

Supporting advanced HPC/HTC scientific workloads with QCG services



Bartosz Bosak, Piotr Kopta, <u>Tomasz Piontek</u> <u>{bbosak, pkopta, piontek}@man.poznan.pl</u> Poznan Supercomputing and Networking Center



A bit of history



http://www.qoscosgrid.org

Our motivation

It is not enough to give scientist access to resources. They need tools and support.



The world has changed... We cannot ignore this fact



Most of the new users need easy and intuitive tools to overcome the initial barrier in accessing the infrastructure



The idea

- QCG is a set of services, tools and libraries providing uniform and efficient access to HPC/HTC resources.
- Integrates many computing resources, but preserves simplicity of execution on a single machine.



- Provides highly efficient mapping for a variety of applications, such as parameter sweep, workflows, MPI or hybrid MPI-OpenMP, GPGPU, and multiscale.
- > Useful and applicable for both experienced and novice users.



Architecture



- > QCG consists of three basic layers:
 - \circ Client tools
 - High-level services (common)
 - Cluster-level services



Architecture (2)





Basic services: QCG-Computing

- Deployed on access nodes of the batch systems (Slurm, SGE, Slurm, torque/maui, LoadLeveler, PBS Pro, Condor, Apple Xgrid)
- Provides remote access to task submission and advance reservation capabilities of LRMS via DRMAA interface (the own DRMAA implementations)
- > Compatible with the OGF HPC Basic Profile specification (JSDL and BES)





Basic services: QCG-Broker

- > Offers capabilities for scheduling and brokering of jobs
- > supports multi criteria approaches taking into consideration many criteria like energy and resources consumption as well as time-to-finish
- > Controls the execution of whole experiments (including workflows and parameter sweep tasks)
- > Provides requested QoS and co-allocates resources
- > Stages in/out files and directories (the main transport mechanism based on gridFTP)





Basic services: QCG-Notification

- > Supports the topic-based publish/subscribe pattern for asynchronous message exchange
- > Serves as the main message bus between the services, applications and the end-user
- Is capable of sending notifications using variety of transport mechanism, including SOAP, SMTP and XMPP (notifications for users)
- > All major QCG clients indirectly use QCG-Notification to support tracking of statuses and progress of executed application (mail, XMPP, QCG-Monitoring)





Client tools & libs



Users are interested in tools not services ...

- > Scientists have different preferences in accessing the infrastructure what results in the necessity to provide them with various tools tailored to their habits and expectations. (Swiss army knife approach often fails).
- > Users very often can accept the fact that their jobs need long time to be finished but they would like to know when they finish.
- > For long running simulations users need a possibility to see partial results and to control the execution.



Client tools: QCG-Client

- > Set of commands modelled on queuing system tools
- > QCG-Simple, JSDL, QCG-XML description dialects
- > Detailed information about running and historical tasks
- > Support for interactive tasks
- > Automatic staging in/out files with gridFTP
- > Single access point to all resources

#QCG note=NAMD apoa1 #QCG host=hydra.icm.edu.pl

#QCG walltime=PT10M #QCG queue=plgrid #QCG nodes=1:12:12 #QCG output=apoa1.output #QCG error=apoa1.error #QCG application=NAMD #QCG argument=apoa1/apoa1.namd #QCG stage-in-file=apoa1.zip #QCG preprocess=unzip apoa1.zip #QCG stage-out-dir=. -> results #QCG notify=xmpp:tomasz.piontek@plgrid.pl #QCG watch-output=mailto:tp@mail,20,ENERGY

QCG

Client tools: QCG-Client (2)

> Submission and control of tasks:

- **qcg-cancel** cancel task(s)
- **qcg-clean** clean the working directories of given tasks
- **qcg-connect** establish interactive session to the task
- **qcg-info** display detailed information about the given tasks
- **qcg-list** list tasks in the system
- **qcg-peek** display ending of (stdout, stderr) streams
- **qcg-refetch** retry/repeat the transfer of output files/directories
- \circ **qcg-sub** submit batch or interactive tasks to be processes by QCG
- > Reservation and control of resources:
 - **qcg-rcancel** cancel reservation(s)
 - **qcg-reserve** reserve resources
 - **qcg-rinfo** display information about the given reservation(s)
 - **qcg-rlist** list reservation in the system
- > Information about resources:
 - **qcg-offer** provides detailed information about all managed resources, e.g. list of available applications, list of available modules, current load



Client tools: QCG-Portal

- > A web client to QCG, based on the Django framework
- > Offers functionality needed to submit, monitor and manage jobs over QCG middleware
- > Authentication/authorization based on OpenID
- > Includes built-in support for the gridFTP protocol
- > Support for job templates
- > In principle easily extendable and adaptable to certain needs
- Several ready application / domain based solutions
 built on top of QCG-Portal



Client tools: QCG-Portal (2)

lobs	s list			Search term	Q E	∠ ▲		
	Selected filters	Finished X Hydra X Moss X	Zeus X			Clear al filters		
			¢ 1	1 > of 1				
	Description	Submission	Start	End	Status	Host		
10	opisowy opis	03/18/2016 3:05 p.m.	03/18/2016 3:05 p	.m. 03/18/2016 3:29 p.m.	FINISHED	hydra	more -	
10		03/16/2016 4:02 p.m.	03/16/2016 4:02 p	m. 03/16/2016 4:28 p.m.	FINISHED	moss	more -	
10	python script	01/25/2016 1:37 p.m.	01/25/2016 1:37 p	m. 01/25/2016 1:38 p.m.	FINISHED	zeus	more -	
		01/25/2016 1:35 p.m.	01/25/2016 1:35 p	.m. 01/25/2016 1:36 p.m.	FINISHED	zeus	more	
10	basic task	01/19/2016 2:29 p.m.	01/19/2016 2:29 p	m. 01/19/2016 2:30 p.m.	FINISHED	hydra	more -	
10	basic task	01/19/2016 2:15 p.m.	01/19/2016 2:15 p	.m. 01/19/2016 2:15 p.m.	FINISHED	hydra	more -	
10	basic task	01/19/2016 2:11 p.m.	01/19/2016 2:11 p	m. 01/19/2016 2:11 p.m.	FINISHED	hydra	more -	
			1	of. 3				
		CosCosGrid Portal Joi Submit job	to Submit job GridFT	p			Maciej T Save as	
	-	CosCosGrid Portal Jet Submit job	ts Submit job GridF7 les Environment Mo Host	p p ontoring Others			Maciej T Save as	
	_	CosCesGrid Portal Jan Submit job Base Resources Fr	ts Submitjob GridF7 Hes Environment Mich Host	of 1 p p p p p p p p p p p p p p p p p p p			Maciej T Save as	
	_	CosCosGrid Portal Jan Submit job Base Resources Fr	is Submit job Grid 71 Ins Environment Mc Host Ourse	of 1 p p p p p p p p p p p p p p p p p p p		•	Maciej T Save as	
	_	CosCosGrid Portal Jan Submit job Bais Resource Fr	s Gubent job Groff ies Enveronment Mr. Host Guese Properties	of 1 p p p p p p p p p p p p p p p p p p p		•	Maciej T Save as	
	_	OosCosGrid Portal Jae Submit job Bait Resources Fr	s Bulent po Gref 7 Ins Environment Mr Host Dosse Properties Modules	of 1 p p p p p p p p p p p p p p p p p p p	nauny151 a		Mackej T Save as	
	-	OosCosGrid Portal Jae Submit job Baic Resources Fr	s Bulent po Gref 7 Ins Environment Mr Host Doses Properties Modules	of 1 p p p p p p p p p p p p p p p p p p p	nauny/16.1 s	•	Macing T Save as	
	-	OosCosGird Portal Jue Submit job Bait Resources Fi	s Bulmit job Gref 7 is Bulmit job Gref 7 Host Guess Properties Modules Processes Nodes topology	of 1 p p p p p p p p p p p p p p p p p p p	nawytâl e		Macinj T Save as	
		OosCosGrid Portal Jud Submit job Baac Presures Fr	s Butent yoo Graf 7 ba Butent yoo Graf 7 ba Environment Mr Host Granse Properties Modules Processes Nodes topology Wall time	of 1 p p p p p p p p p p p p p p p p p p p	enumpy1.6.1 +		Maciej T Savé as	
		OosCosGrid Portal Ju Submit job Bate Resources Fri	s Butent po Graf 7 s Butent po Graf 7 Host Host Droperties Modules Processes Nodes topology Vali time Memory (MB)	of 1 p p p p p p p p p p p p p p p p p p p	nnumpy1.8.1 ×	•	Maciej T Savé as	
		OosCosGrid Portal Ju Submit job Base Resources Pri	s Butent po Gref? s Butent po Gref? Host Properties Nodukes Processes Nodukes topology Wall time Memory per process (MB)	of 1 pp	naurpy1.6.1 c E Seconds		Macking T	



Client tools: QCG-Science-Gateways

Domain or problem oriented, highly specialized portal \geq solutions

			AND ST	A bet	er way to use grid					Add Page	
	Welcome	Simulations	File Manager	SSH Client	Configuration						
NanoC	lient										00
		Nan	otechnology				GRM83		Monit	oring	
Abinit											
Dase:	Total energy and D	OS charts			<+ Backt	to case list	Selected set result	s:			
				Advanc	Advanced mode Download results/visualization fil			es Details:			
Para	ameter set descriptio	n: 4 paramet	erset	Nu	mber of processes:	4	Show visuali	ation chart	i [
		1.			1		Case s	earch	1		
	The	unit cell (acell):	Bohr(a.u.)	× 10.18	10.18	8	Case keywords:				
	The primitive	vectors (rprim):	0.0 0.5	0.5) (1993		L	10.000	1		
			0.5 0.0	0.5		*	Parameter Set keywords:		r.		
	The aton	n types (ntypat):	1				L		1		
	The atomic i	number (znucl):	14								
Trou	ullier-Martins psp (pr	itencials files) :	14si pspnc		Upload	/Assign	Relative difference	of total energy (curen	t set)		
	Definition of the ato	ms - number of atoms (natom):	2	Visualization chart			000				
	Aton	is types (typat):	1.1	BOS chart for Bet 1	005 chart for Set 2	005 chart for 54t 3	DOS chief for Set.4				
The	location of atoms (o	ed) - Fractional	-	120							
	(atomic) coordinate	s, one triplet for each atom:	0.0 0.0 0.0 0.0 0.25 0.25 0.2	5 0	Age by	man	Ma.	-			
	Maximal kinetic e	nergy cut-off. In	8.0	-	Bases	ay(Ha)		2	3 Step (no.)	4	
		Hartree (ecut):	Law	129 -		٨					
E	Definition of the k-po	nt grid (kptopt):	Automatic	12 80				Print chart	Total energy obart (all param	eter sets)	
G	rid based on the prir the reciproca	space (ngkpt):	555	§ ∞		1	An			and states	
Max	imal number of SCF	cycles (nstep):	10		A						
Cut	t-off difference betwe	en evaluations	1.0d-6	\$ 20 /	toop -	A	16				
	oftotal	energy (toldle):		0 0.12 0.14 0	10 0.18 0.2 0.22 0	24 0.28 0.28	0.3 0.32 0.34				
1					Ene	rgs(Ha)					
Darca	to any J Cohome a	antino	anten [] S			et case	8 -9.869				
tinimu	im total energy -8	8726097094126					Iters				
Set	Description	Action		Total energy	Progress	Vis	3 18.87				
1	1 parameter set	Show	Cancel	8.8662238959759	FINISHED		-8.871				
2	2 parameter set	Show	Cancel	8.8718529095013	FINISHED			Total energ	gy chart (all parameter sets)	1	
3	3 parameter set	Show	Cancel	8.8724909739094	FINISHED		-0.072	9.8726723	1302567		
4	4 parameter set	Show	Cancel	8.8725723302567	FINISHED		-8.873				
6	6 parameter set	Show	Cancel	8.8725999848763	FINISHED		1 2	3 4 6	6 7 8 9 10 Set(ea.)	11 12 13 14	16
-	1						1				

POZNAN

INNOWACYJNA GOSPODARKA



Client tools: QCG-Now

- > A desktop (GUI) tool for submission of jobs to computing resources made available with QCG middleware.
- > Available as an installable package for Windows, Linux and OS X.
- > Integrated support for data transfer and application monitoring.
- > Support for job templates and command-line submission (easy integration with other applications).



http://www.qoscosgrid.org/qcg-now



Client tools: QCG-Now(2)

properties Т. Task definition view







Client tools: QCG-Monitoring

- > Monitoring progress of long-lasting jobs
- > Cyclic scanning of output files
- > Current status presented with text messages, tables, graphs and pictures
- > Predefined schemes for different types of applications (e.g. Gaussian)
- > Integrated with QCG-Client, QCG-Portal and QCG-Now



QCG-Monitoring An example monitoring scheme in QCG-Now

> Integration of QCG-Monitoring with QCG-Now (in the testing phase)



QCG

QCG-PilotJob Manager

HTC scenarios are also important ...

- > Schedules and manages a set of tasks on already allocated resources:
 - to eliminate waiting of every single task in a queue
 - to better utilize resources (better resource granularity)
 - \circ to "overcome" local administrative policies
- > Can read requests from various sources: file or network socket (static vs dynamic scenario)
- > Contains API (Python) for communication with user programs





QCG-PilotJob Manager

> Integration of QCG-PJM with EasyVVUQ for uncertainty quantification in the VECMA project



QCG

QCG-PilotJob Manager (future plans)

- > Ongoing work on introducing the concept of a global queue
 - \circ to use resources of many clusters
 - to adjust the amount of resources to the current need of the application by adding new allocations when needed
 - \circ to support exascale examples
- > Integration with the QCG-Monitoring system (almost done)





QCG-Comp-NG

QCG Next Generation services

Transition to new technologies

- > Mico-services approach for HA
- > Unified interface on cluster and 'grid' levels
- > Implemented in Python with modular approach
- > **REST** interfaces
- > New authentication and authorization
- First version deployed and validated at PSNC as a part of external commercial system





{ REST }







QCG-Comp-NG

QCG Next Generation Tools - Portal

Brs										B	Automatic refresh every	30 sec
	Modify	columns										Result cor
		ID	Commissioner	Commissioner role	Arrangement identi	Snapshot version	State	Queue	Thumbnail	Commission time	Last modified	Actions
									WSnow	Relative time	Minelative time	
missioner		9107	13	user	test-integration-sr	2019-04-01-08-57	FINISHED	user	C	15 minutes ago	11 minutes ago	8
	8	9106	11	user	test-integration-X	2019-04-01-08-37	FINISHED	user	C ²	35 minutes ago	31 minutes ago	ê ··
user (8947)		9105	11	user	test-integration-T	2019-04-01-08-17	FINISHED	user	C.	an hour ago	an hour ago	8
adviser 23		9104	13	user	test-integration-b	2019-04-01-07-38	FINISHED	user	C	2 hours ago	2 hours ago	8
user-role (1)		9103	11	user	test-integration-nJ	2019-04-01-07-17	FINISHED	user	C	2 hours ago	2 hours ago	8
angement identifier		9102	9	user	test-integration-IV	2019-04-01-06-57	FINISHED	user	C.	2 hours ago	2 hours ago	8
pshot version		9101	9	user	test-integration-W	2019-04-01-06-37	FINISHED	user	C.	3 hours ago	3 hours ago	8 -
		9100	11	user	test-integration-m	2019-04-01-06-17	FINISHED	user	C	3 hours ago	3 hours ago	8
te		9099	11	user	test-integration-w	2019-04-01-05-38	FINISHED	user	C	4 hours ago	4 hours ago	8
FAILED 632		9098	12	user	test-integration-q	2019-04-01-05-17	FINISHED	user	C ²	4 hours ago	4 hours ago	Ê
SUBMITTED 22		9097	11	user	test-integration-H	2019-04-01-04-57	FINISHED	user	C.	4 hours ago	4 hours ago	8
PENDING 9		9096	13	user	test-integration-d	2019-04-01-04-37	FINISHED	user	C	5 hours ago	5 hours ago	8
COMPLETING 1		9095	11	user	test-integration-g	2019-04-01-04-17	FINISHED	user	G	5 hours ago	5 hours ago	8
PROCESSING 0		9094	9	user	test-integration-K	2019-04-01-03-37	FINISHED	user	6	6 hours ago	ó hours ago	8
UE		9093	11	user	test-integration-V	2019-04-01-03-17	FINISHED	user	C.	6 hours ago	6 hours ago	8
user (8918)		9092	11	user	test-integration-ZI	2019-04-01-02-57	FINISHED	user	C	6 hours ago	6 hours ago	8
default-queue 33		9091	11	user	test-integration-Q	2019-04-01-02-37	FINISHED	user	C	7 hours ago	7 hours ago	8
itaff 26		9090	11	user	test-integration-B	2019-04-01-02-17	FINISHED	user	C ²	7 hours ago	7 hours ago	8
gold-user (1)		9089	12	user	test-integration-TT	2019-04-01-01-57	FINISHED	user	C.	7 hours ago	7 hours ago	â
mportant 1		9088	12	user	test-integration-Zf	2019-04-01-01-38	FINISHED	user	C.	8 hours ago	8 hours ago	8





Success stories



QCG is the primary middleware in Poland



Success stories























Project web page



QCG

OCG - Quality in Cloud and Grid

The QCG middleware (previously known as QosCosGrid) is an integrated system offering advanced job and resource management capabilities to deliver to end-users supercomputer-like performance and structure. By connecting many distributed computing resources together, QCG offers highly efficient mapping, execution and monitoring capabilities for variety of applications, such as parameter sweep, workflows, MPI or hybrid MPI-OpenMP. Thanks to QCG, large-scale applications, multi-scale or complex computing models written in Fortran, C, C++ or Java can be automatically distributed over a network of computing resources with guaranteed QoS. The middleware provides also a set of unique features, such as advance reservation and co-allocation of distributed computing resources.

QCG Middleware

QCG provides:

- · the most efficient remote access to computational resources in a single cluster or many clusters in Poland and Europe. · automatic steering of various types of complex computing experiments ranging from multi-parameter sweep studies to crosscluster executions of parallel applications,
- · fully transparent integration with parallel programming and execution environments like OpenMPI and ProActive located on many computing clusters,
- · support for Quality of Service (e.g. start time) based on advance reservation mechanisms,
- · shorter waiting times and improved resource utilization by hierarchical grid- and local-level job scheduling.

/www.qoscosgrid.org http:/

INNOVATION DRIVEN EFFICIENT COMPUTING



http://www.computing-innovations.org

News

QCG-Now 1.2 [3 January 2019] The refreshed version of QCG-Now has been released and it is available for download from the tool's webpage. The new version introduces a few new features, and improves the general stability of the program

UMD 4.5.0 [8 August 2017]

New versions of QCG-Broker and QCG-Broker Client has been published as part of UMD 4.5.0. The newly released packages has been marked with version number 4.2.0 and provide a support for execution of Array Jobs and possibility to get instant information about resources with the qcgresources command.

OCG-Now 1.0 beta [22 December 2016] QCG-Now - a new multiplatform desktop client for QCG - is already available for testing. It may be downloaded from the ⊕ product's webpage and easily installed on Windows, OS X or Linux, We encourage, especially the QCG-lcon's, users to play with a new software and give us feedback



Thank you!

Contact: piontek@man.poznan.pl