



DIRAC testbed

Bruno Santeramo - 22/11/2011



todo list



1. stop Dirac v5
2. install Dirac v6 (keeping all files and DBs of v5)
3. understand how to configure Cream CEs
 1. enable Direct mode submission where possibile
 2. otherwise use WMS glite submission
4. simple job submission test:
 1. jobs like: "hello world" + sleep(180)
 2. each job is parametric (100 parameters)
 3. in total 10.000 task to be executed
 4. ensure filling mode works properly
5. input data test:
 1. 100-200 MB input data (replicated on 5-6 SEs)
 2. 100-200 jobs to submit
 3. each job is parametric (at least 10 parameters)
 4. in total more than 1000 tasks to be executed
 5. ensure filling mode continue to works properly

Deadline: 20 Nov 2011

Goal: comparative tests begin from 21 Nov 2011



work progress status



1. stop Dirac v5
2. install Dirac v6 (keeping all files and DBs of v5)
3. understand how to configure Cream CEs
 1. enable Direct mode submission where possible
 2. otherwise use WMS glite submission
4. simple job submission test:
 1. jobs like: "hello world" + sleep(180)
 2. each job is parametric (100 parameters)
 3. in total 10.000 task to be executed
 4. ensure filling mode works properly
5. input data test:
 1. 100-200 MB input data (replicated on 5-6 SEs)
 2. 100-200 jobs to submit
 3. each job is parametric (at least 10 parameters)
 4. in total more than 1000 tasks to be executed
 5. ensure filling mode continue to works properly

Deadline: 20 Nov 2011

Goal: comparative tests begin from 21 Nov 2011



simple job submission test



100 simple parametric jobs
echo "hello world"
sleep 180
each job has 100 parameters
total of 10.000 jobs simultaneously submitted

Test results:

100% jobs successfully submitted
100% jobs successfully executed
filling mode properly working
10 sites executed jobs
4.962 pilots per 10.000 jobs (ratio: 2.02)
a single pilot executed up to 5 jobs



input data test



4 test using (not trivial) input file: 100MB, 200MB, 500MB, 1GB

used DFC (DIRAC File Catalog)

files uploaded via `dirac-dms-add-file <LFN> <SE>`

and replicated via `dirac-dms-replicate <LFN> <SE>`

6 replicas of each file

SEs used:

1. GRIF (LAL-USER)
2. GRISU-UNINA (GRISU-UNINA)
3. INFN-NAPOLI-ATLAS (NAPOLI-INFN)
4. RAL-LCG2 (GridPP)
5. UKI-SOUTHGRID-OX-HEP (OXFORD)
6. VICTORIA-LCG2 (UVIC)



input data test



LFN:/superbvo.org/user/b/bsanteramo/test/<FILE>

- srm://grisuse.scope.unina.it:8446/srm/managerv2?SFN=/dpm/scope.unina.it/home/superbvo.org/superbvo.org/user/b/bsanteramo/test/<FILE>
- srm://srm-superb.gridpp.rl.ac.uk:8443/srm/managerv2?SFN=/castor/ads.rl.ac.uk/prod/superb/superbvo.org/user/b/bsanteramo/test/<FILE>
- srm://grid05.lal.in2p3.fr:8446/srm/manager?SFN=/dpm/lal.in2p3.fr/home/superbvo.org/user/b/bsanteramo/test/<FILE>
- srm://t2-dpm-01.na.infn.it:8446/srm/managerv2?SFN=/dpm/na.infn.it/home/superbvo.org/superbvo.org/user/b/bsanteramo/test/<FILE>
- srm://t2se01.physics.ox.ac.uk:8446/srm/managerv2?SFN=/dpm/physics.ox.ac.uk/home/superbvo.org/superbvo.org/user/b/bsanteramo/test/<FILE>
- srm://lcg-se.rcf.uvic.ca:8443/srm/managerv2?SFN=/pnfs/rcf.uvic.ca/data/superbvo/superbvo.org/user/b/bsanteramo/test/<FILE>



input data test



200 simple parametric jobs

md5sum <FILE>

each job has 10 parameters

total of 2.000 jobs simultaneously submitted

4 tests: (100MB, 200MB, 500MB, 1GB)

Test results:

100% jobs successfully submitted

99.99% jobs successfully executed (1/8000 jobs failed)

filling mode properly working

10 sites executed jobs

3.539 pilots per 8.000 jobs (ratio: 2.26)

a single pilot executed up to 5 jobs



simple jobs test



Site	jobs	pilots	ratio
CIT_HEP_CE	249	249	1,00
INFN-T1	5248	1402	3,74
CYFRONET-LCG2	1325	1325	1,00
INFN-FRASCATI	160	160	1,00
INFN-LNL-2	928	193	4,81
INFN-MILANO-ATLASC	546	113	4,83
INFN-NAPOLI-ATLAS	91	91	1,00
INFN-PISA	1219	1219	1,00
UKI-LT2-QMUL	35	11	3,18
UKI-SOUTHGRID-RALPP	199	199	1,00
SUMMARY	10000	4962	2,02



input data test



100MB

Site	jobs	pilots	ratio
CIT_HEP_CE	53	53	1,00
INFN-T1	440	98	4,49
CYFRONET-LCG2	99	99	1,00
INFN-FRASCATI	372	372	1,00
INFN-LNL-2	71	16	4,44
INFN-MILANO-ATLASC	58	20	2,90
INFN-NAPOLI-ATLAS	166	166	1,00
UKI-LT2-QMUL	738	159	4,64
UKI-SOUTHGRID-RALPP	3	3	1,00
SUMMARY	2000	986	2,03

200MB

Site	jobs	pilots	ratio
CIT_HEP_CE	41	41	1,00
INFN-T1	1284	327	3,93
INFN-FRASCATI	68	68	1,00
INFN-LNL-2	56	16	3,50
INFN-MILANO-ATLASC	71	44	1,61
INFN-PISA	141	141	1,00
UKI-LT2-QMUL	339	84	4,04
SUMMARY	2000	721	2,77

500MB

Site	jobs	pilots	ratio
CIT_HEP_CE	63	63	1,00
INFN-T1	1162	286	4,06
INFN-FRASCATI	154	154	1,00
INFN-LNL-2	111	33	3,36
INFN-MILANO-ATLASC	69	54	1,28
INFN-PISA	213	213	1,00
UKI-LT2-QMUL	228	67	3,40
SUMMARY	2000	870	2,30

1GB

Site	jobs	pilots	ratio
CIT_HEP_CE	66	66	1,00
INFN-T1	1035	264	3,92
CYFRONET-LCG2	10	10	1,00
INFN-FRASCATI	87	87	1,00
INFN-LNL-2	148	55	2,69
INFN-MILANO-ATLASC	75	67	1,12
INFN-NAPOLI-ATLAS	80	80	1,00
INFN-PISA	228	228	1,00
UKI-LT2-QMUL	271	105	2,58
SUMMARY	2000	962	2,08



getting info via DIRAC API



DIRAC offers an interesting set of API to retrieve info on jobs

I wrote a simple set of script to retrieve useful info to:

- measure jobs efficiency (status of jobs)
- pilots distribution
- jobs distribution
- filling mode efficiency

each info can be referred:

- per sites
- per CE
- per queue
- per WN



getting info via DIRAC API



and more... (I have not still wrote function)

- submission time spread
- mean execution time
- fastest and slowest pilot/job
- analyze logging info for every pilot/job
 - time to download input
 - time to execute program
 - time to upload output
- more...

analyzing pilot output more info can be retrieved:

- SE used for every input file
- when pilot in filling mode, SEs used for every job executed
- more... (pilot's outputs are very detailed)