



Log block and datanode details in BlockRecoveryWorker

▼ Details

Type:	↑ Improvement	Status:	RESOLVED
Priority:	⚠ Major	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	2.9.0, 2.8.3, 3.0.0
Component/s:	datanode		
Labels:	None		
Target Version/s:	2.8.3		
Hadoop Flags:	Reviewed		

▼ Description

In a recent investigation, we have seen a weird block recovery issue, which is difficult to reach to a conclusion because of insufficient logs.

For the most critical part of the events, we see block recovery failed to `[[commitBlockSynchronization]]` on the NN, due to the block not closed. This leaves the file as open forever (for 1+ months).

The reason the block was not closed on NN, was because it is configured with `dfs.namenode.replication.min = 2`, and only 1 replica was with the latest genstamp.

We were not able to tell why only 1 replica is on latest genstamp.

From the primary node of the recovery (ps2204), `initReplicaRecoveryImpl` was called on each of the 7 DNs the block were ever placed. All DNs but ps2204 and ps3765 failed because of genstamp comparison - that's expected. ps2204 and ps3765 have gone past the comparison (since no exceptions from their logs), but `updateReplicaUnderRecovery` only appeared to be called on ps3765.

This jira is to propose we log more details when `BlockRecoveryWorker` is about to call `updateReplicaUnderRecovery` on the `DataNodes`, so this could be figured out in the future.

```
$ grep "updateReplica:" ps2204.dn.log
$ grep "updateReplica:" ps3765.dn.log
hadoop-hdfs-datanode-ps3765.log.2:{"@timestamp":"2017-09-13T00:56:20.933Z","source_host":"ps3765.example.com","file":"FsDataset:
$ grep "initReplicaRecovery:" ps2204.dn.log
hadoop-hdfs-datanode-ps2204.log.1:{"@timestamp":"2017-09-13T00:56:20.691Z","source_host":"ps2204.example.com","file":"FsDataset:
hadoop-hdfs-datanode-ps2204.log.1:{"@timestamp":"2017-09-13T00:56:20.691Z","source_host":"ps2204.example.com","file":"FsDataset:
$ grep "initReplicaRecovery:" ps3765.dn.log
hadoop-hdfs-datanode-ps3765.log.2:{"@timestamp":"2017-09-13T00:56:20.457Z","source_host":"ps3765.example.com","file":"FsDataset:
hadoop-hdfs-datanode-ps3765.log.2:{"@timestamp":"2017-09-13T00:56:20.457Z","source_host":"ps3765.example.com","file":"FsDataset:
hadoop-hdfs-datanode-ps3765.log.2:{"@timestamp":"2017-09-13T00:56:20.457Z","source_host":"ps3765.example.com","file":"FsDataset:
```

P.S. [HDFS-11499](#) was once suspected, but non-conclusive since we don't have all NN logs to know about the decommission.

▼ Attachments

HDFS-12642.01.patch	2 kB	12/Oct/17 05:16
HDFS-12642.02.patch	2 kB	15/Oct/17 02:47

▼ Activity

▼ [Xiao Chen](#) added a comment - 12/Oct/17 05:29

Patch 1 to add the logs. The added overhead / line counts of logging should be ignorable because recovery does not happen as often as other operations on the DN.



Here is an example of what it looks like from a unit test run:

```
2017-10-11 14:18:00,803 INFO datanode.DataNode (BlockRecoveryWorker.java:logRecoverBlock(324)) -
BlockRecoveryWorker:NameNode at localhost/127.0.0.1:54426
calls recoverBlock(BP-50205560-10.0.0.51-1507756664863:blk_1073741825_1007,
targets=[DatanodeInfoWithStorage[127.0.0.1:54486,null,null],
```

```
DatanodeInfoWithStorage[127.0.0.1:54442,null,null],
DatanodeInfoWithStorage[127.0.0.1:54433,null,null]],
newGenerationStamp=1008)

2017-10-11 14:18:01,149 INFO datanode.DataNode (BlockRecoveryWorker.java:syncBlock(184)) -
BlockRecoveryWorker: block=BP-50205560-10.0.0.51-1507756664863:blk_1073741825_1007, (length=100),
syncList=[block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54486,null,null],
block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54442,null,null],
block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54433,null,null]]

2017-10-11 14:18:01,150 INFO datanode.DataNode (BlockRecoveryWorker.java:syncBlock(271)) -
BlockRecoveryWorker: block=BP-50205560-10.0.0.51-1507756664863:blk_1073741825_1007, (length=100),
participatingList=[block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54486,null,null],
block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54442,null,null],
block:blk_1073741825_1007[numBytes=100,originalReplicaState=RBW] node:DatanodeInfoWithStorage[127.0.0.1:54433,null,null]]
```

  Hadoop QA added a comment - 13/Oct/17 05:15

 -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	13m 35s	trunk passed
+1	compile	0m 57s	trunk passed
+1	checkstyle	0m 41s	trunk passed
+1	mvnsite	1m 6s	trunk passed
+1	shadedclient	10m 22s	branch has no errors when building and testing our client artifacts.
+1	findbugs	1m 50s	trunk passed
+1	javadoc	0m 54s	trunk passed
			Patch Compile Tests
+1	mvninstall	0m 59s	the patch passed
+1	compile	0m 58s	the patch passed
+1	javac	0m 58s	the patch passed
+1	checkstyle	0m 38s	the patch passed
+1	mvnsite	1m 2s	the patch passed
+1	whitespace	0m 0s	The patch has no whitespace issues.
+1	shadedclient	9m 43s	patch has no errors when building and testing our client artifacts.
+1	findbugs	2m 2s	the patch passed
+1	javadoc	0m 49s	the patch passed
			Other Tests
-1	unit	97m 2s	hadoop-hdfs in the patch failed.
+1	asflicense	0m 23s	The patch does not generate ASF License warnings.
		142m 49s	

Reason	Tests
--------	-------

Failed junit tests	hadoop.hdfs.TestReadStripedFileWithMissingBlocks
--------------------	--

Subsystem	Report/Notes
Docker	Image:yetus/hadoop:3d04c00
JIRA Issue	HDFS-12642
JIRA Patch URL	https://issues.apache.org/jira/secure/attachment/12891634/HDFS-12642.01.patch
Optional Tests	asflicense compile javac javadoc mvninstall mvnsite unit shadedclient findbugs checkstyle
uname	Linux fd594502c92b 3.13.0-129-generic #178-Ubuntu SMP Fri Aug 11 12:48:20 UTC 2017 x86_64 x86_64 x86_64 GNU/Linux
Build tool	maven
Personality	/testptch/hadoop/patchprocess/precommit/personality/provided.sh
git revision	trunk / e46d5bb
Default Java	1.8.0_144
findbugs	v3.1.0-RC1
unit	https://builds.apache.org/job/PreCommit-HDFS-Build/21677/artifact/patchprocess/patch-unit-hadoop-hdfs-project_hadoop-hdfs.txt
Test Results	https://builds.apache.org/job/PreCommit-HDFS-Build/21677/testReport/
modules	C: hadoop-hdfs-project/hadoop-hdfs U: hadoop-hdfs-project/hadoop-hdfs
Console output	https://builds.apache.org/job/PreCommit-HDFS-Build/21677/console
Powered by	Apache Yetus 0.6.0-SNAPSHOT http://yetus.apache.org

This message was automatically generated.

▼  **Yongjun Zhang** added a comment - 14/Oct/17 08:31

Thanks [xiaochen](#) for working on this.

Some comments:

1. Include isTruncateRecovery and blockId in the following message

```
LOG.info("BlockRecoveryWorker: block=" + block + ", (length=" + block
    .getNumBytes() + "), syncList=" + syncList);
```

2. Include bestState and newBlock info in the following message:


```
LOG.info("BlockRecoveryWorker: block=" + block + ", (length=" + block
    .getNumBytes() + "), participatingList=" + participatingList);
```

For example

```
LOG.info("BlockRecoveryWorker: block=" + block + ", (length=" + block
    .getNumBytes() + "), bestState=" + bestState.name() +
    ", newBlock=" + newBlock + ", (length=" +
    newBlock.getNumBytes() + "), participatingList=" + participatingList);
```

BTW, I found using TestLeaseRecovery code can help observing the output.

Thanks.

▼  **Xiao Chen** added a comment - 15/Oct/17 02:47

Thanks Yongjun for the review.



Patch 2 addressed the comments. New example logs for convenience:

```
2017-10-14 19:43:25,345 [org.apache.hadoop.hdfs.server.datanode.BlockRecoveryWorker$1@40bdcea2]
INFO datanode.DataNode (BlockRecoveryWorker.java:logRecoverBlock(549)) -
BlockRecoveryWorker: NameNode at localhost/127.0.0.1:52885 calls recoverBlock(BP-342842897-10.0.0.51-1508035399504:blk_107
```

```
targets={DatanodeInfoWithStorage[127.0.0.1:52896,null,null], DatanodeInfoWithStorage[127.0.0.1:52901,null,null], DatanodeInfoWithStorage[127.0.0.1:52888,null,null]}, newGenerationStamp=1004, newBlock=null, isStriped=false)
```

```
2017-10-14 19:43:25,368 [org.apache.hadoop.hdfs.server.datanode.BlockRecoveryWorker$1@40bdcea2] INFO datanode.DataNode (BlockRecoveryWorker: block=BP-342842897-10.0.0.51-1508035399504:blk_1073741827_1003 (length=952), isTruncateRecovery=false) syncList=[block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52896,null,null], block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52901,null,null], block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52888,null,null]]
```

```
2017-10-14 19:43:25,369 [org.apache.hadoop.hdfs.server.datanode.BlockRecoveryWorker$1@40bdcea2] INFO datanode.DataNode (BlockRecoveryWorker: block=BP-342842897-10.0.0.51-1508035399504:blk_1073741827_1003 (length=952), bestState=FINALIZED, newBlock=BP-342842897-10.0.0.51-1508035399504:blk_1073741827_1004 (length=952), participatingList=[block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52896,null,null], block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52901,null,null], block:blk_1073741827_1003[numBytes=952,originalReplicaState=FINALIZED] node:DatanodeInfoWithStorage[127.0.0.1:52888,null,null]])
```

  Hadoop QA added a comment - 15/Oct/17 05:24

 -1 overall

Vote	Subsystem	Runtime	Comment
0	reexec	0m 15s	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	13m 49s	trunk passed
+1	compile	0m 51s	trunk passed
+1	checkstyle	0m 38s	trunk passed
+1	mvnsite	0m 57s	trunk passed
+1	shadedclient	10m 26s	branch has no errors when building and testing our client artifacts.
+1	findbugs	1m 50s	trunk passed
+1	javadoc	0m 46s	trunk passed
Patch Compile Tests			
+1	mvninstall	0m 51s	the patch passed
+1	compile	0m 48s	the patch passed
+1	javac	0m 48s	the patch passed
+1	checkstyle	0m 35s	the patch passed
+1	mvnsite	0m 53s	the patch passed
+1	whitespace	0m 0s	The patch has no whitespace issues.
+1	shadedclient	9m 53s	patch has no errors when building and testing our client artifacts.
+1	findbugs	1m 54s	the patch passed
+1	javadoc	0m 45s	the patch passed
Other Tests			
-1	unit	108m 31s	hadoop-hdfs in the patch failed.
+1	asflicense	0m 19s	The patch does not generate ASF License warnings.
		153m 35s	

Reason	Tests
--------	-------

Failed junit tests	hadoop.hdfs.server.namenode.ha.TestPipelinesFailover
--------------------	--

Subsystem	Report/Notes
Docker	Image:yetus/hadoop:0de40f0
JIRA Issue	HDFS-12642
JIRA Patch URL	https://issues.apache.org/jira/secure/attachment/12892263/HDFS-12642.02.patch
Optional Tests	asflicense compile javac javadoc mvninstall mvnsite unit shadedclient findbugs checkstyle
uname	Linux 16dccb4d8b5a 3.13.0-123-generic #172-Ubuntu SMP Mon Jun 26 18:04:35 UTC 2017 x86_64 x86_64 x86_64 GNU/Linux
Build tool	maven
Personality	/testptch/hadoop/patchprocess/precommit/personality/provided.sh
git revision	trunk / 87ea1df
Default Java	1.8.0_144
findbugs	v3.1.0-RC1
unit	https://builds.apache.org/job/PreCommit-HDFS-Build/21703/artifact/patchprocess/patch-unit-hadoop-hdfs-project_hadoop-hdfs.txt
Test Results	https://builds.apache.org/job/PreCommit-HDFS-Build/21703/testReport/
modules	C: hadoop-hdfs-project/hadoop-hdfs U: hadoop-hdfs-project/hadoop-hdfs
Console output	https://builds.apache.org/job/PreCommit-HDFS-Build/21703/console
Powered by	Apache Yetus 0.6.0-SNAPSHOT http://yetus.apache.org

This message was automatically generated.

Yongjun Zhang added a comment - 15/Oct/17 05:29

Hi xiaochen,

Thanks for the updated patch.

Sorry I did not make it clear earlier, I meant to add both isTruncateRecovery and blockId in my previous comment 1, because

```
long blockId = (isTruncateRecovery) ?
    rBlock.getNewBlock().getBlockId() : block.getBlockId();
```

I suggest we make the message:

```
200 LOG.info("BlockRecoveryWorker: block={ } (length={}),"
201         + " isTruncateRecovery={ }, blockId={ } syncList={ }", block,
202         block.getNumBytes(), isTruncateRecovery, blockId, syncList);
```

Thanks.

Xiao Chen added a comment - 16/Oct/17 02:15

Thanks yzhangal for the additional review.

Sorry I missed the blockId part in your initial comment. But when updating patch 2 I looked at blockId too. 😊

Reason I didn't add it is, blockId is used at 2 places:

- 1. construct newBlock
- 2. updateReplicaUnderRecovery


Since we always log newBlock before iterating through participatingList, blockId can be known from the logged newBlock info logs.

Yongjun Zhang added a comment - 16/Oct/17 04:53 - edited

Makes sense xiaochen.


I found the failed test was reported as [HDFS-12279](#).

+1 on rev2.

▼  [Xiao Chen](#) added a comment - 16/Oct/17 17:32

Thanks Yongjun. Failed test is not related to the change. No tests needed since the change is log-only.

Committing this.

▼  [Xiao Chen](#) added a comment - 16/Oct/17 17:35

Committed to trunk, branch-3.0, branch-2, branch-2.8.

Thanks for the reviews [yzhangal](#).

▼  [Hudson](#) added a comment - 16/Oct/17 18:30

SUCCESS: Integrated in Jenkins build Hadoop-trunk-Commit #13090 (See <https://builds.apache.org/job/Hadoop-trunk-Commit/13090/>)

[HDFS-12642](#). Log block and datanode details in BlockRecoveryWorker. (xiao: rev 21bc85558718490e558c5b3bdb44c9c64eada994)

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

▼ **People**

Assignee:



[Xiao Chen](#)

Reporter:



[Xiao Chen](#)

Votes:



Vote for this issue

Watchers:



Start watching this issue

▼ **Dates**

Created:

12/Oct/17 05:16

Updated:

16/Oct/17 23:54

Resolved:

16/Oct/17 17:35