Journal of Physics: Conference Series **396** (2012) 052014 doi:10.1088/1742-6596/396/5/052014

Operational Experience with the Frontier System in CMS

Barry Blumenfeld[1], Dave Dykstra[2], Peter Kreuzer[3], Ran Du[4], Weizhen Wang[5]

[1] Johns Hopkins University, Baltimore, MD, USA
[2] Fermilab, Batavia, IL, USA
[3,4,5] CERN, Geneva, Switzerland

Email: [1] bjb@jhu.edu [2] dwd@fnal.gov [3] Peter.Kreuzer@cern.ch
[4] Ran.Du@cern.ch [5] Weizhen.Wang@cern.ch

Abstract. The Frontier framework is used in the CMS experiment at the LHC to deliver conditions data to processing clients worldwide, including calibration, alignment, and configuration information. Each central server at CERN, called a Frontier Launchpad, uses tomcat as a servlet container to establish the communication between clients and the central Oracle database. HTTP-proxy Squid servers, located close to clients, cache the responses to queries in order to provide high performance data access and to reduce the load on the central Oracle database. Each Frontier Launchpad also has its own reverse-proxy Squid for caching. The three central servers have been delivering about 5 million responses every day since the LHC startup, containing about 40 GB data in total, to more than one hundred Squid servers located worldwide, with an average response time on the order of 10 milliseconds. The Squid caches deployed worldwide process many more requests per day, over 700 million, and deliver over 40 TB of data. Several monitoring tools of the tomcat log files, the accesses of the Squids on the central Launchpad servers, and the availability of remote Squids have been developed to guarantee the performance of the service and make the system easily maintainable. Following a brief introduction of the Frontier framework, we describe the performance of this highly reliable and stable system, detail monitoring concerns and their deployment, and discuss the overall operational experience from the first two years of LHC data-taking.

1. Introduction

The Frontier [1] framework is used in the CMS (Compact Muon Solenoid) experiment at the LHC to deliver conditions data (including calibration, alignment, and configuration information) to processing clients worldwide. Each central server at CERN, called a Frontier Launchpad, uses tomcat as a servlet container to establish the communication between clients and the central Oracle database. HTTP-proxy Squid servers, located close to clients, cache the responses to queries in order to provide high

performance data access and to reduce the load on the central Oracle database. Each Frontier Launchpad also has its own reverse-proxy Squid for caching.

In order to guarantee the performance of the service and make the system easily maintainable, several monitoring tools to test the availability of remote Squids, to analyze the tomcat log files and the accesses of the Squid on the central Launchpad Servers have been developed.

In this paper, following a brief introduction of the Frontier framework, we describe the load and performance of this highly reliable and stable system, detail monitoring concerns and their deployment, and discuss the overall operational experience from the first two years of LHC data-taking.

2. Frontier Framework

The CMS experiment at the LHC has established an infrastructure using the Frontier framework to deliver conditions data to processing clients worldwide. One key component of the offline deployment is the central server called 'Frontier Launchpad', in which three computers (cmsfrontier1-3) are operated in parallel and provide load sharing through a Round Robin DNS configuration, each with Tomcat and Squid services deployed. Tomcat serves as the servlet container. Each servlet is configured to connect to an instance of the Oracle database; Squid is used for the proxy/caching servers. Figure 1 shows the framework of Frontier Launchpad.

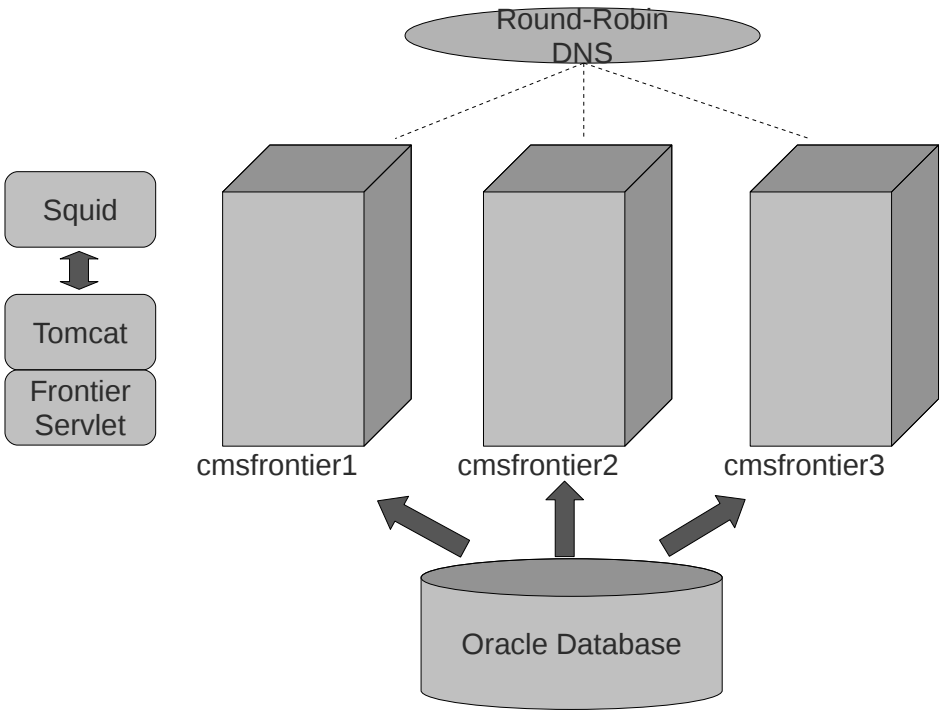

Figure 1: Frontier Launchpad Framework

Squid servers are maintained near clients. Frontier_client converts SQL queries into HTTP requests and sends them over the network. The HTTP requests go through a Squid for the local (T0/T1/T2) site,

and if the result is not in the cache the request is forwarded from there to another Squid on one of the Frontier Launchpads. If the request is not in that cache, the peer Squids are queried to see if any of them have it. If they don't have it, the request is forwarded to a Tomcat process on the same node and to the frontier servlet configured there. The frontier servlet converts the HTTP request back to an SQL query and sends that to an oracle server. The results go back in the reverse direction. Figure 2 shows the Frontier workflow briefly.

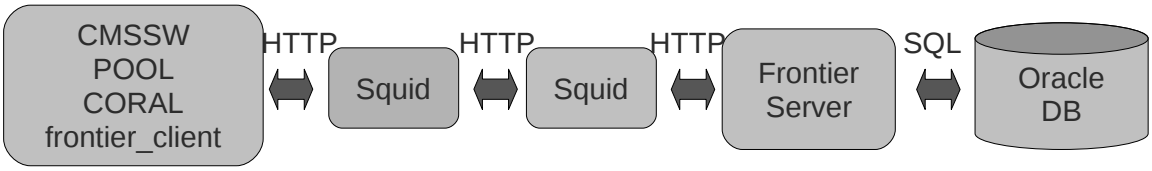

Figure 2: Frontier Workflow

Frontier Workload

Three Frontier Launchpad servers are working in parallel to handle data requests from all over the world. Each launchpad has a similar workload as the other two because of DNS round-robin. These three central servers have been delivering an average of 5 million responses every day since the LHC startup, containing about 40 GB data in total, to more than one hundred Squid servers located worldwide, with an average response time on the order of 10 milliseconds. The Squid caches deployed worldwide process many more requests per day, over 700 million, and deliver over 40 TB of data. Figure 3 shows the workload for the Frontier Launchpads during the past year (until May 10th). The green part stands for the total requests received from all Squids, while the blue line stands for the request passed to tomcat (not cached requests). As was mentioned in section 2, only the requests which can not be found in Squid cache are forwarded to Tomcat. As a result, the blue line should be always lower than the green part.

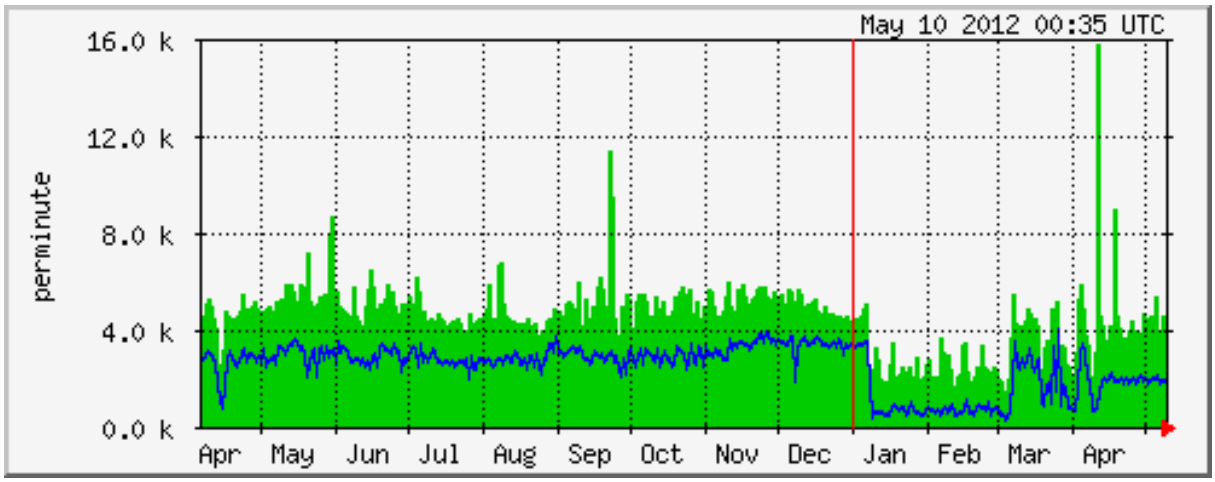

Figure 3: Frontier Launchpads Workload

3. Frontier Monitoring Tools and Operational Experience

Several monitoring tools [2] [3] test the availability of remote Squids, analyze the tomcat log files and the accesses of the Squid on the central Launchpad servers have been developed to guarantee the performance of the service and make the system easily maintainable. In this section, we will introduce seven Frontier monitoring tools in detail. In addition, some operational experience is also included here.

3.1. MRTG

All the Squids of CMS are monitored by an Open Source monitoring tool called MRTG (Multi Router Traffic Grapher) [3] which sends an SNMP request by UDP to every Squid every five minutes. Because these charts are updated every five minutes they always contain the most up-to-date information available. There are three kinds of MRTG plots: Request/Fetch plots, In/Out plots, Object Count plots. Each kind of plot has four different charts according to different lengths of duration displayed: Daily Graph, Weekly Graph, Monthly Graph and Yearly Graph. Figure 3 is the yearly graph for Frontier Launchpad Squids.

In addition to the plots, each MRTG web page contains a useful message that shows the Squid version and the length of time it has been running.

3.2. Squid Not Responding List

The Squid Not Responding List monitoring tool is used to find out which Squids are not currently active. The activeness of one Squid is tested by filtering corresponding MRTG plots web pages: all the Squids whose MRTG web page doesn't contain 'at which time' will be seen as not-responding Squids and added to the not-responding list. At the same time, alarms are sent to the cms-frontier-alarm@cern.ch mailing list. At the present time, alarms are sent for the Launchpad, T0, T1 and T2 sites but not on T3 sites.

When the operator receives such alarms titled 'No XXX Squid', it is better to check corresponding MRTG charts for those alarmed Squids. The reason is that since the MRTG probes are sent over UDP, it is possible that the remote Squid is working but because of temporary network problem the probe failed. If the MRTG plot says the Squid is up and has been for a long time, then you can tell the reason the Squid was not responding was due to a site network problem but the Squid itself was fine. That's why there is another script that generates a more 'visualized' monitoring page for the Squid-not-responding list.

Sometimes, it is particularly useful to look at the "Object count" chart for the Squid. This should normally be continuously green and monotonically increasing, unless the Squid has been restarted or the disk cache has filled up. If there are lots of dead regions (white bands in the green) that can indicate a network problem or a machine that is too heavily loaded to respond.

3.3. SAM Test

There are also Squid and Frontier tests in the Site Availability Monitor (SAM). These tests run every hour, but sometimes have a large latency. It should also be noted that some sites, in particular OSG sites in the US, may use their Squids for other experiments in addition to CMS and for caching things other than Frontier.

If a Squid is not responding, the operator should check the SAM tests for Squid and Frontier to see if it is a site problem or a Squid problem. Sometimes, one may have to wait an hour or two for the SAM tests to be updated. If it is a site problem, usually there is no need to do anything. If the other SAM tests are OK and only the Squid test is failing, the operator should send an e-mail to the site contact which can be found in SiteDB.

3.4. Non-proxy Monitoring Tools

The non-proxy problem refers to the possibility that jobs bypass the local Squid and directly access the Frontier Launchpads, which can cause a heavy load to Frontier Launchpads and low efficiency for the clients. The non-proxy problem is usually caused by inappropriate Squid settings or unexpected Squid downtime. In order to detect such kind of Squids, two monitoring tools were developed. One is called 'Failover Nodes', which detects non-proxy machines every hour, generates a web log and automatically sends an email warning to the administrator of a Squid when jobs at the Squid's site directly access the Frontier Launchpads too many times in an hour. The other tool is called 'Failover Summary', which keeps an hourly history of direct queries in the last three days.

3.5. AWStats

AWStats [5] is an Open Source program which shows detailed information from the log files of the Squids on each Frontier Launchpad. This program updates every hour. In addition to being useful by itself, it is the basis for the Non-Proxy Monitoring Tools.

One can also use AWStats to detect the non-proxy problem manually. In the left hand column, under Hosts, click on Full list. This will give the name of every machine in the world that contacted this Squid so far today. Normally, only site Squid servers should appear here. If one address has hundreds of thousands of hits more than usual, or there are dozens of addresses from the same site, either case means there are some problems with the remote Squid.

3.6. Max Thread Monitoring Tool

The number of threads refers to the number of concurrent requests to one servlet on a Frontier Launchpad. Each servlet has its own limit on the maximum number of concurrent requests, typically 100 or 200 threads.

The Max Thread Monitoring tool was developed to monitor the number of requests to every servlet on each Frontier Launchpad. At the same time, because DB response time has a significant effect on the number of threads, the max thread monitoring tool also monitors DB response time.

Whenever the number of threads on one Frontier Launchpad servlet exceeds a threshold (75% of the maximum allowed for any servlet), an alarm message will be emailed to the operations team so they can investigate the cause. The best method to find out what is happening is to analyze the Tomcat log file, catalina.out. The root cause may be one of the following:

- Database connection time out caused by DB errors
- Excessively time-consuming requests
- Too many requests sent by user jobs
- Too many concurrent jobs sent to a Frontier servlet

If it is caused by DB errors, then the operator should contact the DB administrators to see what happened to the Database. Otherwise, the operator should contact the user who is responsible for the max threads, ask him/her what's the situation and try to make sure it doesn't happen again. If it turns out that the user's job load is legitimate and a new normal, sometimes the Frontier Launchpads need to be reconfigured to be able to handle a larger load than seen previously.

3.7. Squid Source Compare Monitoring Tool

Sites that get condition data from Frontier are supposed to register their Squid(s) in CVS. At the same time, all Squids are monitored by MRTG. Sometimes the Squid information in CVS becomes inconsistent with the information in MRTG (e.g. an administrator doesn't keep CVS up to date, Squids are shared by multiple sites, or Squids have different addresses on private and public networks). In order to make sure the information stays as correct as possible, the Squid Source Comparing Monitoring tool was developed to keep Squid information consistent between CVS and MRTG. It sends an email warning whenever the two sources become different, and it includes the ability to store exceptions for cases where the operations team understands why the two sources are different.

4. Conclusions

Frontier monitoring tools and correspondence alarms are the basis of Frontier operations maintenance work. Squids are so important that several monitoring tools (e.g. MRTG, Squid Not Responding List, SAM Test, Non-proxy Monitoring Tools and the Squid Source Compare Monitoring Tool) were developed to make sure the Squids of all sites are working (or can be detected quickly if not). Every system has its own load limitations, and Frontier is no exception. As a result, load monitoring is another big part of a Frontier operator's work. That's why AWStats and the Max Thread Monitoring Tool were developed.

The alarms generated by monitoring tools have different priorities for the operation team. Generally speaking, the alarms generated because of inactivity of Tomcat/Squid (e.g. 'No XXX Squid' alarm) or the exceeding of monitoring threshold (e.g. Max thread exceeding threshold alarm) are the most urgent, while the non-realtime alarms are usually less urgent (e.g. Squid Source Compare alarm).

More information for Frontier monitoring tools can be found on the home page for Frontier, http://frontier.cern.ch. All the monitoring tools mentioned in this paper can be found on that page, so readers can get a direct impression of these monitoring tools.

5. Acknowledgements

Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.

References

[1] Dave Dykstra and Lee Lueking 2010 Greatly improved cache update times for conditions data with Frontier/Squid *J. Phys.: Conf. Ser.* **219** 072034
[2] https://twiki.cern.ch/twiki/bin/viewauth/CMS/FroNtier_FacOps
[3] https://twiki.cern.ch/twiki/bin/view/CMS/FrontierMonitoringGuide
[4] http://oss.oetiker.ch/mrtg/
[5] http://awstats.sourceforge.net/