Rob,Eric,Sergey,Taylor,Doug,Danila,Tadashi

https://indico.cern.ch/event/637437/

Todo:
Send email for weekly meeting scheduling, add Wei and Xin Zhao (BNL)
Doug & Rob will follow up on Stratum-R

Danila & Titan (notes / questions by Rob)
- Danila's notes:
  https://docs.google.com/document/d/1N_pvCh_UxxZT82cbmQdtqPuyUrJFS4Sxj9e5Pnl
  XNaw/edit#
- Wondering where this software is kept? Who are the developers?
- role of the DTN node and the MultiJobPilot - does all the heavy I/O
- Lots of resources but for short periods --> suggests using Yoda and the Event Service;
  even w/ 10 minutes you can get huge numbers of events processed
- Can we see an architecture of the Titan "site" - showing the login node, the DTN node,
  etc?
- Is the only submission via MPI? What is the local scheduler?
- What sorts of monitoring tools are available? E.g. number of jobs running, etc.
- What decides (and based on what) how many pilots to submit to the DTN node?
- How is Harvester replacing these capabilities?
- Sergei: see CHEP presentations
- Software distribution to Titan
  - ATLAS software:
    - Pacman previously
    - Ayum setup, software dependencies a problem due to RHEL7
      - Are these officially supported by ATLAS?
      - build at cern, tarball, untar on Titan
    - Yes, works, but hate we have yet another method
    - So, someone has to install all production releases and updates by hand,
      on all the Titan worker nodes?
    - Is the software being installed into a shared file system, visible to all the
      nodes, or is it installed specifically on each Titan worker?
    - Grrrr -- alternative linux tools
  - Grid tools:
    - RHEL7.x, very little software installed
    - gfal libs w/ dependencies
    - yum download of rpms in user space
    - wrote custom setup scripts to get the environment setup correctly
    - voms-client w/ dependencies, no big deal
    - Each of these by hand

- - - Asked Horst for packages on Lucille. 150 packages, so won't do it by hand. Took 2,3 days to install everything
    - Rucio probs - pip install, gets messy
    - RG SUGGESTED ACTION ITEM: lets standardize on this crap, automate
    - (sorry, had to step away) -- Sergey: discussing some kind of install tool on Vaho's HPC page. Where is this? What is this?
      - really hate having yet another install method

Taylor's talk (notes/questions by Rob)
- https://indico.cern.ch/event/637437/contributions/2582104/attachments/1454483/2245320/2017-05-05.USATLASHPCFacilities.pdf
- ALRB w/ relocatable CVMFS path
  - please no Pacman!
  - you tell it where the CVMFS base is going to be
  - everything works like a normal grid site
  - Can this work on Titan?
- Lost -- ALRB for grid tools, but Ayum for releases?
- Doug is writing Rucio plugins for Globus. Want in the long term a Globus RSE
- (note - I am still not quite following how ALRB and the usual release will be handled)

Doug - StratumR
- Go to slide 7 of the first talk
- Hosted by MWT2 - entire CVMFS gets unpacked
- Globus transfers to Blue Waters
  - there was a symlink issue - Dave corrects this after the fact
- At Bluewaters we use containers (shifter)
  - looks like /cvmfs
  - Question is how to handle relocatable CVMFS
  - Would the releases break
  - Strength is all releases are handled automatically, and its Globus based which is used by the large HPC centers
- Need to consider the effort needed to support the HPC centers

Rob's questions:
- What is the state of Harvester deployment and development?
- How does Danila's tools at Titan fit into the future? Or are they replaced by Harvester?
- Documentation of deployment for Harvester?
- Where is github harvester? https://github.com/PanDAWMS/panda-harvester
- Are there milestones for the Harvester project? Tadashi does have a GANT Chart, see WFMS meetings:
- 20M hours on Theta - expire end of september
  - cannot run pilots - no outbound connectivity
  - q: is there a standard
-

Harvester Git Management for co-development

Doug: Should we setup a dev branch in git for Harvester so that when one facility uploads a change, they notify other facilities that they need to test the change to ensure it doesn't break local functionality.

Quote from Tadashi - "It is good to use development branches. Actually Danila is already working on his own development branch. I think that you and Taylor also have privilege to make branches in the repository. Concerning the workflow I would prefer the Feature Branch Workflow (GitHub flow) to the Gitflow Workflow since the continuous delivery and deployment model has been used for panda/jedi where there isn't the concept of 'release' and it has been working well. Actually using the agile development model is one of the main recommendations in the WFMS/DDM review. I know that the development environment for harvester is different since it involves more people, but I suppose that the GitHub flow works for harvester as well since larger projects such as github itself are using the workflow and harvester has a plugin structure. It is impossible to break HPCs by adding changes to grid plugins. Of course we should be very careful when adding changes to harvester core, but that is a must regardless of workflow. If multiple developers work on the same plugin for some resources they should know what others are doing and what are requirements for those resources, where weekly WFMS meeting will help. Basically all changes to the master branch should be deployable. If some changes break backward compatibles we should inform in WFMS meetings as well."

https://www.atlassian.com/git/tutorials/comparing-workflows#feature-branch-workflow