFailure080 |
- | Subsystem | Report/Notes |
|------------|----------------------------|
| Docker | Image:yetus/hadoop:14b5c93 |
| JIRA Issue | HDFS-12278 |
- https://issues.apache.org/jira/browse/HDFS-12278

2/6

JIRA Patch URL	https://issues.apache.org/jira/secure/attachment/12881050/HDFS-12278.patch
Optional Tests	asflicense compile javac javadoc mvninstall mvnsite unit findbugs checkstyle
uname	Linux 7231b045032c 3.13.0-116-generic #163-Ubuntu SMP Fri Mar 31 14:13:22 UTC 2017 x86_64 x86_64 x86_64 GNU/Linux
Build tool	maven
Personality	/testptch/hadoop/patchprocess/precommit/personality/provided.sh
git revision	trunk / 63cfc9
Default Java	1.8.0_131
findbugs	v3.1.0-RC1
findbugs	https://builds.apache.org/job/PreCommit-HDFS-Build/20620/artifact/patchprocess/branch-findbugs-hadoop-hdfs-project_hadoop-hdfs-warnings.html
unit	https://builds.apache.org/job/PreCommit-HDFS-Build/20620/artifact/patchprocess/patch-unit-hadoop-hdfs-project_hadoop-hdfs.txt
Test Results	https://builds.apache.org/job/PreCommit-HDFS-Build/20620/testReport/
modules	C: hadoop-hdfs-project/hadoop-hdfs U: hadoop-hdfs-project/hadoop-hdfs
Console output	https://builds.apache.org/job/PreCommit-HDFS-Build/20620/console
Powered by	Apache Yetus 0.6.0-SNAPSHOT http://yetus.apache.org

This message was automatically generated.

▼  [Rushabh Shah](#) added a comment - 09/Aug/17 20:38


TestDFSStripedOutputStreamWithFailure150 and TestDFSStripedOutputStreamWithFailure080 are well known flaky tests.

Both have failed number of times. I ran 4 times each test. Fails with a probability of 50%

TestUnderReplicatedBlocks#testSetRepIncWithUnderReplicatedBlocks is timing out. Tracked via [HDFS-9243](#).

There are no new tests since all the existing test covers the correctness.


Will attach a branch-2.8 patch soon.

▼  [Rushabh Shah](#) added a comment - 09/Aug/17 20:39

Attaching a branch-2.8 patch.

Imports conflict with trunk patch.

trunk patch applies nicely to branch-2.

▼  [Ravi Prakash](#) added a comment - 09/Aug/17 21:37

Hi Rushabh! Slightly on a tangent, but how did you benchmark TreeSet and PriorityQueue? Are you aware of JMH?

▼  [Kihwal Lee](#) added a comment - 09/Aug/17 21:44

+1 looks good.

▼  [Rushabh Shah](#) added a comment - 09/Aug/17 21:45 - edited

but how did you benchmark TreeSet and PriorityQueue

[daryn](#) benchmarked.

He just created lease-like objects and tested renew like methods.

Basically an object with string, int member variable and a comparator.


He created 100,000 such objects and called renew on them and measured via `monotonicTime`.

Daryn: please correct me if I am wrong.


Are you aware of JMH?

I wasn't aware until I read the comment and did web search. Thanks!

But it is very simple to understand that priority queue is not a good data structure if you want to remove any object other than the top one.


▼  [Kihwal Lee](#) added a comment - 09/Aug/17 22:03

I've committed this to trunk, branch-2, branch-2.8 and branch-2.8.2. Thanks for reporting and fixing the issue.

▼  [Rushabh Shah](#) added a comment - 09/Aug/17 22:07

Thanks [kihwal](#) for the review and commit !


Thanks [daryn](#) for the benchmarking the change.

▼  [Hudson](#) added a comment - 09/Aug/17 22:11

SUCCESS: Integrated in Jenkins build Hadoop-trunk-Commit #12156 (See <https://builds.apache.org/job/Hadoop-trunk-Commit/12156/>)

[HDFS-12278](#). LeaseManager operations are inefficient in 2.8. Contributed (kihwal: rev b5c02f95b5a2fcb8931d4a86f8192caa18009ea9)

- (edit) `hadoop-hdfs-project/hadoop-hdfs/src/main/java/org/apache/hadoop/hdfs/server/namenode/LeaseManager.java`

▼  [Hadoop QA](#) added a comment - 09/Aug/17 23:11

 **-1 overall**

Vote	Subsystem	Runtime	Comment
0	reexec	12m 33s	Docker mode activated.
Prechecks			
+1	@author	0m 0s	The patch does not contain any @author tags.
-1	test4tests	0m 0s	The patch doesn't appear to include any new or modified tests. Please justify why no new tests are needed for this patch. Also please list what manual steps were performed to verify this patch.
branch-2.8 Compile Tests			
+1	mvninstall	8m 53s	branch-2.8 passed
+1	compile	0m 47s	branch-2.8 passed with JDK v1.8.0_144
+1	compile	0m 46s	branch-2.8 passed with JDK v1.7.0_131
+1	checkstyle	0m 22s	branch-2.8 passed
+1	mvnsite	0m 59s	branch-2.8 passed
+1	findbugs	2m 11s	branch-2.8 passed
+1	javadoc	0m 44s	branch-2.8 passed with JDK v1.8.0_144
+1	javadoc	1m 0s	branch-2.8 passed with JDK v1.7.0_131
Patch Compile Tests			
+1	mvninstall	0m 45s	the patch passed
+1	compile	0m 38s	the patch passed with JDK v1.8.0_144
+1	javac	0m 38s	the patch passed
+1	compile	0m 41s	the patch passed with JDK v1.7.0_131
+1	javac	0m 41s	the patch passed
+1	checkstyle	0m 18s	the patch passed
+1	mvnsite	0m 49s	the patch passed
+1	whitespace	0m 0s	The patch has no whitespace issues.
+1	findbugs	2m 9s	the patch passed
+1	javadoc	0m 35s	the patch passed with JDK v1.8.0_144
+1	javadoc	0m 57s	the patch passed with JDK v1.7.0_131

			Other Tests
+1	unit	52m 11s	hadoop-hdfs in the patch passed with JDK v1.7.0_131.
+1	asflicense	0m 22s	The patch does not generate ASF License warnings.
		150m 38s	


Reason	Tests
JDK v1.8.0_144 Failed junit tests	hadoop.hdfs.server.blockmanagement.TestRBWBlockInvalidation
	hadoop.hdfs.server.namenode.TestStartup

Subsystem	Report/Notes
Docker	Image:yetus/hadoop:d946387
JIRA Issue	HDFS-12278
JIRA Patch URL	https://issues.apache.org/jira/secure/attachment/12881080/HDFS-12278-branch-2.8.001.patch
Optional Tests	asflicense compile javac javadoc mvninstall mvnsite unit findbugs checkstyle
uname	Linux 060b17dc6b37 3.13.0-123-generic #172-Ubuntu SMP Mon Jun 26 18:04:35 UTC 2017 x86_64 x86_64 GNU/Linux
Build tool	maven
Personality	/testptch/hadoop/patchprocess/precommit/personality/provided.sh
git revision	branch-2.8 / 639380e
Default Java	1.7.0_131
Multi-JDK versions	/usr/lib/jvm/java-8-oracle:1.8.0_144 /usr/lib/jvm/java-7-openjdk-amd64:1.7.0_131
findbugs	v3.0.0
JDK v1.7.0_131 Test Results	https://builds.apache.org/job/PreCommit-HDFS-Build/20623/testReport/
modules	C: hadoop-hdfs-project/hadoop-hdfs U: hadoop-hdfs-project/hadoop-hdfs
Console output	https://builds.apache.org/job/PreCommit-HDFS-Build/20623/console
Powered by	Apache Yetus 0.6.0-SNAPSHOT http://yetus.apache.org


This message was automatically generated.

▼ People

Assignee:

 Rushabh Shah

Reporter:

 Rushabh Shah

Votes:

 0 Vote for this issue

Watchers:

 14 Start watching this issue

▼ Dates

Created:

08/Aug/17 20:44

Updated:

09/Aug/17 23:11

4/19/23, 3:37 PM

[HDFS-12278] LeaseManager operations are inefficient in 2.8. - ASF JIRA

Resolved:

09/Aug/17 22:03