NORDUGRID

Grid Solution for Wide Area Computing and Data Handling

Overview of the ARC6 CE

EGI Clinic, compute services

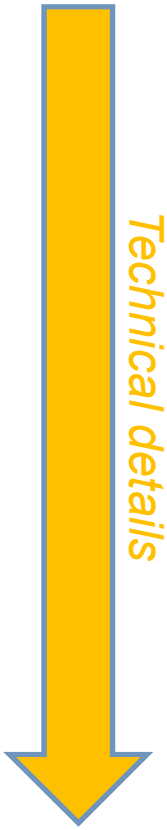
Balázs Kónya

on behalf of the ARC developer team

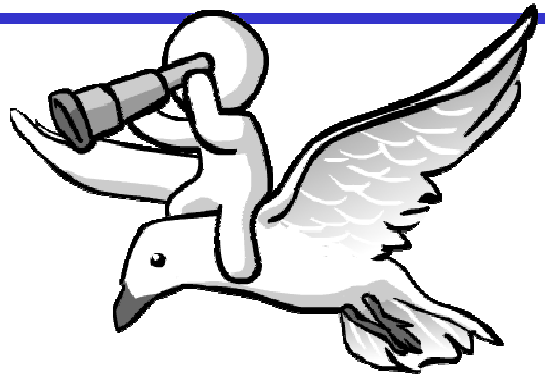
ARC6 clinic's program

1. Birds-eye view of an ARC6 CE

2. Q&A with the ARC dev team



Birds-eye-view



ARC is a „traditional CE”

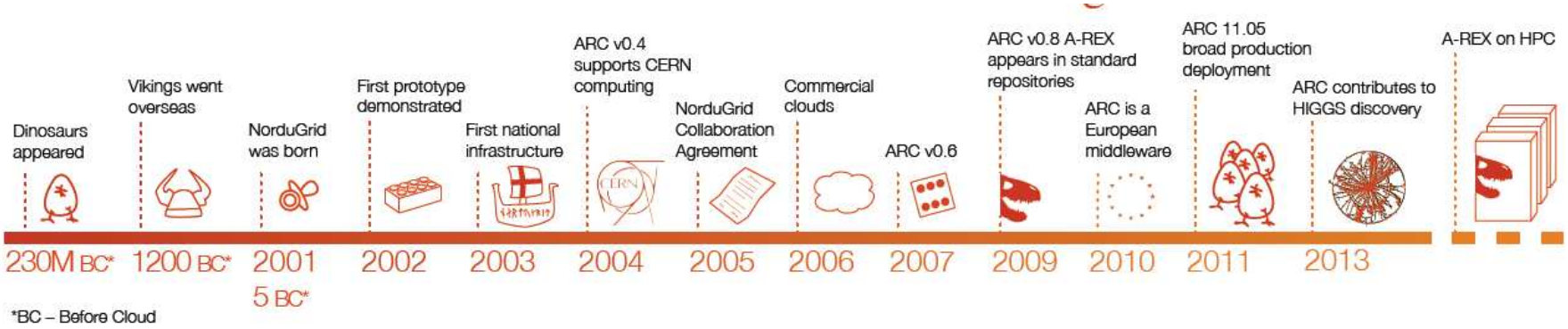
- a generic „CE system”
- one of those batch system gateways
- comes from the „NorduGrid”

When asking the sysadmins:

- *... it works*
- *... it is well-known*
- *... boring, there is nothing exciting around and ARC CE*

If the ARC CE were a car...

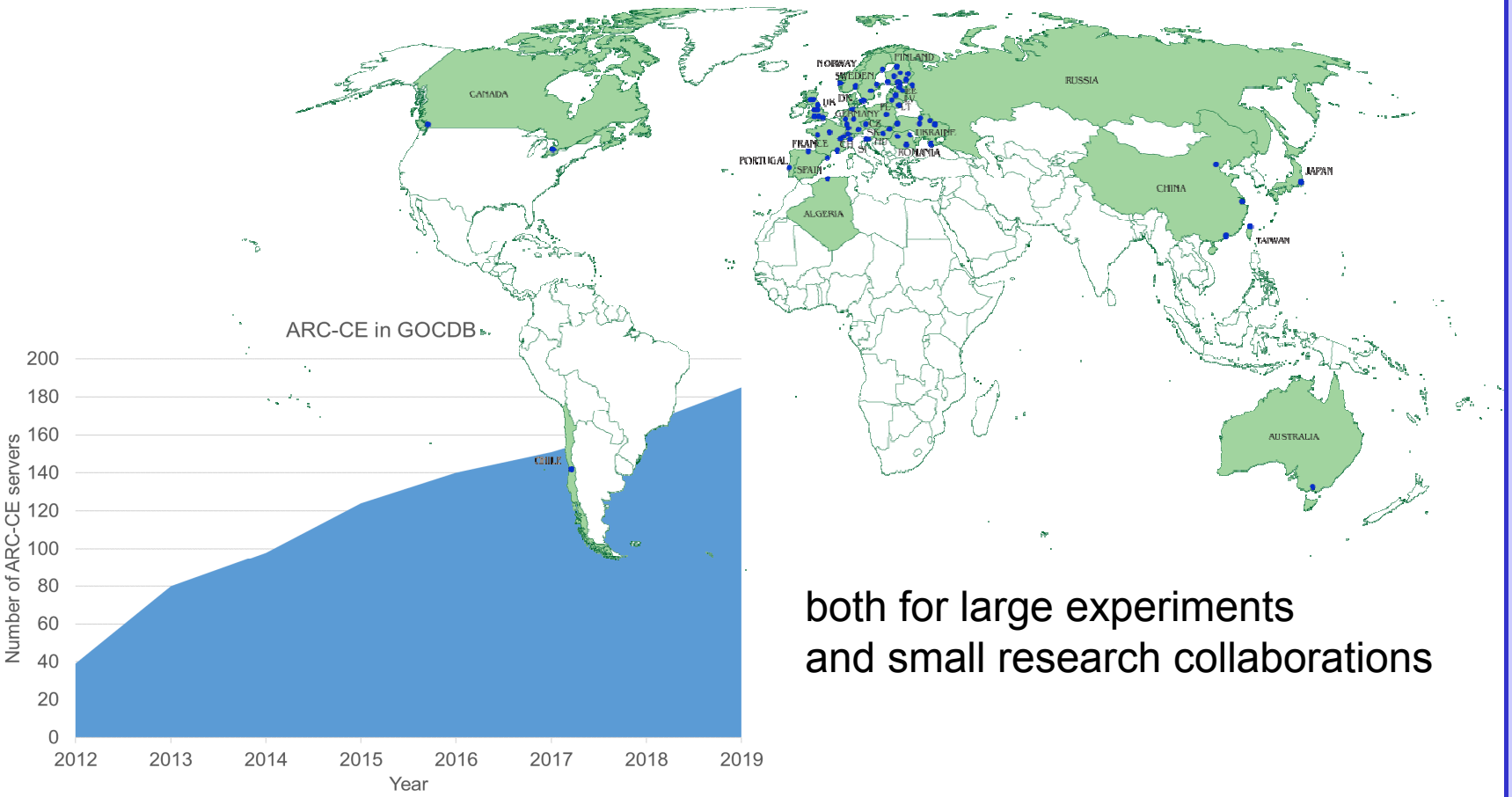
ARC history dates back 5 years BC



DEMONSTRATED LONG-TERM SUPPORT COORDINATED BY THE NORDUGRID COLLABORATION

ABILITY TO RENEW CONTINUOUSLY, ADOPT AND INTEGRATE WITH NEW TECHNOLOGIES

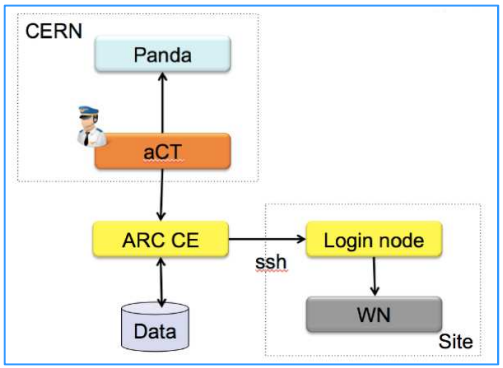
ARC CE is a common choice

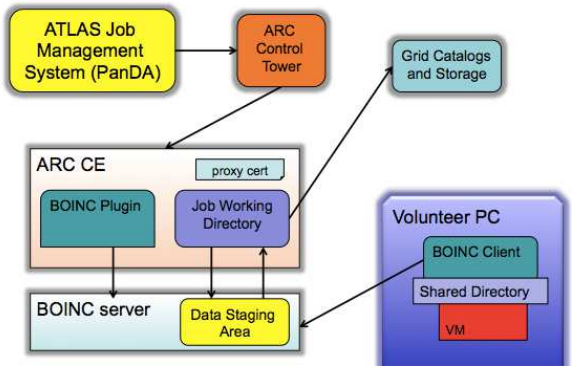


both for large experiments
and small research collaborations

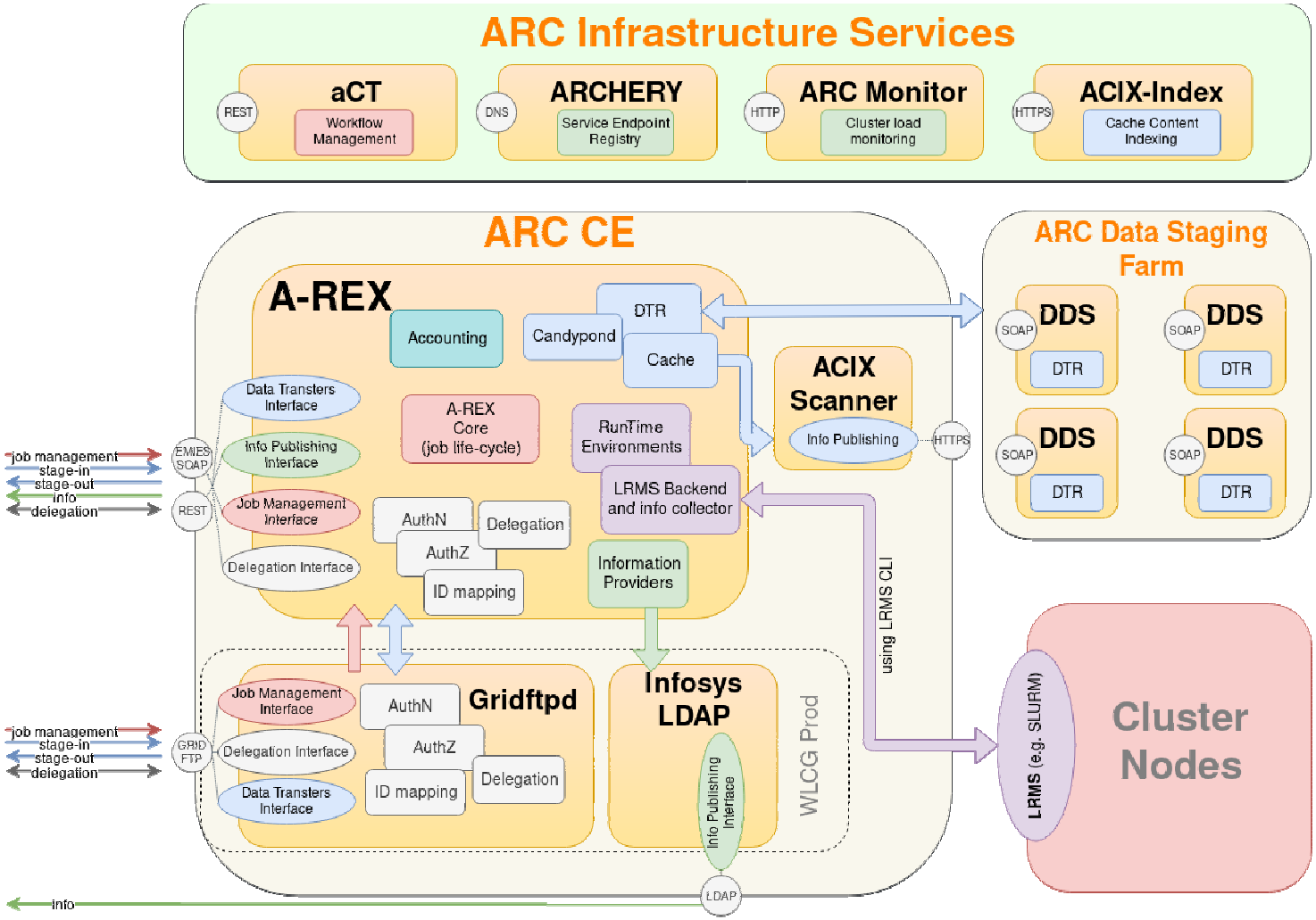
ARC as part of WLCG

- ARC CE is well-known to the WLCG VOs
  - the experiment frameworks know howto „submit to this CE”
  - an ARC CE has very small footprint on WLCG site operation or architecture (e.g. no middleware on the WNs)
- Full infrastructure integration:
  - APEL, ARGUS, BDII (GLUE2 and 1), GOCDB, VOMS,
  - EGI/WLCG monitoring, UMD & GGUS procedures....
- Works both on restricted HPCs and opportunistic (Boinc) nodes ... and of course with normal clusters





ARC CE: internals, interfaces and the infrastructure ecosystem around

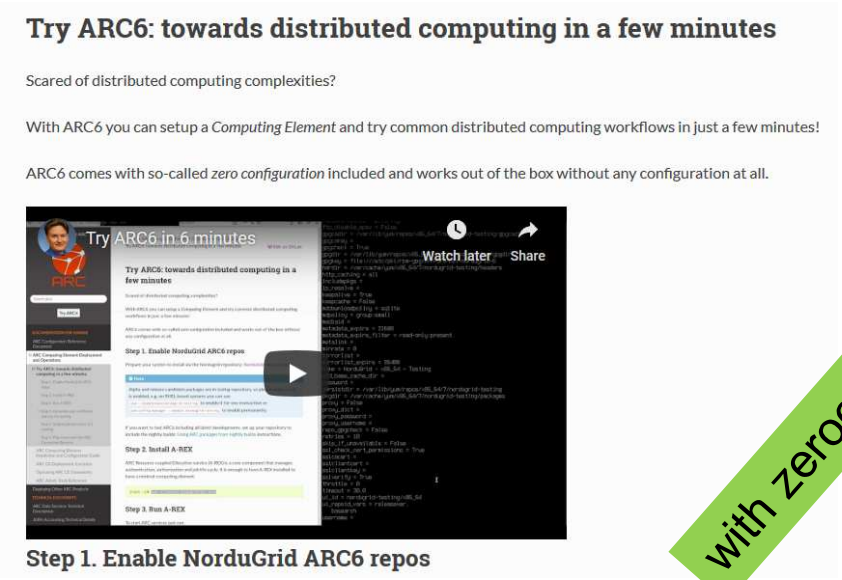


and now comes the interesting part

Power feature 1: quickstart from scratch

Setup and try out a real CE in just a few minutes!

- ARC6 CE in six steps and six minutes



with zeroconfig, defaults, arcctl

- quickguide and video:
  - http://www.nordugrid.org/arc/arc6/admins/try_arc6.html
- later you can use the same setup as a starting point for a production deployment

Power feature 2: online documentation

NordUgrid ARC

Try ARC6

Search docs

DOCUMENTATION FOR ADMINS

ARC Configuration Reference Document

ARC CE Deployment and Operation

ARCHERY Overview and Deployment

ARC Admin Tools Reference

TECHNICAL DOCUMENTS

ARC Data Services Technical Description

JURA Accounting Technical Details

Requirements and design of REST interface

Old Relevant Technical Documents

DOCUMENTATION FOR DEVELOPERS

Implementation Details for Developers

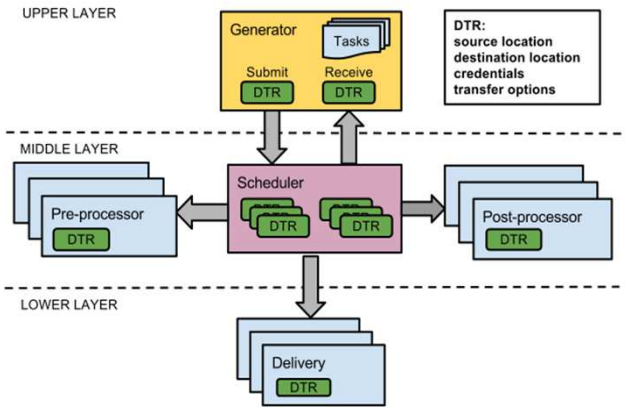
Docs » ARC CE Deployment and Operation

ARC CE Deployment and Operation

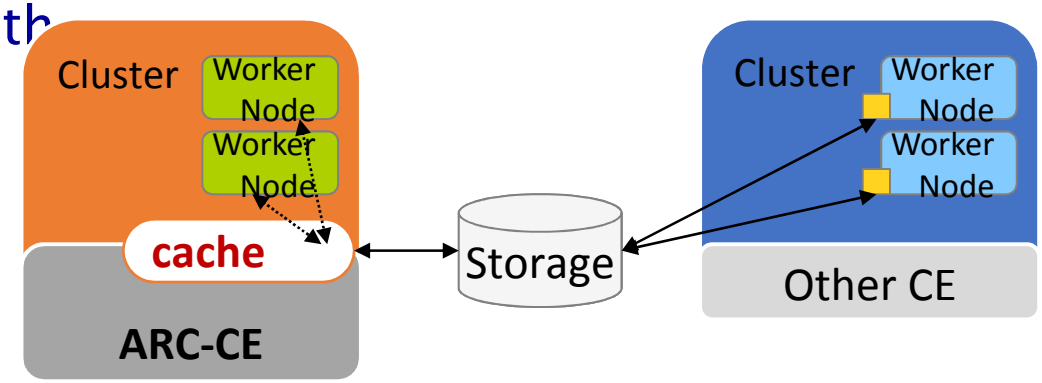
- Try ARC6: towards distributed computing in a few minutes
  - Step 1. Enable NorduGrid ARC6 repos
  - Step 2. Install A-REX
  - Step 3. Run A-REX
  - Step 4. Generate user certificate and key for testing
  - Step 5. Submit job and check it is running
  - Step 6. Play more with the ARC Computing Element
- ARC Computing Element Installation and Configuration Guide
  - Prerequisites
  - Installation
  - Configuration
  - Configure Firewall
  - Enable and Run Services
  - Test Basic Functionality
- ARC5 to ARC6 Migration Guide
  - Quick reference
  - Configuration file
  - Operating ARC services
  - Retiring the EGIIS
- ARC CE Deployment Scenarios
- Operating ARC CE Subsystems
  - ARC6 Packages
  - ARC6 Services
  - Authorization and Mapping rules in ARC6
  - Batch systems support

Power feature 3: data handling

- ARC-CE does **data staging** with DTR and DDS
  - The DTR performs the critical role of transferring input and output data for jobs



- job input **data caching**
  - Improves WN usage efficiency
  - Minimizes bandwidth



- and many more (!)

Power feature 4: ARCCTL

Deployment Automation

- CA Certificates
- VOMS
- Firewall
- Package install

Service Control

manage ARC services based on configuration

Configuration Management

- overview
- validation
- in-line help
- runtime config

Jobs Control

- get info, logs, script
- kill and clean
- ownership
- statistics

`arcctl`

Accounting, Usage Analysis

- who is using CE
- how they using CE
- how much CPU, memory, disc, network

Worker Node Environment

- features available during job runtime
- environment setup
- containers support

Data Staging

- cache control
- transfers stats

Power feature 5: CE overload protection

- **maxjobs A,B,C,D,E** – specifies maximum allowed number of jobs.
  - A: jobs which are not in FINISHED state (jobs tracked in RAM)
  - B: jobs being run (SUBMITTING, INLRMS states)
  - C: jobs being processed per DN
  - D: jobs in whole system
  - E: LRMS scripts limit (jobs in SUBMITTING and CANCELING

- **defaulttl:** how long job directory and job control info is kept
- **maxrerun, wakeupperiod, infoproviders_timelimit, maxtransfertries, maxdelivery, maxprocessor**

Power feature 6: identity mapping,access control

Identity mapping: connect global (grid, VO) identity to local cluster accounts

- Dedicated powerful authz **mapping subsystem** [mapping]
  - map_to_user, map_to_pool
  - map_with_plugin (e.g. for LCMAPS), map_with_file (gridmapfiles are the past)
  - policy_on_map, policy_on_nomap, policy_on_nogroup
  - BUT:any mapping to root -> request processing fails implicitly.

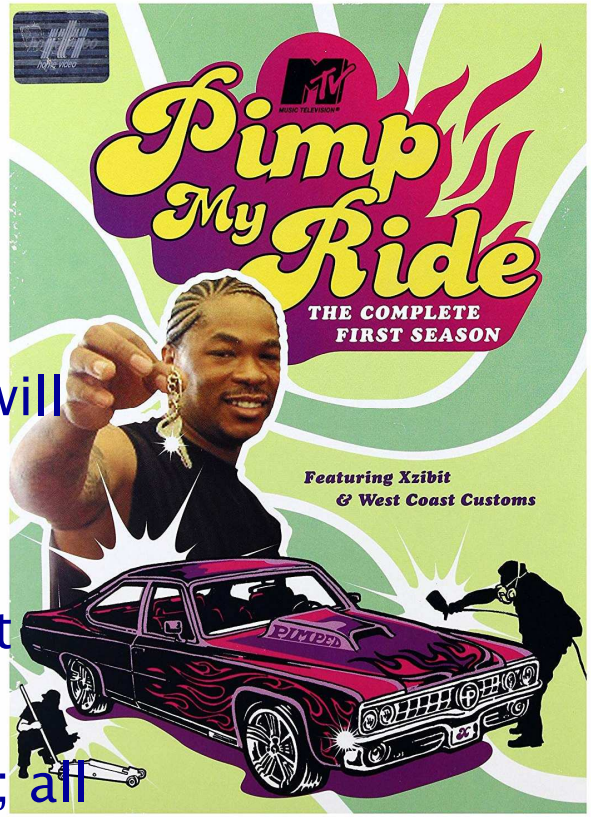
Access control:

- Per service and per queue: allowaccess and denyaccess
- Using the very powerfull **authgroup** concept
  - subject, file, VOMS, plugin, authgroup (for nesting), all

Power feature 6+: A CE crying for tuning!

Built-in official mechanisms for tuning your CE:

- statecallout
  - with state, options, plugin_path, [plugin_arguments]
  - Enables a callout feature of A-REX: every time job goes to state A-REX will run plugin_path executable.
- RunTimeEnvironment framework
  - Setup and advertise WN environment
  - Job script customization by A-REX
  - system, user, default, enabled types; all managed with arcctl tool



Power feature 6++: job traceability

- Very detailed information about the various job lifecycle changes, job information
- “active” job info is published via info interfaces
- Completed job info, job history is kept in local accounting database (AAR record format)
- Job log (history) is available to job owner
- Site admin has full overview of grid jobs (arccctl tool)

Development plans:in the pipeline

- RESTful interface:
http://www.nordugrid.org/arc/arc6/tech/rest/rest.html

- Support for tokens:
http://www.nordugrid.org/arc/arc6/misc/oidc_tokens.html

- New ideas:
ARC DATA service, LRMS-less ARC

ARC FACTSHEET

Latest version: ARC 6.8.1

- Code:
  - https://source.coderefinery.org/nordugrid/arc
- Installable packages:
  - Global Linux repositories (CentOS, Debian, EPEL)
  - Upstream: http://download.nordugrid.org/repos.html
- Documentation:
  - ARC6: complete documentation online at http://www.nordugrid.org/arc/arc6
- Support:
  - Dedicated skype support channel
    - https://join.skype.com/dyf3A6Uutjy2
  - Email lists:
    - nordugrid-discuss@nordugrid.org – generic
    - CERN e-group wlcg-arc-ce-discuss@cern.ch – WLCG-specific
  - Own bugzilla and EGI GGUS:
    - https://bugzilla.nordugrid.org
    - For those familiar with GGUS, submit tickets to “ARC” unit