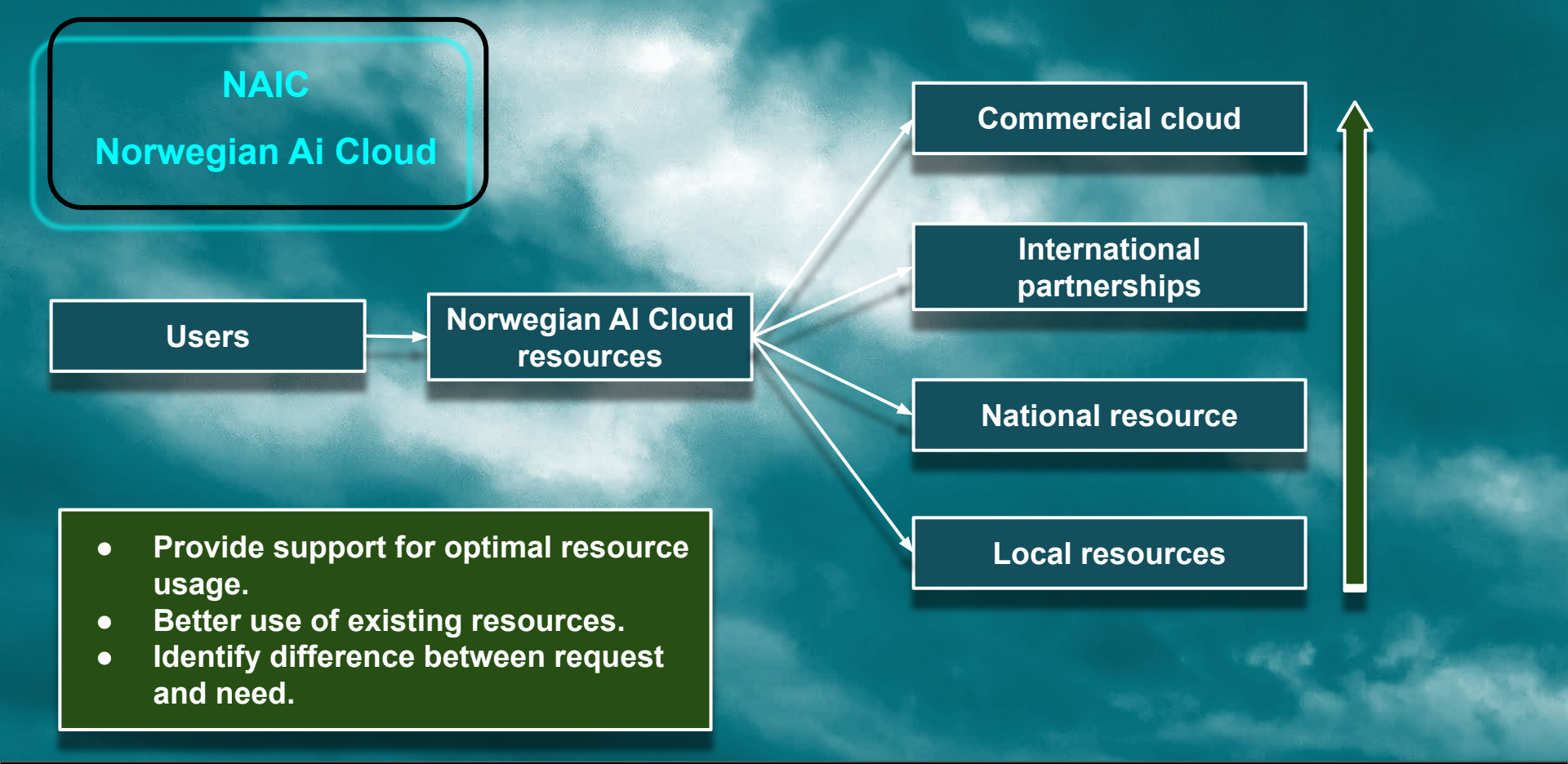
Jobanalyzer + sonar

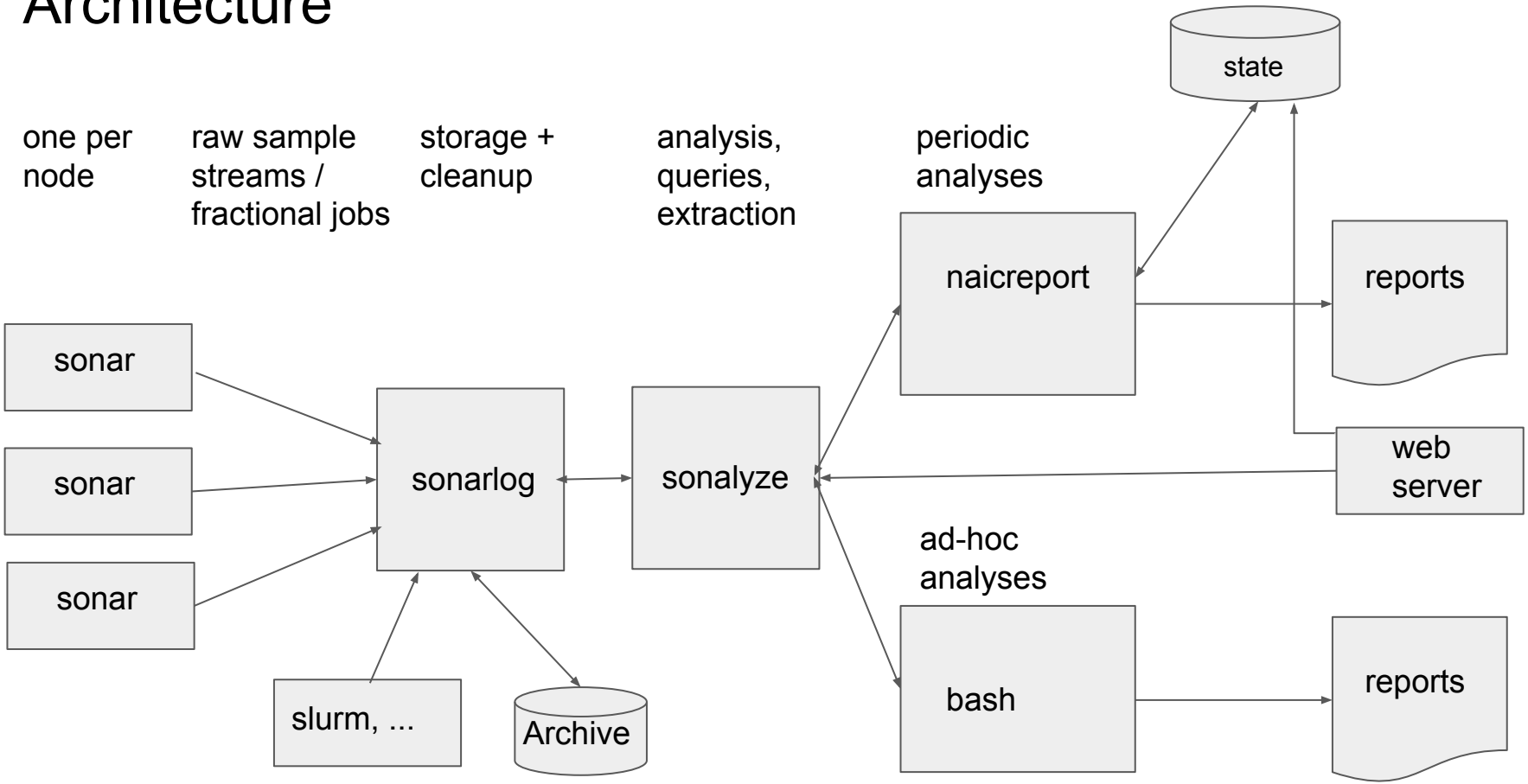
Sample-based systems and jobs monitoring

Lars T Hansen, NAIC / UiO, 30 November 2023

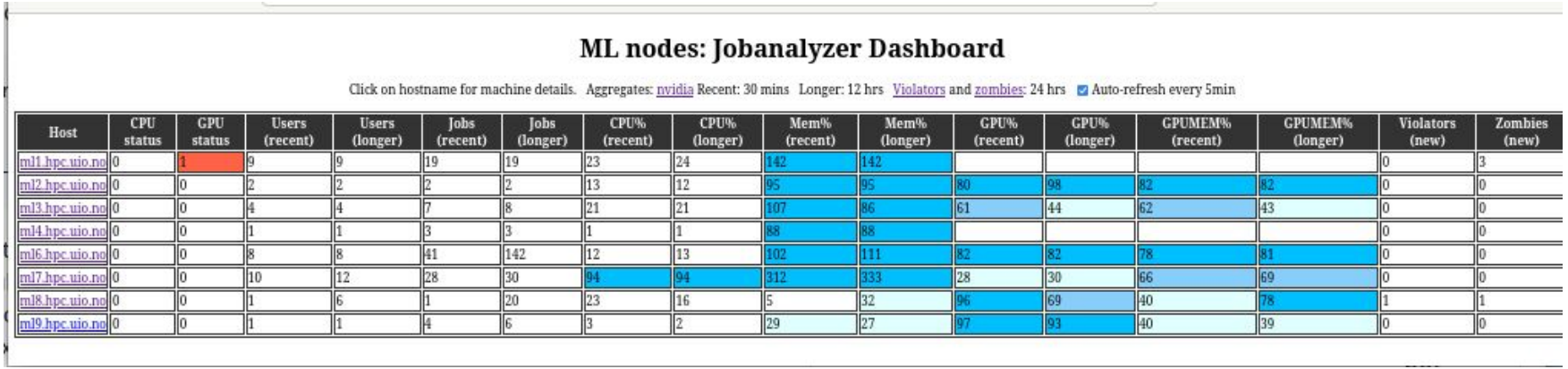


Norwegian AI Cloud

Architecture



Sysadmin dashboard



ML nodes: Jobanalyzer Dashboard

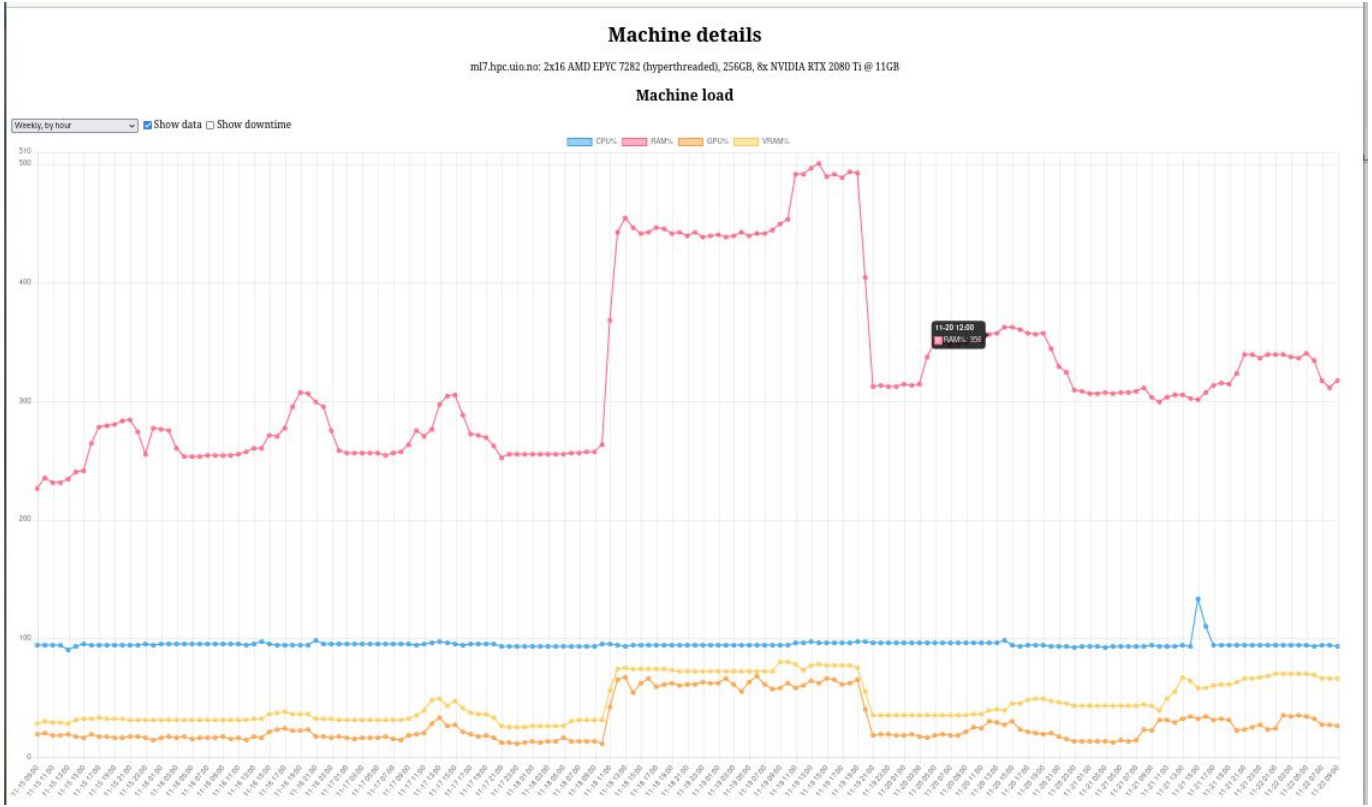
Click on hostname for machine details. Aggregates: nvidia Recent: 30 mins Longer: 12 hrs Violators and zombies: 24 hrs ☑ Auto-refresh every 5min

Host	CPU status	GPU status	Users (recent)	Users (longer)	Jobs (recent)	Jobs (longer)	CPU% (recent)	CPU% (longer)	Mem% (recent)	Mem% (longer)	GPU% (recent)	GPU% (longer)	GPUMEM% (recent)	GPUMEM% (longer)	Violators (new)	Zombies (new)
ml1.hpc.uio.no	0	1	9	9	19	19	23	24	142	142					0	3
ml2.hpc.uio.no	0	0	2	2	2	2	13	12	95	95	80	98	82	82	0	0
ml3.hpc.uio.no	0	0	4	4	7	8	21	21	107	86	61	44	62	43	0	0
ml4.hpc.uio.no	0	0	1	1	3	3	1	1	88	88					0	0
ml6.hpc.uio.no	0	0	8	8	41	142	12	13	102	111	82	82	78	81	0	0
ml7.hpc.uio.no	0	0	10	12	28	30	94	94	312	333	28	30	66	69	0	0
ml8.hpc.uio.no	0	0	1	6	1	20	23	16	5	32	96	69	40	78	1	1
ml9.hpc.uio.no	0	0	1	1	4	6	3	2	29	27	97	93	40	39	0	0

Memory occupancy seems to be high, esp on ml7 (and gpu on ml1 is down)

Has the memory use been high for some time? Let's look.

Individual system utilization



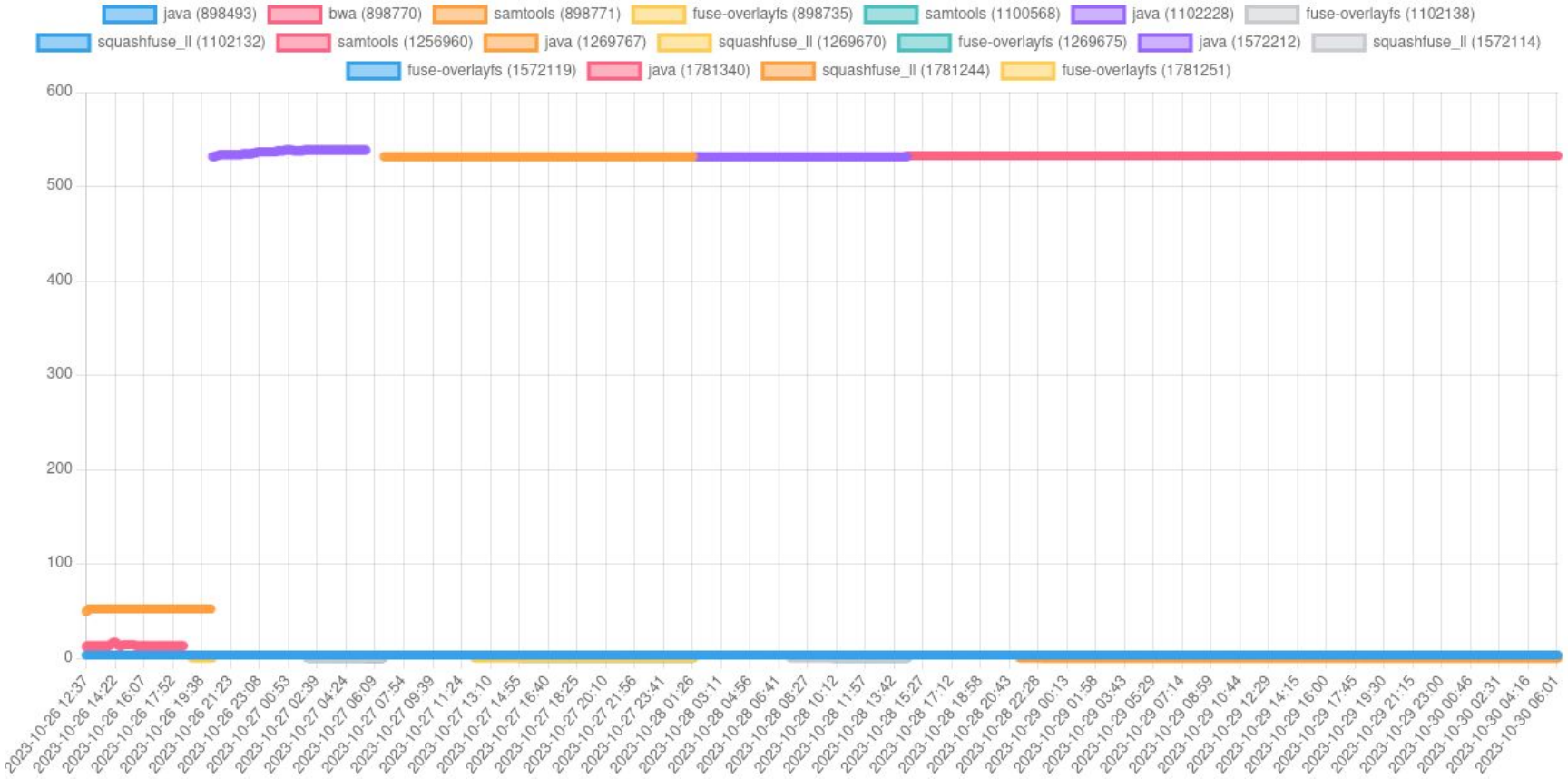
Drill down on a particular host and jobs

Utilization of ML7 - very high memory occupancy and CPU utilization over time

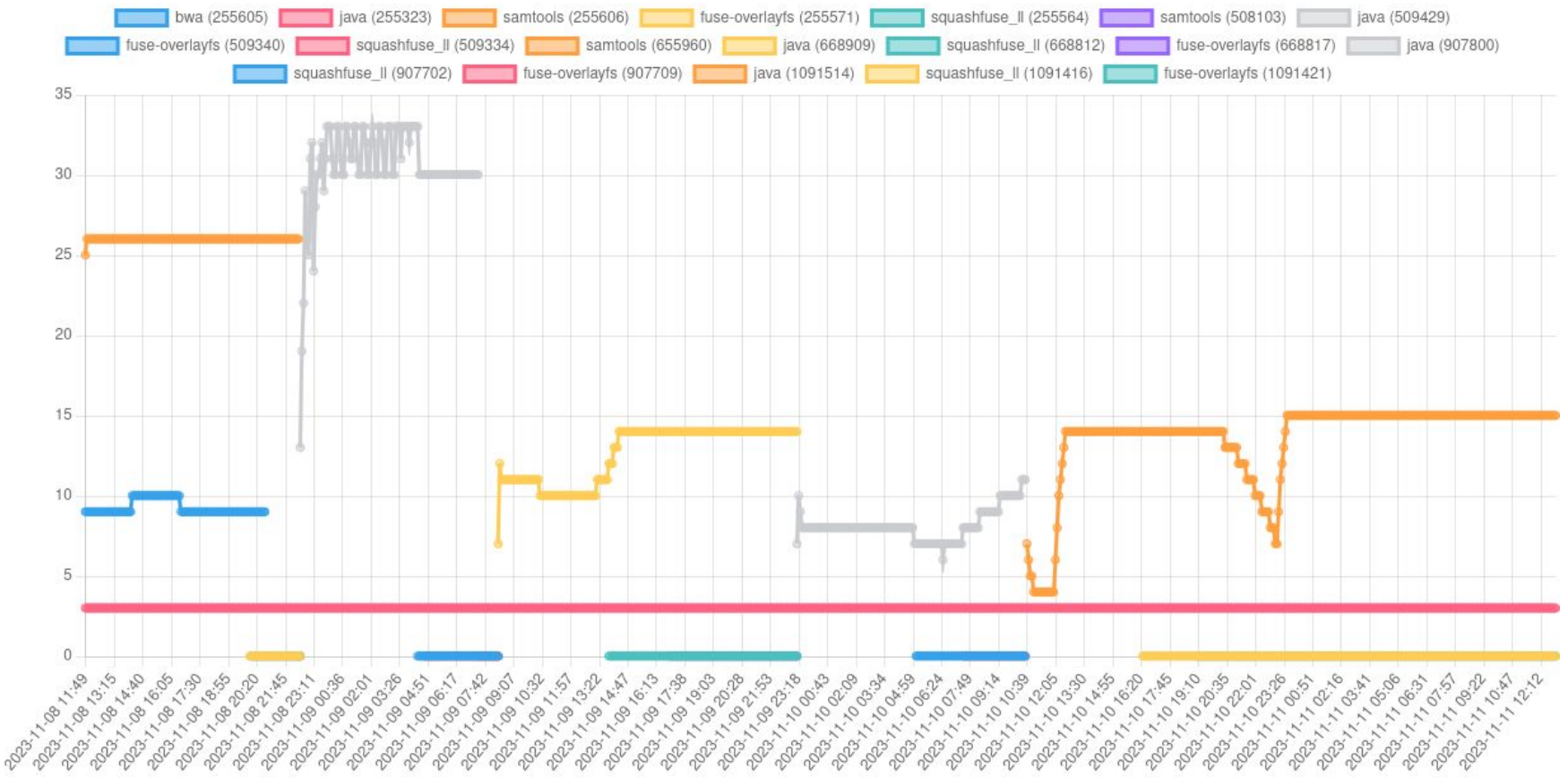
Which are the jobs? We can go to the command line...

```
$ sonalyze jobs --host ml7 -f7d --min-mem-peak 64 -u- --fmt=std,cpu,mem,cmd
```

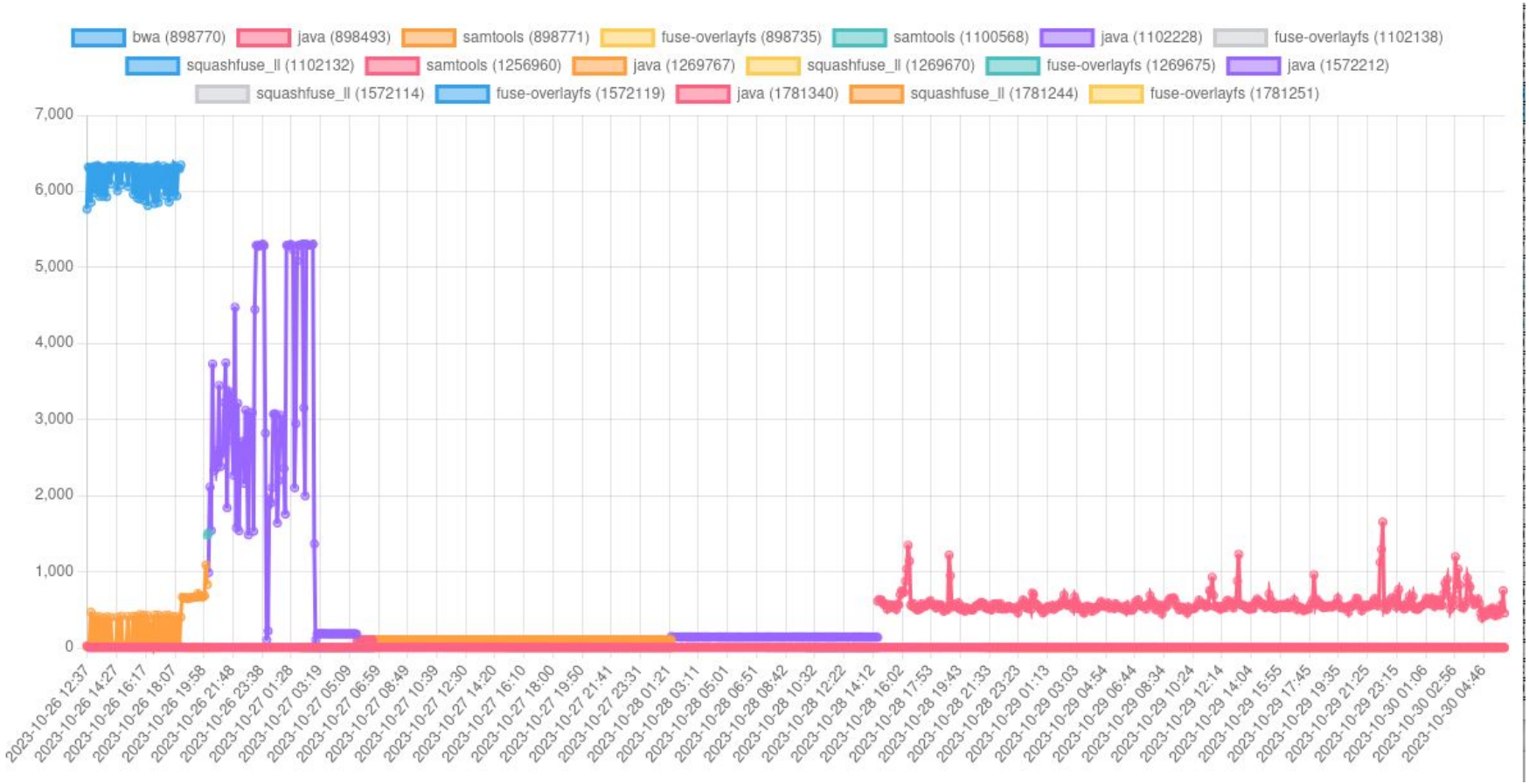
jobm	user	duration	host	cpu-avg	cpu-peak	mem-avg	mem-peak	cmd
3384712>	annammc	6d23h55m	ml7	992	2581	128	192	scripts.train
3386163>	annammc	6d23h55m	ml7	83	116	175	199	scripts.train,scripts.train <defunct>
3669438>	annammc	6d23h55m	ml7	1	1	285	285	scripts.preproc
4146598	hughav	0d 5h20m	ml7	295	565	66	82	python
92681	hughav	0d 3h20m	ml7	121	481	59	83	python3.9
428872	hughav	0d 5h15m	ml7	96	680	60	80	python3.9
751224	johanfag	0d 0h10m	ml7	472	488	71	87	python
757710	johanfag	0d 3h 0m	ml7	490	560	86	87	python
1156151	balintl	1d 9h 0m	ml7	469	537	460	538	python3



Job uses 512GB of RAM for 3 days...



Turns out, it doesn't need much memory at all



There's a tail of 3 days of 1-5 core execution...

Summary

System view and job view (and more)

Resource use of job that's already completed

- "what just happened?"
- job utilization vs requested allocation

Scalability prediction

- "will this run faster on a bigger system?"
- "am i using the right hardware?"

Policy violations

Violators last 30 days

The following jobs have violated usage policy and are probably not appropriate to run on this cluster. The list is recomputed at noon and midnight and goes back four weeks.

By user

User	No. violations	First seen	Last seen
ahmetyi	4	2023-11-28 13:00	2023-11-29 11:02
einarvid	8	2023-11-01 11:05	2023-11-29 11:02
karths	9	2023-11-01 09:15	2023-11-21 13:02
mateuwa	1	2023-11-21 13:10	2023-11-22 15:02
tsauren	8	2023-11-04 15:30	2023-11-11 19:02

Alerts about new violations appear in dashboard and are sent to admins

User-specific policy violation report

ML nodes individual policy violator report

Hi,

This is a message from your friendly UiO systems administrator.

To ensure that computing resources are used in the best possible way,
we monitor how jobs are using the systems and ask users to move when
they are using a particular system in a way that is contrary to the
intended use of that system.

You are receiving this message because you have been running jobs
in just such a manner, as detailed below. Please apply the suggested
remedies (usually this means moving your work to another system).

"ML nodes" individual policy violator report for host ml8.hpc.uio.no

Report generated on Wed Nov 29 2023 15:40:32 GMT+0100 (Central European Standard Time)

User:
  ahmetyi

Policies violated:
  ml-cpuhog:
    Trigger: Job uses more than 10% of system's CPU at peak, runs for at least 10 minutes, and uses no GPU at all
    Problem: ML nodes are for GPU jobs. Job is in the way of other jobs that need GPU
    Remedy: Move your work to a GPU-less system such as Fox or Light-HPC

(Times below are UTC, job numbers are derived from session leader if not running under Slurm)

Host	Job	Policy	First seen	Last seen	CPU% avg	CPU% peak	Mem% avg	Command
ml8	2578076	ml-cpuhog	2023-11-28 13:00	2023-11-29 11:02	42	93	1	java,python3,ud2drs-exe
ml8	2572446	ml-cpuhog	2023-11-28 13:00	2023-11-29 09:02	45	93	1	java,python3,ud2drs-exe
ml8	2561729	ml-cpuhog	2023-11-28 13:00	2023-11-28 19:02	49	93	1	java,python3,ud2drs-exe
ml8	2619101	ml-cpuhog	2023-11-28 13:00	2023-11-28 15:02	45	85	1	java,python3,ud2drs-exe

Status

Operational at UiO

- ML nodes since August, utility is fairly obvious
- Fox since mid-November, utility TBD

Use cases and workflows are still evolving

Prototype - not solid production code

Focus for 2023Q4 and 2024Q1 will be exploration & features, not stability & perf

Where to look

NAIC homepage: https://www.naic.no/

UiO dashboards: http://158.39.48.160

Jobanalyzer source: https://github.com/NAICNO/Jobanalyzer

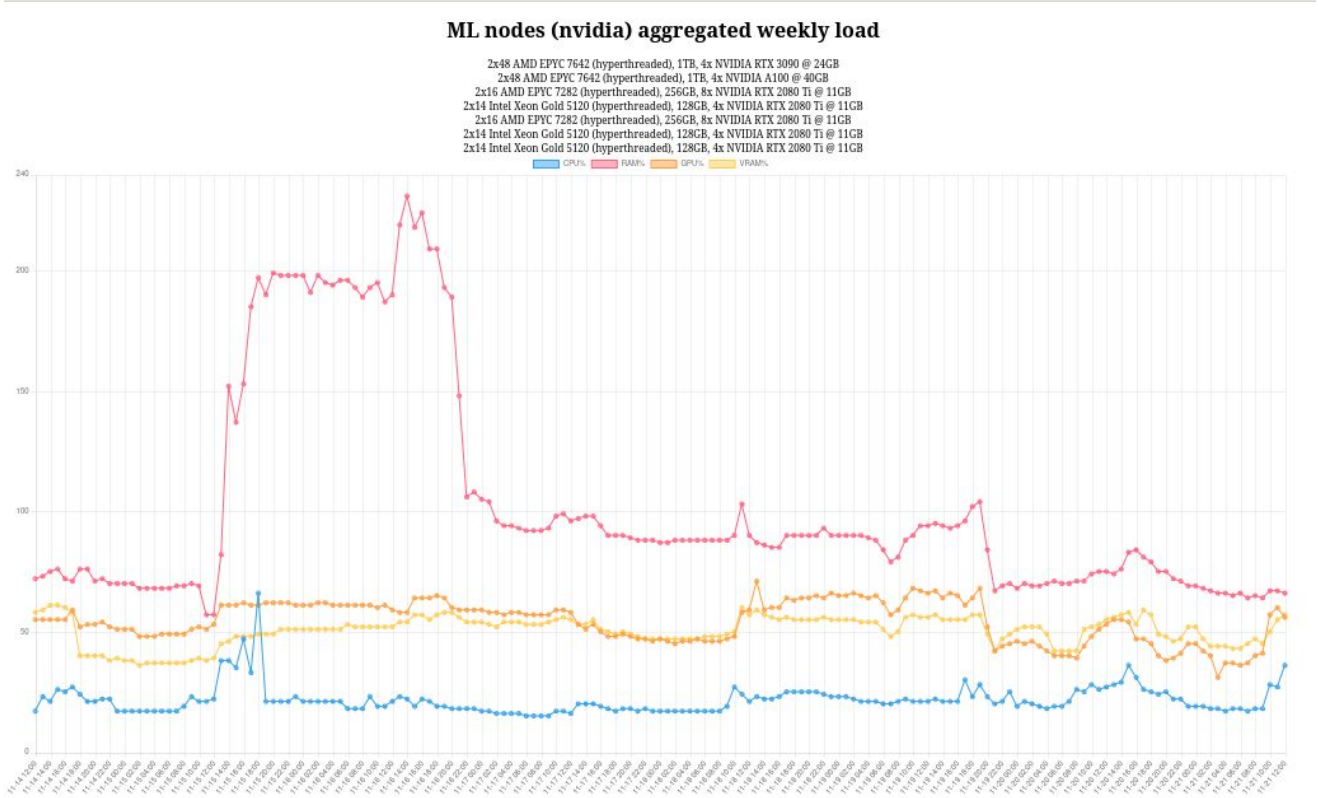
Sonar source: https://github.com/NordicHPC/sonar

ML nodes policy violators

The following jobs have been running significantly outside of policy and are probably not appropriate to run on this cluster. The list is recomputed at noon and midnight and goes back four weeks.

Host	User	Job	Command	First seen	Last seen	CPU peak	CPU% avg	CPU% peak	Mem% avg	Mem% peak
ml4.hpc.uio.no	bendimol	3499948	gcs_server, python, ray::IDLE, ray::ReplayBuff, ray::SelfPlay.c, ray::SharedStor, ray::Trainer.co, raylet	2023-11-09 16:05	2023-11-09 21:02	6	6	11	29	34
ml4.hpc.uio.no	bendimol	3499487	gcs_server, python, ray::IDLE, ray::ReplayBuff, ray::SelfPlay.c, ray::SharedStor, ray::Trainer.co, raylet	2023-11-09 16:05	2023-11-09 21:02	7	4	12	28	34
ml4.hpc.uio.no	bendimol	3483207	python, ray::IDLE, ray::ReplayBuff, ray::SelfPlay.c, ray::SharedStor, ray::Trainer.co	2023-11-09 15:25	2023-11-09 21:02	6	6	11	23	27
ml4.hpc.uio.no	bendimol	3482865	python, ray::IDLE, ray::SelfPlay.c, ray::Trainer.co	2023-11-09 15:25	2023-11-09 21:02	6	3	10	22	27
ml1.hpc.uio.no	annammc	2232869	scripts.train	2023-10-31 21:25	2023-11-09 21:02	27	49	49	128	131
ml1.hpc.uio.no	hermankn	3943720	jupyter-lab	2023-11-08 10:20	2023-11-09 13:02	5	1	11	1	1
ml4.hpc.uio.no	einarvid	2290247	python3	2023-11-06 05:30	2023-11-07 13:02	28	39	45	123	208
ml8.hpc.uio.no	einarvid	2891150	python3	2023-11-06 07:05	2023-11-07 13:02	20	10	11	16	66
ml8.hpc.uio.no	tsauren	2125648	python3	2023-11-06 10:50	2023-11-07 11:02	170	88	90	9	9
ml4.hpc.uio.no	einarvid	1990250	python3	2023-11-05 21:00	2023-11-06 23:02	29	43	47	131	200
ml8.hpc.uio.no	einarvid	202209	python3	2023-11-05 21:05	2023-11-06 23:02	21	9	12	15	65
ml8.hpc.uio.no	tsauren	1584229	python3	2023-11-05 21:50	2023-11-06 21:02	170	68	89	7	9
ml4.hpc.uio.no	einarvid	1917640	python3	2023-11-05 20:25	2023-11-06 19:02	28	41	45	100	113
ml4.hpc.uio.no	einarvid	1056893	python3	2023-11-05 00:15	2023-11-06 15:02	36	52	58	159	261
ml8.hpc.uio.no	tsauren	647655	python3	2023-11-05 15:10	2023-11-06 15:02	56	23	30	13	14
ml8.hpc.uio.no	tsauren	1091062	python3	2023-11-04 17:50	2023-11-06 13:02	187	93	98	10	10
ml8.hpc.uio.no	tsauren	2332979	python3	2023-11-05 13:20	2023-11-06 13:02	45	23	24	6	6
ml8.hpc.uio.no	tsauren	74896	python3	2023-11-04 15:30	2023-11-05 17:02	175	91	92	9	9
ml8.hpc.uio.no	einarvid	2492655	python3	2023-11-04 14:00	2023-11-05 13:02	45	23	24	16	17
ml8.hpc.uio.no	einarvid	2781066	python3	2023-11-01 11:05	2023-11-05 07:02	168	19	88	114	133
ml4.hpc.uio.no	einarvid	4033338	python3	2023-11-01 11:10	2023-11-04 05:02	51	73	81	350	445
ml7.hpc.uio.no	einarvid	3597329	python3	2023-11-01 11:05	2023-11-02 21:02	31	45	50	104	108
ml8.hpc.uio.no	karths	1503174	jupyter-lab	2023-11-01 09:15	2023-11-02 11:02	73	6	39	1	1
ml6.hpc.uio.no	smrashid	1186151	jupyter-lab	2023-10-31 22:05	2023-11-02 01:02	18	1	29	1	1
ml4.hpc.uio.no	bendimol	3842155	raylet, ray::Trainer.co, ray::SharedStor, ray::IDLE, python, ray::SelfPlay.c, ray::ReplayBuff	2023-10-31 10:35	2023-11-01 09:02	6	11	11	29	34
ml4.hpc.uio.no	bendimol	3676778	ray::Trainer.co, ray::SharedStor, ray::ReplayBuff, raylet, ray::CPUActor, ray::SelfPlay.c, gcs_server, ray::IDLE, python	2023-10-30 12:55	2023-10-31 13:02	6	3	11	29	34
ml8.hpc.uio.no	mateuwa	1691852	jupyter-lab	2023-10-30 10:35	2023-10-31 09:02	46	5	24	1	1
ml6.hpc.uio.no	daniehh	212307	python3	2023-10-30 07:00	2023-10-31 07:02	6	11	11	1	1
ml8.hpc.uio.no	einarvid	943573	python3	2023-10-28 15:05	2023-10-29 17:02	30	14	17	14	20
ml3.hpc.uio.no	johanfag	2936544	python	2023-10-28 10:25	2023-10-29 15:02	13	23	24	13	14
ml4.hpc.uio.no	einarvid	3226261	python3	2023-10-28 15:15	2023-10-29 15:02	37	56	60	157	229
ml6.hpc.uio.no	daniehh	4010362	python3	2023-10-28 05:30	2023-10-29 05:02	7	11	12	1	1
ml8.hpc.uio.no	einarvid	1138849	python3	2023-10-25 20:00	2023-10-27 20:02	71	30	38	76	120
ml2.hpc.uio.no	einarvid	2826710	python3	2023-10-23 09:05	2023-10-26 14:02	32	51	58	132	134
ml7.hpc.uio.no	einarvid	1017602	python3	2023-10-23 08:50	2023-10-25 22:02	38	57	61	125	162

Weekly summary of cluster utilization



Memory's a thing (red line), peaking at 230% cluster-wide, generally 80-100%

Sysadmin use cases

System status

Utilization, unused capacity, overload

Guiding users to the right system for their job

Reporting historical system status, uptime, utilization

Reporting the software being used (subsumes Appusage)

(...)

Future implementation

Multiple data sources (system config observers, slurm observers, ...)

No shared disk, send data by message instead

Better database, better representations

Web server can perform on-line queries

Access control: users should only see their own jobs and aggregate info

Scalable UI for Big Systems

Implementation

one per node

raw sample streams (fractional jobs), csv, APPEND ONLY

queries go via sonarlog and sonalyze

periodic analyses