RBs' jobs statistics

JRA2

Pisa, Tuesday, 25 October 2005

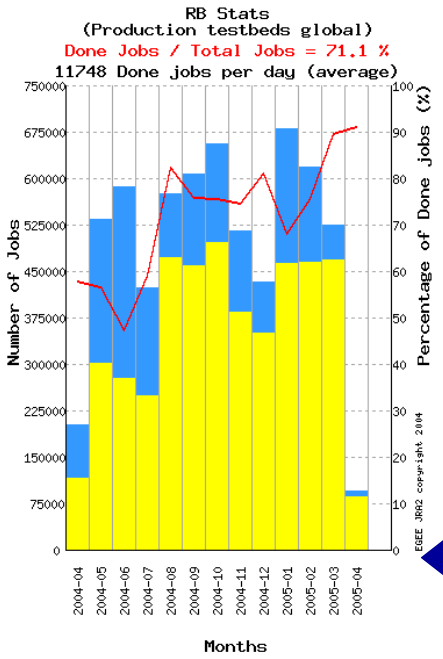
Introduction

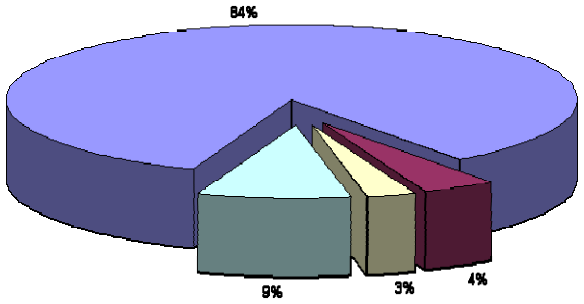
- **Data are collected from RB L&B**

Consequently, jobs submitted out of the scope of the RBs are not taken into account. Complete job throughput is provided by GOC accounting.

  - Currently we gathered
    - *About 60 Go data since January*
    - *More than 6 Millions jobs*
    - *55 RBs collected*

Architecture





intermediate data build
statistics computations

VO1 LDAP directory

VO2 LDAP directory

VOp LDAP directory

VOn LDAP directory

userid

statistics

Database

Raw data

Intermediate data

Statistics data

events

short_fields

long_fields

L&B
RB 1

L&B
RB 2

L&B
RB P

L&B
RB n

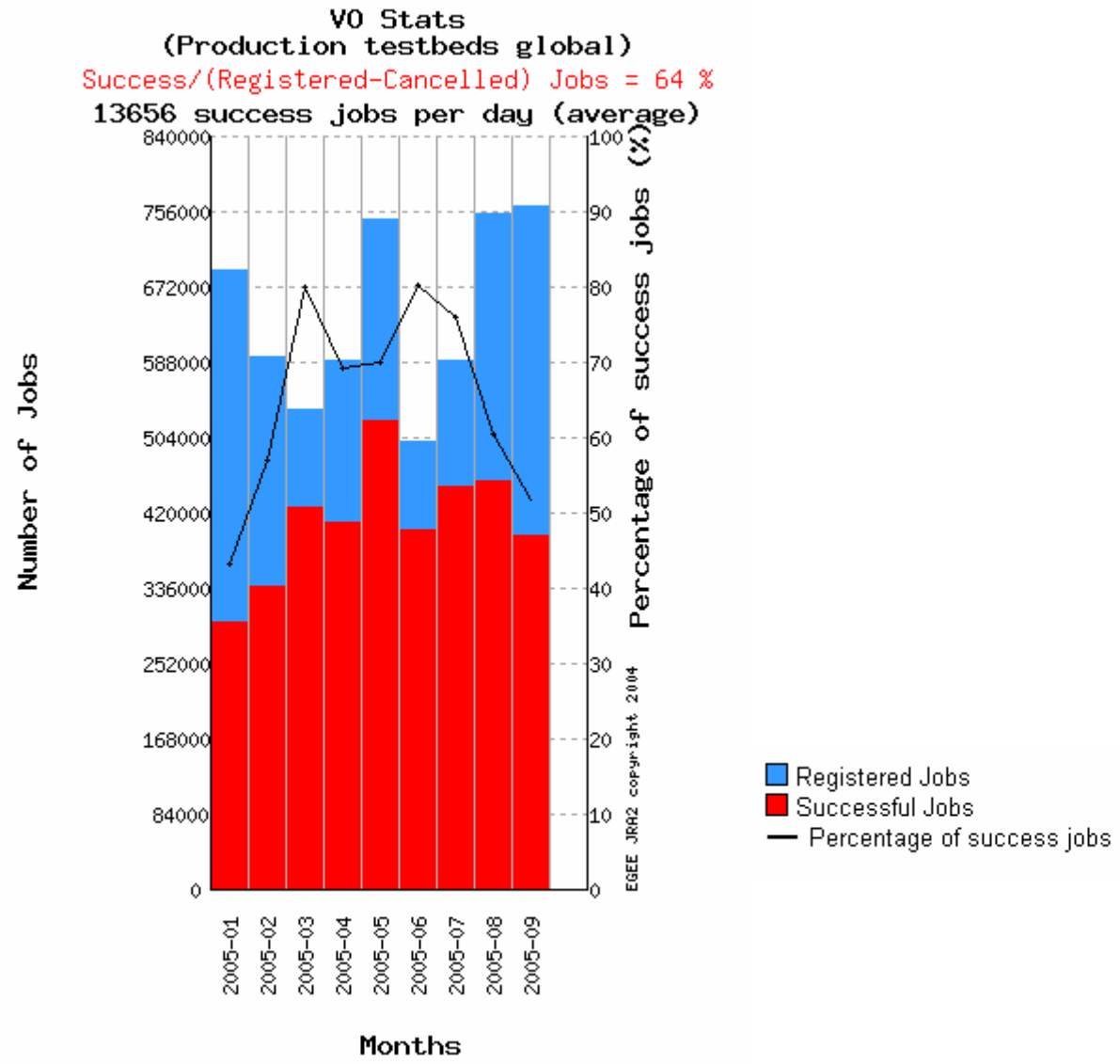
Throughput and success rate 1/2

- **Job throughput and job success (RB point of view)**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/showstatsVO.php

- **Job throughput per month (per VO or per RB)**
  - nb of registered jobs on the RBs, nb of sucessful jobs on the RBs
  - sucessful job = job done and status 'OK'
  - This reflects the RB point of view, not the success from the applications point of view. For instance the file transfer failures are not taken into account.

- **Success rate per month (per VO or per RB)**
  - success rate = nb of success / (nb of registered - nb of cancelled)

Throughput and success rate 2/2



Per site throughput

- **Job throughput distribution per country, site and CE**
  - presents the total values for run jobs from 2005
    - The CE information is know when the job runs
    - From CE name we catch
      - *the site with the mapping table CE-Site from the GOC database*
      - *the country with the ccTLD (country code Top Level Domains)*

MENU

- Austria
- Bulgaria
- Canada
- China
- Cyprus
- Czech Republic
- France
- Germany
- Greece
- Hungary
- India
- Ireland
- Israel
- Italy
- Japan
- Korea Republic of
- Netherlands
- Pakistan
- Poland
- Portugal
- Romania
- Russian Federation
- Singapore
- Slovak Republic
- Spain
- Sweden
- Switzerland
- Taiwan
- United Kingdom

France > IN2P3-CC

month	cclcgceli01.in2p3.fr	cclcgceli02.in2p3.fr	cclcgceli04.in2p3.fr	cclcgceli05.in2p3.fr	cclcgceli07.in2p3.fr	
2005-01	1248	1	0	4	4	1257
2005-02	1932	1659	0	21	0	3612
2005-03	2711	1298	0	4	0	4013
2005-04	4927	1149	6094	0	3582	15752
2005-05	11637	2437	2625	9	1798	18506
2005-06	2406	2764	0	10	0	5180
2005-07	12635	10189	0	60	0	22884
2005-08	21828	0	0	2	0	21830
2005-09	12538	758	0	19	0	13315
2005-10	0	6398	0	3	0	6401

Sites top ten

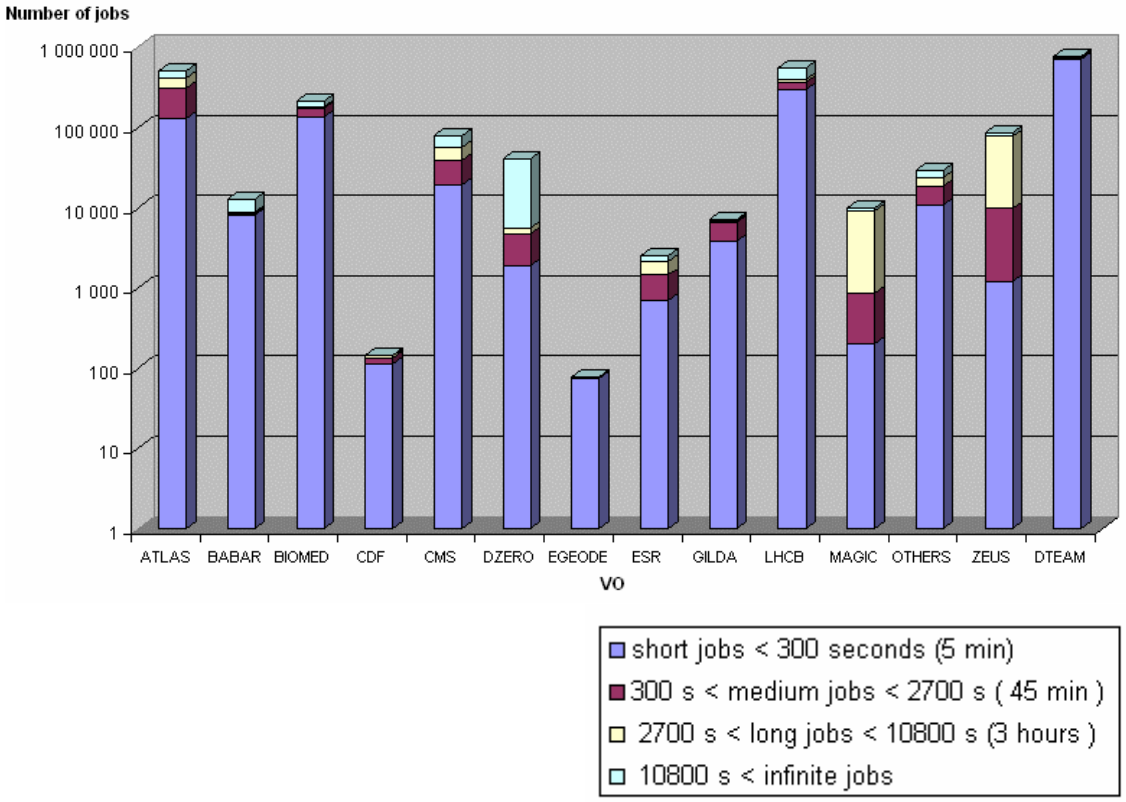
- **Presents the 10 sites of largest throughput**

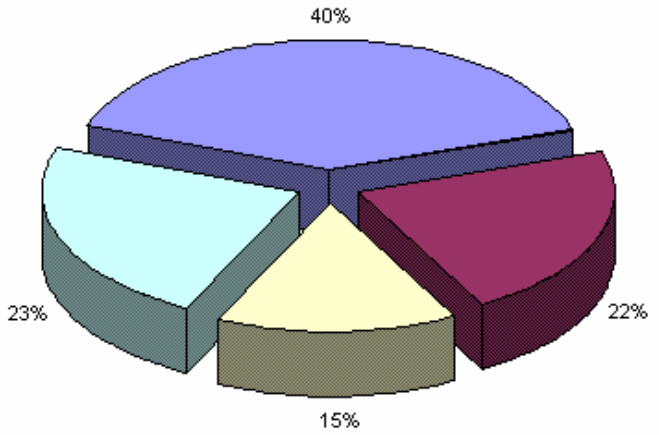
SITES TOP 10

2005-05		2005-06		2005-07	
CNAF-T1	26941	CNAF-T1	56590	RAL-LCG2	70674
RAL-LCG2	22778	RAL-LCG2	32148	RAL-LCG2	57163
IN2P3-LPC	15885	RALPP-LCG	12684	CNAF-T1	25728
CERN-PROD	13783	FNAL-LCG2	11961	RALPP-LCG	16028
INFN-LNL-LCG	13191	IN2P3-LPC	10293	FNAL-LCG2	14914
IN2P3-LPC	12595	RAL-LCG2	9619	RALPP-LCG	14125
RALPP-LCG	11957	NIKHEF-ELPROD	9107	CERN-PROD	12653
IN2P3-CC	11620	DESYPRO	8616	RWTH-Aachen	12374
NIKHEF-ELPROD	11324	INFN-PADOVA	7703	RWTH-Aachen	12241
USC-LCG2	11116	DESYPRO	7516	NIKHEF-ELPROD	11496
2005-08		**2005-09**		**2005-10**	
RAL-LCG2	41658	RAL-LCG2	50107	RAL-LCG2	24501
CNAF-T1	28585	FNAL-LCG2	39662	FNAL-LCG2	18568
FNAL-LCG2	23911	CERN-PROD	20040	CERN-PROD	15171
CERN-PROD	21916	CNAF-T1	16073	CNAF-T1	7598
csTCDie	19072	RALPP-LCG	12823	INFN-BARI	7313
CNAF-T1	15427	IN2P3-CC	12538	IN2P3-CC	6398
csTCDie	15213	INFN-BARI	10226	pic	5960
IN2P3-CC	15084	pic	9166	INFN-PADOVA	4852
CERN-PROD	13646	IN2P3-LPC	8949	UCL-CCC	4849
DESYPRO	13615	IN2P3-LPC	8746	RALPP-LCG	4637

Duration

- **Duration distribution**
  - the duration calculation is possible for successful jobs and when the run time-stamp and the done time-stamp on the CE are both available.

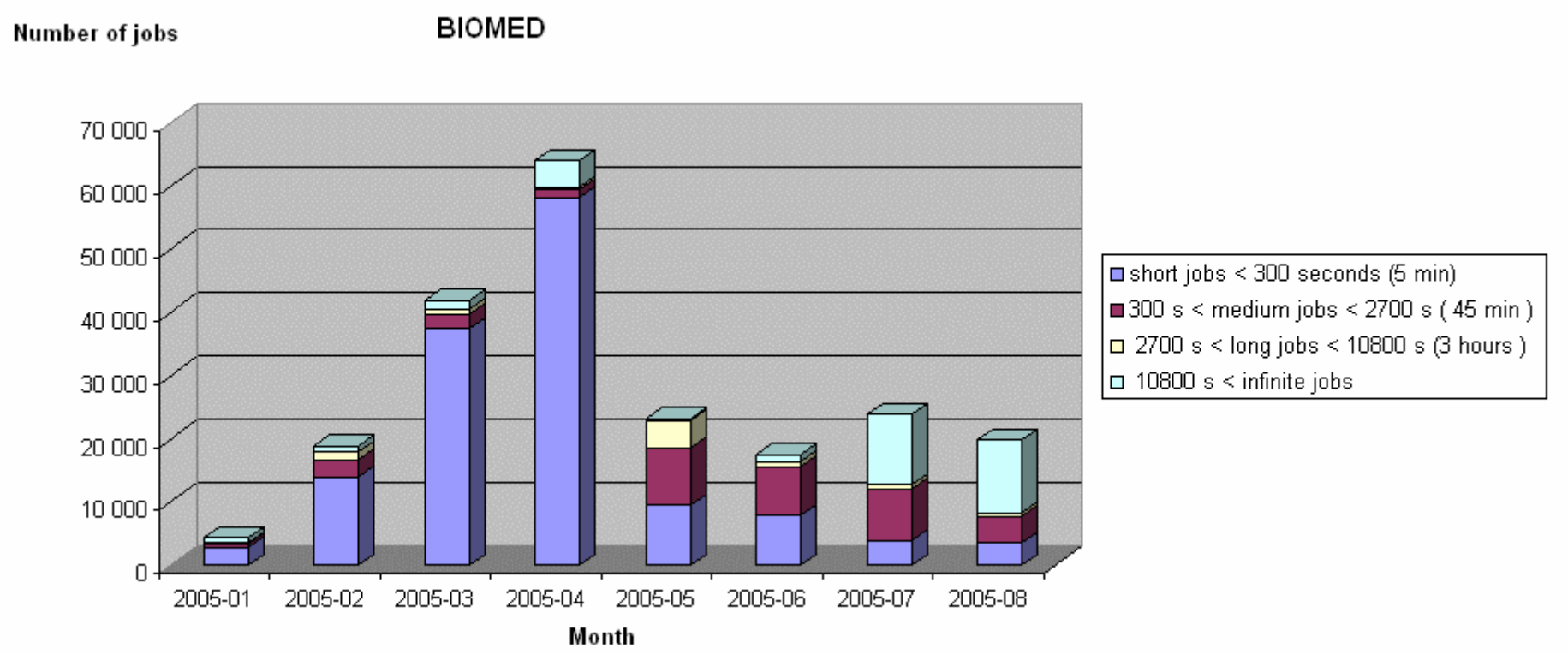




Job duration distribution without Dteam

Duration example

- One can see the largest number of infinite jobs during July and August. They are due to the Wisdom DC.



Job waiting time

- Job execution time and job waiting time for each VO
- http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/efficiencytime.php
  - *D1 : The time stamp when the job is registered on RB by UI*
  - *D2 : The time stamp when the job state is run on CE*
  - *D3 : The time stamp when the job is done on CE*
- Execution time: ET = D3-D2 , Waiting Time: WT = D2-D1.
- ETmax is the max bound for each category (5 mn for short jobs, 45 mn for medium jobs , 3 h for long jobs , and 15 h for infinite jobs).

Overall - Job Waiting Distribution

Type of jobs	Total Number of jobs	WT < ETmax (%)	ETmax < WT < 5 ETmax (%)	WT > 5 ETmax* (%)
infinite	797288	76	6	16
long	348830	75	12	12
medium	361618	65	15	18
short	1586020	63	17	19

Job abort reasons distribution

- **Available for aborted jobs**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/abort_reason.php

ABORT REASONS DISTRIBUTION

The distribution of abort reasons is the following (data have been collected from January to September 2005) :

Number of Abort Reasons	Percentage of this abort Reason	Abort Reason
395088	70 %	Cannot plan:BrokerHelper
91219	16 %	Job proxy is expired.
29589	5 %	Submission to condor failed.
24677	4 %	cannot retrieve previous matches
6169	1 %	cannot retrieve original JDL
6117	1 %	Unable to receive data
5783	1 %	IOException
3149	<1 %	AuthenticationException
2797	<1 %	Failure while executing job wrapper
2104	<1 %	SandboxIOException
896	<1 %	Unable to read data
256	<1 %	JobSizeException: Job Size exceeds limits.
227	<1 %	Cannot plan: JobAdapterHelper
86	<1 %	ProxyRenewalException: Error during Proxy Renewal registration.
22	<1 %	proxy expired
17	<1 %	Globus Ftp API Failure in creating remote Directories.
14	<1 %	JDLParsingException: Error while parsing Jdl string.
12	<1 %	Unable to send data
7	<1 %	Error during proxy renewal registration
3	<1 %	Cannot create condor submit file

Active users

- **Active users per VO**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/users_active.php?choix=Overall
  - A user has been considered as active when he has launched more than 10 jobs in the month.
  - If he has launched at least one job and less than 10, he is considered as inactive.

Overall

Month	Nb of active users	Nb of inactive Users
2005-01	143	160
2005-02	180	172
2005-03	210	190
2005-04	166	139
2005-05	190	122
2005-06	149	150
2005-07	255	136
2005-08	185	126

Future

- **Job provenance will offer us a more comfortable collect of data (easier, efficient, reliable) and a mean to obtain the real and sure "job history" record (middleware linked) we need to provide statistics**
  - we are preparing the changes

- **We are willing to refine our statistics and provide new ones**
  - e.g. abort reasons to be refined
  - e.g. total duration estimate for successful jobs to be built
  - new schema summarizing job throughput, job duration and job success

- **In collaboration with GOC accounting people we would like to evaluate the job throughput using other jobs submission mechanisms (direct access to the CE, Dirac…)**

Urls summary

- **JRA2 statistics**
  - http://egee-jra2.web.cern.ch/EGEE-JRA2/QoS/JobsMetrics/JobMetrics.htm
- **Job throughput and job success**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/showstatsVO.php
- **Duration distribution**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/durationmonthlyVO.php
- **Per site throughput**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/sites_Ce.php
- **Waiting time**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/efficiencytime.php
- **Abort reasons**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/abort_reason.php
- **Active users**
  - http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/users_active.php?choix=Overall