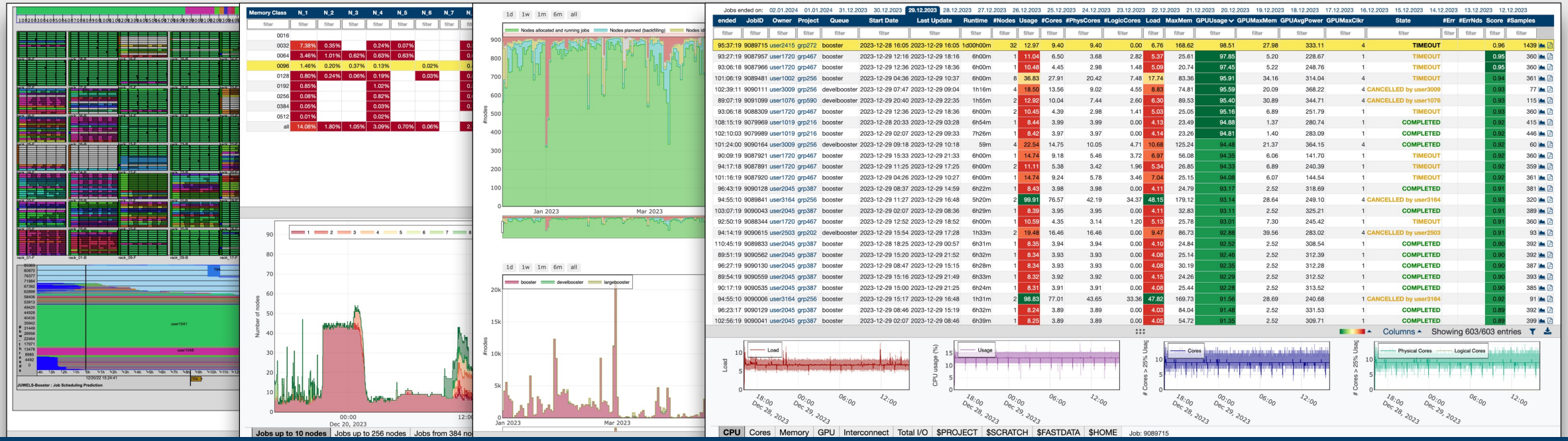


LLview: Reporting Platform for the Exascale Era

ADAC17 Symposium, Helsinki, Finland

03.09.2025 I FILIPE GUIMARÃES | HPC SUPPORT, JÜLICH SUPERCOMPUTING CENTRE

Introduction

monitor *verb*

observe and check the progress or quality of (something) over a period of time;

keep under systematic review;

maintain regular surveillance over;

Introduction

monitor *verb*

observe and check the progress or quality of (something) over a period of time;

keep under systematic review;

maintain regular surveillance over;

HPC → Slurm, Prometheus (+Grafana), Nvidia DCGM, …

Introduction

monitor *verb*

observe and check the progress or quality of (something) over a period of time;

keep under systematic review;

maintain regular surveillance over;

HPC → Slurm, Prometheus (+Grafana), Nvidia DCGM, …

Operational data analytics (ODA)

analysis of existing and real-time operation metrics of the business

focus on improving the operational nature of a business or entity

Introduction

monitor *verb*

observe and check the progress or quality of (something) over a period of time;

keep under systematic review;

maintain regular surveillance over;

HPC → Slurm, Prometheus (+Grafana), Nvidia DCGM, …

Operational data analytics (ODA)

analysis of existing and real-time operation metrics of the business

focus on improving the operational nature of a business or entity

Introduction

monitor *verb*

observe and check the progress or quality of (something) over a period of time;

keep under systematic review;

maintain regular surveillance over;

HPC → Slurm, Prometheus (+Grafana), Nvidia DCGM, …

Operational data analytics (ODA)

analysis of existing and real-time operation metrics of the business

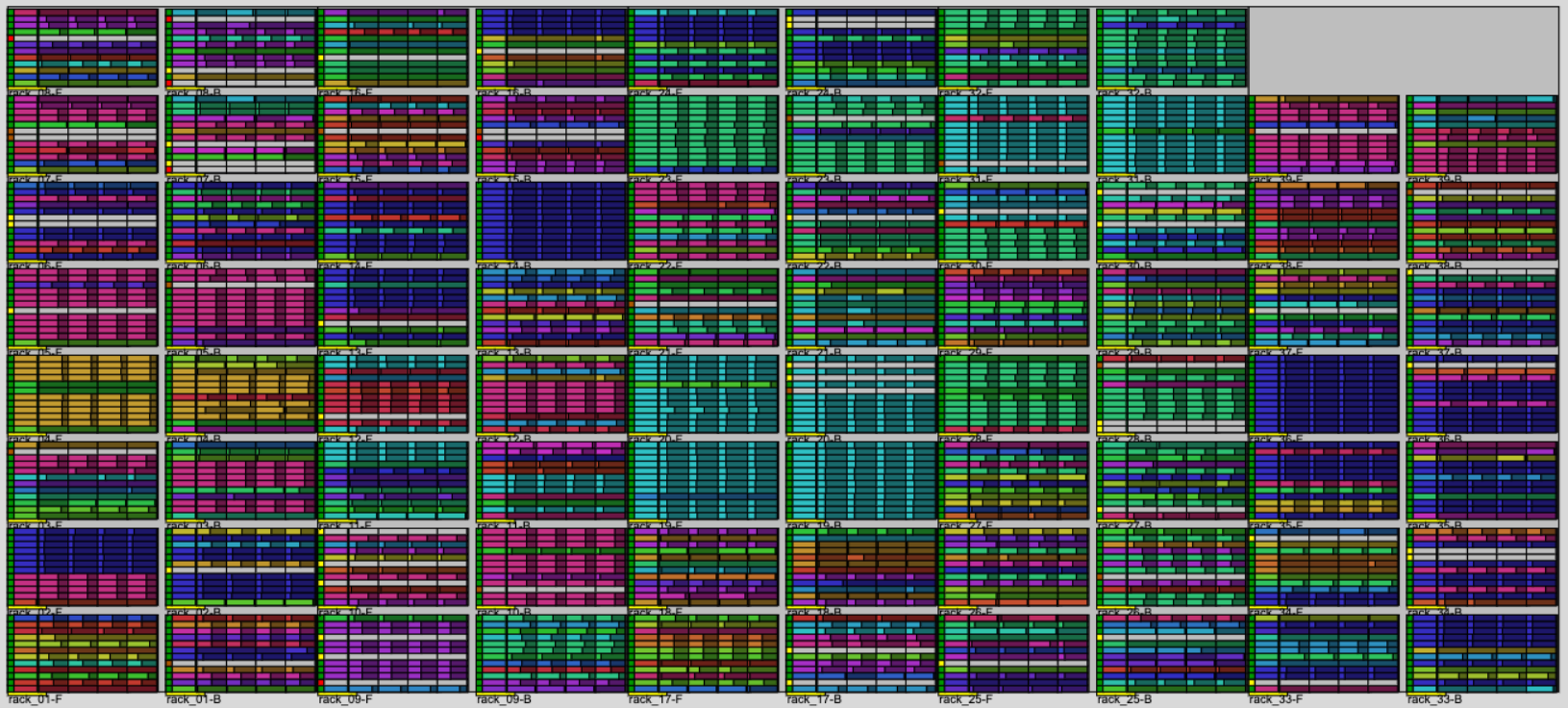
focus on improving the operational nature of a business or entity

Reporting platform

Focuses on delivering structured, contextualised, role-specific reports to the different audiences

Challenges at the exascale

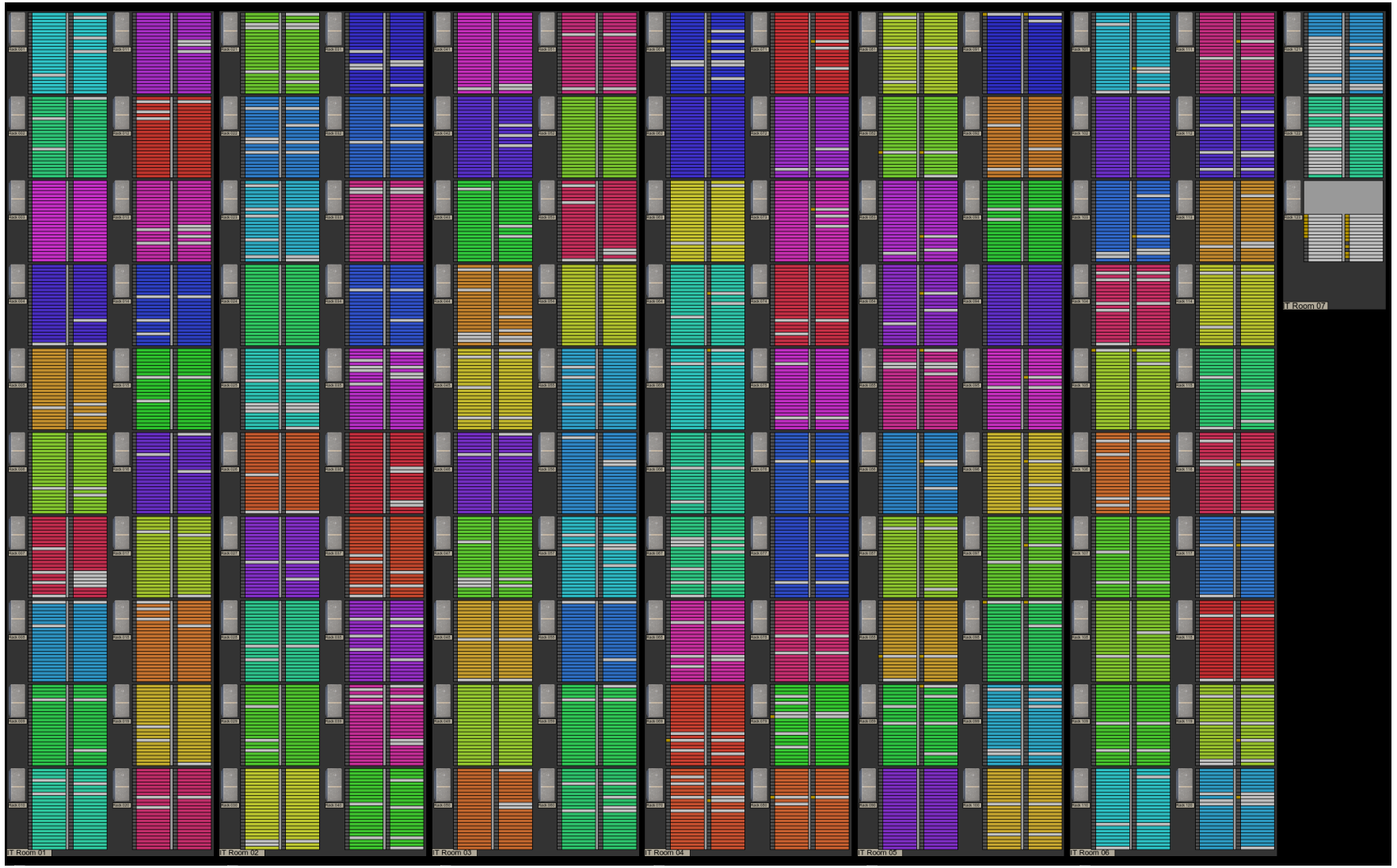
JUWELS Booster



~900 nodes

Challenges at the exascale

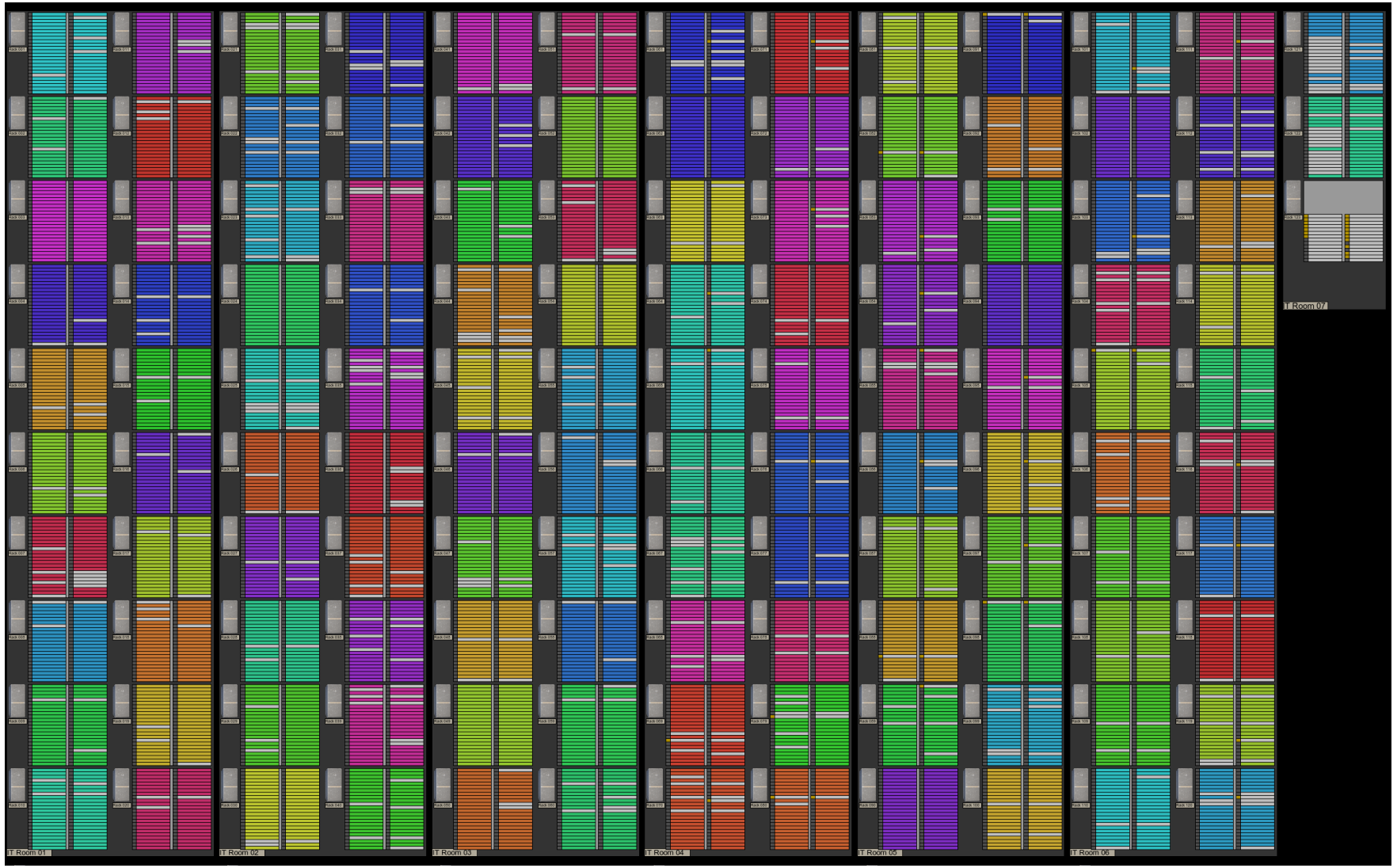
JUPITER



~6000 nodes

Challenges at the exascale

JUPITER



~6000 nodes

Lots of data!

- ➡ **Longer to process**
- ➡ **Heavier reports**
- ➡ **Difficult to present**

Why do we need a reporting platform?

Why do we need a reporting platform?

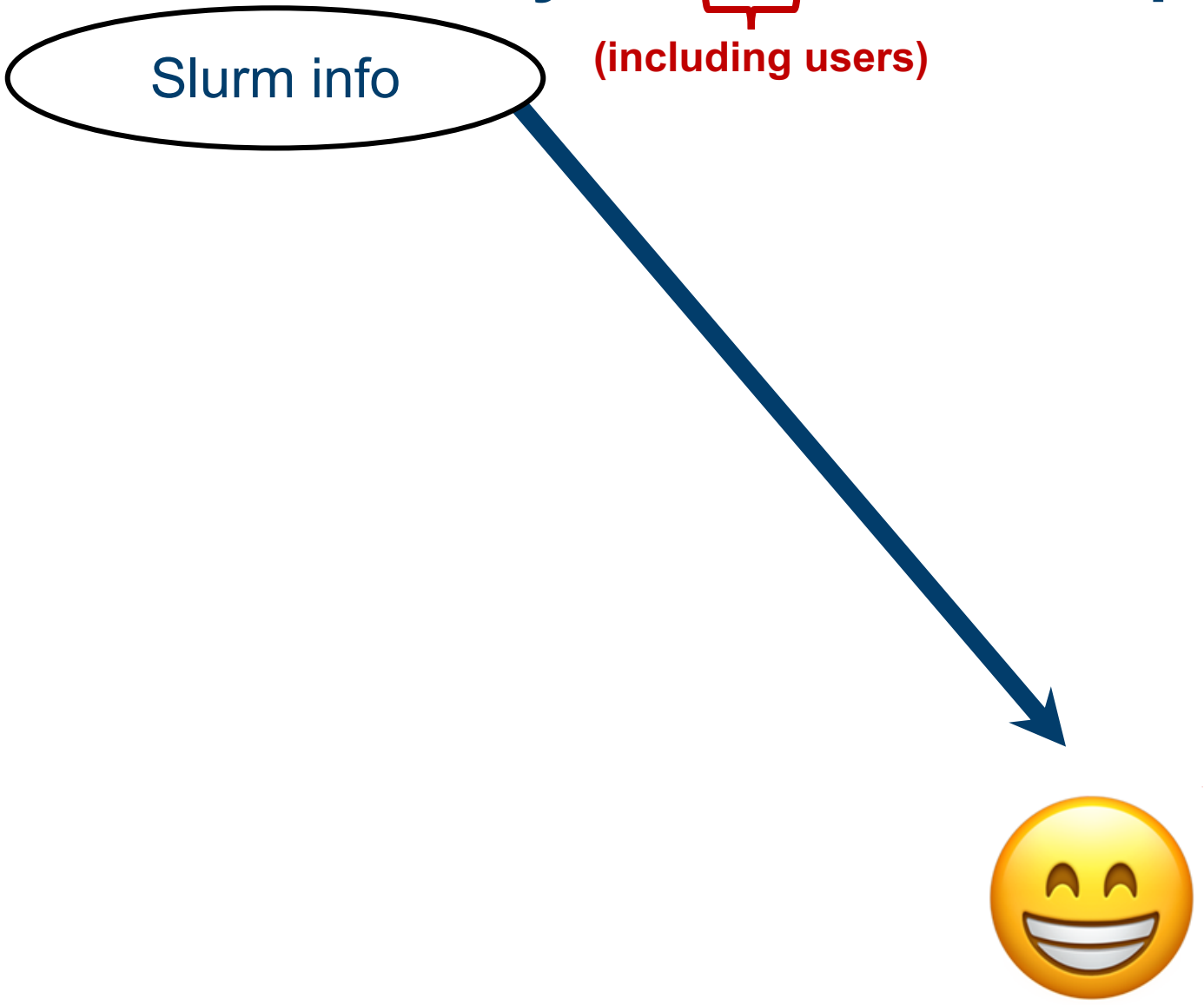
(including users)

Why do we need a reporting platform?

(including users)

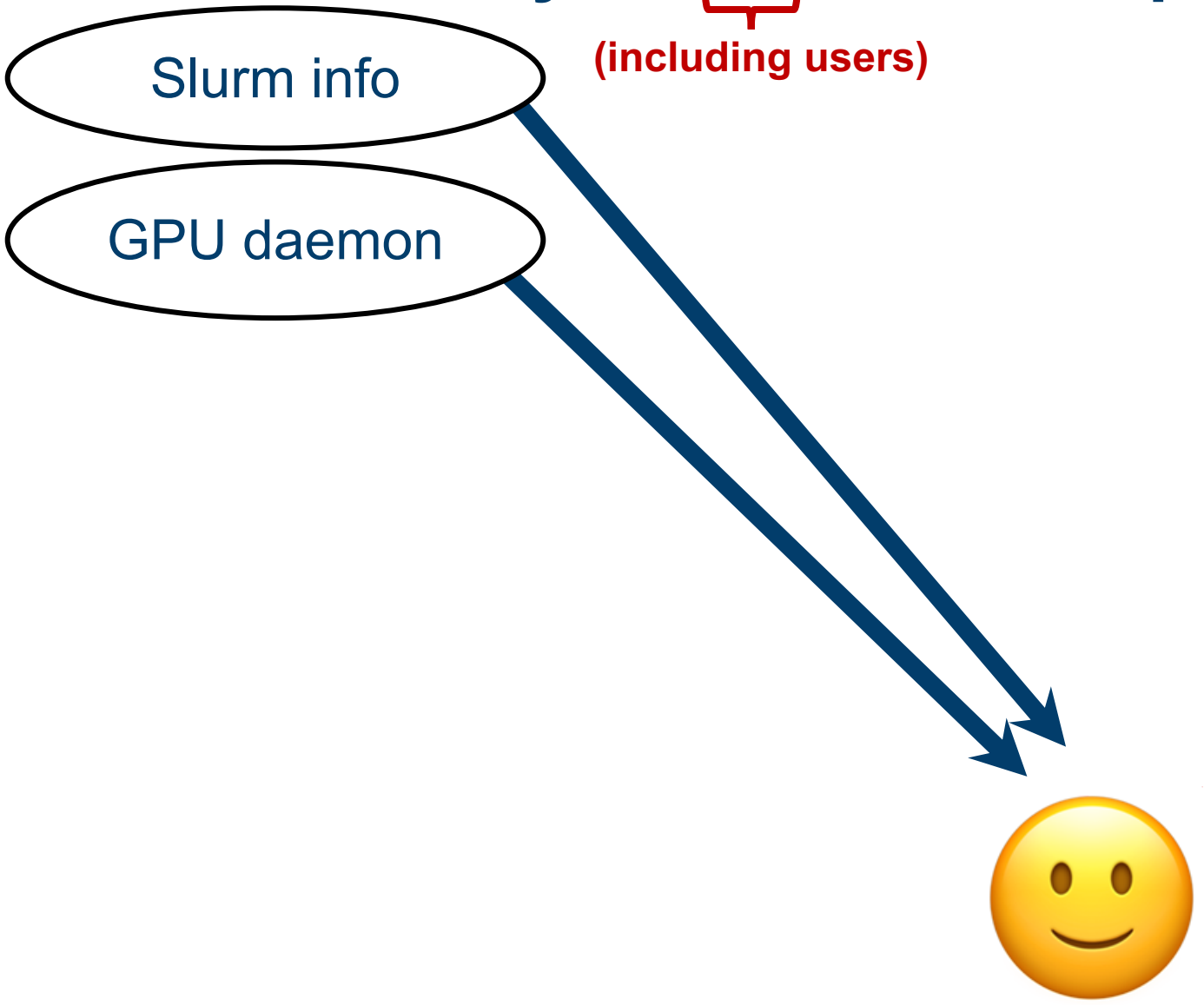
Why do we need a reporting platform?

(including users)



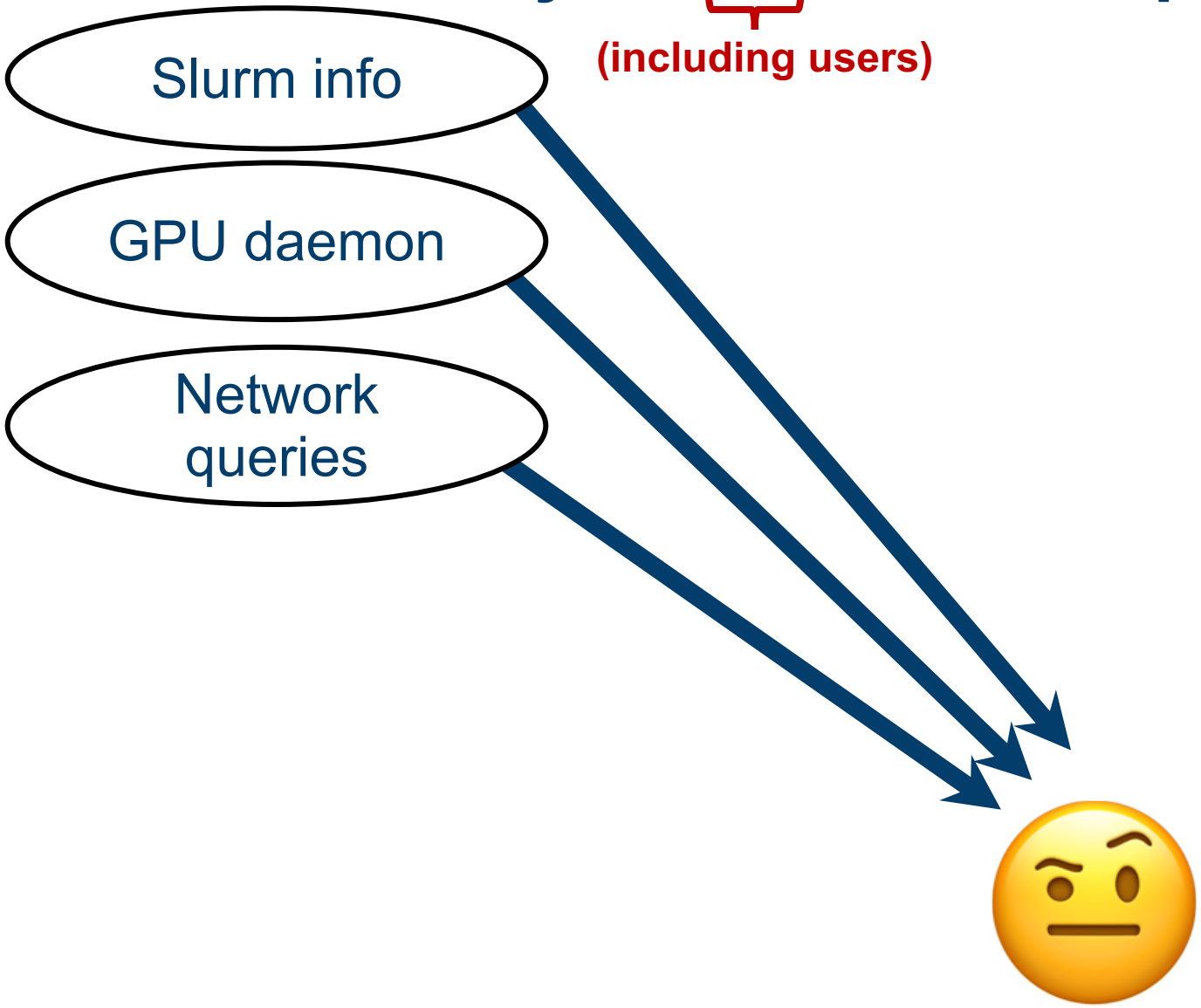
Why do we need a reporting platform?

(including users)



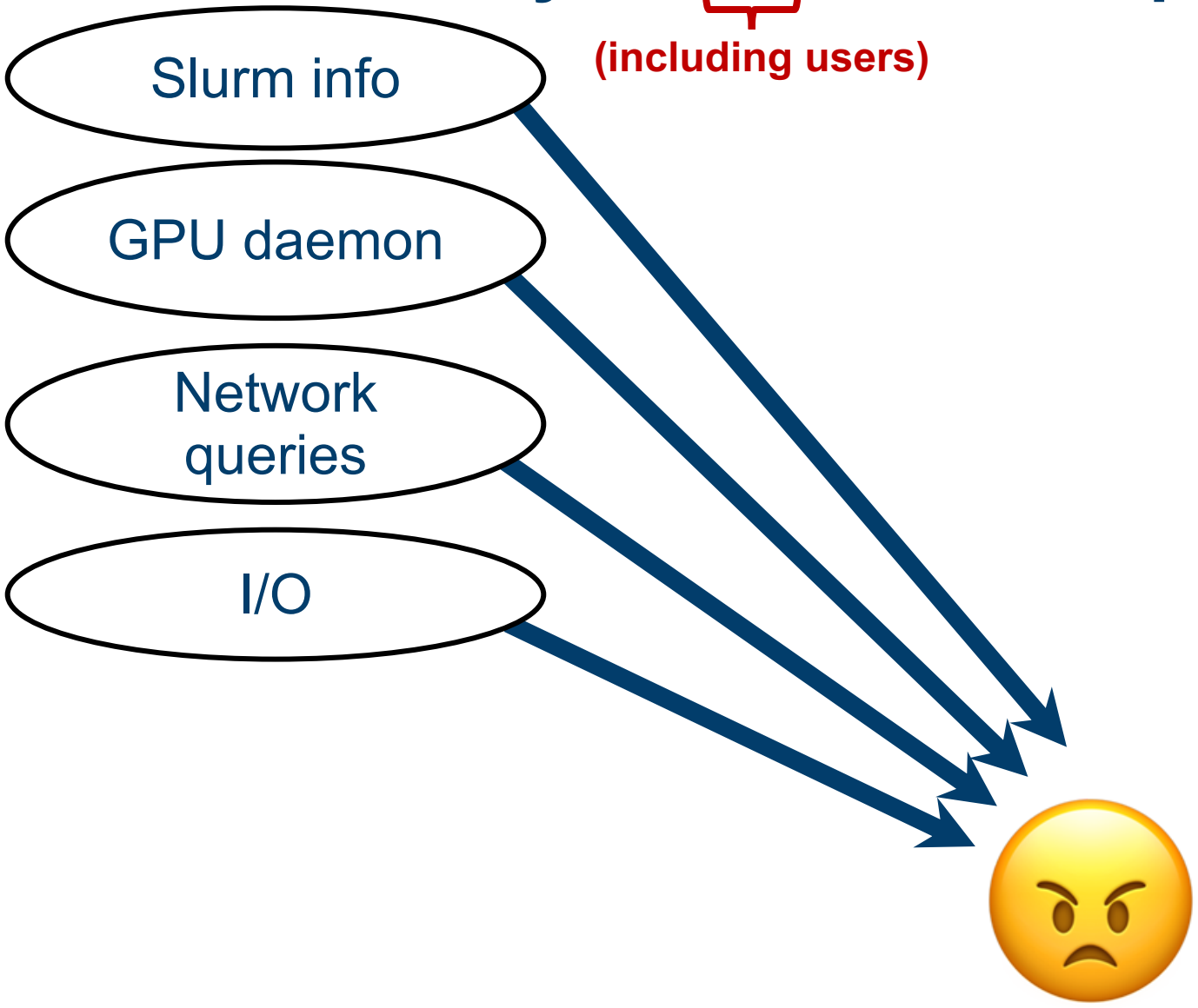
Why do we need a reporting platform?

(including users)



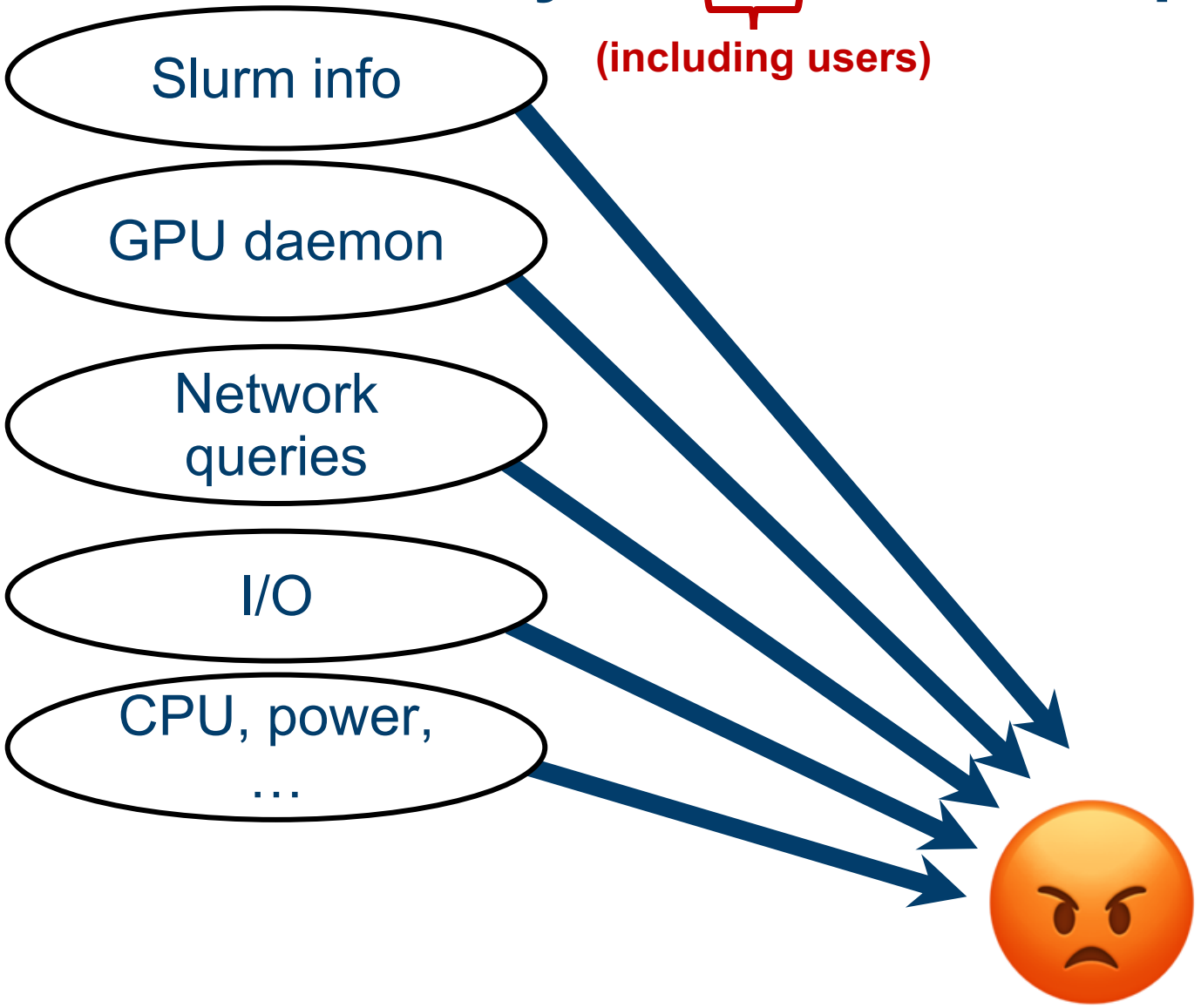
Why do we need a reporting platform?

(including users)



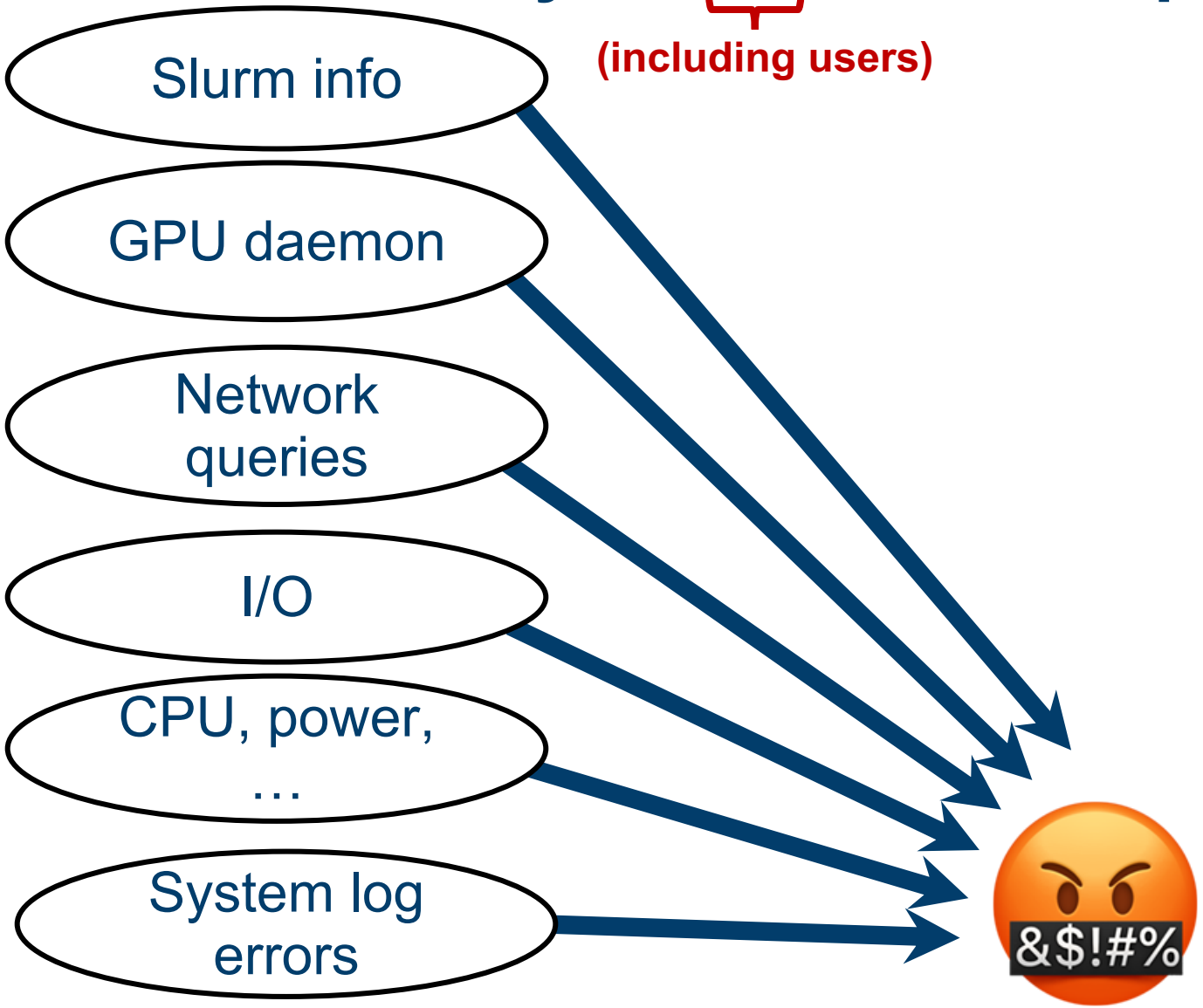
Why do we need a reporting platform?

(including users)



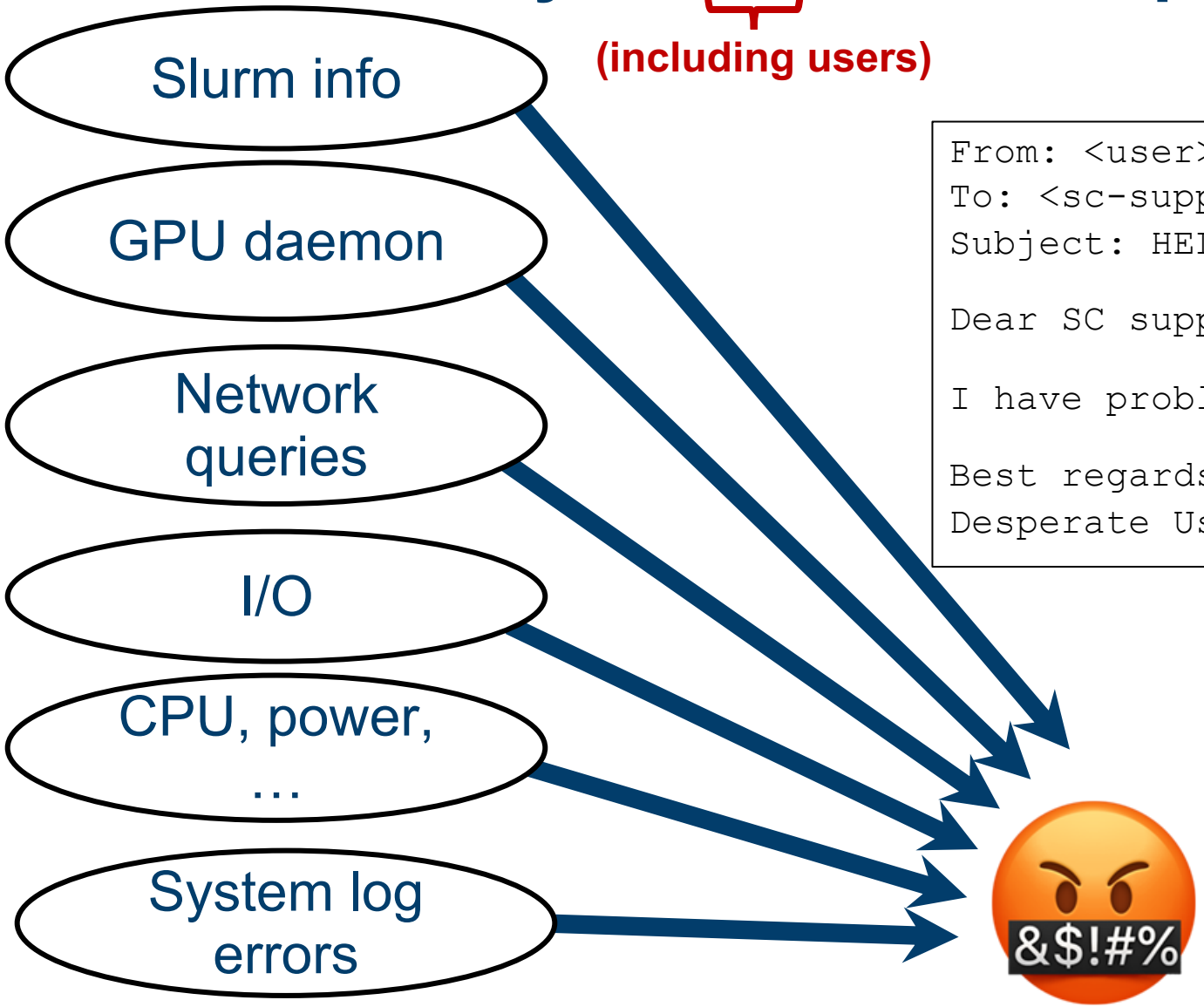
Why do we need a reporting platform?

(including users)



Why do we need a reporting platform?

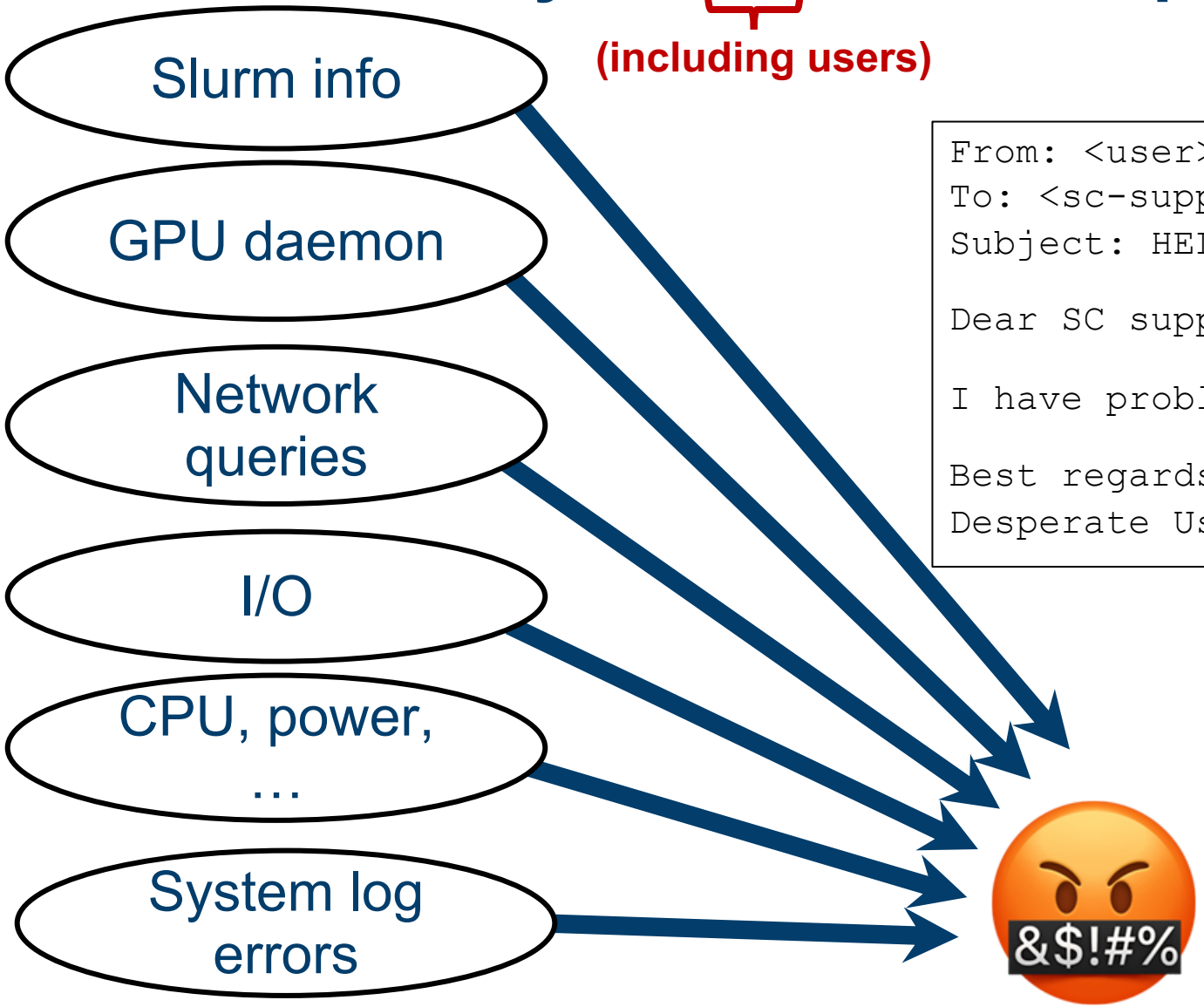
(including users)




```
From: <user>
To: <sc-support>
Subject: HELP!

Dear SC support,

I have problems! Help!

Best regards,
Desperate User
```

Why do we need a reporting platform?

(including users)




```
From: <user>
To: <sc-support>
Subject: HELP!

Dear SC support,

I have problems! Help!

Best regards,
Desperate User
```

```
From: <sc-support>
To: <admins>
Subject: Fw: HELP!

Dear Admins,

We have problems! Help!

Best regards,
Desperate Support
```

Why do we need a reporting platform?

(including users)

- Slurm info
- GPU daemon
- Network queries
- I/O
- CPU, power, …
- System log errors

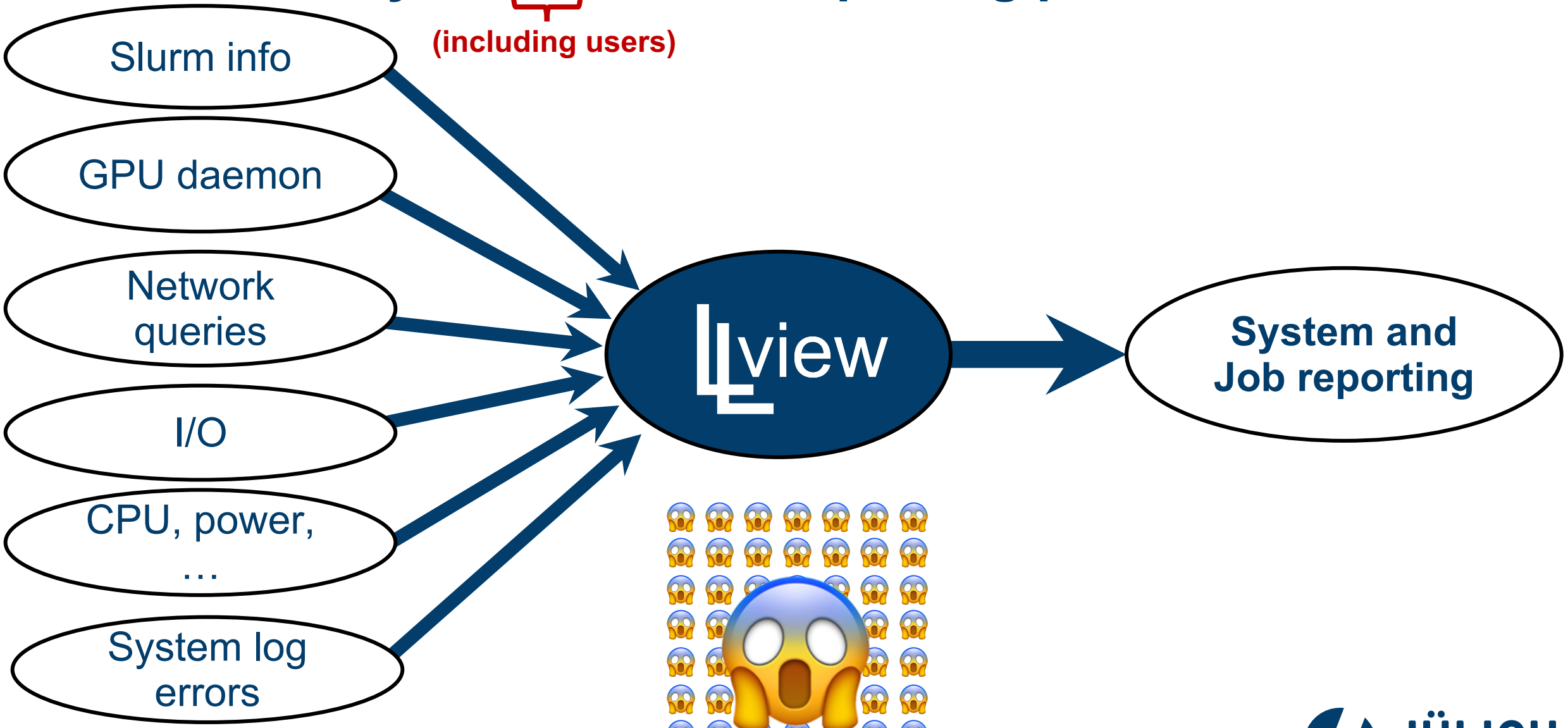
```
From: <user>
To: <sc-support>
Subject: HELP!

Dear SC support,

I have problems! Help!

Best regards,
Desperate User
```

```
From: <sc-support>
To: <admins>
Subject: Fw: HELP!

Dear Admins,

We have problems! Help!

Best regards,
Desperate Support
```

*ratios not in scale

Why do we need a reporting platform?

(including users)



Why do we need a reporting platform?

(including users)

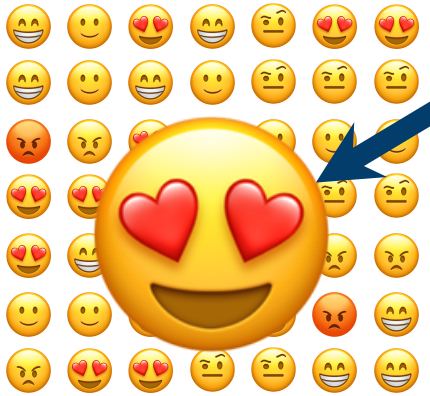
- Slurm info
- GPU daemon
- Network queries
- I/O
- CPU, power, ...
- System log errors

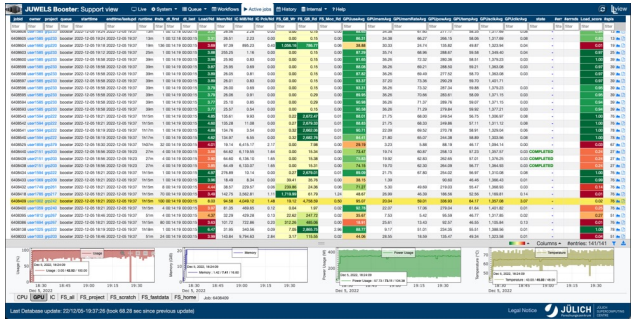
→ LLview →

System and Job reporting

Updates every ~1min using existing sources

- Negligible overhead!
- Scalable!





Why do we need a reporting platform?

(including users)

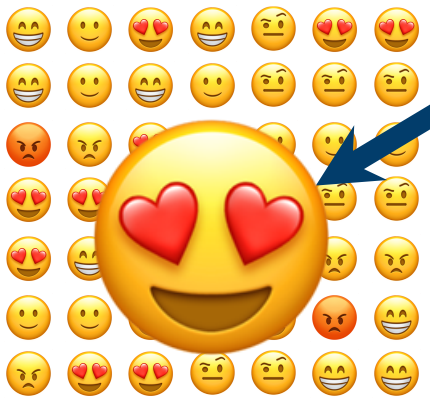
- Slurm info
- GPU daemon
- Network queries
- I/O
- CPU, power, ...
- System log errors

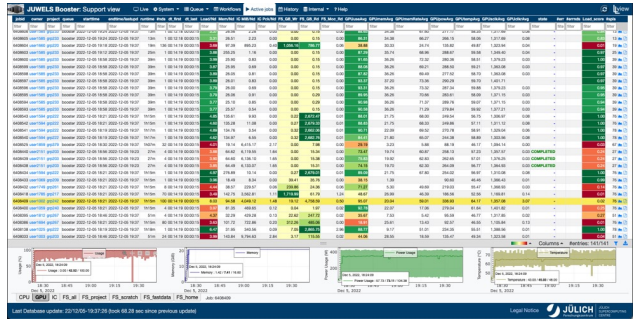
Supporters

System and Job reporting

Updates every ~1min using existing sources

- Negligible overhead!
- Scalable!





Why do we need a reporting platform?

(including users)

Slurm info

GPU daemon

Network queries

I/O

CPU, power, ...

System log errors

Supporters

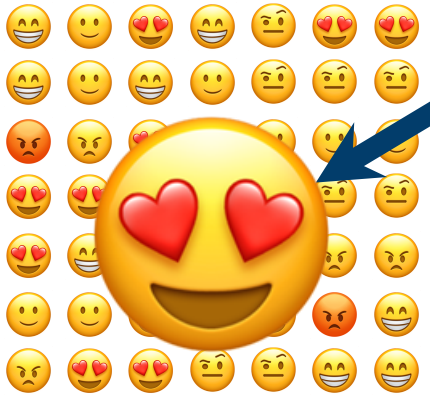
Admins

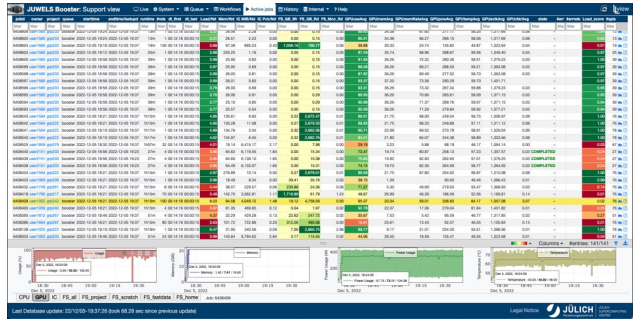
Updates every ~1min using existing sources

→ Negligible overhead!

→ Scalable!

System and Job reporting





Characteristics and features of LLview

- Developed in JSC since 2004
- Job reports available for all users since 2019
- Perl, Python, Javascript

- Near-real time information (also in HTML and PDF reports)
- Role-based access (Users, PI/PAs, mentors, support)
- Modular/extendable design
- Separation between data processing and visualisation
  - ➢ Static web portal (no query to internal DB is initiated by the user)

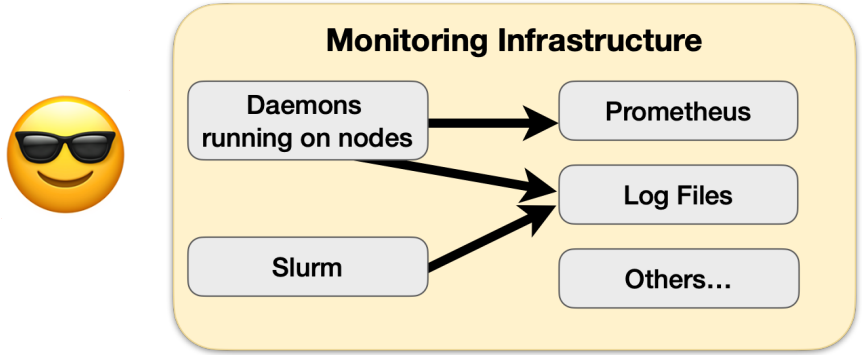
Job information

- Basic job information
- Computing time spent/estimation
- Job Status
- CPU: Load, Core, Memory usage
- GPU: Usage, Power, Temperature, …
- I/O statistics
- Network traffic
- Node list (with interconnected cells)
- System error messages
- Timeline

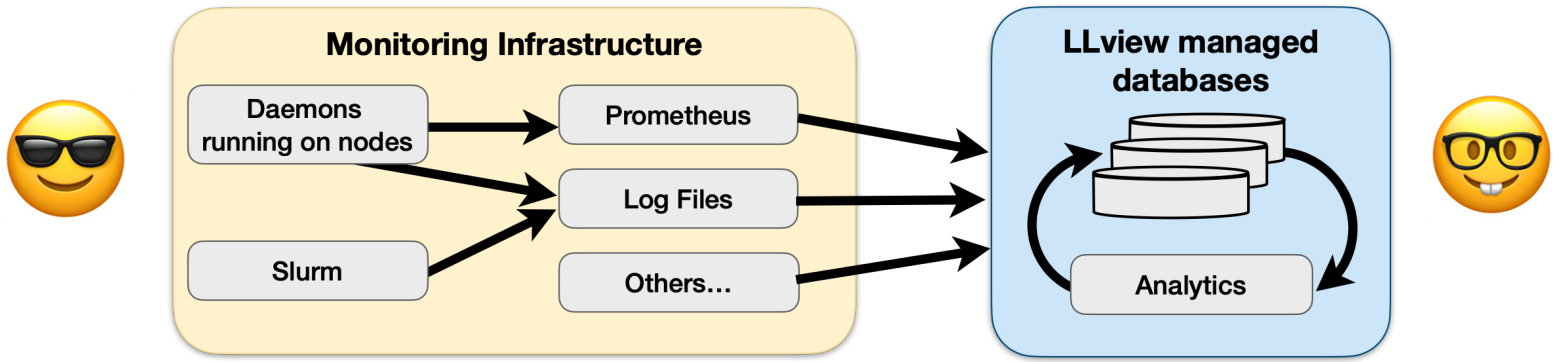
System information

- System usage overview
- Node usage, Errors

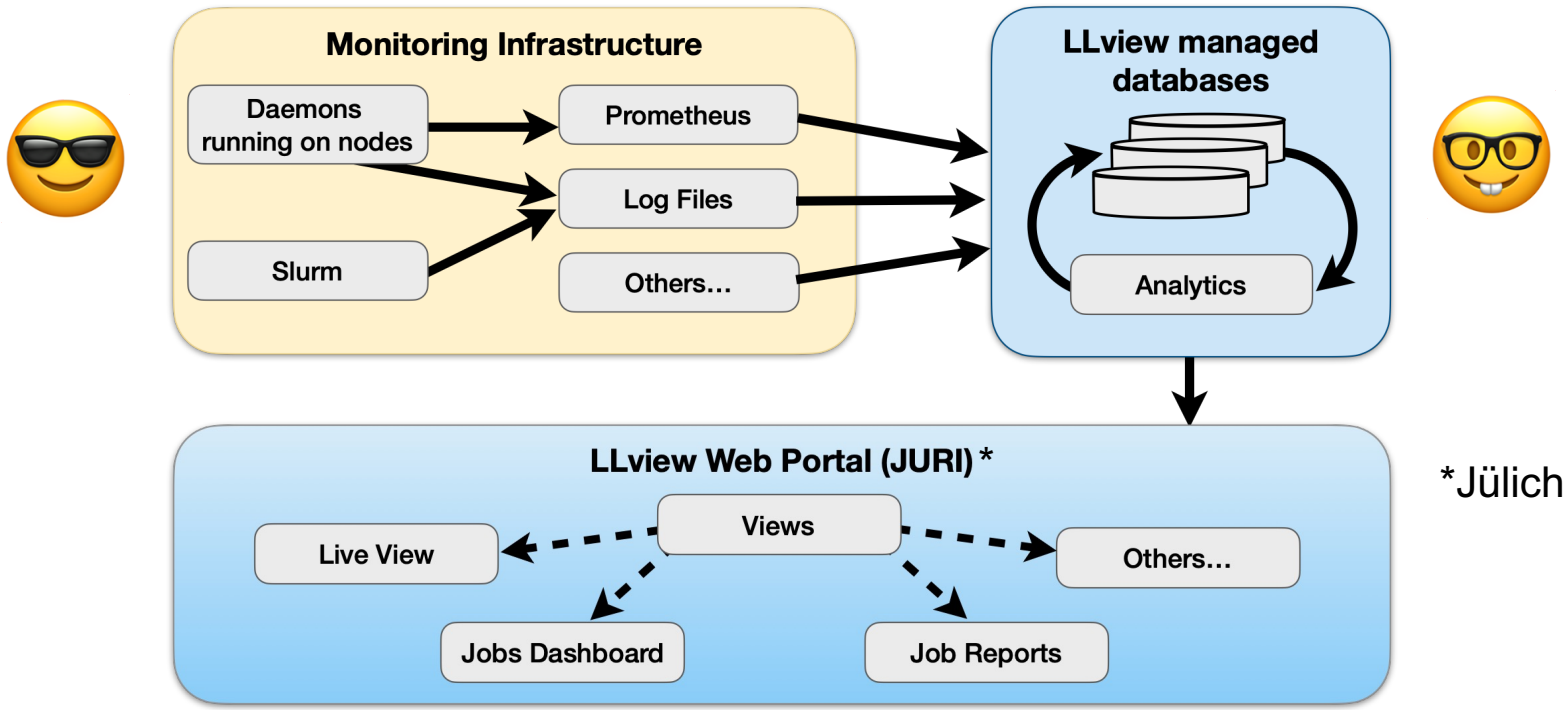
Simplified Scheme of LLview



Simplified Scheme of LLview

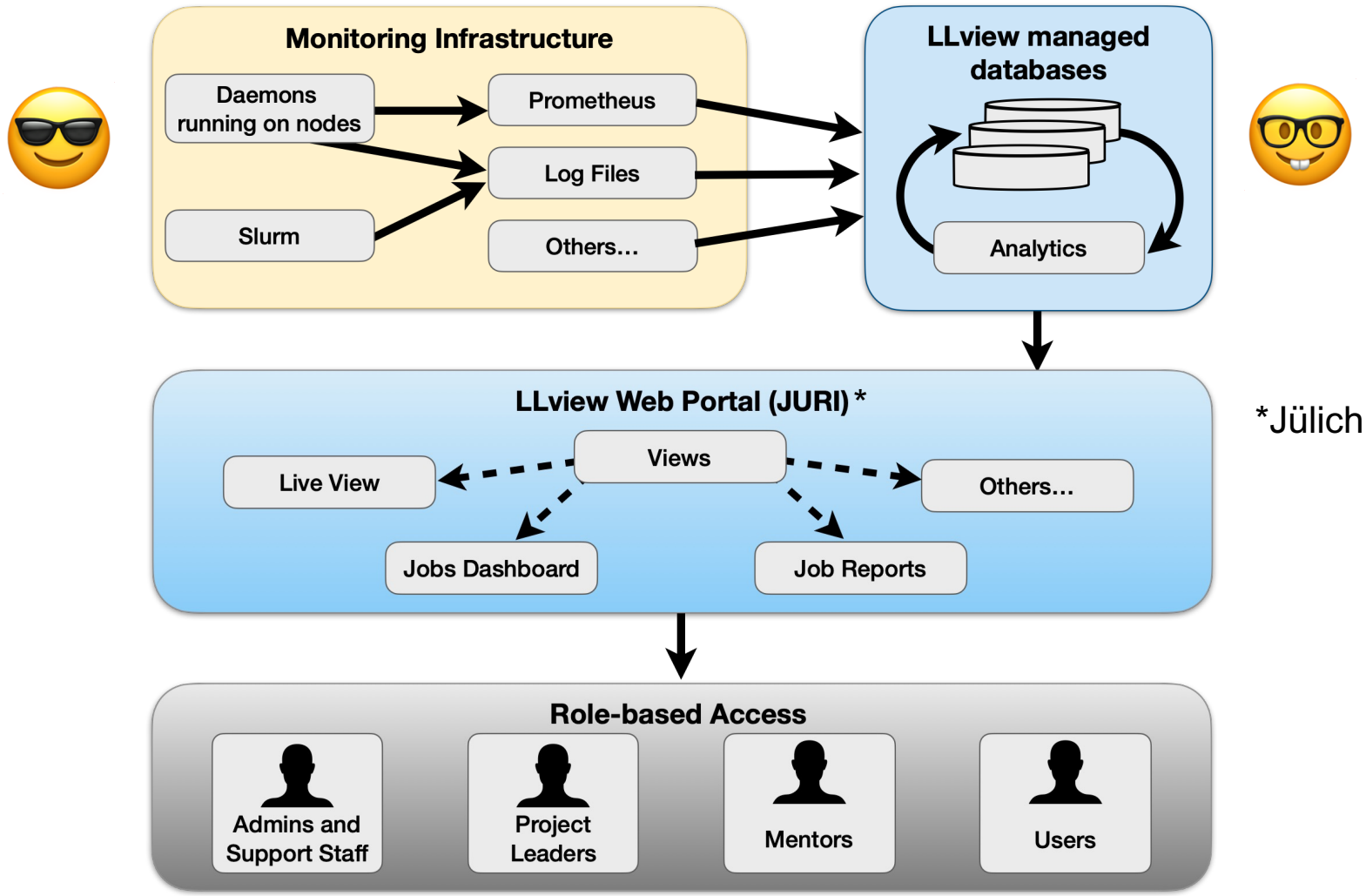


Simplified Scheme of LLview



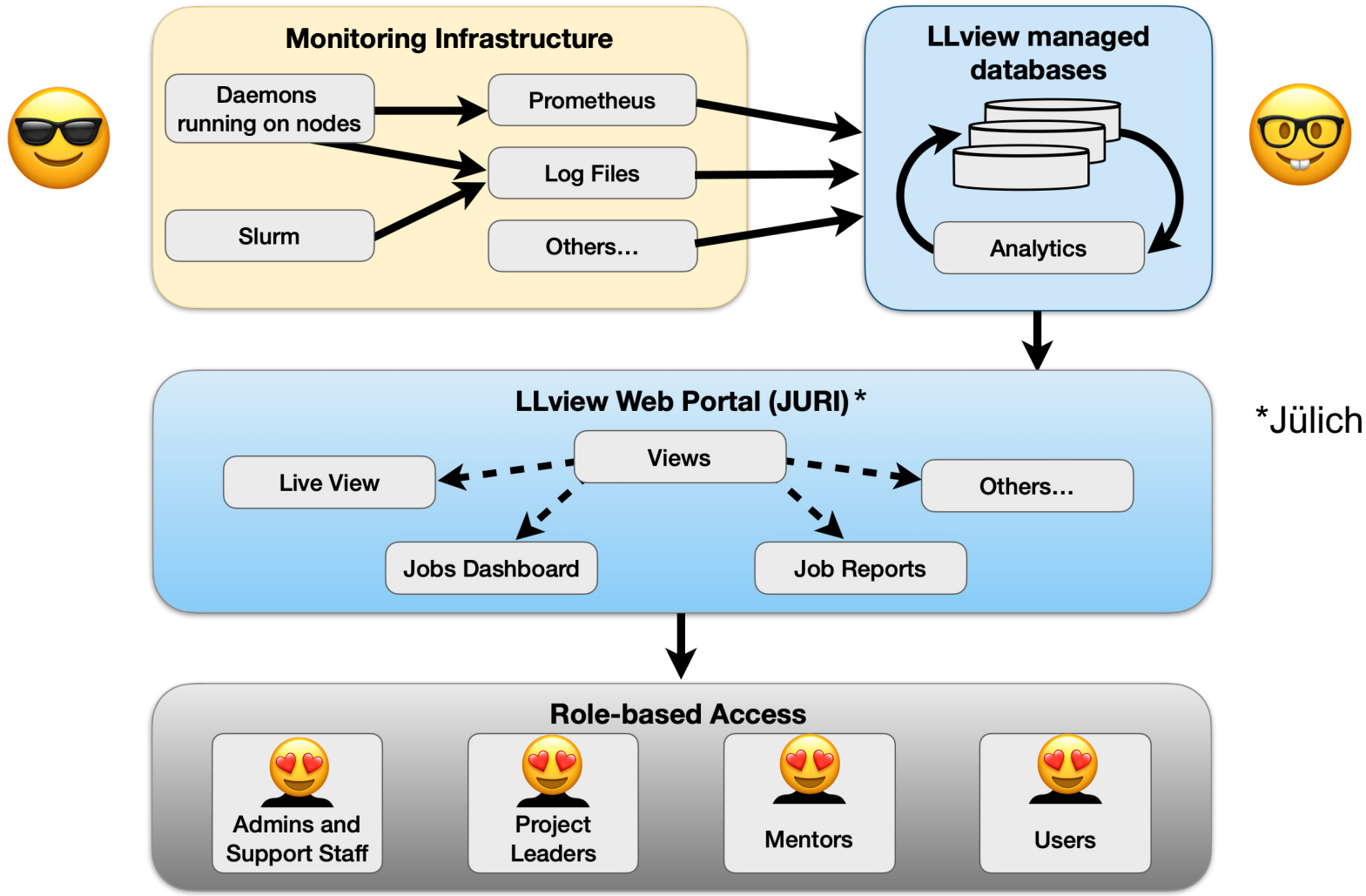
*Jülich Reporting Interface

Simplified Scheme of LLview



*Jülich Reporting Interface

Simplified Scheme of LLview



*Jülich Reporting Interface

More Detailed Scheme of LLview

LLview (DA)
Data Assimilation

*1/min**
* depending on the data source

Data Center
User DB

every 15min

LLview (DB)

sqlite3

Time Series 24h | *Job Statistics* 3 weeks | *Global Statistics* > 1 year

YAML

Configuration file(s)

1/day

1/day

LLview (long term DB)*

sqlite3

Aggr. data > 1 year

External Analytic Tools (AI)

~0.5 / min

Job Reporting Web Portal

3 weeks

~ 0.5 / min

REST API* Server

1 week

Update rate

Data lifetime

** in development*

More Detailed Scheme of LLview

LLview (DA)
Data Assimilation

*1/min**
* depending on the data source

LLview (DB)

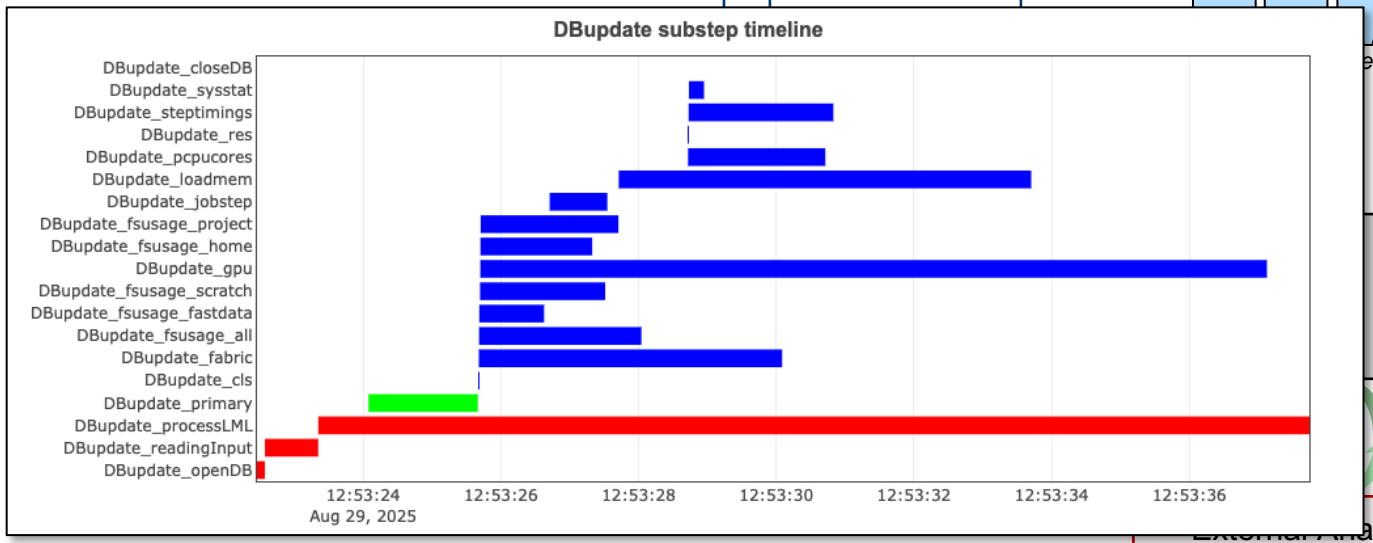
sqlite3

Time Series	Job Statistics	Global Statistics
24h	3 weeks	> 1 year

YAML

e(s)

DB)*

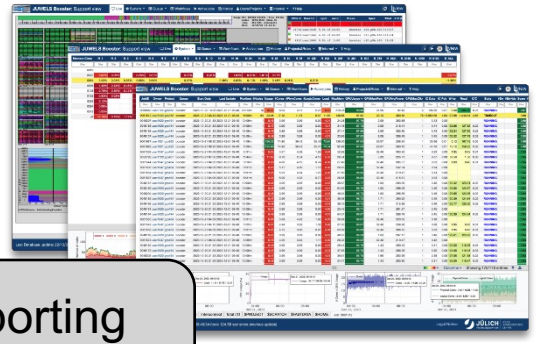


External Analytic Tools (AI)

~0.5 / min

Job Reporting Web Portal

3 weeks



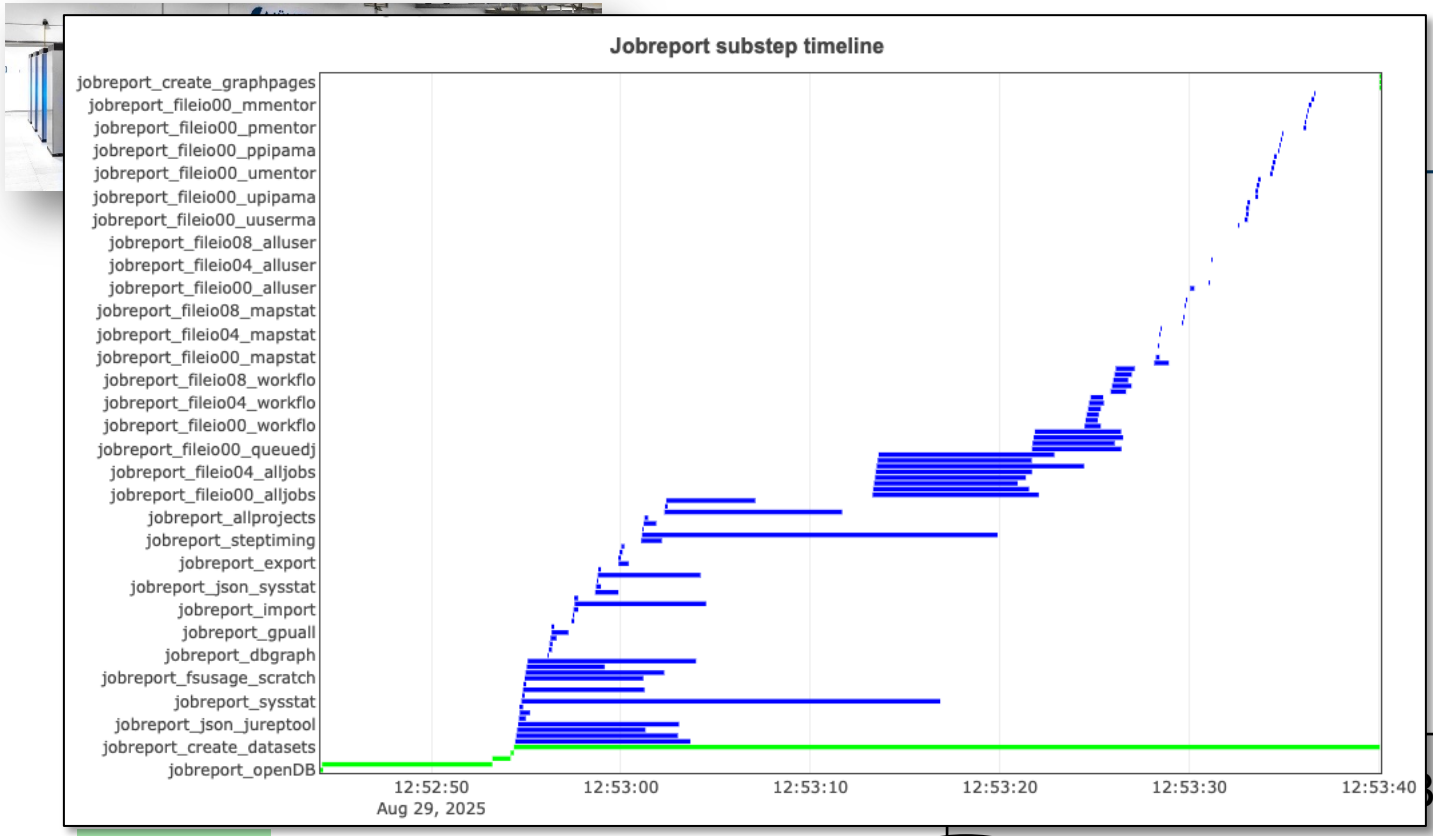
~ 0.5 / min

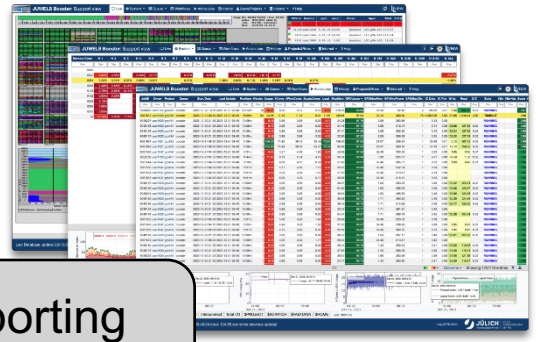
REST API* Server

1 week

* in development

More Detailed Scheme of LLview





Update rate

Data lifetime

~0.5 / min

Job Reporting Web Portal

3 weeks

~ 0.5 / min

REST API* Server

1 week

sqlite3

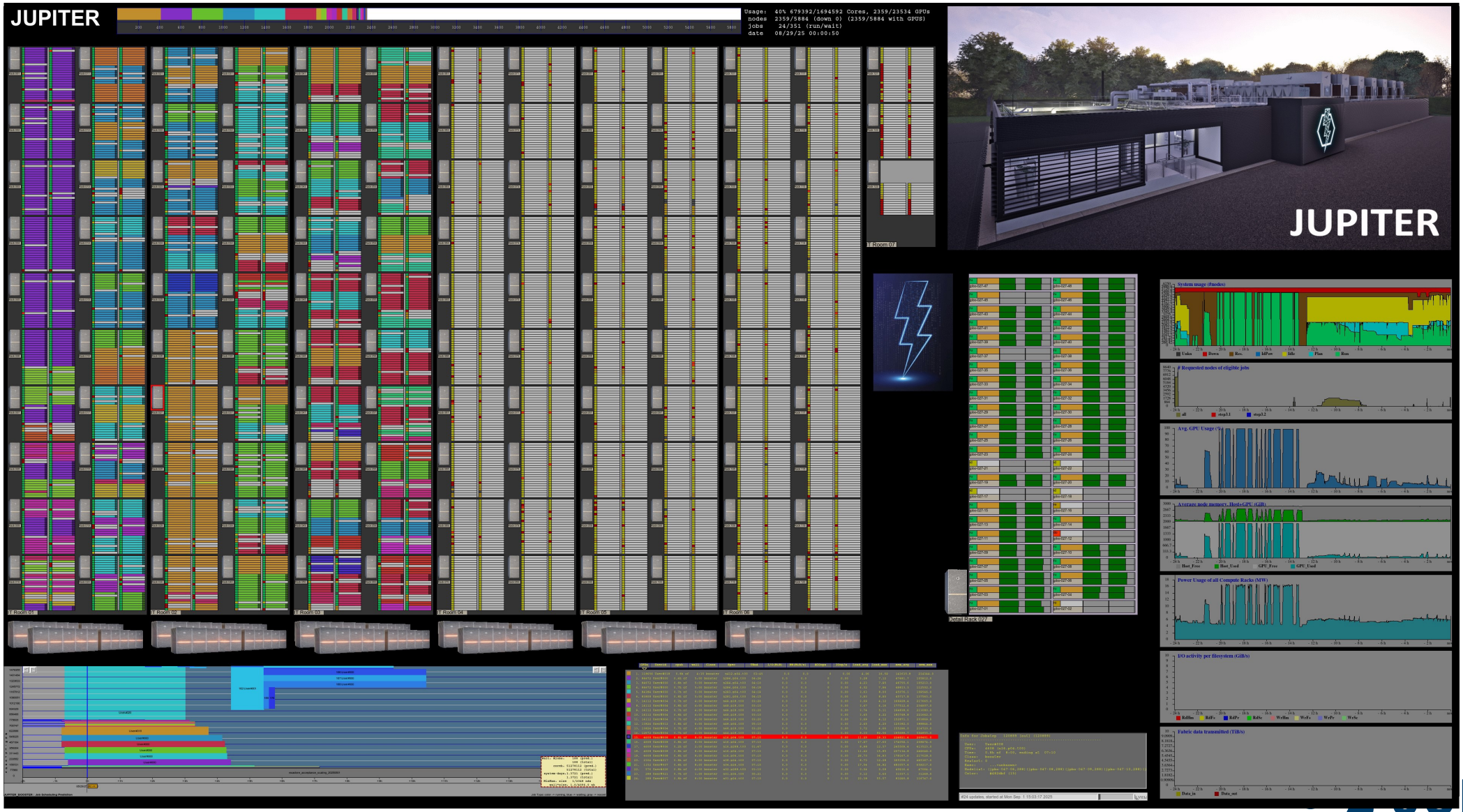
Aggr. data

> 1 year

External Analytic Tools (AI)

** in development*

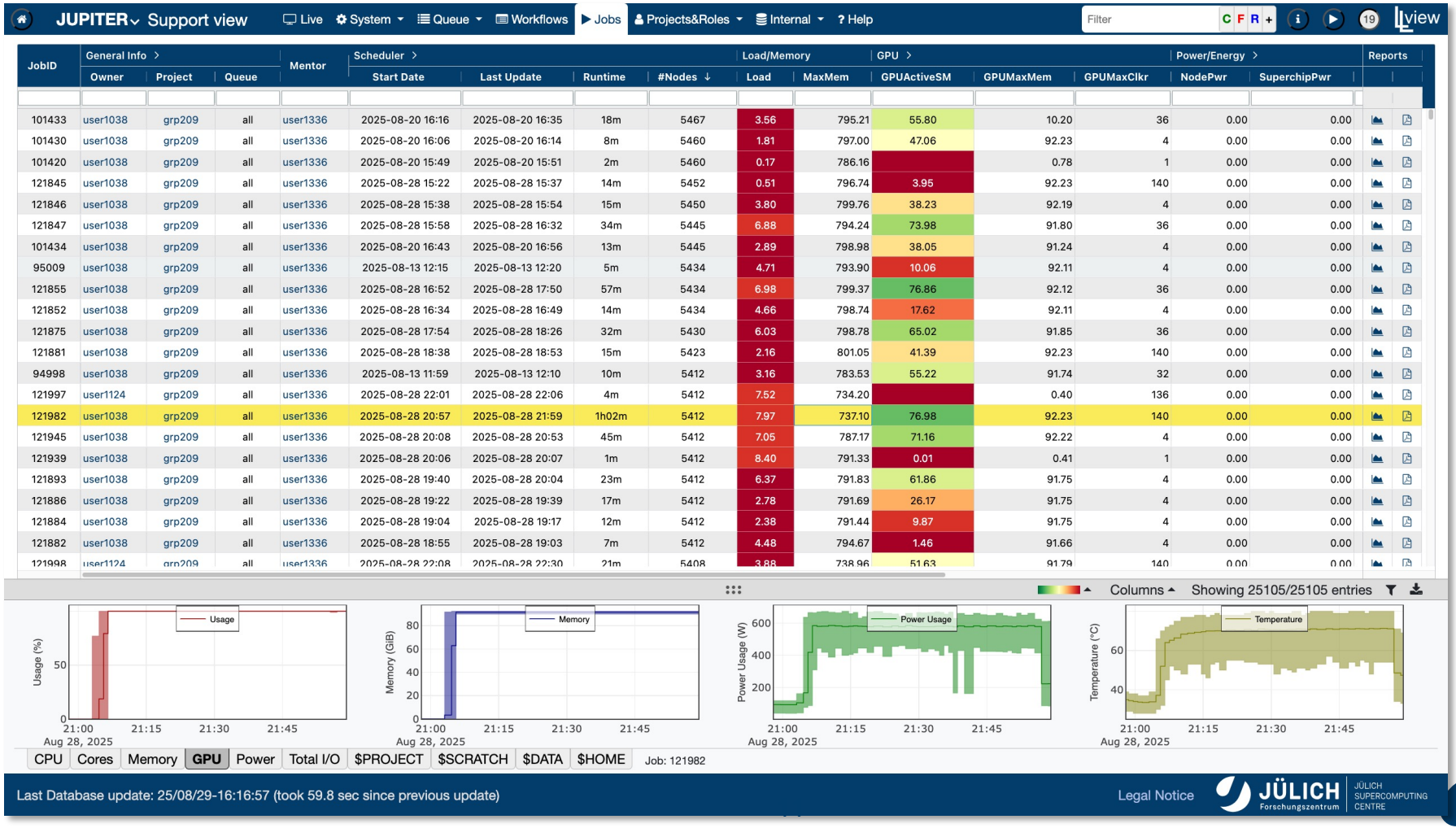
Live view shows current usage



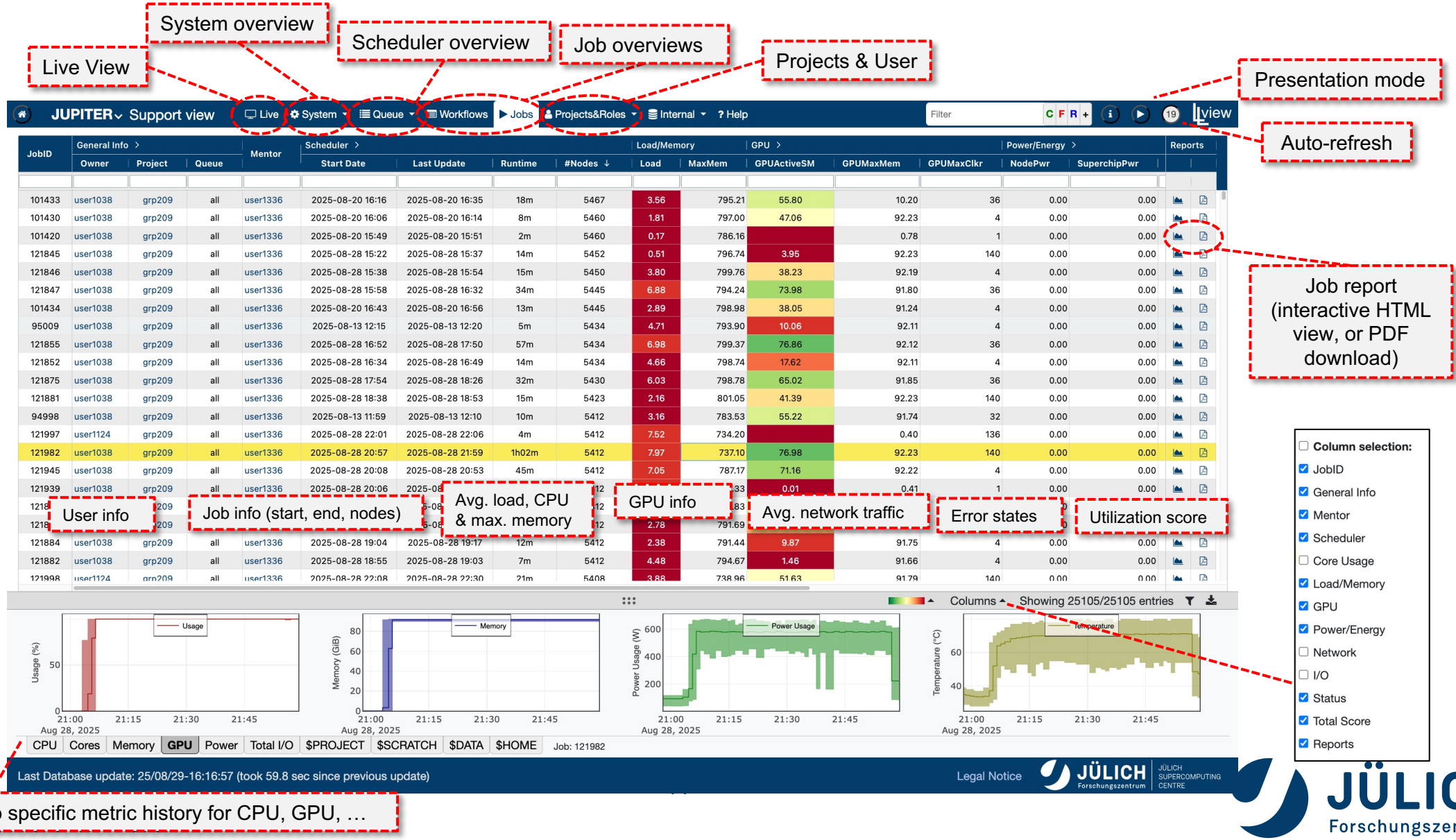
(LLview is also useful for Quantum Computers)

JÜLICH
Forschungszentrum

Dashboard with job-focused information



Dashboard with job-focused information

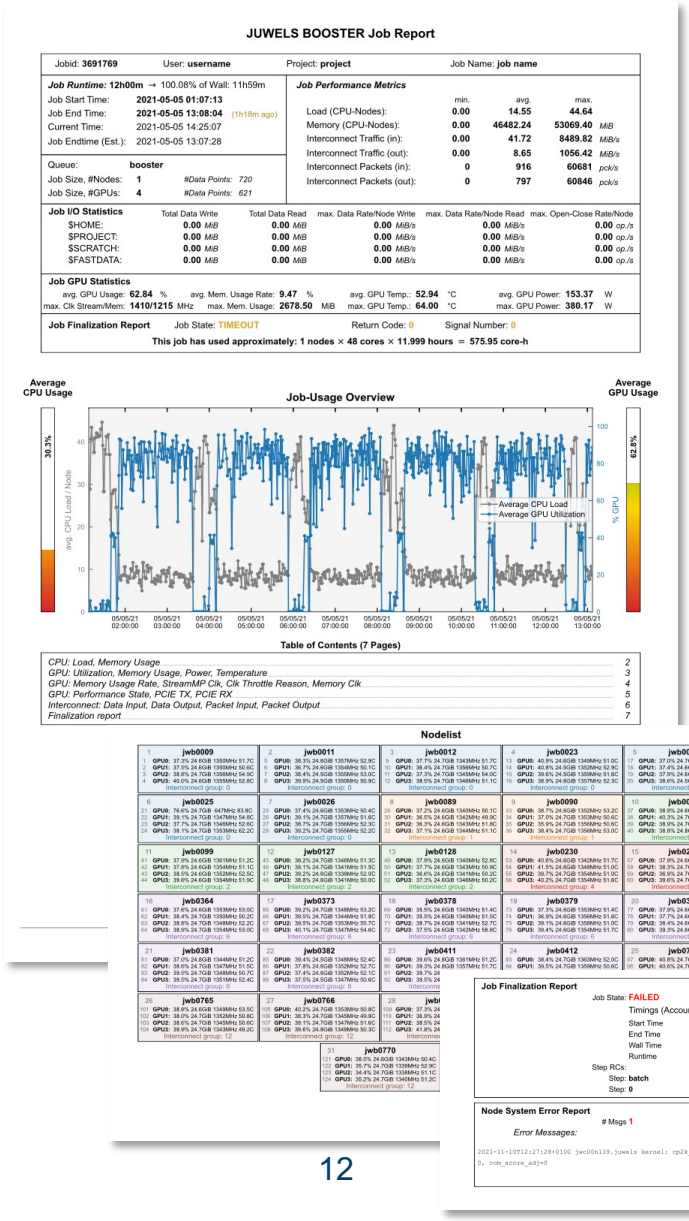


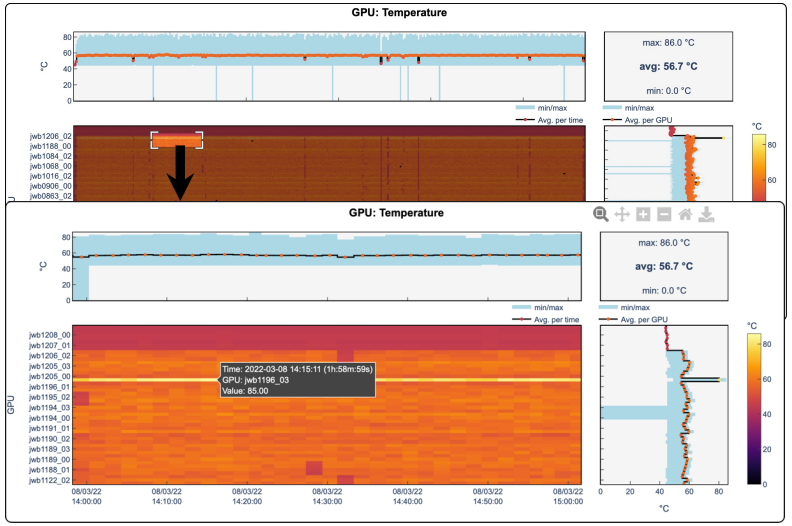
JobID	General Info			Mentor	Scheduler				Load/Memory		GPU			Power/Energy		Reports
	Owner	Project	Queue		Start Date	Last Update	Runtime	#Nodes	Load	MaxMem	GPUActiveSM	GPUMaxMem	GPUMaxClkr	NodePwr	SuperchipPwr	
101433	user1038	grp209	all	user1336	2025-08-20 16:16	2025-08-20 16:35	18m	5467	3.56	795.21	55.80	10.20	36	0.00	0.00	
101430	user1038	grp209	all	user1336	2025-08-20 16:06	2025-08-20 16:14	8m	5460	1.81	797.00	47.06	92.23	4	0.00	0.00	
101420	user1038	grp209	all	user1336	2025-08-20 15:49	2025-08-20 15:51	2m	5460	0.17	786.16		0.78	1	0.00	0.00	
121845	user1038	grp209	all	user1336	2025-08-28 15:22	2025-08-28 15:37	14m	5452	0.51	796.74	3.95	92.23	140	0.00	0.00	
121846	user1038	grp209	all	user1336	2025-08-28 15:38	2025-08-28 15:54	15m	5450	3.80	799.76	38.23	92.19	4	0.00	0.00	
121847	user1038	grp209	all	user1336	2025-08-28 15:58	2025-08-28 16:32	34m	5445	6.88	794.24	73.98	91.80	36	0.00	0.00	
101434	user1038	grp209	all	user1336	2025-08-20 16:43	2025-08-20 16:56	13m	5445	2.89	798.98	38.05	91.24	4	0.00	0.00	
95009	user1038	grp209	all	user1336	2025-08-13 12:15	2025-08-13 12:20	5m	5434	4.71	793.90	10.06	92.11	4	0.00	0.00	
121855	user1038	grp209	all	user1336	2025-08-28 16:52	2025-08-28 17:50	57m	5434	6.98	799.37	76.86	92.12	36	0.00	0.00	
121852	user1038	grp209	all	user1336	2025-08-28 16:34	2025-08-28 16:49	14m	5434	4.66	798.74	17.62	92.11	4	0.00	0.00	
121875	user1038	grp209	all	user1336	2025-08-28 17:54	2025-08-28 18:26	32m	5430	6.03	798.78	65.02	91.85	36	0.00	0.00	
121881	user1038	grp209	all	user1336	2025-08-28 18:38	2025-08-28 18:53	15m	5423	2.16	801.05	41.39	92.23	140	0.00	0.00	
94998	user1038	grp209	all	user1336	2025-08-13 11:59	2025-08-13 12:10	10m	5412	3.16	783.53	55.22	91.74	32	0.00	0.00	
121997	user1124	grp209	all	user1336	2025-08-28 22:01	2025-08-28 22:06	4m	5412	7.52	734.20		0.40	136	0.00	0.00	
121982	user1038	grp209	all	user1336	2025-08-28 20:57	2025-08-28 21:59	1h02m	5412	7.97	737.10	76.98	92.23	140	0.00	0.00	
121945	user1038	grp209	all	user1336	2025-08-28 20:08	2025-08-28 20:53	45m	5412	7.05	787.17	71.16	92.22	4	0.00	0.00	
121939	user1038	grp209	all	user1336	2025-08-28 20:06	2025-08		12		33	0.01	0.41	1	0.00	0.00	
1218		209				5-08		12		83						
1218		209				5-08		12	2.78	791.69						
121884	user1038	grp209	all	user1336	2025-08-28 19:04	2025-08-28 19:17	12m	5412	2.38	791.44	9.87	91.75	4	0.00	0.00	
121882	user1038	grp209	all	user1336	2025-08-28 18:55	2025-08-28 19:03	7m	5412	4.48	794.67	1.46	91.66	4	0.00	0.00	
121998	user1124	grp209	all	user1336	2025-08-28 22:08	2025-08-28 22:30	21m	5408	3.88	738.96	51.63	91.79	140	0.00	0.00	

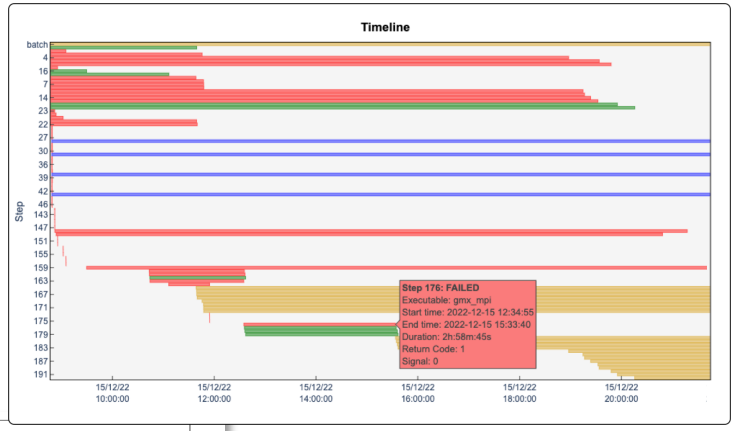
JÜLICH Forschungszentrum

Job reports contain detailed information

- Basic job information
- General average usage
- Computing time spent/estimation
- Status
- CPU, Core
  - Usage, Load, Memory usage
- GPU
  - Usage, Power, Temperature, …
- I/O
- Network traffic
- Nodelist (with interconnected cells)
- Error messages
- Timeline

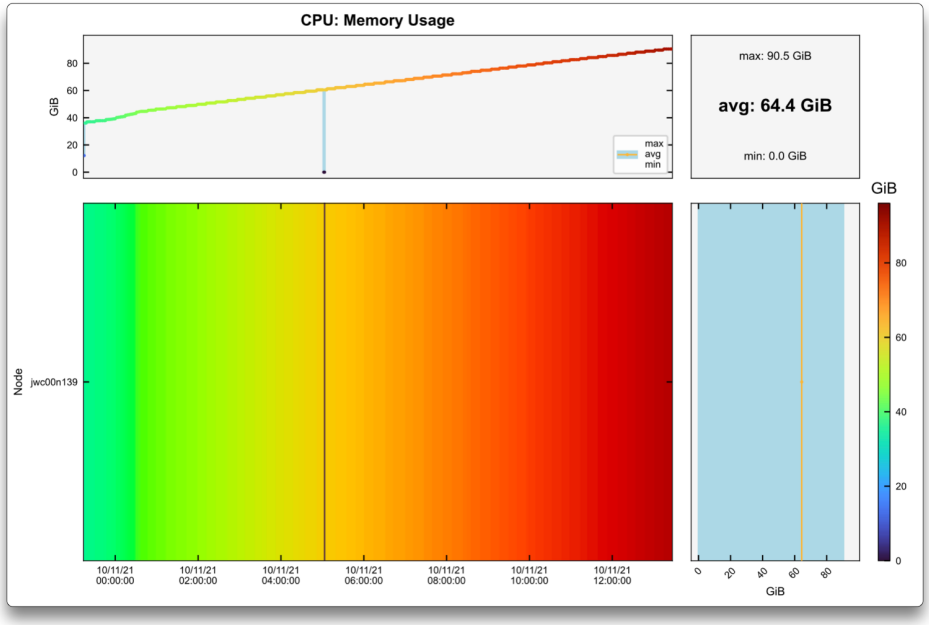


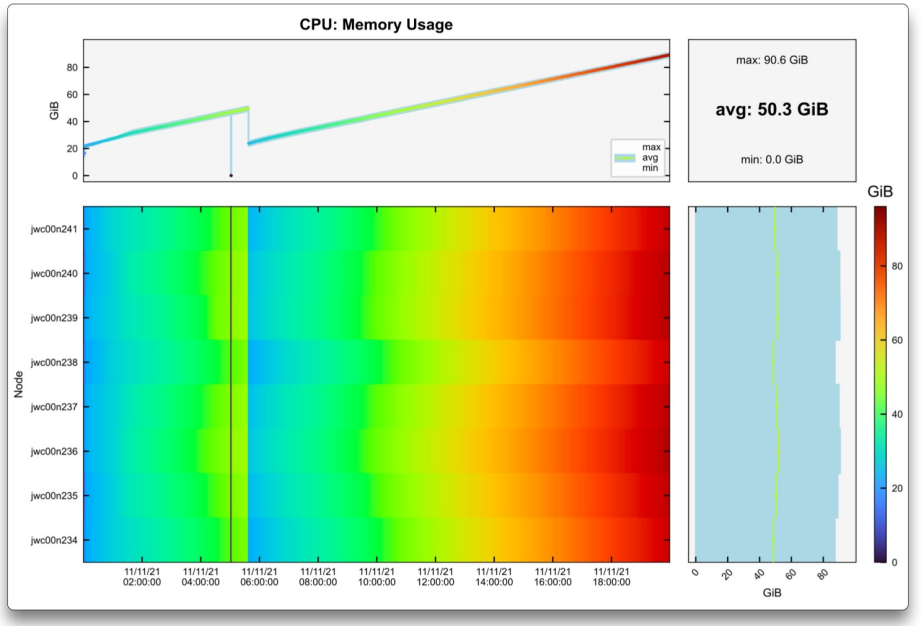


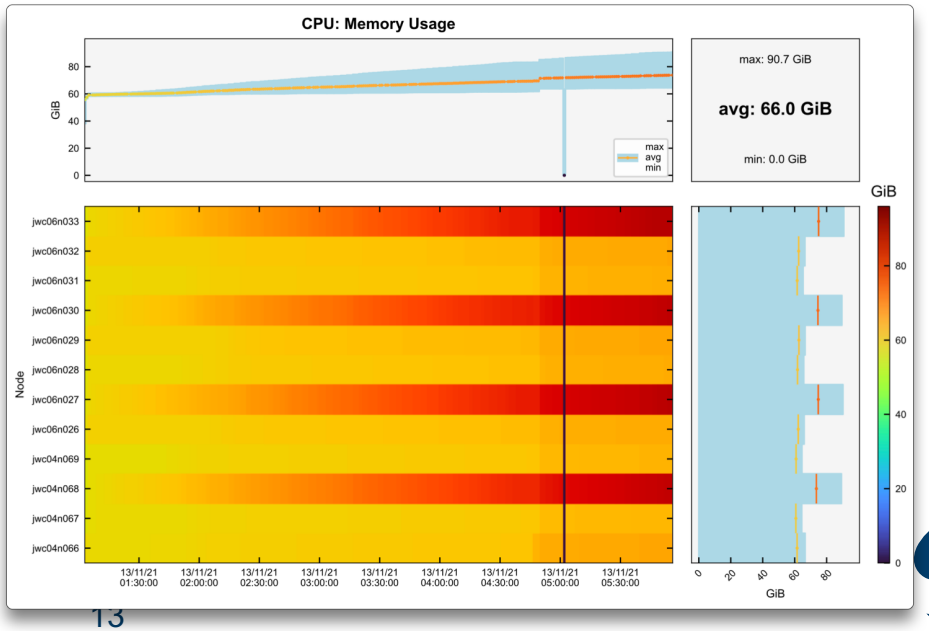


Examples

Memory leak







Examples

System Error Logs

Job Finalization Report	Job State: **FAILED**	Return Code: **9**	Signal Number: **0**
System Error Report	# Msgs: **2** # Nodes: **1**	**(Out-of-memory)**	

This job has used approximately: 1 nodes × 48 cores × 0.186 hours = 8.93 core-h

System Error Report

Msgs: **2** # Nodes: **1**

Error Messages:

```
2025-04-24T19:21:29+0200 jwb0163.juwels kernel: python invoked oom-killer:
gfp_mask=0x140dca(GFP_HIGHUSER_MOVABLE!__GFP_COMP!__GFP_ZERO), order=0, oom_score_adj=0
2025-04-24T19:21:29+0200 jwb0163.juwels kernel: Out of memory: Killed process 1556003 (python) total-
vm:902572524kB, anon-rss:496358356kB, file-rss:89856kB, shmem-rss:19200kB, UID:xxxxx pgtables:973944kB
oom_score_adj:0
```

Examples

System Error Logs

Job Finalization Report	Job State: **TIMEOUT**	Return Code: **0**	Signal Number: **0**
System Error Report	# Msgs: **2** # Nodes: **1**	**(Flipping Link)**	

This job has used approximately: 16 nodes × 48 cores × 11.973 hours = 9195.26 core-h

System Error Report

Msgs: **2** # Nodes: **1**

Error Messages:

```
2025-06-02T03:24:56+0200 jwb0155.juwels pshealthcheck[184436]: [EE] ERROR: mlx5_2 local port (LID
4667, port 1): LinkDownedCounter = 8 (threshold = 1)
2025-06-02T03:24:56+0200 jwb0155.juwels pshealthcheck[184436]: [EE] ERROR: mlx5_2 remote port (LID
297, port 34): LinkDownedCounter = 8 (threshold = 1)
```

Examples

System Error Logs

Job Finalization Report	Job State: **FAILED**		Return Code: **0**	Signal Number: **9**
System Error Report	# Msgs: **59**	# Nodes: **59**	**(Node Error)**	

This job has used approximately: 60 nodes × 48 cores × 11.328 hours = 32624.64 core-h

System Error Report

Msgs: **59** # Nodes: **59**

Error Messages:

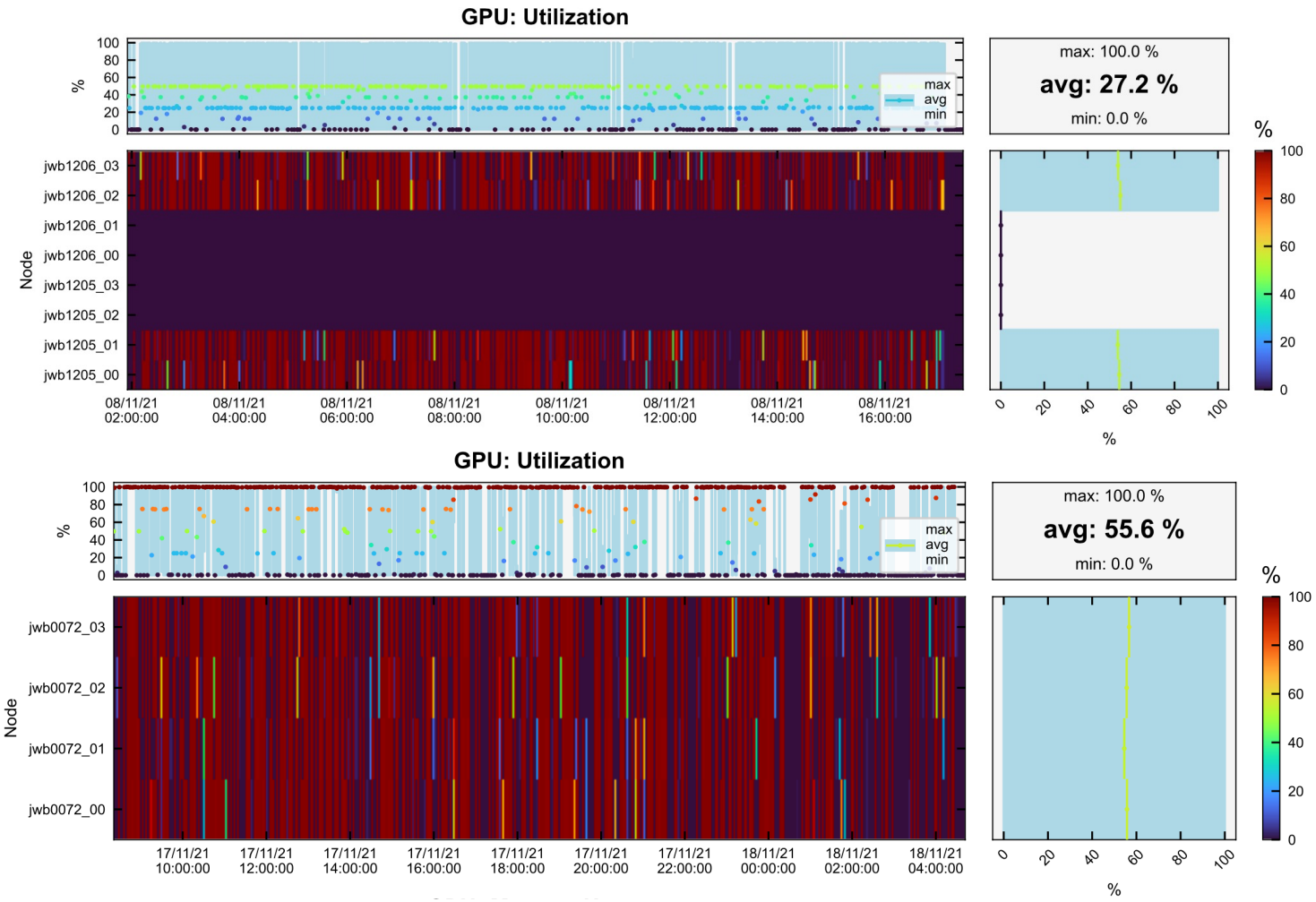
```
2025-06-03T10:50:14+0200 jwb0011.juwels psid[63607]: psslurm: nodeDownAlloc: node 3764 in allocation
11584692 state A_RUNNING is down
2025-06-03T10:50:15+0200 jwb0025.juwels psid[80835]: psslurm: nodeDownAlloc: node 3764 in allocation
11584692 state A_EPILOGUE is down
2025-06-03T10:50:15+0200 jwb0036.juwels psid[62353]: psslurm: nodeDownAlloc: node 3764 in allocation
11584692 state A_EPILOGUE is down
2025-06-03T10:50:15+0200 jwb0041.juwels psid[64695]: psslurm: nodeDownAlloc: node 3764 in allocation
11584692 state A_EPILOGUE is down
2025-06-03T10:50:15+0200 jwb0054.juwels psid[62378]: psslurm: nodeDownAlloc: node 3764 in allocation
11584692 state A_EPILOGUE is down
```

Examples

Inadequate Resource Usage

2 nodes (8 GPUs) using only 2 GPUs each

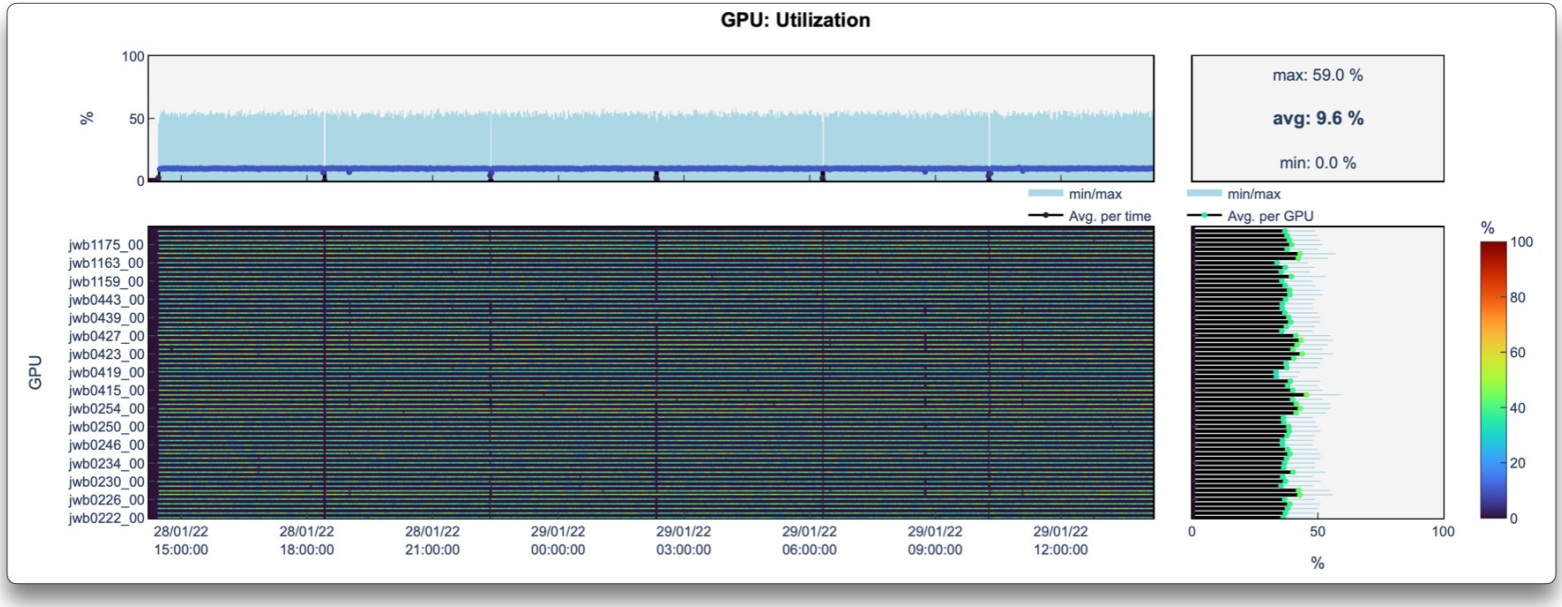
1 nodes using all 4 GPUs



Examples

Inadequate Resource Usage

1 GPU per node

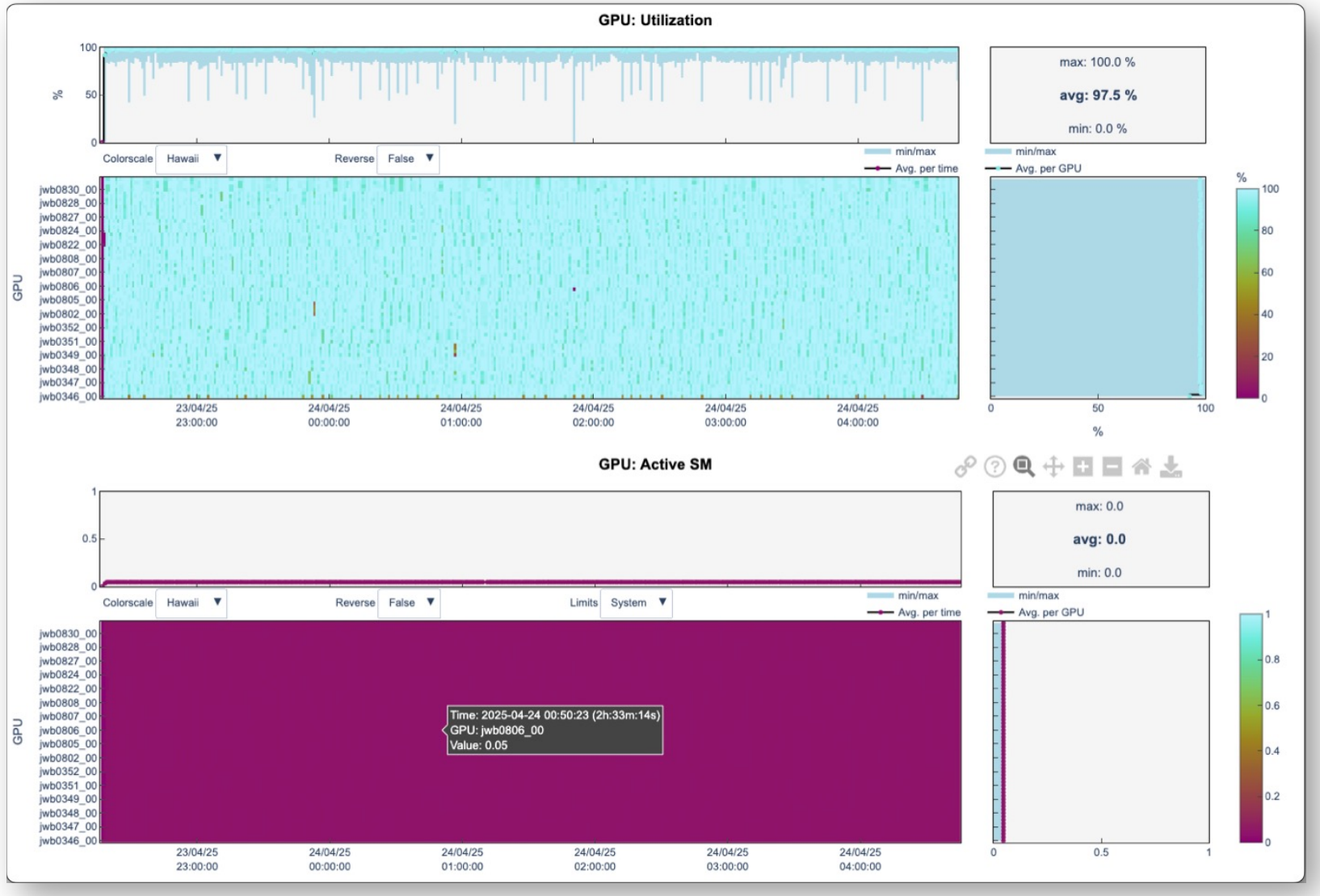


Examples

Inadequate Resource Usage

GPU Utilization ~100%

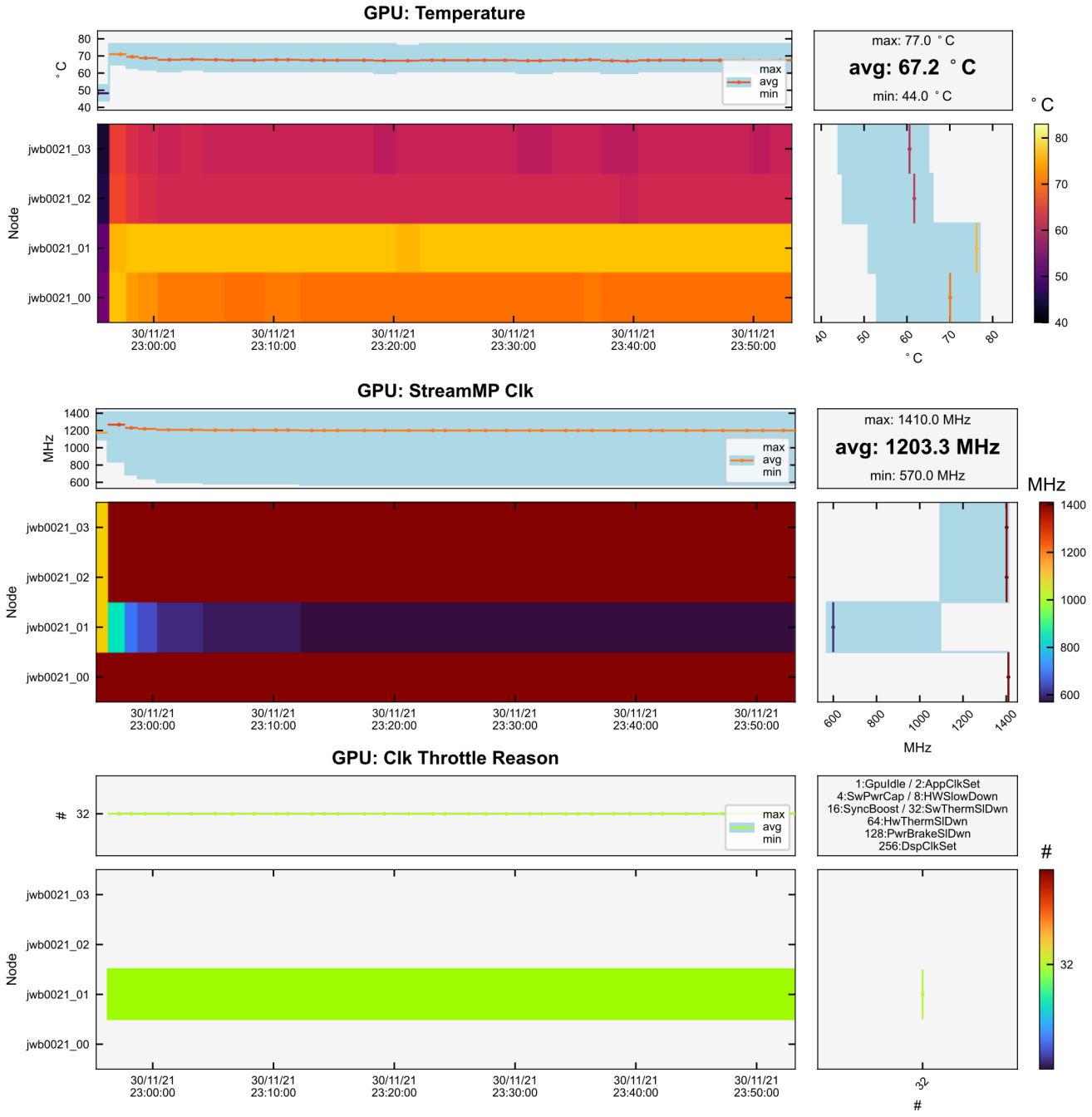
GPU Active SM ~5%



Examples

Hardware Issues

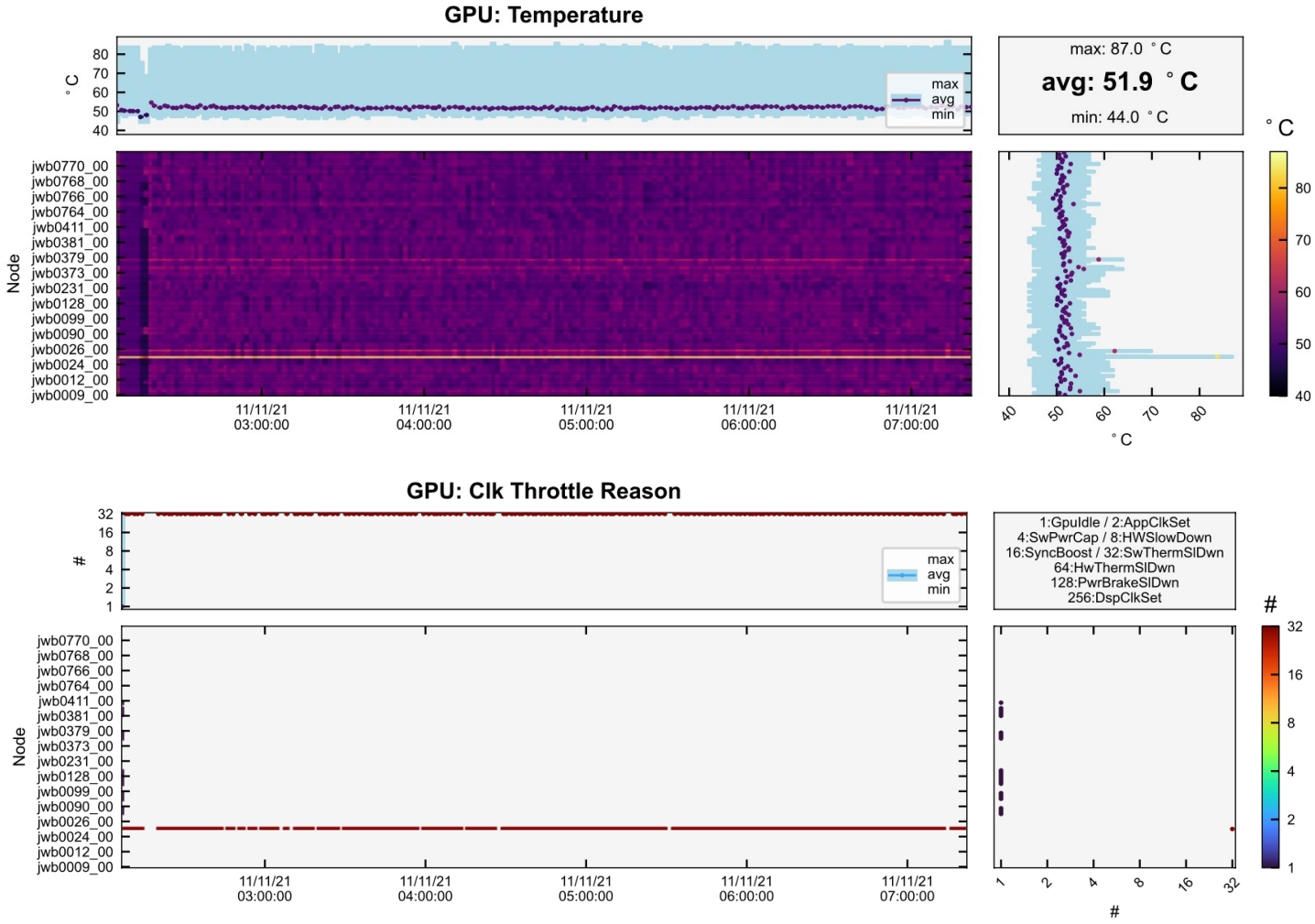
4 GPUs:



Examples

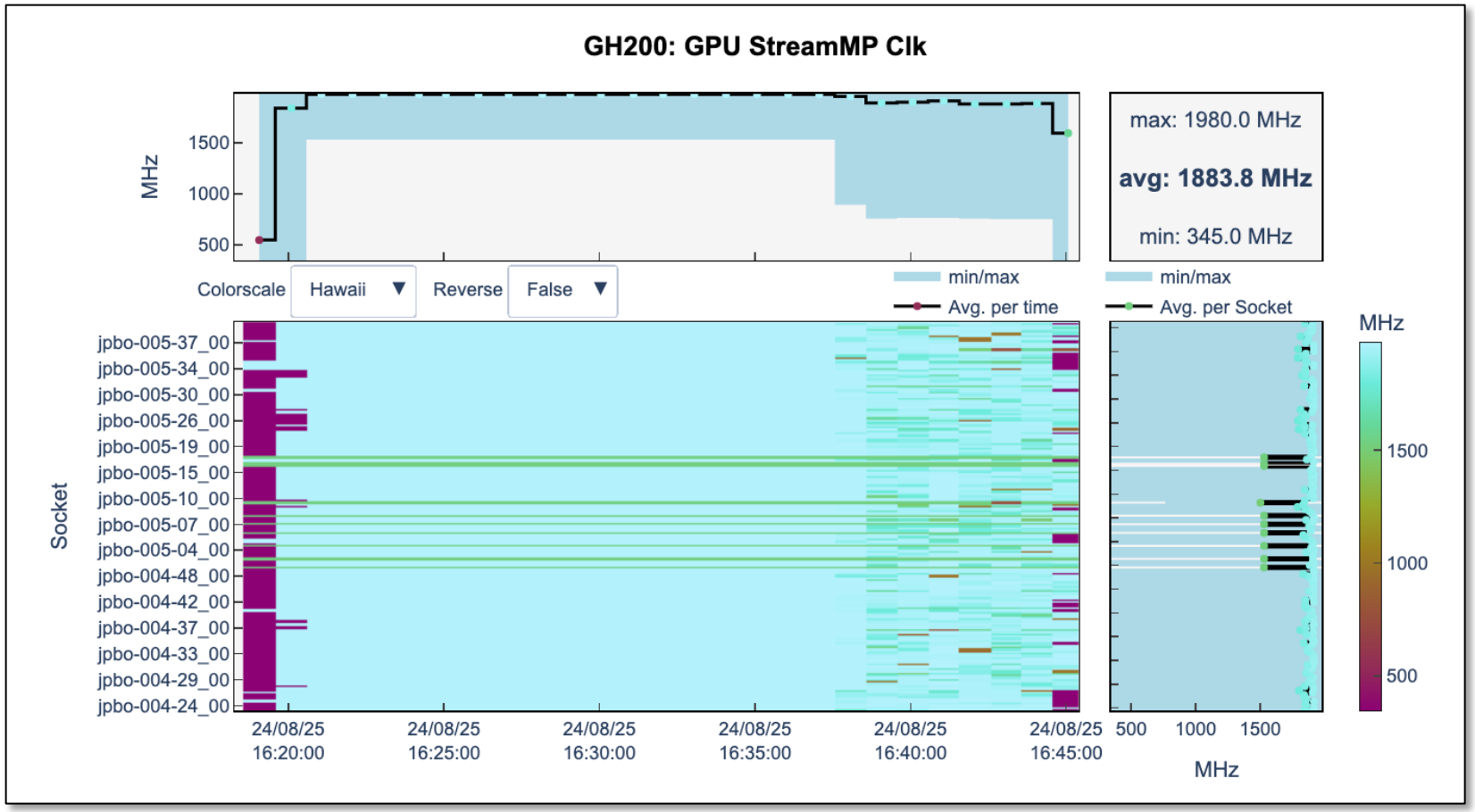
Hardware Issues

128 GPUs:



Examples

Low GPU Frequency



Current steps and Outlook

- Faster and optimized updates for JUPITER
  - Bigger cluster
  - Improved workloads
  - Optimized SQL queries
  - Further parallelization

Current steps and Outlook

- Faster and optimized updates for JUPITER
  - Bigger cluster
  - Improved workloads
  - Optimized SQL queries
  - Further parallelization

Current steps and Outlook

- Faster and optimized updates for JUPITER
  - Bigger cluster
  - Improved workloads
  - Optimized SQL queries
  - Further parallelization
- Generalisations, new metrics and data sources
- Queue view and analysis
- REST-API
- Long-term analysis
- Workflow and project reports
- …

JUPITER

EUPEX
European Pilot for Exascale

SEANERGYS

Conclusions

LLview…

- …helps to identify (user and system) software and hardware issues
- …enables users find problems on their own
- …brings support to a higher level
- …reduces the need for communication between users and support as well as support and admins
- …improves the overall system utilisation
- …is ready for exascale!

Conclusions

Thank you!

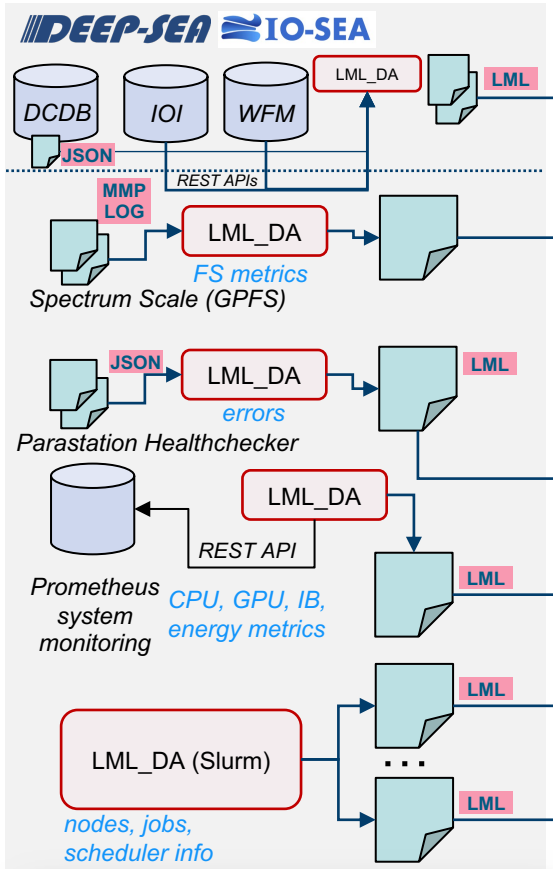
LLview…

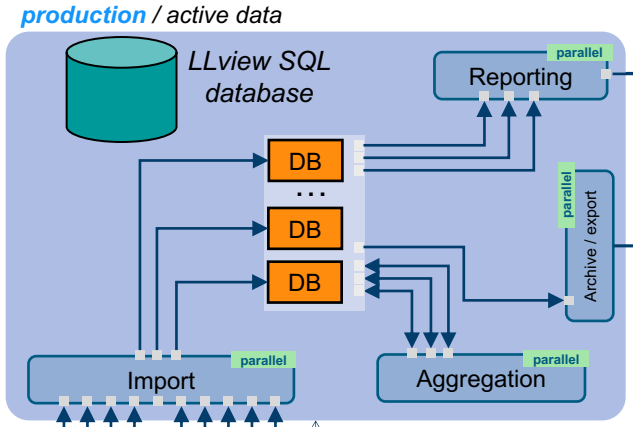
- …helps to identify (user and system) software and hardware issues
- …enables users find problems on their own
- …brings support to a higher level
- …reduces the need for communication between users and support as well as support and admins
- …improves the overall system utilisation
- …is ready for exascale!

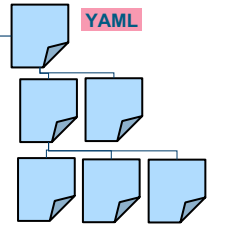
- Open source GPLv3 Licence github.com/FZJ-JSC/llview
- Website: llview.fz-juelich.de
- Contact: llview.jsc@fz-juelich.de

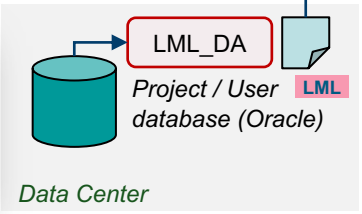
Extra slides

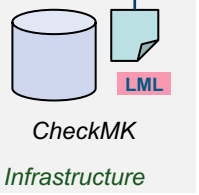
Even More Detailed Scheme of LLview

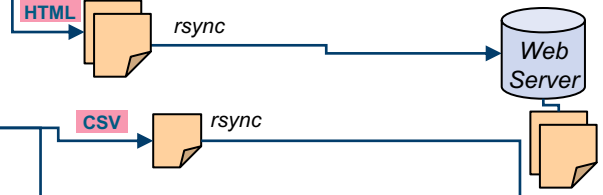


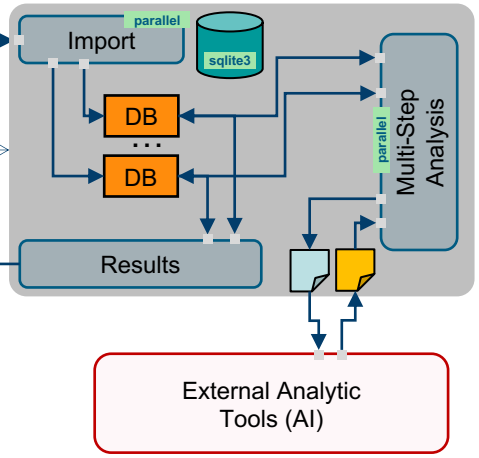


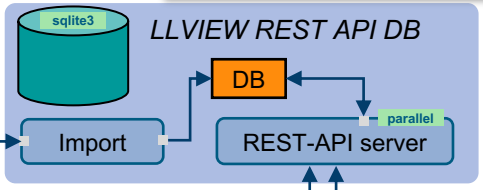


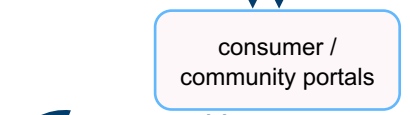


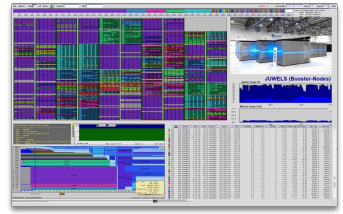












* in development

Examples

Inadequate Resource Usage

Usage	#Cores	#PhysCores	#LogicCores	Load
<50	filter	filter	filter	filter
2.90	0.01	0.01	0.00	512.26
1.37	0.37	0.27	0.10	477.35
4.07	0.01	0.01	0.00	395.68
0.68	0.00	0.00	0.00	158.31
3.00	3.73	3.73	0.00	149.31
18.90	23.98	23.98	0.00	128.64
10.49	0.00	0.00	0.00	128.32
18.91	24.00	24.00	0.00	127.77
18.86	23.94	23.94	0.00	127.68
18.88	23.98	23.98	0.00	127.68
30.34	78.75	78.75	0.00	124.96

Examples

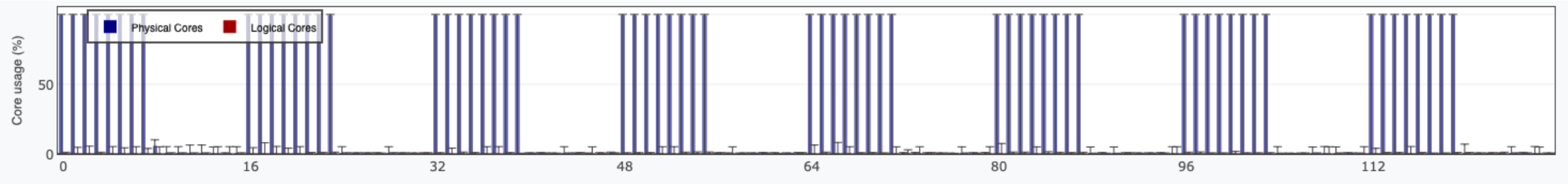
Inadequate Resource Usage

Usage	#Cores	#PhysCores	#LogicCores	Load
<50	filter	filter	filter	filter
2.90	0.01	0.01	0.0	512.26
1.37	0.37	0.27	0.1	477.35
4.07	0.01	0.01	0.0	395.68
0.68	0.00	0.00	0.00	158.31
3.00	3.73	3.73	0.00	149.31
18.90	23.98	23.98	0.00	128.64
10.49	0.00	0.00	0.00	128.32
18.91	24.00	24.00	0.00	127.77
18.86	23.94	23.94	0.00	127.68
18.88	23.98	23.98	0.00	127.68
30.34	78.75	78.75	0.00	124.96

Examples

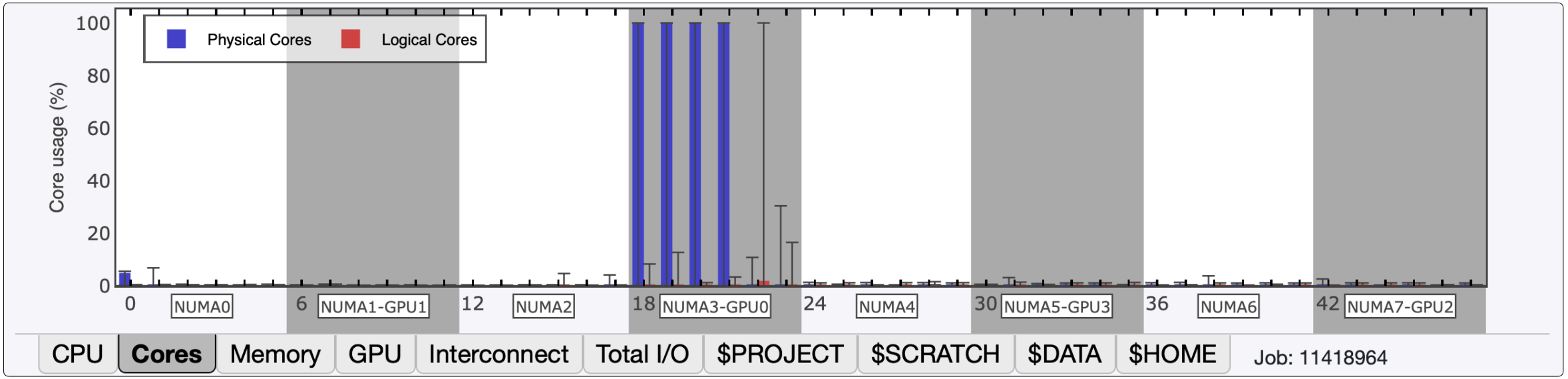
Inadequate Resource Usage

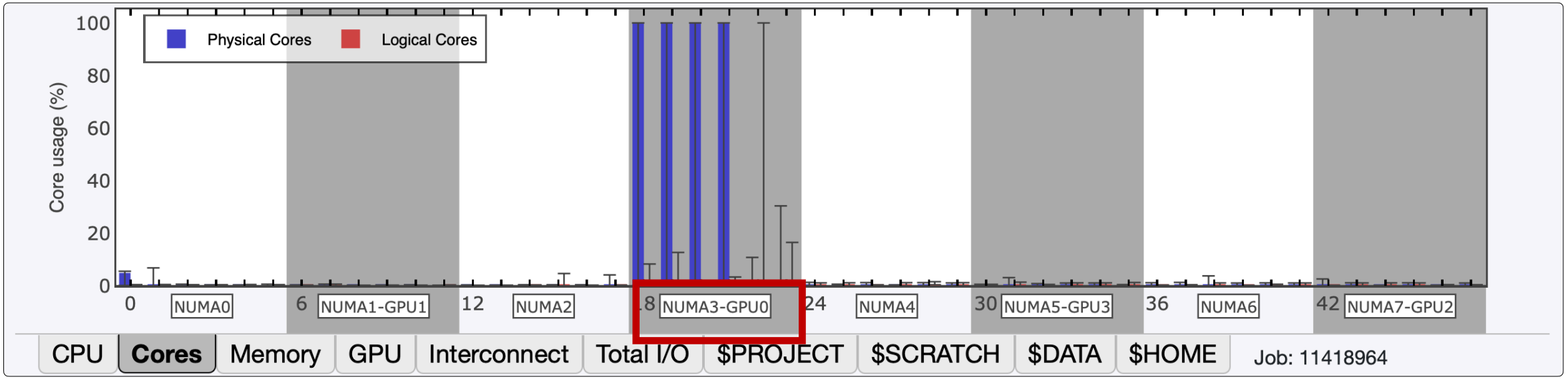
Usage	#Cores	#PhysCores	#LogicCores	Load
<50	filter	filter	filter	filter
2.90	0.01	0.01	0.0	512.26
1.37	0.37	0.27	0.1	477.35
4.07	0.01	0.01	0.0	395.68
0.68	0.00	0.00	0.00	158.31
3.00	3.73	3.73	0.00	149.31
18.90	23.98	23.98	0.00	128.64
10.49	0.00	0.00	0.00	128.32
18.91	24.00	24.00	0.00	127.77
18.86	23.94	23.94	0.00	127.68
18.88	23.98	23.98	0.00	127.68
30.34	78.75	78.75	0.00	124.96



Examples

Inadequate Resource Usage



Examples

Inadequate Resource Usage