



do not check replication if the block is under construction

Details

Type:	Bug	Status:	CLOSED
Priority:	Critical	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	2.8.0, 2.7.3, 2.6.4, 3.0.0-alpha1
Component/s:	None		
Labels:	None		
Target Version/s:	2.8.0, 2.7.3, 2.6.4		
Hadoop Flags:	Reviewed		

Description

When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.

If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.

Furthermore, after closing the appended file, in FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication. If some of nodes are decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...

Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.

Attachments

HDFS-9600-branch-2.6.patch	6 kB	30/Dec/15 09:11
HDFS-9600-branch-2.7.patch	7 kB	30/Dec/15 09:11
HDFS-9600-branch-2.patch	7 kB	30/Dec/15 09:11
HDFS-9600-v1.patch	0.9 kB	25/Dec/15 07:09
HDFS-9600-v2.patch	5 kB	28/Dec/15 11:09
HDFS-9600-v3.patch	7 kB	29/Dec/15 05:33
HDFS-9600-v4.patch	7 kB	30/Dec/15 03:42

Issue Links

is related to

HDFS-10810 Setreplication removing block from underconstrcution temporarily when batch IBR is enabled.	RESOLVED
HDFS-9586 listCorruptFileBlocks should not output files that all replications are decommissioning	PATCH AVAIL...




Activity



[Phil Yang](#) created issue - 25/Dec/15 06:58

[Phil Yang](#) made changes - 25/Dec/15 07:09

Field	Original Value	New Value
Attachment		HDFS-9600-v1.patch [12779493]

Field	Original Value	New Value
 Phil Yang made changes - 25/Dec/15 07:09		
Status	Open [1]	Patch Available [10002]
 Phil Yang made changes - 25/Dec/15 07:17		
Description	<p>When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.</p> <p>If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high and actually they are not missing.</p> <p>Furthermore, after closing the appended file, in FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication and because of some of nodes is decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...</p> <p>Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.</p>	<p>When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.</p> <p>If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.</p> <p>Furthermore, after closing the appended file, in FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication and because of some of nodes is decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...</p> <p>Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.</p>
 Phil Yang made changes - 25/Dec/15 07:17		
Description	<p>When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.</p> <p>If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.</p> <p>Furthermore, after closing the appended file, in</p>	<p>When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.</p> <p>If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.</p> <p>Furthermore, after closing the appended file, in</p>

FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication and because of some of nodes is decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...

Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.

FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication. If some of nodes is decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...

Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.

 Phil Yang made changes - 25/Dec/15 07:18

Description

When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.

If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.

Furthermore, after closing the appended file, in FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication. If some of nodes is decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...

Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.

When appending a file, we will update pipeline to bump a new GS and the old GS will be considered as out of date. When changing GS, in BlockInfo.setGenerationStampAndVerifyReplicas we will remove replicas having old GS which means we will remove all replicas because no DN has new GS until the block with new GS is added to blockMaps again by DatanodeProtocol.blockReceivedAndDeleted.

If we check replication of this block before it is added back, it will be regarded as missing. The probability is low but if there are decommissioning nodes the DecommissionManager.Monitor will scan all blocks belongs to decommissioning nodes with a very fast speed so the probability of finding missing block is very high but actually they are not missing.

Furthermore, after closing the appended file, in FSNamesystem.finalizeNodeFileUnderConstruction, it will checkReplication. If some of nodes are decommissioning, this block with new GS will be added to UnderReplicatedBlocks map so there are two blocks with same ID in this map, one is in QUEUE_WITH_CORRUPT_BLOCKS and the other is in QUEUE_HIGHEST_PRIORITY or QUEUE_UNDER_REPLICATED. And there will be many missing blocks warning in NameNode website but there is no corrupt files...

Therefore, I think the solution is we should not check replication if the block is under construction. We only check complete blocks.

 Phil Yang made changes - 25/Dec/15 07:38

Link

This issue is related to [HDFS-9586](#) [[HDFS-9586](#)]

 Phil Yang made changes - 28/Dec/15 11:09

Attachment

[HDFS-9600-v2.patch](#) [12779638]

 Phil Yang made changes - 29/Dec/15 05:13

Attachment


[HDFS-9600-v3.patch](#) [12779753]

 Phil Yang made changes - 29/Dec/15 05:16


Attachment		HDFS-9600-v3.patch [12779753]
<hr/>		
 Phil Yang made changes - 29/Dec/15 05:33		
Attachment		HDFS-9600-v3.patch [12779756]
<hr/>		
 Phil Yang made changes - 30/Dec/15 03:42		
Attachment		HDFS-9600-v4.patch [12779897]
<hr/>		
 Phil Yang made changes - 30/Dec/15 09:11		
Attachment		HDFS-9600-branch-2.patch [12779932]
Attachment		HDFS-9600-branch-2.7.patch [12779933]
Attachment		HDFS-9600-branch-2.6.patch [12779934]
<hr/>		
 Vinayakumar B made changes - 07/Jan/16 06:30		
Fix Version/s		2.7.3 [12333995]
Fix Version/s		2.6.4 [12334241]
Hadoop Flags		Reviewed [10343]
Resolution		Fixed [1]
Status		Resolved [5]
Patch Available [10002]		
<hr/>		
 Vinayakumar B made changes - 07/Jan/16 06:43		
Fix Version/s		2.8.0 [12329057]
<hr/>		
 Vinod Kumar Vavilapalli made changes - 25/Aug/16 23:00		
Status		Closed [6]
Resolved [5]		
<hr/>		
 Andrew Wang made changes - 30/Aug/16 01:21		
Fix Version/s		3.0.0-alpha1 [12335732]
<hr/>		
 Mingliang Liu made changes - 13/Sep/16 21:08		
Link		This issue is related to HDFS-10810 [HDFS-10810]

▼ People

Assignee:

 Phil Yang

Reporter:

 Phil Yang

Votes:

0

Vote for this issue

Watchers:

14

Start watching this issue

▼ Dates

Created:

25/Dec/15 06:58

Updated:

13/Sep/16 21:08

Resolved:

07/Jan/16 06:30