

TAJO-1166

hyunsik

2014-11-12

Contents

Chapter 1

Root issue TAJO-1166

1.1 Summary

S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT

1.2 Description

```
1 mvn clean install -DskipTests -Dhadoop.version=2.6.0-SNAPSHOT
```

The above command will cause the following error:

```
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: cannot find symbol
[ERROR]     symbol: class S3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S3Integ
error: cannot find symbol
[ERROR]     symbol:   class FileSystemStore
        location: class SmallBlockS3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/Integ
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: cannot find symbol
[ERROR]     symbol: class FileSystemStore
```

```

/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class InMemoryFileSystemStore
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class InMemoryFileSystemStore
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class InMemoryFileSystemStore
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class FileSystemStore
    location: class Holder
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/TestI
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: package org.apache.hadoop.fs.s3 does not exist
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: cannot find symbol
[ERROR]   symbol:   class FileSystemStore
    location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class Block
    location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class FileSystemStore
    location: class S3OutputStream

```

```

/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: cannot find symbol
[ERROR]   symbol:   method setConf(Configuration)
      location: class SmallBlockS3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: no suitable method found for makeQualified(SmallBlockS3FileSystem)
[ERROR]       method Path.makeQualified(URI,Path) is not applicable
          (actual and formal argument lists differ in length)
      method Path.makeQualified(FileSystem) is not applicable
          (actual argument SmallBlockS3FileSystem cannot be converted to FileSystem
by method invocation conversion)
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: no suitable method found for makeQualified(SmallBlockS3FileSystem)
[ERROR]       method Path.makeQualified(URI,Path) is not applicable
          (actual and formal argument lists differ in length)
      method Path.makeQualified(FileSystem) is not applicable
          (actual argument SmallBlockS3FileSystem cannot be converted to FileSystem
by method invocation conversion)
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: no suitable method found for makeQualified(SmallBlockS3FileSystem)
[ERROR]       method Path.makeQualified(URI,Path) is not applicable
          (actual and formal argument lists differ in length)
      method Path.makeQualified(FileSystem) is not applicable
          (actual argument SmallBlockS3FileSystem cannot be converted to FileSystem
by method invocation conversion)
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: method mkdirs in class SmallBlockS3FileSystem cannot be applied to
given types;
[ERROR]   required: Path,FsPermission
      found: Path
      reason: actual and formal argument lists differ in length
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   method getConf()
      location: class SmallBlockS3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   variable statistics
      location: class SmallBlockS3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor
error: method does not override or implement a method from a supertype
[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/stor

```



```

[ERROR] /Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/IT
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class INode
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/TestI
error: cannot find symbol
[ERROR]   symbol:   class S3FileSystem
        location: class TestFileSystems
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/TestI
error: incompatible types
[ERROR]   required: Object
        found:     SmallBlockS3FileSystem
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol
[ERROR]   symbol:   class Block
        location: class S3OutputStream
/Users/hyunsik/Code/tajo/tajo-9/tajo-storage/src/test/java/org/apache/tajo/storage/s3/S
error: cannot find symbol

```

1.3 Attachments

No attachments

1.4 Commits

1. Commit **3d485ec** by **JaeHwa Jung** (2014-11-27): TAJO-1166: S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT. (jaehwa)

Closes #262

1.5 Comments

1. **jhkim**: It seems that the s3 packages moved to sub-level HADOOP-11074, HADOOP-10400

I found the netflix contribution. it looks good to me
<http://www.slideshare.net/zhenxiao/presto-atnetflix>
<http://techblog.netflix.com/2014/10/using-presto-in-our-big-data-platform.html>

2. **hyunsik:** +1 for Jinho's idea.

In Hadoop 2.6.0-SNAPSHOT, S3 related classes seem to be moved to a separate maven module, named hadoop-aws. It may be hard for Tajo to support multiple Hadoop versions. Actually, we can solve this problem if we use both maven profile ({{-P}}) and maven {{-D}} option. But, instead, as Jinho said, it would be great to have our own S3 implementation based on some other implementation. Netflix contributions look promising.

3. **blrunner:** Hi guys.

Thanks for your ideas and I also agree with yours. And I found two problems fixing this issue.

The first is aws dependency. As Hyunsik said, S3 module separated to hadoop-aws. The second is yarn containerId primitive change. In hadoop-2.6.0, containerId primitive type changed from int to long.

Thus, this issue need to separate S3 dependency improvement and hadoop-2.6.0 bump up. For the first, I'll improve S3 dependency with netflix contribution sooner or later.

For reference, I uploaded small patch to my private repository as follows.

- <https://github.com/blrunner/tajo/tree/TAJO-1166>

You can build it as follows:

```
1 mvn clean install -Dhadoop.version=2.6.0-SNAPSHOT
```

4. **blrunner:** I've just removed aws-sdk dependencies. And I replaced S3FileSystem to LocalFileSystem for S3 unit test case in temporary.

5. **blrunner:** [~jhkim], thanks for your quick review.

I've just committed it to master branch.

6. **hudson:** SUCCESS: Integrated in Tajo-master-build #464 (See [<https://builds.apache.org/job/Tajo-master-build/464/>])

TAJO-1166: S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT. (jaehwa) (blrunner: rev 3d485ecb0112af12258a5a2bdc4a400b8df4fae8)

- * tajo-core/src/main/java/org/apache/tajo/master/DefaultTaskScheduler.java
- * tajo-core/src/main/java/org/apache/tajo/master/ContainerProxy.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/LocalTaskEvent.java
- * tajo-core/src/main/java/org/apache/tajo/worker/TaskRunnerHistory.java
- * CHANGES

- * tajo-core/pom.xml
- * tajo-storage/src/test/java/org/apache/tajo/storage/s3/INode.java
- * tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerContainerId.java
- * tajo-storage/src/test/java/org/apache/tajo/storage/s3/S3OutputStream.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryUnitAttempt.java
- * tajo-core/src/main/java/org/apache/tajo/worker/TajoResourceAllocator.java

- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactoryPBIImpl.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/impl/pb/TajoContainerIdPBIImpl.java
- * tajo-storage/src/test/java/org/apache/tajo/storage/s3/InMemoryFileSystemStore.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoContainer.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/TaskAttemptAssignedEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/rm/TajoRMContext.java
- * tajo-core/src/main/java/org/apache/tajo/master/rm/WorkerResourceManager.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryMasterManagerService.java
- * tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerResourceManager.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/QueryUnitAttemptScheduleEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/TaskRequestEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/LazyTaskScheduler.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryInProgress.java
- * tajo-core/src/test/java/org/apache/tajo/master/rm/TestTajoResourceManager.java
- * tajo-core/src/main/proto/ResourceTrackerProtocol.proto
- * tajo-core/src/main/java/org/apache/tajo/master/event/SubQueryContainerAllocationEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerContainer.java
- * tajo-core/src/main/java/org/apache/tajo/master/TajoContainerProxy.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecords.java
- * tajo-core/src/main/java/org/apache/tajo/worker/AbstractResourceAllocator.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoContainerId.java
- * tajo-core/src/main/proto/TajoWorkerProtocol.proto
- * tajo-core/src/main/java/org/apache/tajo/worker/TaskRunner.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/SubQuery.java
- * tajo-storage/src/test/java/org/apache/tajo/storage/s3/SmallBlockS3FileSystem.java
- * tajo-core/src/main/proto/TajoMasterProtocol.proto
- * tajo-core/src/main/java/org/apache/tajo/worker/ResourceAllocator.java
- * tajo-storage/src/main/java/org/apache/tajo/storage/StorageManager.java
- * tajo-core/src/main/proto/QueryMasterProtocol.proto
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactory.java
- * tajo-core/src/main/java/org/apache/tajo/master/LaunchTaskRunnersEvent.java
- * tajo-storage/src/test/java/org/apache/tajo/storage/TestFileSystems.java
- * tajo-core/src/main/proto/ContainerProtocol.proto
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoConverterUtils.java
- * tajo-core/src/main/java/org/apache/tajo/master/TaskRunnerGroupEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/TajoMasterService.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactoryProvider.java

7. **hudson:** FAILURE: Integrated in Tajo-master-CODEGEN-build #106 (See [<https://builds.apache.org/job/Tajo-master-CODEGEN-build/106/>])

TAJO-1166: S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT. (jaehwa) (blrunner: rev 3d485ecb0112af12258a5a2bdc4a400b8df4fae8)

- * tajo-core/pom.xml
- * tajo-core/src/main/java/org/apache/tajo/worker/TajoResourceAllocator.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/SubQueryContainerAllocationEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/TaskRunnerGroupEvent.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoContainer.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryInProgress.java
- * tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryUnitAttempt.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactory.java
- * tajo-core/src/main/java/org/apache/tajo/master/event/QueryUnitAttemptScheduleEvent.java
- * tajo-core/src/main/java/org/apache/tajo/worker/AbstractResourceAllocator.java
- * tajo-core/src/main/java/org/apache/tajo/master/container/TajoContainerId.java

```

* tajo-core/src/main/java/org/apache/tajo/master/container/impl/pb/TajoContainerIdPBImp.java
* tajo-storage/src/test/java/org/apache/tajo/storage/s3/S3OutputStream.java
* tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerContainerId.java
* tajo-core/src/main/proto/ContainerProtocol.proto
* tajo-storage/src/test/java/org/apache/tajo/storage/s3/InMemoryFileSystemStore.java
* tajo-core/src/main/proto/TajoWorkerProtocol.proto
* tajo-core/src/main/java/org/apache/tajo/master/rm/TajoRMContext.java
* tajo-core/src/main/java/org/apache/tajo/master/ContainerProxy.java
* tajo-core/src/main/proto/ResourceTrackerProtocol.proto
* CHANGES
* tajo-core/src/main/java/org/apache/tajo/master/TajoContainerProxy.java
* tajo-core/src/main/java/org/apache/tajo/master/event/TaskRequestEvent.java
* tajo-core/src/main/java/org/apache/tajo/master/rm/WorkerResourceManager.java
* tajo-storage/src/test/java/org/apache/tajo/storage/s3/INode.java
* tajo-core/src/main/java/org/apache/tajo/master/container/TajoConverterUtils.java
* tajo-storage/src/main/java/org/apache/tajo/storage/StorageManager.java
* tajo-core/src/main/proto/TajoMasterProtocol.proto
* tajo-core/src/main/java/org/apache/tajo/master/querymaster/SubQuery.java
* tajo-core/src/main/java/org/apache/tajo/worker/TaskRunner.java
* tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactoryPBImp.java
* tajo-core/src/main/java/org/apache/tajo/worker/TaskRunnerHistory.java
* tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecords.java
* tajo-core/src/main/java/org/apache/tajo/master/LazyTaskScheduler.java
* tajo-core/src/main/java/org/apache/tajo/master/querymaster/QueryMasterManagerService.java
* tajo-core/src/test/java/org/apache/tajo/master/rm/TestTajoResourceManager.java
* tajo-core/src/main/java/org/apache/tajo/worker/ResourceAllocator.java
* tajo-core/src/main/java/org/apache/tajo/master/DefaultTaskScheduler.java
* tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerResourceManager.java
* tajo-core/src/main/java/org/apache/tajo/master/TajoMasterService.java
* tajo-core/src/main/java/org/apache/tajo/master/event/LocalTaskEvent.java
* tajo-core/src/main/java/org/apache/tajo/master/event/TaskAttemptAssignedEvent.java
* tajo-core/src/main/proto/QueryMasterProtocol.proto
* tajo-storage/src/test/java/org/apache/tajo/storage/s3/SmallBlockS3FileSystem.java
* tajo-core/src/main/java/org/apache/tajo/master/container/TajoRecordFactoryProvider.java
* tajo-core/src/main/java/org/apache/tajo/master/rm/TajoWorkerContainer.java
* tajo-core/src/main/java/org/apache/tajo/master/LaunchTaskRunnersEvent.java
* tajo-storage/src/test/java/org/apache/tajo/storage/TestFileSystems.java

```

1.6 Pull requests

1.6.1 Pull request 262

Title: TAJO-1166: S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT

Author: blrunner

Date: 2014-11-19

Status: closed

Comments:

1. **blrunner** (2014-11-20): As I mentioned, I replaced S3FileSystem to LocalFileSystem for unit test. But in this case, we can't test split size because of result data size is less than HDFS default block size. So, I added "fs.local.block.size" to storage-site.xml for unit test. I found that this branch build had succeed on hadoop-2.5.1 and hadoop-2.6.0-SNAPSHOT.

2. **blrunner** (2014-11-20): For reference, I also tested some TPC-H queries on a real cluster. In addition, all test has passed successfully.
3. **jinossy** (2014-11-27): @blrunner
Could you add comment to record that we borrow this class from YARN? It would be useful for other guys to maintain the code in the future.
4. **blrunner** (2014-11-27): Hi @jinossy

Thanks for your review. I added a comment to indicate the source for yarn and I fixed a few document bugs.
5. **jinossy** (2014-11-27): +1 Looks good to me.
Ship it!

1.6.2 Pull request 260

Title: TAJO-1166: S3 related storage causes compilation error in Hadoop 2.6.0-SNAPSHOT

Author: blrunner

Date: 2014-11-19

Status: closed

Comments:

1. **blrunner** (2014-11-19): I found that commit logs are mixed with another patch. I'll create another request.