

Retour d'expérience

Un plan

Des ressources

Schéma de fonctionnement


HTCondor

- Config, Processus, Interfaces
- Gestion d'une hiérarchie de groupes

ARC

- Procesus, Interfaces
- Config, Accounting, Monitoring

Ressources




ADVANCED RESOURCE CONNECTOR

About ARC	Download	Documents	Releases
Nightlies	Testing portal	Report bugs	Support

Current release: ARC v15.03u18





For users and sysadmins

- Download and installation instructions:
 - ARC client installation
 - ARC CE installation
 - ARC CE configuration examples
- Migration of ARC CE from ARC 0.8.x to 11.05
- NorduGrid Wiki
- User support (via GGUS)
- Bugzilla issue tracking system (direct line to developers)
- FAQ.
- Posters, logos etc
- Release Notes
- Roadmap



For developers

- ARC SDK documentation
- ARC Code repository
 - Trac browser
 - Direct access to ARC code repository
- Rules
 - Coding rules
 - Documentation writing instructions
 - ARC SVN instructions
 - Bugzilla rules
 - Release procedures
- Tests portal: automatic revision, functional and performance tests
- Bugzilla issue tracking system
- NorduGrid Wiki
- Nightly build status
- Weekly TCG meetings
- Manuals
- Release Notes



<http://www.nordugrid.org/arc>



Computing with HTCondor™

Our goal is to develop, implement, deploy, and evaluate mechanisms and policies that support [High Throughput Computing \(HTC\)](#) on large collections of distributively owned computing resources. Guided by both the technological and sociological challenges of such a computing environment, the [Center for High Throughput Computing](#) at UW-Madison has been building the open source [HTCondor distributed computing software](#) (pronounced 'aitch-tee-condor') and related technologies to enable scientists and engineers to increase their computing throughput.

Note: The HTCondor software was known as 'Condor' from 1988 until its name changed in 2012. If you are looking for Phoenix Software International's software development and library management system for z/VSE or z/OS, click [here](#).

[Home](#) | [News](#) | [Download](#) | [Publications](#) | [Contact Us](#)

Google Custom S

Latest News [RSS](#)

> HTCondor powers Marshfield Clinic project on disease genetics
May 16, 2018

> HTCondor 8.7.8 released!
May 10, 2018

> HTCondor 8.6.11 released!
May 10, 2018

▼ HTCondor 8.7.7 released!
March 13, 2018

The HTCondor team is pleased to announce the release of HTCondor 8.7.7. This development series release contains new features that are under development. This release contains all of the bug fixes from the 8.6.10 stable release. Enhancements in the release include: condor_ssh_to_job now works with Docker Universe jobs; A 32-bit condor_shadow is available for Enterprise Linux 7 systems; Tracks and reports custom resources, e.g. GPUs, in the job ad and user log; condor_q_unmatchable reports jobs that will not match any slots; Several updates to the parallel universe; Spaces are now allowed in input, output, and error paths in submit files; in DAG files, spaces are now allowed in submit file paths. Further details can be found in the [Development Version History](#) and the [Stable Version History](#). HTCondor 8.7.7 binaries and source code are available from our [Downloads](#) page.

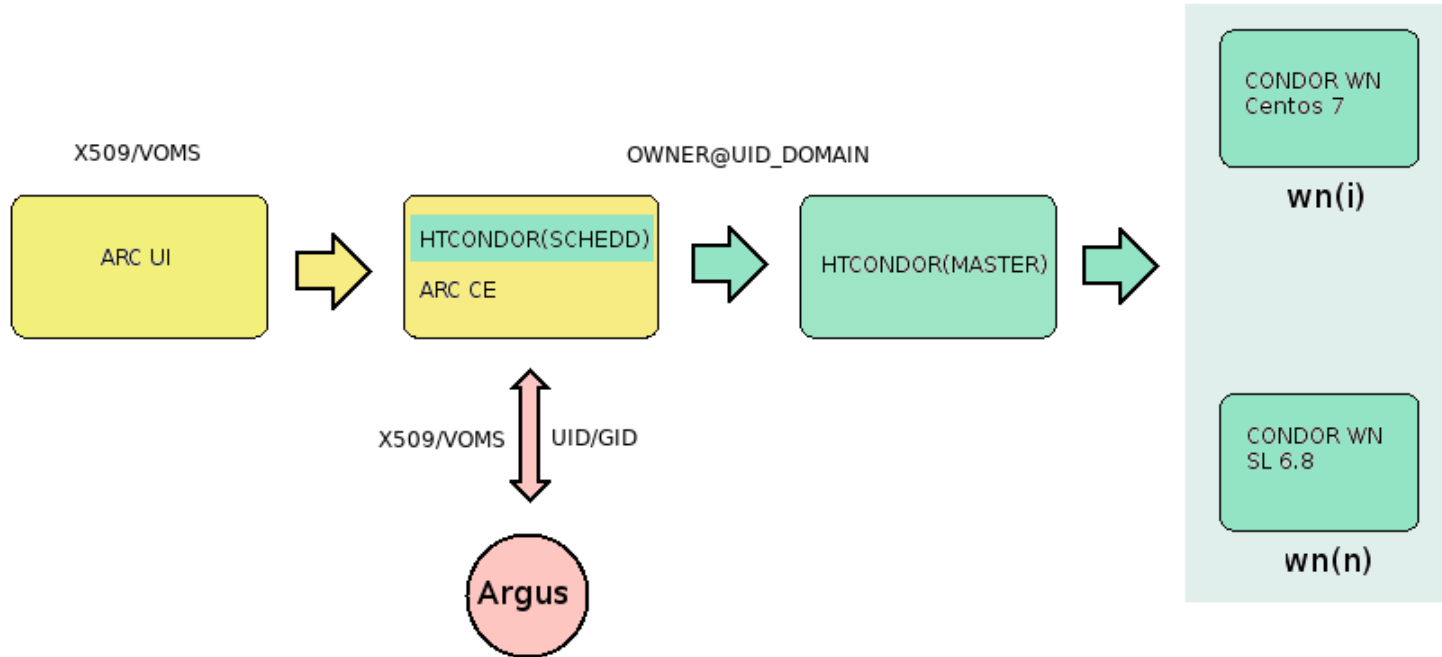
> HTCondor 8.6.10 released!
March 13, 2018

> HTCondor Week 2018 Registration Open
February 28, 2018

[More News >](#)

<http://research.cs.wisc.edu/htcondor>

Le cluster – Schéma



Sous Centos7

Des dépôts a configurer:

- epel-release
- htcondor-stable-rhel7.repo
(<http://research.cs.wisc.edu/htcondor/yum/>)
- Dans le serveur Arc
 - Nordugrid-release
<https://download.nordugrid.org>
15.03-1.el7.centos (latest)
 - umd-4

Des paquets a installer :

- **Condor** (tout nœud)
8.6.10-1/11-1
- Dans le serveur Arc :
 - Un certificat serveur
 - EGI IGTF release
 - **Nordugrid-compute-element**
15.03-1.el7.centos latest
 - des packages UMD4 :
 - ✓ argus-pep, argus-pep-common, etc
 - ✓ Apel-server
 - ✓ lcms, lcms-pugins, etc
- Dans les Wns
 - nordugrid-arc-clients

Notre démarche

- Suivre ARC HTCondor Basic Install
 - https://www.gridpp.ac.uk/wiki/ARC_HTCondor_Basic_Install
- Valider Condor

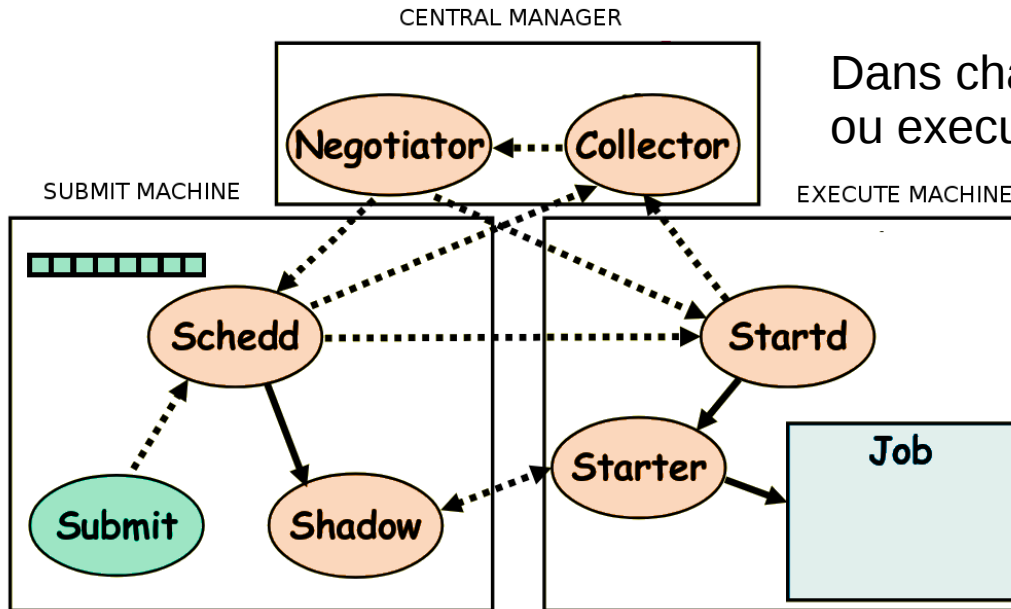
A partir du « master condor»

 - Tester l'envoi de jobs, prévoir un utilisateur Unix « test » dans toute machine avec condor installée.
- Valider la chaîne Arc + Condor avec un mapping statique.

A partir d'une UI (avec le client arc installée), il faudra prévoir (dans le arc)

 - un grid-mapfile avec une ligne "/O=GRID-FR/C=FR/O=CNRS/OU=CPPM/CN=XXX YYY ZZZ" test
 - Un user « test » (arc, condor, wn's)
- Voms
- Valider Argus
 - Avec un proxy interroger le serveur Argus avec le client arc-lcmaps
- Configurer pools (arc, condor, wns)
- Declarer le service dans la local BDII et GOCDB
- Monitoring, Accounting, etc

HTCONDOR Machines-Processus



Dans chaque type de nœud submit, master ou execute la configuration de condor diffère.

Éditer **DAEMON_LIST**
en conséquence

DAEMON_LIST = COLLECTOR, MASTER, NEGOTIATOR, SCHEDD, STARTD

HTCondor et dynamic slots

```
SLOT_TYPE_1 = cpus=100%,disk=100%,swap=100%  
SLOT_TYPE_1_PARTITIONABLE = TRUE  
NUM_SLOTS = 1  
NUM_SLOTS_TYPE_1 = 1
```

```
[root@marcce01 config.d]# condor_status -server
```

Name	OpSys	Arch	LoadAv	Memory	D
slot1@marwn68.in2p3.fr	LINUX	X86_64	1.000	11850	
slot1_1@marwn68.in2p3.fr	LINUX	X86_64	0.000	512	
slot1_2@marwn68.in2p3.fr	LINUX	X86_64	0.000	512	
slot1_3@marwn68.in2p3.fr	LINUX	X86_64	0.000	512	
slot1_4@marwn68.in2p3.fr	LINUX	X86_64	0.460	512	
slot1_5@marwn68.in2p3.fr	LINUX	X86_64	1.000	512	
slot1_6@marwn68.in2p3.fr	LINUX	X86_64	1.000	512	
slot1_7@marwn68.in2p3.fr	LINUX	X86_64	1.000	512	
slot1_8@marwn68.in2p3.fr	LINUX	X86_64	1.000	512	

HTCondor et multicore

```
SUBMIT_REQUIREMENT_NAMES = slots
```

```
SUBMIT_REQUIREMENT_slots = (RequestCpus == 1) || \  
    (RequestCpus == 8 && x509UserProxyVOName "atlas")
```

```
SUBMIT_REQUIREMENT_slots_REASON = "Only 1core requirements are accepted"
```

HTCONDOR fragmentation

```
DAEMON_LIST = $(DAEMON_LIST) DEFRAG
```

```
DEFRAG_INTERVAL = 3600
```

```
# one x startd daemon
```

```
DEFRAG_DRAINING_MACHINES_PER_HOUR = 1.0
```

```
# max number of whole machines ( default Cpus == TotalCpus && Offline!=True )
```

```
DEFRAG_MAX_WHOLE_MACHINES = 20
```

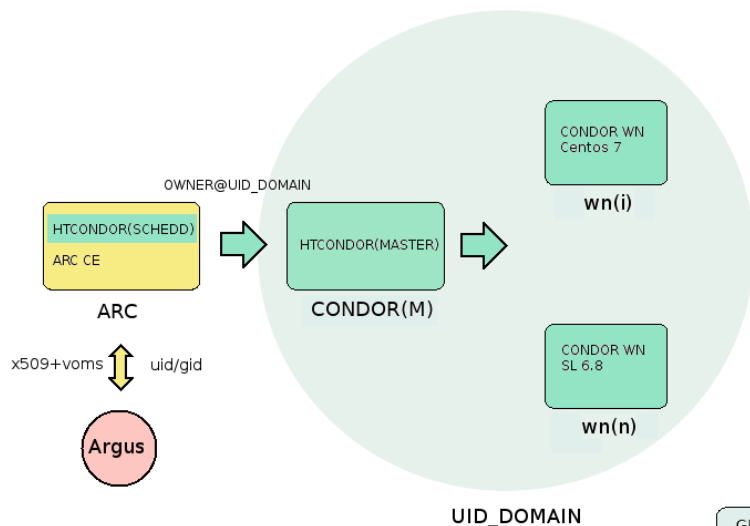
```
# the maximum number of draining machines
```

```
DEFRAG_MAX_CONCURRENT_DRAINING = 10
```

```
# which machines are already operating as whole machines
```

```
DEFRAG_WHOLE_MACHINE_EXPR = ((Cpus == TotalCpus) || (Cpus >= 8))
```

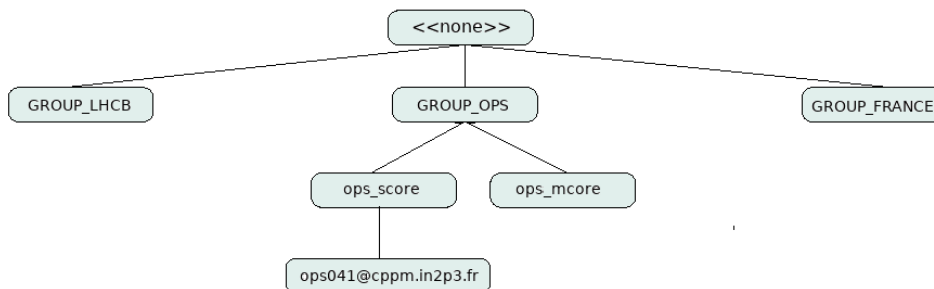
HTCondor Hierarchical groups



Pour Condor un utilisateur est de la forme **OWNER@UID_DOMAIN**.

Tout utilisateur au départ a la même priorité et quota.

Condor offre la possibilité d'agrouper des utilisateurs en agissant sur le classad « AccountingGroup » et ensuite configurer des quotas et priorités



 OWNER@UID_DOMAIN

 GROUP.SUBGROUP.OWNER@UID_DOMAIN

HTCondor groups, subgroups

OWNER@UID_DOMAIN => GRUP.SUBGROUP.OWNER@UID_DOMAIN

Ex : france008@cppm.in2p3.fr => group_FRANCE.france_score.france008@cppm.in2p3.fr

On doit modifier le classad :

SUBMIT_EXPRS = \$(SUBMIT_EXPRS) **VAcctGroup, VAcctSubGroup, AccountingGroup**

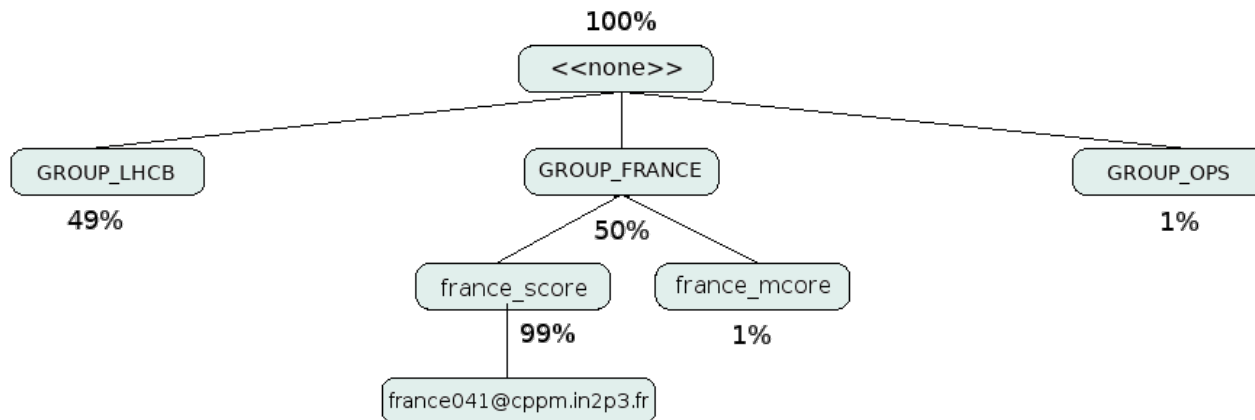
Contribution de différents classads :

- VAcctGroup <= **x509UserProxyVOName**
- VAcctSubGroup <= substr(**owner**) + score/mcore(**RequestCPUS**)
- AccountingGroup <= VAcctGroup.VAcctSubGroup.**owner@UID_DOMAIN**

Ex :

x509UserProxyVOName (france)	=>	group_FRANCE
owner (france008) + RequestCpus	=>	france_score/france_mcore
Owner @ UID_DOMAIN (*)	=>	france008@cppm.in2p3.fr

HTCondor quota



GROUPS, QUOTAS, SURPLUS

GROUP_NAMES=group_FRANCE,group_FRANCE.france_score,group_FRANCE.france_mcore

GROUP_QUOTA_DYNAMIC_group_FRANCE = 0.5

GROUP_QUOTA_DYNAMIC_group_FRANCE.france_score = 0.99

GROUP_QUOTA_DYNAMIC_group_FRANCE.france_mcore = 0.01

GROUP_ACCEPT_SURPLUS_group_FRANCE = False

GROUP_ACCEPT_SURPLUS_group_FRANCE.france_score = True

GROUP_ACCEPT_SURPLUS_group_FRANCE.france_mcore = True

HTCondor quota et surplus

Group User Name	Config Quota	Use Surplus	Effective Priority	Priority Factor	Res In Use	Total Usage (wghted-hrs)	Time Since Last Usage	Requested Resources
group_DTEAM.dte_mcore	0.70	Regroup		1000.00	0	6.70	9+21:53	0
group_DTEAM.dte_score	0.30	Regroup		100.00	0	1552.00	8+16:49	0
dte177@cppm.in2p3.fr			50.00	100.00	0	0.84	8+16:49	
group_FRANCE	0.50	Regroup		1000.00	0	24.24	34+03:28	0
group_FRANCE.france_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_FRANCE.france_score	0.99	Regroup		200.00	0	18.69	0+22:07	0
france186@cppm.in2p3.fr			100.00	200.00	0	10.21	7+23:41	
france044@cppm.in2p3.fr			100.00	200.00	0	0.98	0+22:07	
france130@cppm.in2p3.fr			100.00	200.00	0	7.50	47+00:08	
group_OPS.ops_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_OPS.ops_score	0.99	Regroup		300.00	0	12.04	0+00:04	1
ops036@cppm.in2p3.fr			150.00	300.00	0	1.20	0+06:13	
ops049@cppm.in2p3.fr			150.00	300.00	0	1.20	0+00:31	
ops018@cppm.in2p3.fr			150.00	300.00	0	6.29	8+16:42	
ops010@cppm.in2p3.fr			150.26	300.00	0	3.35	0+00:04	
group_DTEAM	0.15	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION	0.00	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION.forrzk_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION.forrzk_score	0.99	Regroup		600.00	0	4.65	4+15:30	0
forrzk152@cppm.in2p3.fr			300.00	600.00	0	4.65	4+15:30	
group_HIGHPRIO	0.00	Regroup		1000.00	0	0.00	17506+13:1	0

Output commande « condor_userprio »

HTCondor priorité

Deux possibilités :

Avec la commande « condor_userprio »

ex :

```
#condor_userprio -setfactor 600
```

```
#condor_userprio -setfactor 600 group_FRANCE.france_score.france148@cppm.in2p3.fr
```

Persistent dans la config (Condor File)

```
GROUP_PRIO_FACTOR_group_FORMATION = 1000.0
```

```
GROUP_PRIO_FACTOR_group_FORMATION.forrzk_score = 600.0
```

```
GROUP_PRIO_FACTOR_group_FORMATION.forrzk_mcore = 1000.0
```

HTCondor group priorité

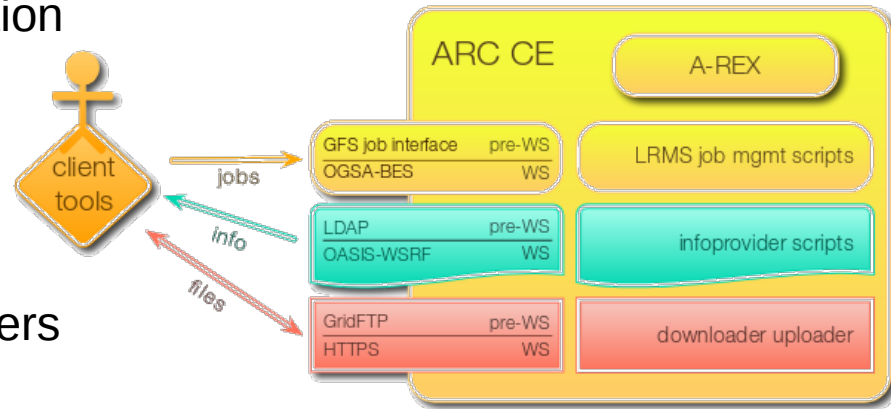
Group User Name	Config Quota	Use Surplus	Effective Priority	Priority Factor	Res In Use	Total Usage (wghted-hrs)	Time Since Last Usage	Requested Resources
group_DTEAM.dte_mcore	0.70	Regroup		1000.00	0	6.70	9+21:53	0
group_DTEAM.dte_score	0.30	Regroup		100.00	0	1552.00	8+16:49	0
dte177@cppm.in2p3.fr			50.00	100.00	0	0.84	8+16:49	
group_FRANCE	0.50	Regroup		1000.00	0	24.24	34+03:28	0
group_FRANCE.france_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_FRANCE.france_score	0.99	Regroup		200.00	0	18.69	0+22:07	0
france186@cppm.in2p3.fr			100.00	200.00	0	10.21	7+23:41	
france044@cppm.in2p3.fr			100.00	200.00	0	0.98	0+22:07	
france130@cppm.in2p3.fr			100.00	200.00	0	7.50	47+00:08	
group_OPS.ops_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_OPS.ops_score	0.99	Regroup		300.00	0	12.04	0+00:04	1
ops036@cppm.in2p3.fr			150.00	300.00	0	1.20	0+06:13	
ops049@cppm.in2p3.fr			150.00	300.00	0	1.20	0+00:31	
ops018@cppm.in2p3.fr			150.00	300.00	0	6.29	8+16:42	
ops010@cppm.in2p3.fr			150.26	300.00	0	3.35	0+00:04	
group_DTEAM	0.15	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION	0.00	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION.forrzk_mcore	0.01	Regroup		1000.00	0	0.00	17506+13:1	0
group_FORMATION.forrzk_score	0.99	Regroup		600.00	0	4.65	4+15:30	0
forrzk152@cppm.in2p3.fr			300.00	600.00	0	4.65	4+15:30	
group_HIGHPRIO	0.00	Regroup		1000.00	0	0.00	17506+13:1	0

Output commande « condor_userprio »

ARC Processus et Interfaces

A-REX (the execution service)

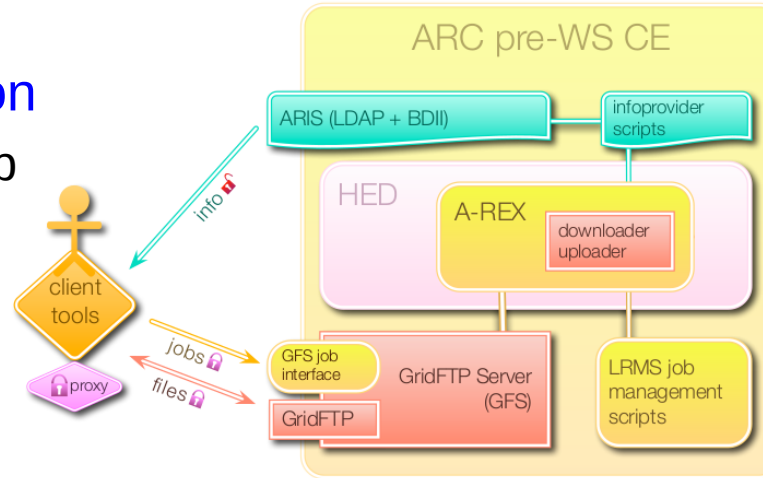
- Accepter les demandes d'exécution
- Executer dans le batch system local
- Surveiller le statut de chaque travail.
- Pré and post traitement des fichiers



ARC Interfaces

Deux interfaces de soumission

- pre-web service Ldap/gridftp
- web serviceHttps



```
[carranza@martb08 exo4]$ arcinfo marcce01.in2p3.fr
Computing service: marcce01 (IN2P3-CPPM) (production)
Information endpoint: ldap://marcce01.in2p3.fr:2135/Mds-Vo-Name=local,o=grid
Information endpoint: ldap://marcce01.in2p3.fr:2135/Mds-Vo-Name=resource,o=grid
Information endpoint: ldap://marcce01.in2p3.fr:2135/o=glue
Information endpoint: https://marcce01.in2p3.fr:60000/arex
Information endpoint: https://marcce01.in2p3.fr:60000/arex
Submission endpoint: https://marcce01.in2p3.fr:60000/arex (status: ok, interface: org.ogf.bes)
Submission endpoint: https://marcce01.in2p3.fr:60000/arex (status: ok, interface: org.ogf.glue.emies.a
Submission endpoint: qsiftp://marcce01.in2p3.fr:2811/jobs (status: ok, interface: org.nordugrid.gridft
```

ARC - Configuration

Un fichier « arc.conf » avec différents sections permettant de configurer services et processus notamment :

[common]

... réseau, sécurité et LRMS

[cluster]

... authorized vos, cluster info, etc

[grid-manager]

... A-REX, comportement des jobs, directories, accounting (JURA)

[gridftpd]

... serveur pour le gridftp protocole, ports, mapping, etc

[gridftp/jobs]

... interface de soumission (web, pre-web)

[infosys]

... système d'information, le format de l'information publiée par le serveur.

[queue/xxx]

..... configuration de queues, architecture, condor « requirements », etc

ARC - ARGUS interaction

Dans la section « gridftpd » du fichier arc.conf

[gridftpd]

...

```
unixmap="* lcmaps liblcmaps.so /usr/lib64 /etc/lcmaps/lcmaps.db vomms"
```

```
unixmap="nobody:nobody all"
```

Dans le fichier lcmaps.db

```
pepc = "lcmaps_c_pep.mod"
```

```
--pep-daemon-endpoint-url https://margus.in2p3.fr:8154/authz"
```

```
--resourceid http://authz-interop.org/xacml/resource/resource-type/arc"
```

```
--actionid http://glite.org/xacml/action/execute"
```

```
--capath /etc/grid-security/certificates/"
```

```
--certificate /etc/grid-security/hostcert.pem"
```

```
--key /etc/grid-security/hostkey.pem"
```

ARC - Accounting

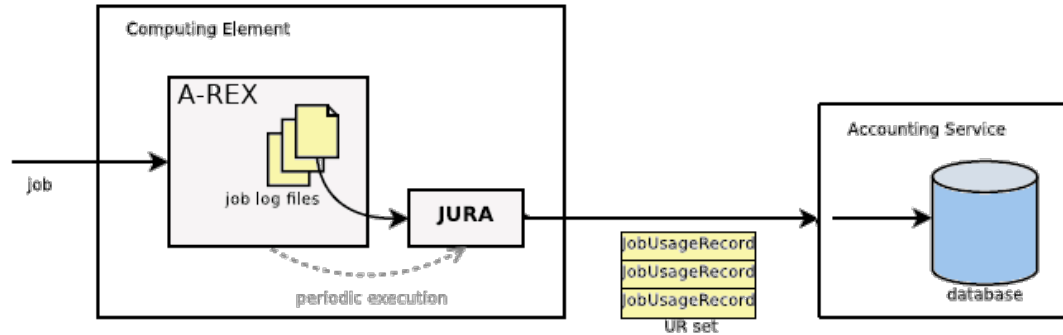
A-REX – Execution service

Pour chaque job génère des logs.

JURA, job usage report of ARC

Génère « job usage records » UR

- Usage Record 2.0 (Computing Accounting Record) XML format
- Capable d'envoyer les enregistrements directement vers a SGAS service ou vers un APEL service.



Notre choix

- En reprenant le travail de F.Schaer – CEA
- Un script dans la crontab récupère chaque UR et le dépose dans une file de messages (MQ)
- Un service « apeldbloader » lit dans la MQ et insère chaque message dans la base de données de l'accounting national

ARC - Monitoring

- Déclarer le service
- ✓ dans la local BDII
- ✓ dans la GOC DB (only monitored)
- Install nordugrid-arc-nagios-plugins (wn's)
- Un environment /etc/arc/runtime/ENV/PROXY

Current Network Status
 Last Updated: Thu May 17 13:39:43 CEST 2018
 Updated every 90 seconds
 Nagios Core™ 4.3.1 - www.nagios.org
 Logged in as O-GRID-FRC-FRC-CNRS/OU=CPPMCA/Juan Carlos Carranza
 View History For This Host
 View Notifications For This Host
 View Service Status Detail For All Hosts

Host Status Totals

Up	Down	Unreachable	Pending
1	0	0	0
All Problems All Types			
0	1	0	0

Service Status Totals

Ok	Warning	Unknown	Critical	Pending
10	0	0	1	0
All Problems All Types				
1	0	0	11	0

Service Status Details For Host 'marcce01.in2p3.fr'

Host	Service	Status	Last Check	Duration	Attempts	Status Information	
marcce01.in2p3.fr	org.nordugrid.ARC-CE-ARIS	OK	05-17-2018 13:37:43	576 2h 49m 16s	1/3	1 cluster (nordugrid-arc-5.4.2), 2 queues (active, active)	
	org.nordugrid.ARC-CE-IGTF-ops	CRITICAL	05-17-2018 12:09:18	48d 21h 16m 29s	2/2	IGTF-1.91, 3 days old, all present. - SHA Fingerprint failed for IRAN-GRID.	
	org.nordugrid.ARC-CE-SRM-result-ops	OK	05-16-2018 12:06:41	3d 22h 25m 28s	1/2	Job succeeded. - JID: gslfp://marcce01.in2p3.fr:2811/jobs/9MMAMDMF48cswiWsaJ7oEwRUmABFKDmABFKDmBOPVdMABFKDmEUsG	
	org.nordugrid.ARC-CE-SRM-submit-ops	OK	05-17-2018 12:57:43	0d 11h 41m 58s	1/2	Job submitted.	
	org.nordugrid.ARC-CE-result-ops	OK	05-17-2018 12:09:18	76d 23h 56m 34s	1/2	Job succeeded. - JID: gslfp://marcce01.in2p3.fr:2811/jobs/OTEOdmagUdsrWsaJ7oEwRUmABFKDmABFKDmABFKDmABFKDmABKlqm - JID: gslfp://marcce01.in2p3.fr:2811/jobs/OTEOdmagUdsrWsaJ7oEwRUmABFKDmABFKDmABFKDmABKlqm	
	org.nordugrid.ARC-CE-srm-ops	OK	05-16-2018 12:06:41	61d 20h 25m 3s	1/2	Service OK.	
	org.nordugrid.ARC-CE-submit-ops	OK	05-17-2018 12:47:43	6d 18h 39m 41s	1/2	Job submitted.	
	org.nordugrid.ARC-CE-sw-csh-ops	?	OK	05-17-2018 12:09:18	86d 8h 19m 34s	1/2	Found working csh.
	org.nordugrid.ARC-CE-sw-gcc-ops	?	OK	05-17-2018 12:09:18	154d 19h 34m 44s	1/2	Found GCC version 4.8.5.
	org.nordugrid.ARC-CE-sw-perl-ops	?	OK	05-17-2018 12:09:18	154d 19h 34m 44s	1/2	Found Perl version 5.16.3.
	org.nordugrid.ARC-CE-sw-python-ops	?	OK	05-17-2018 12:09:18	154d 19h 34m 44s	1/2	Found Python version 2.7.5.

```
#!/bin/bash
x509_cert_dir="/etc/grid-security/certificates"
case $1 in
  0) mkdir -pv $joboption_directory/arc/certificates/
      cp -rv $x509_cert_dir/$joboption_directory/arc
      cat ${joboption_controldir}/job.${joboption_gridid}.proxy \
          $joboption_directory/user.proxy
      ;;
  1) export X509_USER_PROXY=$RUNTIME_JOB_DIR/user.proxy
      export X509_USER_CERT=$RUNTIME_JOB_DIR/user.proxy
      export X509_CERT_DIR=$RUNTIME_JOB_DIR/arc/certificates
      ;;
  2) :
      ;;
esac
```

Des liens

Basic Install

https://www.gridpp.ac.uk/wiki/ARC_HTCondor_Basic_Install

Exemple Arc/Condor Cluster

https://www.gridpp.ac.uk/wiki/Example_Build_of_an_ARC/Condor_Cluster

Show how Liverpool runs multicore Jobs (Stephen Jones)

<https://indico.cern.ch/event/467075/contributions/1143835/attachments/1236390/1815626/mc.pdf>

How Liverpool adopted ARC / HTCondor Combo to build a Grid Cluster (Stephen Jones)

<https://indico.cern.ch/event/467075/contributions/1143835/attachments/1236390/1815626/mc.pdf>

Multicore job RAL (Andrew Lahiff, Alastair Dewhurst, John Kelly)

<https://www.slideserve.com/hanzila/multi-core-jobs-at-the-ral-tier-1>

Pour finir

- Mapping, Accounting
- Nagios ops ok, atlas ok
- LCG
 - Atlas en progrès
 - arcproxy issue job pilotes (solved)
 - Pilotes score/mcore ~ ok
 - Lhcb
- Dirac (rfc proxies, nommage des queues ARC)
- A faire (monitoring local)
- A comprendre (defrag réservation)
- Des choses a tester (cgroups)

Merci

