

DataDirectTM
N E T W O R K S
I N F O R M A T I O N I N M O T I O N TM

Simplifying Lustre Operations with Chukwa and Hadoop

LAD2012 Paris, France

Travis Dawson

Abstract

Lustre is complicated. It is commonly deployed across dozens of servers, supporting thousands of clients. Each of these components is constantly generating logs providing insight into operations, performance, errors and recoveries within the environment. A Lustre environment that generates millions of events a day is not uncommon.

A system administrator can quickly become overwhelmed in trying to associated events from separate servers and create a complete picture of how health the environment is, proactively respond to problems and enable visibility for developers and users to the environment.

Enter Chukwa. Chukwa is a distributed tool for log gathering, storage and analysis that runs on Apache Hadoop, leveraging the Hadoop Distributed File System. Chukwa enables administrators to intelligently locate patterns in their environments, respond and correlate events.

This presentation will cover how Chukwa can be implemented on top of Apache Hadoop to facilitate the gathering, correlation, analysis and reporting of large numbers of events across a Lustre environment.

Lustre Operational Struggles

- ▶ Lots of Log Messages
 - Storage System
 - Hardware
 - Lustre/Application
- ▶ Messages are hard to decipher
 - Change version to version
- ▶ ‘Hidden’ and linked errors
 - Errors/Warnings that hide other problems
- ▶ Reactive vs Proactive
 - So many logs only able to concentrate on failures and total errors

Current Lustre Logging/Monitoring

- ▶ SyslogNG
- ▶ Rationalized PrintK
 - John Hammond TACC
 - Used to decipher messages
- ▶ SEC
 - Simple/Statistical Event Correlation
 - Perl
- ▶ Nagios
 - Sometimes Multiple Instances to handle the load
- ▶ RRDTool
 - Pretty Pictures

One really big machine to monitor them all

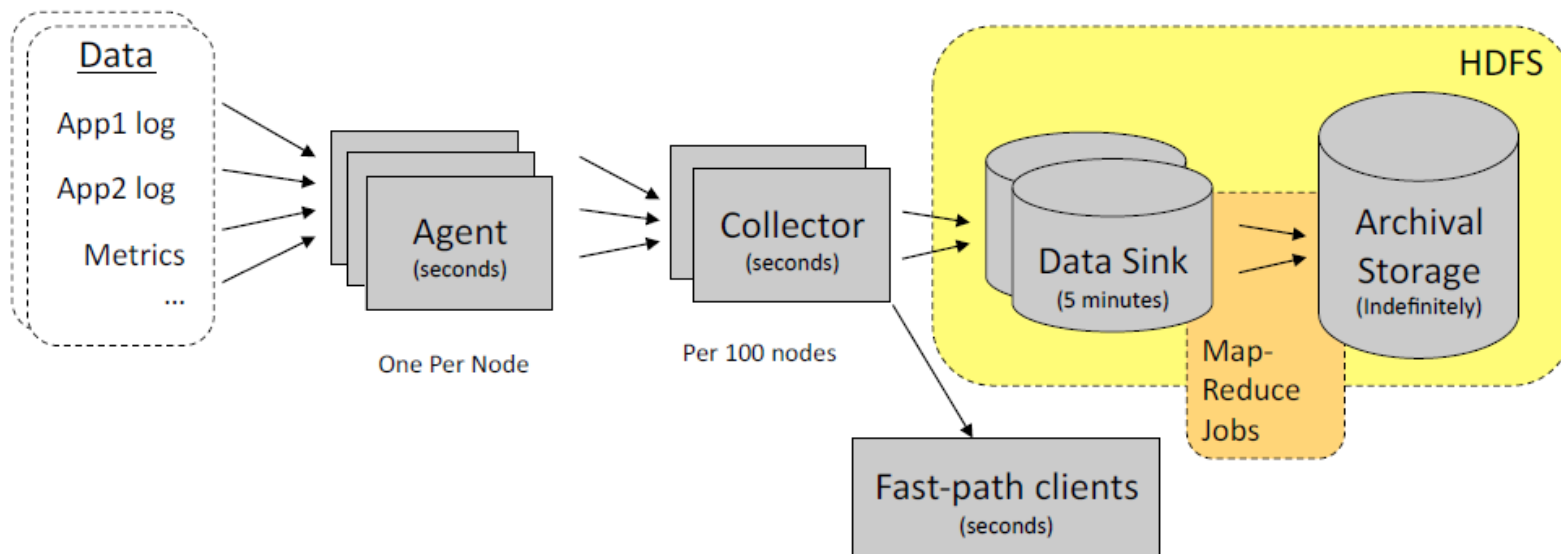
Introducing Chukwa

- ▶ Open Source Data Collection System
 - <http://incubator.apache.org/chukwa>
- ▶ Built on top of HDFS and Map/Reduce (Hadoop)
 - Hadoop Scalability & Robustness
- ▶ Built in Visualization toolkit (HICC)
 - Widgets
 - Alerts
 - Pretty Graphs



Chukwa Architecture

- ▶ Host Agent (low overhead)
 - Low/No impact on applications
- ▶ Fast-Path and Reliable Delivery
 - Alerting as well as Archival
 - Scalable to over **200MB/sec**
- ▶ HICC Visualization
 - Built in Graphing and Widgets



Visualization/Reporting/Alerting



Lustre Logs

```
Nov 30 17:47:41 uiy601 attrd_updater: [7930]: info: Invoked: attrd_updater -n pingd-tcp -v 4 -d 60
Nov 30 17:48:06 uiy601 ntpd[5861]: synchronized to 10.0.6.173, stratum 4
Nov 30 17:49:45 uiy601 attrd_updater: [8059]: info: Invoked: attrd_updater -n pingd-tcp -v 4 -d 60
Nov 30 17:51:06 uiy601 kernel: Lustre: 7063:0:(mds_lov.c:1191:mds_notify()) MDS scratch-MDT0000: in recovery, not resetting orphans on scratch-OST0003_UUID
Nov 30 17:51:06 uiy601 kernel: Lustre: scratch-OST0003-osc: Connection restored to service scratch-OST0003 using nid 10.0.6.176@tcp.
Nov 30 17:51:06 uiy601 kernel: Lustre: 7063:0:(mds_lov.c:1191:mds_notify()) MDS scratch-MDT0000: in recovery, not resetting orphans on scratch-OST0004_UUID
Nov 30 17:51:06 uiy601 kernel: Lustre: scratch-OST0004-osc: Connection restored to service scratch-OST0004 using nid 10.0.6.176@tcp.
Nov 30 17:51:49 uiy601 attrd_updater: [8190]: info: Invoked: attrd_updater -n pingd-tcp -v 4 -d 60
Nov 30 17:52:02 uiy601 cib: [6014]: info: cib_stats: Processed 228 operations (175.00us average, 0% utilization) in the last 10min
Nov 30 17:52:19 uiy601 kernel: LustreError: 7667:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 300s
Nov 30 17:52:19 uiy601 kernel: LustreError: 7667:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf81061f5cc000 x1386922024894473/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672039 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:52:25 uiy601 kernel: LustreError: 7668:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 293s
Nov 30 17:52:25 uiy601 kernel: LustreError: 7668:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf81030f78fc00 x1386922024894476/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672045 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:52:32 uiy601 kernel: LustreError: 7669:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 286s
Nov 30 17:52:32 uiy601 kernel: LustreError: 7669:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf810310f36800 x1386922024894477/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672052 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:52:39 uiy601 kernel: LustreError: 7670:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 279s
Nov 30 17:52:39 uiy601 kernel: LustreError: 7670:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf8103086b1c00 x1386922024894478/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672059 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:52:46 uiy601 kernel: LustreError: 7671:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 272s
Nov 30 17:52:46 uiy601 kernel: LustreError: 7671:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf810312dcc000 x1386922024894479/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672066 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:53:06 uiy601 kernel: LustreError: 7672:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.6@tcp (a64b13b1-9fc4-abf5-37f5-5860ef26fb03):
1 clients in recovery for 252s
Nov 30 17:53:06 uiy601 kernel: LustreError: 7672:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf810619797800 x1386922073128969/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672086 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:53:15 uiy601 kernel: LustreError: 7673:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.3@tcp (8de758f3-5e08-4d53-bef3-c952a2f7cc53):
1 clients in recovery for 243s
Nov 30 17:53:15 uiy601 kernel: LustreError: 7673:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf8103177a7c00 x1386922024894482/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672095 ref 1 fl Interpret:0/0 rc -16/0
Nov 30 17:53:34 uiy601 kernel: LustreError: 7676:0:(ldlm_lib.c:944:target_handle_connect()) scratch-MDT0000: denying connection for new client 10.0.6.6@tcp (a64b13b1-9fc4-abf5-37f5-5860ef26fb03):
1 clients in recovery for 225s
Nov 30 17:53:34 uiy601 kernel: LustreError: 7676:0:(ldlm_lib.c:944:target_handle_connect()) Skipped 2 previous similar messages
Nov 30 17:53:34 uiy601 kernel: LustreError: 7676:0:(ldlm_lib.c:1919:target_send_reply_msg()) @@@ processing error (-16) req@afcbf81061fed2400 x1386922073128973/t0 o38-><?>@<?>:0/0 lens
368/264 e 0 to 0 dl 1322672114 ref 1 fl Interpret:0/0 rc -16/0
```


Analyzing Lustre Logs

- ▶ Groups of “xx recoverable clients remain” and “temporarily refusing client connection from xx.xx.xx.xx@tcp” messages
 - XX gets lower and lower
 - What other messages appear around that time (scheduled cron job taking up CPU)

- ▶ Eviction Root Cause Analysis
 - Network issues (correlate with packet drops)
 - BUG (look for related errors, Kernel Messages)
 - Random Gremlins (stop feeding them)

- ▶ Any other ideas?

Big Data Analytics on Lustre Logs

- ▶ Machine Logs are Ideal for Analytics
 - Little variability, easy to parse (mostly)
- ▶ REAL event correlation
 - Storage Subsystem
 - Hardware
 - Application
- ▶ Find the Root Cause of Issues
 - How does a slight issue here result in a failure there
- ▶ Detect subtle errors using data mining and machine learning techniques
 - Xu et al “Detecting Large Scale System Problems by Mining Console Logs” 22nd ACM SOSP, 2009
- ▶ Proactive vs Reactive

Future work

- ▶ **Deeper integration**
 - Blktrace/Seekwatcher for IO Performance
 - Deeper Lustre Logging/PrintK
- ▶ **Better Analytics**
 - Taking advantage of more Data Mining Research
 - “Detecting Large-Scale System Problems by Mining Console Logs”
- ▶ **Vampir Log Analytics**
 - Take in Vampir Log Traces
 - Plot them either internally or externally
 - Display using HICC

Putting it all together

- ▶ Scalable
- ▶ Capable
- ▶ Reactive -> Proactive

Questions



- ▶ Chukwa: A system for reliable large-scale log collection
 - <http://www.cs.berkeley.edu/~asrabkin/drafts/chukwa-lisa.pdf>

- ▶ Apache Org Project Pages
 - <http://incubator.apache.org/chukwa/>
 - <http://wiki.apache.org/hadoop/Chukwa>

- ▶ Detecting Large-Scale System Problems by Mining Console Logs
 - <http://www.cs.berkeley.edu/~jordan/papers/xu-et-al-sosp09.pdf>