Item 267

## git\_comments:

- 1. Go to brute force mode of scanning all partitions
- Use Incremental Mode for finding partition paths. \* @param cleanMetadata \* @param newInstantToRetain \* @return
- 3. \*\* Scan and list all paritions for cleaning. \* @return all partitions paths for the dataset. \* @throws IOException
- 4. \* \* Return partition paths for cleaning by commits mode. \* @param instantToRetain Earliest Instant to retain \* @return list of partitions \* @throws IOException
- 5. We no longer write empty cleaner plans when there are not enough commits present
- 6. NOTE: See CleanPlanner#getFilesToCleanKeepingLatestCommits. We explicitly keep one commit before earliest commit

#### git\_commits:

1. summary: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode (#1576) message: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode (#1576)

#### github\_issues:

#### github\_issues\_comments:

#### github\_pulls:

1. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

2. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

3. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. 4. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

5. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. label: code-design

6. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

7. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. label: code-design

8. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

9. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

10. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

**body:** We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

11. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. . title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

13. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. 14. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. 15. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

16. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

17. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

18. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

19. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning.

20. title: [HUDI-850] Avoid unnecessary listings in incremental cleaning mode

body: We are unnecessarily scanning partitions for cleans\_by\_commit mode when there are no sufficient commits accumulated for cleaning. label: code-design

# github\_pulls\_comments:

- 1. Let me resolve the conflicts and ping back
- 2. # [Codecov](https://codecov.io/gh/apache/incubator-hudi/pull/1576?src=pr&el=h1) Report > Merging [#1576](https://codecov.io/gh/apache/incubator-hudi/pull/1576?src=pr&el=h1) Report > Merging [#1576](https://codecov.io/gh/apache/incubator-hudi/pull/1576](h hudi/pull/1576?src=pr&el=desc) into [master](https://codecov.io/gh/apache/incubator-hudi/commit/c4b71622b90fc66f20f361d4c083b0a396572b75&el=desc) will \*\*decrease\*\* coverage by `0.00%`. > The diff coverage is `80.00%`. [![Impacted file tree graph](https://codecov.io/gh/apache/incubator-

hudi/pull/1576/graphs/tree.svg?width=650&height=150&src=pr&token=VTTXabwbs2)] (https://codecov.io/gh/apache/incubator-hudi/pull/1576?src=pr&token=VTTXabwbs2)] (https://codecov.io/gh`diff @@ Coverage Diff @@ ## master #1576 +/- ## ================== ================== - Coverage 71.81% 71.81% -0.01% 

=========================== + Hits 11880 11885 +5 + Misses 3932 3931 -1 - Partials 730 733 +3 ``` | [Impacted Files]

[...g/apache/hudi/table/action/clean/CleanPlanner.java](https://codecov.io/gh/apache/incubator-hudi/pull/1576/diff?src=pr&el=tree#diff $a HV kaS1 jbGllbnQvc3 JjL21 haW4 vamF2YS9 vcmcvYXBhY2 hlL2 h1ZGkvdGFibGUvYWN0 aW9 uL2 NsZWFuL0 NsZWFuUGxhbm5 lci5 qYXZh) \mid `86.13\% h1ZGkvdGFibGUvYWN0 aW9 uL2 NsZWFuL0 NsZWFuUGxhbm5 lci5 qYXZh) \mid `86.13\% h1ZGkvdGFibGUvYWN0 aW9 uL2 NsZWFuL0 NsZWFuUGxhbm5 lci5 qYXZh) \mid `86.13\% h1ZGkvdGFibGUvYWN0 aW9 uL2 NsZWFuUGxhbm5 lci5 qYXZh) lci5 qYXZh uL2 NsZWFuUGxhbm5 lci5 qYXhbm5 lci5 qYXhbm5$ <80.00%> (-1.57%)` | `0.00 <0.00> (Ø)` | | | [...le/action/rollback/BaseRollbackActionExecutor.java](https://codecov.io/gh/apache/incubator-hudi/pull/1576/diff? src=pr&el=tree#diff-

aHVkaS1jbGllbnQvc3JjL21haW4vamF2YS9vcmcvYXBhY2hlL2h1ZGkvdGFibGUvYWN0aW9uL3JvbGxiYWNrL0Jhc2VSb2xsYmFja0FjdGlvbkV4ZWN1dG9yhdV4`70.83% <0.00%> (-6.95%)`|`0.00% <0.00%> (ø%)`|||[...n/java/org/apache/hudi/index/hbase/HBaseIndex.java](https://codecov.io/gh/apache/incubatorhudi/pull/1576/diff?src=pr&el=tree#diff-

aHVkaS1jbGllbnQvc3JjL21haW4vamF2YS9vcmcvYXBhY2hlL2h1ZGkvaW5kZXgvaGJhc2UvSEJhc2VJbmRleC5qYXZh) | `82.77% <0.00%> (-0.48%)` | `0.00% <0.00%> (ø%) | | | | | ...able/action/restore/BaseRestoreActionExecutor.java](https://codecov.io/gh/apache/incubator-hudi/pull/1576/diff? src=pr&el=tree#diff-

[`93.75% < 0.00% (+3.12%)` | `0.00% < 0.00% > (@%)` | | | [...src/main/java/org/apache/hudi/metrics/Metrics.java](https://codecov.io/gh/apache/incubator $hudi/pull/1576/diff?src=pr\&el=tree\#diff-aHVkaS1jbGllbnQvc3JjL21haW4vamF2YS9vcmcvYXBhY2hlL2h1ZGkvbWV0cmljcy9NZXRyaWNzLmphdmE=) | `67.56% < 0.00% > (+10.81%)` | `0.00% < 0.00% > (\varphi \)` | | [...g/apache/hudi/metrics/InMemoryMetricsReporter.java](https://codecov.io/gh/apache/incubator-hudi/pull/1576/diff?src=pr&el=tree#diff-$ 

aHVkaS1jbGllbnQvc3JjL21haW4vamF2YS9vcmcvYXBhY2hlL2h1ZGkvbWV0cmljcy9Jbk1lbW9yeU1ldHJpY3NSZXBvcnRlci5qYXZh) | `80.00% < 0.00% > (+40.00%)` | `0.00% < 0.00% > (0.00%) | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0.00% < 0.00% > (0.00%)` | `0

3. @vinothchandar: Ready for review

4. @vinothchandar: There is a followup on fixing the quality of unit-tests (filed HUDI-857). Otherwise, good to review and merge.

## github\_pulls\_reviews:

1. Collections.emptyList()?

2. **body:** can we break this method up into two, based on the policy? it feels very overloaded now..

label: code-design

3. **body:** do we have enough tests that now test when a valid cleaner plan?

label: test

- 4. Yes, this particular check is when inserting first batch of data. There are other checks present which validates the effect of valid clean operations
- 5. Done
- 6. Done
- 7. move to separate lines?
- 8. we don't need the second check anymore, due to the switch above?
- 9. rename: `getPartitionPathsForIncrementalCleaning()`?
- 10. rename: `getPartitionPathsForFullCleaning()`
- 11. Fixed
- 12. Done
- 13. done
- 14. done

#### jira\_issues:

1. summary: Avoid unnecessary listings in incremental cleaning mode

**description:** Came up during https://github.com/apache/incubator-hudi/issues/1552 Even with incremental cleaning turned on, we would have a scenario where there are no commits yet to clean, but we end up listing needlessly

2. summary: Avoid unnecessary listings in incremental cleaning mode

**description:** Came up during https://github.com/apache/incubator-hudi/issues/1552 Even with incremental cleaning turned on, we would have a scenario where there are no commits yet to clean, but we end up listing needlessly

label: code-design

### jira\_issues\_comments: