Feasibility study of a PROOF cluster at a CMS Tier 3 site

ROB SNIHUR
University of Nebraska, Lincoln

1 Introduction – basic problem

We wish to analyze data produced in the CMS experiment at CERN's LHC in the most efficient manner possible. Here we describe a feasibility study of a PROOF (Parallel ROOt Facility) [1] cluster for CMS data analysis at a Tier 3 site. PROOF/ROOT has been optimized for the analysis of data from high energy physics collisions (or events) and its main features are transparency, scalability, and adaptability. It can also be adapted to run on multiple sites simultaneously via the GRID (see Fig. 1). PROOF clusters can be and have been implemented on OSG sites.

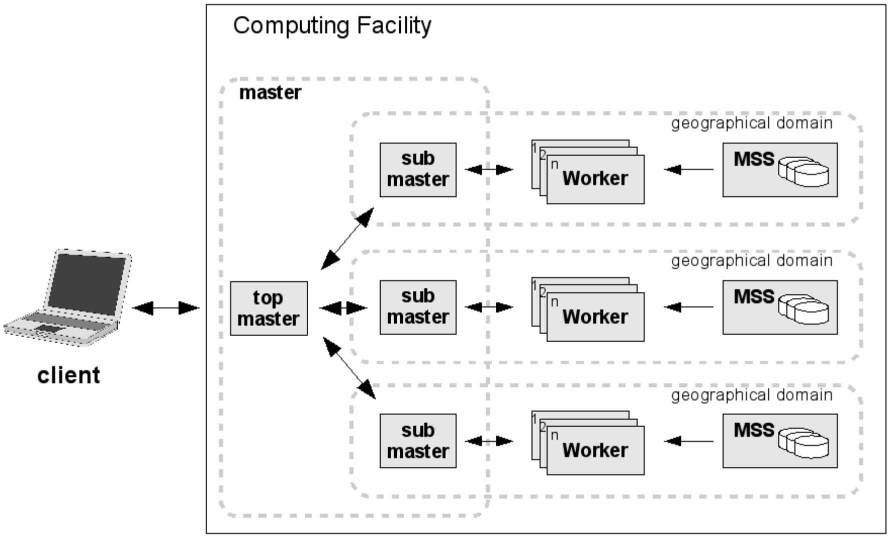

Figure 1: Multi-tier master-worker architecture

2 Conceptual solution

The basic idea of PROOF is to run ROOT in parallel using a master-worker architecture, as shown in the figure above. Each worker reads in a subset of the events, analyzes them sequentially and collects results (in histograms, for example). The results are delivered back to the master where they are merged together, then final results are delivered to the client.

The PROOF system employs event-level parallelism [2] (see Fig. 2), which dynamically determines the amount of work (called packets) distributed to the worker nodes. This ensures that all workers finish their assigned tasks at approximately the same time. While we spent a significant amount of time learning efficient ways to split up our workload in the OSG Summer School, the PROOF system includes such optimizations automatically.

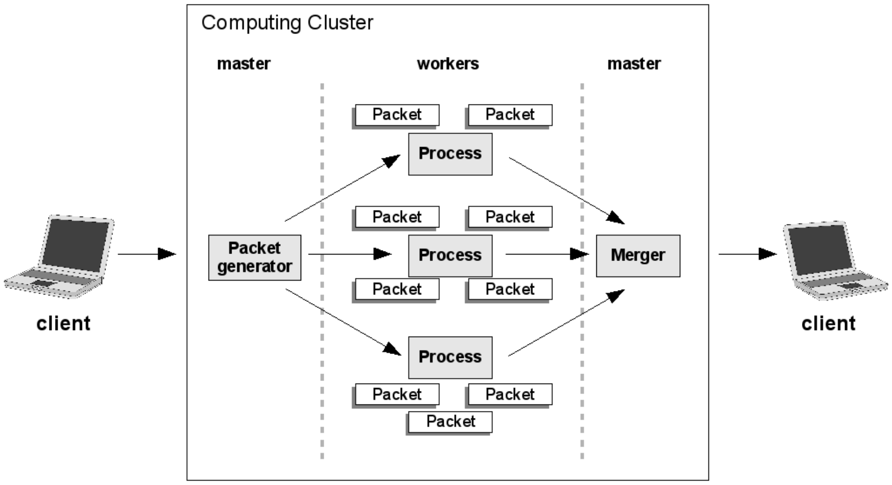

Figure 2: Execution flow illustrating event level parallelism.

Another optimization of note is the ability to bring the job to the data. The workers will preferentially read and analyze data files which are stored on local disks attached to the worker nodes. This reduces the overall network utilization within the cluster.

3 Proposal

As a first phase, we propose to run a simple CMS analysis on the PROOF cluster at the CMS Tier 2 site at Purdue [3]. Two types of analyses will be compared: IO-intensive and CPU-intensive. It is assumed that memory consumption will be below 1 GB per core so that existing resources at Purdue will be sufficient. We will use the most stripped-down data-format available within CMS (which may or may not be optimized for PROOF analyses), and select a dataset that already exists at Purdue (it's already in ROOT format). The dataset must be large enough so that the overhead for each job on a PROOF worker is negligible; we guess that about 100 GB should be sufficient. We will benchmark the performance of the analysis and try to compare against a conventional analysis (using the full CMS framework, or possibly "Framework Lite"). Other parameters to investigate include the number of worker nodes, and the number of jobs executing on the PROOF master.

We will also attempt to use the ATLAS cluster in Madison, Wisconsin, where PROOF has been integrated with condor [4]. This will require copying the dataset from Purdue and storing it at Madison.

4 Future work

There are a few obvious next phases for this proposal:

1. Evaluate what would be different at a Tier 3 site compared to the Purdue Tier 2 site. For instance, a Tier 3 site may choose not to enable GRID services and may simply have a batch cluster; the PROOF cluster should be easy to set up in this case. The cluster can be configured such that conventional (e.g. condor) jobs are preempted, and the PROOF tasks get highest priority in an ”interactive” mode.

2. Build a PROOF cluster at a Tier 3 site [5]. Possible extension: use the xrootd distributed filesystem.

3. Read data from remote sites via the xrootd WAN project now underway within CMS [6].

4. Federate PROOF clusters among Tier 3 sites. In general, the workers would be located on remote sites.

5 Summary

The use of a PROOF cluster offers several possible benefits for data analysis within the CMS experiment, especially for Tier 3 sites. The feasibility study proposed here has several phases. The first phase will benchmark a simple analysis work flow on the PROOF cluster at the Purdue CMS Tier 2 site. Subsequent phases would evaluate and implement PROOF clusters at CMS Tier 3 sites, and federate them together.

References

[1] Chapter 25 of ROOT Guide
ftp://root.cern.ch/root/doc/chapter25.pdf
PROOF web site
http://root.cern.ch/drupal/content/proof

[2] http://root.cern.ch/drupal/content/event-level-parallelism

[3] PROOF at Purdue T2
http://www.physics.purdue.edu/Tier2/content/proof-example

[4] PROOF-Condor Integration
http://www.cs.wisc.edu/condor/PCW2008/condor_presentations/ganis_proof.pdf

[5] how to set up a proof cluster:
http://root.cern.ch/drupal/content/proof-installation
http://root.cern.ch/drupal/content/standard-proof-installation-cluster-machines

[6] xrootd WAN
https://twiki.cern.ch/twiki/bin/view/Main/CmsXrootdArchitecture