ARC, plans and releases

Jon Kerr Nilsen
ARC release manager
NeIC/Univ. of Oslo

NordForsk

Nordic e-Infrastructure Collaboration

Outline

- What is ARC?
- What are we working on?
- How do we release it?

What is ARC?

- A middleware, an advanced tool for connecting resources, a software developed by NorduGrid...
- ARC is a set of products
  - ARC, ARC-docs, ARC-Nagios-plugins, CANL-C++
  - **Not** ARC CE, ARC core, ARC Clients, ARC InfoSys and ARC gridftp
    - (Just an EMI mix-up of products and product teams)
- Broad platform support
  - Services on RHEL/SL and Debian/Ubuntu
  - Clients on RHEL/SL, Debian/Ubuntu, MacOSX and Windows

What are we working on?

- Targeting HPC clusters
- ARC Control Tower
- Windows GUI
- Code improvements:
  - startup script cleanup
  - investigating the aggregated accounting feature for Jura
  - client-side error messages review
  - client scalability improvements (job.xml replaced with db)

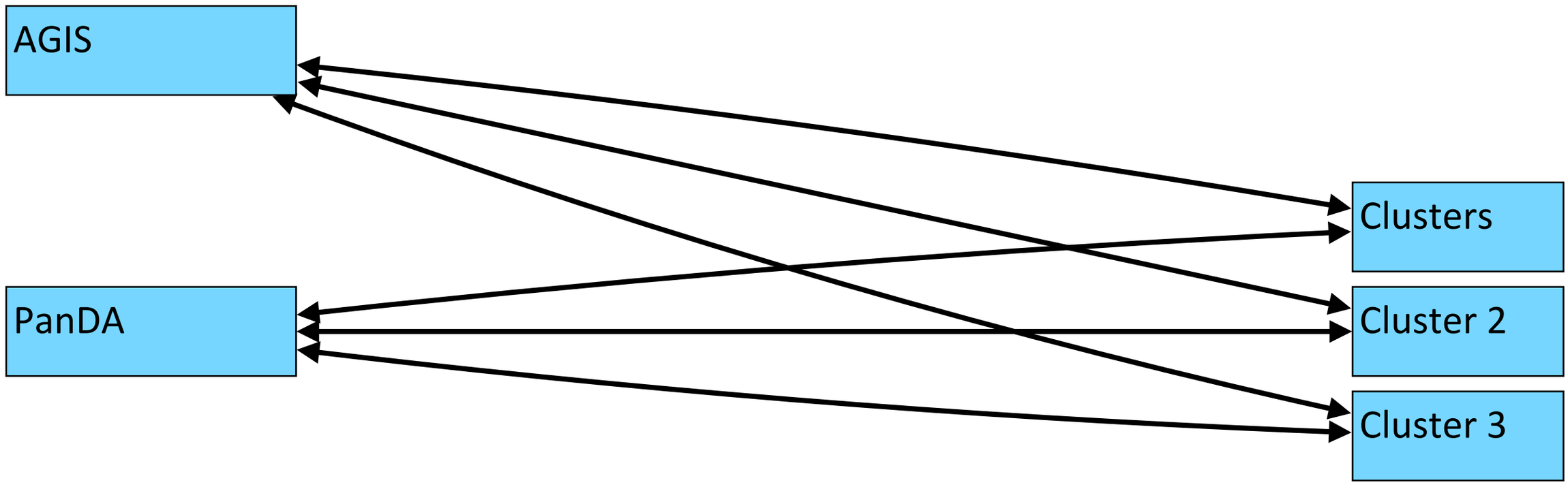
ARC on HPC in ATLAS

- Current HPCs:
  - Abel, NO
    - 800 cores guaranteed (T1 pledge)
    - 3.5k cores opportunistic
  - Abisko, SE
    - up to 2k cores opportunistic
- Future:
  - All WLCG SNIC/Sweden clusters will be decommissioned this year
  - T1/T2 pledges will be provided exclusively through HPC clusters
  - Room for extensive opportunistic usage
- Further future: US HPC resources?

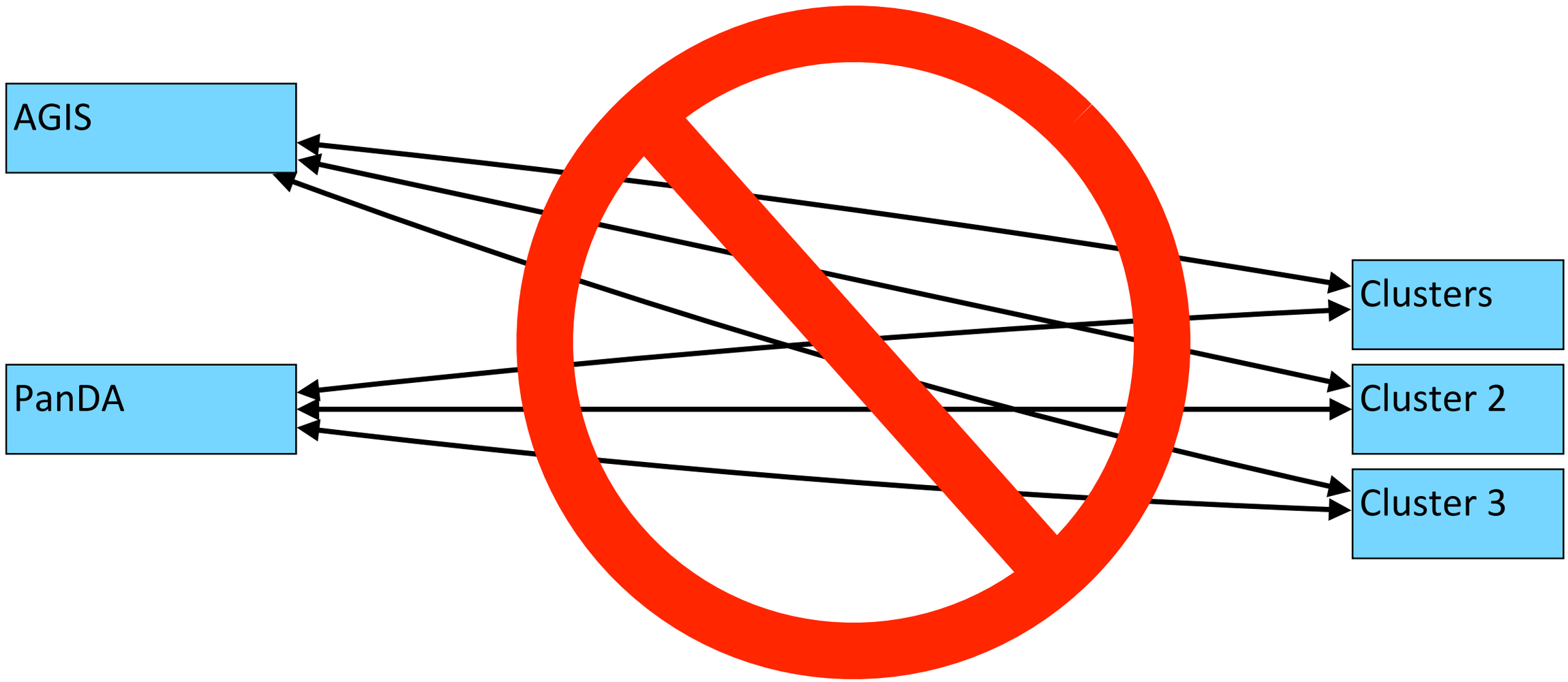
HPC requirements

- Inputs on shared FS, outputs to shared FS
- Transfer agent
- Limited worker node network access
- cvmfs partial syncing to shared FS
- Complex batch scheduling (multi-core, multi-node)
- Limited Grid frontend rights
- No or limited middleware on nodes
- In some cases, big batch jobs (tasks)
- Access/authorization - grid cert might not be enough, depends on site policies

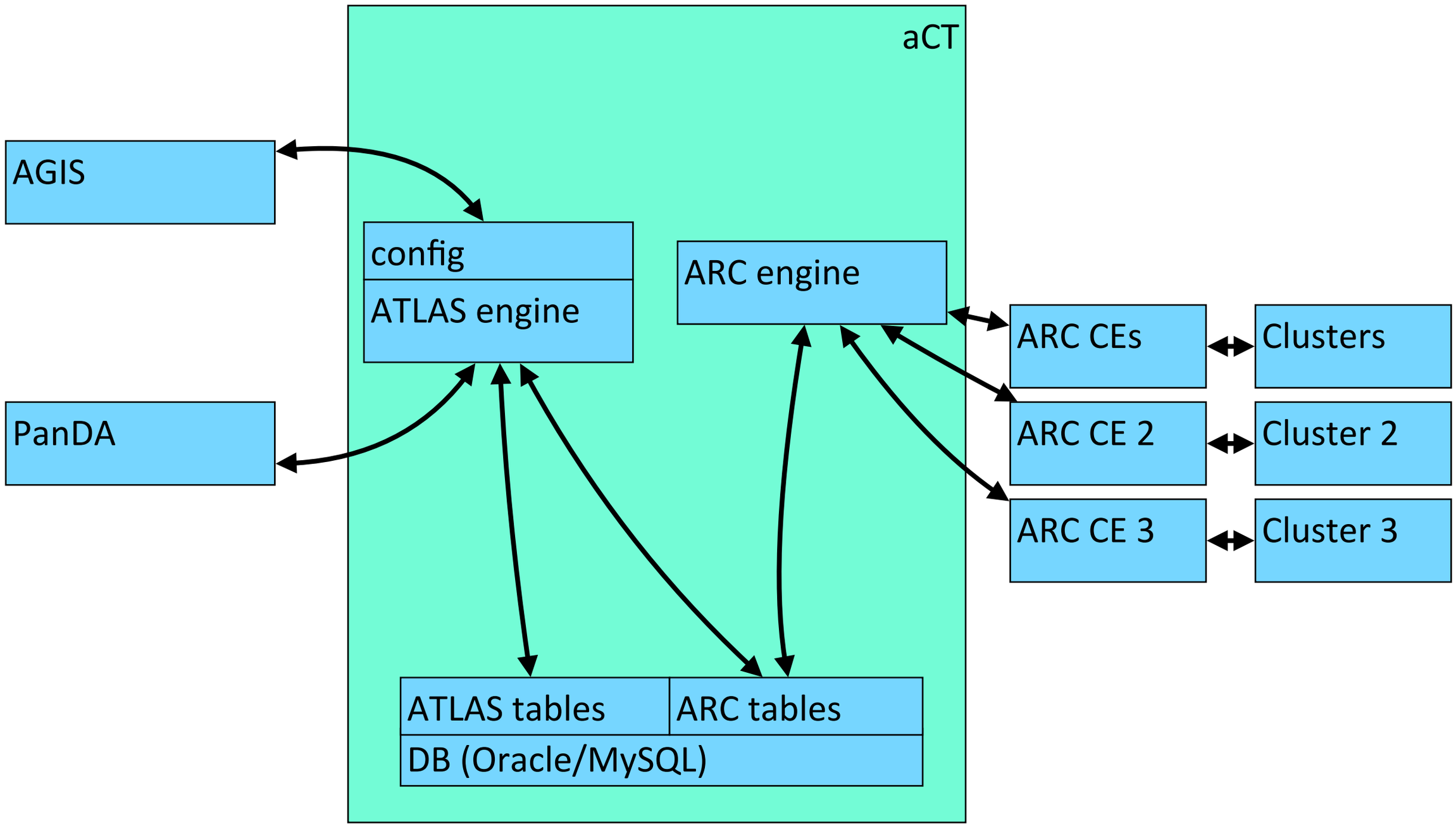
ARC Control Tower



ARC Control Tower

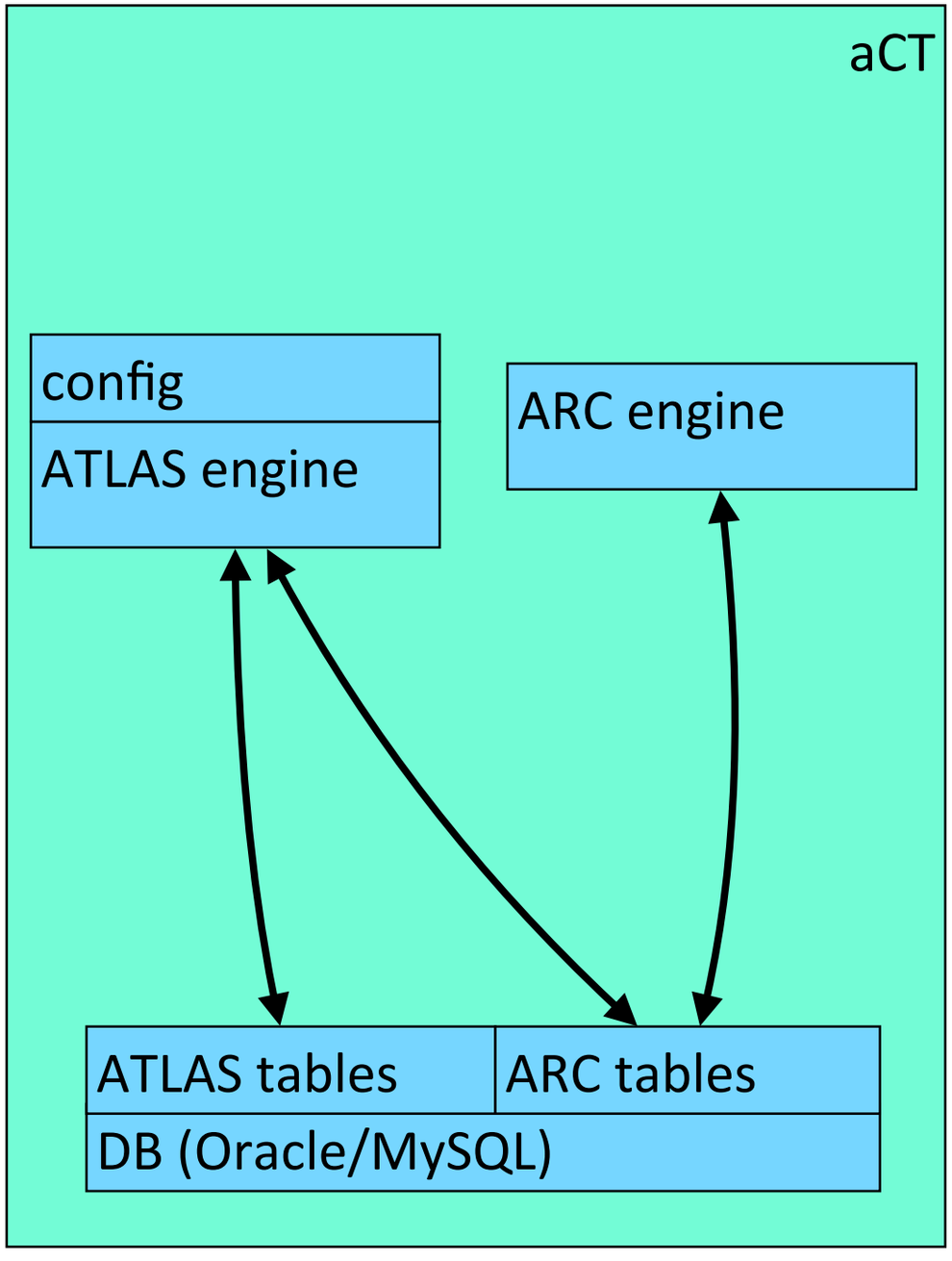


ARC Control Tower



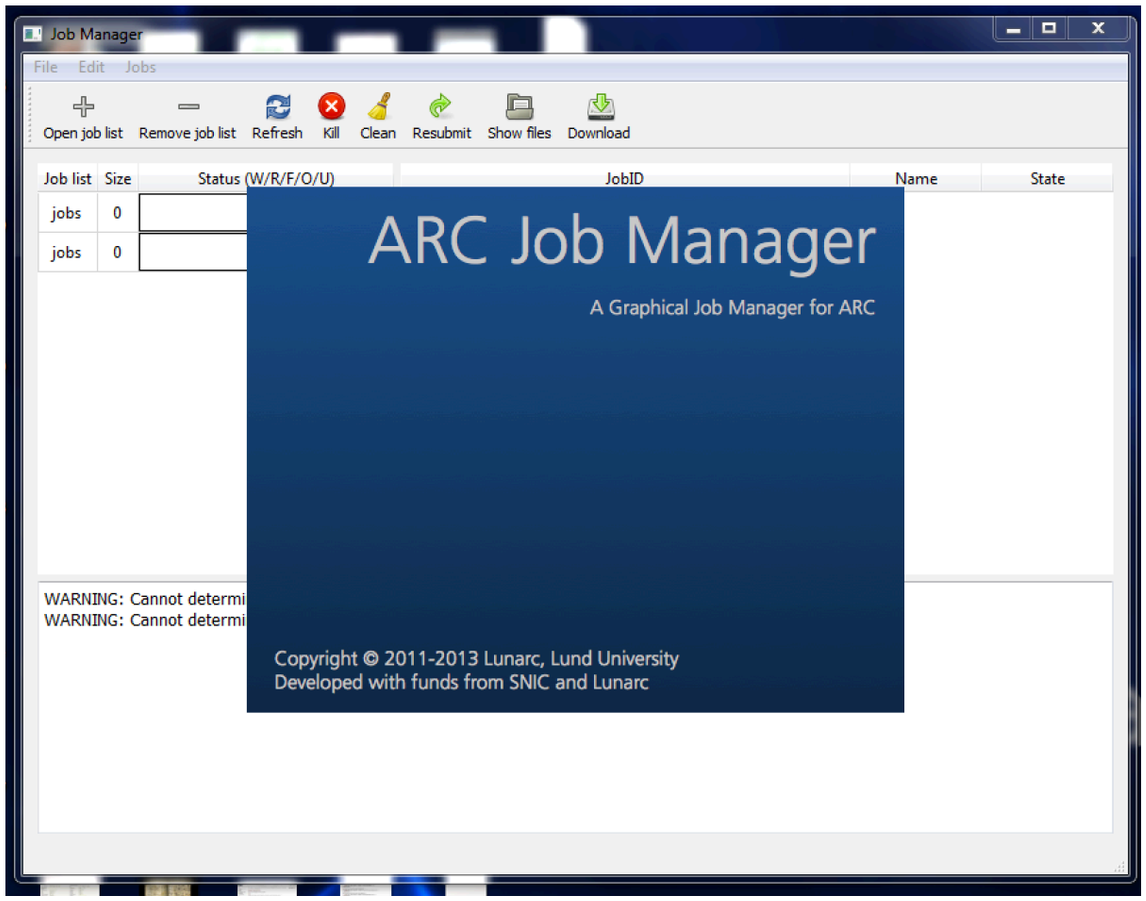
ARC Control Tower

- Major code clean-up
  - New DB engine, ARC part clearly separated, uses ARC 3.0 API instead of ARC 0.8 CLI, proper init scripts, separated source dir from run dir...
- aCT will run at CERN (or directly at HPC cluster if needed/ requested)
- Pulls jobs from PanDA
- Pushes jobs through ARC CEs to clusters
- No pilot jobs
  - but we run exactly the same pilotcode wrapper as the others, just in preloaded way and not network access



Windows GUI

- ARC clients on Windows existed for several years
- Not easy to convince Windows users of the virtues of command line tools
- Solution: ARC GUI
- Clean interface for managing jobs
- ARC Storage Explorer for drag-and-drop use of all grid stores
- NSS-enabled – can use firefox/ thunderbird credentials store
- ARC Storage Explorer fully functional, still some tweaks needed for job management



Other improvements

- startup script cleanup
  - current infosys startup scripts are tricky to get to work
  - planned for cleanup before next major release
- investigating the aggregated accounting feature for Jura
  - Jura fully support non-aggregated EMI CAR records
  - Aggregated accounting feature is under investigation (is it an EGI requirement?)
- client-side error messages review
  - For several technical reasons (propagation of third-party error messages), current client-side error messages can be a bit confusing
  - Extensive review of error messages is ongoing
- client scalability improvements (job.xml replaced with db)
  - Reading xml file containing info on large number of jobs is not optimal
  - DBs do the same job much more efficiently – replacement coming in next major release

How do we release it?

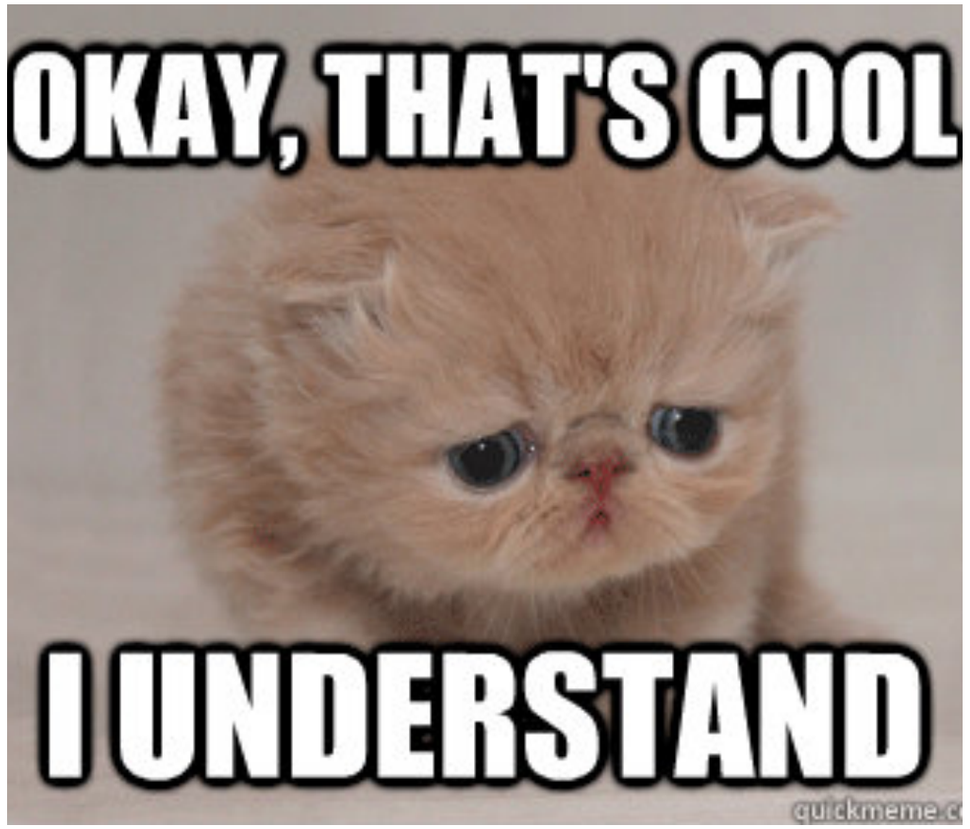
- ARC is developed in SVN
  - Developers submit to trunk
  - When it's time for release changes are ported to release branch
- Major release once a year or if API has changed in an important fix
- Minor/bugfix releases when needed
- When release manager is satisfied that all needed changes are ported (and branch still builds), release is tagged, branch built and submitted to various repositories
- Basic functional testing, deployment testing is run a couple of days before tagging
- If major release, a couple of release candidates might be needed

How do we release it?

- NorduGrid ARC release naming
  - Major releases are named NorduGrid ARC YY.MM (year and month of release)
    - Minor releases are numbered sequentially NorduGrid ARC YY.MMuX
    - Package versions are numbered the usual way (major-minor-bugfix)
      - Independent of release naming
  - Current release: NorduGrid ARC 13.02u3
  - Current packages: arc-3.0.3, arc-doc-1.3.3, arc-meta-1.0.7, arc-nagios-plugins-1.6.0, canl-c++-1.0.0
- Releases in NorduGrid repos, EPEL, Debian, EMI, UMD

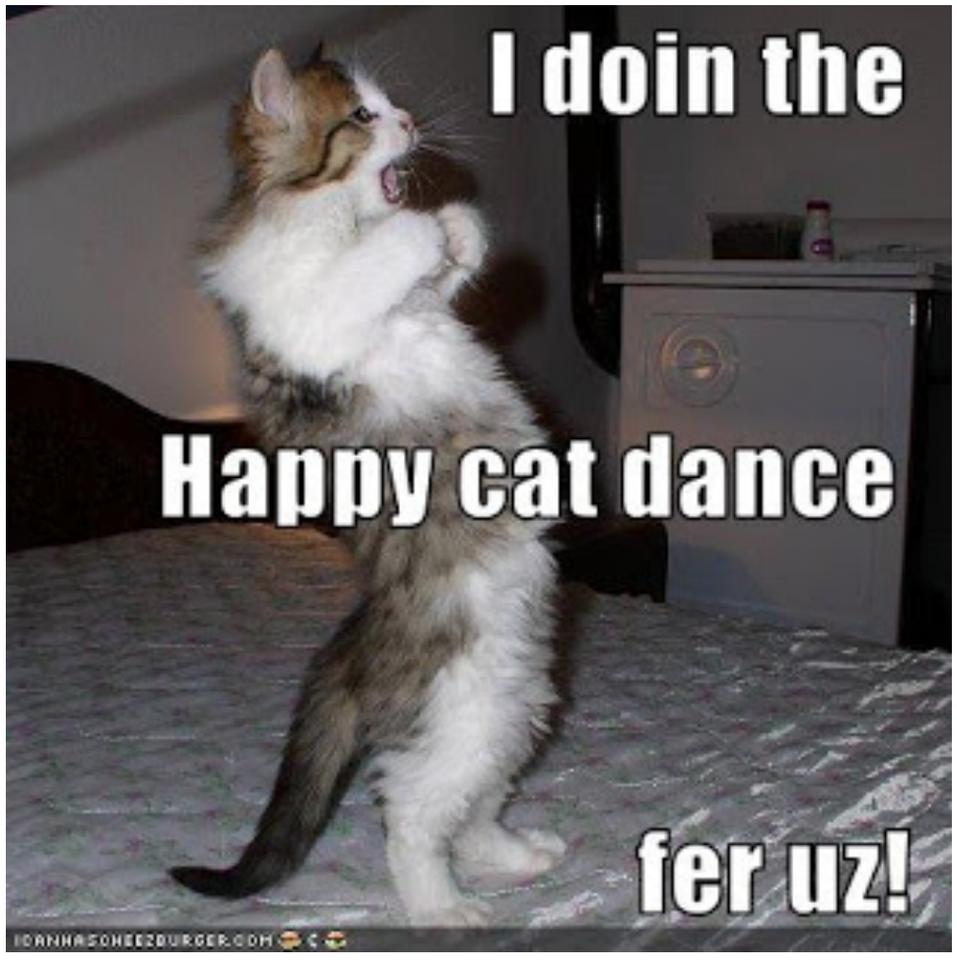
How do we release it? ARC and UMD - current procedure

- ARC is released to NG (NorduGrid) repos
- Then submitted to EPEL (and Debian)
- Then Cristina takes packages from NG and/or EPEL and puts in EMI repos
- and submits some xml thingy to UMD
- then testing is started within UMD
- and if any problems are found in UMD testing it's probably time for another ARC update release



How do we release it? ARC and UMD - preferred procedure

- ARC is released to NG (NorduGrid) repos
- Then submitted to EPEL (and Debian)
- new packages goes to EPEL testing for two weeks testing period
  - if no negative remarks within two weeks, packages go to EPEL stable
- UMD runs testing (by enabling EPEL testing and installing ARC) during EPEL testing period
  - Testers can give feedback through https://admin.fedoraproject.org/updates/FEDORA-EPEL-2013-10468
- If no negative remarks, packages are moved to EPEL stable and pulled to UMD repos
- Cristina no longer needs to submit strange xml forms to get ARC released in UMD



ARC and UMD preferred procedure

- What about meta-packages?
- EPEL does not support meta-packages
- but does support *groups* which is for ARC purposes just as good
- ARC meta-packages are currently released in NG repo
- Three possible solutions:
  - UMD adds support for groups as alternative to meta-packages
  - Meta-packages go through EMI to UMD (not Cristina-friendly)
  - UMD pulls meta-packages directly from NG repo (what about testing?)

Conclusions

- ARC is a set of products (just not the set defined by EMI)
- With aCT, ARC is moving even more towards HPC resources
- Windows GUI is coming soon
- ARC is released in many channels
  - Releases to NorduGrid repo first
  - Other repos should pull from EPEL
- UMD procedures can benefit from EPEL

Questions?

ARC Control Tower – What's new

- Major clean-up
  - current version evolved from Prof. Alex Read's project for learning Python programming
  - Andrej Filipcic took over development and babysitting
  - David Cameron and myself have taken over the ARC part of the development
- Clean separation of ATLAS part and ARC part
- New DB
  - current version uses sqlite – too slow, and not useable in production environment
  - Adds support for MySQL and soon Oracle – makes it possible to run multiple instances on multiple hosts
- Uses ARC 3.0 API
  - Current version uses ARC 0.8 CLI...