

- the new generation of the LHCb grid software



A.Tsaregorodtsev, CPPM-IN2P3-CNRS, Marseille

For the LHCb DIRAC team

26 March, CHEP 2009, Prague











DIRAC development team

A. Tsaregorodtsev, N. Brook, A. Casajus Ramo, Ph. Charpentier, J.Closier, G.Cowan, R.Graciani Diaz, E.Lanciotti, Z.Mathe, R.Nandakumar, S.Paterson, V.Romanovsky, R.Santinelli, A.C.Smith, M. Seco Miguelez, A.Zhelezov

Science & Technology Facilities Council











DIRAC overview

- Highlights of the new DIRAC features
- Performance
- Conclusion







DIRAC overview

- DIRAC is a distributed data production and analysis system used by the LHCb experiment
 - Includes workload and data management components
 - Was developed originally for the MC data production tasks
 - Extended to data processing and user analysis
 - + The goal was:
 - integrate all the heterogeneous computing resources available to LHCb
 - Minimize human intervention at LHCb sites





DIRAC project review

- In 2005-2006 the design and implementation of various DIRAC systems was reviewed:
 - Independent experts
 - Non DIRAC
 - Non LHCb
 - Multiple recommendations were collected in the final 35 pages review report

	AC review report
	LHCb Computing Note
Revision: Reference: Created:	1.5 LHCb -2006-04 COMP 26 January 2006
Last modified: Prepared By: P.Mato-Vila	16 March 2009 J-P.Baud, Ph.Charpentier (Editor), J.Closier, R.Graciani, A.Maie





Review outcomes

- Base technologies are OK
 - Python, MySQL
 - Light service/client protocols
- The general architecture was approved
 - Services based architecture
 - Modular design, pluggable VO specific components
 - Pilot Job based WMS





Review outcomes

- Main recommendations
 - DIRAC security model to be complete and following grid standards
 - Possibility to apply LHCb VO policies
 - Build in more redundancy to cope with the grid resources instabilities
 - Comprehensive monitoring of production and user activities
 - Improve project management, separate generic and VO specific parts
- The outcome of the review process was the new generation DIRAC3 system
 - Essentially rewritten from scratch to meet the review recommendations
 - Took long as the same team was supporting ongoing operations





- DIRAC is covering all the LHCb needs in the distributed data processing
 - Data export from the experiment pit to CERN off-line storage
 - Automatic data distribution to Tier-1 centers
 - Automatic creation and submission of the data reconstruction jobs
 - Automatic distribution of the analysis data
 - Full management of the MC data production
 - Full support for the user analysis jobs
- Different subsystems built in the same framework
 - Reuse of technical solutions in different subsystems
 - + A concerted team of developers sharing experience





- All the communications between the distributed components in the DIRAC framework (*DISET*) are secure
 - + X509, GSI security standards
 - Fine grained authorization rules
 - Per individual user FQAN
 - Per service interface method
 - Per job
- Full featured proxy management system
 - Proxy storage and extension
 - Support for multiuser pilot jobs
 - Limited proxies
 - Special tokens to limit the number of served proxies

see [90] – R.Graciani ,A.Casajus

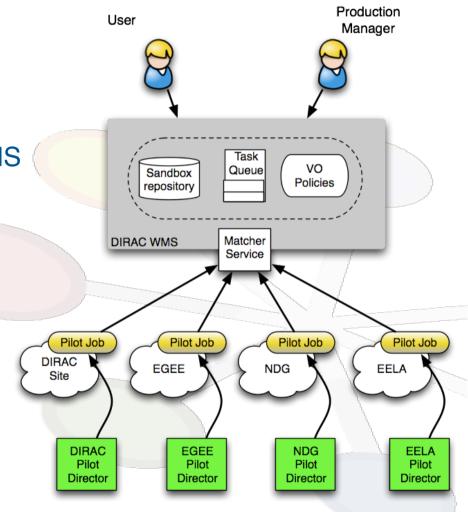




WMS: Pilot Jobs

- Jobs are submitted with credentials of their owner (VOMS proxy)
- Pilot Jobs are submitted by specific Directors to a Grid WMS
- The Pilot Job fetches the User Job and proxy
- The User Job is executed with its owner's proxy used to access SE, catalogs, etc

see [108] – R.Graciani et al

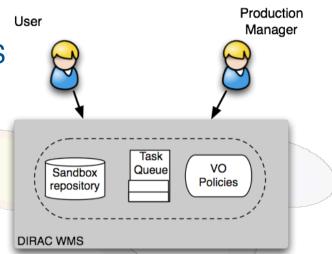






WMS: applying VO policies

- In DIRAC both User and Production jobs are now treated by the same WMS
- This allows to apply efficiently policies for the whole VO
 - Assigning Job Priorities for different groups and activities
 - Static group priorities are used currently
 - More powerful scheduler can be plugged in
 - demonstrated with MAUI scheduler



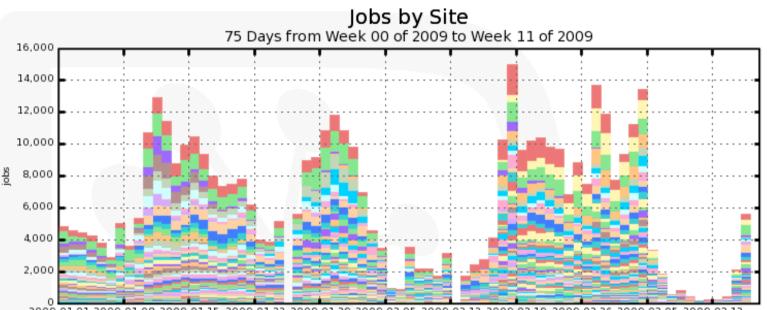
- The VO policies application in the central Task Queue dictates the use of Multiuser Pilot Agents
- DIRAC fully supports this mode of operation
 - Multiuser Pilots Jobs submitted with a special "pilot" VOMS role
 - Using glexec on the WNs to track the identity of the payload owner

see [439] – S.Paterson et al [278] – M.Cattaneo et all





Performance



2009-01-01 2009-01-08 2009-01-15 2009-01-22 2009-01-29 2009-02-05 2009-02-12 2009-02-19 2009-02-26 2009-03-05 2009-03-12

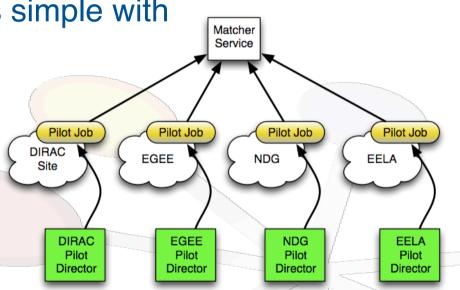
- DIRAC performance measured in the recent production and FEST'09 runs
 - Up to 15K concurrent jobs in ~120 distinct sites
 - One mid-range central server hosting DIRAC services
 - Further optimizations to increase capacity are possible
 - Hardware, database optimizations, service load balancing, etc





WMS: using heterogeneous resources

- Including resources in different grids and standalone clusters is simple with Pilot Jobs
 - Needs a specialized Pilot
 Director per resource type
 - Demonstrated with NDG and EELA grid sites
 - Users just see new sites appearing in the job monitoring



💱 🔹 Systems 🔻 Jobs 🔻 Data 🔨 🕅	Web 🔻												Sele
JobMonitoring		Select All 📃 Selec	t Non	е									
Selections -		Jobid		Status	MinorStatus	ApplicationStatus	Site	JobName	LastUpdate [UTC]	LastSignOfLife [UT	SubmissionTime [U	Owner	
DIRAC Site:		6824		Done	Execution Complet	Unknown	LCG.CERN.ch	Unknown	2009-02-21 23:51	2009-02-21 23:51	2009-02-21 23:45	atsareg	
All		6823		Done	Execution Complet	Unknown	LCC DIC as	Unknown	2009-02-21 23:52	2009-02-21 23:52	2009-02-21 23:45	atsareg	
Status:		6822		Done	Execution Complet	Unknown	EELA.CIEMAT.es	Uni nown	2009-02-21 22:38	2009-02-21 22:38	2009-02-21 22:16	atsareg	
All		6821		Done	Execution Complet	Un nown	LCG.TCD.ie	Unknown	2009-02-21 22:24	2009-02-21 22:24	2009-02-21 22:16	atsareg	
Minor status:		6820		Done	Execution Complet	Unknown	LCG.TCD.ie	Unknow	2009-02-21 22:18	2009-02-21 22:18	2009-02-21 22:16	atsareg	
All		6819		Done	Execution Complet	Unknown	EELA.CIEMAT.es	Unktown	2009-02-21 22:29	2009-02-21 22:29	2009-02-21 21:43	atsareg	
Application status:		6818		Done	Execution Complet	Unknown	LOO TODIL	Unknown	2009-02-21 22:02	2009-02-21 22:02	2009-02-21 21:43	atsareg	
All		6817		Done	Execution Complet	Unknown	LCG.TCD.ie	Unknown	2009-02-21 22:20	2009-02-21 22:20	2009-02-21 21:43	atsareg	



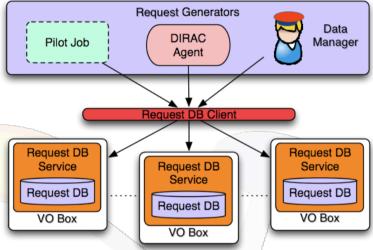


DIRAC

Request Management system

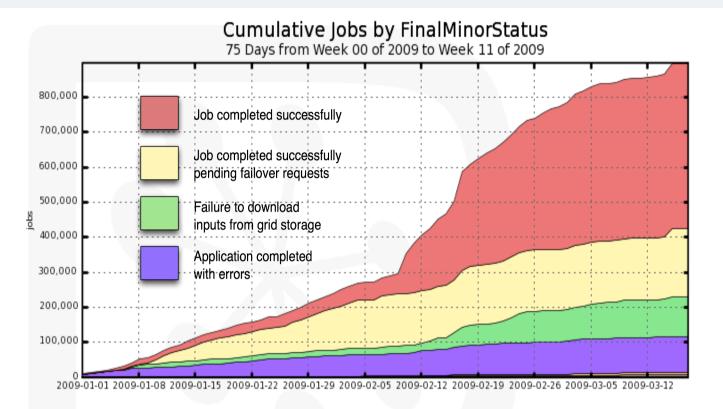
- A new Request Management System (RMS) to accept and execute asynchronously any kind of operation that can fail
 - Data upload and registration
 - Job status and parameter reports
- Request are collected by RMS instances on VO-boxes at 7 Tier-1 sites
 - Extra redundancy in VO-box availability
- Requests are forwarded to the central Request Database
 - For keeping track of the pending requests
 - For efficient bulk request execution







Failover redundancy



- Using failover system allows to recover almost all the successful jobs with incomplete operations
 - + up to 20% of the total number of jobs
- Applied initially for production jobs
 - Now is being extended also for the user jobs





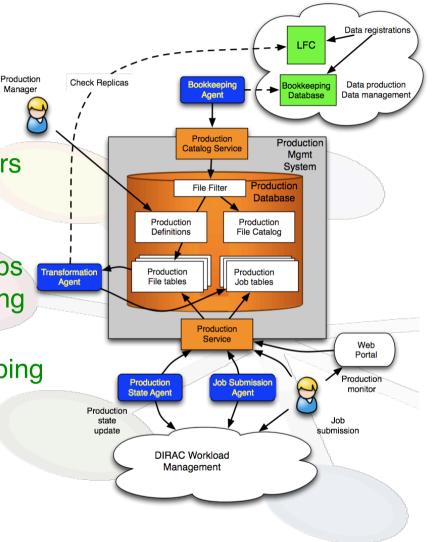


DIRAC: High level services

Manager

- Production Management built on top of the DIRAC WMS and DMS
 - Data requests formulated by users are processed and monitored using Web based tools
 - Automatic data reconstruction jobs creation and submission according to predefined scenarios
 - Interfaced to the LHCb Bookkeeping (metadata) Database
 - Built using the DISET framework

see [419] – A.Zhelezov et al

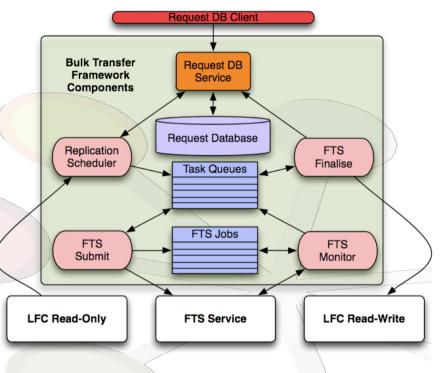






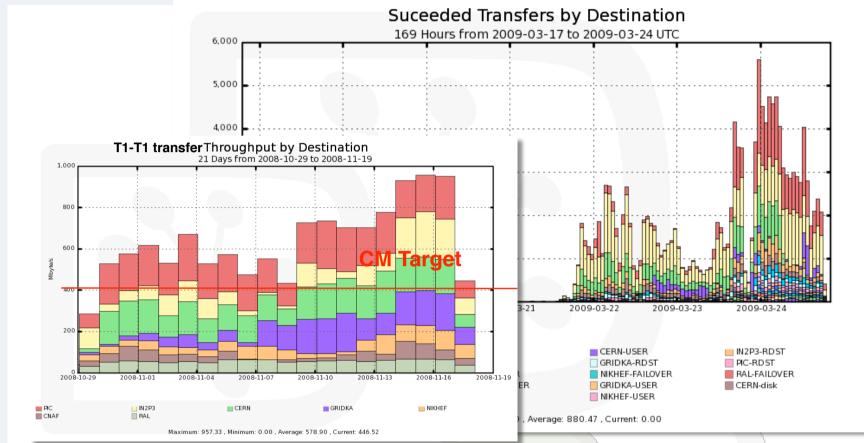
Data Management System

- All the Data Distribution operations
 - Pit to CERN transfers
 - + T0-T1 transfers
 - T1-T1 transfers
- Based on the Request Management System
 - Automatic transfer scheduling
 - Full monitoring of ongoing operations
- Using FTS for bulk data transfers
 - + Full failure recovery
- Comprehensive checks of data integrity in SEs and File Catalogs



Lнср





Extensively tested in a serious of tests (CCRC, FEST'09)
 Proven to support the LHCb Computing Model targets





DIRAC: Secure Web Portal

- Web portal with intuitive desktop application like interface
- Monitoring and control of all activities
 - User job monitoring and manipulation
 - Data production controls
 - DIRAC Systems configuration
- Secure access
 - Standard grid certificates
 - Fine grained authorization rules
- This work is in progress

see [369] – A.Casajus, M.Sapunov





Web Portal: example interfaces

DIRAC

					💱 🔹 Systems 🔻 Jobs 🔻 Proc	furtion T Data T Web T	
							Corport of
				💱 T Systems 🔻 Jobs 🔻 Data 🔻 W	/eb T		
				Text actions	😑 슬 LHCb Configuration		
				View configuration as text			
				Download configuration	G G Systems		LCG.CERN.ch
					WorkloadManagement		
				Modification actions	RequestManagement		Site Info
				Redownload configuration data from server	DataManagement DroductionManagement		Status: Allowed
😵 🔹 Systems 🔹 Jobs 🔹 I	Data T Web T			Show differences with server	Dogging	And the second second	Location: 6.0458° E, 46.2325° N
				Commit configuration	H Monitoring	The second second	Category: T0
JobMonitoring	K Select All				Accounting Bookkeeping		More Information
Selections	Jobid		Status		Framework	and the second se	
DIRAC Site:	Logging info for Jo		ng Requests		HCP		
All	Source	Status MinorStatus	App		🗄 🧰 Stager		
Status:	JobManager F	Received Job accepter			Resources Operations		
Completed	JobPath F	Received False	Uni	I	⊕ 🚞 Website	and the second and the second	e and a second se
Minor status:	JobSanity (Checking JobSanity	Uni	l i i i i i i i i i i i i i i i i i i i			
Pending Requests	-	Checking JobScheduli			 DefaultGroup = Ihcb_user DefaultProxyLifeTime = 432000 		and the second
Application status:			ubmissic Unl		H Clisers		
All		Matched Assigned Matched Job Receive	Uni d by Age: Uni		🕀 🧰 Groups		
Owner:		Matched Installing Sol			Hosts OMSMapping		Share a 🐙 🖉 a shekara 🕅
		Matched Submitted To		I	Contendpping	009-03-15 18:12 paterson	5
JobGroup:	JobWrapper F	Running Downloading	InputSa Uni	l i i i i i i i i i i i i i i i i i i i		009-03-15 18:12 paterson	
00004608		Running Application	Uni			y GridResourceBroker for last week 🛞	
Date:	-	Running Application	Exe				
YYYY-mm-dd		Running Application		uss v35r1 step 1 Sun Mar 15 2009 22:			
JobID:		Running Application		uss v35r1 Success Mon Mar 16 2009 01:		Pilots by GridReso 169 Hours from 2009-03-09	
	1894723	Completed Pend	ng Requests	Job Finished Succe LCG.Glasgow.uk	00004608_000285: 2009-03-16 01:56	800	
	1894722	Completed Pend	ng Requests	Job Finished Succe LCG.Durham.uk	00004608_000285: 2009-03-16 01:27	700	· · · · · · · · · · · · · · · · · · ·
	1894721	Completed Pend	ng Requests	Job Finished Succe LCG.Glasgow.uk	00004608_000285: 2009-03-16 01:53	600	
	1894670	Completed Pend	ng Requests	Job Finished Succe LCG.Glasgow.uk	00004608_000285: 2009-03-16 01:33		
	1894669	Completed Pend	ng Requests	Job Finished Succe LCG.CSCS.ch	00004608_000285: 2009-03-16 02:50	500	·····
	1894665	Completed Pend	ng Requests	Job Finished Succe LCG.Glasgow.uk	00004608_000285 2009-03-16 01:26	± 400	
	1894663			Job Finished Succe LCG.USC.es	00004608_000285 2009-03-16 03:07		
📀 Submit 🖉 Reset				Job Finished Succe LCG.CSCS.ch	00004608_000285_2009-03-16 02:57	300	
Global Sort	+ 1894661	— ·		Job Finished Succe LCG.RAL-HEP.uk	00004608_000285_2009-03-16 02:56	200	malden indiana indiana
Current Statistics	+ 1894660			Job Finished Succe LCG CSCS ch	00004608 000285 2009-03-16 02:50	In the second	
Global Statistics		e 63 of 639 🕨 🔰 🔅 It			and a second second second second		
jobs > Job monitor						0 2009-03-10 2009-03-11 2009-03-12 2009-03-13	2009-03-14 2009-03-15 2009-03-16

wms203.cem.ch h03.pic.es wms-1-fzk.gridka.de graspol.nikhef.nl wms216.cern.ch wms-3-fzk.gridka.de rb01.pic.es graszode.nikhef.nl wms010.cnaf.infn.it kgwms01.gridpp.rl.ac.uk wms006.cnaf.infn.it wms-2-fzk.gridka.de wms.grid.sara.nl kgwms02.gridpp.rl.ac.uk





Other new services

- Bookkeeping Service
 - Moved into the DIRAC framework
 - + see [449] Z.Mathe, E.Lanciotti
- Monitoring Service
 - Monitoring DIRAC components
- System Logging Service
 - Collecting error reports from all the components
 - Error analysis and alarm setting alarms
- Notification Service
 - Notifying users and experts on predefined events





Conclusions

- DIRAC time proven general architecture is now realized on a solid base technology
- It covers all the LHCb tasks for the distributed data processing
- DIRAC demonstrates the performance adequate to the LHCb needs for real Data Taking
- DIRAC has evolved to a general purpose grid system, is considered to be applied for other grids and application domains



DIRAC

http://dirac.cern.ch

22