

PROOF on Demand*

P. Malzacher, A. Manafiov

GSI, Darmstadt, Germany

PROOF [1] is an extension of ROOT [2] enabling interactive analysis of large sets of files in parallel on clusters of computers. Normally users get PROOF provided by administrators as a pre-installed shared cluster. To avoid certain disadvantages of “static” PROOF clusters PoD [3] has been developed.

PROOF on Demand is a tool-set, which dynamically sets up a PROOF cluster at a user request, on any resource management system (RMS).

PoD is a user oriented product with an easy to use GUI and a command-line interface. It is fully automated, and no administrative privileges, special knowledge or pre-configured nodes are required to use it.

Features

- **Easy to use.** The process of installation is very simple and fully automated. PoD works out of the box. Its distribution contains preconfigured modules and everything users need to just immediately start to work with it right after the installation.
- **GUI & Command-line.** PoD provides a simple and intuitive graphics user interface in order to simplify access to its functionality. For user’s convenience there is also a command line interface, it helps to manage a PoD cluster remotely or use it in a batch mode.
- **Native PROOF connections.** Whenever possible, PoD setups direct PROOF connections between nodes. It results in a full functional PROOF cluster. Users get native speed and the whole range of PROOF features. To use native connections an incoming traffic must be allowed on PoD workers for a defined port. Otherwise PoD uses packet-forwarding algorithms.
- **Packet-forwarding.** When worker nodes are behind a firewall then PoD uses its packet-forwarding algorithms to maintain the PROOF traffic. The algorithms are very efficient, there will be no speed penalty, but some PROOF functions are limited.
- **Multiuser/-core environment.** PoD implements automatic port mapping algorithms to properly handle cases when several users start PoD instances (servers/workers) on the same machine. PoD also automatically manages situations when multiple PoD workers are started on the same node. Private PoD instances can't disturb each other.
- **Different job managers.** PoD supports different job managers via a plug-in system. It is a very easy to extend system. PoD is currently shipped with the following plug-ins: SSH, LSF (Load Sharing Facility), PBS Pro/OpenPBS/Torque (Portable Batch System), Grid Engine (Oracle/Sun Grid Engine), Condor, gLite.

Use case

The main use case of PoD is to set up a distributed PROOF cluster on the Grid and/or local RMS. 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 xproofd 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 pod-agent 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

As PoD matures as a product, it is used more and more as a standard for setting up dynamic PROOF clusters in many different institutions in HEP community. Additionally there are already several Cloud based installations, which use PoD as a PROOF cluster solution. With PoD there is no need to maintain a dedicated PROOF analysis facility. PoD users create themselves private dynamic PROOF clusters on general purpose batch farms, Grid or Cloud systems.

Upcoming versions of PoD are going to support an out-of-server user interface. Users will be able to select a remote computer acting as PoD server (PROOF master). In this case PoD UI will be just a lightweight control center and could run on different OS types. Also an AliEn plug-in is going to be developed in collaboration with the ALICE Offline team. This cooperation will help PoD to provide a fast interactive PROOF experience on the AliEn Grid.

References

- [1] The Parallel ROOT Facility (PROOF) <http://root.cern.ch/drupal/content/proof>
- [2] An Object Oriented Data Analysis Framework (ROOT) <http://root.cern.ch>
- [3] PROOF on Demand (PoD) <http://pod.gsi.de>