PROOF on Demand[*]

P. Malzacher, A. Manafov, V. Penso, C. Preuss, and K. Schwarz

GSI, Darmstadt, Germany

The Parallel ROOT [1] Facility, PROOF [2], is an extension of ROOT enabling interactive analysis of large sets of ROOT files in parallel. PROOF on Demand, PoD [3], is a set of utilities, which allows starting a PROOF cluster at user request on any resource management system. Installation is simple and doesn't require administrator privileges, and all the processes run in user space. PoD gives users, who don't have a centrally-administrated static PROOF cluster at their institute, the possibility to enjoy the full power of interactive analysis with PROOF.

Overview

PoD is a set of utilities, scripts, and configuration files. It provides a possibility to set up a PROOF cluster dynamically on Grid and batch systems. It consists of the following main components:

- pod-agent, a lightweight, standalone C++ daemon. Responsible to control the cluster of PROOF workers and optionally acts as package forwarder when there is a firewall between server and workers. Also pod-agent has a number of additional useful features which help to start, to process, and to control a PROOF interactive analysis.

- pod-console, a standalone C++ application. It provides a GUI and aims to simplify the usage of pod-agent and PoD configuration files. It is a PoD management console.

- PoDWorker script, a generic job script, executed on remote machines.

- PoD utilities, these are default job scripts for plug-ins, a number of configuration files and helper utilities.

Use Case

The main use case of PoD is to set up a distributed PROOF cluster on the Grid and/or local resource management system. The first thing, a user has to do, is to start the server side processes on a central machine; in terms of PoD it is a user interface machine. The next step is to submit PoD jobs to worker nodes. As soon as a job arrives at a remote worker node, it automatically configures the environment and starts all needed client services including an xrootd worker and a pod-agent in client mode. In case the pod-console is used as session management tool, each new

connection is immediately reflected in the GUI. When the instantiated PROOF workers of all submitted PoD jobs are connected, or when the user is satisfied with the number of connected worker processes, the PROOF analysis can be processed as if on a native PROOF cluster. The user then starts a ROOT session, e.g. on the private laptop, and connects to the PROOF master, registers the data, and runs the analysis script. Since PROOFAgent can manage disconnects, the user can also disconnect from ROOT, restart the ROOT session, and reconnect to the same PROOF cluster without having to resubmit the PoD jobs.

Summary

The main features of PoD:

- One click installation. The process of installation of PoD is fully automated and in user space.

- Easy to use. PoD provides a simple graphics user interface in order to simplify access to its functionality.

- It works out of the box. The PoD distribution contains everything users need to just immediately start using PoD right after the installation.

- Supports different job submission front-ends. Pod-console supports job manager plug-ins, therefore it is very easy to extend PoD to support different kinds of resources. PoD currently supports gLite [4] and LSF [5]. In future versions more plug-ins will be implemented, such as SSH [6] and PBS [7].

- PoD supports reconnections. This is an essential feature of PoD. Whatever happens with the ROOT/PROOF analysis sessions, even a crash, users don't need to resubmit PoD jobs. PoD will automatically refresh the environment and reconnect its workers.

References

[1] An Object Oriented Data Analysis Framework (ROOT) http://root.cern.ch

[2] The Parallel ROOT Facility (PROOF) http://root.cern.ch/drupal/content/proof

[3] PROOF on Demand (PoD) http://pod.gsi.de

[4] gLite http://glite.web.cern.ch/glite/

[5] Load Sharing Facility (LSF) http://www.platform.com/

[6] Secure Shell (SSH) http://www.openssh.com/

[7] Portable Batch System (PBS) http://www.pbsgridworks.com