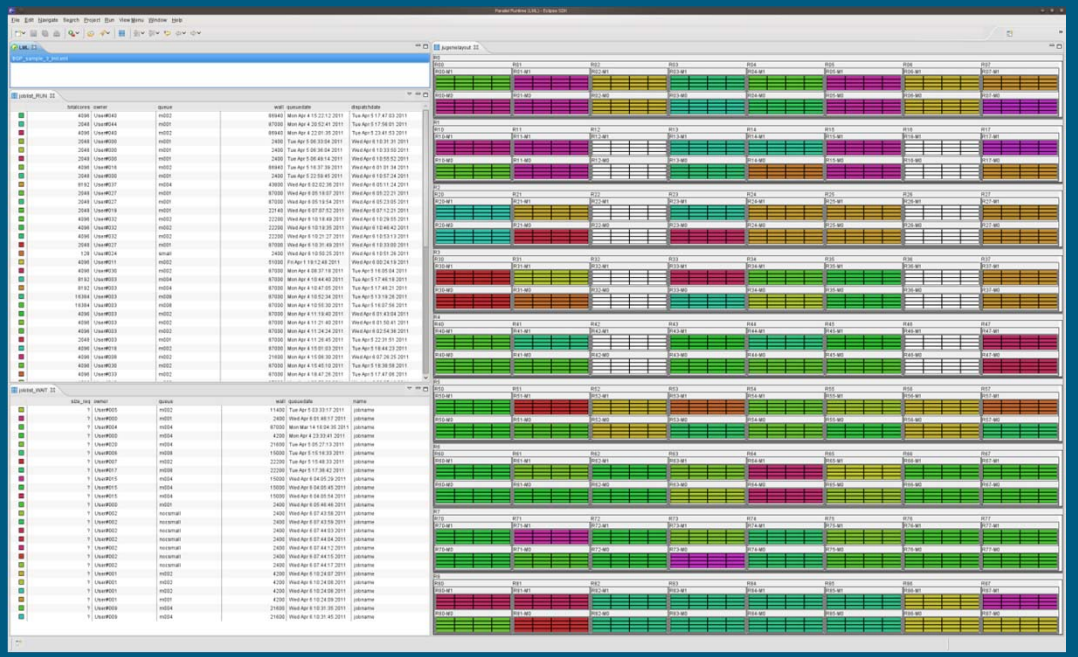
Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

eclipse
PTP

Scalable System Monitoring

with Eclipse Parallel Tools Platform



Wolfgang Frings
Jülich Supercomputing Centre

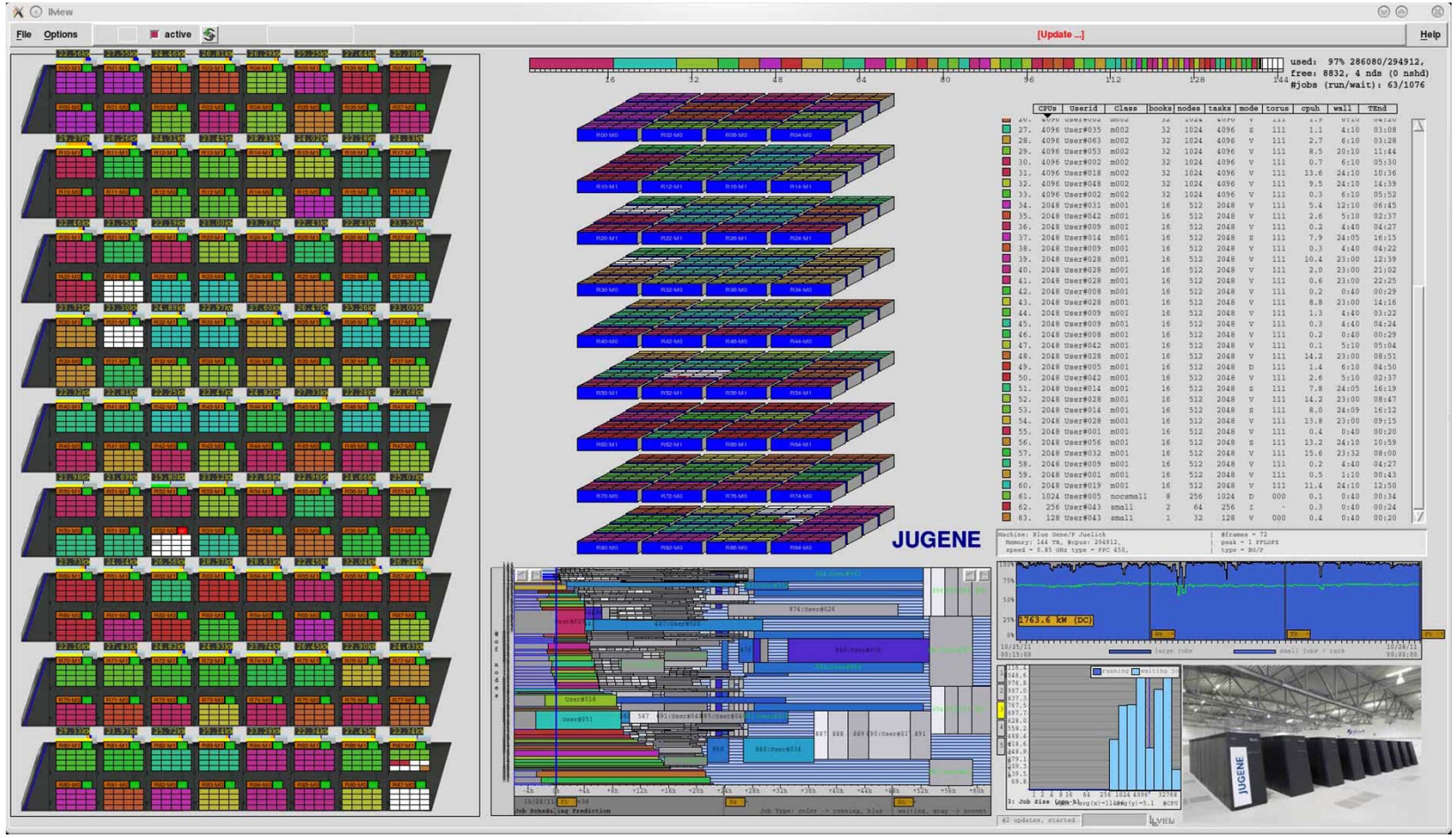
September 2012, CHANGES Workshop

W.Frings@fz-juelich.de

Why System Monitoring in an IDE?

- Eclipse/PTP: Development, execution and debugging of parallel programs on remote systems
- Information about remote system status:
  - Where is my job running?
  - What's going on remote system?
  - Why is my job not running?
- Monitoring parallel programs on remote system:
  - Eclipse/PTP version of llq / qstat
    - → LLview batch system monitoring tool
- LLview: graphical monitoring, mapping of jobs to system resources

Sample: llview client, JUGENE



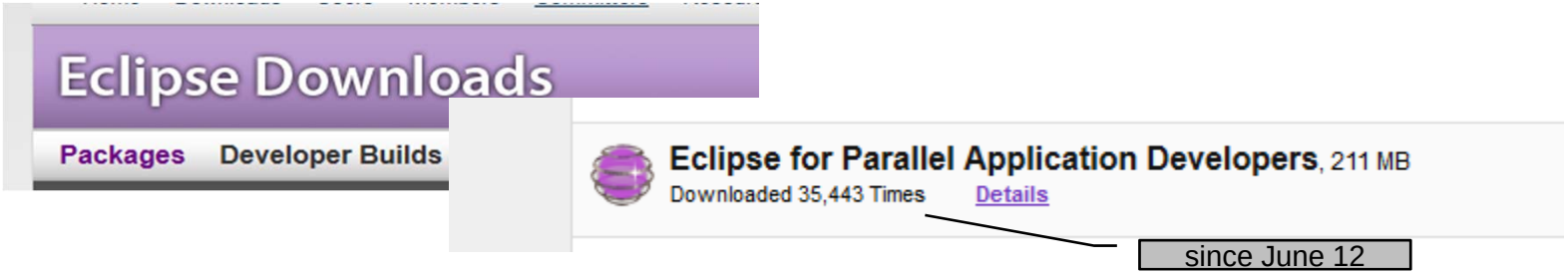
LLview and Eclipse PTP

PTP: Eclipse Parallel Tools Platform

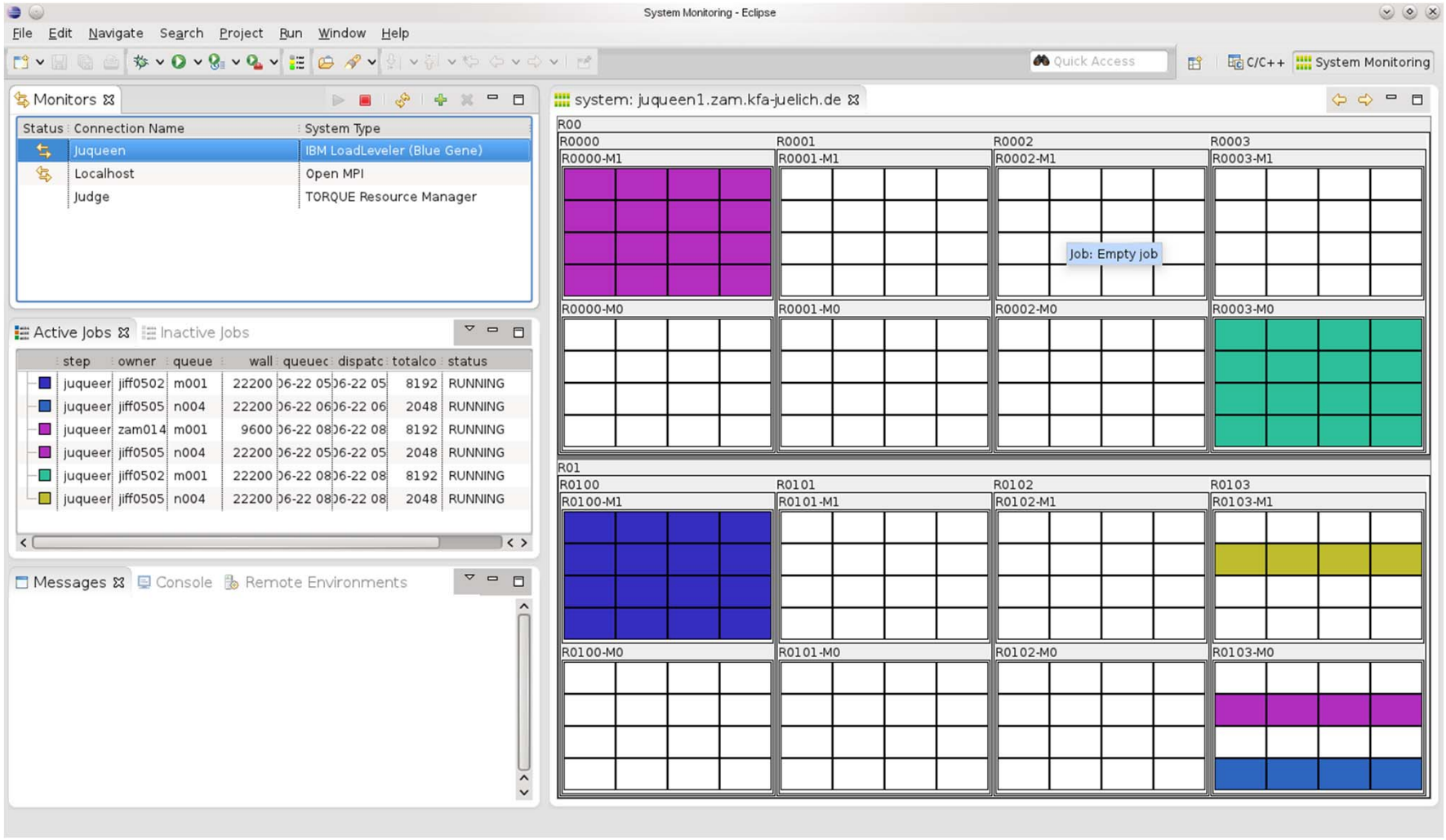
- Remote system monitoring
- New implementation using LML, LLview components and LML adapters

Project „A Scalable Development Environment for Peta-Scale Computing“

- 3-years project, started September 2009, DOE funded
- JSC contribution:
  - *System and application monitoring*
  - *LLview components integration, part of Eclipse/PTP (since June 2011)*



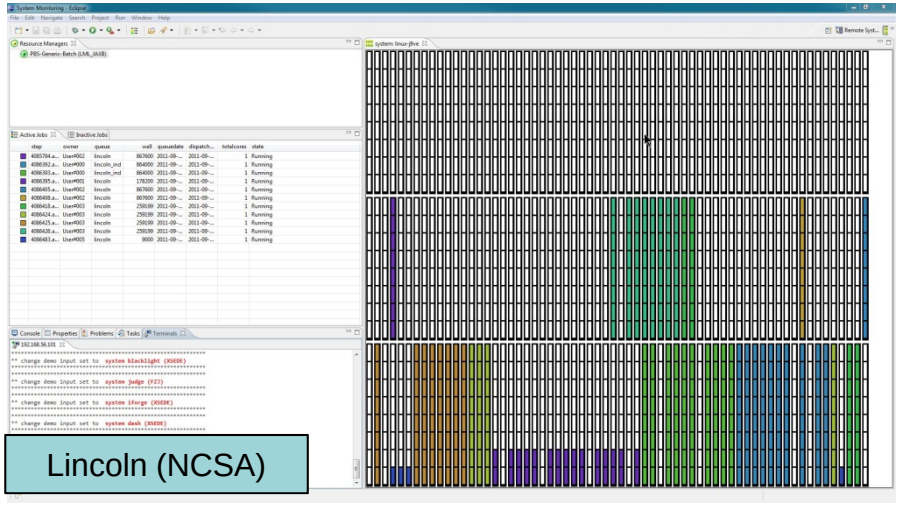
PTP Monitoring: Example JUQUEEN



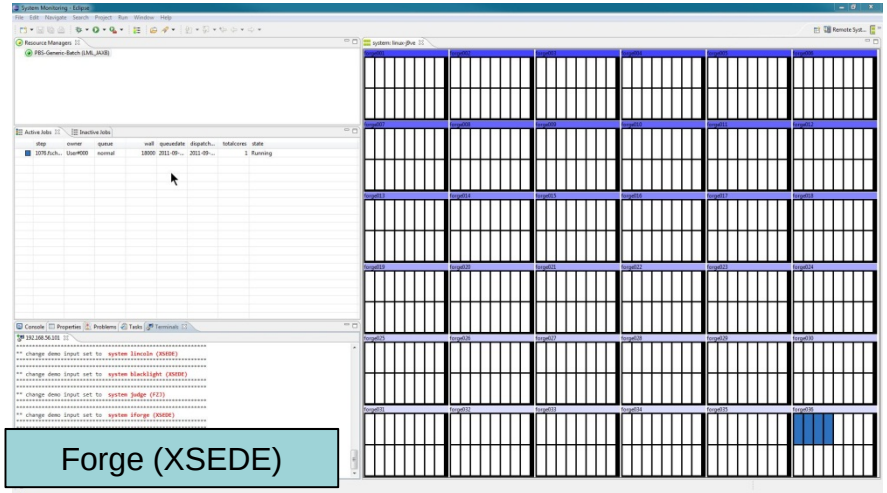
PTP Monitoring: Further Examples

Trestles (XSEDE)

Juropa (Juelich)

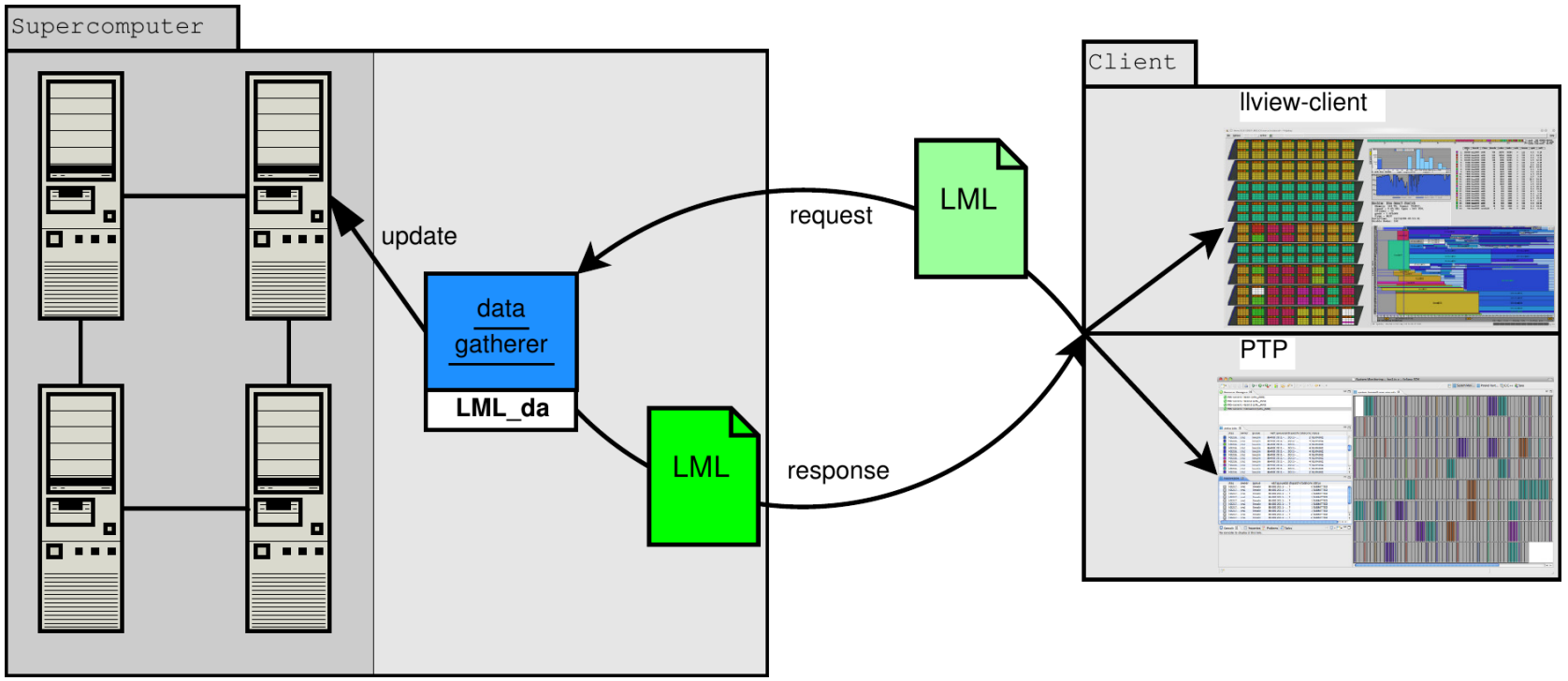


Lincoln (NCSA)



Forge (XSEDE)

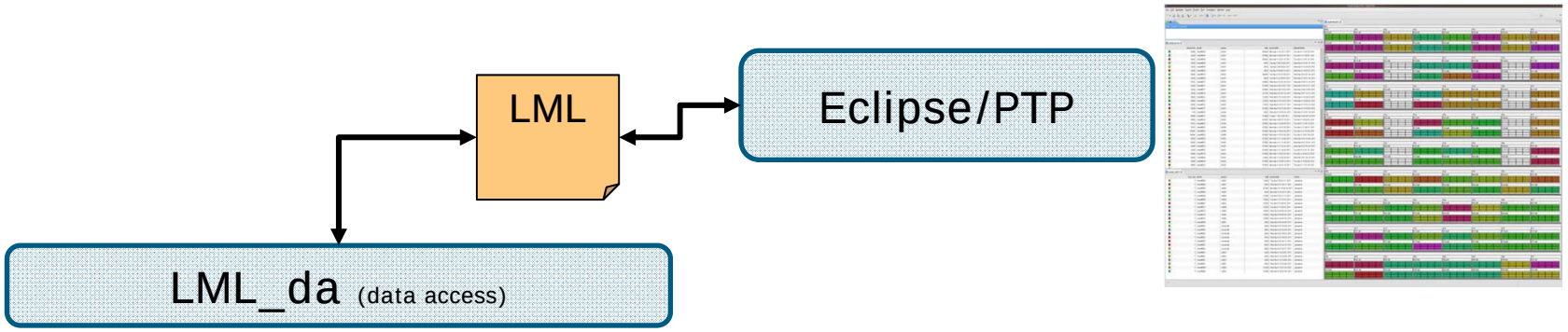
Monitoring System Dataflow



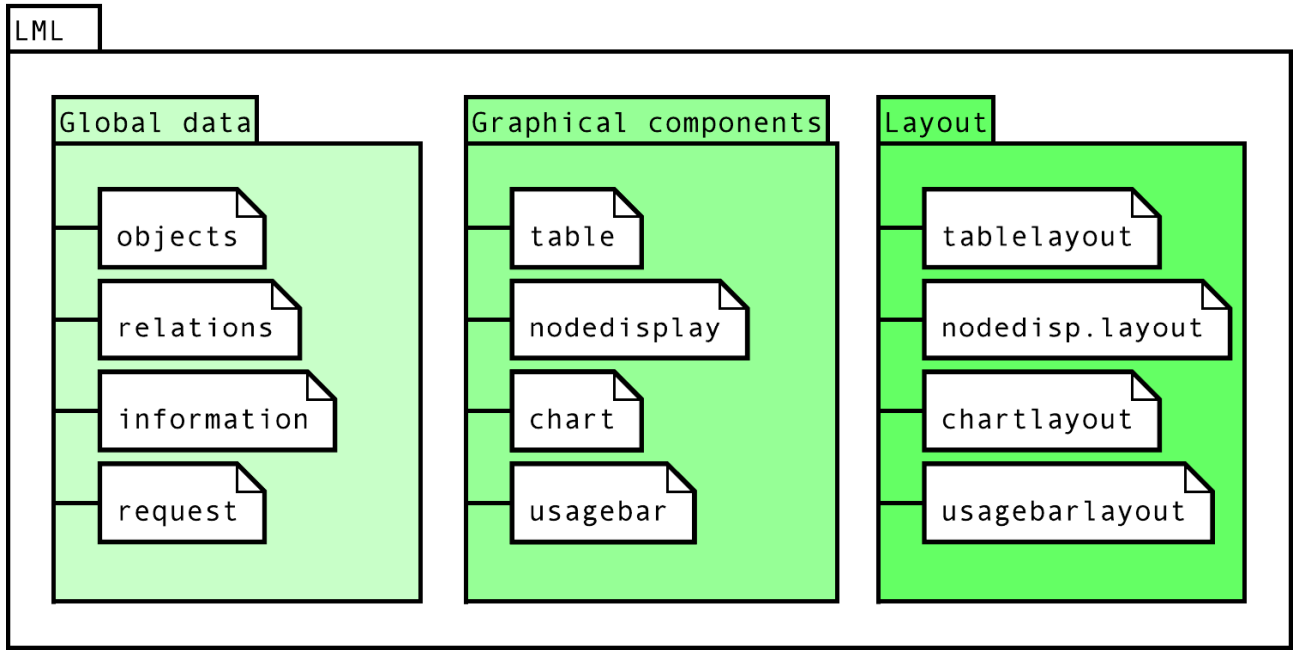
- LML_da
  - Implements distributed workflow for data acquisition and data processing
- LML-Request: describes data to visualize in client
  - Implements server-side filtering and support for level-of-details

Large-scale system Markup Language

- LML: Markup language for description of **supercomputer's status**
- Interface between LML_da and visualization clients
- Describes all graphical components available in LLview (job-list, nodedisplay, charts …)
- **Logical an system independent description** of current status, which can easily be converted into graphical output
- Implemented in XML, validation against **XML-Schema**

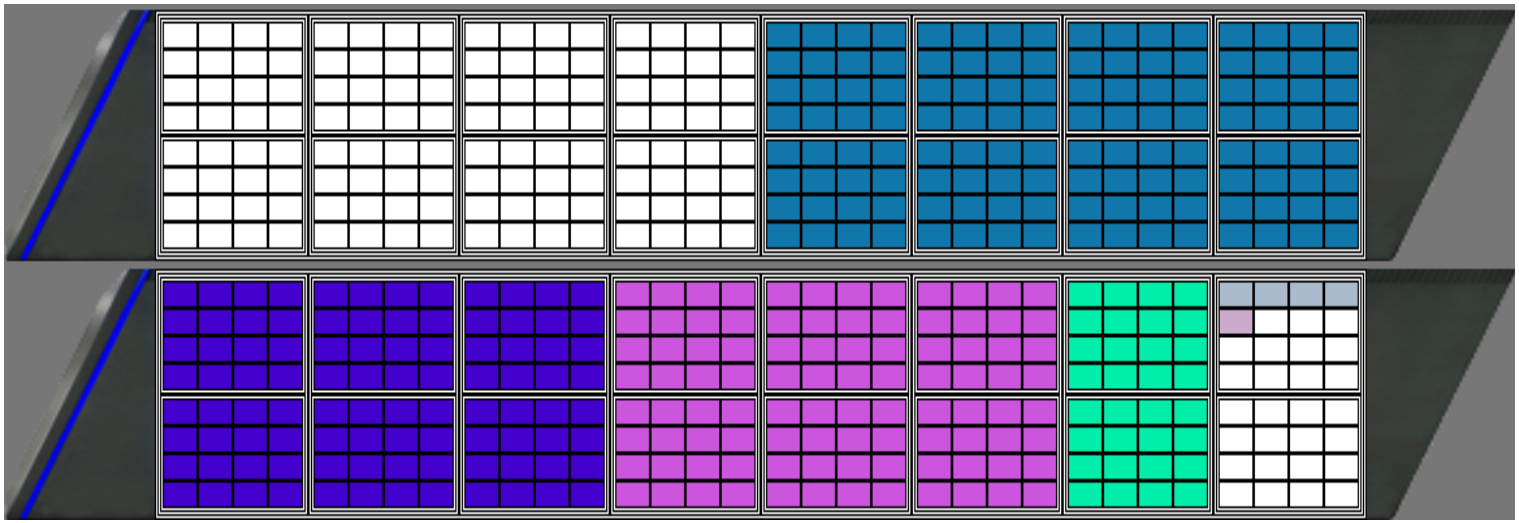


LML Structure



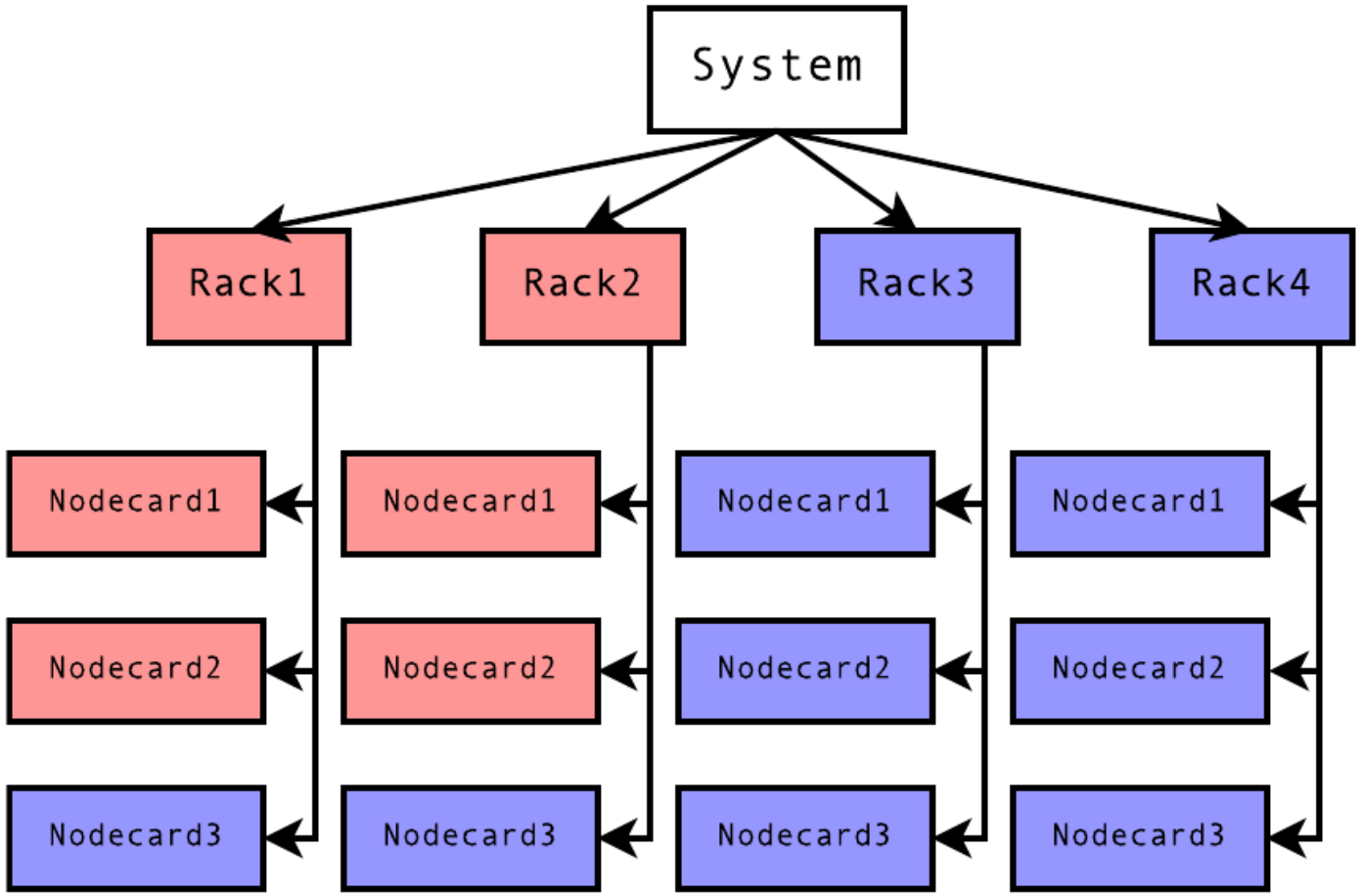
- **Global data**: intermediate data format, scheduling objects
- **Graphical components**: data for visualization components
- **Layout**: hints for visualization

Node Display



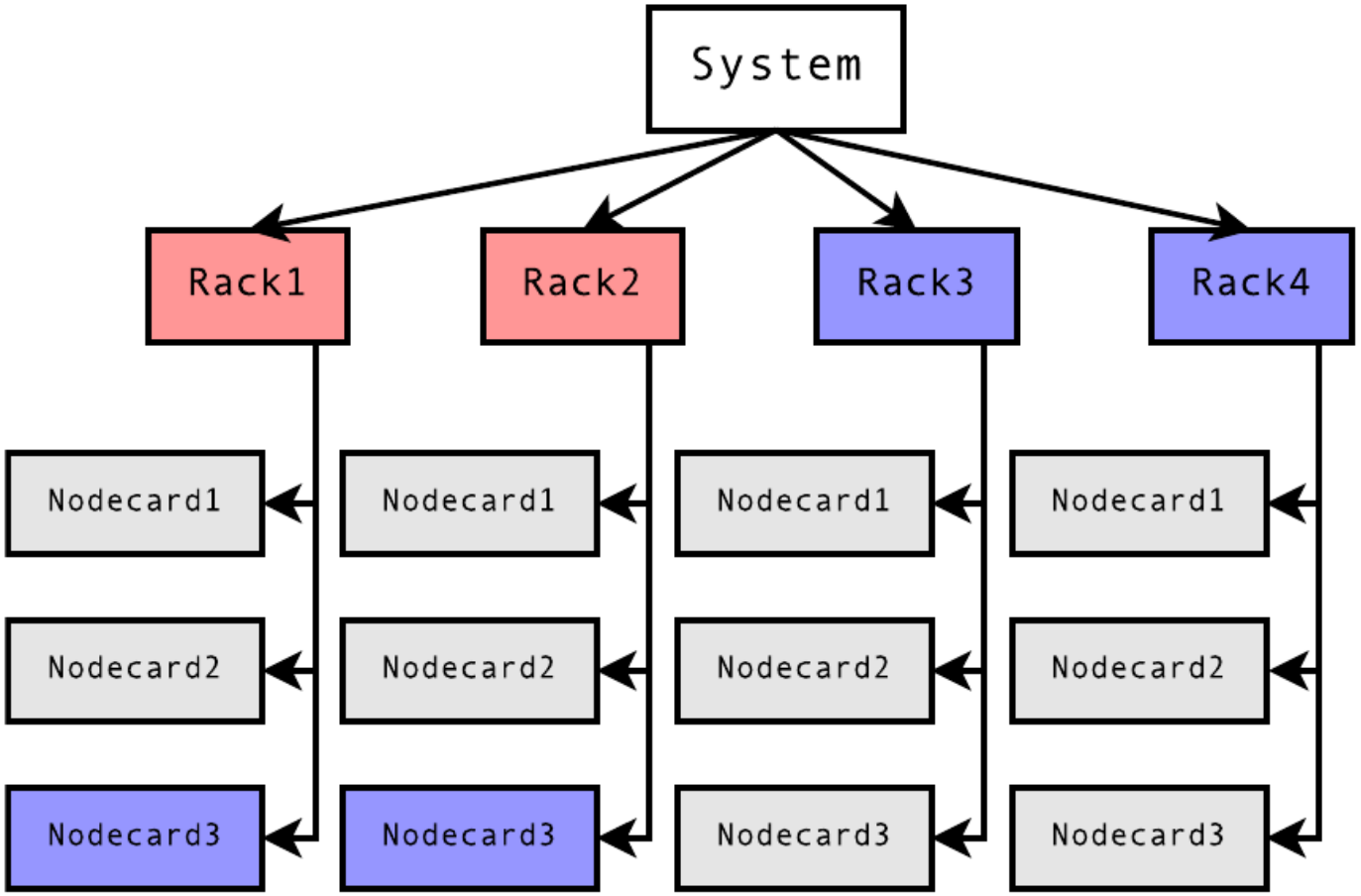
- Description of the supercomputer's architecture
- Maps jobs to compute resources (nodes, CPUs)
- **Challenge**: description of large systems
  (e.g. JUQUEEN: 131072 CPUs)
- **Targets**: data compression, avoid redundancy

Node Display – Compression

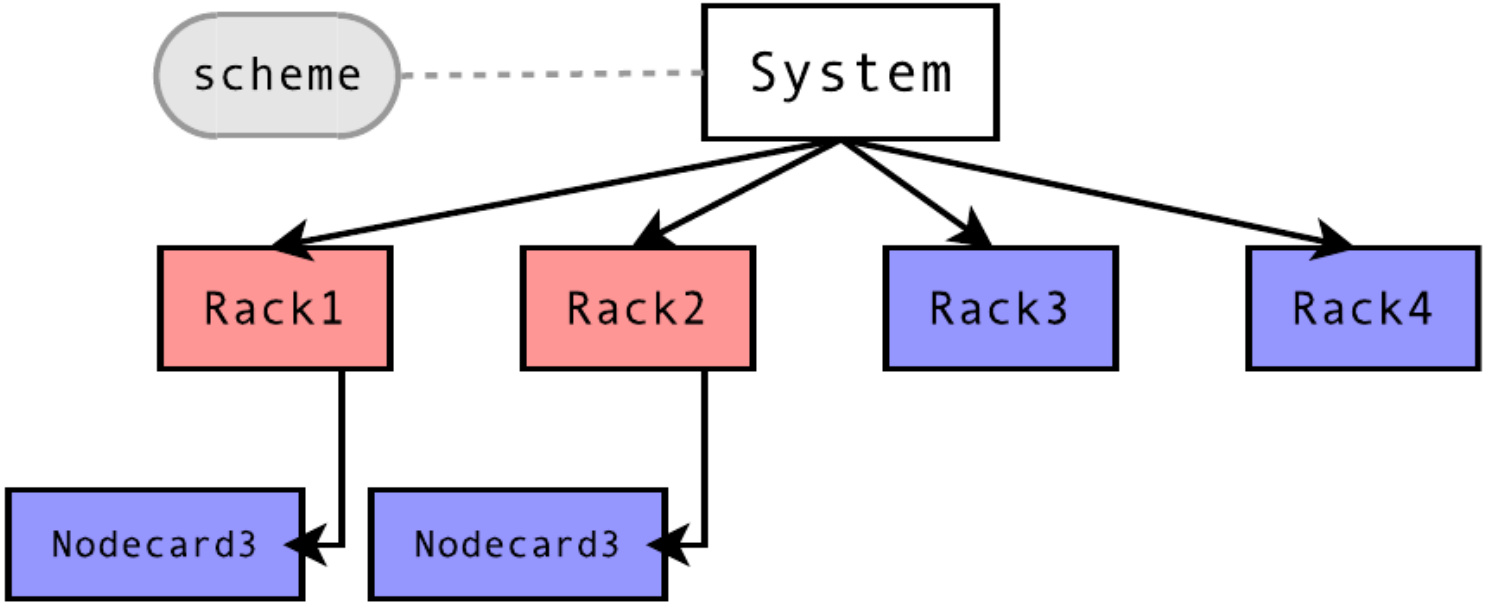


Node Display – Compression

Attribute inheritance

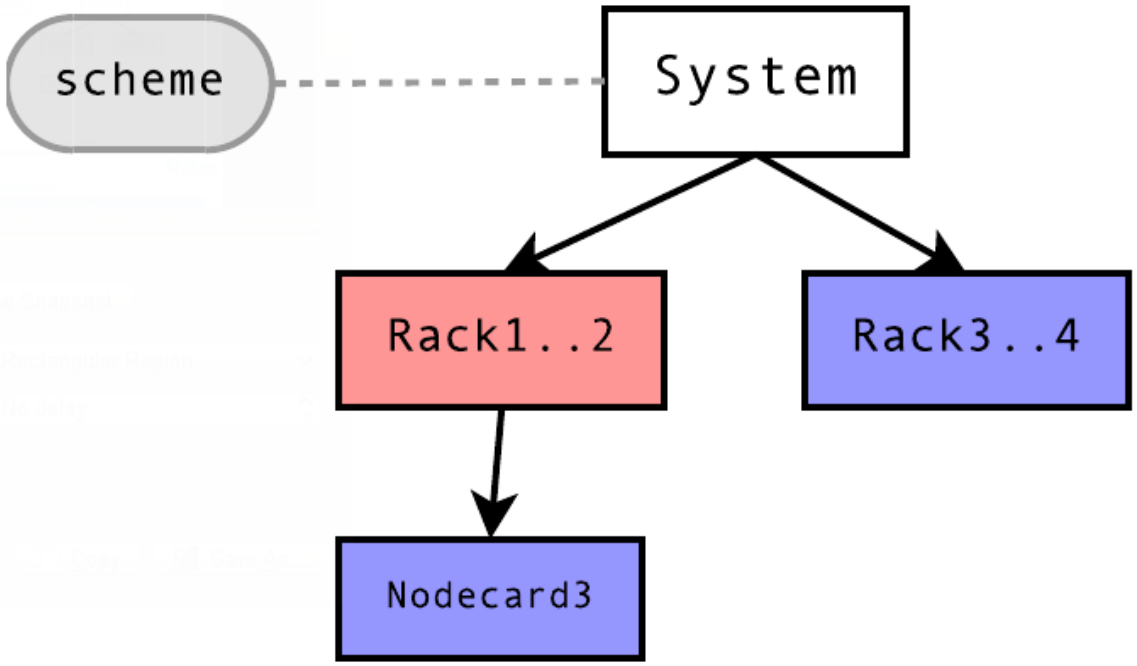


Node Display – Compression



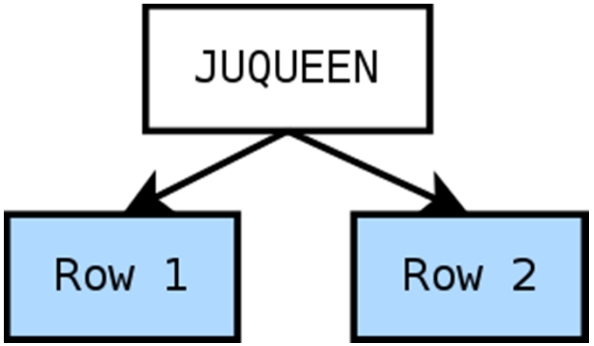
- **Scheme**: defines architecture of empty system

Node Display – Compression
Ranges

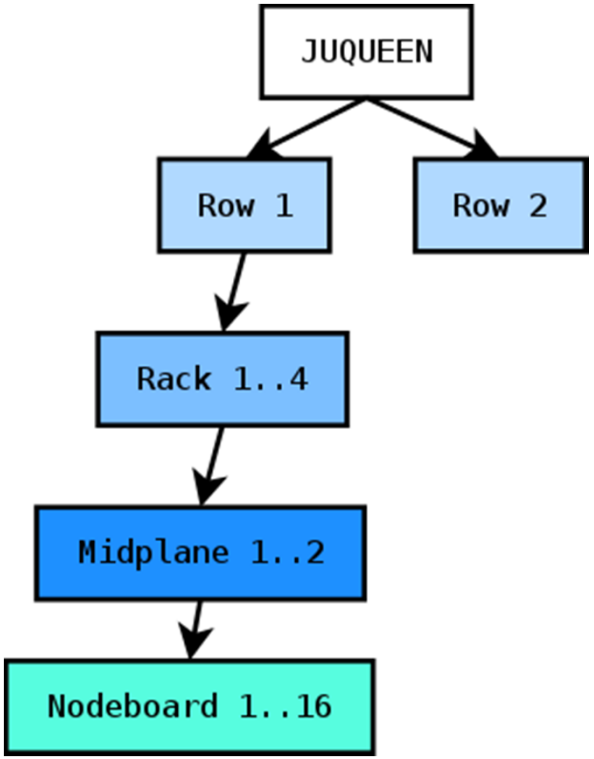


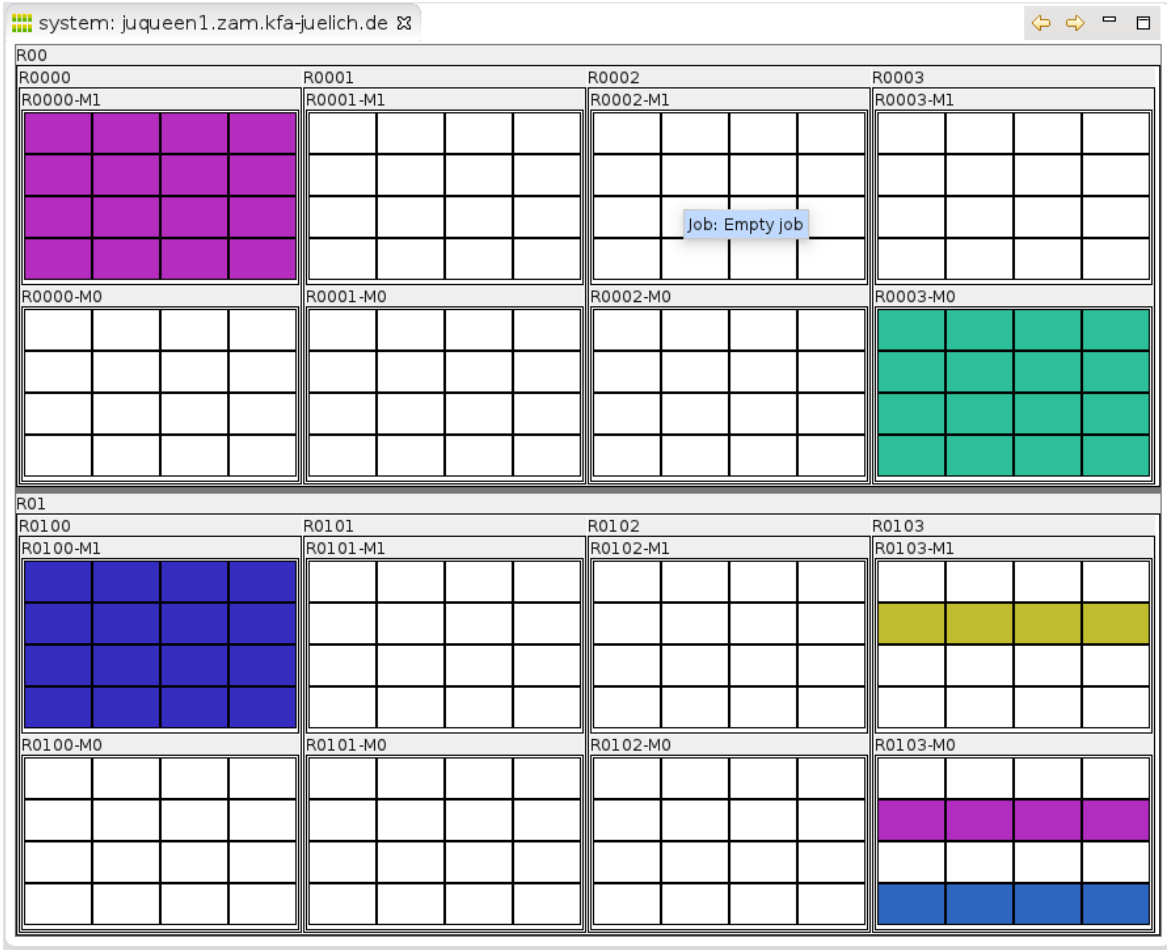
→ 3 objects instead of 16

Scalable Visualization – Row level

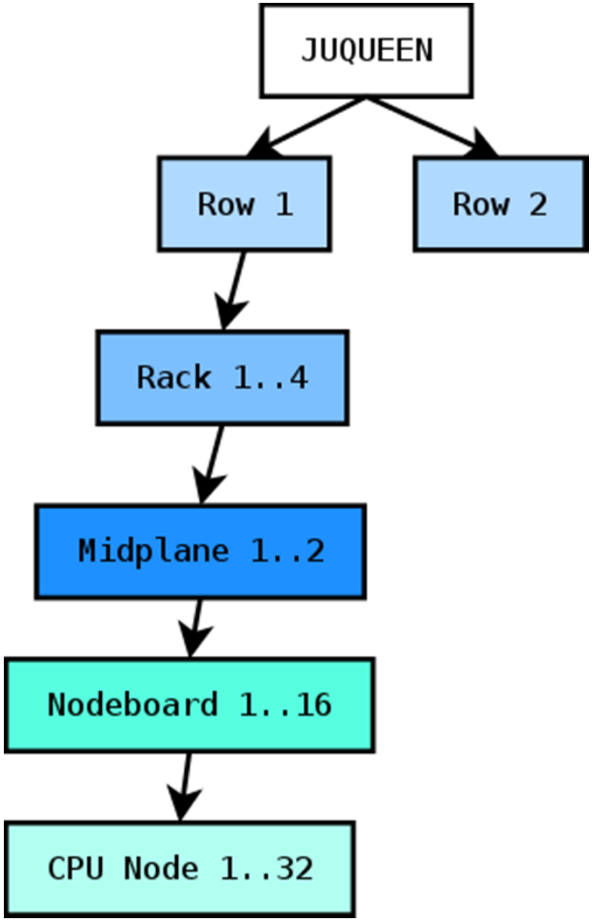


Scalable Visualization – Nodeboard level





Scalable Visualization – CPU Node level



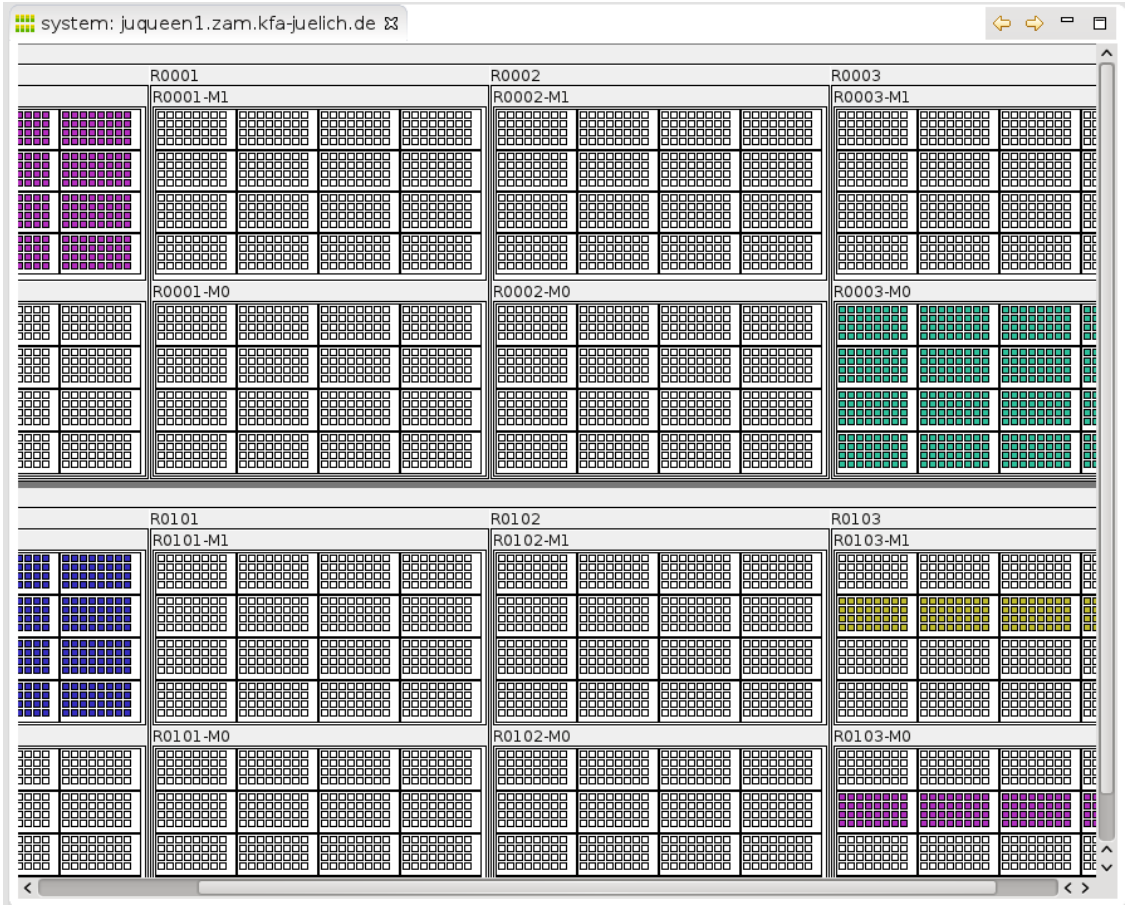
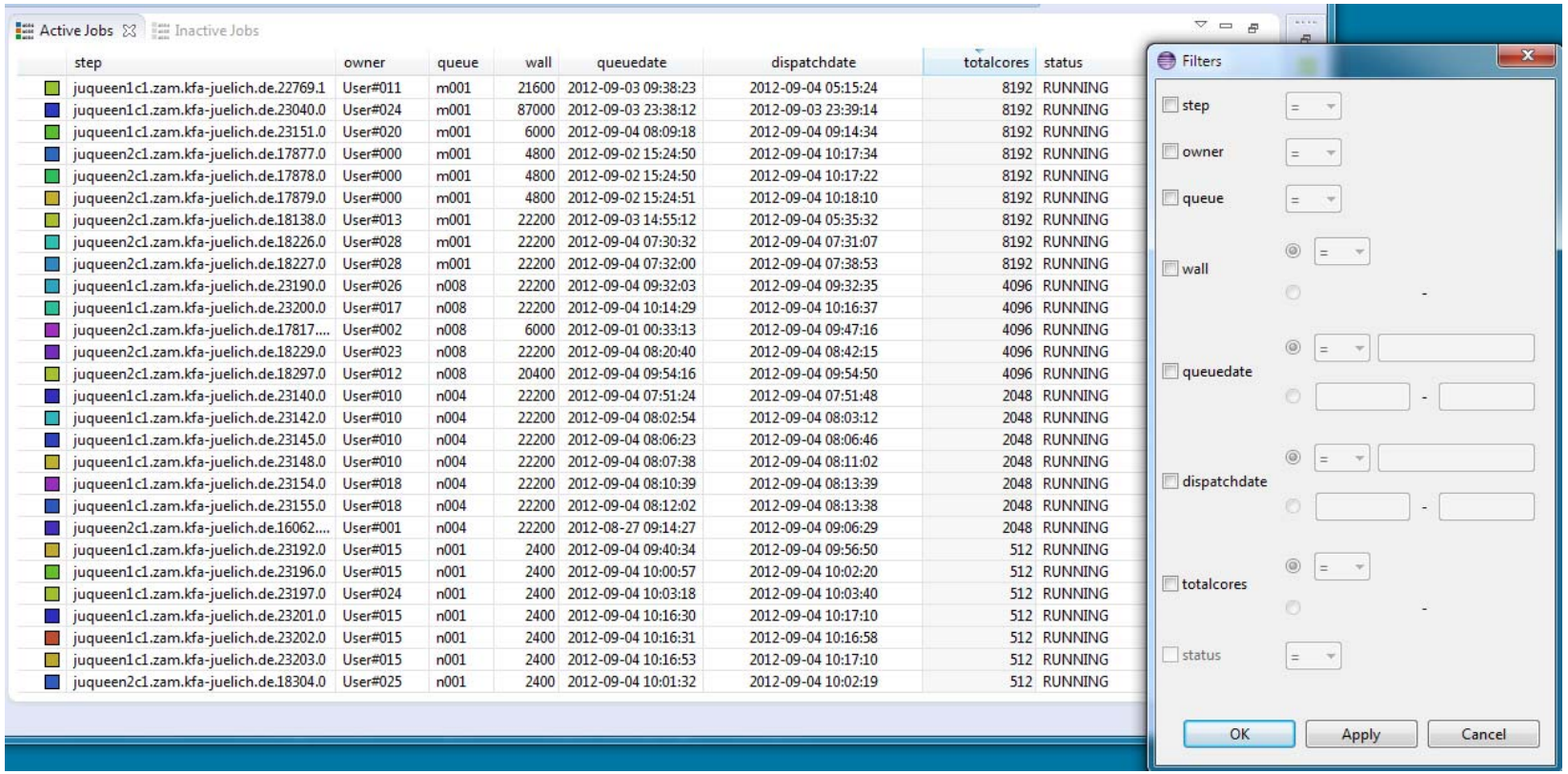


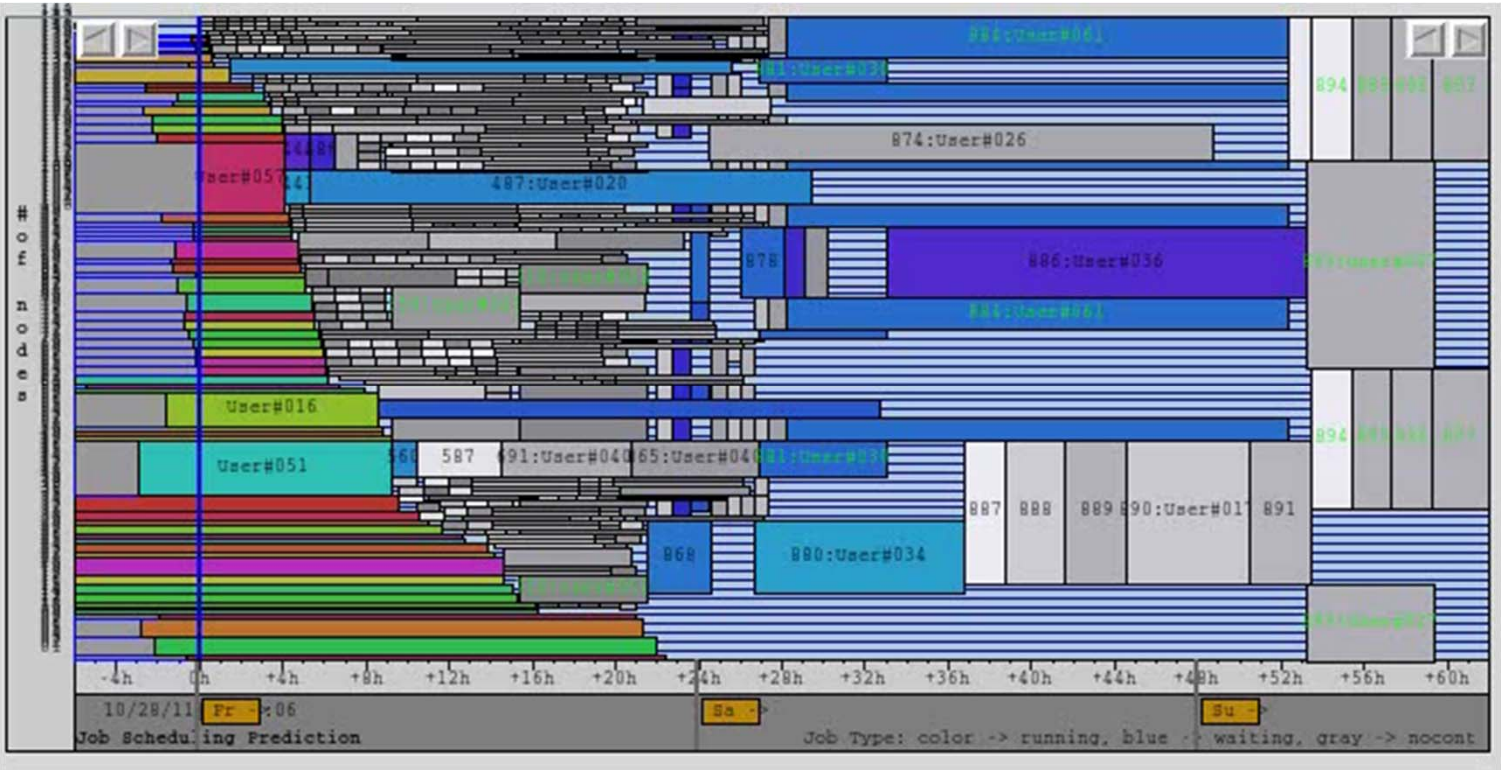
Table Filtering

- Scalability of job data display
- Filtering by specification of attributes values/ranges
- Server- and client-side filtering



Predicition of system usage (JuFo)

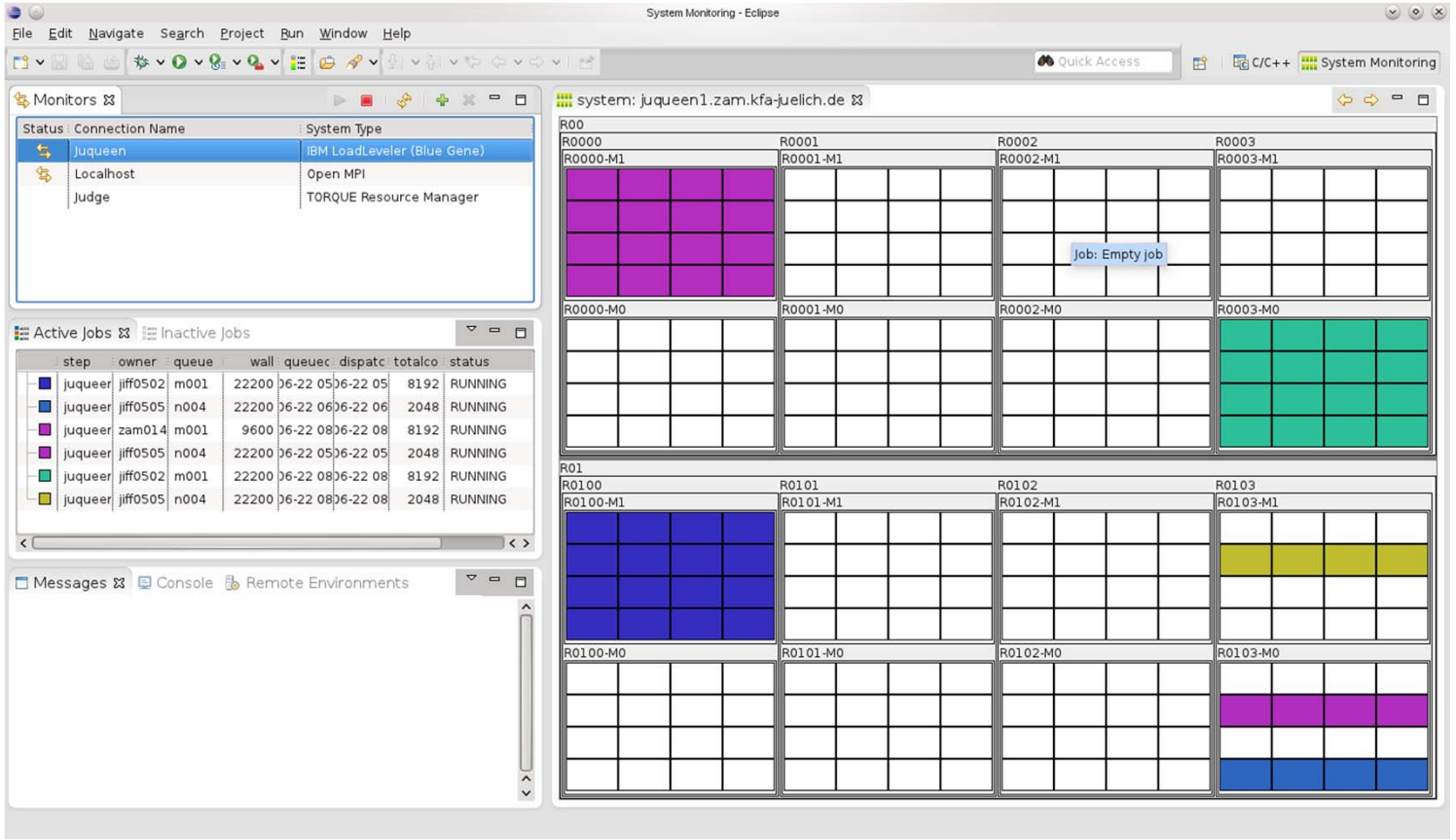
- Online simulation for **job schedulers** on supercomputers (JuFo: Juelich Forecast)
- Independent module using **intermediate LML** as data interface
- Target system independent prediction of job start dates
- Visualisation: Gantt chart



Conclusion

- Eclipse PTP:
  - Integration tool to develop parallel codes on remote system
    - *Code* + *Execution* + *Debugging*
  - System monitoring to control program execution
- Scalability
  - Scalable data acquisition
  - Scalable data format
  - Scalable data presentation
- Flexibility and portability
  - System independent data format
  - Small system and scheduler related drivers
  - Eclipse/PTP integration

Questions?


System Monitoring - Eclipse
File Edit Navigate Search Project Run Window Help
Quick Access
C/C++
System Monitoring
Monitors
Status Connection Name System Type
Juqueen IBM LoadLeveler (Blue Gene)
Localhost Open MPI
Judge TORQUE Resource Manager
Active Jobs
Inactive Jobs
step owner queue wall queuec dispatc totalco status
juqueer jiff0502 m001 22200 06-22 05 06-22 05 8192 RUNNING
juqueer jiff0505 n004 22200 06-22 06 06-22 06 2048 RUNNING
juqueer zam014 m001 9600 06-22 08 06-22 08 8192 RUNNING
juqueer jiff0505 n004 22200 06-22 05 06-22 05 2048 RUNNING
juqueer jiff0502 m001 22200 06-22 08 06-22 08 8192 RUNNING
juqueer jiff0505 n004 22200 06-22 08 06-22 08 2048 RUNNING
Messages
Console
Remote Environments
system: juqueen1.zam.kfa-juelich.de
R00
R0000
R0001
R0002
R0003
R0000-M1
R0001-M1
R0002-M1
R0003-M1
Job: Empty job
R0000-M0
R0001-M0
R0002-M0
R0003-M0
R01
R0100
R0101
R0102
R0103
R0100-M1
R0101-M1
R0102-M1
R0103-M1
R0100-M0
R0101-M0
R0102-M0
R0103-M0