

Infrastructure - Story #3349

LogAggregation needs to be more fault tolerant

2012-10-19 22:29 - Robert Waltz

Status:	Closed	Start date:	2012-10-19
Priority:	Normal	Due date:	2013-01-05
Assignee:	Robert Waltz	% Done:	100%
Category:	d1_log_aggregation	Estimated time:	0.00 hour
Target version:	Sprint-2012.50-Block.6.4		
Story Points:			

Description

Staging test of logAggregation failed with only 10% of records processed with the following error

```
[ERROR] 2012-10-06 05:43:57,221 (LogAggregatorTask:retrieve:289) LogAggregatorTask-urn:node:mnStageUCSB1 <?xml version="1.0" encoding="UTF-8"?>
```

```
class javax.net.ssl.SSLPeerUnverifiedException: peer not authenticated
```

regardless of the cause, logAggregation needs a more fault tolerant mechanism to harvest records.

Original Algorithm

b = begin date

e = end date

t = total record count

l = limit of records to retrieve

p = total records processed

```
harvestRecords()
{
do
{
recordstoProcess = queryLogRecords(b,e,l);
total = recordstoProcess.totalRecordsFound;
if (recordstoProcess.isNotEmpty())
{
process logRecords;
// maintain the date of the last record processed
update lastLogRecordProcessedDate;
p += recordstoProcess.getCount(); //count should be <= l
}
} while (p < t);
}
```

First time trial with algorithm from current implementation:

05:53:09 began

22:16:49 (failed due to connection problem!)

16:23:40 total hours at time of failure: retrieved 575000 count=1000 total=2306641

By these numbers it would take 2 2/3 days to complete > 2M records

Divide and Conquer Algorithm

b = begin date

e = end date

t = total record count

l = limit insertions

```
// recursively recordstoProcess may become huge memory problem
```

```
retrieveLogRecords( b , e)  
recordstoProcess = queryLogRecords;  
if (c > l && b != e)  
{  
    e1 = median(b,e);  
    recordstoProcess += retrieveLogRecords( b , e1);  
    recordstoProcess += retrieveLogRecords( e1, e);  
}  
return recordstoProcess;
```

```
// use explicit stacking mechanism and loops instead of recursion  
// earliest records should always be pushed on the top of the stack;
```

```
b = earliest date to process;  
e = now;  
l = 5000;
```

```
harvestRecords()  
{  
    new DateStack(b,e);  
    do  
    {  
        logRecords = null;  
        try_again = false;  
        try  
        {  
            logRecords = retrieveLogRecords();  
        } catch (exception TryAgain e) {  
            try_again = true;  
        } catch (exception SomethingBadHappened e) {  
            try_again=true; X number of times before giving up  
        }  
        if (logRecords.isEmpty)  
        {  
            process logRecords;  
            // maintain the date of the last record processed  
            update lastLogRecordProcessedDate;  
        }  
    } while (DateStack.count > 0 || try_again);  
}
```

```
retrieveLogRecords()  
{  
    (b,e) = pop DateStack();  
    records = queryLogRecords(b,e);  
    t = records.total();  
    if (t > l && b != e)  
    {  
        e1 = median(b,e);  
        push DateStack(e1,e);  
        push DateStack(b,e1);  
        throw TryAgain;  
    }  
    else  
    {  
        return records  
    }  
}
```

```
stats of current run:
```

[INFO] 2012-10-19 20:45:49,558 (DivideAndConquer:getRecords:94) retrieving from start=0

[INFO] 2012-10-19 22:05:56,186 (DivideAndConquer:doit:80) Total Harvested Log count 575586

01:20:07 to retrieve 575586 records thus far

shows about 92% reduction in processing time

Subtasks:

Task # 3350: Test different algorithms to prove performance improvement and fault toler...	Closed
Task # 3351: Integrate new harvesting prodecure into LogAggregatorTask	Closed

History

#1 - 2012-10-19 22:38 - Robert Waltz

- Target version changed from Sprint-2012.46-Block.6.3 to Sprint-2012.44-Block.6.2
- Due date changed from 2012-12-01 to 2012-11-10

#2 - 2012-10-19 22:39 - Robert Waltz

- Due date changed from 2012-11-10 to 2012-10-27
- Target version changed from Sprint-2012.44-Block.6.2 to Sprint-2012.41-Block.6.1

#3 - 2012-10-22 16:29 - Robert Waltz

- Description updated

#4 - 2012-10-24 17:24 - Robert Waltz

- Status changed from New to In Progress

#5 - 2012-10-29 16:34 - Robert Waltz

- Target version changed from Sprint-2012.41-Block.6.1 to Sprint-2012.44-Block.6.2
- Due date changed from 2012-10-27 to 2012-11-10

#6 - 2012-12-04 14:13 - Robert Waltz

- Target version changed from Sprint-2012.44-Block.6.2 to Sprint-2012.50-Block.6.4
- Due date changed from 2012-11-10 to 2013-01-05

#7 - 2013-01-08 18:27 - Robert Waltz

- Status changed from In Progress to Closed