

- github_issues:**

- github_issues_comments:**

- github_pulls:**

- [illegible]

- [illegible]

summary.html The following link was added on 5/13 and goes over some use cases <https://gist.github.com/keith-turner/f6f7ca661d88b935b74f57dbdeeb07d3> The following link was added on 6/1 and documents a test I ran on a cluster with this code. <https://gist.github.com/keith-turner/29b72da97acad53c395dd0724b65f1c>
label: documentation

- [illegible]

66. **title:** Fixes #564 adds support multiple compaction executors

body: This change adds support for multiple compaction executors and multiple concurrent compactons per tablet. The best way to understand these changes is to look at the documentation at [core/src/main/java/org/apache/accumulo/core/spi/compaction/package-info.java](https://keith-turner.github.io/apidocs-accumulo-564/org/apache/accumulo/core/spi/compaction/package-info.java) The javadocs mentioned above are hard to read in source form and are rendered at the link below. <https://keith-turner.github.io/apidocs-accumulo-564/org/apache/accumulo/core/spi/compaction/package-summary.html> The following link was added on 5/13 and goes over some use cases <https://gist.github.com/keith-turner/f6f7ca661d88b935b74f57dbdeeb07d3> The following link was added on 6/1 and documents a test I ran on a cluster with this code. <https://gist.github.com/keith-turner/29b72dfa97acad53c395dd0724b65f1c>

github_pulls_comments:

- > Would there be value in marking the new properties/APIs as "experimental" until it bears out after a release or two? That would entail keeping all the old code and then jamming this new code into the old code. I don't think that would help achieve stability or maintainability. There are actually a lot of big changes that have built up for 2.1.0. I think we need to stop accepting new features for 2.1.0 at some point and start heavily testing it and only fixing bugs. > There are so many new classes. The number of new classes might make the compaction code a bit more logical, but it's a big shift from the current code, and very hard for a newcomer to quickly familiarize themselves with all the code. Did you look at the javadoc link I posted? The current plugins break down by data and execution. The CompactionSelector, CompactionConfigurer, and iterators are only concerned with user data and can be set per table or per compaction. The CompactionPlanner and CompactionDispatcher are only concerned with executing compactons. Planners are configured at the system level and dispatchers are configured per table. Users can set execution hints when initiating a compaction through the API and these hints are made available to the dispatcher and planner. If this is helpful, I can work this into the javadoc. > It might help to organize a video chat or slide show to go over all the code changes at a high level, so that other Accumulo devs can quickly assess the new design. I can do a chat on slack. One goal of that could be to determine how the written documentation could be improved. > I kind of liked the idea of setting a single compaction strategy property that encapsulates all the configuration This is very vague and I don't know what this would concretely look like. We could discuss this in the chat. If you have more specific ideas, it would help to write them up somewhere.
- body:** @ctubbsii something else that I eventually need to write is user facing documentation. The documentation written so far is a bit more internal. The old WIP PR #1589 does have a bit of user facing documentation in its opening comment that would be useful to look at though. It shows how to set configure different compaction services and make tables use them. The properties have changed a bit, but all of the concepts are the same.
label: documentation
- There have not been any comments on this in a while, does anyone have any objections to me merging this? In other branches in my fork I have been working on following work for #1609, #1611, and #1612. I would like to merge this so I can submit PRs for those, which are much smaller relative to this. I have [run a test of this code](<https://gist.github.com/keith-turner/29b72dfa97acad53c395dd0724b65f1c>) on a cluster and plan to run more test. I would like to wait to run subsequent test after finishing the follow on work.
- body:** @milleruntime I found all the TODOs and addressed/removed them.
label: code-design

github_pulls_reviews:

- Why disable these from the full logger by default?
- body:** I don't necessarily see the second reason to be a problem. Binding these into a single entity, can be nice for modularizing the user code to make it more reusable, and maintainable in a separate user-controlled repo. Under the new paradigm, what's the best way for users to combine their overall compaction strategy so they can maintain it separately, and just drop it in when needed? Would they just serialize the `CompactionConfig` in some way?
label: code-design
- ```suggestion this.className = requireNonNull(className);```
- Should the CompactionStrategy interface exist in an SPI package? Otherwise, it seems we haven't completely solved the use of internal types here.
- ```suggestion this.options = Map.copyOf(requireNonNull(opts)); return this;```
- Since `className` and `options` have setters, this class is mutable, and changing it can result in unpredictable behavior if stored in data structures that use `hashCode` and `equals`. Can this class be refactored to be immutable?
- body:** I'm confused by this javadoc. I don't think this javadoc is sufficient for me to understand how it is intended to be used, without additional outside documentation. Can you provide accompanying code for the compression example? Also, I'm wondering if we can come up with a better name for this. I might be able to think of something once I get a better grasp on how it's supposed to be used.
label: documentation
- Can this suppression occur more narrowly, rather than for the entire class?
- Where do these numbers come from? An inline comment would be useful.
- Presumably, these new properties are for major compactons only. The property prefix could be made consistent with other `tserver.compaction.major` properties.
- I don't know, it was probably annoying me as I spent a lot of time looking at the logs while working on this. That change should not be made in this PR, I will pull it out.
- body:** This entire class is utility code for a deprecated class.
label: code-design
- I don't remember. I may have taken a git commit hash or they may be random. The intent is to be random.
- [CompressionConfigurer]
(<https://github.com/apache/accumulo/blob/bd206be8ed50ebf1fee76dc8cf1e9820ef8ca5b8/core/src/main/java/org/apache/accumulo/core/client/admin/compaction/CompactionConfigurer.java>) is an example. It overrides the table property for compression type when the sum of input files exceeds a certain size.
- My motivation was in the past we had two compaction strategies included with Accumulo, one that filtered compaction candidates based on max size and another that changed the compression type based on input files sizes. These had nothing to do with each other, but only one could be configured at a time.
- That should not be mentioned, that is a copy and paste bug. I will fix that.
- The javadoc for Map.copyOf says it will throw an NPE if its input is null. So the requireNonNull could be dropped.
- body:** In that case, it too can be marked `@Deprecated` instead of merely suppressing the deprecation.
label: code-design
- body:** I think other non-deprecated code calls this to deal with compaction strategies. I was trying to put the code that does this in one place and in order to avoid suppression elsewhere.
label: code-design
- removed in ecc79e3
- Made immutable in ecc79e3
- fixed in ecc79e3
- fixed in ecc79e3
- fixed in ecc79e3
- body:** Would this be more appropriately named as a factory? It seems to be creating a compaction configuration - so maybe something like CompressionConfigFactory?
label: code-design
- I agree the current name is not great. I am not too excited about CompressionConfigFactory either. One thing that occurred to me while running a recent test on a cluster of this PR is that in addition to thinking of a new class name, we also need to think about the table property names. Below are table props I set when running the test. ```table.compaction.configurer=org.apache.accumulo.core.client.admin.compaction.CompressionConfigurer table.compaction.configurer.opts.large.compress.threshold=100M table.compaction.configurer.opts.large.compress.type=gz```
- @ctubbsii does my answer address your concern?
- @ctubbsii did the answer and the slack chat we had resolve this or do you think there is still something to be done?
- Done in 8d024d8
- body:** Typo ```suggestion * This class selects which files a user compaction will compact. It can also be configured per table```
label: documentation
- body:** Open issue for TODO
label: requirement
- body:** Open Issue for TODO
label: requirement

33. **body:** Open issue for TODO
label: requirement

34. **body:** Open issue for TODO
label: requirement

35. **body:** Open issue for TODO
label: requirement

36. **body:** You don't want warnings for this class? It is no longer used with this change.
label: code-design

37. I think these names weren't updated after previous changes. ``suggestion private PluginConfig selectorConfig = UserCompactionUtils.DEFAULT_SELECTOR;
 ``

38. ``suggestion private PluginConfig configurerConfig = UserCompactionUtils.DEFAULT_CONFIGURER; ``

39. ``suggestion * imported files. Accumulo estimates that bulk imported files have zero entries. The second option ``

40. ``suggestion "Compaction planner for metadata table"), ``

41. ``suggestion "Options for the table compaction configurer"), ``

42. ``suggestion "A configurable dispatcher that decides what compaction service a table should use."), ``

43. **body:** API javadoc needs description.
label: documentation

44. I think these names weren't updated after previous changes. (see other suggestion)

45. Looks like no changes to this file

46. **body:** Not sure what this means, javadoc could be improved.
label: documentation

47. **body:** Javadoc description needed for public facing SPI.
label: documentation

48. ``suggestion /** ``

49. Should this class be in this package?

50. Should have brief description.

51. ``suggestion * default it dispatches to a compaction service named default. * * <p> * The following schema is supported for configuration options. ``

52. ``suggestion * The following is a description of each functional component. * * * Compaction Manager A non pluggable component within the tablet server that brings all * other components together. The manager will route compactables to compaction services. For each * kind of compaction, an individual compactable will be routed to a single compaction service. For ``

53. This related to removal for merging minor compactions?

54. Is all the work for removing merging minor compactions done here?

55. Yes

56. done in 5aa9672

57. done in 5aa9672

58. done in 5aa9672

59. good catch, fixed in 5aa9672

60. yeah, eliminating merging minor compaction eliminated the need of dealing with a file being merged and in use by scans.

jira_issues:

1. **summary:** Consistency of minor compaction and major compaction differ
description: Minor compactions add files to !METADATA table and then update the tablets in memory list of files. Major compactions do the opposite, update in memory list then !METADATA. In the case of machine failures, this could lead to a user see the result of major compaction iterators and then not seeing it. Minor compactions do not have this issue. See ACCUMULO-6
2. **summary:** Consistency of minor compaction and major compaction differ
description: Minor compactions add files to !METADATA table and then update the tablets in memory list of files. Major compactions do the opposite, update in memory list then !METADATA. In the case of machine failures, this could lead to a user see the result of major compaction iterators and then not seeing it. Minor compactions do not have this issue. See ACCUMULO-6
3. **summary:** Consistency of minor compaction and major compaction differ
description: Minor compactions add files to !METADATA table and then update the tablets in memory list of files. Major compactions do the opposite, update in memory list then !METADATA. In the case of machine failures, this could lead to a user see the result of major compaction iterators and then not seeing it. Minor compactions do not have this issue. See ACCUMULO-6
4. **summary:** Consistency of minor compaction and major compaction differ
description: Minor compactions add files to !METADATA table and then update the tablets in memory list of files. Major compactions do the opposite, update in memory list then !METADATA. In the case of machine failures, this could lead to a user see the result of major compaction iterators and then not seeing it. Minor compactions do not have this issue. See ACCUMULO-6
5. **summary:** Consistency of minor compaction and major compaction differ
description: Minor compactions add files to !METADATA table and then update the tablets in memory list of files. Major compactions do the opposite, update in memory list then !METADATA. In the case of machine failures, this could lead to a user see the result of major compaction iterators and then not seeing it. Minor compactions do not have this issue. See ACCUMULO-6

jira_issues_comments:

1. Any update to this issue? Considering how long it has been open, is it still a concern?
2. There has been no interest in this ticket in the last three years.
3. This is an open bug. Just because no one has worked on it doesn't mean that it isn't relevant.