

outils de monitoring des VOs LHC

Benjamin Gaidioz (CERN, IT), pour l'equipe
« dashboard »

problématique

- Utilisation de la grille par une VO :
 - Stockage de ses données sur plusieurs sites, calcul distribué, etc.
 - Parfois plusieurs middleware (les expériences ATLAS et CMS au CERN).
 - Ses propres systèmes :
 - Distribution des données,
 - Soumission de jobs
 - Ses propres outils de monitoring:
 - généralement chaque système (pour les opérateurs).
 - donc transverse aux outils classiques du middleware.

Job monitoring

- Installe pour ATLAS, CMS, LHCb, ALICE
- Et standalone installation par VleMED (Amsterdam).

sort by user ▼

sort by user

sort by site

sort by vosite

sort by cloud

sort by tier

sort by submissiontool

sort by dataset

sort by application

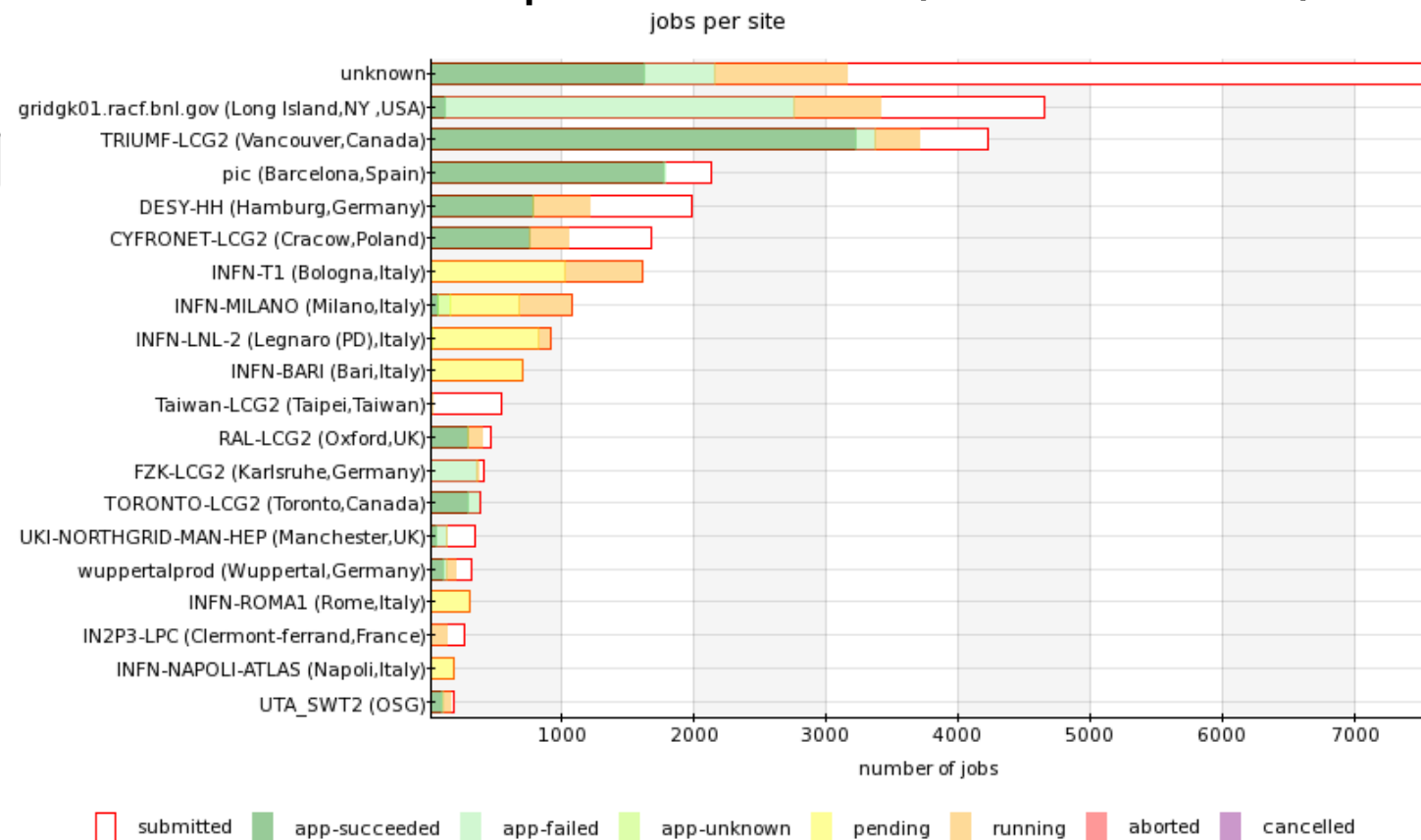
sort by rb

sort by ce

sort by activity

sort by grid

sort by country



monitoring des VOs

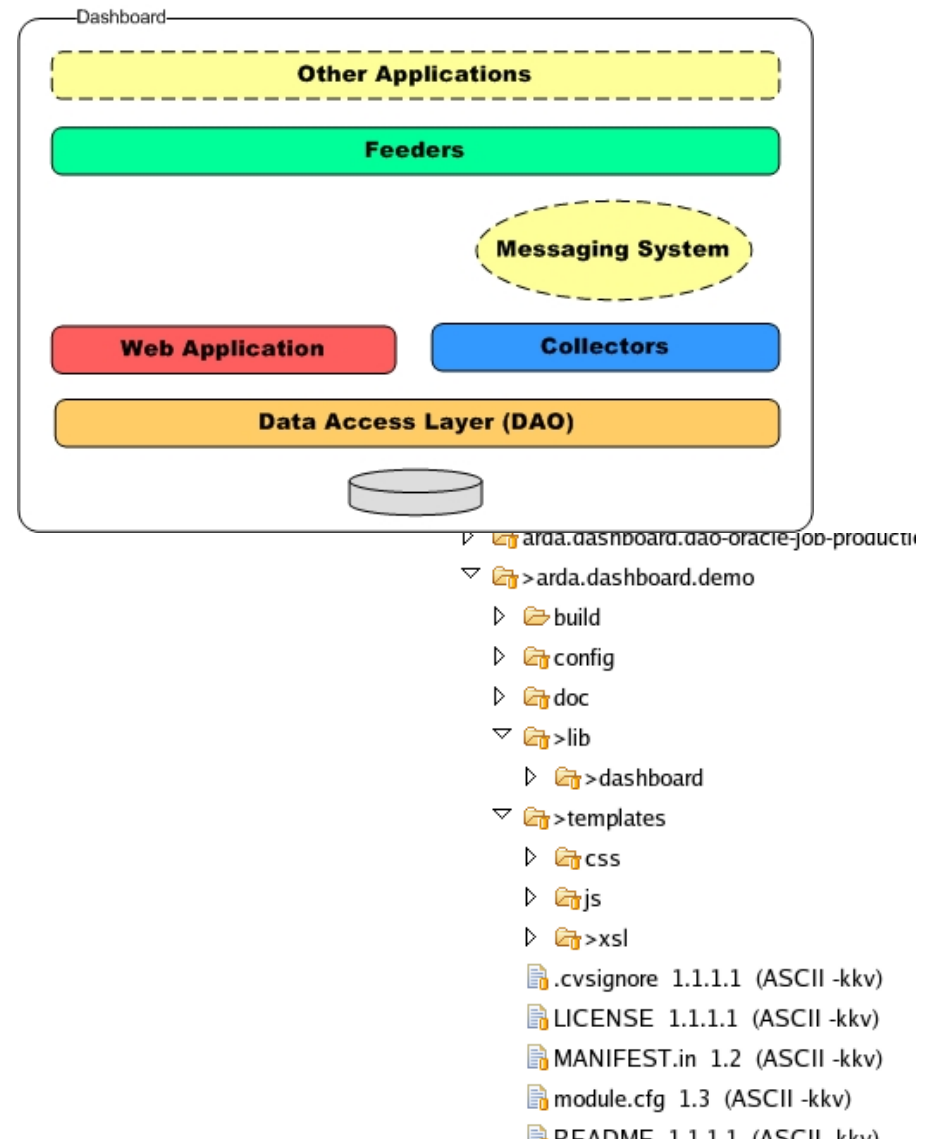
- Dysfonctionnements du grid parfois correles a l'activite
 - replication des donnees utilisant un certain SE,
 - production utilisant une certaine installation logicielle
 - Besoin de monitoring des « erreurs systeme » (grid) :
 - Resource broker,
 - Stockage,
 - Computing elements
 - Instructions pour les operateurs specifiques au systeme.

Dashboard

- Projet CERN/EGEE
 - Applications de monitoring du grid pour les Vos,
 - Melange d'informations de monitoring :
 - Middleware,
 - Logiciels des Vos
 - Certaines applications utilisees par plusieurs Vos.
 - <http://dashboard.cern.ch>

Dashboard software

- Effort de faire les choses « bien »
- Releases, guide développeur, user's guide
- Framework commun pour les applications,
- Utilisation en dehors du dashboard:
 - Monitoring et config FTS,
 - ATLAS DQ2 (build system),
 - Interface web SAM



Dashboard software

- Fonctionalites interessantes :
 - Flexibilite en output (HTML, XML, CSV, PNG, etc.),
 - API python pour lire et publier, outils de ligne de commande

```
[lxplus237] /afs/cern.ch/user/b/bgaidioz > dashb-job-summary -f text/plain
```

name	sub	unk	pend	run	term	done	canc	abort	g-unk	grid%	succ	fail	a-unk	app%	d/s	all-%
production	45621	15577	0	6130	23914	0	0	0	23914	0.00	19588	4326	0	81.91	0	0.00
analysis	1443	0	565	70	808	327	9	9	463	97.32	529	262	17	66.88	285	87.16
unknown	1111	0	155	162	794	720	44	27	3	96.39	3	0	791	100.00	0	0.00

```
[lxplus237] /afs/cern.ch/user/b/bgaidioz > dashb-job-summary -f text/plain --sortby=rb --activity=unknown
```

name	sub	unk	pend	run	term	done	canc	abort	g-unk	grid%	succ	fail	a-unk	app%	d/s	all-%
rb01.pic.es	452	0	99	100	253	252	0	1	0	99.60	0	0	253	0.00	0	0.00
rb121.cern.ch	222	0	14	38	170	169	0	1	0	99.41	0	0	170	0.00	0	0.00
rb106.cern.ch	128	0	4	5	119	114	0	5	0	95.80	0	0	119	0.00	0	0.00
lcgrb01.gridpp.rl.ac.uk	124	0	15	1	108	52	43	13	0	80.00	0	0	108	0.00	0	0.00
lcgrb02.gridpp.rl.ac.uk	89	0	19	7	63	62	1	0	0	100.00	0	0	63	0.00	0	0.00
lcgrb03.gridpp.rl.ac.uk	70	0	1	0	69	69	0	0	0	100.00	0	0	69	0.00	0	0.00
lcg005.ihep.ac.cn	10	0	0	2	8	1	0	7	0	12.50	0	0	8	0.00	0	0.00
lapp-rb01.in2p3.fr	10	0	1	9	0	0	0	0	0	0.00	0	0	0	0.00	0	0.00
lb103.cern.ch	5	0	2	0	3	0	0	0	3	0.00	3	0	0	0.00	0	0.00
rb-fzk.gridka.de	1	0	0	0	1	1	0	0	0	100.00	0	0	1	0.00	0	0.00

```
[lxplus237] /afs/cern.ch/user/b/bgaidioz >
```

```
- <item>
  <terminated>581</terminated>
  <name>SARA-MATRIX (Amsterdam, The Netherlands)</name>
  <unknown>0</unknown>
  <app-failed>0</app-failed>
  <submitted>584</submitted>
  <app-succeeded>0</app-succeeded>
  <running>1</running>
  <done>581</done>
  <aborted>0</aborted>
```

Exemple : ATLAS (experience LHC)

DDM: replication des donnees

- Unite de donnee d'ATLAS : dataset
 - Ensemble logique de fichiers
 - Doit etre replique integralement sur un site
- Service de replication :
 - ATLAS DQ2
 - Systeme distribue (Voboxes) enregistrant les requetes de replication, transferts FTS des fichiers, etc.
 - Plusieurs instances (production, tests, etc., differents SE)
- Monitoring :
 - Failures des SE, disk full, etc.
 - Actions : essayer quelques lignes de commande, envoyer un GGUS ticket, etc.

ATLAS Distributed Data Management

Overview

Dataset Info

Page Help

User Guide

Feedback

OVERVIEW

OVERVIEW Activity

Activity in Last Hour

Activity in Last 4 Hours

Activity in Last 24 Hours

Activity in Last 7 Days

Activity in Last 30 Days

Activity in ...

Cloud Activity

ASGC Cloud

BNL Cloud

CERN Cloud

CNAF Cloud

FZK Cloud

LYON Cloud

NDGF Cloud

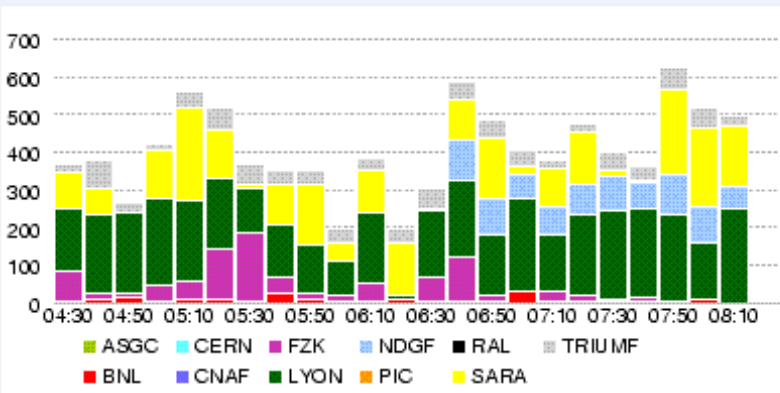
PIC Cloud

RAL Cloud

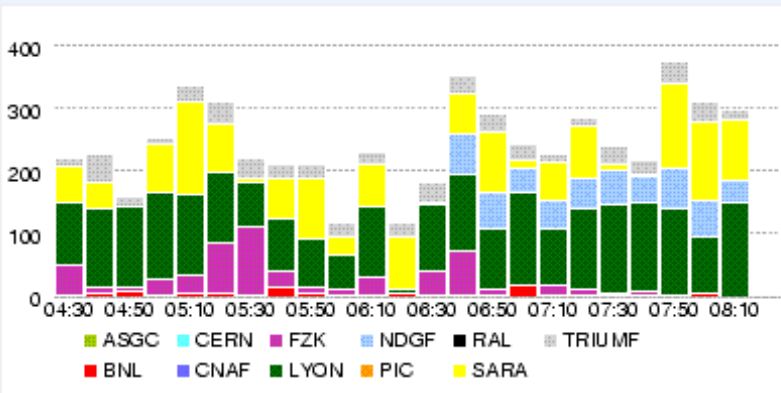
SARA Cloud

TRIUMF Cloud

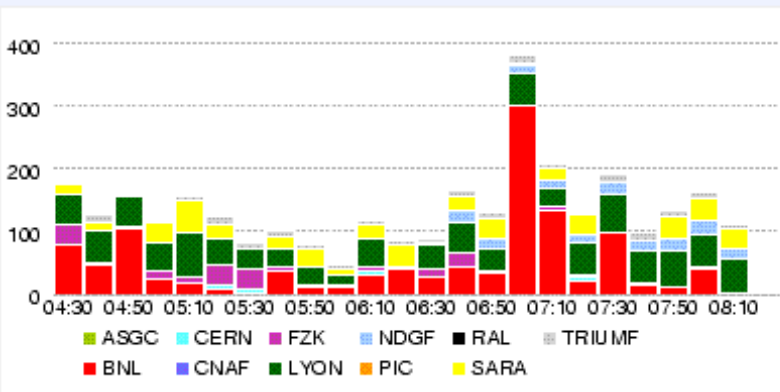
Throughput (MB/s)



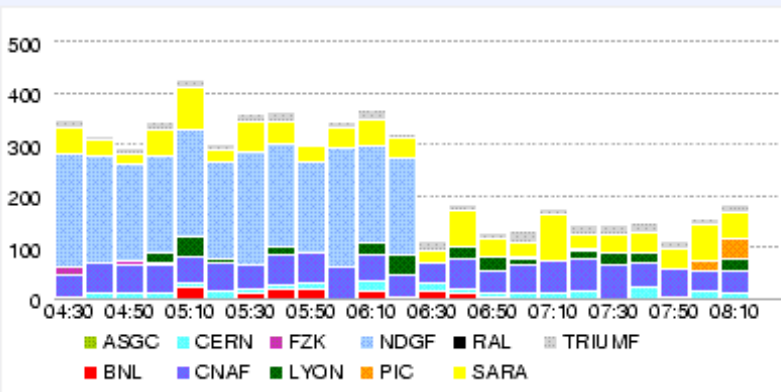
Data Transferred (GBytes)



Completed File Transfers



Total Number Errors



Activity Summary (Last 4 Hours)

Click on the cloud name to view list of sites

Cloud	Transfers					Loca
	Efficiency	Avg Throughput	Files Done	Datasets Done	Transfer	
ASGC	0%	0 MB/s	0		0	0
BNL	92%	6 MB/s	1177		108	0
CERN	15%	0 MB/s	50		280	0
CNAF	0%	0 MB/s	0		1617	0
FZK	90%	40 MB/s	203		22	0
LYON	0%	0 MB/s	0		0	0
NDGF	0%	0 MB/s	0		0	0
PIC	0%	0 MB/s	0		0	0
RAL	0%	0 MB/s	0		0	0
SARA	0%	0 MB/s	0		0	0
TRIUMF	0%	0 MB/s	0		0	0

Prodsys : traitements des donnees du detecteur

- Environ 40000 grid jobs par jour :
 - Sur plusieurs middleware,
 - Utilise les SE, CE
- Efficacite tres sensible aux :
 - « site errors » :
 - Effet « trou noir » (milliers de jobs qui plantent apres une seconde),
 - Timeouts
 - ... mais aussi les « software errors » d'ATLAS :
 - Plantage de leur framework
 - Instructions speciales pour les operateurs

OVERVIEW

view

by grid

by cloud

by dest_cloud

by executortype

by executor

by site

by cluster

by task

select grid

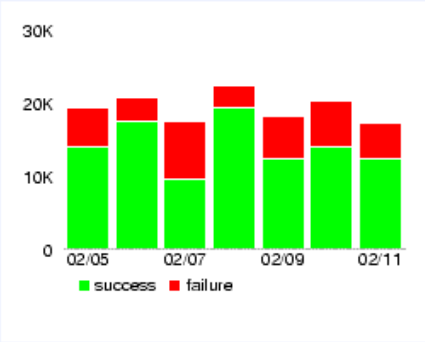
LCG

OSG

None

Nordugrid

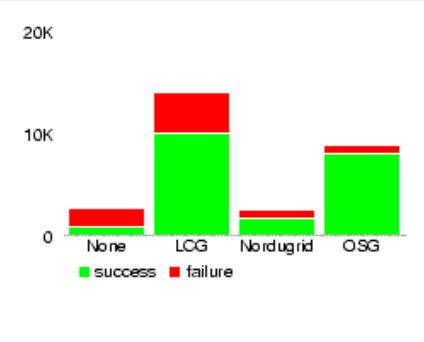
jobs last seven days



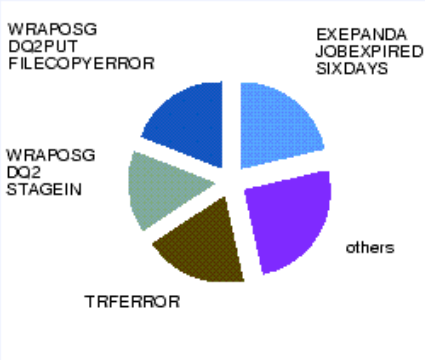
queued jobs last seven days



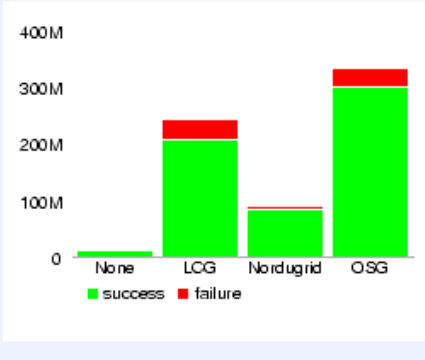
jobs (last 24h)



errors (jobs, last 24h)



walltime (seconds, last 24h)



grid	pending	stalled (>14 days)	running	success	failure
LCG	13775	0	33451	9838	4046
OSG	7882	0	5648	7893	789
None	5679	0	201	750	1818
Nordugrid	37	0	383	1648	783
total	27373	0	39683	20129	7436

CRITICAL

WARNING

NORMAL

GOOD

NO_ACTIVITY

Interet pour les sites

- Du point de vue d'ATLAS, certains sites sont inutilisables :
 - Un module python manque,
 - Stockage qui ne fonctionne pas,
 - Etc.
- Visible dans le monitoring :
 - Error code du job d'ATLAS,
 - Erreur de DQ2 relative a un service du site, etc.
- Toute cette information est dans une application dashboard.

Utilisation par les sites

- ATLAS DDM :
 - Les sites admins des tier1 d'ATLAS peuvent simplement voir la performance de leur « ATLAS cloud » (concept ATLAS),
 - Pareil pour les tier2 (inclus dans un « ATLAS cloud »)
 - Limite :
 - Erreur associee a la destination du transfert, pas forcement la faute du site destination !

Utilisation par les sites

- ATLAS prodsys :
 - Une page « summary » :
 - Nombre de jobs par jour (success/failure),
 - Erreurs principales
 - Limite :
 - Erreur pas forcément une erreur du site.

Integration avec le site monitoring

- Problemes :
 - Trop de pages web a visiter,
 - Pas toujours la faute du site :
 - Classification des erreurs difficile.
- Solution (travaux en cours) :
 - Les operateurs de la VO envoient un GGUS ticket,
 - Alertes par e-mail (un peu facile mais bon depart),
 - Integration avec Nagios (voir les travaux du WLCG monitoring working group) :
 - Sensor Nagios pour les composants ATLAS
 - Alerte Nagios avec un lien vers le dashboard

Conclusions

- Le monitoring VO est bon a prendre pour les sites
 - Les erreurs « systemes » sont monitorees,
 - Satisfaction des utilisateurs
 - Mais souvent : difficile d'identifier les erreurs des sites.
- Concertation avec ATLAS pour un « alert system »
 - Une integration au dashboard serait la bienvenue
 - Plutot que des scripts qui tournent n'importe ou.
 - Escalation flexibility:
 - Page web, e-mail vers les operateurs, RSS, messaging en general,
 - Fabric monitoring: SAM vers NAGIOS, SLS (Lemon),