

TAJO-287

hyunsik

2013-10-27

# Contents

# Chapter 1

## Root issue TAJO-287

### 1.1 Summary

Improve Fragment to be more generic

### 1.2 Description

The current Fragment is only for a file. This patch improves Fragment to be more generic. First of all, I've changed Fragment to an interface and the original Fragment to FileFragment respectively. FragmentProto is changed to contain a table name and a bytestring which contains an storage-dependent contents. Then, the added FragmentConvertor transforms FragmentProto to a specified Fragment instance. It would be very useful to represent various fragment types like a row range of Hbase and database tables.

### 1.3 Attachments

1. [TAJO-287\\_2.patch](#)
2. [TAJO-287\\_3.patch](#)
3. [TAJO-287.patch](#)

### 1.4 Commits

1. Commit **3c22d3e** by **Hyunsik Choi** (2013-10-28): TAJO-287: Improve Fragment to be more generic. (hyunsik)
2. Commit **7dc8de2** by **Hyunsik Choi** (2013-10-27): TAJO-287: Refactor TableDesc, TableMeta, and Fragment. (hyunsik)

### 1.5 Comments

1. **jihoonson:** ++1 for this issue. It has a great extensibility. I have a couple of things that we need to discuss.  
\* FileFragment has getStartKey() and getEndKey() which are inherited from Fragment. These functions return a start offset and a length, respectively. But, their names are not intuitive to suppose what their values mean. How about add getStartOffset() and getLength() function to only FileFragment?

\* In Task, there are some commented out codes and you left a comment as "TODO - to be fixed". What does the comment mean?

2. **hyunsik:** I will handle the TODO that I missed.

Because getStartOffset and getLength will be called in file scanners, your suggestion is reasonable. I expected that the generic start key and end key may be used in QueryMaster without the fragment specific code. However, I can't imagine any usage cases now. Now, I'm going to accept your suggestion.

Thank you for the nice review.

3. **hyunsik:** I've uploaded the second patch that reflects Jihoon's suggestion.

4. **jihoonson:** I'll look at the patch this afternoon.

5. **jihoonson:** +1.

This patch looks good to me.

It also passes 'mvn clean verify'.

6. **hyunsik:** Thanks for the review. I've just committed it to master.

7. **hudson:** SUCCESS: Integrated in Tajo-trunk-postcommit #541 (See [https://builds.apache.org/job/Tajo-trunk-postcommit/541/])

TAJO-287: Improve Fragment to be more generic. (hyunsik) (hyunsik: https://git-wip-

us.apache.org/repos/asf?p=incubator-tajo.git&a=commit&h=3c22d3eba437e9130ee74de8f935c7fc330b1217)

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/v2/CSVFileScanner.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestHashJoinExec.java

\* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/rcfile/TestRCFile.java

\* tajo-catalog/tajo-catalog-common/src/main/proto/CatalogProtos.proto

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/fragment/FileFragment.java

\* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/worker/TaskAttemptContext.java

\* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/v2/TestStorages.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/RawFile.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/RowFile.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestFullOuterMergeJoinExec.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/trevni/TrevniAppender.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestRightOuterHashJoinExec.java

\* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/TestCompressionStorages.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/storage/TestFragment.java

\* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/master/querymaster/QueryUnit.java

\* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/query/QueryUnitRequestImpl.java

\* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/v2/TestCSVScanner.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestPhysicalPlanner.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/fragment/Fragment.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestSortExec.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/fragment/FragmentConvertor.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestLeftOuterNLJoinExec.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/Storage.java

\* tajo-catalog/tajo-catalog-common/src/main/java/org/apache/tajo/catalog/json/TableMetaAdapter.java

\* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/query/QueryUnitRequest.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/worker/TestRangeRetrieverHandler.java

\* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/storage/TestRowFile.java

\* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/Fragment.java

- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/index/TestSingleCSVFileBSTIndex.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/planner/physical/BSTIndexScanExec.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/AbstractStorageManager.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/client/ResultSetImpl.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/storage/TestFileFragment.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/StorageManager.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestBSTIndexExec.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestHashSemiJoinExec.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestFullOuterHashJoinExec.java
- \* tajo-core/tajo-core-storage/src/main/resources/storage-default.xml
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/v2/StorageManagerV2.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestMergeJoinExec.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestRightOuterMergeJoinExec.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/planner/PhysicalPlannerImpl.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/StorageManagerFactory.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/master/event/SubQuerySucceedEvent.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestLeftOuterHashJoinExec.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/MergeScanner.java
- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/index/TestBSTIndex.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/master/querymaster/SubQuery.java
- \* CHANGES.txt
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/FileAppender.java
- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/v2/TestCSVCompression.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/worker/Task.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/v2/RCFileScanner.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/util/IndexUtil.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestNLJoinExec.java
- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/TestStorageManager.java
- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/TestStorages.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/planner/physical/ExternalSortExec.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/FileScanner.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestExternalSortExec.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/v2/FileScannerV2.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestBNLJoinExec.java
- \* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/planner/physical/TestHashAntiJoinExec.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/trevni/TrevniScanner.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/master/querymaster/Repartitioner.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/engine/planner/physical/SeqScanExec.java
- \* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/master/TaskSchedulerImpl.java
- \* tajo-core/tajo-core-storage/src/test/java/org/apache/tajo/storage/TestMergeScanner.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/rcfile/RCFileWrapper.java
- \* tajo-core/tajo-core-storage/src/main/java/org/apache/tajo/storage/CSVFile.java
- \* tajo-core/tajo-core-storage/src/test/resources/storage-default.xml

## 1.6 Pull requests

No pull requests