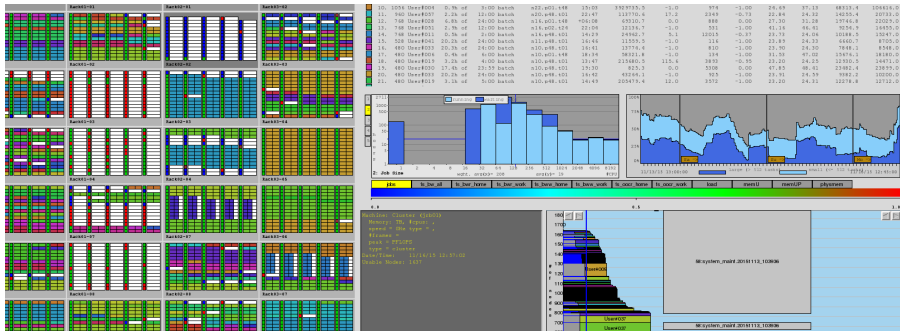




SYSTEM MONITORING AND JOB REPORTS WITH LLVIEW

May 29, 2018 | Carsten Karbach | JSC/Jülich Supercomputing Centre

MOTIVATION

- Is my job running?
- When will it start?
- How is the current load?
- How is my job placed?

Id	Owner	Submitted	ST	PRI	Class	Running On
juqueen1c1.192822.0	curioni	11/7 11:22 I	50	systemall		
juqueen1c1.192824.0	curioni	11/7 11:23 I	50	systemall		
juqueen1c1.192825.0	curioni	11/7 11:23 I	50	systemall		
juqueen1c1.192835.0	curioni	11/7 13:37 I	50	systemall		
juqueen4c1.82115.0	hhhh150	11/4 19:13 I	50	m018		
juqueen1c1.191375.0	hdu180	11/4 23:25 I	50	m008		
juqueen1c1.192826.0	jzam1159	11/7 11:27 I	50	m008		
juqueen1c1.192844.0	jhp0902	11/7 14:00 I	50	m008		
juqueen2c1.190024.0	pra08801	11/3 04:31 I	50	m004		
juqueen4c1.81936.0	jias1304	11/3 09:49 I	50	m004		
juqueen4c1.81937.0	jias1304	11/3 09:49 I	50	m004		
juqueen4c1.81938.0	jias1304	11/3 09:49 I	50	m004		
juqueen4c1.81939.0	jias1304	11/3 09:50 I	50	m004		
juqueen2c1.190675.0	jzam0420	11/5 13:42 I	50	m004		
juqueen1c1.190354.10	hwu081	11/3 12:22 I	50	m002		
juqueen2c1.190857.0	jiff4605	11/6 10:50 I	50	m002		
juqueen2c1.190971.0	hhh073	11/7 10:11 I	50	m002		
juqueen2c1.190972.0	grs30007	11/7 10:12 I	50	m002		
juqueen3c1.79205.0	j1nb3309	11/7 11:56 I	50	m002		
juqueen1c1.192846.0	jzam0435	11/7 14:18 I	50	m002		
juqueen4c1.80786.3	hbo273	10/28 11:49 I	50	m001		
juqueen2c1.189769.2	hbo381	11/1 23:48 I	50	m001		
juqueen1c1.190640.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190643.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190642.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190641.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190647.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190646.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190645.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190644.0	hch02r	11/4 07:03 I	50	m001		
juqueen1c1.190648.0	hch02r	11/4 07:03 I	50	m001		
juqueen3c1.78904.7	j1kp0501	11/5 09:39 I	50	m001		
juqueen2c1.190615.0	hhh045	11/5 09:55 I	50	m001		
juqueen2c1.190617.0	hhh045	11/5 09:58 I	50	m001		
juqueen4c1.82192.0	hhh045	11/5 14:42 I	50	m001		
juqueen1c1.82197.0	hhh045	11/5 15:48 I	50	m001		
juqueen2c1.190713.0	j1kp0403	11/5 16:27 I	50	m001		
juqueen2c1.190715.0	j1kp0403	11/5 16:35 I	50	m001		
juqueen2c1.190716.0	j1kp0403	11/5 16:41 I	50	m001		
juqueen1c1.191700.0	talahde	11/5 16:47 I	50	m001		
juqueen2c1.190718.0	j1kp0403	11/5 16:48 I	50	m001		
juqueen2c1.190723.0	j1kp0403	11/5 16:56 I	50	m001		
juqueen1c1.191701.0	talahde	11/5 16:59 I	50	m001		
juqueen2c1.190725.0	j1kp0403	11/5 17:11 I	50	m001		

More...

WHY SYSTEM MONITORING?

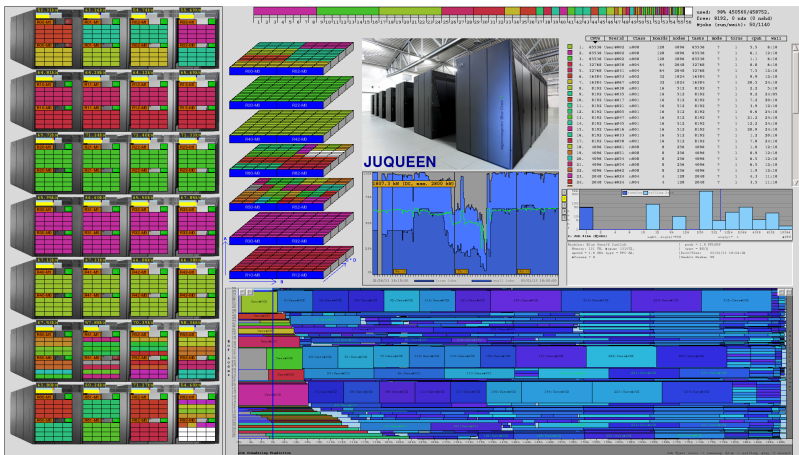
- For users
 - Controlling own running and waiting jobs
 - Planning job submissions
 - Use of idling resources
- For administrators
 - Global overview of system utilization
 - Throughput optimization
 - Batch system configuration optimization
 - Adaptive change of scheduling parameters

⇒ LLview

- Compact display of all usage data in one window
- Easy access to system's status data
- Interactive display for linking information
- Open Source (BSD-style)
- Available for all JSC systems

LLVIEW

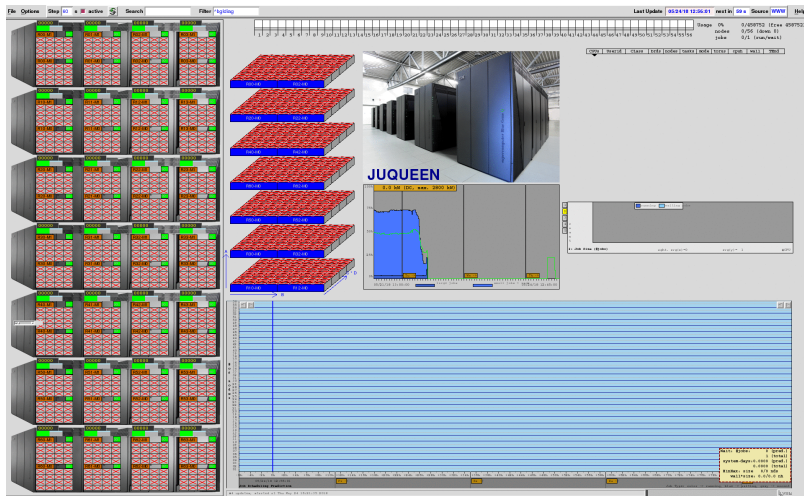
→ Visualizes supercomputer status on a single screen



Source: Screenshot LLview for JUQUEEN

LLVIEW

→ Visualizes supercomputer status on a single screen

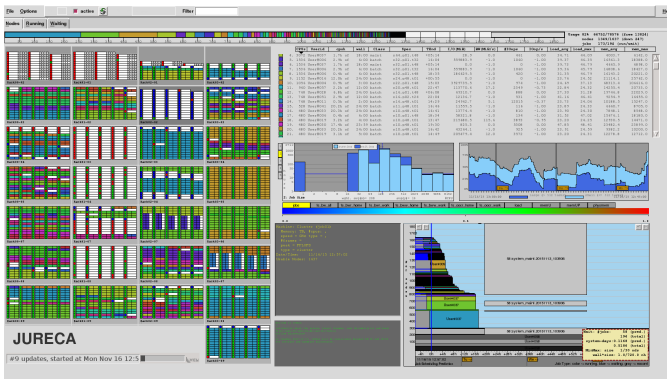


Source: Screenshot LLview for JUQUEEN
May 29, 2018

Slide 3115

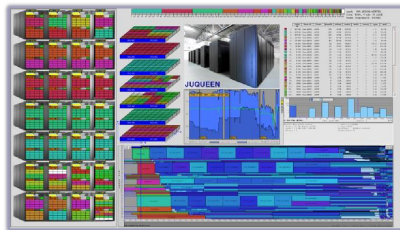
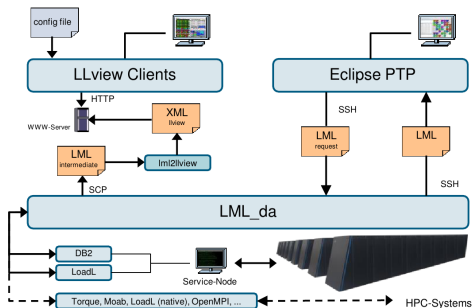
LLVIEW EXAMPLE: JURECA

- High fragmentation
- Heterogeneous
- Batch System: SLURM
- Ongoing development



Source: Screenshot LLview for JURECA

LLVIEW ARCHITECTURE

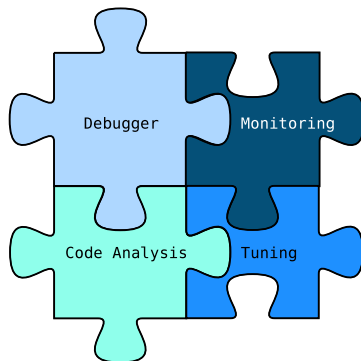


- Client-Server architecture, LML_da as backend
- Clients: Perl-Tk, PTP, Webinterface
- Platform independent: works on Windows, Mac and Linux
- Wide range of supported batch systems, minimal effort for extension
- Minor performance impact on monitored system, only **central batch** system is queried

PTP – PARALLEL TOOLS PLATFORM

What is PTP?

- **IDE** for parallel application development
- Based on **Eclipse**
- **Open-source** project
- Developers:
IBM, U.Oregon, UTK,
Heidelberg University, NCSA,
UIUC, JSC, ...



- **JSC-PTP tutorials** → <http://www.fz-juelich.de/SharedDocs/Downloads/IAS/JSC/EN/PTP/JSCPTPJunqueen.html>
- **PTP Download** → <http://www.eclipse.org/downloads/eclipse-packages/>, **Eclipse** for Parallel Application Developers

HOW TO START THE LLVIEW CLIENT? I

- Four options to start LLview, sorted by effort to get started

Option 1: via SSH

```
ssh -X karbach@jureca
llview
```

Option 2: Webinterface

- Screenshots of LLview updated every minute (static)
- Link (static): <https://llview.fz-juelich.de/LLweb/jureca/Image.html>
- Access secured by JSC webservice accounts
 - register at dispatch
 - request access to LLview via jsc-dispatch@fz-juelich.de

HOW TO START THE LLVIEW CLIENT? II

Option 3: VNC

- start VNC server on JURECA with
`vncserver -profile vis`
- tunnel VNC traffic to local system
- start VNC viewer
- click on LLview links
- detailed step by step guide [here](#)

Option 4: Local installation

- [Download and install](#) the LLview client locally

JOB REPORTS

What can I get?

- Detailed reports on active and finished jobs
- Job summary and time-based diagrams of many job metrics such as load, memory usage, I/O
- Use cases: development, production checks, job health check, light-weight performance analysis

Details

- Includes metrics for CPUs and GPUs
- No instrumentation needed, data is retrieved from IPMI once per minute
- Development in progress

ACCESS

- login: `https://llview.fz-juelich.de/LLweb/jureca/jobreport/login.php`
- Authenticate with webservice account, register [here](#)
- Only available for JURECA

Info Statistics:
Account / Group: user1090 / grp290
Time/date of data: 17/01/21 - 01:00:35
#jobs running/24h/2weeks: 8 / 0 / 125
☒ Data submitted

Jobs ended before 24 hours, max 2 weeks

[Show table metrics description](#)

job ended in Newswatch: 24.00 to 136.00 hours, 72% jobs total

[illegible]

- Lists for active and finished jobs
- Updates after 1-5 minutes
- Colormap for quick evaluation of metrics
- PDF column on the right leads to detailed job report

DETAILED JOB REPORT AS PDF

Job-Info (Jureca Batch, GPU, Booster nodes)

Jobid: 4127061	User: user1009	Group: grp209	Date/Time of job data: 17/11/09-10:09:02
Job runtime: 1h36m --> 80.00% of wall: 2h00m	Job Performance metrics		
Job starttime: 17/11/09-08:32:51	Load (CPU-Nodes):	min 0.01	avg 41.01
Job last timestamp: 17/11/09-10:09:05 (running)	Memory (CPU-Nodes):	3103.00	3200.28
Job endtime (est.): 17/11/09-10:33:01	Interconnect Traffic (in):	0.00	0.68 MB/s
Queue: batch	Interconnect Traffic (out):	0.03	0.50 MB/s
Job Size: #Nodes: 16	Interconnect Packets (out):	22	215 jobs
Interconnect Packets (out):	22	215	jobs
Job I/O statistics	Total Data Write	Total Data Read	max. Data rate/Node Write
SWORK:	64.44 MB	0.00 MB	0.00 MB/s
	5.63 MB	0.14 MB/s	0.09 MB/s
			max. Open Close Rate/Node
			1.60 ops
			3.78 ops

Job-Usage Overview

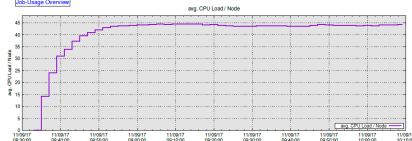


Table of Contents (6 pages)

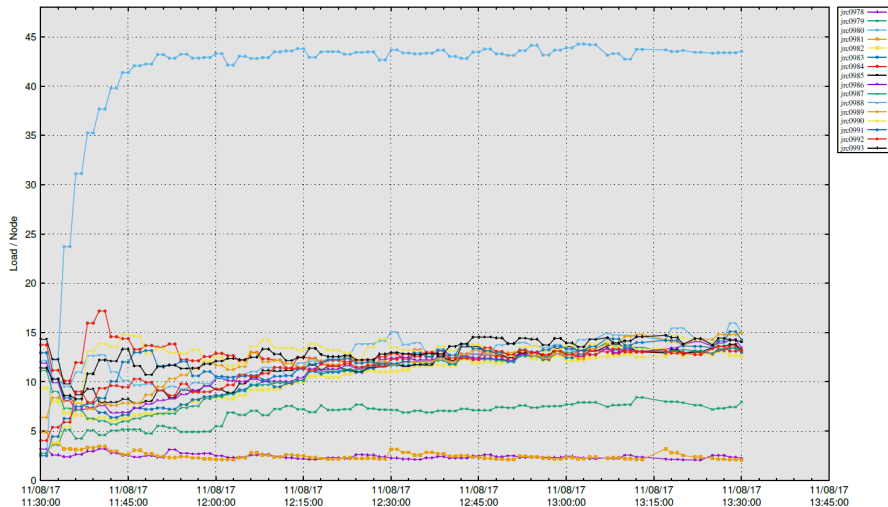
Load and Memory	2
Data transfer/node on inter-communication fabric	3
I/O usage history on WORK	4
Load and Memory history (MODE view (nodes 0..16))	5
I/O usage history on WORK (MODE view (nodes 0..16))	6

- Diagrams for all available metrics per job over runtime
- Overall diagrams and diagrams per compute node

EXAMPLE JOB – BAD LOAD BALANCING

CPU Nodes: Load [nodes 0 .. 16]

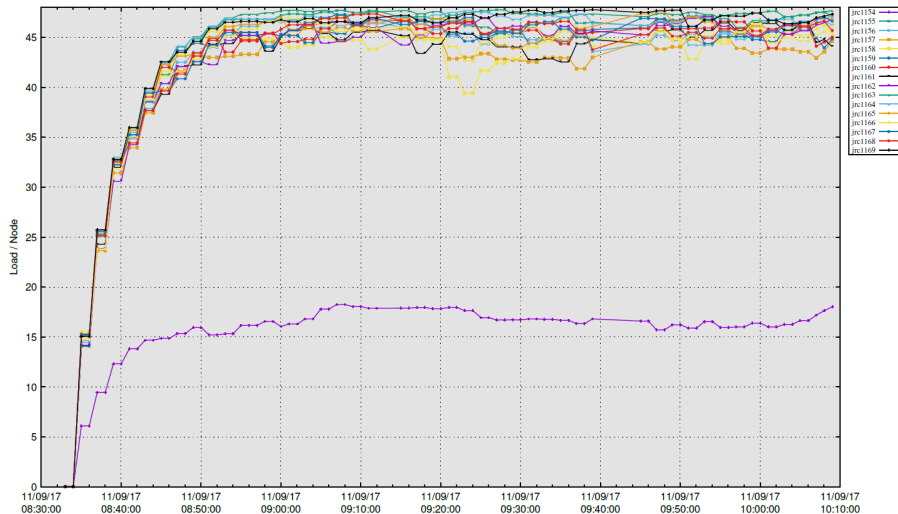
Total Load: 193.01, Average/Node: 12.06, Max/Node: 44.28



EXAMPLE JOB – GOOD LOAD BALANCING

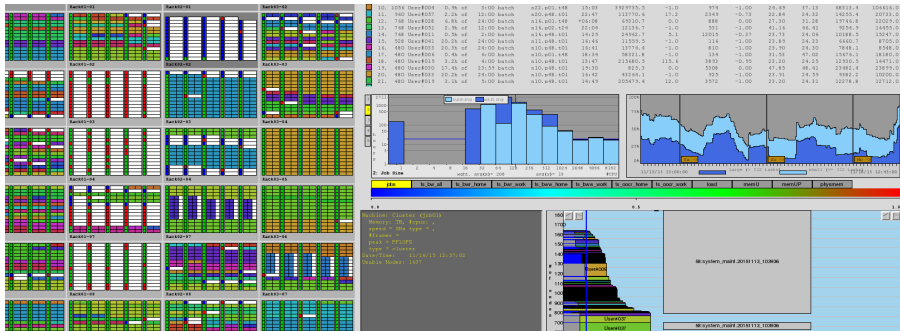
CPU Nodes: Load [nodes 0 .. 16]

Total Load: 656.16, Average/Node: 41.01, Max/Node: 47.78



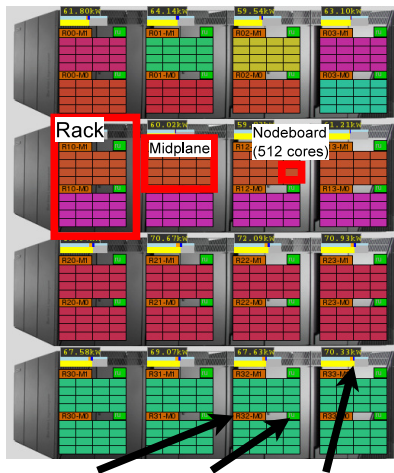
CONTACT

- **E-mail:**
`llview.jsc@fz-juelich.de`
- **LLview** → `http://www.fz-juelich.de/jsc/llview`
- **Job Reports** → `https://llview.fz-juelich.de/LLweb/jureca/jobreport/login.php`
- **JSC-PTP tutorials** → `http://www.fz-juelich.de/SharedDocs/Downloads/IAS/JSC/EN/PTP/JSCPTPJunqueen.html`
- **PTP Download** →
`http://www.eclipse.org/downloads/eclipse-packages/`



Part III: Appendix – LLview Components

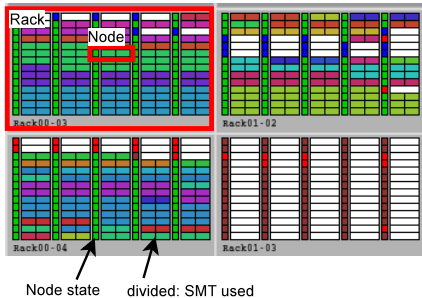
NODE DISPLAY



- Compute resources
- Job distribution
- White = Idle, Colored = running
- Node name
- Node status
- Level of detail

Midplane name Midplane state Rack power usage

NODE DISPLAY



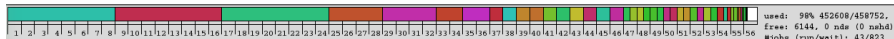
- Compute resources
- Job distribution
- White = Idle, Colored = running
- Node name
- Node status
- Level of detail

JOB LIST

	CPUs	Userid	Class	boards	nodes	tasks	mode	torus	cpuh	wall	
	1.	65536	User#009	m008	128	4096	65536	?	1	1.1	6:10
	2.	65536	User#002	m008	128	4096	65536	?	1	1.5	12:10
	3.	65536	User#009	m008	128	4096	65536	?	1	3.1	6:10
	4.	32768	User#024	m004	64	2048	32768	?	1	1.1	12:10
	5.	32768	User#019	m004	64	2048	32768	?	1	1.5	8:10
	6.	16384	User#039	m002	32	1024	16384	?	1	15.4	24:10
	7.	16384	User#061	m002	32	1024	16384	?	1	1.1	24:10

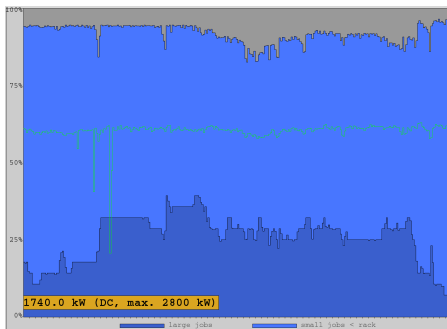
- List of running jobs
- Most important attributes per job
- Sort by clicking on the column header
- Identifying color next to each job entry

USAGEBAR



- Summary of system load
- Job size decreases from left to right
- White space shows idling resources
- Unit for JUQUEEN is midplanes,
for JURECA nodes

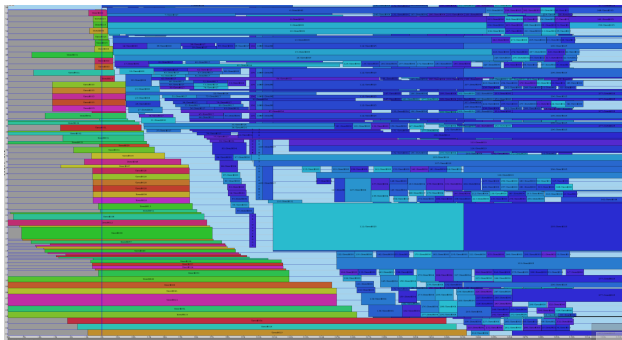
HISTORY



- 3-day load history
- Often divided into small and large jobs
JUQUEEN (1 midplane), JURECA (512 tasks)
- Mouse-Over for detailed information
- Green line for special history value
JUQUEEN (power)

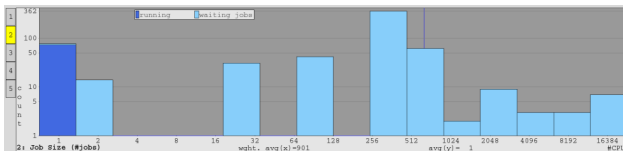
PREDICTION

- Scheduler prediction based on submitted jobs
- Wall clock limit as job duration
- Blue = predicted job, Colored = running jobs
- Each rectangle one job,
x-axis = time, y-axis = nodes/midplanes
- Use of idle times
- More transparent scheduling
- JUQUEEN: self-implemented, JURECA: use of SLURM's prediction



STATISTICS

- Statistic overview on system status
- Histograms on job size, wait time, queue load
- Highly configurable, define x-axis/y-axis domain, logarithmic/linear scale
- Overlaid diagrams for waiting/running jobs



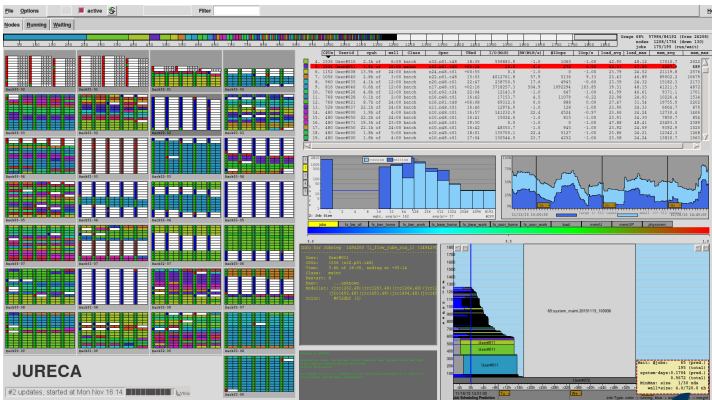
INFOBOX

```
Machine: Blue Gene/Q Juelich      | peak = 5.9 PFLOPS  
Memory: 448 TB, #cpus: 458752,   | type = BG/Q  
speed = 1.6 GHz type = PPC A2,    | Date/Time: 04/02/13 17:58:02  
#frames = 28                     | Usable Nodes: 56
```

- Shows details on the currently focused object
- Mouse-Over triggers to display detailed data on the focused job/node/system/diagram

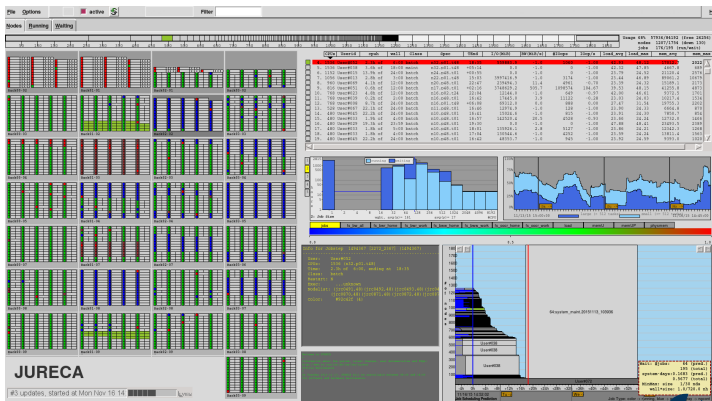
INTERACTION

- **Mouse-Over** jobs highlights job rectangles for the selected job in all components and shows details on the job in infobox



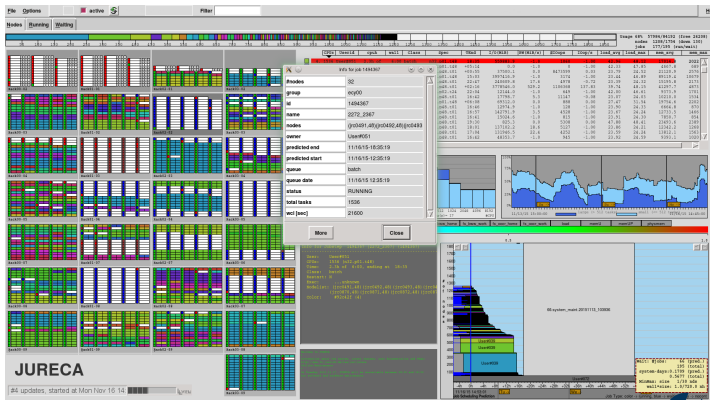
INTERACTION

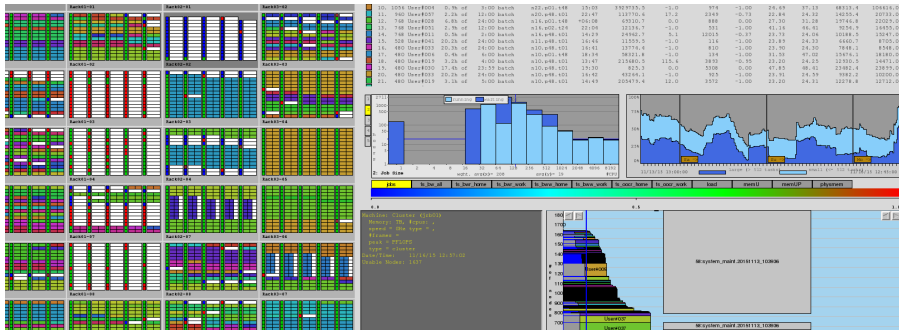
- **Mouse-Over** jobs highlights job rectangles for the selected job in all components and shows details on the job in infobox
- **Mouse-Down** (Hold) removes color for all other jobs, only the selected job is colored



INTERACTION

- **Mouse-Over** jobs highlights job rectangles for the selected job in all components and shows details on the job in infobox
- **Mouse-Down** (Hold) removes color for all other jobs, only the selected job is colored
- **Double-Click** show job detail dialog
- Mouse interaction helps to link the information shown in all components





Part IV: Appendix – LLview Customization

LLVIEW

- LLview is **highly customizable** due to numerous options
- Settings specific to HPC system type
- Start through option menu of main window or `Ctrl+o`
- Most options have **immediate effect** in the main window
- Some will become active at next start of LLview (e.g. [Data source](#) change)
- LLview layouts use **absolute** positioning
- You can use arrow keys to add/subtract one on numeric values
- Use Page up/Page down keys to add/subtract a bigger step on numeric values

LLVIEW CONFIGURATION FILES

Three configuration file locations (highest priority first):

- 1 anywhere on your file system passed to LLview with the `-rcfile` option
 - 2 local `.llview.rc` configuration file in current directory or in HOME directory of the user
 - 3 `llview.rc` in the installation directory of LLview. This file contains the system-wide settings
- Configuration files contain all LLview options
 - You can change them in any text editor or via the LLview Option window

LLVIEW OPTIONS

General	Elements	LocalData	WWW	llqxml	Info	Status	UsageBar	Nodes	NodeBox	LogView	Joblist	His	
WWW: from Web-Server llqxml: Execute local command LocalData: tar file on local machine													
verbose		☐ on/off		demo version		☐ on/off		show +/- buttons					☐ on/off
Node selection regexp			.*		Anonymise user names for demonstration purposes. LLview restart required								
Job selection regexp (uid)			^bgldiag										
RC		☒ on/off		RC_id		_juqueen							
Height		640		Width		1220							
Height (Lines)		61											
Geometry													
Canvas Color		grey85											
Update		☒ on/off		Update time (s)		60							
auto play		☐ on/off		Autoplay Step (s)		5							
Mark Color		red											
Mark Width		2											
version		1.3											
no frame for mainwindow				☐ on/off				notimestate					☐ on/off

GENERAL OPTIONS

- General options for the main window
- Choose your preferred data source ([Web-Server](#), [LocalData](#) or local command)
- [demo version](#): anonymise usernames (for public display)
- [Job selection regexp](#): filtering jobs by regular expressions
- Customize [Height](#) and [Width](#) of the main window
- [Canvas Color](#): background, [Mark Color](#): color for marking job in job list
- Choose time until next update
- [auto play](#) lets LLview mark different jobs automatically (for public display)

LLVIEW ELEMENT OPTIONS

General	Elements	LocalData	WWW	llqxml	Info	Status	UsageBar	Nodes	NodeBox	LocalData
!!! changes on following options have only effect !!! !!! after save options and restart llview !!!										
show usage bar ☒ on/off					show nodes ☒ on/off					
show joblist ☒ on/off					show running ☐ on/off					
show waiting ☒ on/off					show graph ☐ on/off					
show histogram ☒ on/off					show status ☒ on/off					
show info ☒ on/off					show history ☒ on/off					
show partition (BG) ☐ on/off					show reservations (BG) ☐ on/off					
show prediction of usage ☒ on/off					show usage history ☒ on/off					

- Choose, which Elements to show
- For end users: components like **joblist**, **info**, **nodes** and **prediction** etc.
- Changes take effect after **restarting** LLview

LLVIEW NODE OPTIONS

General	Elements	LocalData	WWW	llqxml	Info	Status	UsageBar	Nodes	NodeBox	LogView	Joblist	Histogram	Fe
X position	5	Y position	0	Height	640	Width	380						
Box Margin West	0	Box Margin East	-2	Box Margin North	-0	Box Margin South	-7						
Draw border	☒ on/off	Debug Layout	☐ on/off	BOX Color	grey85								
Twin View	☐ on/off	View Type	Both	Selector X	655	Selector Y	16	max select #	3				
Usagebars	☐ on/off												
Node attr.	jobs	min	-	max	-								
colmap	☐ on/off	colmap x	0	colmap y	0	colmap width	30	colmap height	100				
col map vertical	☒ on/off	number format	%2.1f	colmap unit	☐ on/off	colmap scale factor	0.6						
colmap font	Monospace												
use User Layout	☐ on/off	Layout	(rack:R00-M0,R00-M1,width=70,height=330,order=down,stack=down,frame=yes,fill=grey50,bo										
Racks per Row	4	Rack gap X	11	Rack gap Y	11	named racks	☐ on/off						
Font (State)	-*-Helvetica-Medium-R-Normal--*-80-*-*-*-*												
Font (Action)	-*-Helvetica-Medium-R-Normal--*-100-*-*-*-*												
Font (NodeName)	-*-Helvetica-Medium-R-Normal--*-80-*-*-*-*												
Font (SiteName)	-*-Helvetica-Bold-R-Normal--*-240-*-*-*-*												
Font (Power)	-adobe-Courier-Medium-R-Normal--08-100-75												
show InOUT	☐ on/off	InOut pos. X	858	InOut pos. Y	419								
show Logo	☐ on/off	Logo pos. X	9	Logo pos. Y	0	Image name	lib/images/JUGENE_logo						
show Site Name	☒ on/off	Name pos. X	622	Name pos. Y	389								
Color	darkblue	Site Name	JUQUEEN										

NODE OPTIONS

- Customize **Height**, **Width** and **Margins** of the node display
- Logical node view available for BG systems e.g. JUQUEEN ⇒ **Twin View** places adjacent midplanes next to each other in torus network
- Options for the **Twin View** are available in a new option subfolder ⇒ **Log View**
- **Node attr.:** show scalar data e.g. temperature or power usage
- Show logo or site name

LLVIEW HISTOGRAM OPTIONS

General	Elements	LocalData	WWW	llqxml	Info	Status	UsageBar	Nodes	NodeBox	LogView	Joblist	Histogram
---------	----------	-----------	-----	--------	------	--------	----------	-------	---------	---------	---------	-----------

Diagram 1	Diagram 2	Diagram 3	Diagram 4	Diagram 5	Diagram 6	Diagram 7	Diagram 8
-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

diagram title	Job Wait Time		
Jobselection	ALLSEP	Legend offset X	180
		Legend offset Y	2
X-Axis data	QUEUE TIME		
Y-Axis data	COUNT		
Stepwidth (xdata)	12		
Log x data	LINEAR		
Log y data	LOG10		
Format X	day	Format Y	%3d
Format AVG X	day	Format AVG Y	%3d
Fill Color	darkblue	Fill Color (Run)	darkgreen

Global Option for all diagrams

Display Diagram	Diagram 1		
# diagrams	5	autoplay delay	10
posx	244	posy	587
Height	179	Width	300
Font	-adobe-courier-medium-r-normal-*-*60-*-*-*		
BoldFont	-adobe-courier-bold-r-normal-*-*70-*-*-*		
number Font	*-Courier-Medium-R-Normal-*-*120-*-*-*		
Padding left	30	Padding right	50
		Padding bottom	25
		Padding top	15
Button width	10	Legend width	135
		Legend height	15

HISTOGRAM OPTIONS

- Each histogram shows distribution for a single job attribute, e.g. waiting time
- You can configure up to 8 histograms
- Jobs are grouped into discrete classes, y-axis shows count of jobs in each class
- Y-axis may also show number of CPUs, CPU hours or job duration
- Scaling may be linear or logarithmic
- **Auto play** is available for public display