Workflow Generator and Tracking at the Rescue of Distributed Processing.

Automating the Handling Grid Production for the STAR Collaboration

LAURET, Jérôme
HAJDU, Levente (Presenter)
DIDENKO, Lidia

CHEP 09 Conference

BROOKHAVEN NATIONAL LABORATORY

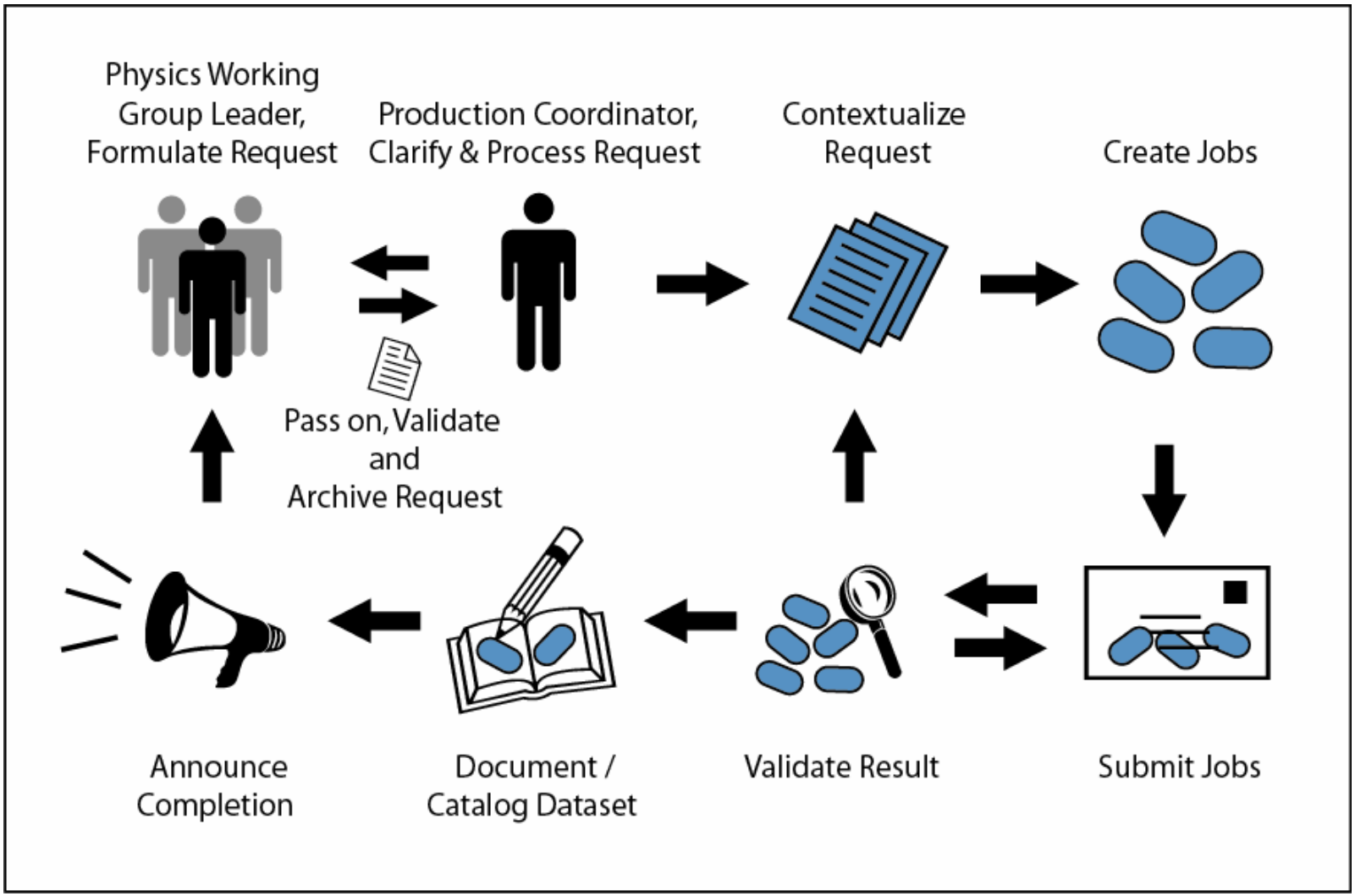
Office of Science
U.S. DEPARTMENT OF ENERGY

STAR

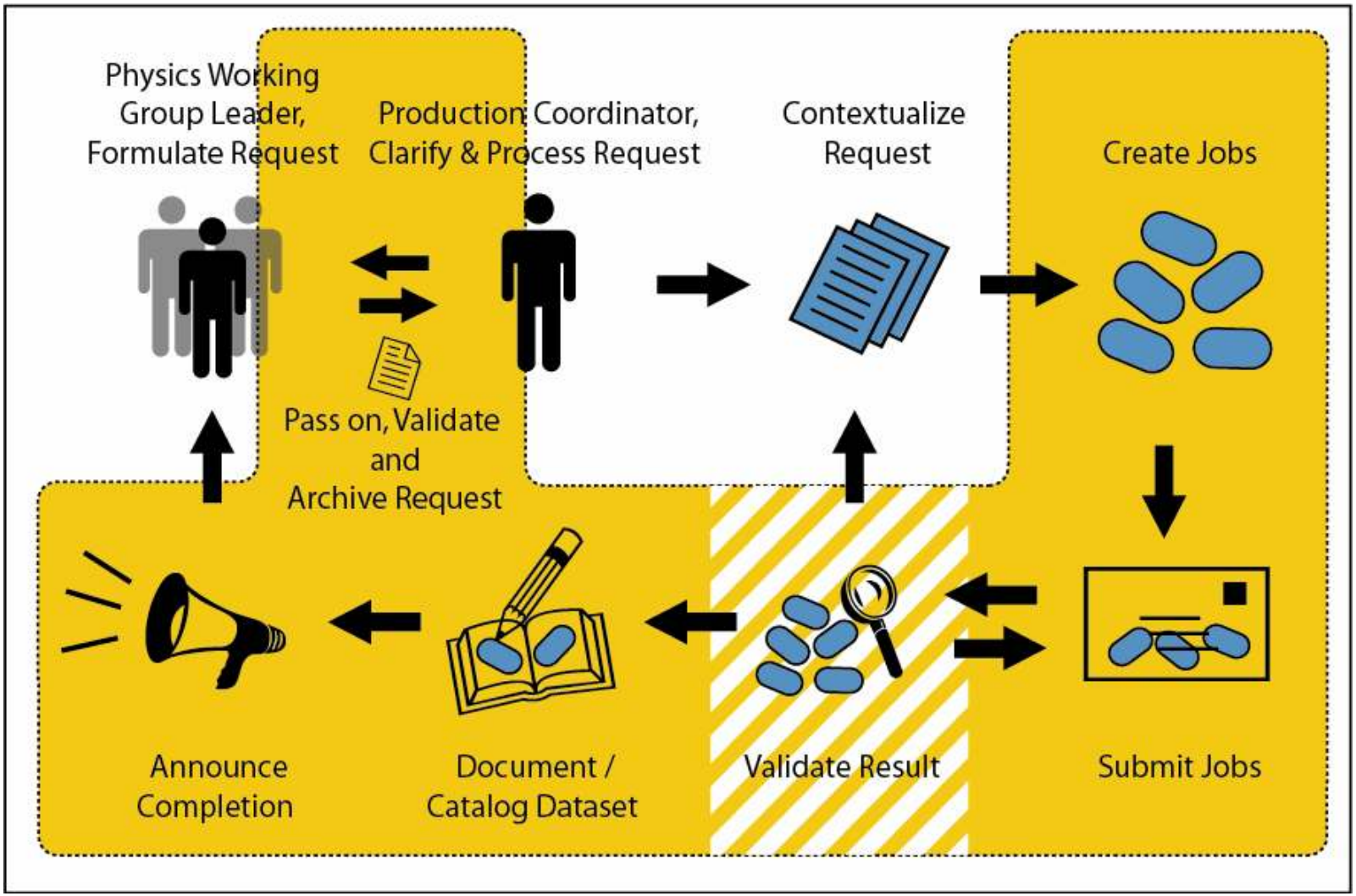
Outline

- The Production Cycle
- Automat-able Aspects of the production cycle
- The Production level software tools of the production coordinator
  - Simulation Production Request API
  - STAR Unified Meta Scheduler
  - The Job Feeder
    - Feeder flow chart
    - Redirection example
  - The Job Scanner
    - Simplified Example
  - Resubmission
- Detailed Overall Diagram
  - Charts derived from information collected by the job scanner
    - Dataset accumulation plot
    - Causes of failure plot
    - Percent success after resubmission plot
    - Data Transfer plot

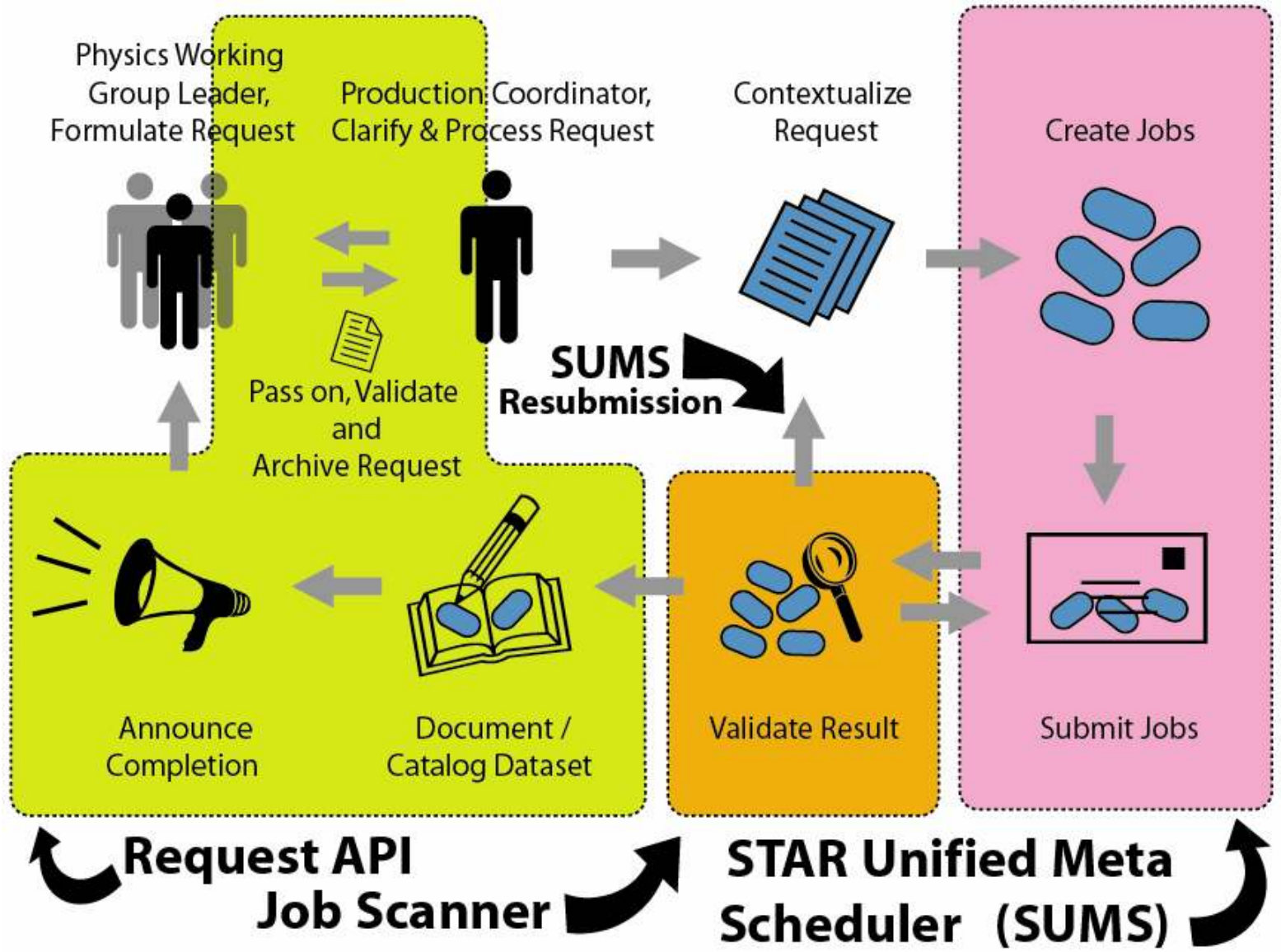
Production Cycle



Production Cycle (Automate-able)

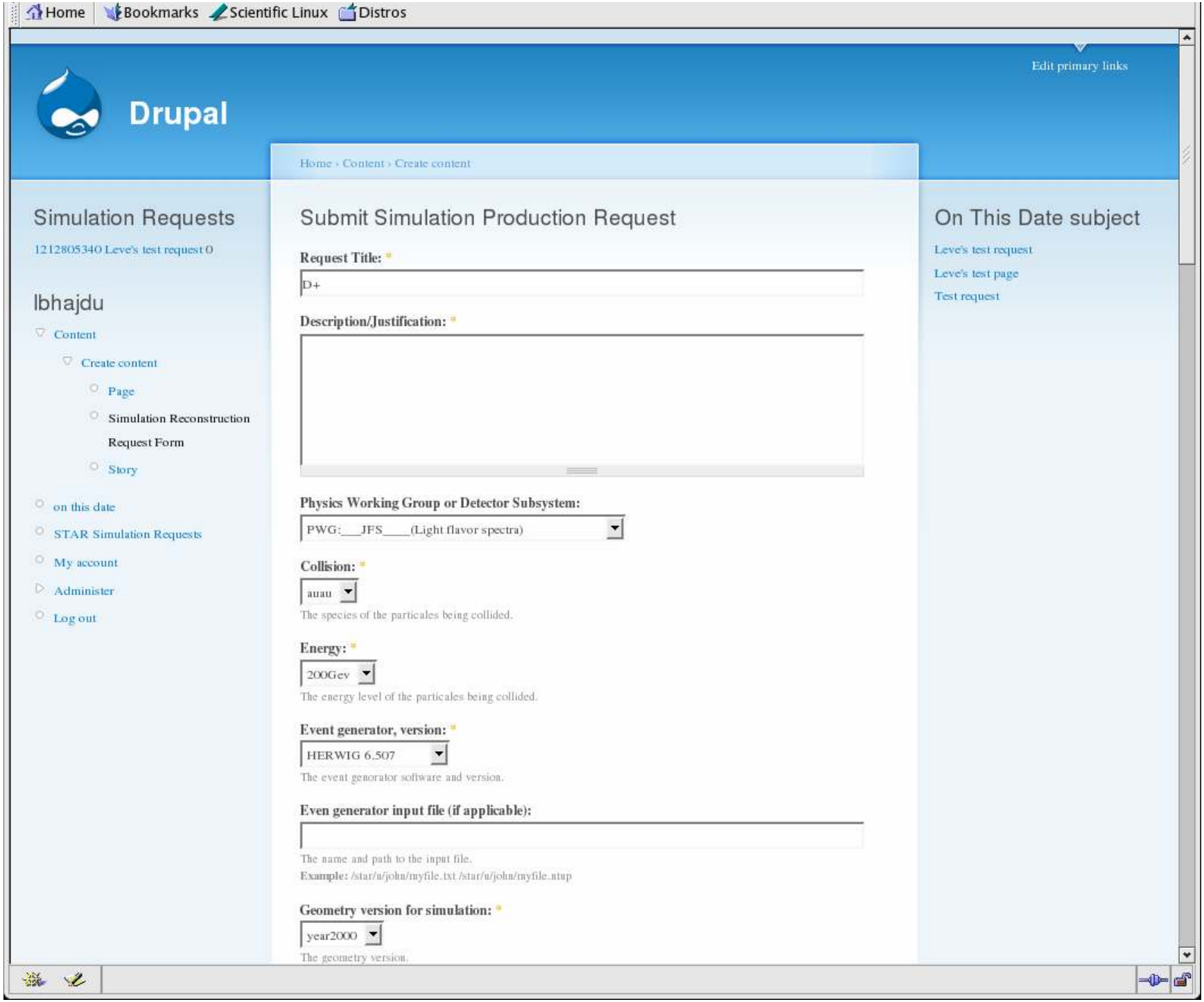


Production Cycle (Software Tools)



Simulation Request API

- Request API
  - Advanced language allows for flexible/complex Input validation
  - Requests are fully documented and reproducible (database back-end)
  - Custom rolls
    - **Users** - can view
    - **PWGC** (*physics working group convener*) -can submit request
    - **Production manager -** can do everything (submit, modify, delete, update statues)
- Drupal benefits
  - Content Management system used in STAR
  - Written in PHP - lots of expertise to maintain custom modules
  - Drupal provides access level rights (ACL).
  - Self-consistent with STAR web support plans

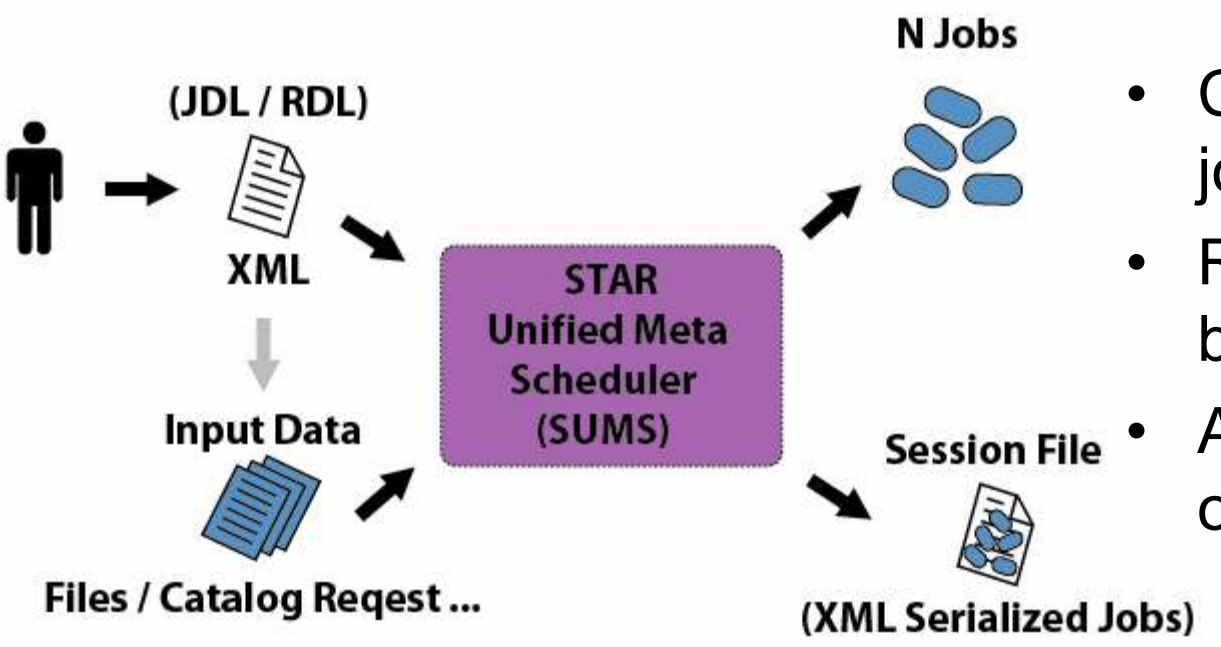

Home
Bookmarks
Scientific Linux
Distros
Drupal
Edit primary links
Home › Content › Create content
Simulation Requests
1212805340 Leve's test request 0
lbhajdu
Content
Create content
Page
Simulation Reconstruction Request Form
Story
on this date
STAR Simulation Requests
My account
Administer
Log out
Submit Simulation Production Request
Request Title: *
D+
Description/Justification: *
Physics Working Group or Detector Subsystem:
PWG:___JFS____(Light flavor spectra)
Collision: *
auau
The species of the particales being collided.
Energy: *
200Gev
The energy level of the particales being collided.
Event generator, version: *
HERWIG 6.507
The event genorator software and version.
Even generator input file (if applicable):
The name and path to the input file.
Example: /star/u/john/myfile.txt /star/u/john/myfile.ntup
Geometry version for simulation: *
year2000
The geometry version.
On This Date subject
Leve's test request
Leve's test page
Test request

SUMS (STAR Unified Meta Scheduler)

- Discussed in:

CHEP06 Meta-configuration for dynamic resource brokering: the SUMS approach

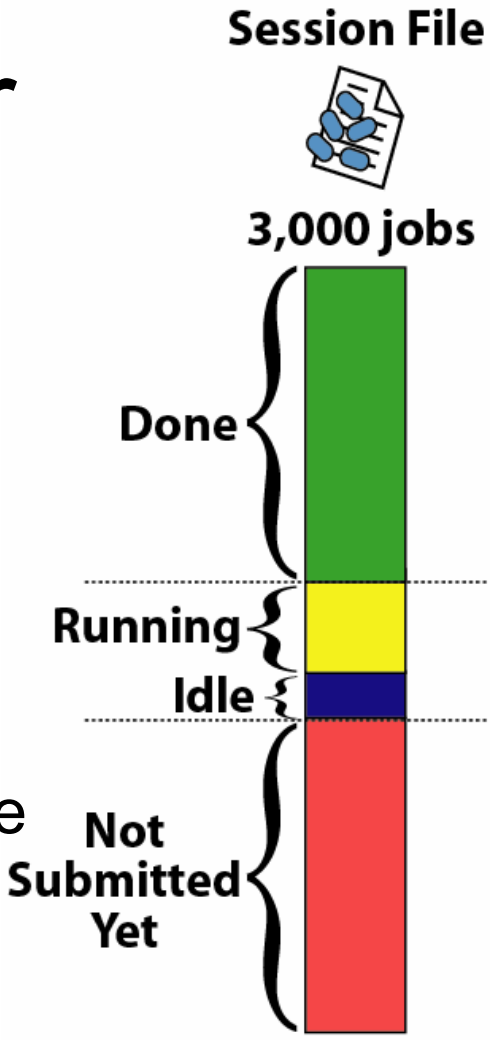
CHEP04



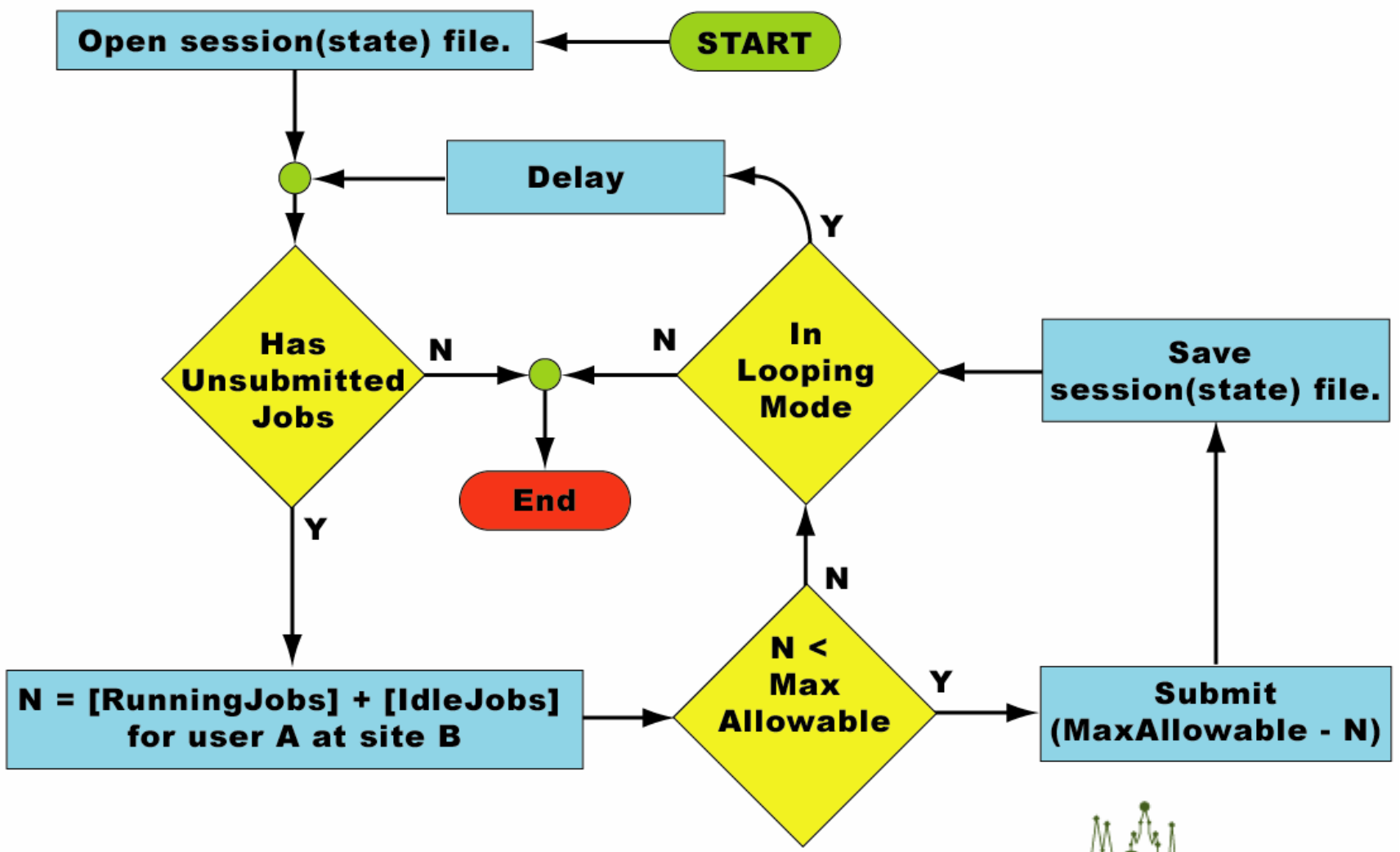
- Converts the users xml job descript into *N* jobs.
- Resolves dataset binding
- Allows for resubmitting of failed jobs

The Job Feeder

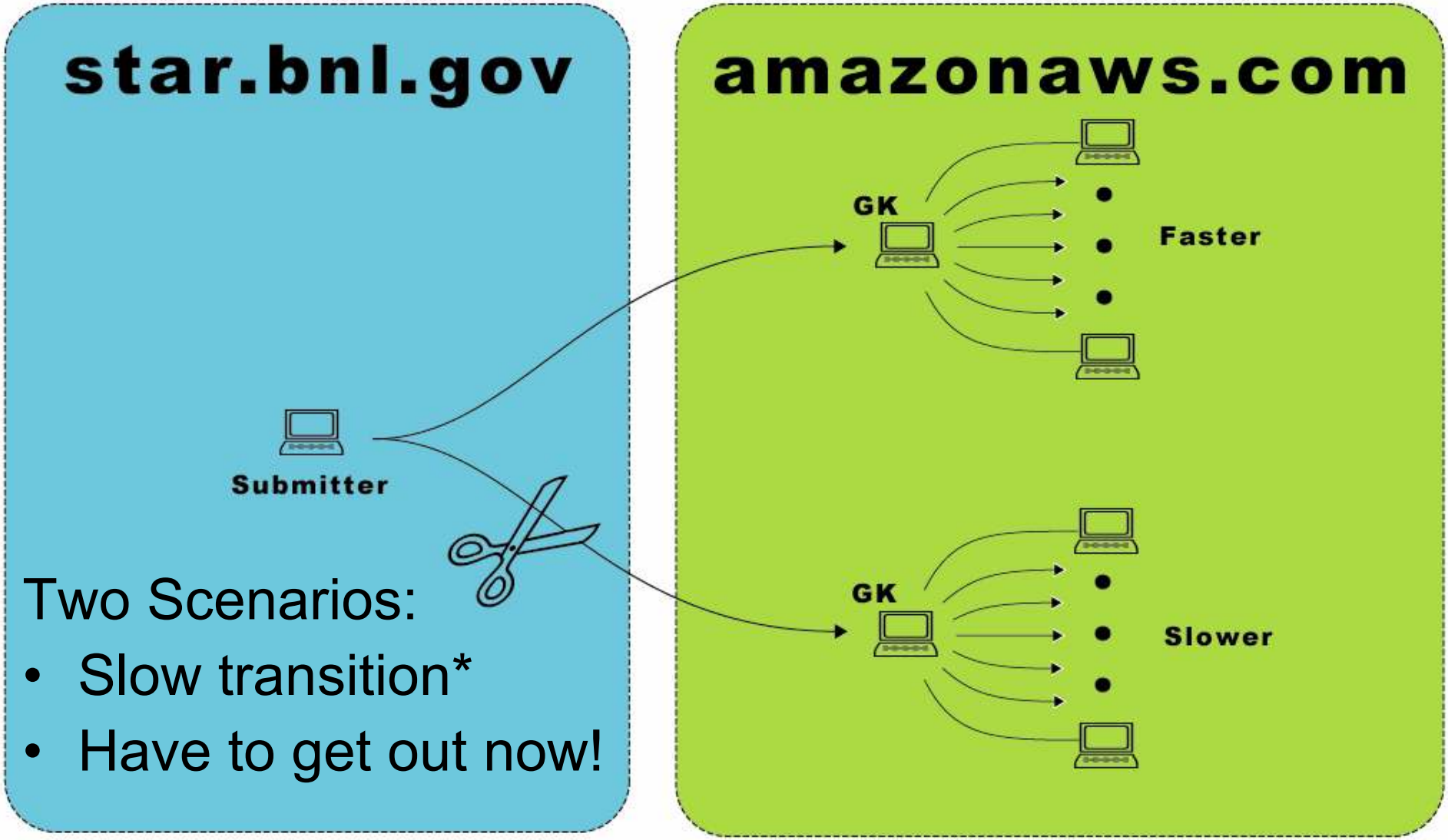
- Keeps queue topped up, does not submit all jobs at once.
- Increases scalability
- Automates submission
- Keeps queue at max occupancy
- Allows for fine tuning of resource usage
  - Ask for only N slots at a time
- Reduces recovery time from cluster failure
  - Limits number of jobs exposed to failure
  - Can be used to prevent submission to black-hole sites (feeding too quickly)



The Job Feeder



The Job Feeder
(Redirection After Submission)

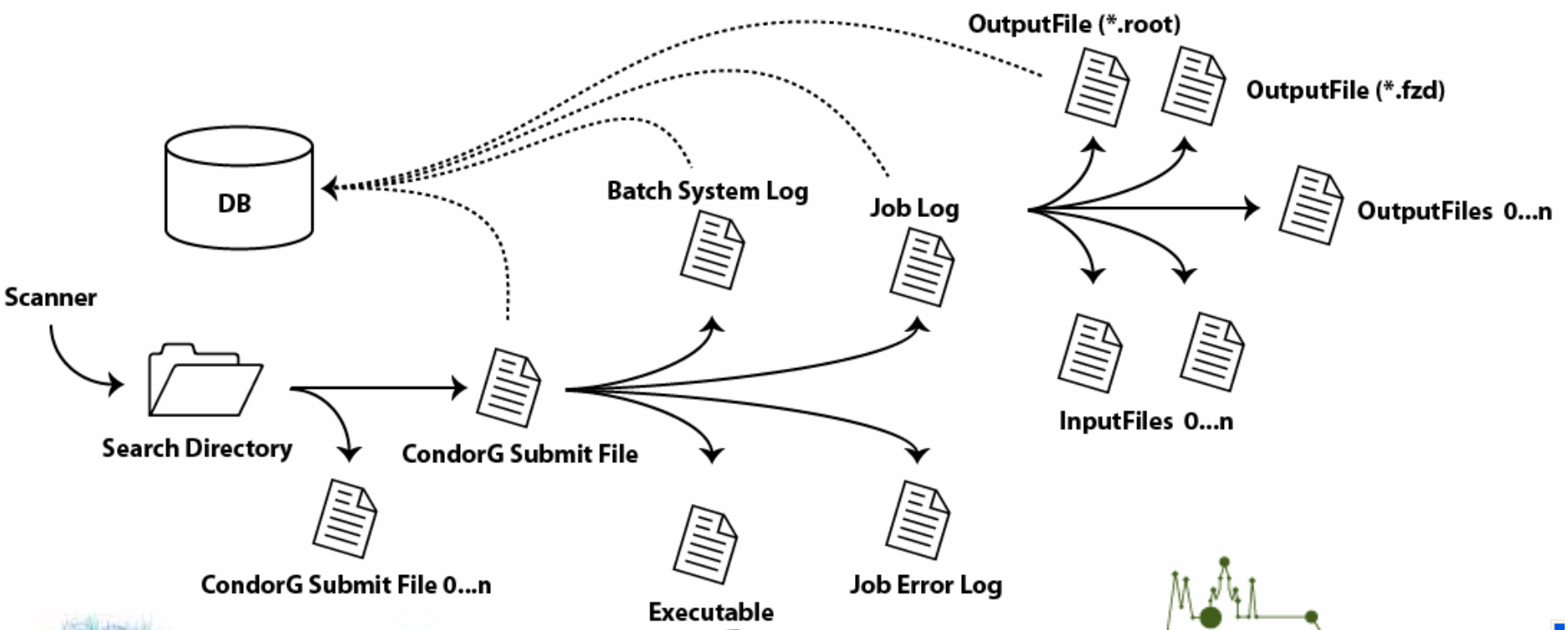


Two Scenarios:

- Slow transition*
- Have to get out now!

The Job Scanner

- Records the state of jobs
- Scans standard job structure (parses the job tree)
  - Pattern recognition of standard success and error states from keywords in log files (Example: “segmentation violation”)
  - Input/Out file sizes and MD5sums are captured.

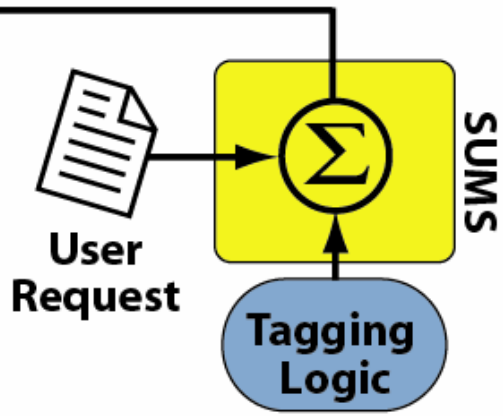


Simplified Example

Executable .csh

```
echo "SUMS_RECO_START"
root4star -b -l -q bfc.C\(${Events},\"trs,srs,ssd,fss,y2007g ... ${INPUT}\"\)
echo "SUMS_RECO_END"

echo "SUMS_OUTPUT_FILE_NAME"="${finalDirectory}/$SUMS_OUTPUT_FILE
echo "SUMS_OUTPUT_FILE_MD5="`md5sum $SUMS_OUTPUT_FILE`
echo "SUMS_OUTPUT_FILE_CPSTART="`/bin/date`
globus-url-copy -p 25 file:$SCRATCH/${SUMS_OUTPUT_FILE}
            gsiftp://stargrid.rcf.bnl.gov${finalDir}/${SUMS_OUTPUT_FILE}
echo "SUMS_OUTPUT_FILE_CPEND="`/bin/date`
```



Scanner Config File

#DB col name	Sring to mattch	Data Type	File Type	Primary key
jobID_MD5	"^setenv REQUESTID ([A-F0-9]{32})$"	STRING	"^.*\.csh$"	PK
processID	"^setenv JOBINDEX ([0-9]*)$"	INT	"^.*\.csh$"	PK
submitAttempt	"^setenv SUBMITATTEMPT ([0-9]*)$"	INT	"^.*\.csh$"	PK
outputMD5	"^SUMS_OUTPUT_FILE_MD5=([A-F0-9]{32})$"	STRING	"^.*\.(out|log)$"	
cpStart	"^SUMS_OUTPUT_FILE_CPSTART=(.*)$"	DATE	"^.*\.(condorg)$"	
cpEnd	"^SUMS_OUTPUT_FILE_CPEND=(.*)$"	DATE	"^.*\.(out|log)$"	

Output File (stdout)

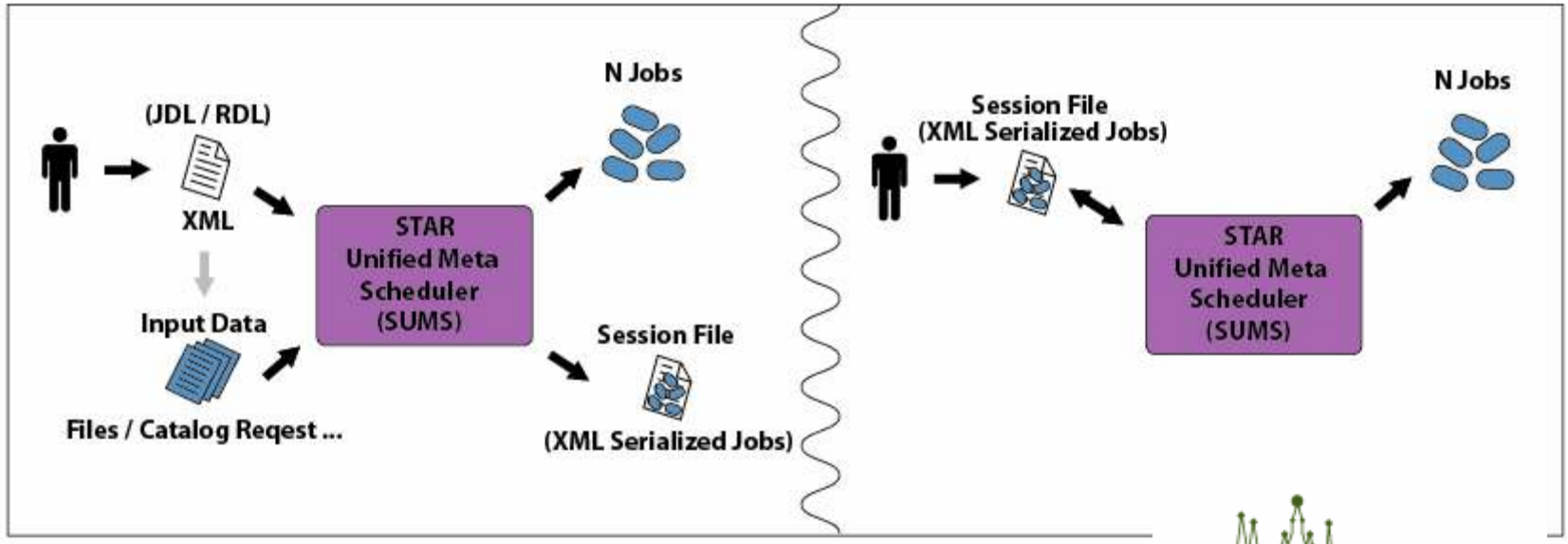
```
...
SUMS_OUTPUT_FILE_NAME= /star/data08/a/b/c.tags.root
SUMS_OUTPUT_FILE_MD5=4BF58A76682503014E404A96DC94D9D3
SUMS_OUTPUT_FILE_CPSTART=Thu Feb 5 19:47:28 PST 2009
SUMS_OUTPUT_FILE_CPEND=Thu Feb 5 19:50:45 PST 2009
...
```

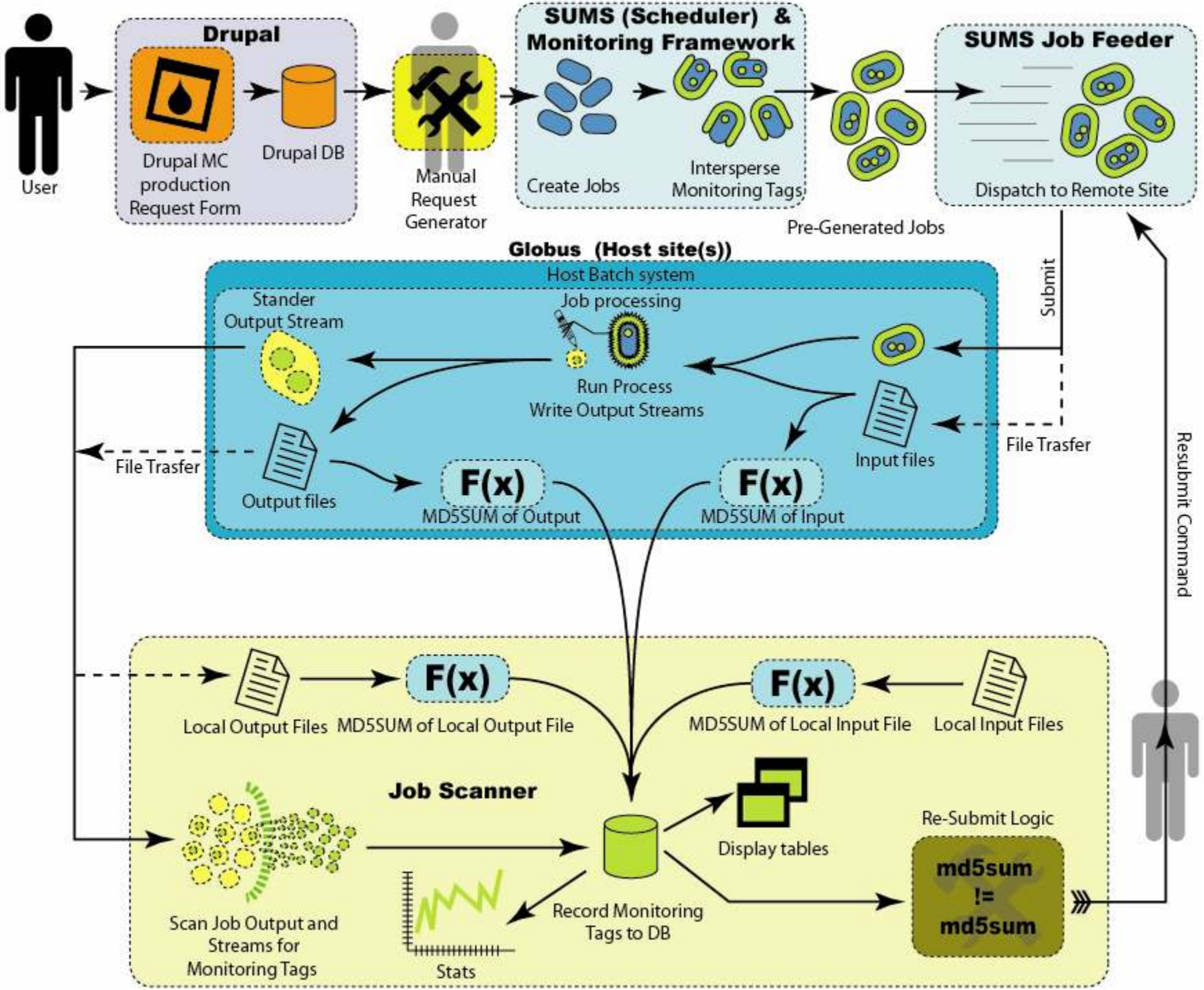
4BF58A76682503014E404A96DC94D9D3

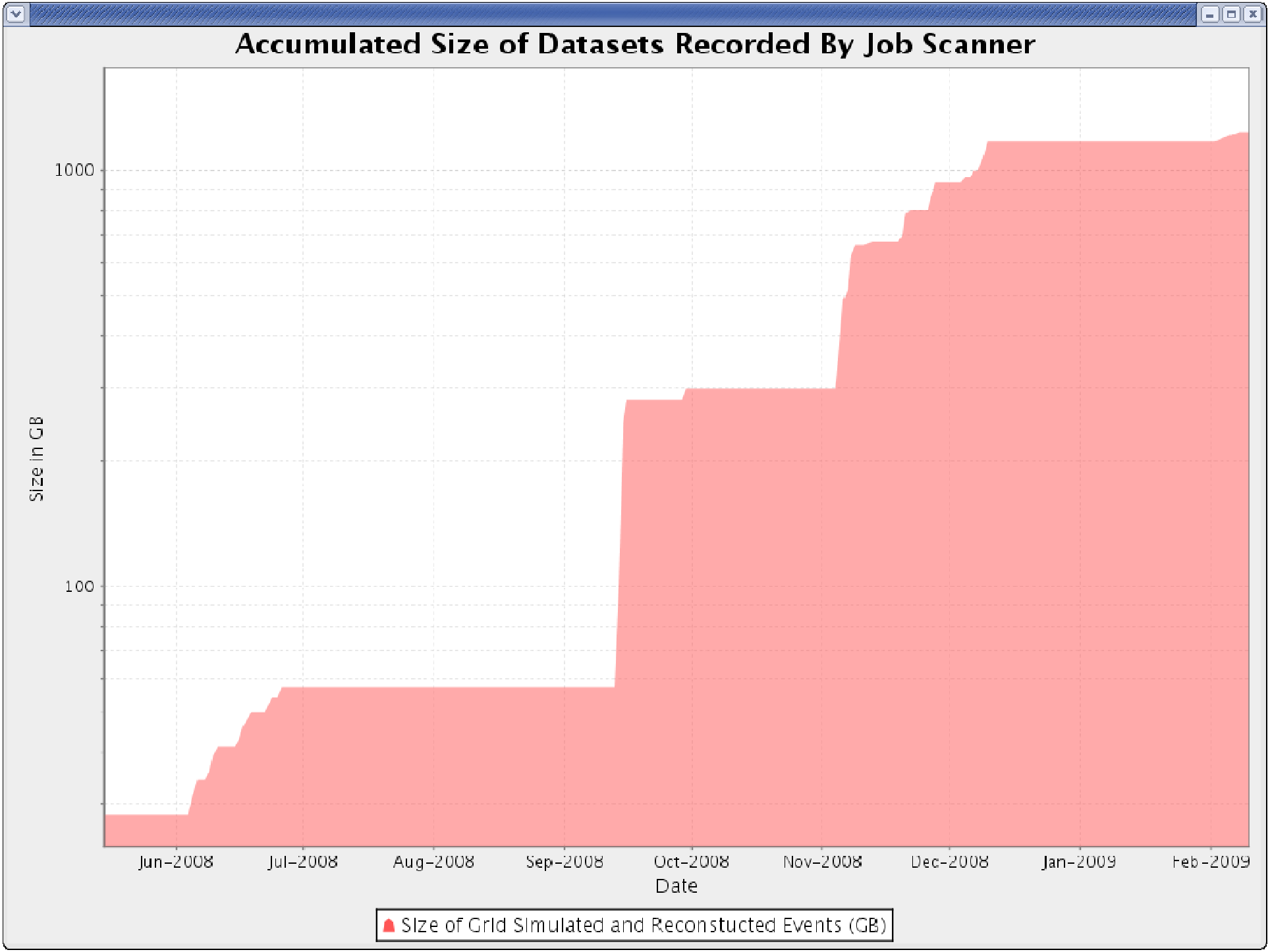
Resubmission

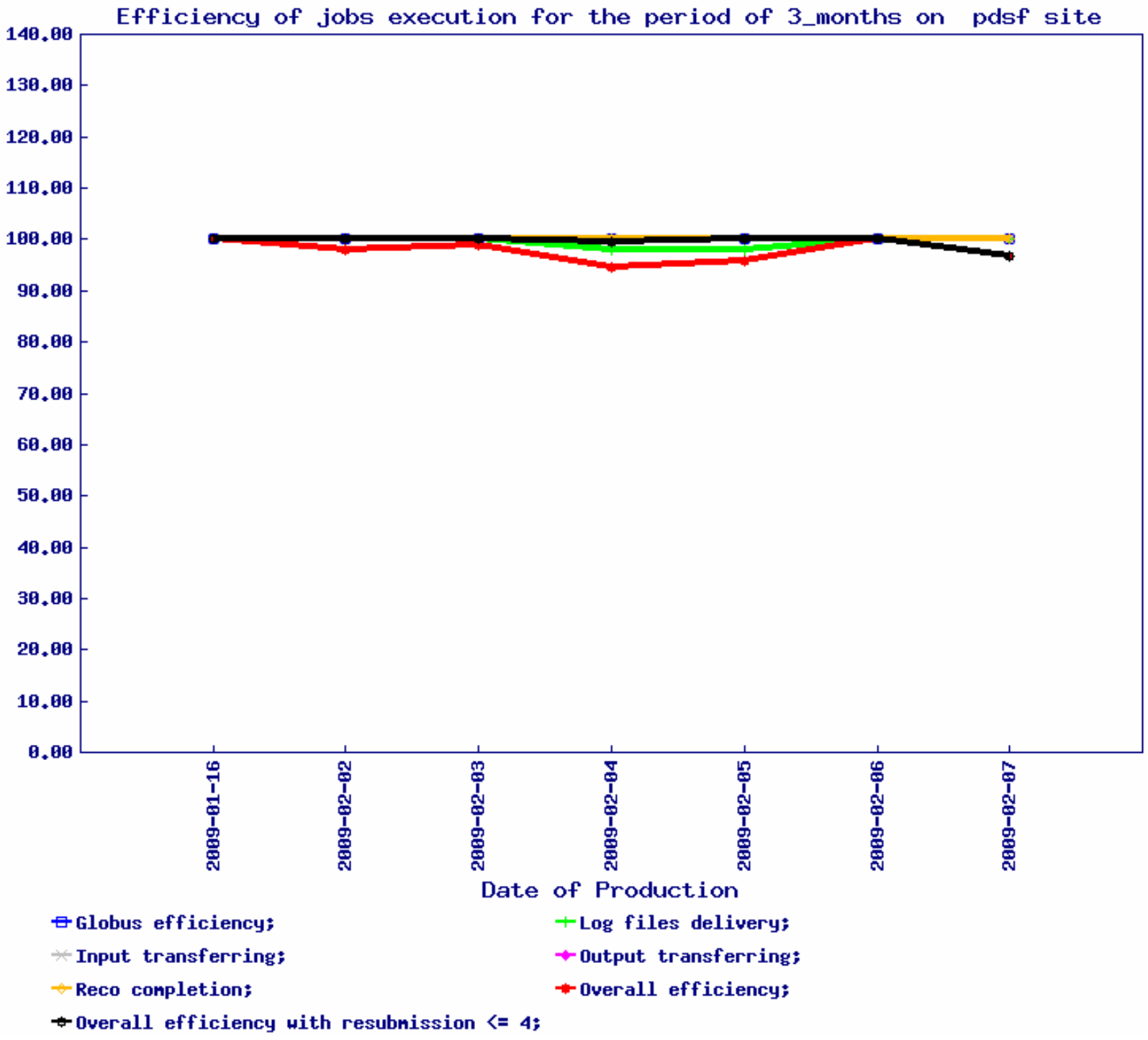
- During the initial submission where the users request description is transformed into jobs, the jobs are also written into an XML serialized file.
- Should any job(s) fail the user can feed this file back to the scheduler and resubmit exactly only the failed jobs.

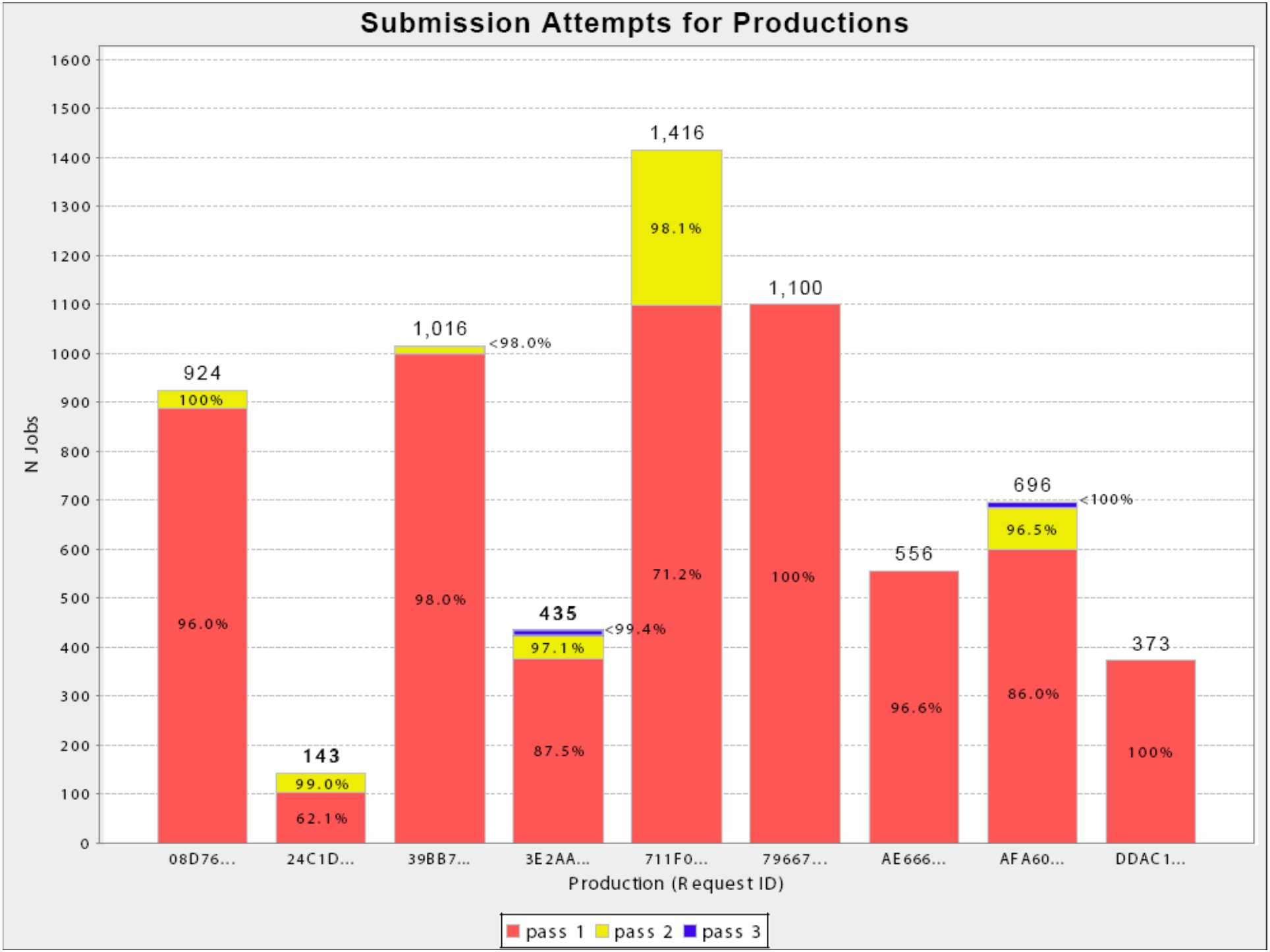
sums-submit –policy zenEC2 myRequest.xml	sums-submit -resubmit 1,2,357 D9CC6.session.xml sums-submit -resubmit 56-85 D9CC6.session.xml

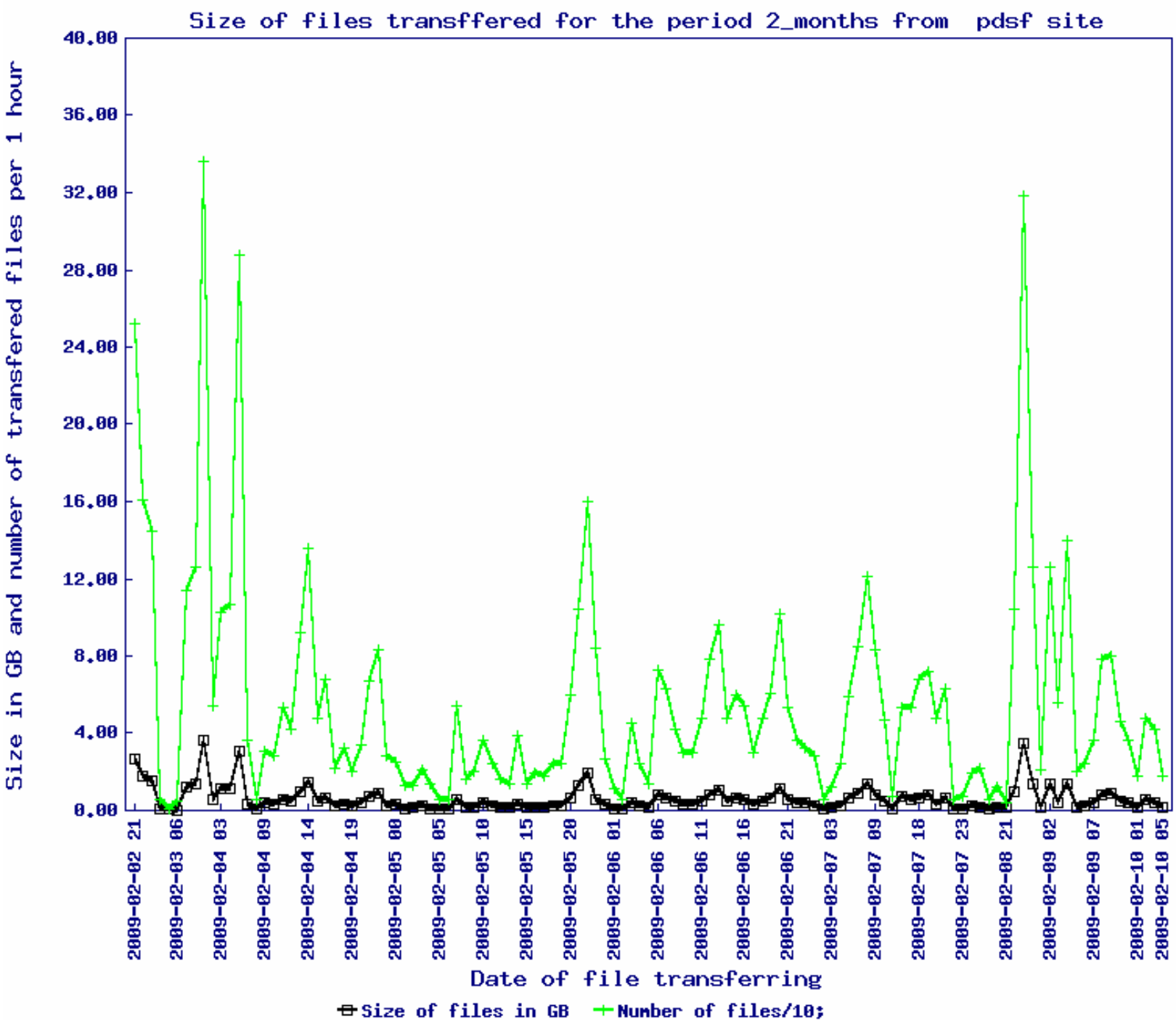



User
Drupal
Drupal MC production Request Form
Drupal DB
Manual Request Generator
SUMS (Scheduler) & Monitoring Framework
Create Jobs
Intersperse Monitoring Tags
Pre-Generated Jobs
SUMS Job Feeder
Dispatch to Remote Site
Submit
Globus (Host site(s))
Host Batch system
Stander Output Stream
Job processing
Run Process
Write Output Streams
Input files
File Trasfer
File Trasfer
Output files
F(x)
MD5SUM of Output
F(x)
MD5SUM of Input
Resubmit Command
F(x)
Local Output Files
MD5SUM of Local Output File
F(x)
MD5SUM of Local Input File
Local Input Files
Job Scanner
Display tables
Re-Submit Logic
md5sum != md5sum
Scan Job Output and Streams for Monitoring Tags
Stats
Record Monitoring Tags to DB


Accumulated Size of Datasets Recorded By Job Scanner
1000
100
Size in GB
Jun-2008
Jul-2008
Aug-2008
Sep-2008
Oct-2008
Nov-2008
Dec-2008
Jan-2009
Feb-2009
Date
Size of Grid Simulated and Reconstucted Events (GB)

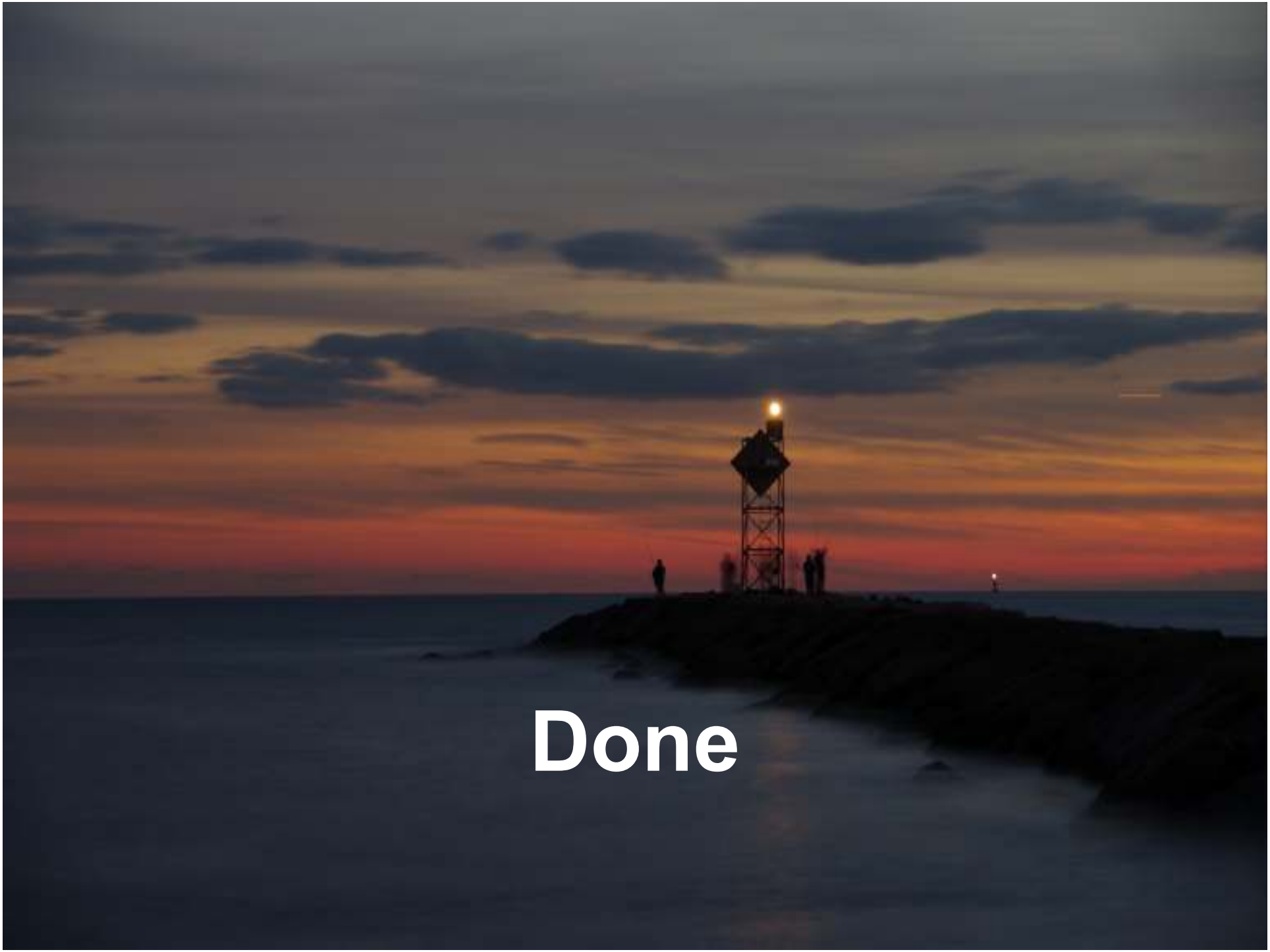

Efficiency of jobs execution for the period of 3_months on pdsf site
140.00
130.00
120.00
110.00
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
2009-01-16
2009-02-02
2009-02-03
2009-02-04
2009-02-05
2009-02-06
2009-02-07
Date of Production
Globus efficiency;
Log files delivery;
Input transferring;
Output transferring;
Reco completion;
Overall efficiency;
Overall efficiency with resubmission <= 4;


Submission Attempts for Productions
N Jobs
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
924
100%
96.0%
143
99.0%
62.1%
1,016
<98.0%
98.0%
435
<99.4%
97.1%
87.5%
1,416
98.1%
71.2%
1,100
100%
556
96.6%
696
<100%
96.5%
86.0%
373
100%
08D76...
24C1D...
39BB7...
3E2AA...
711F0...
79667...
AE666...
AFA60...
DDAC1...
Production (Request ID)
pass 1
pass 2
pass 3


Size of files transffered for the period 2_months from pdsf site
Size in GB and number of transfered files per 1 hour
40.00
36.00
32.00
28.00
24.00
20.00
16.00
12.00
8.00
4.00
0.00
2009-02-02 21
2009-02-03 06
2009-02-04 03
2009-02-04 09
2009-02-04 14
2009-02-04 19
2009-02-05 00
2009-02-05 05
2009-02-05 10
2009-02-05 15
2009-02-05 20
2009-02-06 01
2009-02-06 06
2009-02-06 11
2009-02-06 16
2009-02-06 21
2009-02-07 03
2009-02-07 09
2009-02-07 18
2009-02-07 23
2009-02-08 21
2009-02-09 02
2009-02-09 07
2009-02-10 01
2009-02-10 05
Date of file transferring
Size of files in GB
Number of files/10;

Conclusions

- Much of the production cycle is automat-able
  - someone still needs to oversee the process as a whole.
- A production level suite of Independent components allows for essentially piping jobs between each component to achieve different manipulations.
  - **Simulation Production Request API** – provides request validation and request documentation for reproducibility.
  - **SUMS scheduler** generates and submits jobs as well as producing session files for later (re)submission and feeding of jobs.
  - The **Feeder** takes overhead off the infrastructure without decreasing throughput while keeping production manageable.
  - The **Scanner,** catalogs and identifies failed jobs for resubmissions (finds the needle in a haystack).
    - Far superior to the quality that can be achieved, inspecting the jobs by hand.
    - Cataloging is provided and statistics can be extracted
  - The resubmission system, which is a subcomponent of SUMS, makes resubmission of many jobs simple with out needing the original files and scripts.
- This simplistic approach if applied diligently can get us a long way.


Done

Drupal

Simulation Requests

1212805340 Leve's test request 0

User login

Username: *

Password: *

Log in

- Create new account
- Request new password

Navigation

- Content
- on this date

Test request

Fri, 12/19/2008 - 22:03 — lbhajdu

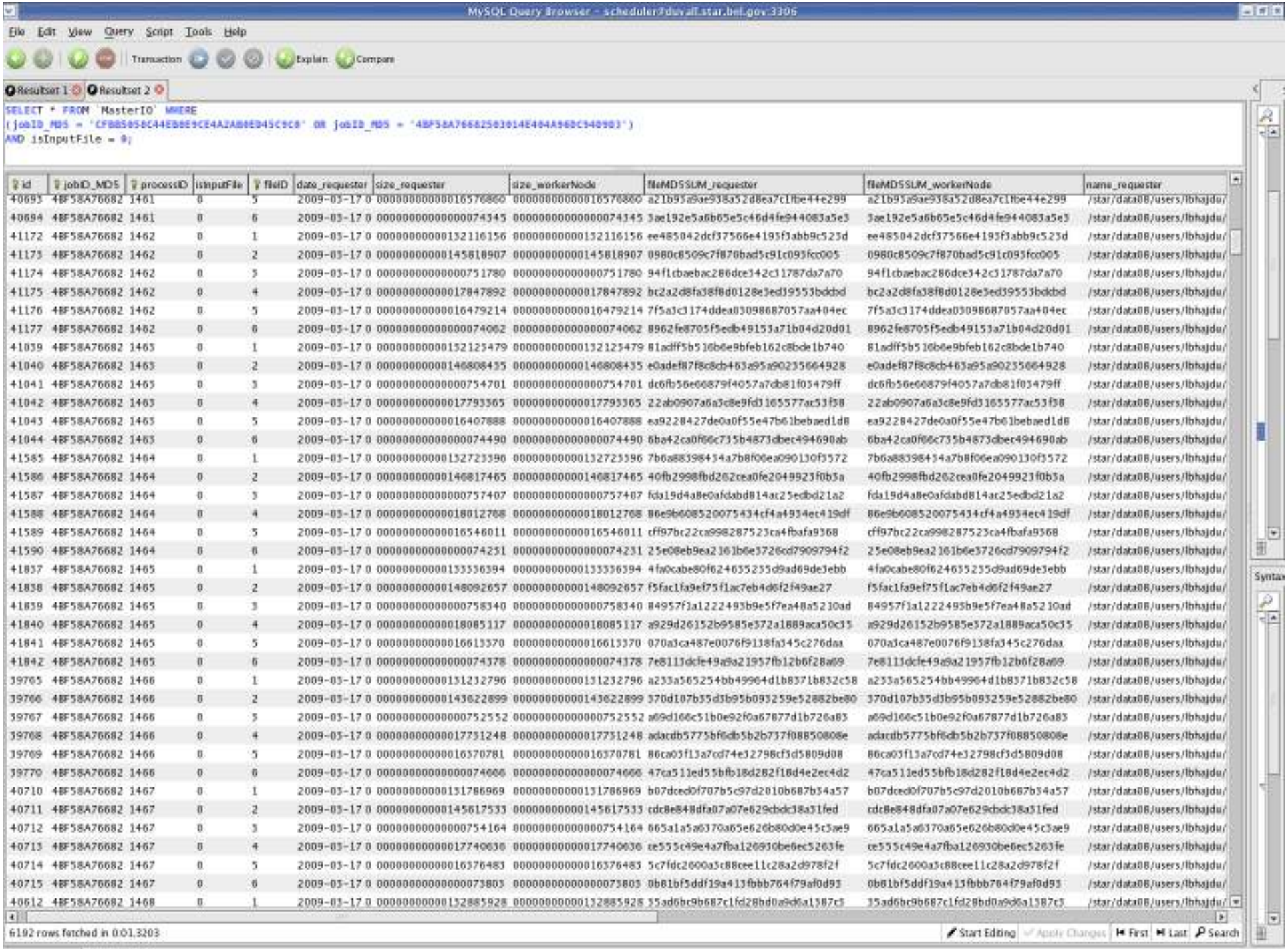
Title:	Test request
Description:	this is a test request
PWG:	SPIN
collision:	dAu
energy:	130Gev
event generator:	MEVSIM
event generator InputFile:	
geometry for simu:	y2003a
geometry for reco:	y2003a
magnetic field:	FIELDOFF
event generator parameters:	
Eta interval:	
Z vertex:	0
Z distribution:	
Z sigma:	+-20cm
pt Low:	
pt High:	
centrality:	NA
embedded particles:	0
embedded particles per event:	0
embedded particles Pt High:	

On This Date subject

Leve's test request

Leve's test page

Test request


MySQL Query Browser - scheduler@duvall.star.bnl.gov:3306
File Edit View Query Script Tools Help
Transaction
Explain
Compare
Resultset 1
Resultset 2
SELECT * FROM `MasterIO` WHERE
(jobID_MD5 = 'CFBB5058C44EB8E9CE4A2AB8ED45C9C8' OR jobID_MD5 = '4BF58A766825030148404A96DC940903')
AND isInputFile = 0;
id jobID_MD5 processID isInputFile fileID data_requester size_requester size_workerNode fileMD5SUM_requester fileMD5SUM_workerNode name_requester
40693 4BF58A76682 1461 0 5 2009-03-17 0 0000000000016576860 0000000000016570860 a21b93a9ae938a52d8ea7c1fbe44e299 a21b93a9ae938a52d8ea7c1fbe44e299 /star/data08/users/lbhajdu/
40694 4BF58A76682 1461 0 6 2009-03-17 0 0000000000000074345 0000000000000074345 3ae192e5a6b65e5c46d4fe944083a5e3 3ae192e5a6b65e5c46d4fe944083a5e3 /star/data08/users/lbhajdu/
41172 4BF58A76682 1462 0 1 2009-03-17 0 0000000000132116156 0000000000132116156 ee485042dcf37566e4193f3abb9c523d ee485042dcf37566e4193f3abb9c523d /star/data08/users/lbhajdu/
41173 4BF58A76682 1462 0 2 2009-03-17 0 0000000000145818907 0000000000145818907 0980c8509c7f870bad5c91c093fcc005 0980c8509c7f870bad5c91c093fcc005 /star/data08/users/lbhajdu/
41174 4BF58A76682 1462 0 3 2009-03-17 0 0000000000000751780 0000000000000751780 94f1cbaebac286dce342c31787da7a70 94f1cbaebac286dce342c31787da7a70 /star/data08/users/lbhajdu/
41175 4BF58A76682 1462 0 4 2009-03-17 0 0000000000017847892 0000000000017847892 bc2a2d8fa38f8d0128e3ed39553bddbd bc2a2d8fa38f8d0128e3ed39553bddbd /star/data08/users/lbhajdu/
41176 4BF58A76682 1462 0 5 2009-03-17 0 0000000000016479214 0000000000016479214 7f5a3c3174ddea03098687057aa404ec 7f5a3c3174ddea03098687057aa404ec /star/data08/users/lbhajdu/
41177 4BF58A76682 1462 0 6 2009-03-17 0 0000000000000074062 0000000000000074062 8962fe8705f5edb49153a71b04d20d01 8962fe8705f5edb49153a71b04d20d01 /star/data08/users/lbhajdu/
41039 4BF58A76682 1463 0 1 2009-03-17 0 0000000000132123479 0000000000132123479 81adff5b516b6e9bfeb162c8bde1b740 81adff5b516b6e9bfeb162c8bde1b740 /star/data08/users/lbhajdu/
41040 4BF58A76682 1463 0 2 2009-03-17 0 0000000000146808435 0000000000146808435 e0adef87f8c8db463a95a90235664928 e0adef87f8c8db463a95a90235664928 /star/data08/users/lbhajdu/
41041 4BF58A76682 1463 0 3 2009-03-17 0 0000000000000754701 0000000000000754701 dc6fb56e66879f4057a7db81f03479ff dc6fb56e66879f4057a7db81f03479ff /star/data08/users/lbhajdu/
41042 4BF58A76682 1463 0 4 2009-03-17 0 0000000000017793565 0000000000017793565 22ab0907a6a3c8e9fd3165577ac53f38 22ab0907a6a3c8e9fd3165577ac53f38 /star/data08/users/lbhajdu/
41043 4BF58A76682 1463 0 5 2009-03-17 0 0000000000016407888 0000000000016407888 ea9228427de0a0f55e47b61bebaed1d8 ea9228427de0a0f55e47b61bebaed1d8 /star/data08/users/lbhajdu/
41044 4BF58A76682 1463 0 6 2009-03-17 0 0000000000000074490 0000000000000074490 6ba42ca0f66c735b4873dbec494690ab 6ba42ca0f66c735b4873dbec494690ab /star/data08/users/lbhajdu/
41585 4BF58A76682 1464 0 1 2009-03-17 0 0000000000132723596 0000000000132723596 7b6a88398434a7b8f06ea090130f3572 7b6a88398434a7b8f06ea090130f3572 /star/data08/users/lbhajdu/
41586 4BF58A76682 1464 0 2 2009-03-17 0 0000000000146817465 0000000000146817465 40fb2998fbd262cea0fe2049923f0b3a 40fb2998fbd262cea0fe2049923f0b3a /star/data08/users/lbhajdu/
41587 4BF58A76682 1464 0 3 2009-03-17 0 0000000000000757407 0000000000000757407 fda19d4a8e0afdabd814ac25edbd21a2 fda19d4a8e0afdabd814ac25edbd21a2 /star/data08/users/lbhajdu/
41588 4BF58A76682 1464 0 4 2009-03-17 0 0000000000018012768 0000000000018012768 86e9b608520075434cf4a4934ec419df 86e9b608520075434cf4a4934ec419df /star/data08/users/lbhajdu/
41589 4BF58A76682 1464 0 5 2009-03-17 0 0000000000016546011 0000000000016546011 cff97bc22ca998287523ca4fbafa9368 cff97bc22ca998287523ca4fbafa9368 /star/data08/users/lbhajdu/
41590 4BF58A76682 1464 0 6 2009-03-17 0 0000000000000074231 0000000000000074231 25e08eb9ea2161b6e3726cd7909794f2 25e08eb9ea2161b6e3726cd7909794f2 /star/data08/users/lbhajdu/
41837 4BF58A76682 1465 0 1 2009-03-17 0 0000000000133336394 0000000000133336394 4fa0cabe80f624635255d9ad69de3ebb 4fa0cabe80f624635255d9ad69de3ebb /star/data08/users/lbhajdu/
41838 4BF58A76682 1465 0 2 2009-03-17 0 0000000000148092657 0000000000148092657 f5fac1fa9ef75f1ac7eb4d6f2f49ae27 f5fac1fa9ef75f1ac7eb4d6f2f49ae27 /star/data08/users/lbhajdu/
41839 4BF58A76682 1465 0 3 2009-03-17 0 0000000000000758340 0000000000000758340 84957f1a1222493b9e5f7ea48a5210ad 84957f1a1222493b9e5f7ea48a5210ad /star/data08/users/lbhajdu/
41840 4BF58A76682 1465 0 4 2009-03-17 0 0000000000018085117 0000000000018085117 a929d26152b9585e372a1889aca50c35 a929d26152b9585e372a1889aca50c35 /star/data08/users/lbhajdu/
41841 4BF58A76682 1465 0 5 2009-03-17 0 0000000000016613370 0000000000016613370 070a3ca487e0076f9138fa345c276daa 070a3ca487e0076f9138fa345c276daa /star/data08/users/lbhajdu/
41842 4BF58A76682 1465 0 6 2009-03-17 0 0000000000000074378 0000000000000074378 7e8113dcfe49a9a21957fb12b6f28a69 7e8113dcfe49a9a21957fb12b6f28a69 /star/data08/users/lbhajdu/
39765 4BF58A76682 1466 0 1 2009-03-17 0 0000000000131232796 0000000000131232796 a233a565254bb49964d1b8371b832c58 a233a565254bb49964d1b8371b832c58 /star/data08/users/lbhajdu/
39766 4BF58A76682 1466 0 2 2009-03-17 0 0000000000143622899 0000000000143622899 370d107b35d3b95b093259e52882be80 370d107b35d3b95b093259e52882be80 /star/data08/users/lbhajdu/
39767 4BF58A76682 1466 0 3 2009-03-17 0 0000000000000752552 0000000000000752552 a69d166c51b0e92f0a67877d1b726a83 a69d166c51b0e92f0a67877d1b726a83 /star/data08/users/lbhajdu/
39768 4BF58A76682 1466 0 4 2009-03-17 0 0000000000017731248 0000000000017731248 adacdb5775bf6db5b2b737f08850808e adacdb5775bf6db5b2b737f08850808e /star/data08/users/lbhajdu/
39769 4BF58A76682 1466 0 5 2009-03-17 0 0000000000016370781 0000000000016370781 86ca03f13a7cd74e32798cf3d5809d08 86ca03f13a7cd74e32798cf3d5809d08 /star/data08/users/lbhajdu/
39770 4BF58A76682 1466 0 6 2009-03-17 0 0000000000000074666 0000000000000074666 47ca511ed55bfb18d282f18d4e2ec4d2 47ca511ed55bfb18d282f18d4e2ec4d2 /star/data08/users/lbhajdu/
40710 4BF58A76682 1467 0 1 2009-03-17 0 0000000000131786969 0000000000131786969 b07dced0f707b5c97d2010b687b34a57 b07dced0f707b5c97d2010b687b34a57 /star/data08/users/lbhajdu/
40711 4BF58A76682 1467 0 2 2009-03-17 0 0000000000145617533 0000000000145617533 cdc8e848dfa07a07e629cbdc38a31fed cdc8e848dfa07a07e629cbdc38a31fed /star/data08/users/lbhajdu/
40712 4BF58A76682 1467 0 3 2009-03-17 0 0000000000000754164 0000000000000754164 665a1a5a6370a65e626b80d0e45c3ae9 665a1a5a6370a65e626b80d0e45c3ae9 /star/data08/users/lbhajdu/
40713 4BF58A76682 1467 0 4 2009-03-17 0 0000000000017740636 0000000000017740636 ce555c49e4a7fba12693be6ec5263fe ce555c49e4a7fba12693be6ec5263fe /star/data08/users/lbhajdu/
40714 4BF58A76682 1467 0 5 2009-03-17 0 0000000000016376483 0000000000016376483 5c7fdc2600a3c88cee11c28a2d978f2f 5c7fdc2600a3c88cee11c28a2d978f2f /star/data08/users/lbhajdu/
40715 4BF58A76682 1467 0 6 2009-03-17 0 0000000000000073803 0000000000000073803 0b81bf5ddf19a413fbbb764f79af0d93 0b81bf5ddf19a413fbbb764f79af0d93 /star/data08/users/lbhajdu/
40612 4BF58A76682 1468 0 1 2009-03-17 0 0000000000132885928 0000000000132885928 35ad6bc9b687c1fd28bd0a9d6a1387c3 35ad6bc9b687c1fd28bd0a9d6a1387c3 /star/data08/users/lbhajdu/
6192 rows fetched in 0.0132s
Start Editing
Apply Changes
First
Last
Search
Query finished.
Syntax