# Infrastructure - Task #8179

Story # 8175 (Closed): Solr crashing with out of memory error

## review log aggregation code for memory leaks and efficiency

2017-09-05 16:45 - Dave Vieglais

| | | | |
|---|---|---|---|
| **Status:** | Closed | **Start date:** | 2017-09-05 |
| **Priority:** | Normal | **Due date:** | |
| **Assignee:** | Dave Vieglais | **% Done:** | 100% |
| **Category:** | d1_log_aggregation | **Estimated time:** | 0.00 hour |
| **Target version:** | | | |
| **Milestone:** | None | **Story Points:** | |
| **Product Version:** | * | | |

**Description**

Solr memory use on the CN running log aggregation always appears higher than on the other CNs.

This suggests that log aggregation imposes a high load on the Solr index or there may be some inefficiencies of memory leaks in the log aggregation service.

The goal of this task is to review the log aggregation implementation and correct where necessary.

### Associated revisions

**Revision 19050 - 2017-12-18 19:29 - Rob Nahf**

refs: #8179. in trunk, fixed resource leak in LogAggretationScheduleManager.init() method.

**Revision 19050 - 2017-12-18 19:29 - Rob Nahf**

refs: #8179. in trunk, fixed resource leak in LogAggretationScheduleManager.init() method.

**Revision 19051 - 2017-12-18 21:08 - Rob Nahf**

refs: #8179. in trunk, fixed resource leak in LogHarvesterTask.harvest method that was not closing FileReaders in a finally block. The CSVParser is Autoclosable, so in theory under regular operation, resources were getting closed, and so resource leaks only would be happening when exceptions thrown. Refactored to use the try-with-resources construct.

**Revision 19051 - 2017-12-18 21:08 - Rob Nahf**

refs: #8179. in trunk, fixed resource leak in LogHarvesterTask.harvest method that was not closing FileReaders in a finally block. The CSVParser is Autoclosable, so in theory under regular operation, resources were getting closed, and so resource leaks only would be happening when exceptions thrown. Refactored to use the try-with-resources construct.

### History

**#1 - 2017-12-18 15:57 - Rob Nahf**

*- Category changed from d1_indexer to d1_log_aggregation*

**#2 - 2017-12-18 18:13 - Rob Nahf**

see https://wiki.apache.org/solr/SolrPerformanceProblems for point of reference. It gives a good overview of potential problem areas.

There's an interesting point about potential problems with the zookeeper database on the same disk as the solr database. This might be a pathway for one solr index (query index) to indirectly impact another (the log index).

Given multiple causes, it would be good to note specific issues and symptoms.

**#3 - 2017-12-18 18:23 - Rob Nahf**

Not directly related to memory, but looked at the SolrClientManager code that submits the updates. The submitBeans call wraps the apache HttpSolrClient.addBeans method. The logs show decent update speed.

```
[ INFO] 2017-11-28 18:00:45,889 (SolrClientManager:submitBeans:42) SolrClientManager-urn:node:mnSandboxUCSB2 submitting 500 records to index
[DEBUG] 2017-11-28 18:00:45,889 (SolrClientManager:submitBeans:45) SolrClientManager-urn:node:mnSandboxUCSB2 locked
[ INFO] 2017-11-28 18:00:47,406 (SolrClientManager:submitBeans:50) SolrClientManager-urn:node:mnSandboxUCSB2 Completed submission
[DEBUG] 2017-11-28 18:00:47,406 (SolrClientManager:submitBeans:53) SolrClientManager-urn:node:mnSandboxUCSB2 unlocked
[DEBUG] 2017-11-28 18:00:47,407 (LogHarvesterTask:harvest:430) LogHarvesterTask-urn:node:mnSandboxUCSB2 publishing 500 items to Solr
[ INFO] 2017-11-28 18:00:47,407 (SolrClientManager:submitBeans:42) SolrClientManager-urn:node:mnSandboxUCSB2 submitting 500 records to index
[DEBUG] 2017-11-28 18:00:47,407 (SolrClientManager:submitBeans:45) SolrClientManager-urn:node:mnSandboxUCSB2 locked
[ INFO] 2017-11-28 18:00:48,563 (SolrClientManager:submitBeans:50) SolrClientManager-urn:node:mnSandboxUCSB2 Completed submission
[DEBUG] 2017-11-28 18:00:48,563 (SolrClientManager:submitBeans:53) SolrClientManager-urn:node:mnSandboxUCSB2 unlocked
[DEBUG] 2017-11-28 18:00:48,563 (LogHarvesterTask:harvest:430) LogHarvesterTask-urn:node:mnSandboxUCSB2 publishing 500 items to Solr
[ INFO] 2017-11-28 18:00:48,564 (SolrClientManager:submitBeans:42) SolrClientManager-urn:node:mnSandboxUCSB2 submitting 500 records to index
[DEBUG] 2017-11-28 18:00:48,564 (SolrClientManager:submitBeans:45) SolrClientManager-urn:node:mnSandboxUCSB2 locked
[ INFO] 2017-11-28 18:00:49,523 (SolrClientManager:submitBeans:50) SolrClientManager-urn:node:mnSandboxUCSB2 Completed submission
[DEBUG] 2017-11-28 18:00:49,523 (SolrClientManager:submitBeans:53) SolrClientManager-urn:node:mnSandboxUCSB2 unlocked
[DEBUG] 2017-11-28 18:00:49,523 (LogHarvesterTask:harvest:430) LogHarvesterTask-urn:node:mnSandboxUCSB2 publishing 133 items to Solr
[ INFO] 2017-11-28 18:00:49,524 (SolrClientManager:submitBeans:42) SolrClientManager-urn:node:mnSandboxUCSB2 submitting 133 records to index
[DEBUG] 2017-11-28 18:00:49,524 (SolrClientManager:submitBeans:45) SolrClientManager-urn:node:mnSandboxUCSB2 locked
[ INFO] 2017-11-28 18:00:49,876 (SolrClientManager:submitBeans:50) SolrClientManager-urn:node:mnSandboxUCSB2 Completed submission
[DEBUG] 2017-11-28 18:00:49,876 (SolrClientManager:submitBeans:53) SolrClientManager-urn:node:mnSandboxUCSB2 unlocked
[ INFO] 2017-11-28 18:00:49,881 (LogHarvesterTask:harvest:470) LogHarvesterTask-urn:node:mnSandboxUCSB2 Latested Harvested Log Entry Date 2017-11-27T23:48:51.023+0000
```

## [ INFO] 2017-11-28 18:00:49,881 (LogHarvesterTask:harvest:474) LogHarvesterTask-urn:node:mnSandboxUCSB2 Complete

```
[ INFO] 2017-11-28 18:02:33,169 (SolrClientManager:submitBeans:42) SolrClientManager-urn:node:mnSandboxORC1 submitting 17 records to index
[DEBUG] 2017-11-28 18:02:33,169 (SolrClientManager:submitBeans:45) SolrClientManager-urn:node:mnSandboxORC1 locked
[ INFO] 2017-11-28 18:02:33,437 (SolrClientManager:submitBeans:50) SolrClientManager-urn:node:mnSandboxORC1 Completed submission
[DEBUG] 2017-11-28 18:02:33,437 (SolrClientManager:submitBeans:53) SolrClientManager-urn:node:mnSandboxORC1 unlocked
[ INFO] 2017-11-28 18:02:33,443 (LogHarvesterTask:harvest:470) LogHarvesterTask-urn:node:mnSandboxORC1 Latested Harvested Log Entry Date 2017-11-27T23:46:19.679+0000
[ INFO] 2017-11-28 18:02:33,443 (LogHarvesterTask:harvest:474) LogHarvesterTask-urn:node:mnSandboxORC1 Complete
```

depending on the number of beans submitted to update, it averages 2-3ms per bean. For smaller commits, it's less efficient, so there's likely some fixed-time overhead in the commit.

**#4 - 2017-12-18 19:34 - Rob Nahf**

*- Status changed from New to In Progress*

*- % Done changed from 0 to 30*

found one resource leak - an unclosed inputStream that probably is inconsequential as it likely occurs only once at LogAggregationScheduleManager init.

**#5 - 2017-12-18 21:10 - Rob Nahf**

found some more potential resource leaks in LogHarvesterTask.harvest method.  The FileReaders were not being closed within a finally block.
switched to using try-with-resources construct to be more legible.

**#6 - 2019-01-22 19:02 - Dave Vieglais**

*- % Done changed from 30 to 100*

*- Status changed from In Progress to Closed*