Subject: Re: GlueX on CVMFS
From: Richard Jones <richard.t.jones@uconn.edu>
Date: 02/14/2018 04:18 PM
To: Mark Ito <marki@jlab.org>
CC: Sean Dobbs <sdobbs@fsu.edu>

Hello Mark and all,

I have the following packages installed on oasis (cvmfs) that mount automatically inside my singularity container whenever it starts up. I list below the name of the package and its disk footprint, followed by the number of versions currently installed. To estimate our needed footprint, we should come up with estimates for how many versions of each we will need, probably more than what I have.

- /cvmfs/oasis.opensciencegrid.org/gluex/amptools 7.4M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/ccdb 264M [2]
- /cvmfs/oasis.opensciencegrid.org/gluex/cernlib 125M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/Diracxx 8.4M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/evio 8.2M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/geant4 1.4G [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/gluex_root_analysis 2.7M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/group 365M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/hdds 41M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/hdgeant4 1.4G [2]
- /cvmfs/oasis.opensciencegrid.org/gluex/hd_utilities 861K [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/jana 499M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/rcdb 31M [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/resources 1.4G [*]
- /cvmfs/oasis.opensciencegrid.org/gluex/root 1.2G [1]
- /cvmfs/oasis.opensciencegrid.org/gluex/sim-recon 25G [2]
- /cvmfs/oasis.opensciencegrid.org/gluex/xerces-c 404M [1]

The resources directory contains a number of magnetic field maps and sqlite files that are compatible with installed releases of the software packages.

-Richard Jones

On Wed, Feb 14, 2018 at 3:13 PM, Mark Ito <marki@jlab.org> wrote:

> We have 17 versions of sim-recon on the disk now, going back to August of last year. We could get away with far fewer. See attached file for the list.
>
> On 02/14/2018 09:47 AM, Sean Dobbs wrote:

> Richard,
>
> Sure, I understand your organizational approach here. I think that we will want 4-5 sim-recon versions available to support the majority of analysis work. Maybe reducing the disk footprint needed is another argument for further subdividing sim-recon (although granted if not done smartly it could make the problem worse!).
>
> I'll let Mark describe how many versions are available on the group disk - my impression is that the available versions are limited by the available disk space, not what the actual working set is.
>
> ---Sean
>
> On Tue, Feb 13, 2018 at 10:47 PM Richard Jones <richard.t.jones@uconn.edu> wrote:
>> Sean and Mark,
>>
>> If we try to build everything inside a single container then any time a single line in a script changes, the whole container needs to be duplicated to support the new version. My singularity container on cvmfs is 1.1GB. The sim-recon and other software stack elements live in the oasis portion of cvmfs, where they are appear in a regular directory tree structure. Right now I have 2 versions of sim-recon, two versions of hdgeant4, one version of geant4lib and root, etc. The total footprint is 32GB. That could grow somewhat without raising a red flag, maybe to 50-75GB but eventually we would have to justify it. How many versions are you maintaining right now in the /group area on cue?
>>
>> -Richard Jones
>>
>> On Tue, Feb 13, 2018 at 5:22 PM, Sean Dobbs <sdobbs@fsu.edu> wrote:
>>> Hi Richard,
>>>
>>> I was thinking about some of your comments in the container discussion on Friday, and I think you have the right idea about container size, but in regards to distribution over CVMFS, I was wondering how many sim-recon versions can be supported at once in the current system? We're getting to the point where people may be using several fixed versions of the software to analyze different data sets (e.g., 2016 vs. 2017), and also to explore other detector configurations (e.g. the FCAL insert), so I wonder how aggressive we have to be in pruning the available set of software.
>>>
>>> Cheers,
>>> Sean