

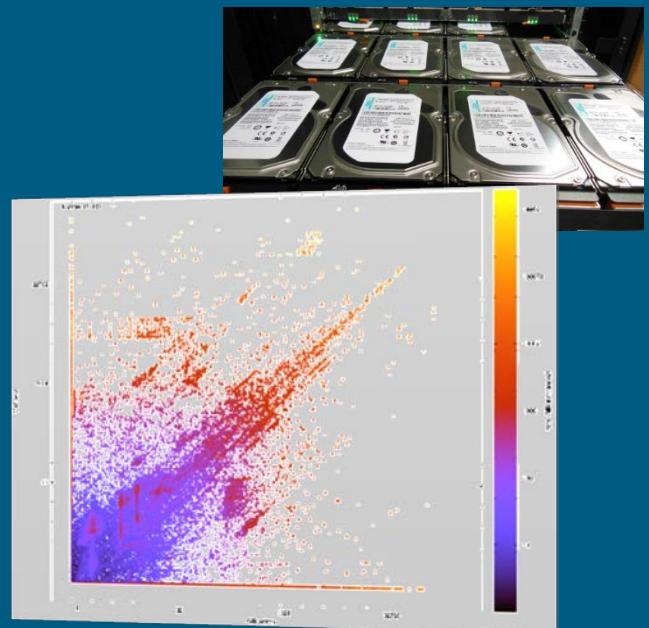
# Job-based I/O-monitoring with LLview

- I/O Monitoring with LLview & GPFS
- I/O Workloads on JURECA

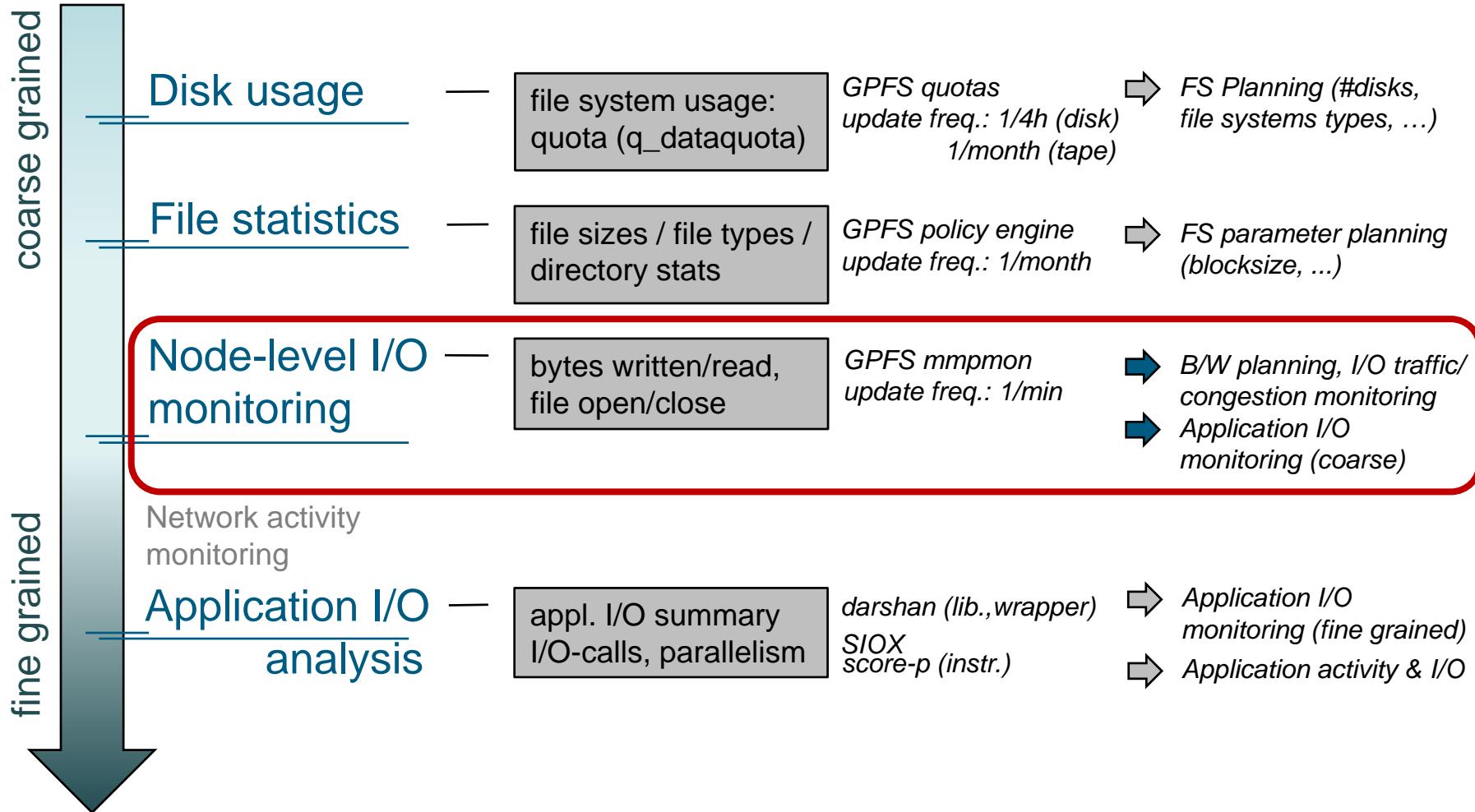
*Wolfgang Frings*

[W.Frings@fz-juelich.de](mailto:W.Frings@fz-juelich.de)

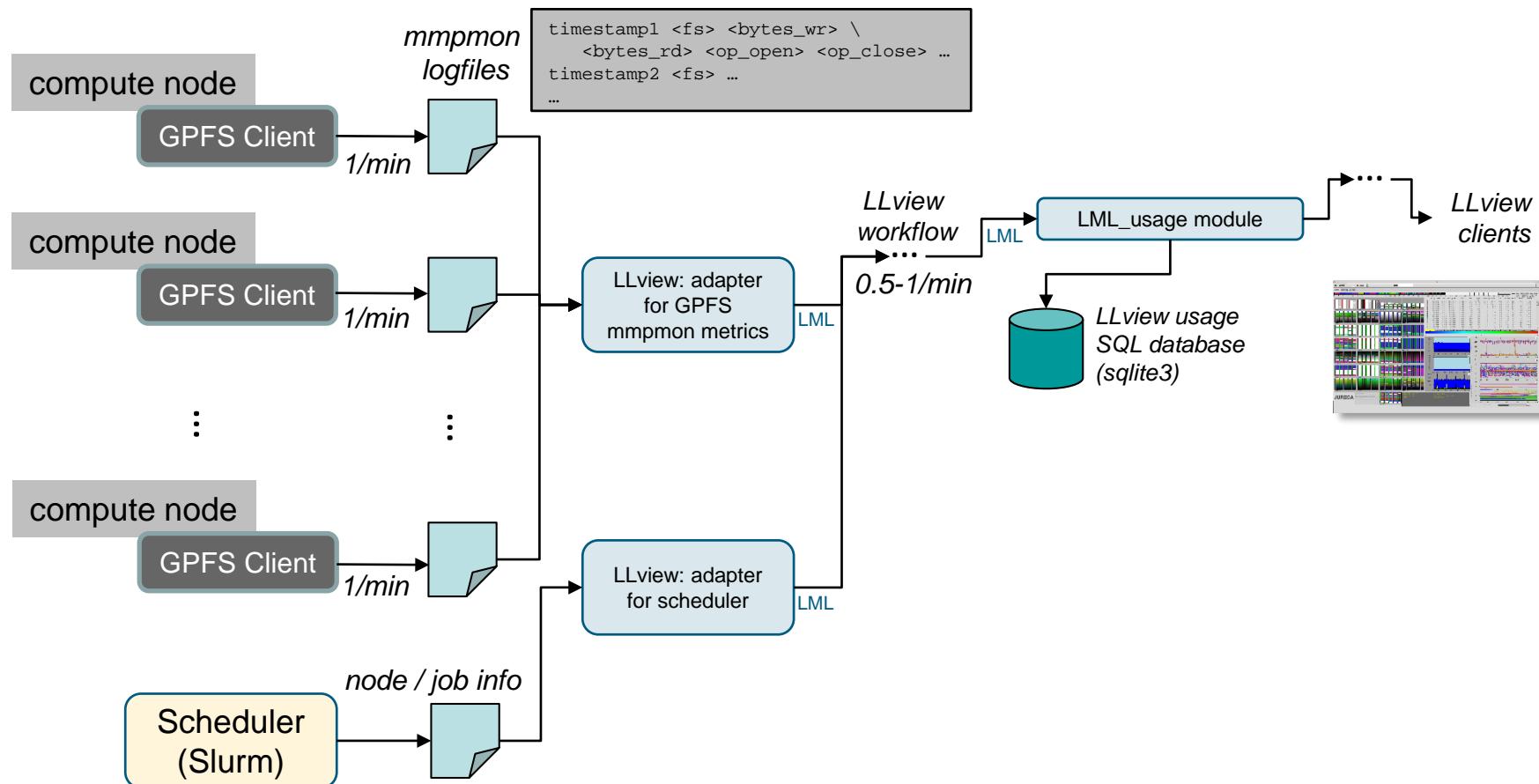
Jülich Supercomputing Centre



# I/O Usage Monitoring and I/O Statistics

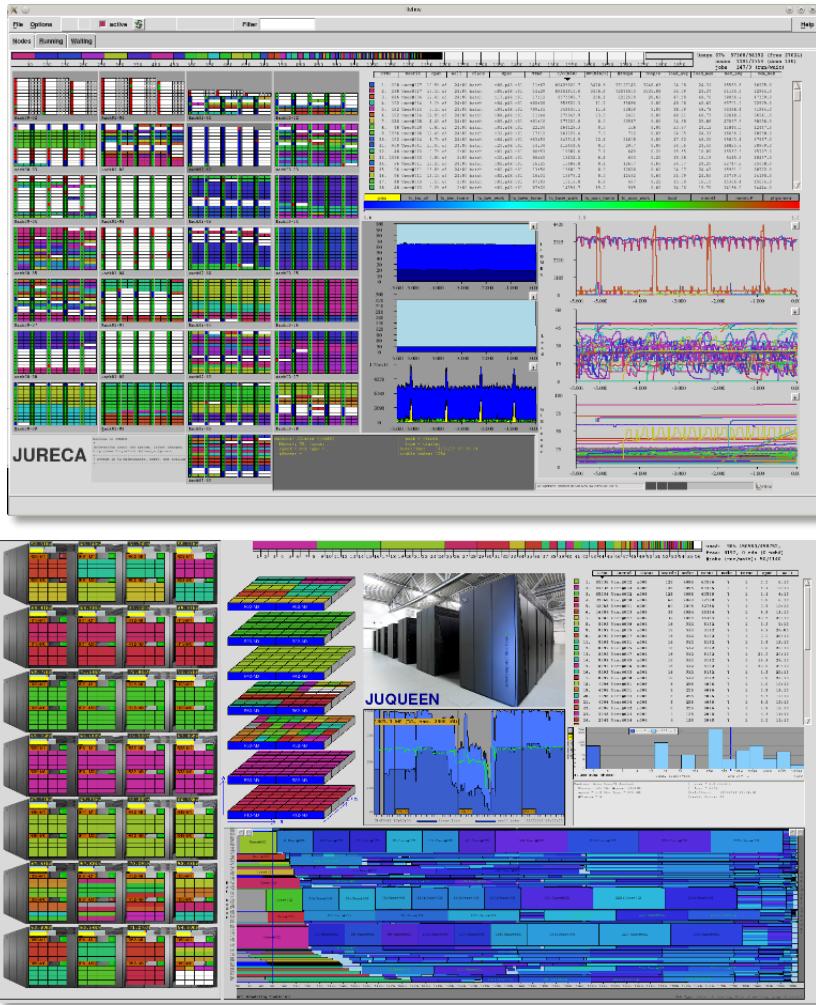


# GPFS mmpmon Data Workflow in LLview



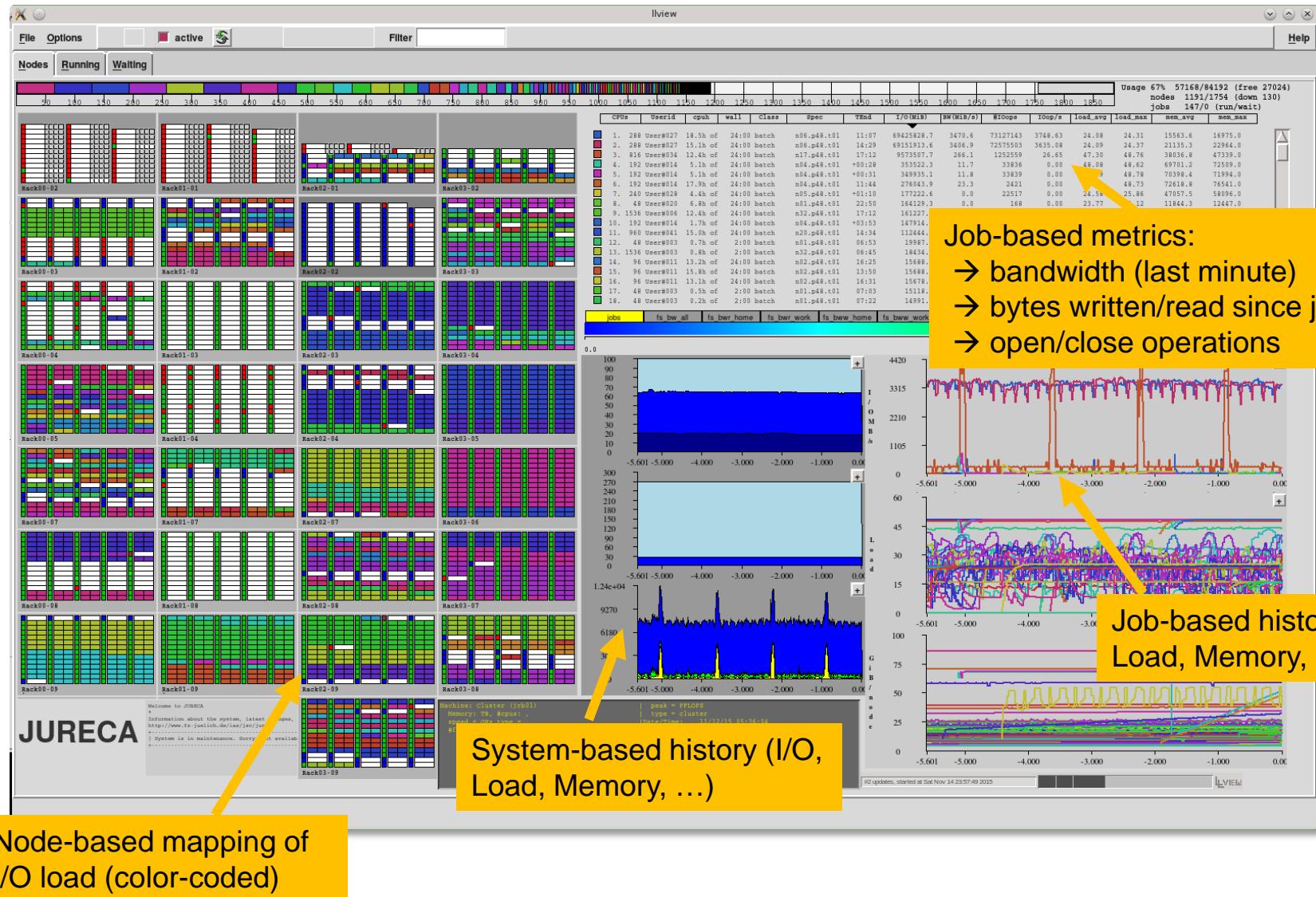
# LLview: User-level Monitoring

- Efficient supervision of node usage, running jobs, statistics, history
- Monitoring of energy consumption, load, memory usage, **I/O usage**, GPU usage
- Interactive and mouse-sensitive
- Main data source: batch scheduler, runtime system
- Web-based HTML5 client (in preparation)
- Fully customizable, fast and portable client-server application
- Integrated into **Eclipse/PTP**
- Support for various resource manager, incl. **LoadLeveler**, IBM Blue Gene, Cray ALPS, PBSpro, Torque, SLURM, Grid Engine and LSF



*LLview download: (open source)  
<http://www.fz-juelich.de/jsc/llview>*

# LLview: Node & File System Metrics



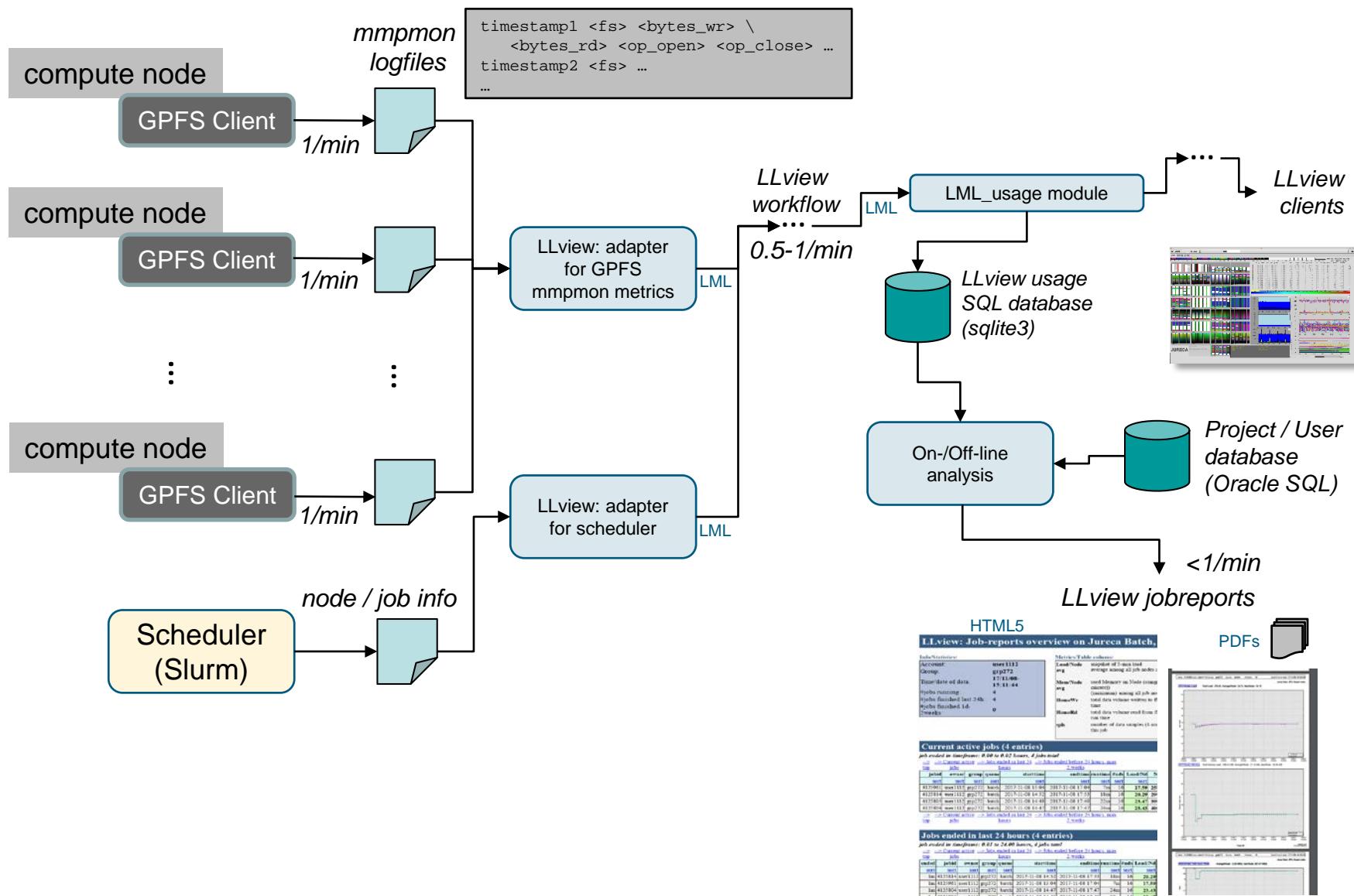
Node-based mapping of  
I/O load (color-coded)

Job-based metrics:  
 → bandwidth (last minute)  
 → bytes written/read since job start  
 → open/close operations

Job-based history (I/O,  
Load, Memory, ...)

System-based history (I/O,  
Load, Memory, ...)

# GPFS mmpmon Data Workflow in LLview



# LLview: Job reports

- Detailed reports on active and finished jobs
- Job summary and time-based diagrams
- Use cases: development, production checks, job health check, light-weight performance analysis
- Metrics per Node/GPU:
  - CPU load (5-min avg.)
  - Interconnect (MB/s in/out, packets in/out)
  - Memory usage (GiB)
  - I/O (per FS: GiB, GiB/s write/read, #Open/Close)
  - GPU (Usage, Memory, Power, Temperature, ...)
- Data Sources: SLURM, GPFS, IB, GPU
- Lists for active and finished jobs (< 2 weeks)
- Updates after 1-5 minutes
- Colormap for quick evaluation of metrics
- Link to PDF: detailed job report

LLview: Job-reports overview on Jureca Batch,

Info/Statistics:		Metrics/Table columns:	
Account:	user1112	Load/Node	snapshot of 5-min load
Group:	grp272	avg	average among all job nodes
Time/date of data:	17/11/08-15:11:44	Mem/Node	used Memory on Node (comp. minute))
#jobs running:	4	avg	(maximum) among all job nodes
#jobs finished last 24h:	4	HomeWr	total data volume written to file system
#jobs finished 1d-2 weeks:	0	HomeRd	total data volume read from file system
		spls	number of data samples (1-min) for this job

Current active jobs (4 entries)

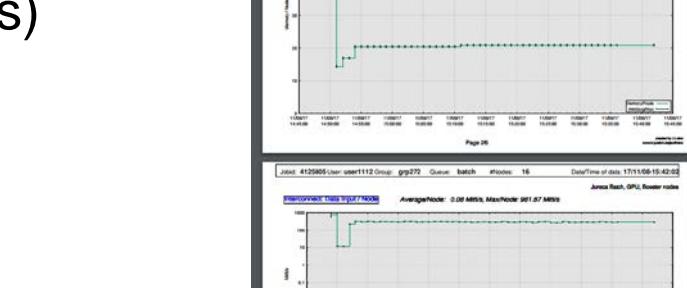
job ended in timeframe: 0.00 to 0.02 hours, 4 jobs total					
--> --> Current active --> Jobs ended in last 24 hours --> Jobs ended before 24 hours, max 2 weeks					
top	jobid	owner	group	queue	start
	4125961	user1112	grp272	batch	2017-11-08
	4125814	user1112	grp272	batch	2017-11-08
	4125805	user1112	grp272	batch	2017-11-08
	4125804	user1112	grp272	batch	2017-11-08

Jobs ended in last 24 hours

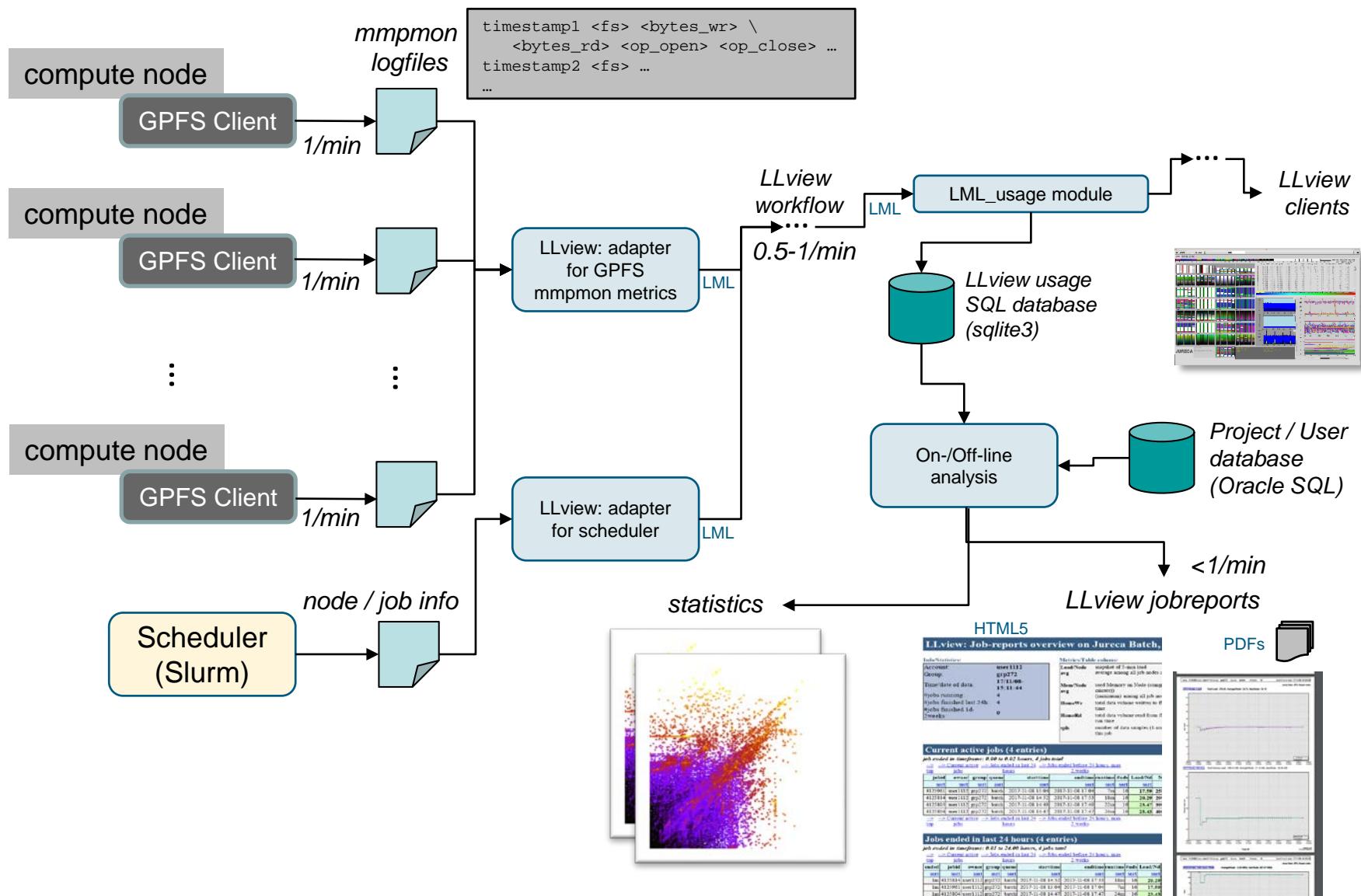
job ended in timeframe: 0.01 to 24.00 hours					
--> --> Current active --> Jobs ended in last 24 hours					
top	ended	jobid	owner	group	queue
	1m	4125814	user1112	grp272	batch
	1m	4125961	user1112	grp272	batch
	1m	4125805	user1112	grp272	batch
	1m	4125804	user1112	grp272	batch

--> --> Current active --> Jobs ended in last 24 hours

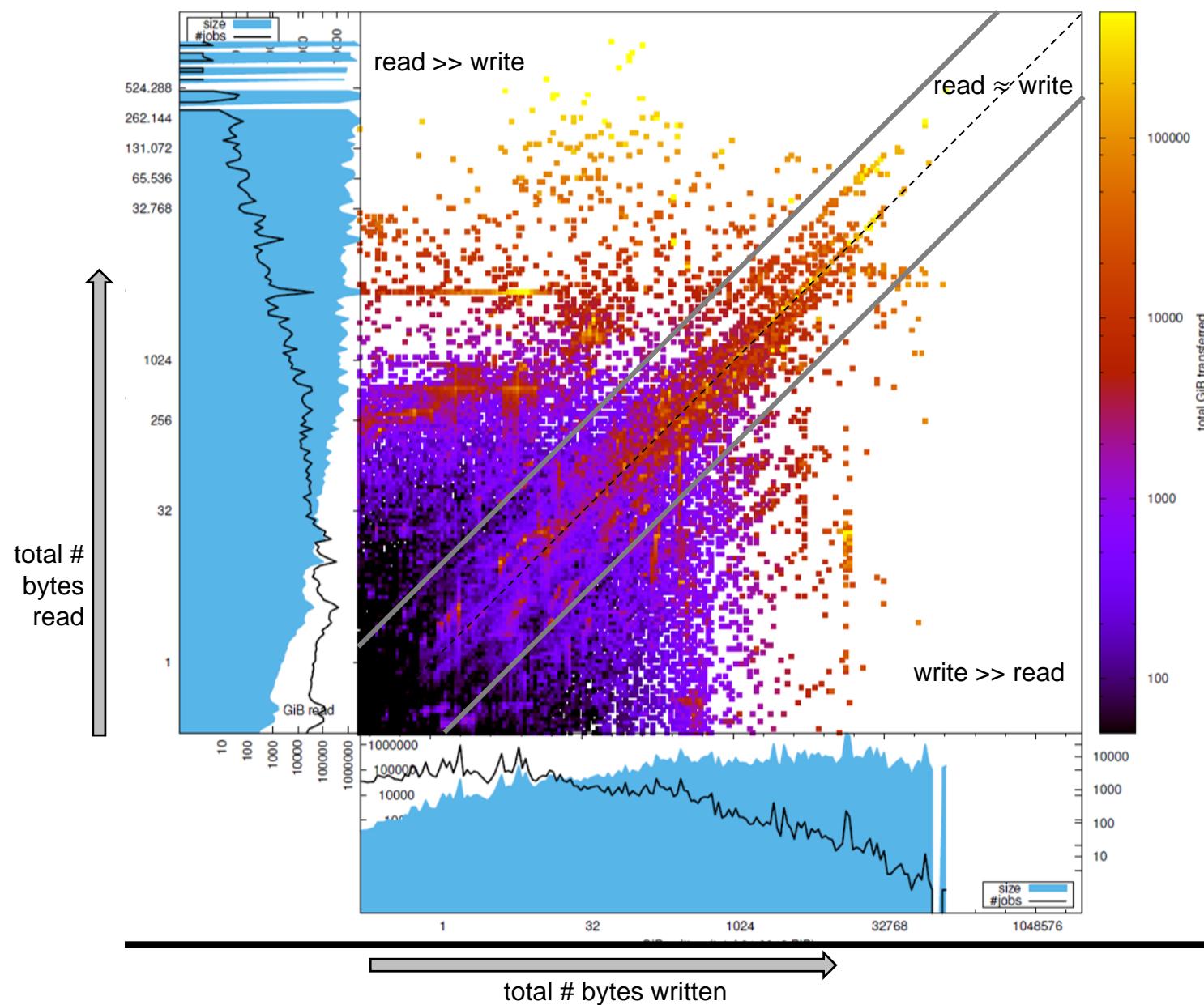
top jobs hours



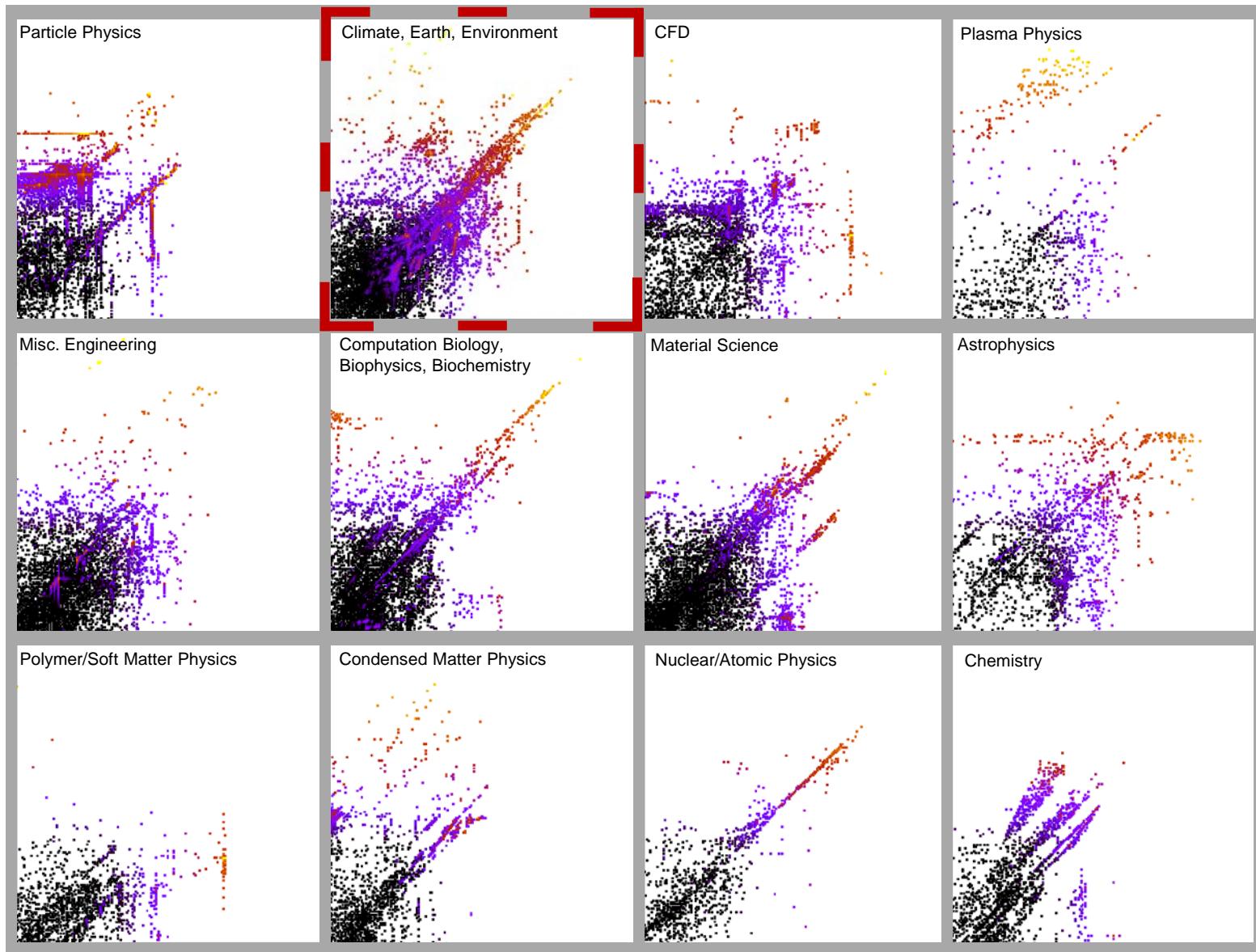
# GPFS mmpmon Data Workflow in LLview



# I/O Workload: Jureca

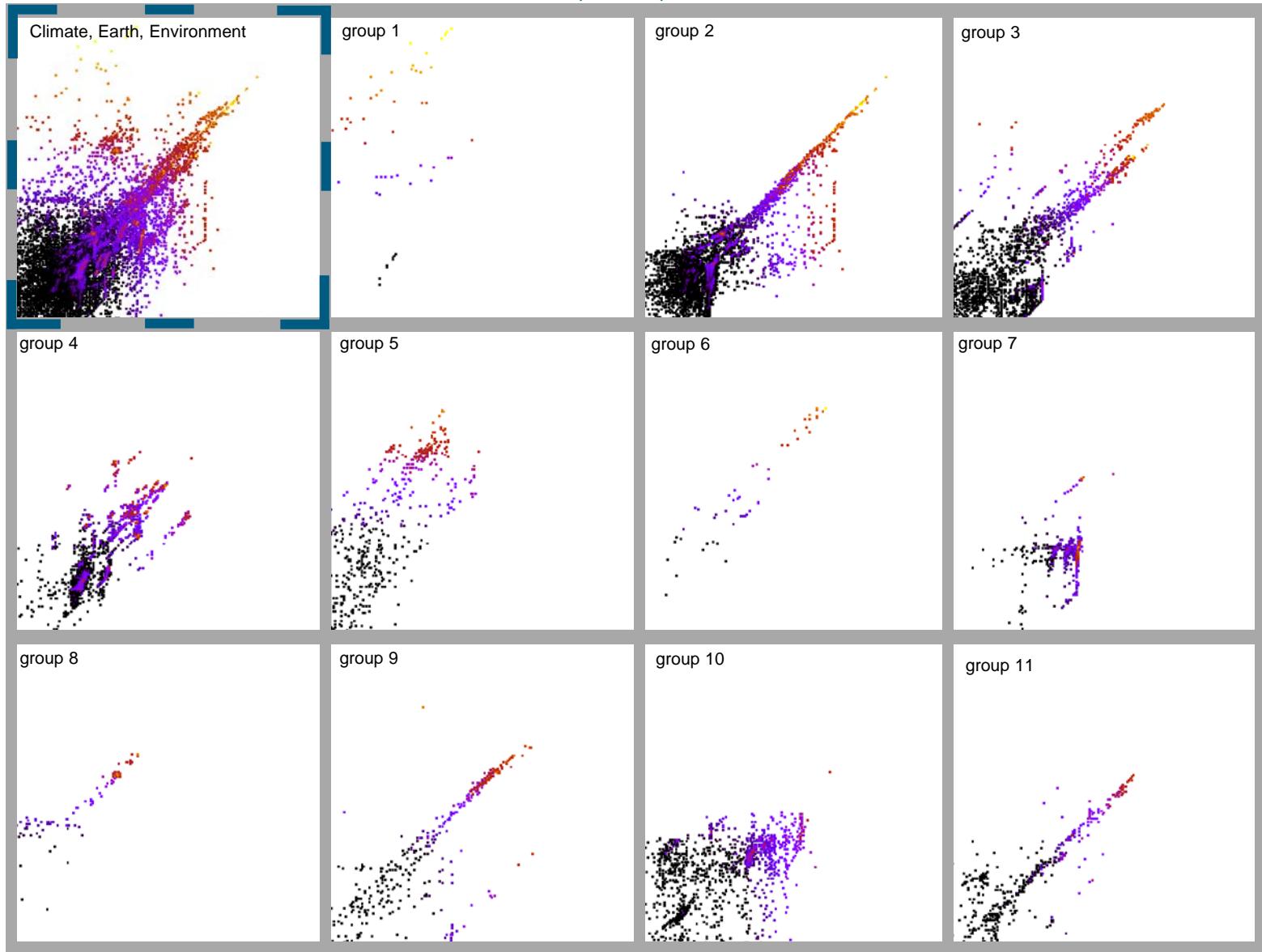


# I/O Workload: Jureca by Topic (WORK)



# I/O Workload: Jureca by Group

Climate, Earth, Environment



# Conclusion & Outlook

- I/O Workload statistics and analysis on different levels
  - Concepts for analysis of job-based mmpmon data;  
On-line and off-line analysis
  - Classification (e.g. Research topic, Group, Account)
- User Access to I/O statistics
  - I/O Reports access within user portal
  - Automatic job reports after job end
- Next steps
  - Automatic Identification of *non-well behaving* jobs
  - Additional (derived) metrics
    - *Temporal locality: Continuous, Burst, ...*
    - *Spatial locality: Parallel I/O schemes*
  - Web interface: statistics, interactive graphs