

Item 122

git_comments:

git_commits:

1. **summary:** [SPARK-26003][SQL][2.4] Improve SQLAppStatusListener.aggregateMetrics performance
message: [SPARK-26003][SQL][2.4] Improve SQLAppStatusListener.aggregateMetrics performance
This PR is to cherry-pick <https://github.com/apache/spark/pull/23002> to Spark 2.4 --- ## What changes were proposed in this pull request? In `SQLAppStatusListener.aggregateMetrics`, we use the `metricIds` only to filter the relevant metrics. And this is a Seq which is also sorted. When there are many metrics involved, this can be pretty inefficient. The PR proposes to use a Set for it. ## How was this patch tested? NA Closes #23002 from mgaido91/SPARK-26003. Closes #25860 from gatorsmile/cherrypickSPARK-26003. Authored-by: Marco Gaido <marcogaido91@gmail.com> Signed-off-by: Dongjoon Hyun <dhyun@apple.com>

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [SPARK-26003] Improve SQLAppStatusListener.aggregateMetrics performance
body: ## What changes were proposed in this pull request? In `SQLAppStatusListener.aggregateMetrics`, we use the `metricIds` only to filter the relevant metrics. And this is a Seq which is also sorted. When there are many metrics involved, this can be pretty inefficient. The PR proposes to use a Set for it. ## How was this patch tested? NA

github_pulls_comments:

1. **[Test build #98683 has finished]
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/98683/testReport>)** for PR 23002 at commit [`7e79041`]
(<https://github.com/apache/spark/commit/7e790412ed6409fdda96216dde7f4f408bb04a57>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
2. cc @cloud-fan @vanzin
3. LGTM, also cc @gengliangwang
4. **[Test build #98730 has finished]
(<https://amplab.cs.berkeley.edu/jenkins/job/SparkPullRequestBuilder/98730/testReport>)** for PR 23002 at commit [`031d512`]
(<https://github.com/apache/spark/commit/031d512b84e0b84a1876c098e0842f13d37c38e8>). * This patch passes all tests. * This patch merges cleanly. * This patch adds no public classes.
5. thanks, merging to master!

github_pulls_reviews:

1. Actually this one can be merged into `metricTypes`.
2. Consider also change the following `flatMap` / `filter` / `groupBy` into `while` loop
3. not sure what you mean here. Why should we use a `while` loop?
4. If the metrics is large, then using a while loop can reduce the number of traversal loops. And it is not complicated to do it in the code here.
5. I am also fine with the current code here.
6. yes, we can save 1 traversal, but I am not sure it is worth honestly... This approach seems cleaner to me.

jira_issues:

jira_issues_comments: