PROOF on Demand

A. Manafov, P. Malzacher

GSI, Darmstadt, Germany

PROOF on Demand (PoD) [1][2] is a tool-set that automats and dramatically simplifies the process of creating private PROOF [3] clusters.

PoD dynamically sets up a PROOF cluster at a user's request on any resource management system (RMS). It provides a plug-in based system, in order to use different job submission front-ends.

PoD plug-ins

Currently PoD supports the following plug-ins: gLite, LSF, PBS (PBSPro/OpenPBS/Torque), Grid Engine (OGE/SGE), Condor, LoadLeveler, and SSH plug-ins. PoD makes it possible just within a few seconds to get a private PROOF cluster on those systems. If there is no RMS, then the SSH plug-in can be used, which dynamically turns a bunch of machines to PROOF workers. The SSH plug-in has been widely improved during a 2012 development cycle.

In 2012 the plug-in set has been also extended to support PanDA [4 a distributed software system developed for the ATLAS experiment. The PanDA PoD plug-in is already used by ATLAST and CMS users.

Binary payload attachment

Recently a new major step in PoD development has been made. To simplify PoD work packages a multilayer binary payload attachment (BPL) has been introduced.

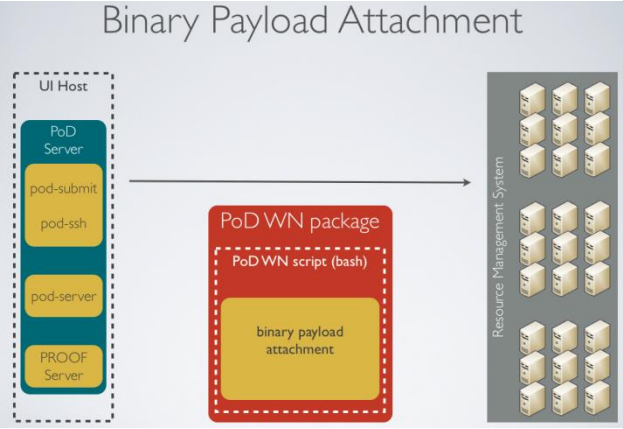

Figure 1: A multilayer binary payload attachment.

Now PoD doesn't require shared file systems on worker nodes or input sandboxes. There is only a single worker script, which is submitted to worker nodes and which contains everything PoD workers need, including precompiled binaries, configuration files, and user resources.

BPL also helped to simplify all plug-ins of PoD, as requirements for PoD workers were dramatically reduced.

User community

PoD is being very successfully used by German ALICE collaboration and has been chosen as a standard PROOF tool for FAIR computing [5].

The tool-set is also official distributed to all Tier3 ATLAS sites via standard ATLAS software deployment machinery [6].

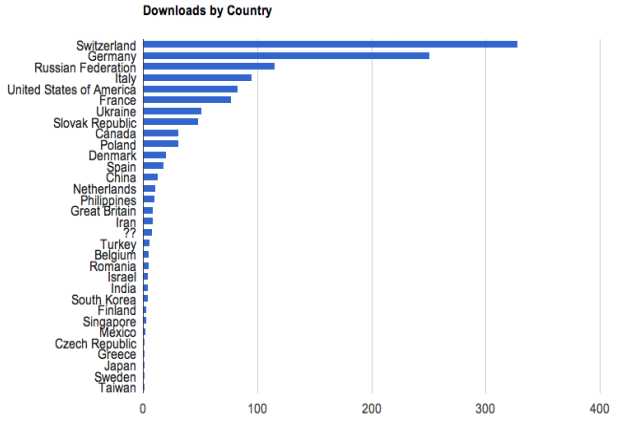

Figure 2: ~1300 downloads from 31 countries (since June, 2010), with 220 downloads/year from unique IPs. It's 20% more unique IPs and additionally 9 new countries than in 2011.

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 the HEP community. 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.

References

[1] A. Manafov et al, "PROOF on Demand", PHN-IS-IT-07, GSI Scientific Report 2010.

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

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

[4] ATLAS PanDA, https://twiki.cern.ch/twiki/bin/view/Atlas/Panda

[5] M. Al-Turany et al, "PROOF integration in FairRoot, GSI", Scientific Report 2011.

[6] ATLAS Wiki, https://twiki.atlas-canada.ca/bin/view/AtlasCanada/ManageTier3SW