

git_comments:

1. De-dupe/merge if needed
2. partition using the insert partitioner
3. * * Inserts the given HoodieRecords, into the table. This API is intended to be used for normal * writes. *
* This implementation skips the index check & is able to leverage benefits such as * small file handling/blocking alignment, as with upsert(), by profiling the workload * * @param records HoodieRecords to insert * @param commitTime Commit Time handle * @return JavaRDD[WriteStatus] - RDD of WriteStatus to inspect errors and counts
4. * * Perform the ultimate IO for a given inserted (RDD) partition * * @param partition * @param recordIterator * @param partitioner
5. Lots of inserts such that file1 is updated and expanded, a new file2 is created.
6. Inserts => will write file1
7. Second, set of Inserts should just expand file1
8. setup the small file handling params hold upto 20 records max
9. this writes ~500kb

git_commits:

1. **summary:** Split insert() into insert() & bulkInsert() (#69)
message: Split insert() into insert() & bulkInsert() (#69) - Behaviour change for existing insert() users - Made the current insert() implementation, as something to use for bulkInsert() - Normal inserts now share a lot of code with upsert, which provides benefits like small file handling - Refactored/Cleaned up code in HoodieWriteClient for reuse - Added a unit test, switching few tests to call bulkInsert() and few to call insert()
label: code-design

github_issues:

1. **title:** Hoodie insert() should be implemented via insert partitioner
body: Right now, we dont have benefits like small insert handling due to this
2. **title:** Hoodie insert() should be implemented via insert partitioner
body: Right now, we dont have benefits like small insert handling due to this
label: code-design

github_issues_comments:

github_pulls:

1. **title:** Split insert() into insert() & bulkInsert()
body: - Behaviour change for existing insert() users - Made the current insert() implementation, as something to use for bulkInsert() - Normal inserts now share a lot of code with upsert, which provides benefits like small file handling - Refactored/Cleaned up code in HoodieWriteClient for reuse - Added a unit test, switching few tests to call bulkInsert() and few to call insert() Fixes #29
label: code-design

github_pulls_comments:

github_pulls_reviews:

jira_issues:

jira_issues_comments: