Migration to ARC experience

M. Svatoš, J. Chudoba, P. Vokáč

NorduGrid 2019

11-14.6.2019

Introduction

- motivation of migration
  - previous batch system (Torque+Maui) not scaling well
  - HTCondor replacement needed CE which supports it well
  - ARC-CE recommended
    `https://twiki.cern.ch/twiki/bin/view/AtlasComputing/SitesSetupAndConfiguration#Computing_Element`
- CREAM-CE+Torque replaced by ARC-CE+HTCondor
- additional machines added to submit jobs to remote HPC
- version 6 of ARC-CE is being tested

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

ARC-CE:Local batch system submission

- two ARC-CE machines
- in production since the beginning of 2018
- computing farm
  - about 9k cores
  - fairshare between ATLAS, ALICE, Auger, CTA, Dune, etc.
  - on CentOS7
- HTCondor batch system, version 8.8.2
  - better container support
  - munge authentization
  - option for job submission warning
  - better admin defined error messages with *_REASON (SUBMIT_REQUIREMENT_*_REASON, SYSTEM_PERIODIC_REMOVE_REASON)

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

ARC-CE:Local batch system submission

- ARC-CE, version 5.4.3
  - minimalist configuration
  - authorization using ARGUS
  - only one production ARC-CE queue (+ some testing queues)

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

ARC-CE:Local batch system submission

Issues, improvements:

- republishing tool to APEL (publishing done by the HTCondor instead)
- ALICE support
  - ALICE is using deprecated BDII GLUE 1 schema → after every update, ARC-CE needs to be patched again to support ALICE
  - `https://bugzilla.nordugrid.org/show_bug.cgi?id=3544`
  - `perl` scripts sometimes have undefined variables causing data not to be published to infosys
- HTCondor support
  - e.g. information come from parsing of `condor_q` output instead of usage of condor API (it happened around 8.5.7; format of `condor_q` can change)
  - `https://source.coderefinery.org/nordugrid/arc/issues/61`
- inode problems
  - extreme number of files in `/var/spool/log/jobstatus`

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

ARC-CE:Local batch system submission

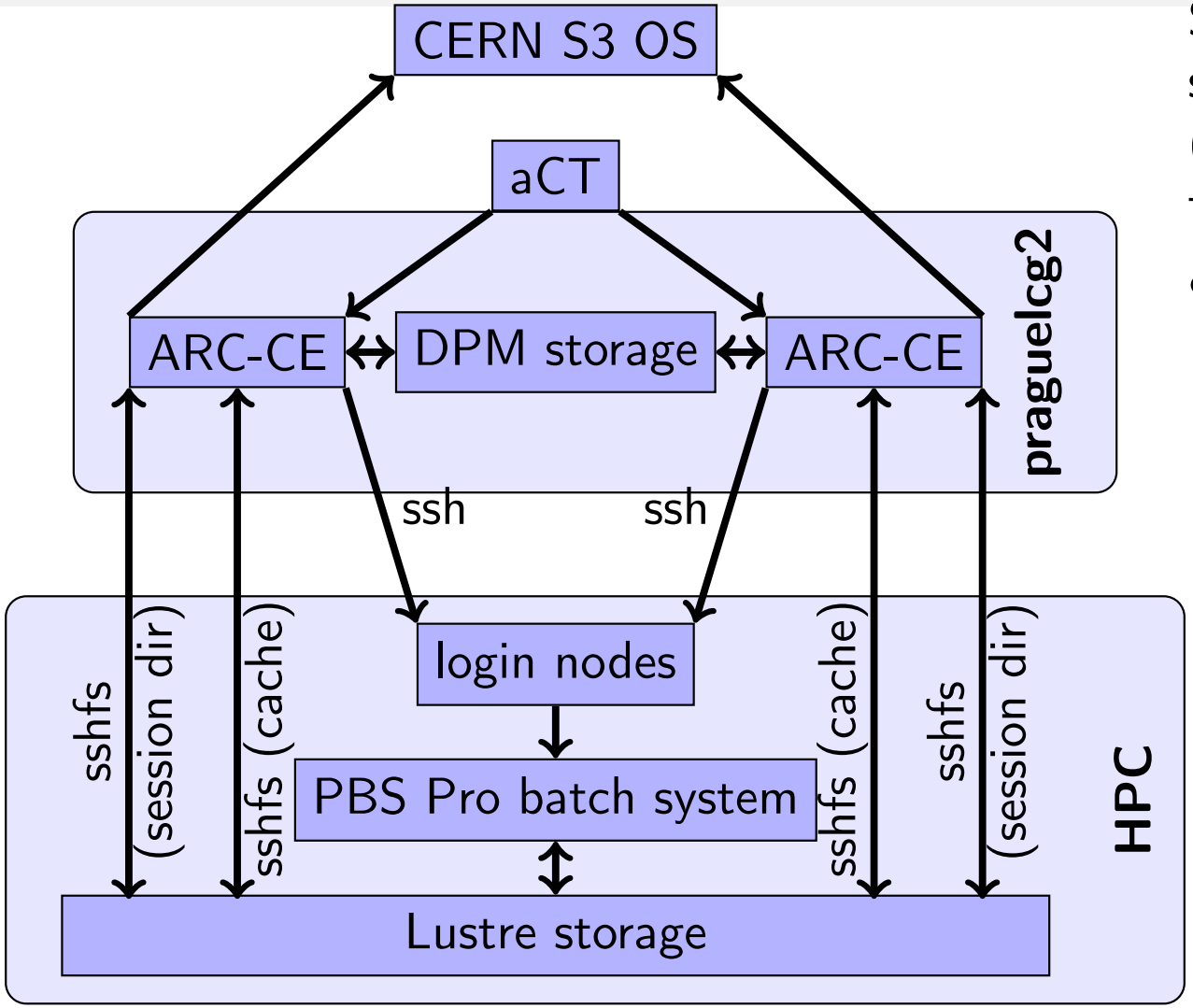
Issues, improvements:

- how to fix inconsistent data when disk runs out of free space
- how to disable automatic job resubmission
  - problem for ATLAS grid jobs
  - `https://bugzilla.nordugrid.org/show_bug.cgi?id=3799`

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

ARC-CE:Remote HPC submission



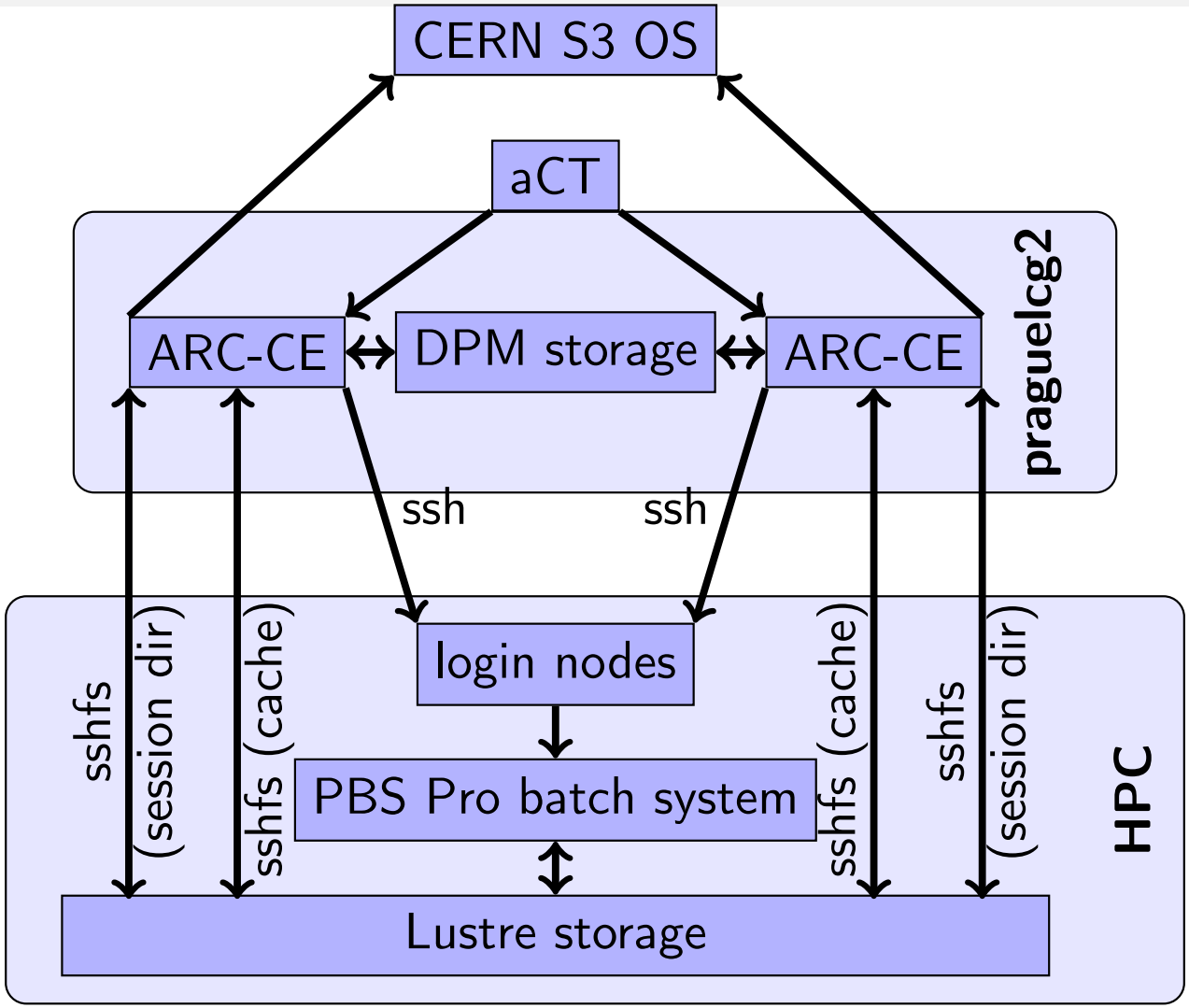
Salomon HPC job submission workflow (in production since the end of 2017):

- the ARC Control Tower (aCT) submits job description into one of the ARC-CE machines located at the computing center of the Institute of Physics of the Czech Academy of Sciences

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



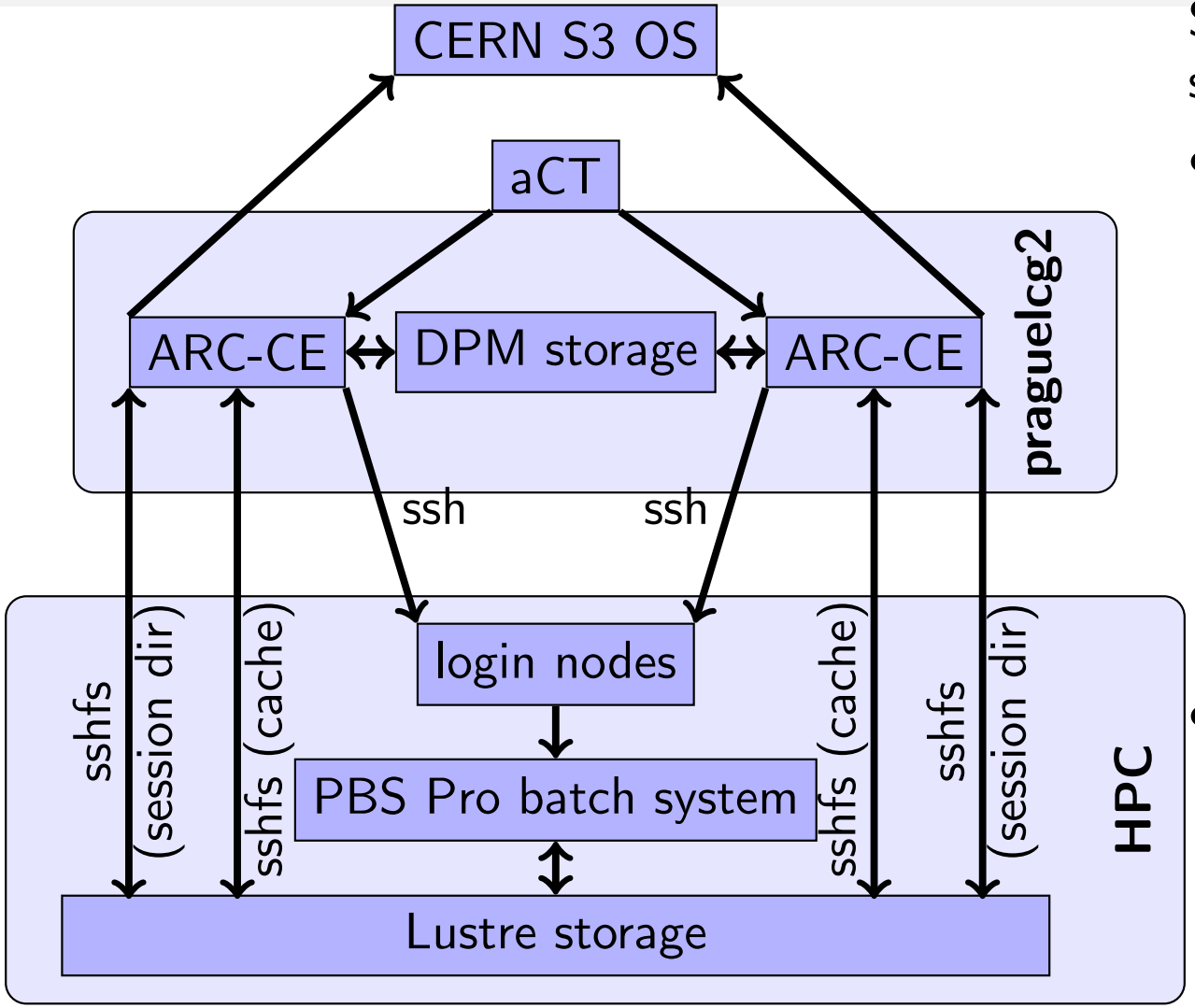
Salomon HPC job submission workflow:

- the ARC-CE translates the job description into a PBS script
- the ARC-CE puts necessary scripts into the session directory which is shared with scratch on Salomon via sshfs

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



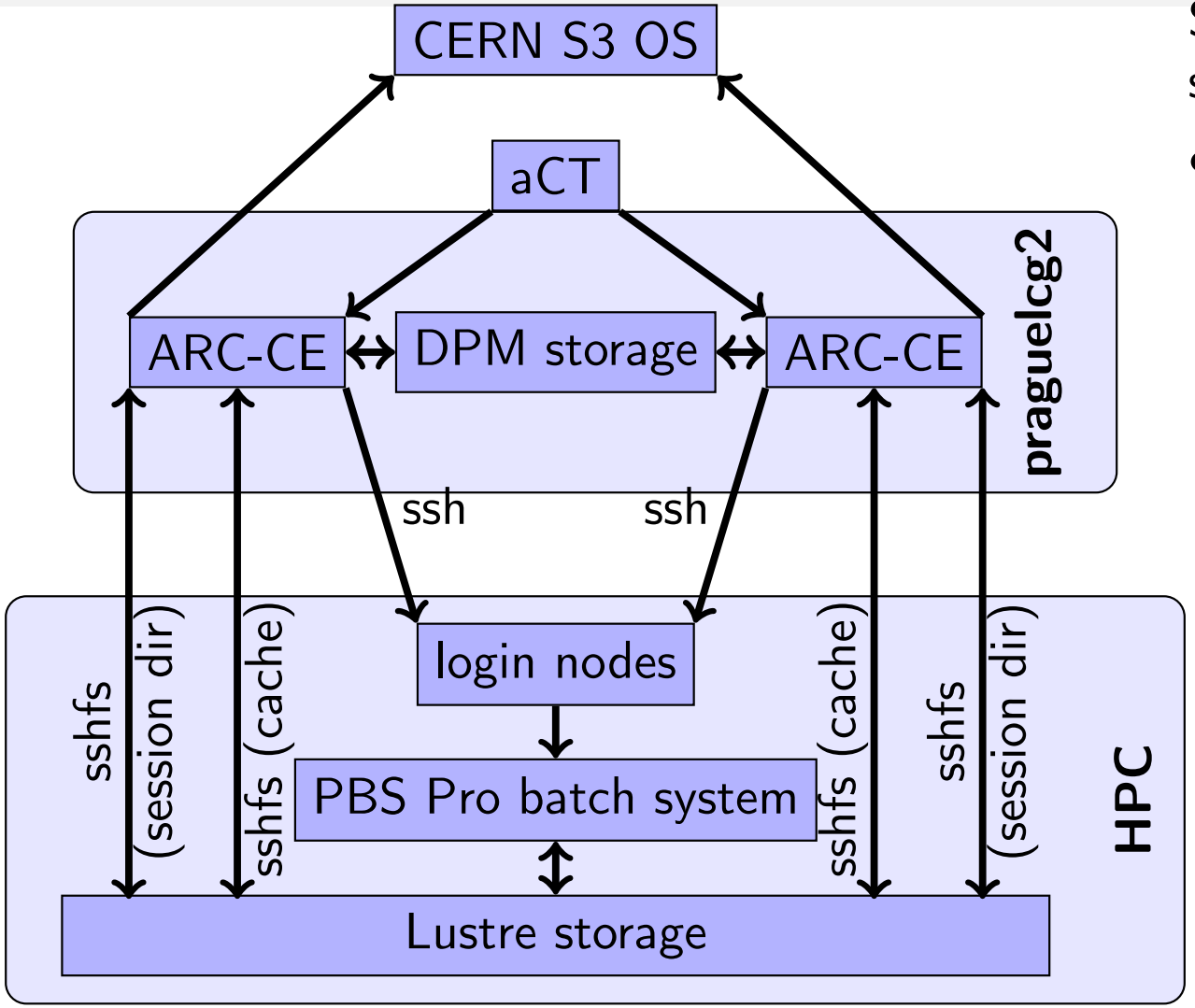
Salomon HPC job submission workflow:

- the ARC-CE gets input files - either it links them to the session dir from a cache dir (also on the scratch of Salomon) or copies them there from local DPM storage
- the ARC-CE submits a job to the PBS via ssh connection to login node

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



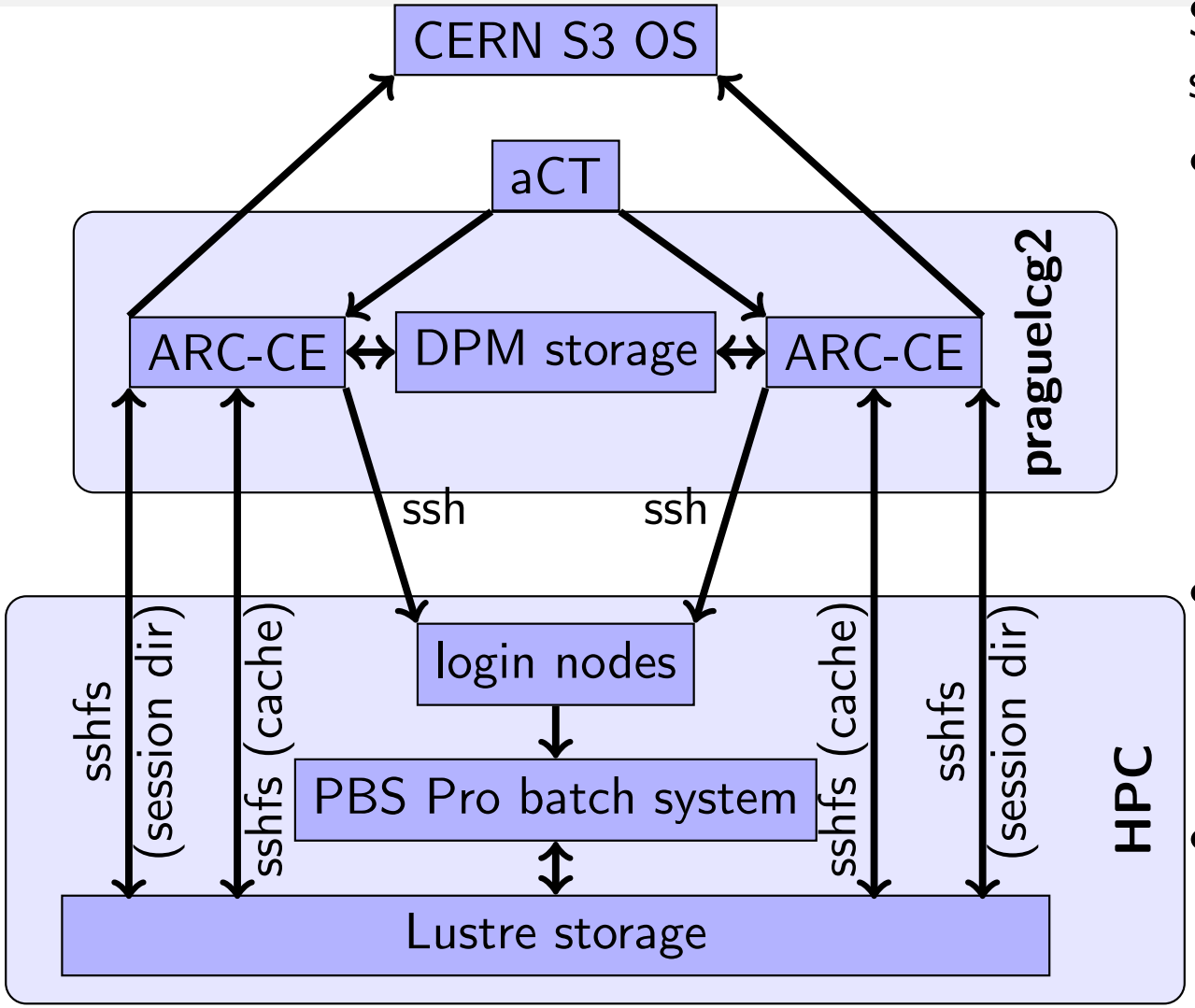
Salomon HPC job submission workflow:

- running job uses software stored on scratch
  - the number of jobs limit on Salomon is 100 jobs per user; each ARC-CE submits as a different user

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research, Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



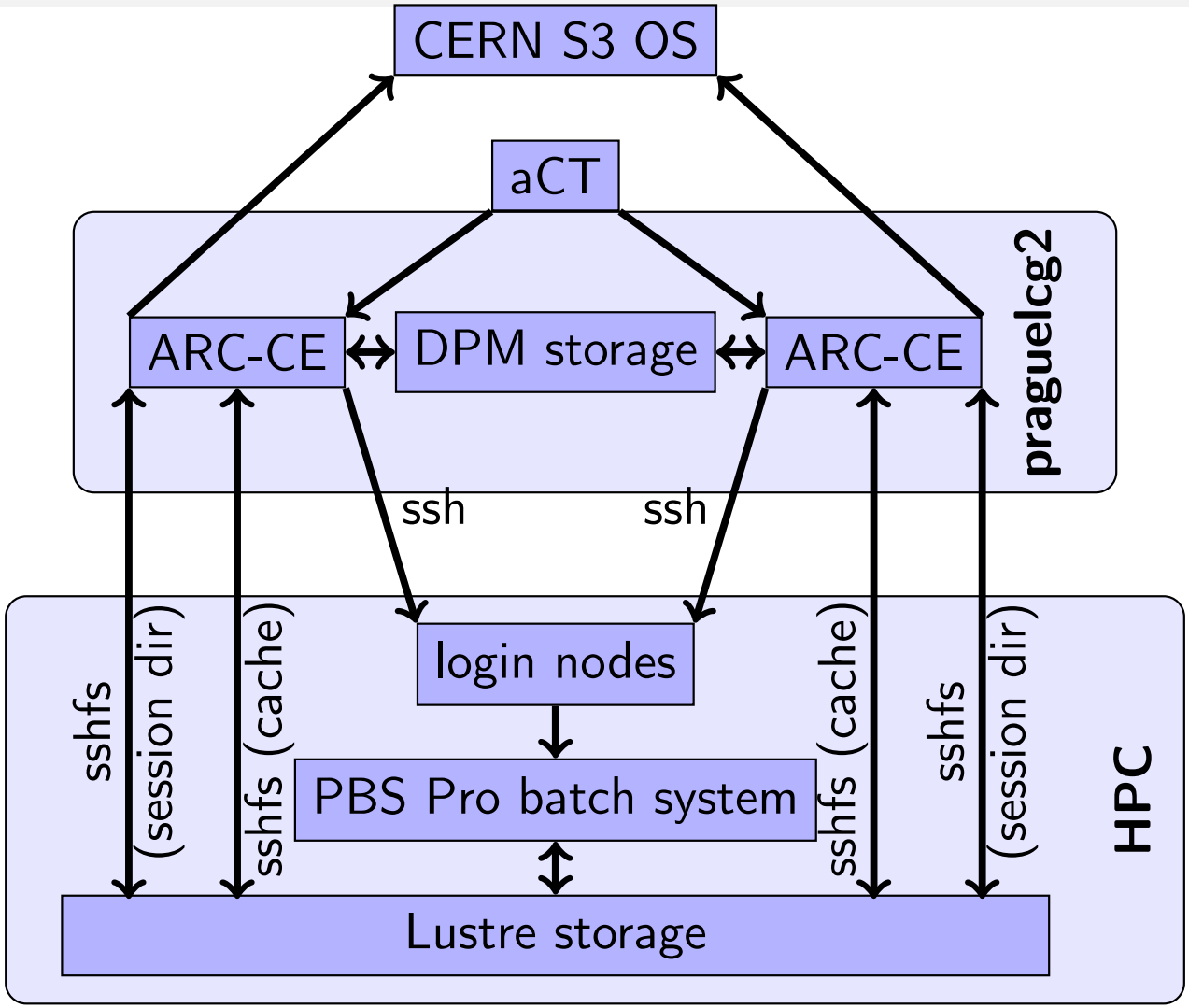
Salomon HPC job submission workflow:

- when the job finishes, the output and logs are located in the session dir and is accessed by the ARC-CE via sshfs
- the job output and log is copied to S3 Object Store in CERN in Geneva
- a copy of the log is stored in the local DPM storage

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



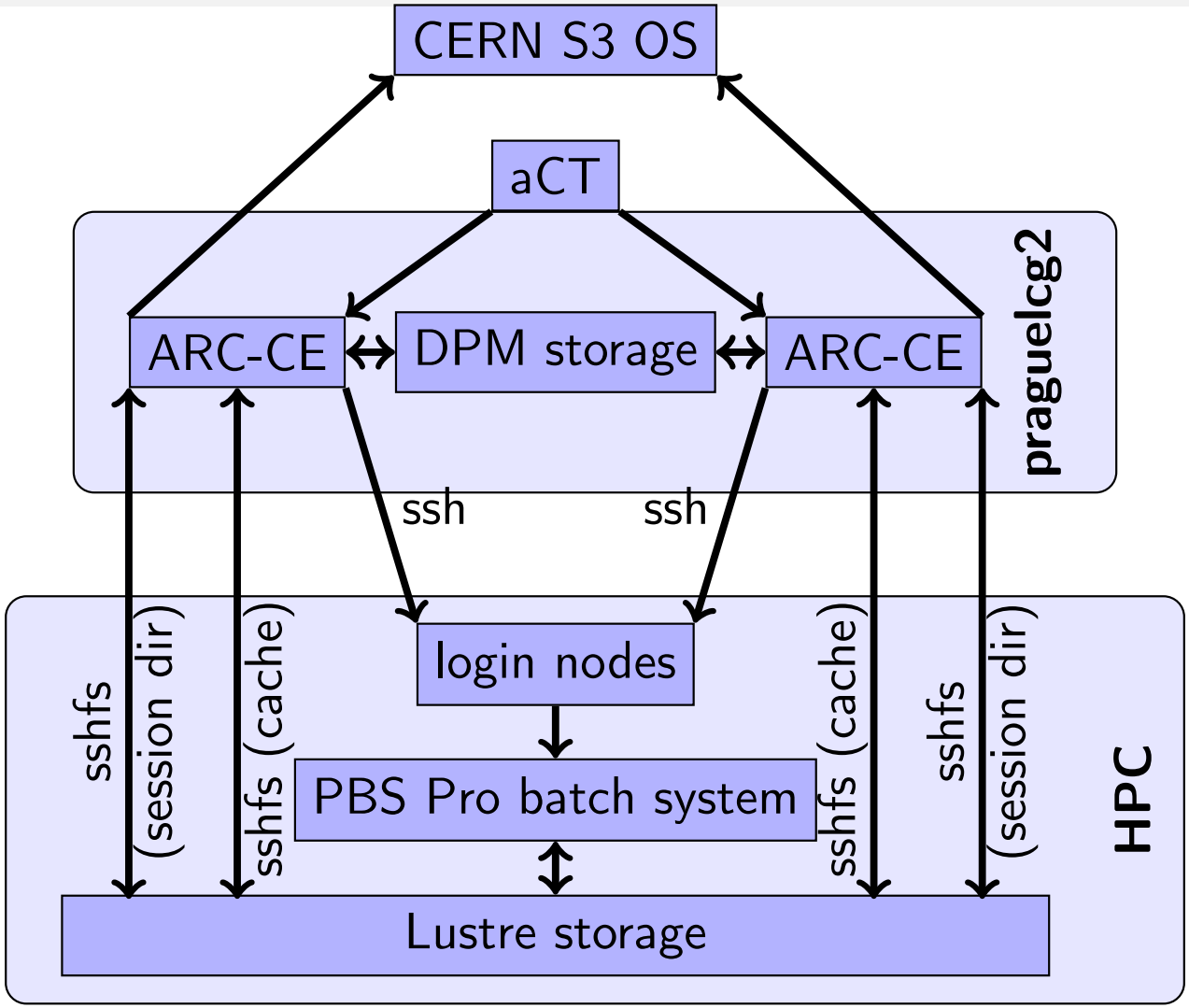
Issues, improvements:

- `submit-pbs-job` made for Torque but recent PBSpro has a different syntax
  - workaround: hard-coding HPC specific values
  - `https://bugzilla.nordugrid.org/show_bug.cgi?id=3696`

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research, Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



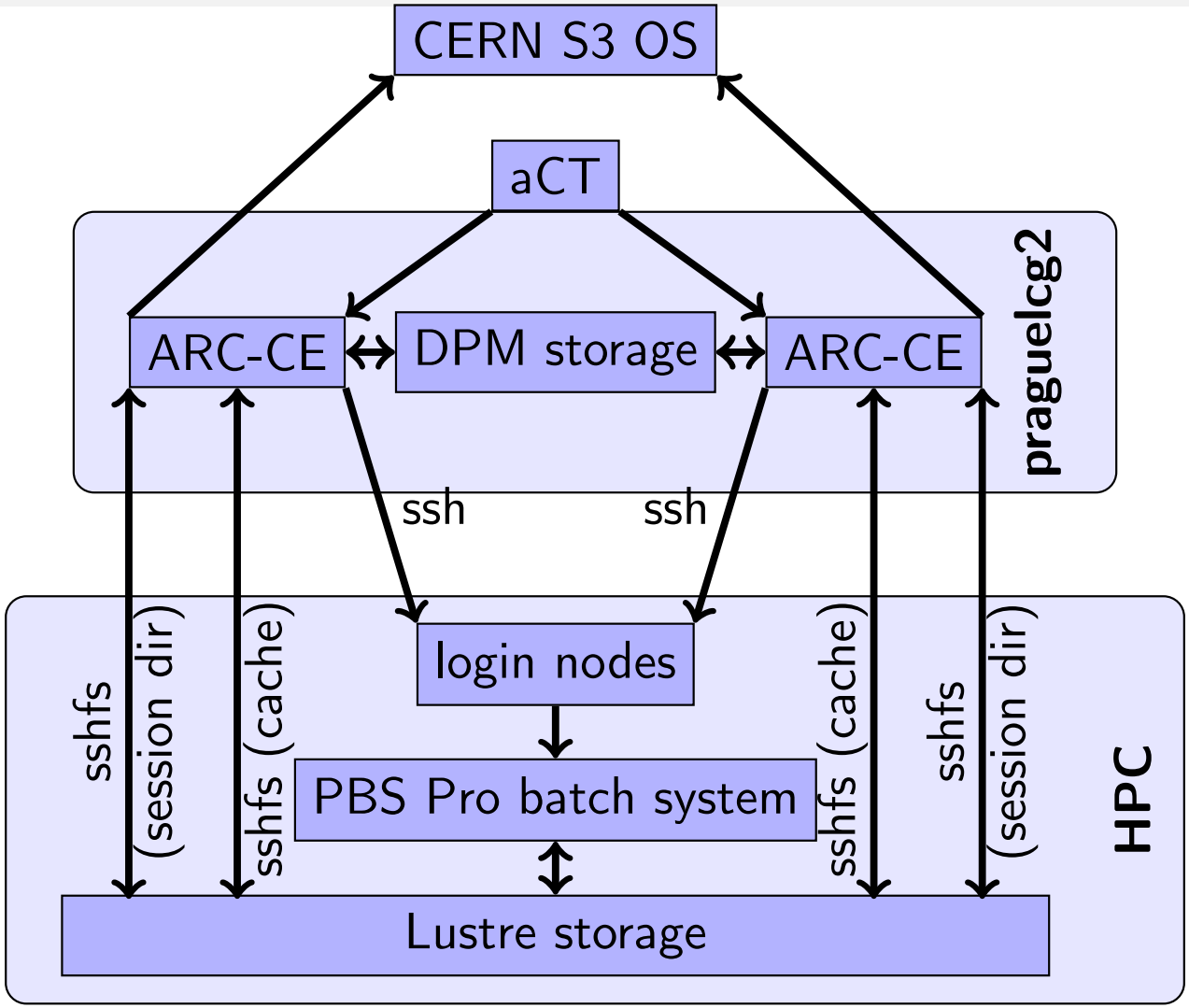
Issues, improvements:

- HPC has limit on number of interactions with PBS server and ARC-CE exceeds them (which leads to failed submission)
  - workaround: put sleep into `qstat`, `pbsnodes`, and `qmgr`

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research, Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC-CE:Remote HPC submission



Issues, improvements:

- values for EGI accounting of ATLAS EventService jobs are wrong – ?

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research, Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

ARC6 testing

- arc1.farm.particle.cz on version 6 since beginning of May
  - version 6.0rc5
  - in the beginning, jobs were failing because of misplaced proxy - `https://bugzilla.nordugrid.org/show_bug.cgi?id=3824`
  - first, a workaround implemented in HTCondor
  - later, ARC-CE script fixed locally
  - used for ATLAS, ALICE, Auger
- subjectively:
  - nicer deamon management interface (`arcctl`)
  - better RTE management
  - cleaner configuration

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Summary

- ARC-CE is configurable enough to be able to submit jobs to local batch system as well as to remote HPC
  - but there is a learning curve
- version 6 of the ARC-CE being tested
  - smooth running (except problem with proxy location)
- implementation using `python` API (when possible) instead of `perl` scripts would be nice

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS