

Virtualization of Tier 2 site and LCG Grid Data Distributed Management

Mihajlo Mudrinic



HEP 4XOVA VITC*



<http://www.vin.bg.ac.yu/hep>



HEA

EROWA

VITTA



Deutsches Elektronen-Synchrotron
in der Helmholtz-Gemeinschaft



ATLAS group at DESY
<http://www-atlas.desy.de/general.html>



The H1 Experiment at HERA
http://www-h1.desy.de/general/home/intra_home.html



ilc international linear collider

<http://www.linearcollider.org/cms/>

Outline:

Introduction: "It would be very nice if..." Case for Virtualization.

Development of web services (SOAP protocol) to automate Monte Carlo production for H1 experiment at DESY – Hamburg.

Xen hypervisor and work on Virtualization of Tier-2 site at DESY-Zeuthen.

Intro to the LCG Data Distributed Management.

- *Virtualization is a framework or methodology of dividing the resources of a computer into multiple execution environments.*
- *A layer of software that provides the illusion of a "real" machine to multiple instances of "virtual machines".*



HEA

EXOVA

VITCX



•“It would be very nice if...” →/usr/sbin/xenguest-install.py -ks filename.ks

•Virtual machines can be used to consolidate the workloads of several under-utilized servers to fewer machines, perhaps a single machine (server consolidation). Related benefits are savings on hardware, management, and administration of the server infrastructure.

•Virtual machines can encapsulate the entire state of a running system: you can save the state, examine it, modify it, reload it, and so on.

• Virtual machines can be used to run multiple operating systems simultaneously: different versions, or even entirely different systems, which can be on hot standby

•The need to run legacy applications is served well by virtual machines. A legacy application might simply not run on newer hardware and/or operating systems.



HEAD

EXOSYSTEM

VITC*



MC simulation:



Physics Working Groups

- Rare & Exotic Processes
- Inclusive Measurements
- Hadronic final State & QCD
- Diffraction Physics
- Heavy Flavour

H1 MC Coordinator

```
ID 123456
PARENT_ID 3793
INPUT_FILE_NAME
/acs/mc/djangoh14/DJANGO14.Q...
OUTPUT_FILE_NAME
WORKING_GROUP HaQ
EVENTS 5168483
RUN_PERIOD 2006_e-
BEAM_TILT YES
...
```

*Message from
H1 MC Coordinator*

MySQL Database

Creation of steering and conf. files.
Creation of:

- Input files.
- Job script (JDL) files.
- Upload of input files to the Grid SE.

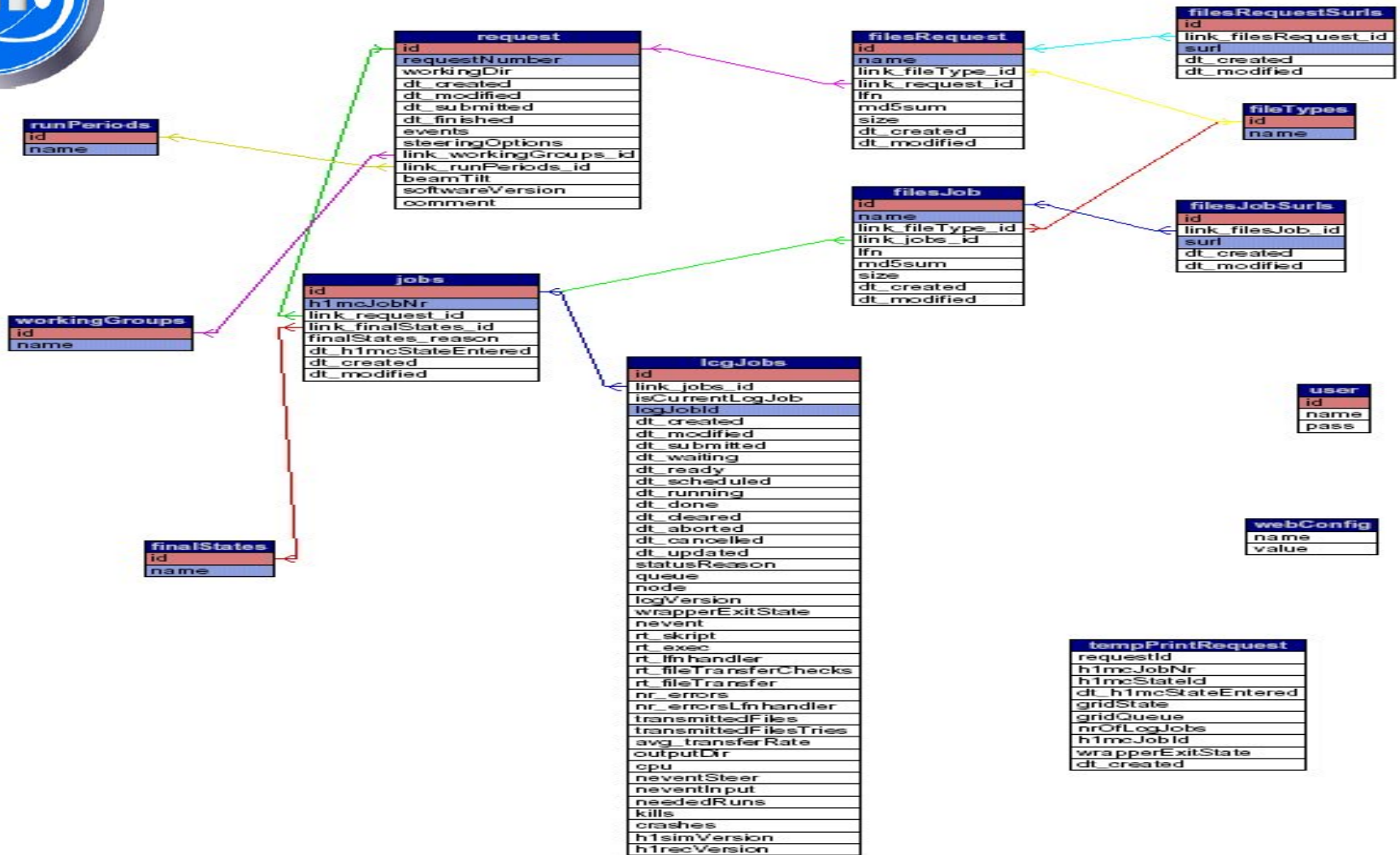


HEA ← XOV A VITCX





H1 MC Grid Production Database



The database is using “linked” tables to store all information about the user job request and subsequent jobs running on LCG grid.



H E A G X O V A V I T T C *



People



Cracow group:

Jacek Tunau, Staszek Mikocki (HI MC Coordinators),

Antek Cyz (production)

local Farm

Production group:

Alexander Fomenko, Bogdan Lobodzinski, Eberhardt Wunsch,

part time: Nina Loktionowa, Titi Preda,

development:

Elena Bystritskaja, Bogdan Lobodzinski, Mihajlo Mudrinic

GRID

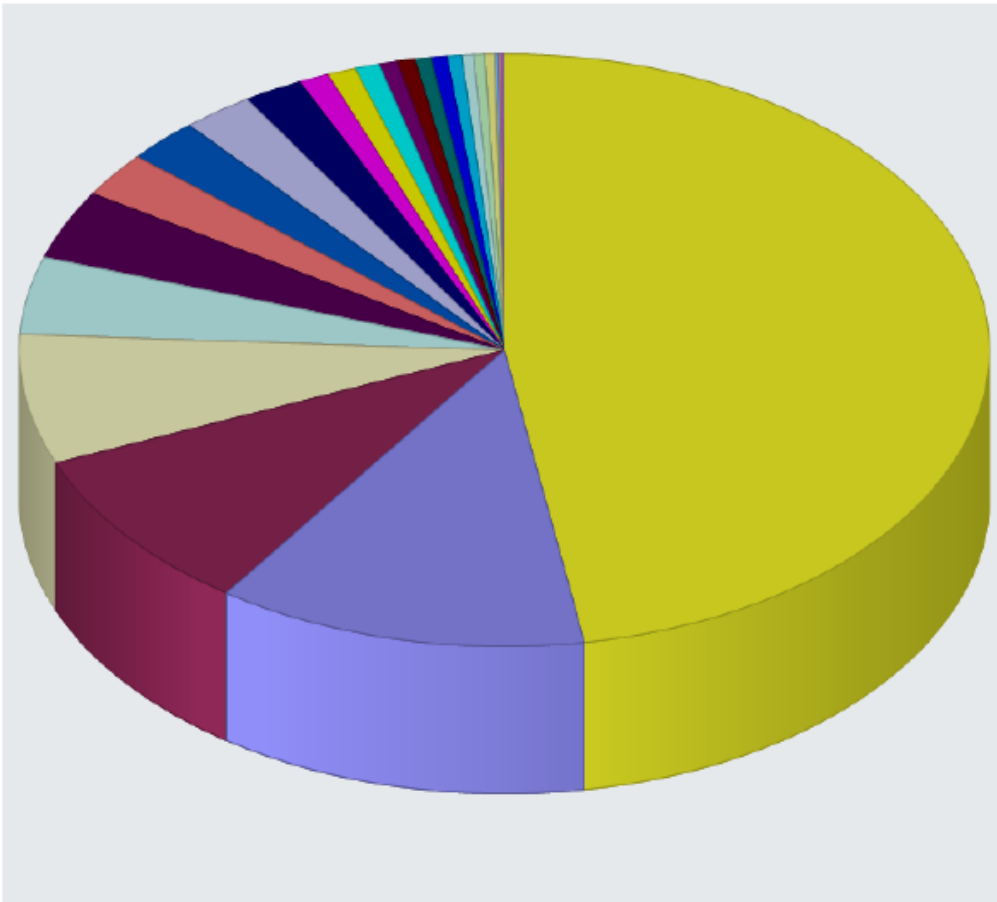


HEA GROVA VITIC*





Events done in 2007 per LCG site



224.902.954	desy.de (Hamburg):
58.671.200	ifh.de (DESY-Zeuthen):
43.001.140	hep.manchester.ac.uk (Manchester)
33.584.458	gridpp.rl.ac.uk (RAL)
20.062.109	lanacs.ac.uk (Lancaster)
17.942.735	physik.uni-dortmund.de (Dortmund)
11.705.141	physics.ox.ac.uk (Oxford)
11.395.233	projects.cscs.ch (CSCS)
11.215.309	brunel.ac.uk (Brunel)
9.365.966	pp.rl.ac.uk (RAL tier2)
4.731.686	pp.rhul.ac.uk (RAL)
4.533.738	iihe.ac.be (Brussels)
3.904.967	mrs.grid.cnrs.fr (Marseille)
2.900.000	cyf-kr.edu.pl (Krakow)
2.811.374	ph.liv.ac.uk (Liverpool)
2.598.960	jlnr.ru (Dubna)
2.514.984	polgrid1.in2p3.fr (Palaiseau)
2.282.500	cclcgceli*.in2p3.fr (Lyon)
1.859.980	hep.ntua.gr (Athens)
1.757.016	nipne.ro (Bucharest)
1.495.010	ph.bham.ac.uk (Birmingham)
550.000	lal.in2p3.fr (Orsay)
540.168	hep.ph.ic.ac.uk (London)
427.997	epcc.ed.ac.uk (Edinburgh)
80.000	farm.particle.cz (Praque)
10.000	saske.sk (Kosice)



HEA

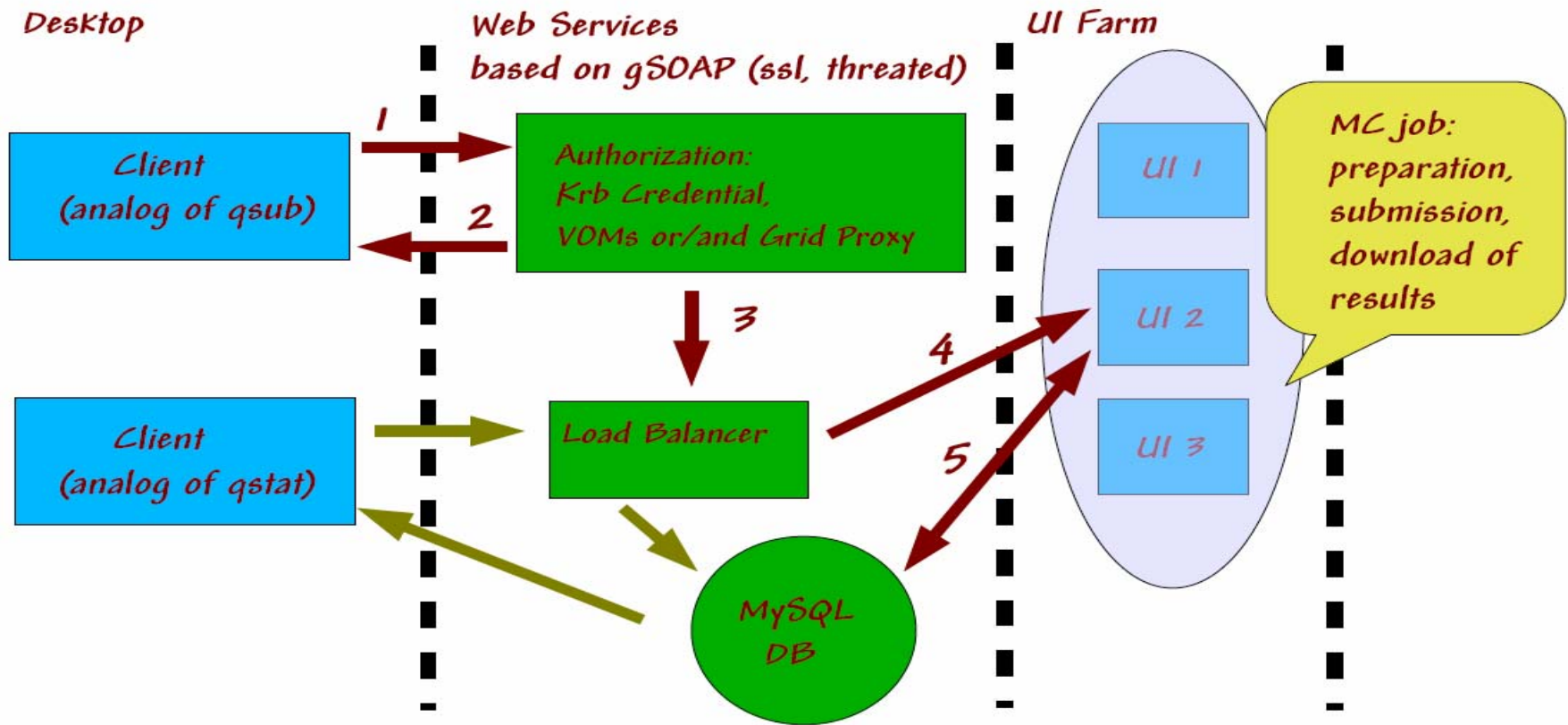
EXONA

VITCX





- The application is based on client-server (two-tiers) architecture.
- Client and Server communicate using Simple Object Access Protocol (SOAP).
- Server will offer web service to a client (alias qsub and qstat)
- The User is sending request for MC job simply by editing text file.
- Server is submitting job to H1 Monte Carlo Grid Batch System and providing the user with receipt showing job ID's (one is local to application, one from Torque and from H1 MC Grid Production).



HEP @ CERN





Down to Earth Client Side

HOST



SERVER



Key Distribution Center (KDS) + Ticket Granting Service (TGS)
Database of principals (users)

MC client.conf

\$HOME/38_mcjobReceipt_03.09.07:20:43

```
##### ID #####
ID 123456
##### PARENT_ID #####
PARENT_ID 3793
##### INPUT FILE NAME #####
INPUT_FILE_NAME /acs/mc/djangoh14/DJANGO14.Q.....
##### OUTPUT_FILE_NAME #####
OUTPUT_FILE_NAME
##### WORKING_GROUP #####
WORKING_GROUP HaQ
##### EVENTS #####
EVENTS 5168483
##### RUN_PERIOD #####
RUN_PERIOD 2006_e-
##### STEERING #####
.....
.....
```

```
=====
RECEIPT Job No. 38::Date: day.mo.year:hour:min ->03.09.07:20:43
Your MonteCarlo Job Request is Accepted
And Soon Will be Processed
-----
Torque JobID ::11.h1farm00.desy.de
Use :: /h1/torque/bin/qstat -f <Torque JobID> to check status
*****
Soon Your Job with Request ID::123456 will be registered and submitted
to LCG
Use :: lcgjobstat -n 38 to query database and check if your request has
running jobs
After you find out that the request has running jobs, you will be able to
obtain full info by typing
:: lcgjobstat -j 38::
OR
:: lcgjobstat -r 123456::
```

Collecting and sending Kerb. and client.conf files using as an additional security level Secure Socket Layer (SSL) protocol.

- Receive job ID's receipt and store it.



HEADERS AND FOOTERS



Down to Earth Server Side



- Running on Apache Web Server as CGI Application.
- Apache Web Server is supporting multi threads and SSL protocol.
- Store user's Kerb. Credential and MC Job request.
- Make batch_script based on batch template and user MC job request.
- Using the user's Kerberos credential run MC job.
- Send back JobID, Torque JobID and Request JobID.

SERVER

Apache SSL

/<path>/server.pem

/<path>/serverkey.pem

```
/x01/usr/MC.web_service/  
Directory BATCH_TEMPLATE  
file .lastjobID  
file configfilesserver.conf  
file logfileserver.log  
directory 38_03.09.07:20:43_mudrim  
kerbfile  
mcjobrequest  
batch_file  
JobIDS
```



HEAD

← X O V A V I T C *



Down to Earth Job statistics



Usage: ./lcgjobstat [options]

- a, --Show List of All Requests from MySQL database
- j, --Show List of All LCG Jobs for the USER Job Number
- n, --Show Status of New Request/OR Compact Info for the USER Job Number
- r, --Show List of All LCG Jobs for the Request Number
- h, --help Print this information.

./lcgjobstat -a

=====**List of All LCG Job Requests Sorted by Request Number**=====

Req. Nr.	new	running	done	failed	succeeded	received	broken	finished	mute	reqDate	comment
4848	--		18	----		182	----		----	2007-09-02 11:47:34	AF, ver 35: R96700; S36400
4847	--		2	----		198	----		----	2007-09-01 13:27:32	AF, ver 35: R96700; S36400
4845	--		----		106	----		----	----	2007-08-31 16:47:57	AF, ver 35: R96700; S36400
123456	9	----		----		----		----	----	2007-08-31 15:43:58	BL, ver 36: R96500; S36000
4844	--		----		20	----		----	----	2007-08-31 14:11:06	AF, ver 36: R96500; S36000
4843	--		----		21	----		----	----	2007-08-31 13:31:05	AF, ver 36: R96500; S36000
4749	--		----		260	----		----	----	2007-08-30 15:55:21	AF, ver 36: R96500; S36000
4727	--		2	----		318	----		----	2007-08-30 14:33:08	BL, ver 36: R96500; S36000
4838	--		----		40	----		----	----	2007-08-29 11:17:21	AF, ver 35: R96700; S36400

./lcgjobstat -r 4848

JobNr.	h1mcJobId	h1mcStateId	GridState	wrapperExitState	gridQueue	nrOfLcgJobs	dt_h1mcStateEntered
1	105708	succeeded	Cleared	ERRORS_OCCURRED	hepgrid2.ph.liv.ac.uk:2119/jobmanager-lcgpbs-hone	1	2007-09-03 07:54:59
2	105709	succeeded	Cleared	ERRORS_OCCURRED	lcgce01.jinr.ru:2119/jobmanager-lcgpbs-hone	2	2007-09-02 22:49:52
3	105710	succeeded	Cleared	ERRORS_OCCURRED	lcg-ce1.ifh.de:2119/jobmanager-lcgpbs-hone_blade	1	2007-09-03 00:36:10
4	105711	succeeded	Cleared	ERRORS_OCCURRED	heplnx206.pp.rl.ac.uk:2119/jobmanager-lcgpbs-hone	1	2007-09-03 03:45:11
5	105712	succeeded	Cleared	ERRORS_OCCURRED	lcg-ce0.ifh.de:2119/jobmanager-lcgpbs-hone	1	2007-09-03 04:10:17
6	105713	succeeded	Cleared	ERRORS_OCCURRED	marseillece01.mrs.grid.cnrs.fr:2119/jobmanager-pbs-hone	1	2007-09-03 05:37:46



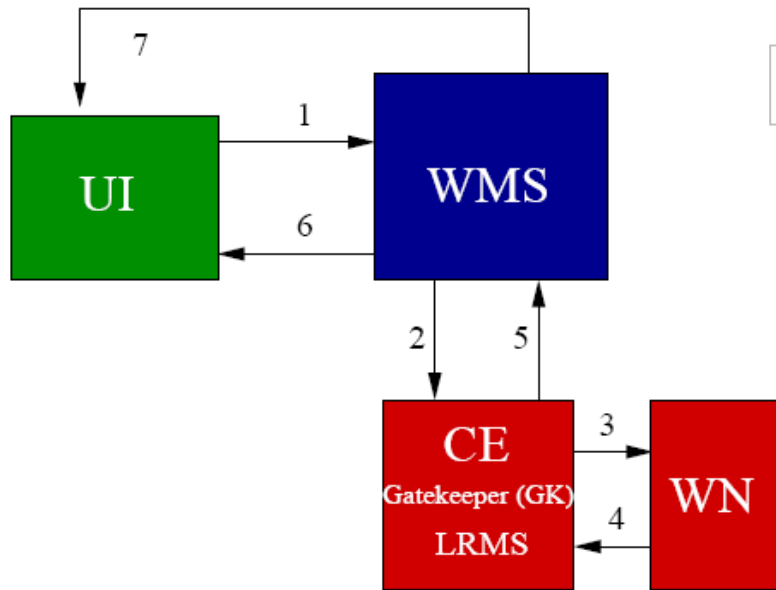
HEP G X O V A V I T C X



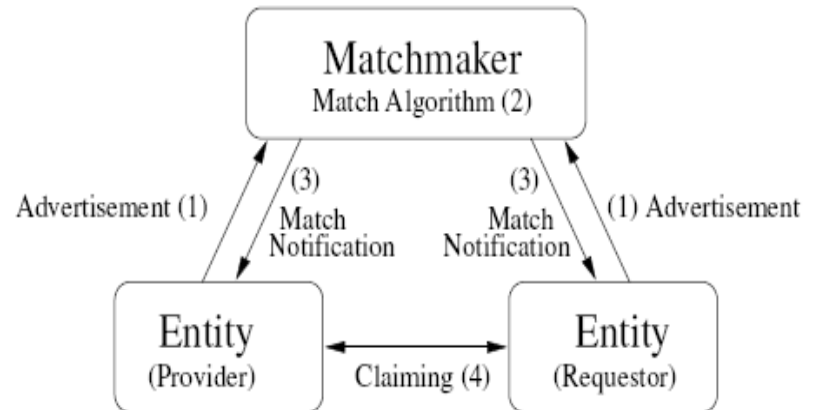
Why Simple Object Access Protocol (SOAP) and gSOAP?

- Encode parameters and return values in XML and to pass them over HTTP.
- Way to remotely make calls on objects (methods) sitting on server.
- Existing gSOAP toolkit based on C/C++ header file with specification of remote methods produce WSDL* document describing Web Service.
- Easy to make skeleton for a client application from WSDL file.

*web service description language



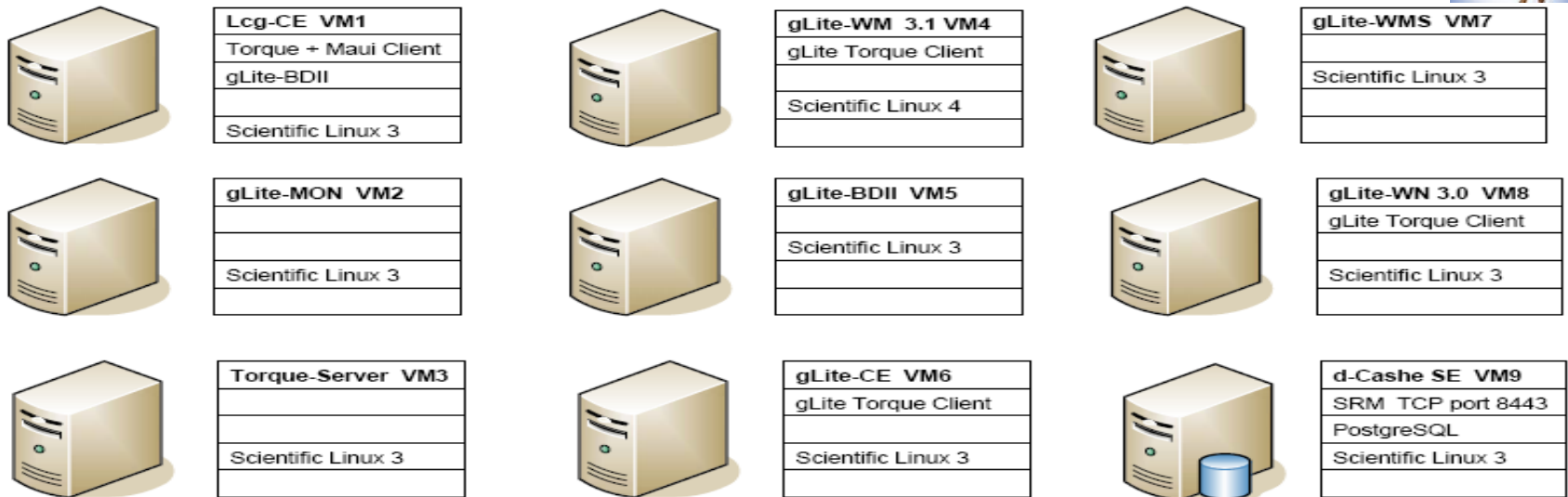
•WMPProxy Service:: UI → gLite WMS



HEA ← X O V A V I T C *



Tier 2 Virtual Site Figure 2.a



Xen Virtualization Figure 2.b

8/9/2007

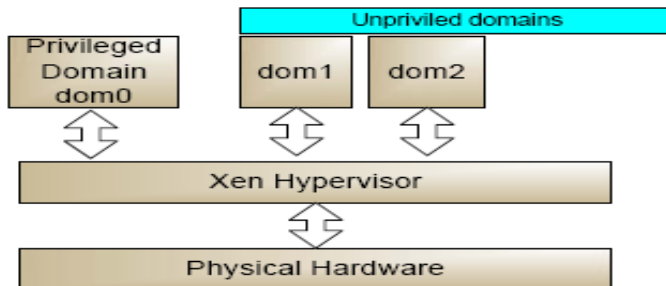


Figure 2.c

8/9/2007

A job submission and the job output retrieval on Virtual LCG Tier-2 site

```

>glite-wms-job-submit -d delID1234 -o ids -r vm6.ifh.de:2119/blah-pbs-atlas job.jdl
>glite-job-status -i ids
*****
BOOKKEEPING INFORMATION:
Status info for the Job : https://brutus-vm7.ifh.de:9000/_9gHoHA57H0vSYefHjJ63Q
Current Status: Done (Success)
Exit code: 0
Status Reason: Job terminated successfully
Destination: brutus-vm6.ifh.de:2119/blah-pbs-atlas
Submitted: Wed Jun 27 10:52:04 2007 CEST
*****
>glite-wms-job-output --dir /afs/ifh.de/user/m/mudrim/GRID/JOBS -i ids
  
```

ATLAS group at DESY-Zeuthen :: Kai Leffham, Andreas Haupt and Mihajlo Mudrinic



HEA G X O V A V I T C *



The Atlas Data Distributed Management (DDM)

DDM Architecture:

– Datasets Catalogs

- Single sets of catalog hosting at CERN (Repository, Content, Location, Subscription)

– Site Services

- These are a collection of agents, each performing a distinct task.
- Movement of Data to sites :: Scheduled transfers (FTS)
- Registering files in the local storage catalogs once they are successfully transferred into the site or registering new dataset locations in the central catalogs once subscriptions are complete.

Datasets – what are they:

– Contain files

- Logical File Name, GUID, metadata

– Have Versions

- Files can be added/removed between versions

– Has Unique Identifiers

- Per dataset, per Version :: Dataset/Version Unique Identifier (DUID/VUID)

– Hierarchies

- Datasets which contain other Datasets



HEA 4 X O V A V I T C *



CENTRAL CATALOGS

LOCAL CATALOGS

Dataset Content Catalog

Dataset Location Catalog

Dataset Name:
<mydataset>

Content:
<guid1> <lfn1>
<guid2> <lfn2>
<...> <...>

Dataset Name:
<mydataset>

Dataset Location:
CNAF,LYON

Entries:
<guid1> </.../lfn1>
<guid2> </.../lfn2>
<...> <...>

CNAF LFC

•In LCG, the local replica catalog is the LCG File Catalog (LFC).
•Currently all 'Local' catalogs are deployed at each ATLAS T1



HEAD FROM A WITH X



DDM Monitoring and Submission

[Panda monitor](http://gridui02.usatlas.bnl.gov:25880/server/pandamon/query)

<http://gridui02.usatlas.bnl.gov:25880/server/pandamon/query>

Atlas DDM

<http://dashb-atlas-data.cern.ch/dashboard/request.py/site>

2007-06-12 16:32:30 [T0.D.run011442.AOD.AOD09](#)

COMPLETE

DUID: d22aea15-f3c5-47d1-8cba-
ce96eecd0646

UUID: d92db46a-15bf-48f4-83a5-
e3ab63dbb7e4

CREATION: 2007-06-12 15:51:22

VERSION: 1

Click on the *Last Update* cell of each row for details of file on site

<i>Last Update</i>	<i>Logical File Name</i>	<i>State</i>
2007-06-12 16:32:29	T0.D.run011442.AOD.AOD09._00001.1	FILE_DONE
GUID: 1452afe8-e358-4b09-a70c-33bd96eff135 CREATION: 2007-06-12 15:52:34		
FILE SIZE: 3600000000 bytes		
SURL: srm://globe-door.ifh.de/pnfs/ifh.de/acs/grid/atlas/T0/sc4tier0/06/12/T0.D.run011442.AOD.AOD09._00001.1		
CHECKSUM: md5:f5a618e1c98b8d36a9bd4cf063bdf447		



HEAD

← X O ↘ A

✓ I H C X



Conclusions:

– H1 Monte Carlo Production

- Simulation of a large number of events in 2007.
- Improvements are on the way – Development of Web Services.

– Virtualization

- Virtualization, in a sense of multiple virtual guest operating systems running on a single host supporting platform, will be future development strategy of large computing centres.

- Easy to test different middleware suites.

– The ATLAS DDM

- The DDM is a system in constant development.
- The DDM operational group at each Tier-1 site is responsible to point the problems appearing in DDM operation and propose possible solutions.



HEA

EROPA

VITCX

