Docker, Monitoring and SLURM Specific Visualisations

QNIBTerminal @ work

Agenda

- Docker in a Nutshell
- QNIBx
  - Terminal
  - Monitoring
  - Inventory
- SLURM Autogenerated Dashboards

About Me

- Christian Kniep
  - @CQnib, christian@qnib.org

About Me

- Christian Kniep
  - @CQnib, christian@qnib.org
- >10y Iteration
  - SysAdmin, SysOps, SysEngineer, R&D Engineer
  - DevOps @Locafox (hyper-scale web-service)

About Me

- Christian Kniep
  - @CQnib, christian@qnib.org
- >10y Iteration
  - SysAdmin, SysOps, SysEngineer, R&D Engineer
  - DevOps @Locafox (hyper-scale web-service)
- Founder of QNIB Solutions
  - Holistic System Management
  - Containerization of SysOps and Workload
  - Consultancy / Software Design & Development

Docker in a Nutshell

Multiple Guests

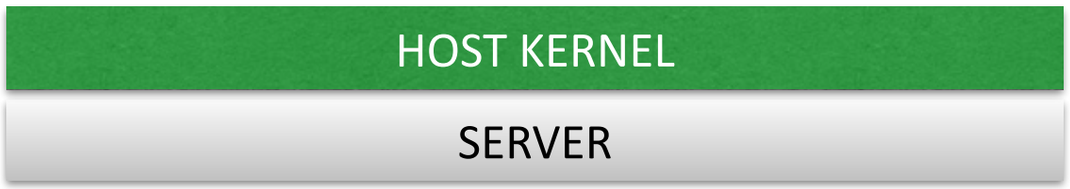
SERVER

Traditional Virtualisation

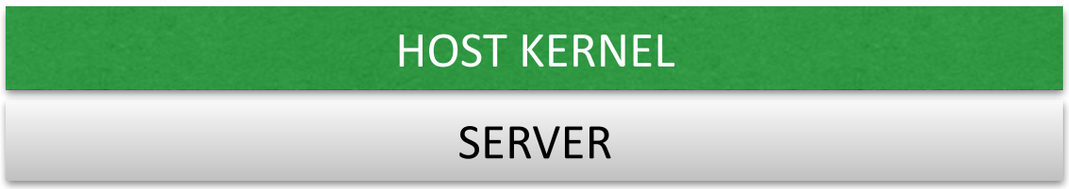
SERVER

Containerisation

Multiple Guests

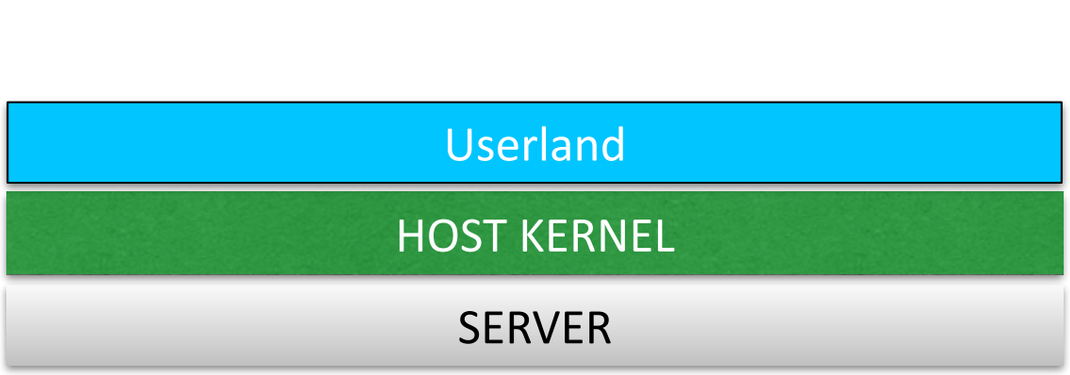


Traditional Virtualisation

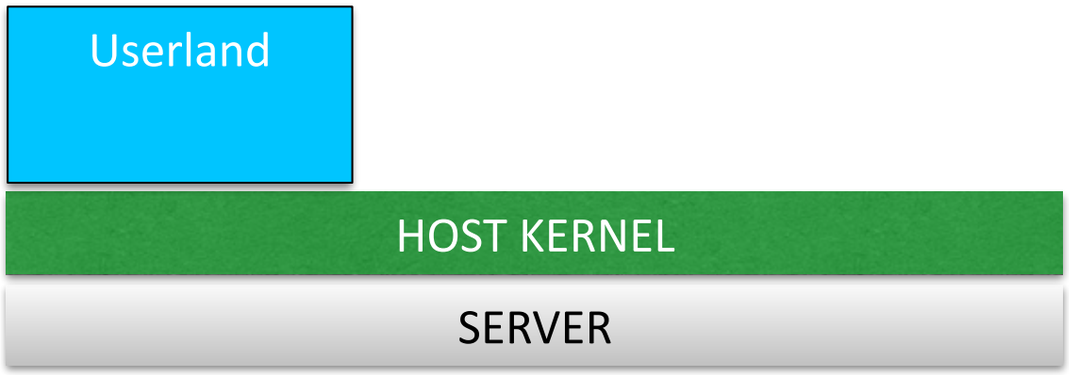


Containerisation

Multiple Guests

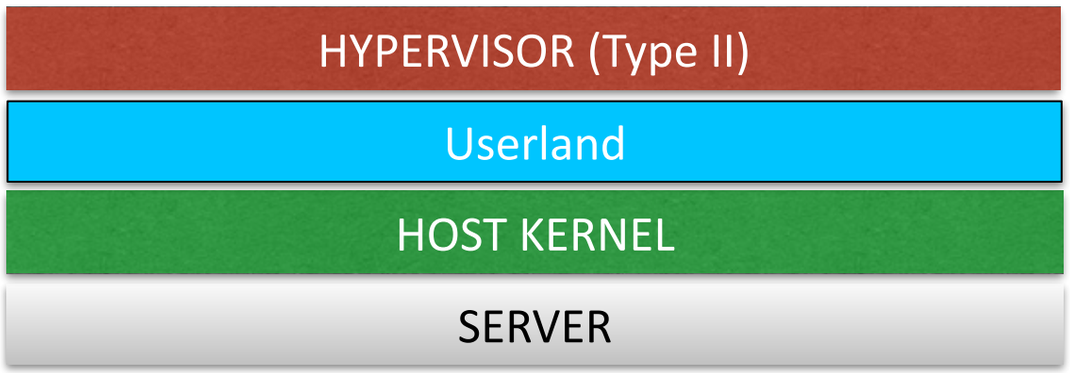


Traditional Virtualisation

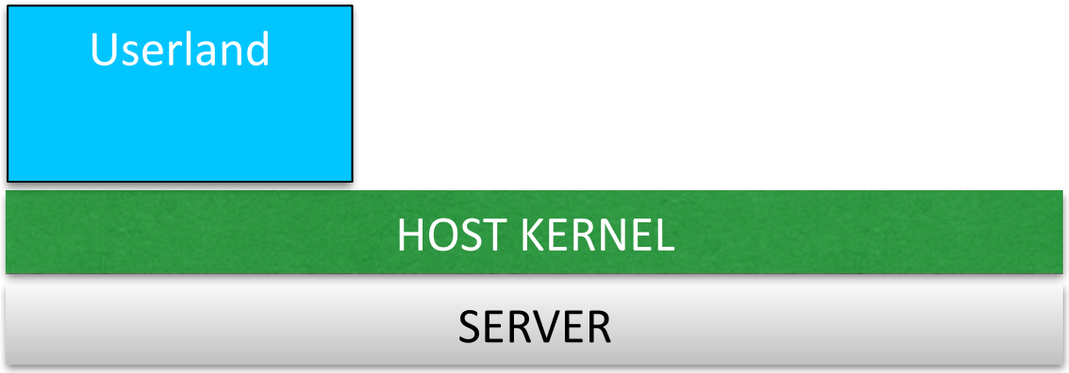


Containerisation

Multiple Guests

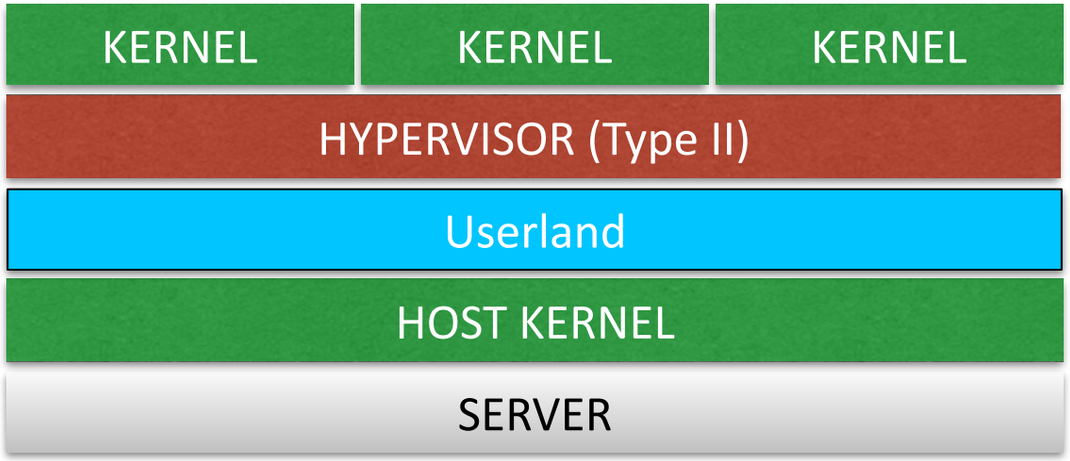


Traditional Virtualisation

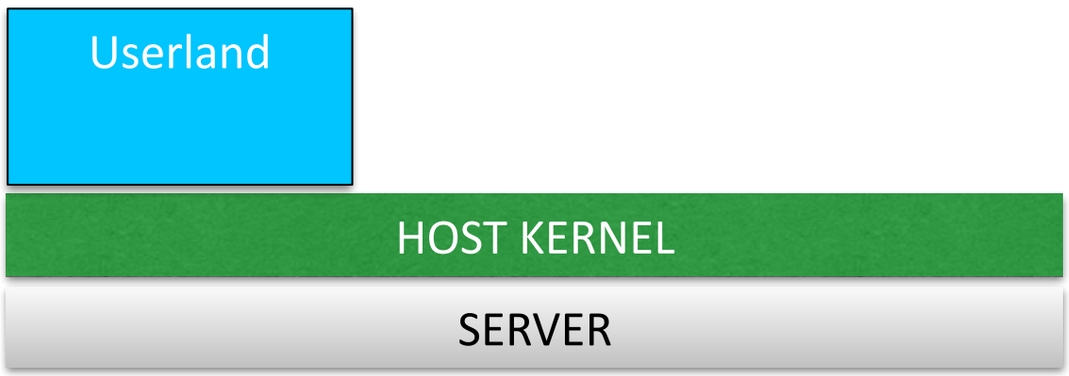


Containerisation

Multiple Guests

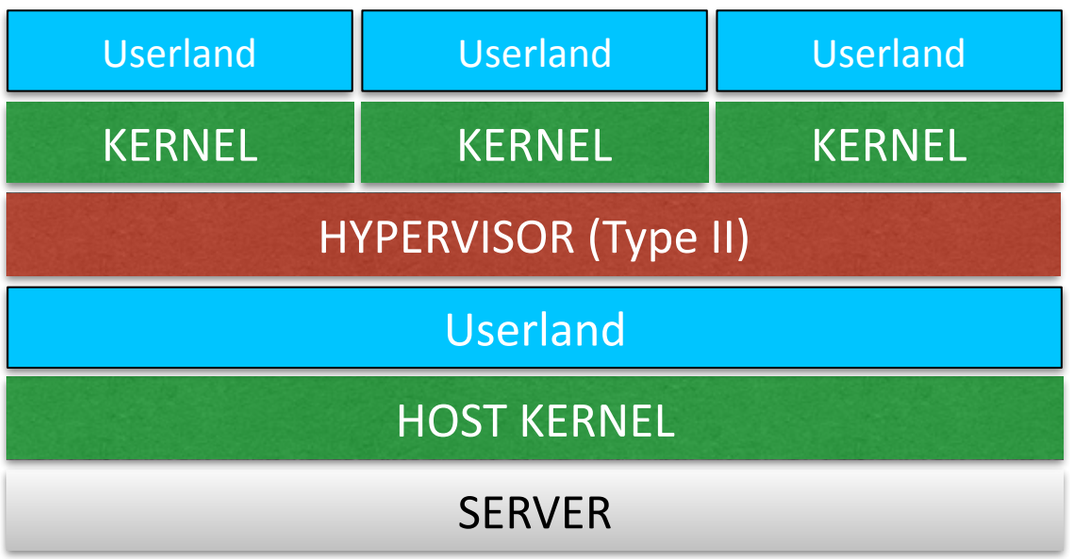


Traditional Virtualisation

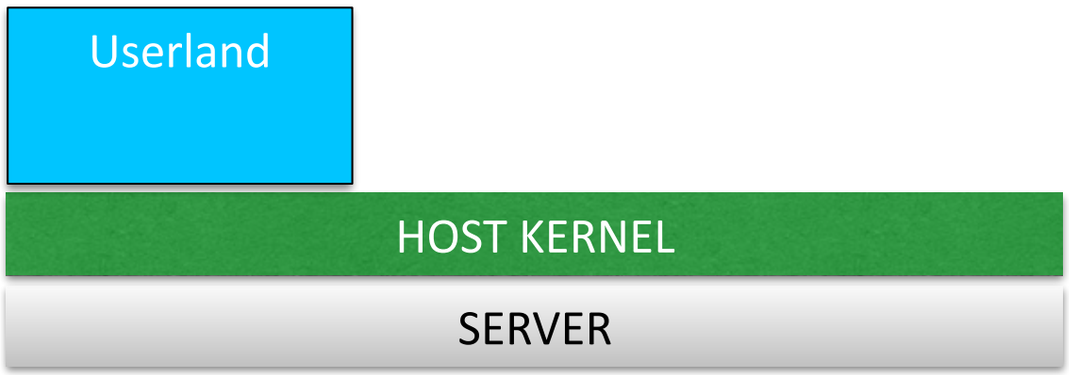


Containerisation

Multiple Guests

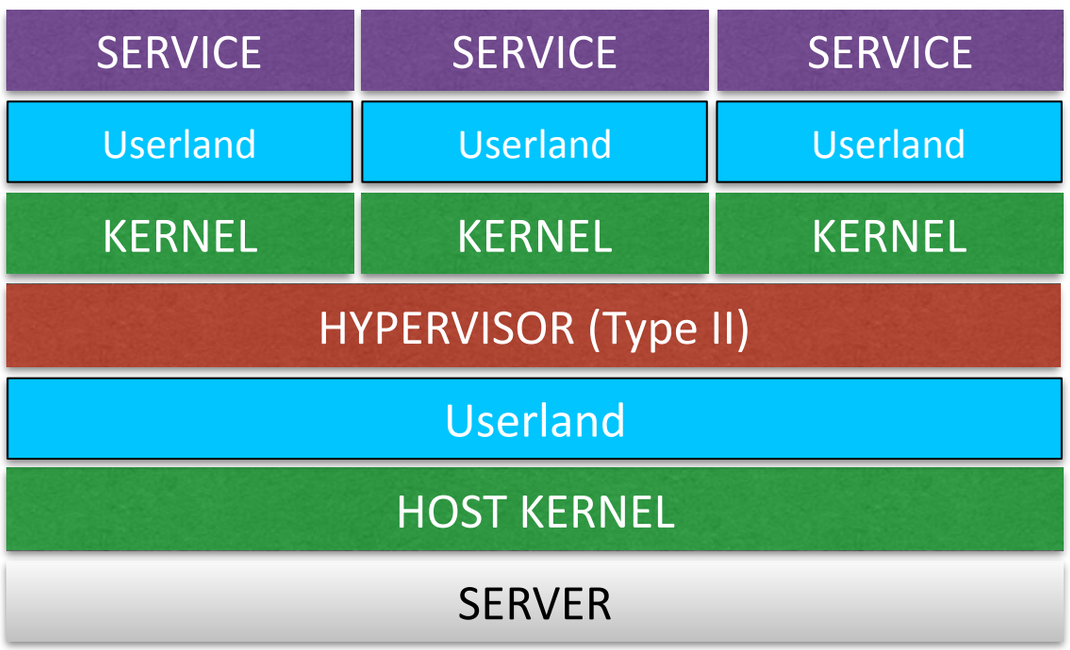


Traditional Virtualisation

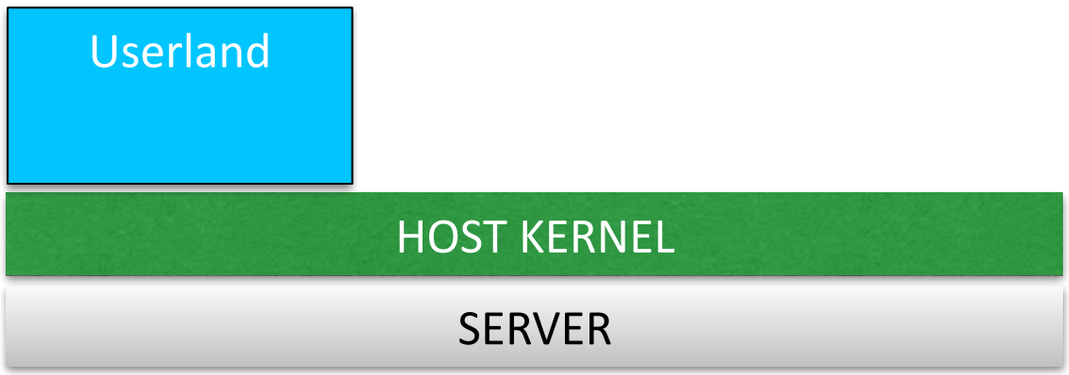


Containerisation

Multiple Guests

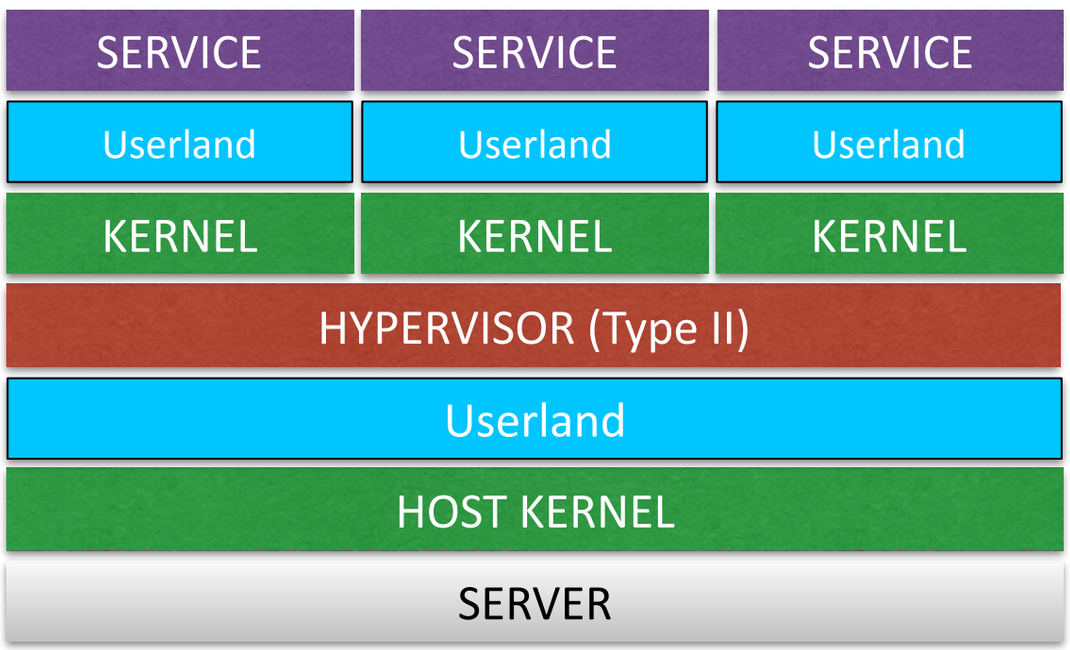


Traditional Virtualisation

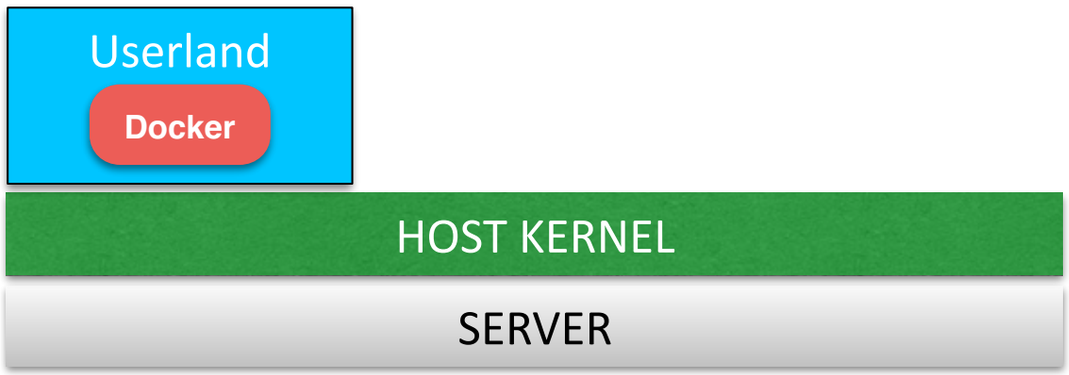


Containerisation

Multiple Guests

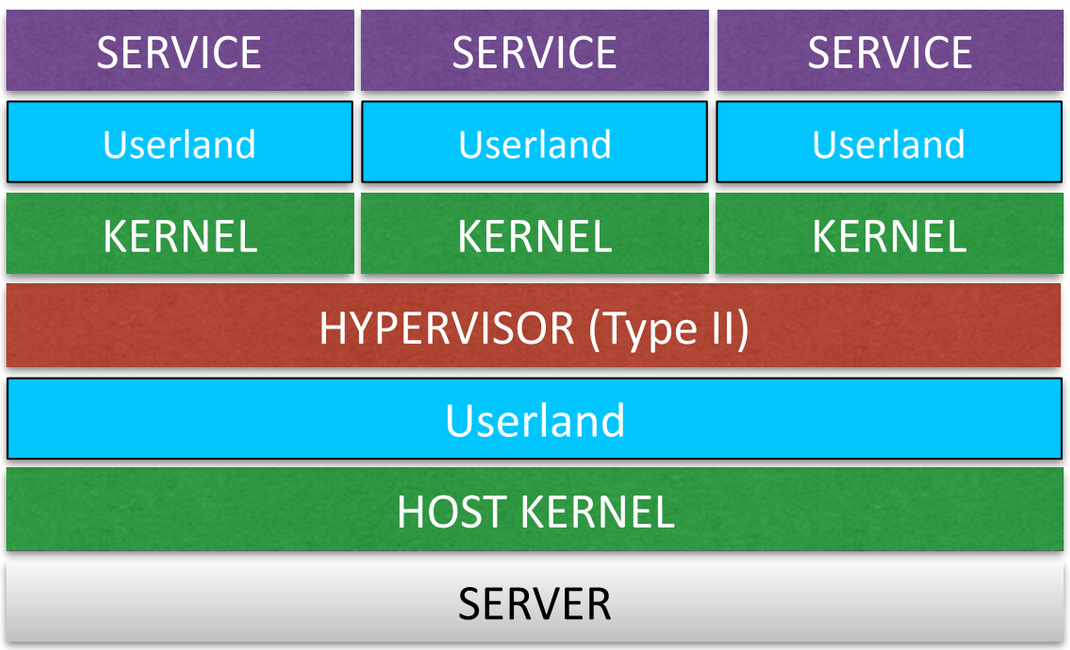


Traditional Virtualisation

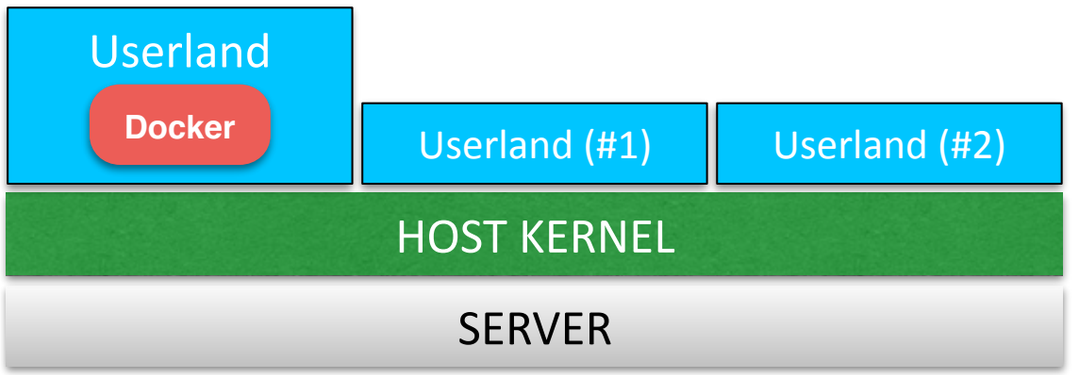


Containerisation

Multiple Guests

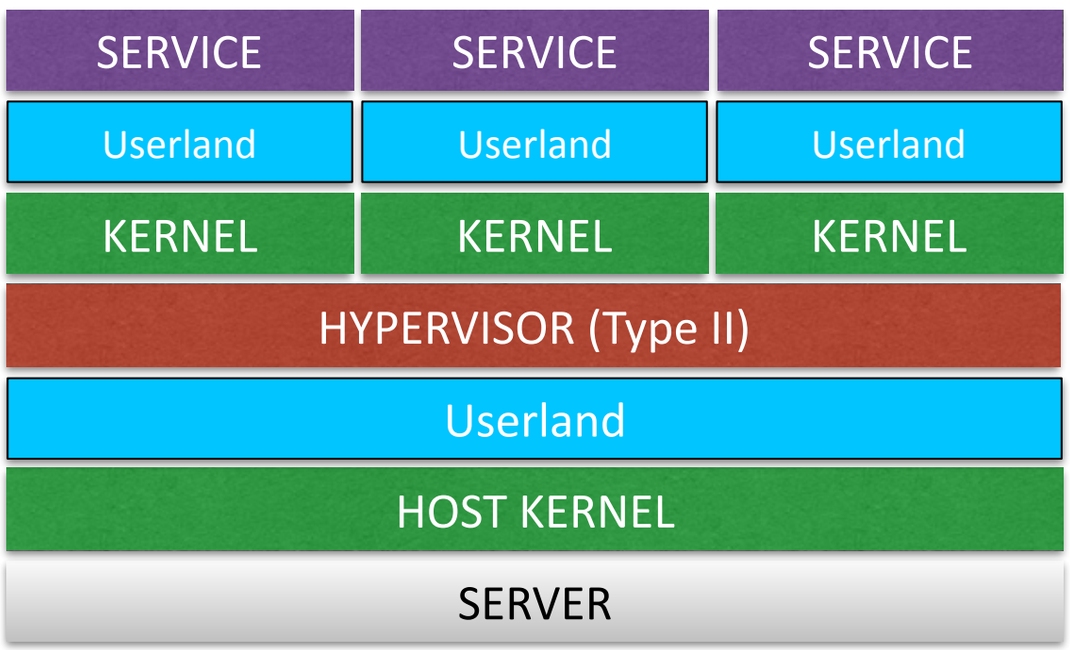


Traditional Virtualisation

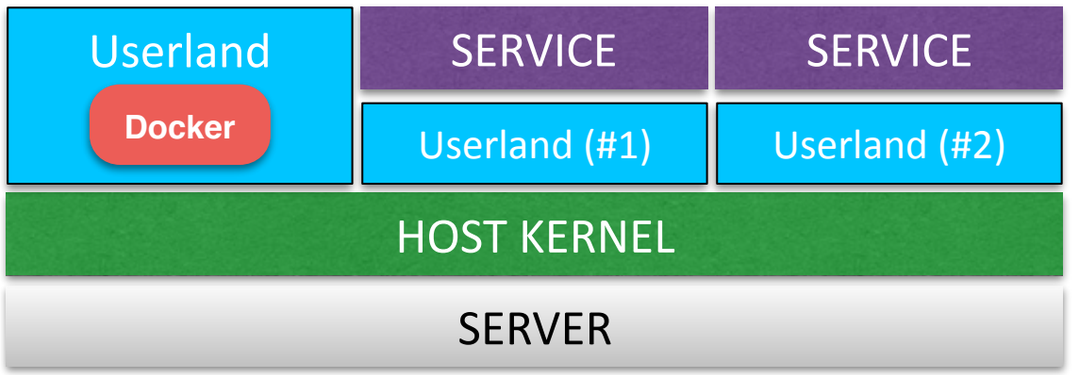


Containerisation

Multiple Guests



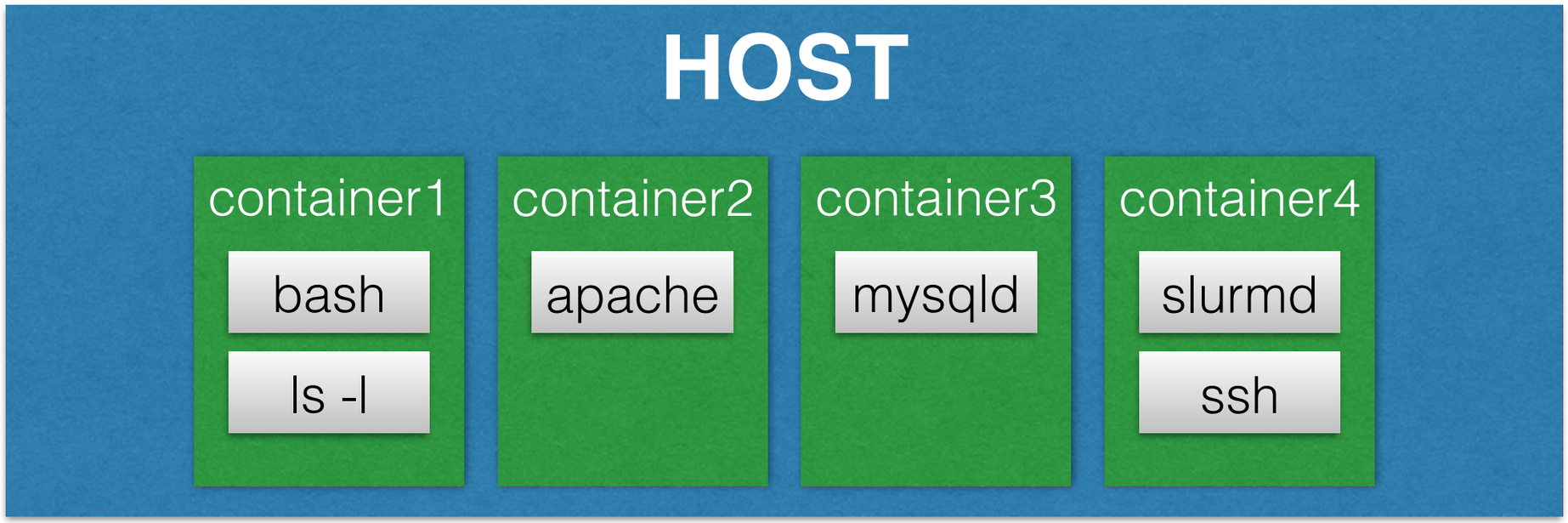
Traditional Virtualisation



Containerisation

Docker Internal View

- Containers are ‘grouped processes’
  - isolated by Kernel Namespaces (PID, network, mount, …)
  - resource restrictions applicable through CGroups

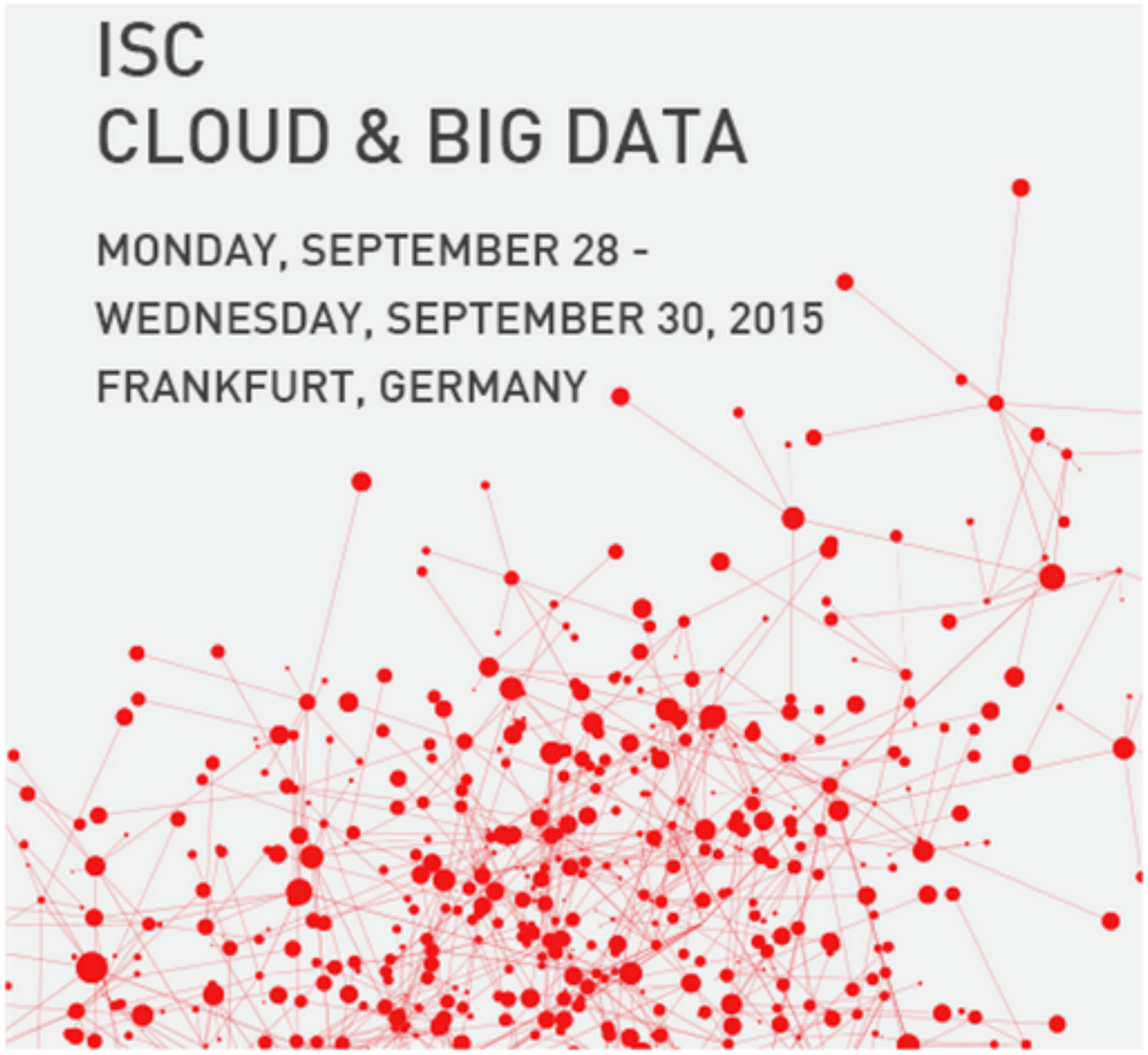


Docker Workshop

- 1/2 Day, July 16th @ISC High Performance
  - Deep dive into the talking points
  - How Docker might impact System Operations & HPC Applications
  - Further discussion beyond what I am talking about today

Docker Workshop #2

- Full Day, September 28th @ISC Cloud&BigData



QNIBTerminal

QNIBTerminal

- Framework of system container to spin up stacks
  - SLURM

QNIBTerminal

- Framework of system container to spin up stacks
  - SLURM

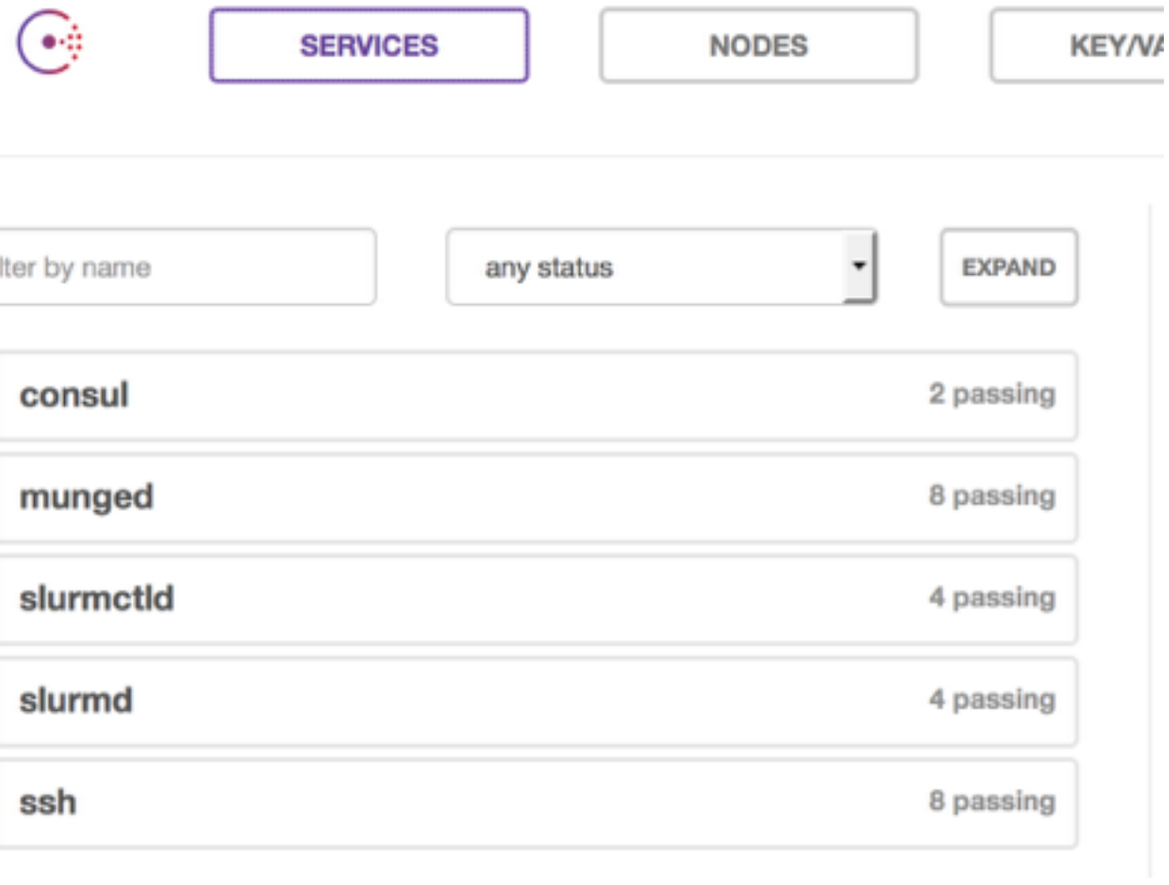
```
$ fig up -d
Creating qnibterminalslurm_consul_1...
Creating qnibterminalslurm_slurmctld_1...
Creating qnibterminalslurm_slurmd_1...
$
```

```
consul:
    image: qnib/consul
    ports:
     - "8500:8500"
    dns: 127.0.0.1
    hostname: consul
    privileged: true

slurmctld:
    image: qnib/slurmctld
    ports:
    - "6817:6817"
    links:
    - consul:consul
    dns: 127.0.0.1
    hostname: slurmctld
    privileged: true

slurmd:
    image: qnib/slurmd
    links:
    - consul:consul
    - slurmctld:slurmctld
    volumes:
    - /tmp/chome/:/chome/
    dns: 127.0.0.1
    privileged: true
```

QNIBTerminal

- Framework of system container to spin up stacks
  - SLURM

```
$ fig up -d
Creating qnibterminalslurm_consul_1...
Creating qnibterminalslurm_slurmctld_1...
Creating qnibterminalslurm_slurmd_1...
$
```




```
consul:
    image: qnib/consul
    ports:
     - "8500:8500"
    dns: 127.0.0.1
    hostname: consul
    privileged: true

slurmctld:
    image: qnib/slurmctld
    ports:
    - "6817:6817"
    links:
    - consul:consul
    dns: 127.0.0.1
    hostname: slurmctld
    privileged: true

slurmd:
    image: qnib/slurmd
    links:
    - consul:consul
    - slurmctld:slurmctld
    volumes:
    - /tmp/chome/:/chome/
    dns: 127.0.0.1
    privileged: true
```

QNIBTerminal

- Framework of system container to spin up stacks
  - SLURM

```
$ fig up -d
Creating qnibterminalslurm_consul_1...
Creating qnibterminalslurm_slurmctld_1...
Creating qnibterminalslurm_slurmd_1...
$
```

```
consul:
    image: qnib/consul
    ports:
     - "8500:8500"
    dns: 127.0.0.1
    hostname: consul
    privileged: true
```

SERVICES NODES KEY/VA

1
```
$ fig up -d
Creating qnibterminalslurm_consul_1...
Creating qnibterminalslurm_slurmctld_1...
Creating qnibterminalslurm_slurmd_1...
```
2
```
$ fig scale slurmd=5
Starting qnibterminalslurm_slurmd_2...
Starting qnibterminalslurm_slurmd_3...
Starting qnibterminalslurm_slurmd_4...
Starting qnibterminalslurm_slurmd_5...
$
```
3
```
$ docker exec -ti qnibterminalslurm_slurmd_1 bash
bash-4.2# sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
qnib*        up   infinite      5   idle 7d594c9273fc,84d5793853a0,c0b19d3a7a4c,c969ae2180d1,cfc38e9b09d2
```

ssh 8 passing

```
    volumes:
    - /tmp/chome/:/chome/
    dns: 127.0.0.1
    privileged: true
```

QNIBMonitoring

QNIBMonitoring

- Current monitoring systems do not connect
  - overlaying metrics with log events
  - use/build inventory system to provide connections usually hidden
  - users perspective and scope/context/background

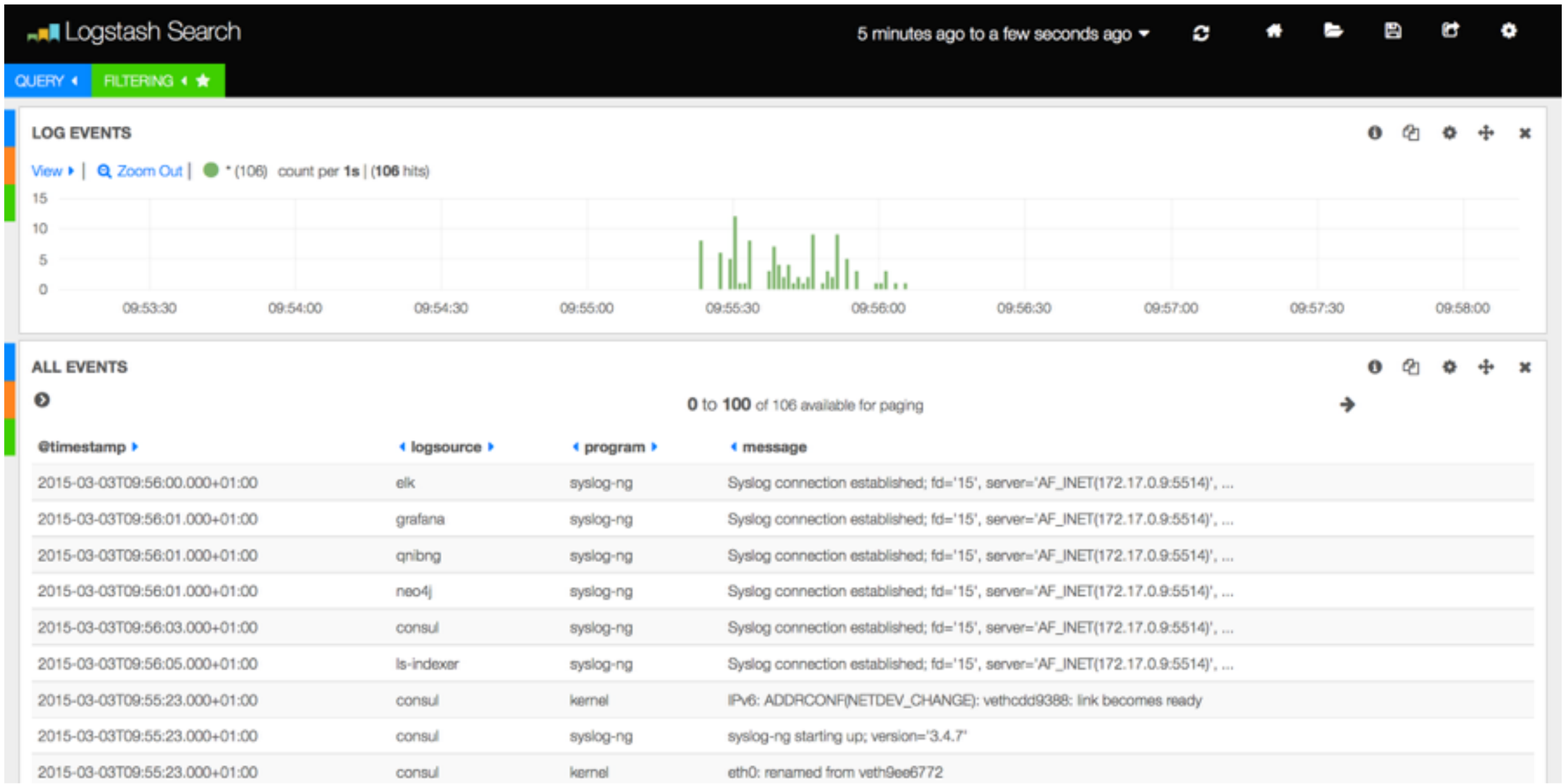
QNIBMonitoring

- Current monitoring systems do not connect
  - overlaying metrics with log events
  - use/build inventory system to provide connections usually hidden
  - users perspective and scope/context/background

- QNIBMonitoring provides
  - open metrics system (system / application metrics, log aggregates)
  - log event framework, consuming/processing/visualise events
  - auto discovery / configuration through consul

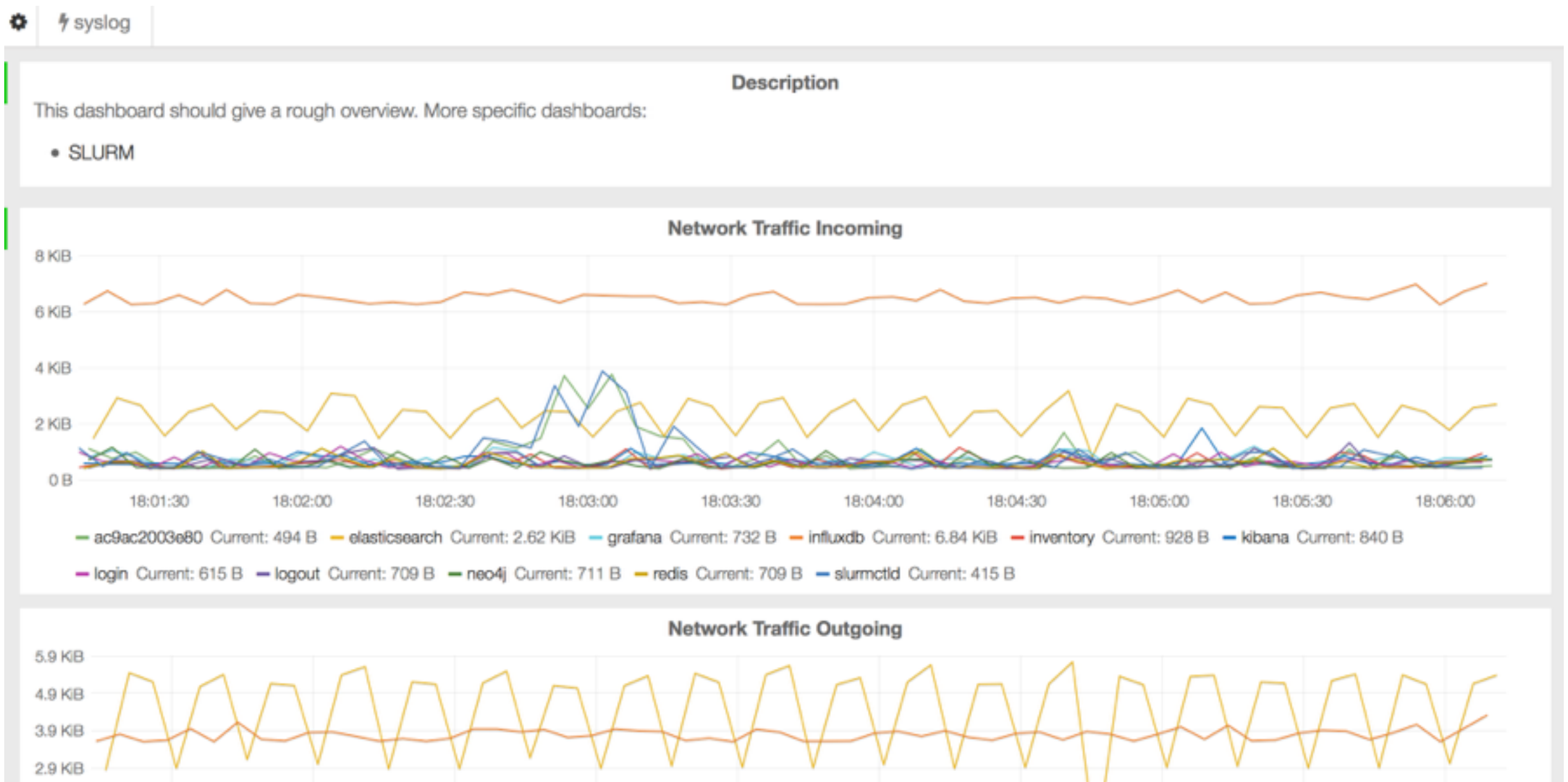
QNIBMonitoring

- Logstash (Log/Event Monitoring)



QNIBMonitoring

- Grafana (Performance Monitoring)



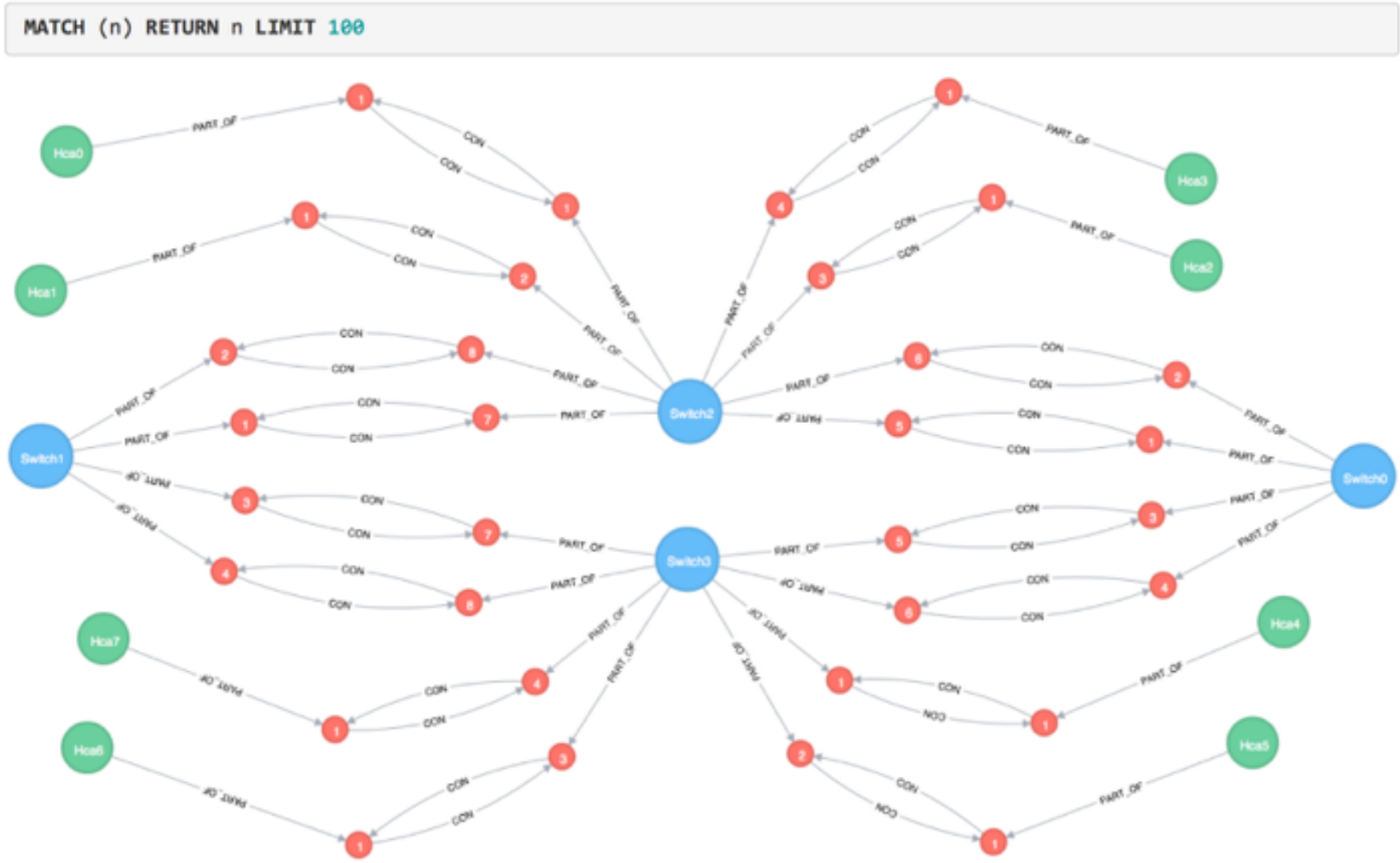
QNIBMonitoring

- Overlay Metrics w/ Events

QNIBInventory

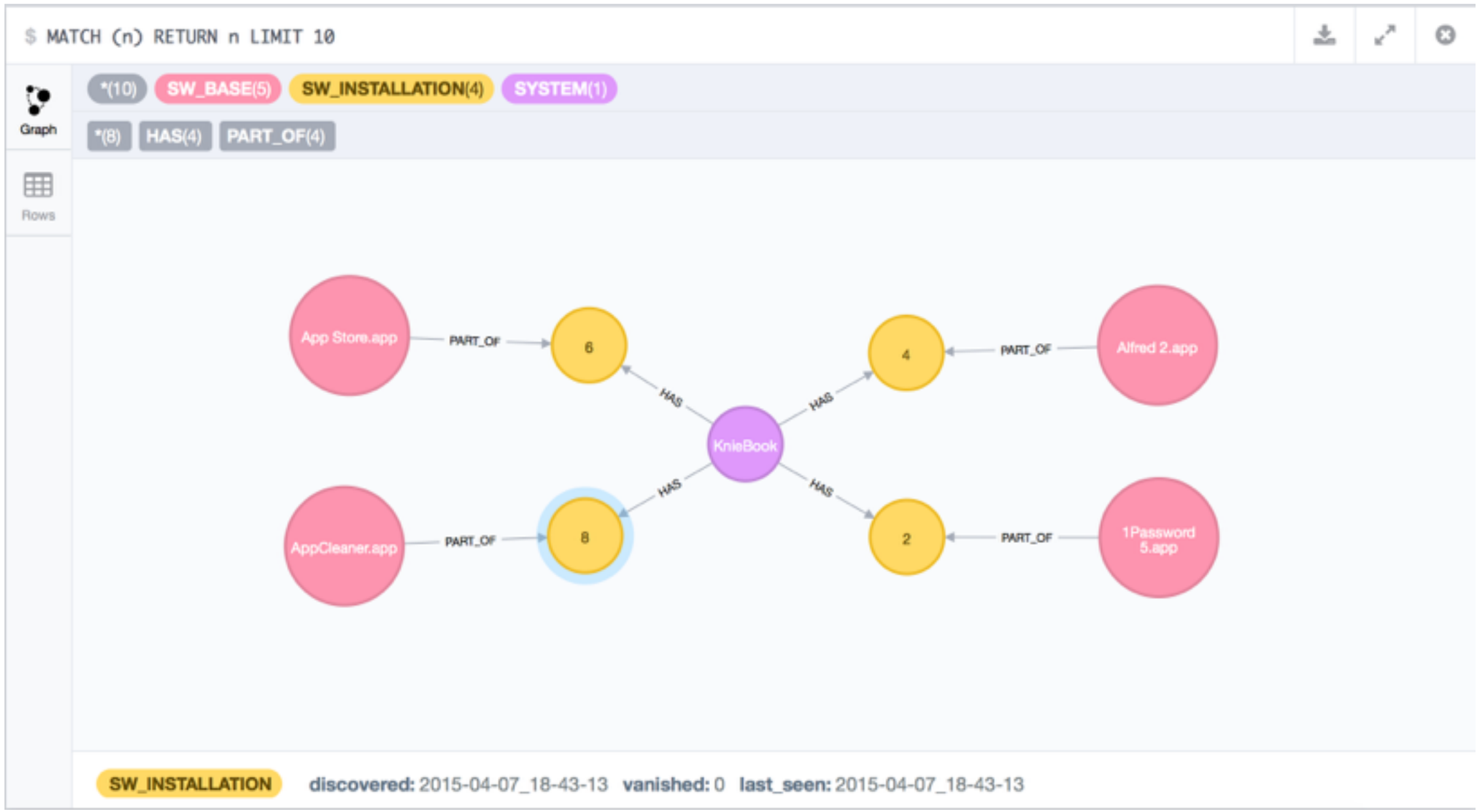
QNIBInventory

- Network Topology



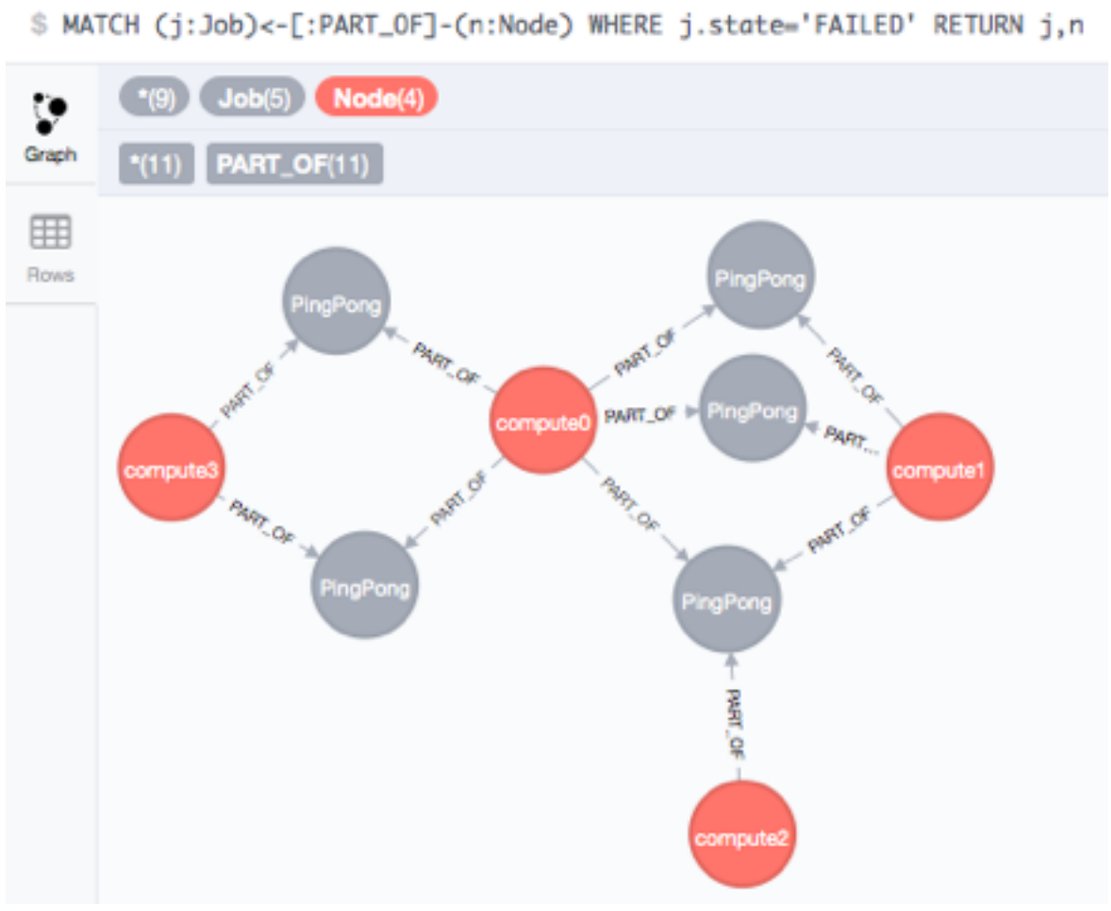
QNIBInventory

- Installed Software



QNIBInventory

- SLURM Cluster



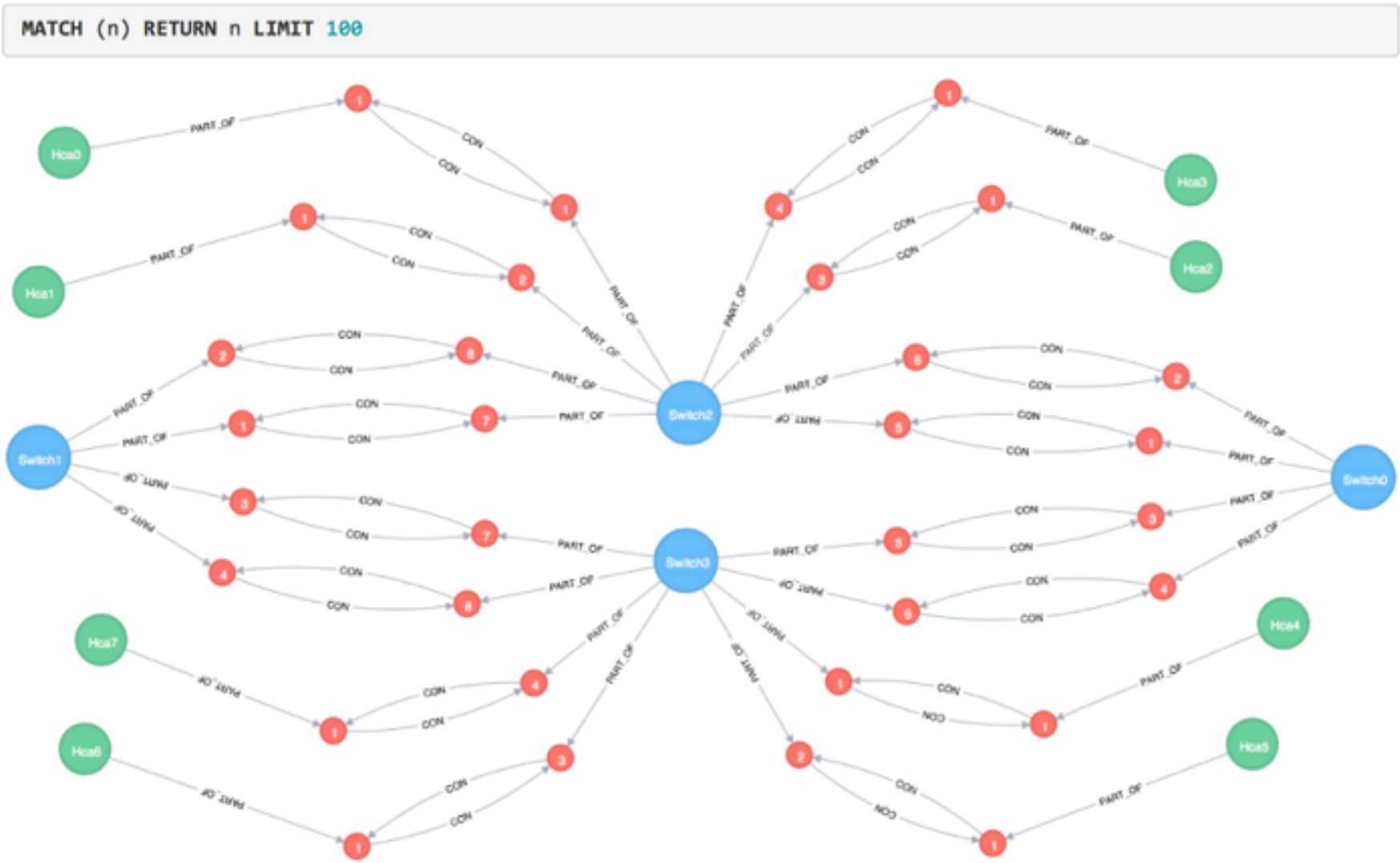
QNIBInventory

- Enrich Log/Events

message	2	Lid 7 assigned to port 1 of host Hca2
metric		lid.assigned
node_name		Hca2
old_msg	1	osm_lid_mgr_process_subnet: Assigned port 0x0000000000100005, new LID [7,7]
osm_func		osm_lid_mgr_process_subnet
plain_guid		100005

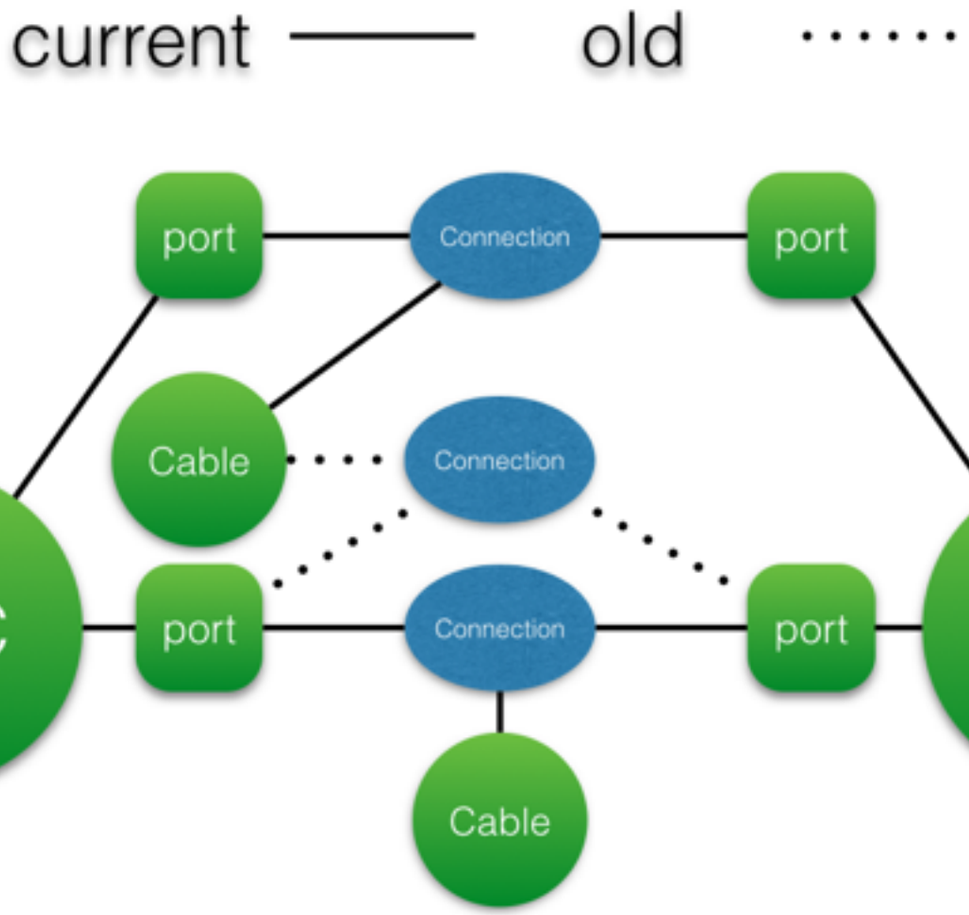
QNIBInventory

- Enrich Log/Events
- Help visualise connections



QNIBInventory

- Enrich Log/Events
- Help visualise connections
- Build up history



Cluster Use-Case

Context Sensitive Dashboards

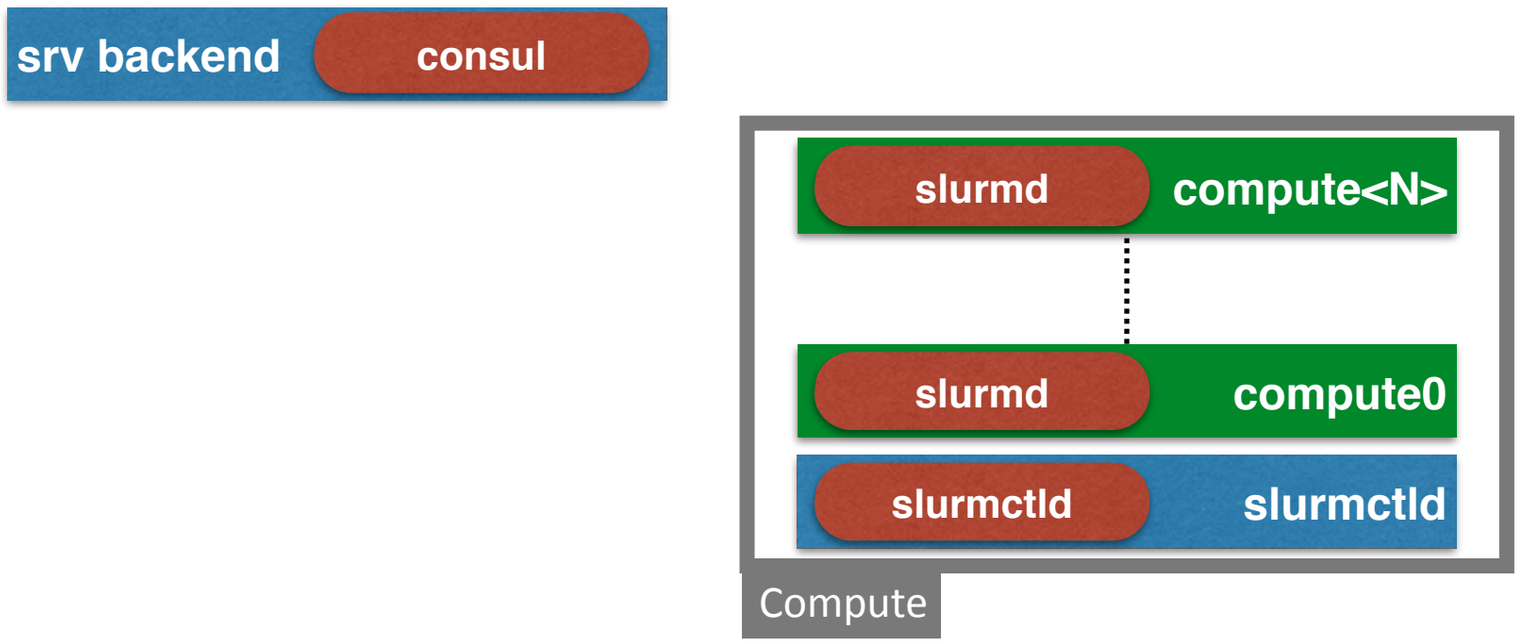
- Multiple backgrounds have to be considered
  - Enduser (Engineer, Software Developer, Scientist)
  - Operation Personel
  - Management
- Psychology plays important role
  - Local rationality / context
  - 10.000ft Overview vs. verifying hypothesis vs. Reporting
  - Empower users to extend their domain knowledge by providing toolset

Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload

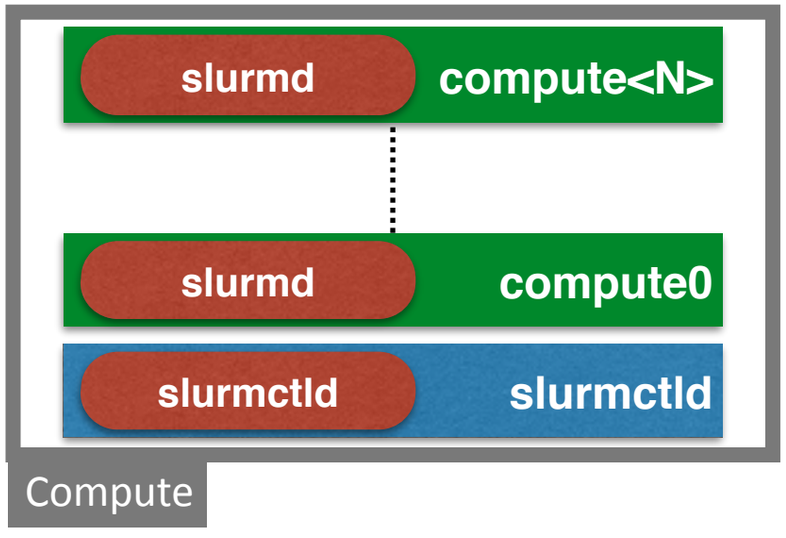
Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload



Cluster Usecase

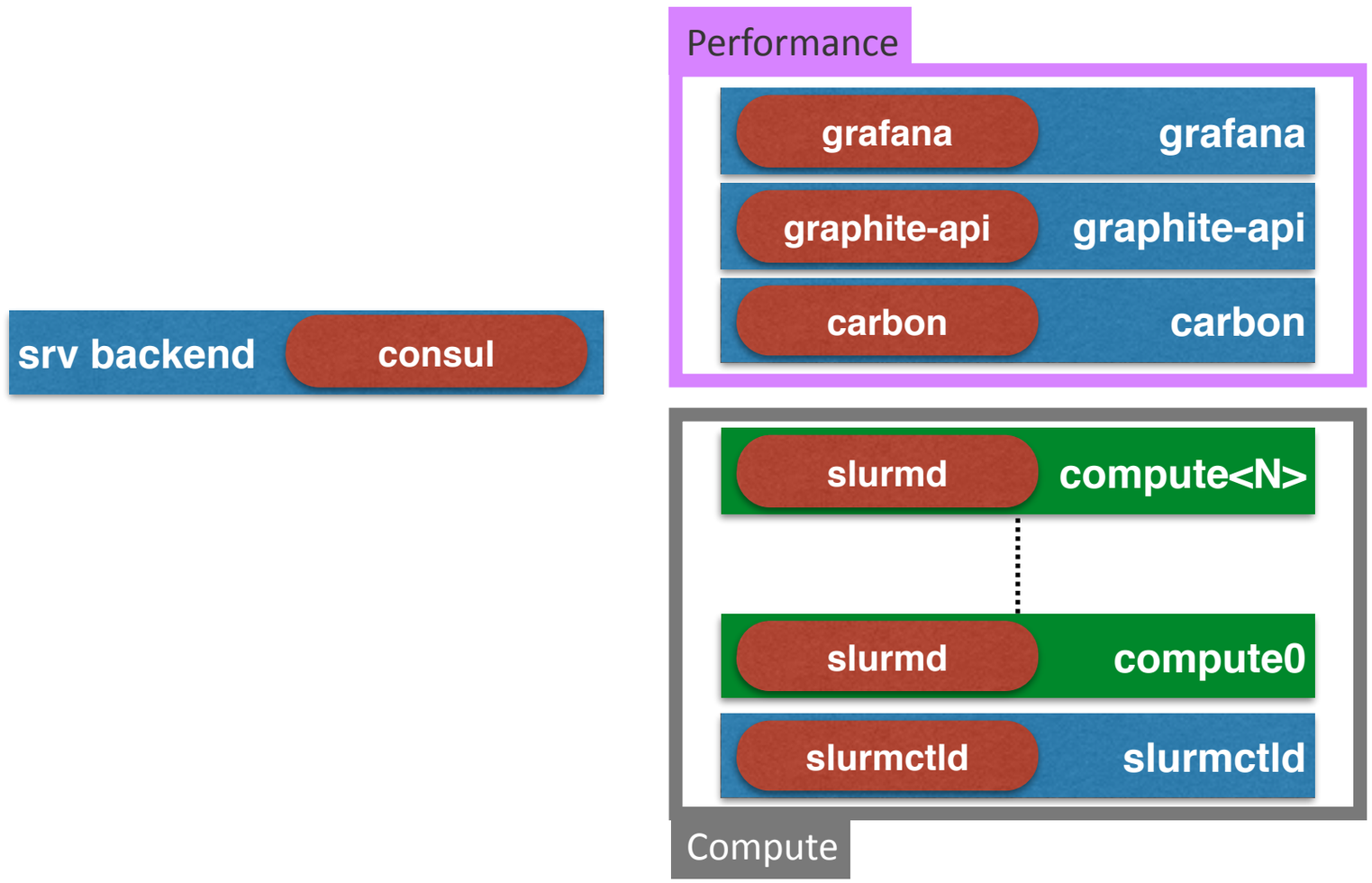
- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload




```
consul:
    image: qni
    ports:
     - "8500:8
    dns: 127.0
    hostname:
    privileged

slurmctld:
    image: qni
    ports:
    - "6817:68
    links:
    - consul:c
    dns: 127.0
    hostname:
    privileged

slurmd:
    image: qni
    links:
    - consul:c
    - slurmctl
    volumes:
    - /tmp/cho
    dns: 127.0
    privileged
```

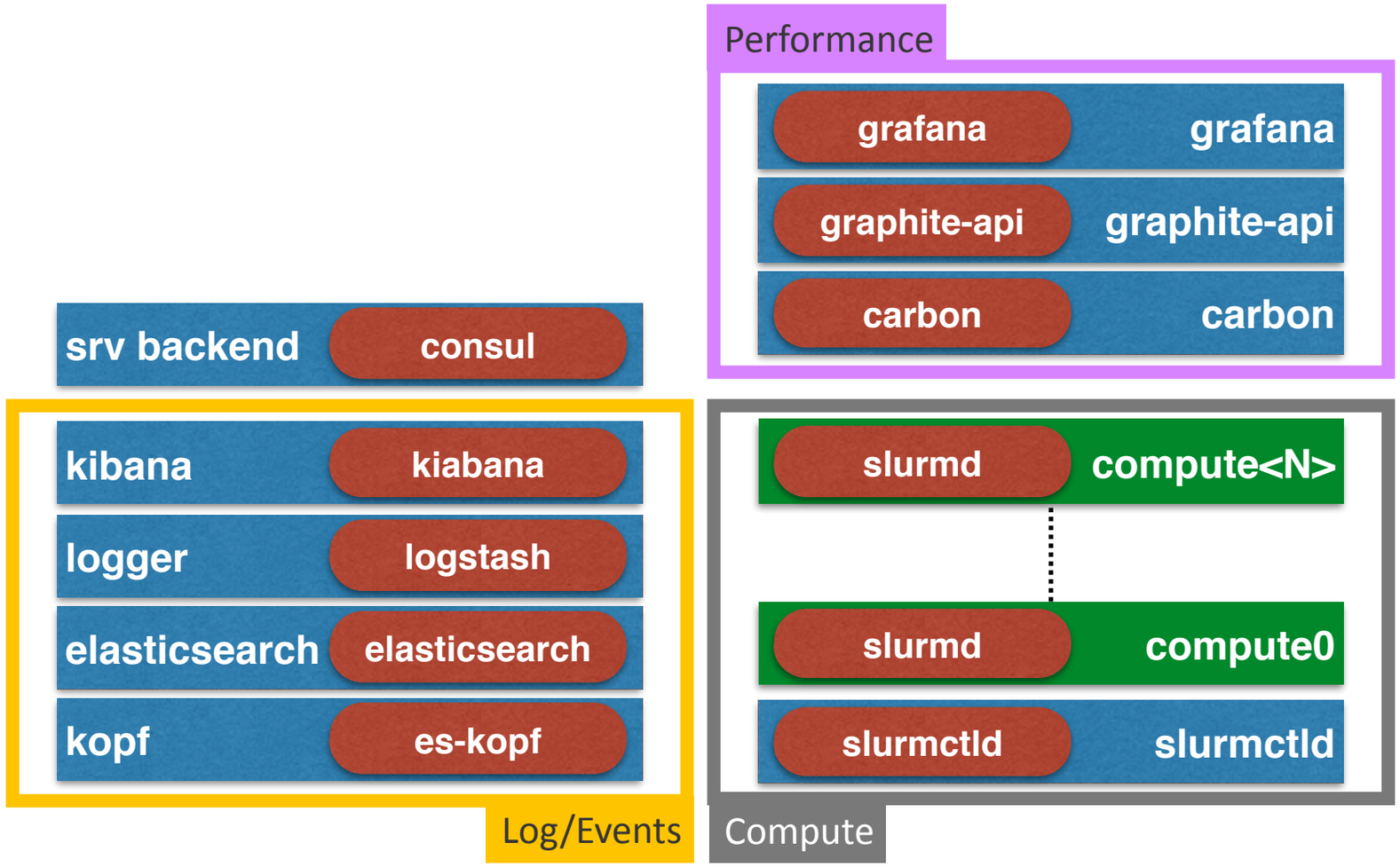
Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload



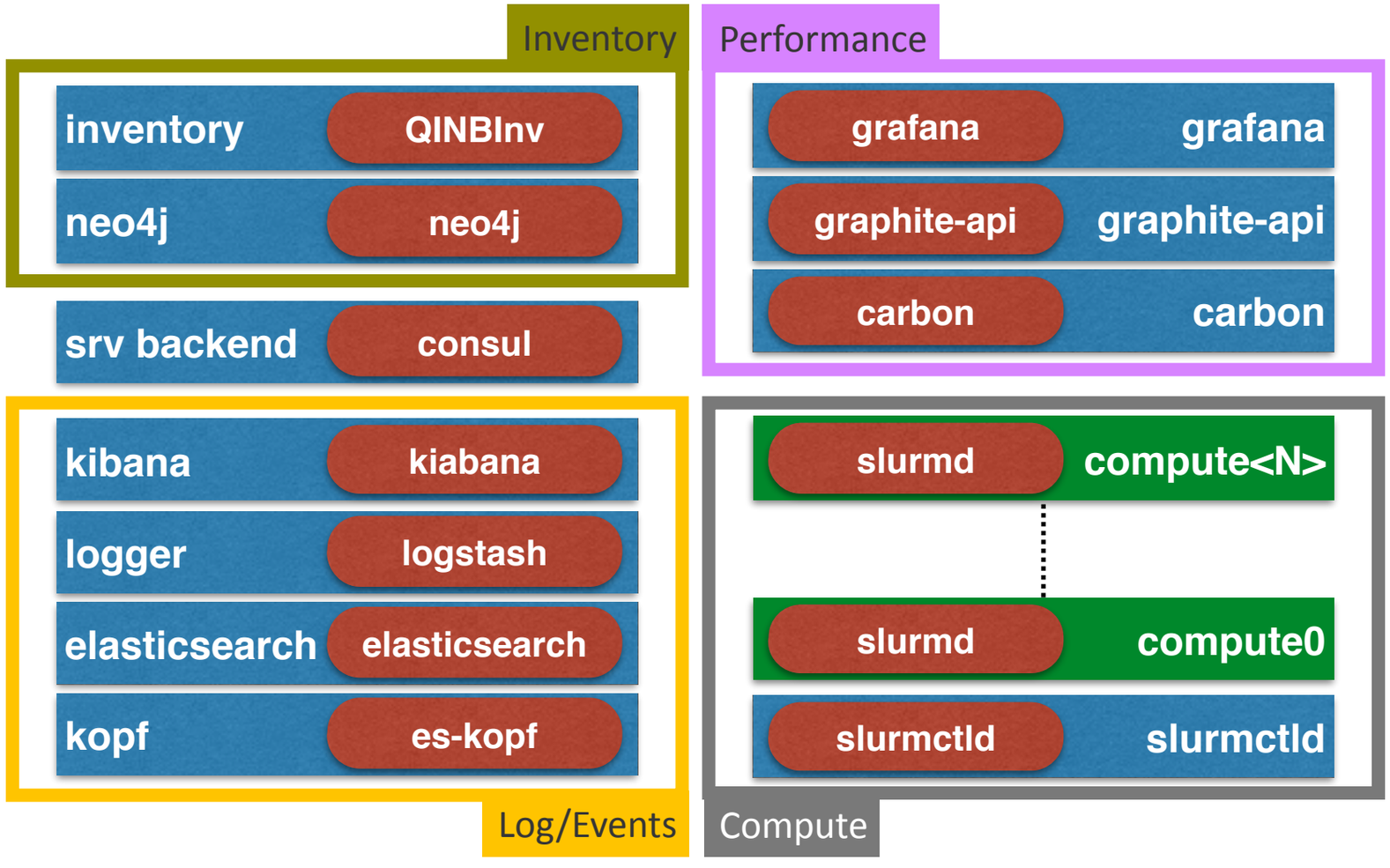
Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload



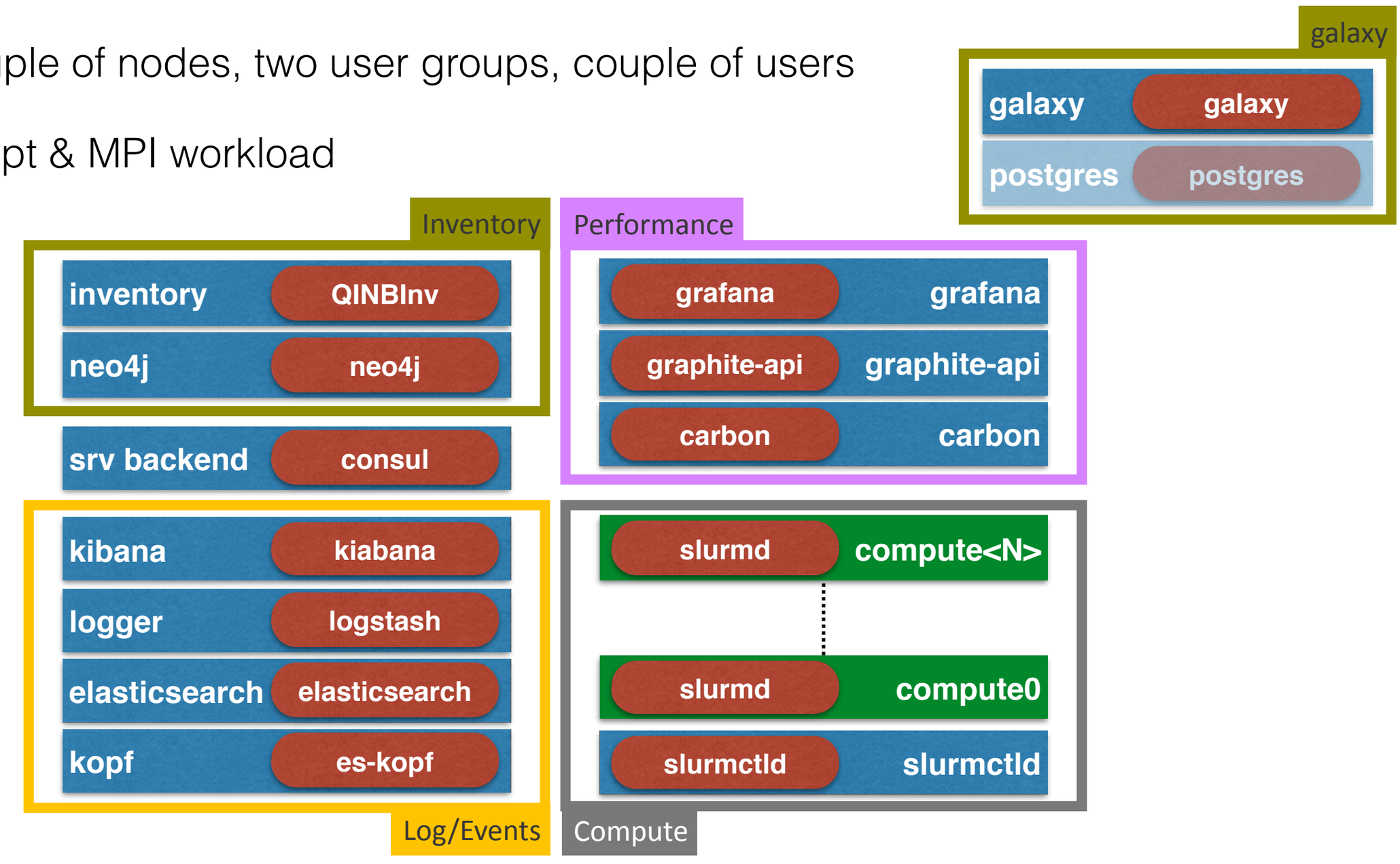
Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload



Cluster Usecase

- Small SLURM cluster
  - couple of nodes, two user groups, couple of users
  - script & MPI workload



Management Context

- Live cluster Status
  - Utilisation per cluster / user / user-group
  - SLA met by SysOps
  - Most common jobs, misbehaving enduser

Management Context

- Live cluster Status
  - Utilisation per cluster / user / user-group
  - SLA met by SysOps
  - Most common jobs, misbehaving enduser
- Reports
  - per day / user / job-type / …

Management Context

- Live cluster Status
  - Utilisation per cluster / user / user-group
  - SLA met by SysOps
  - Most common jobs, misbehaving enduser
- Reports
  - per day / user / job-type / …
- Capacity Planning
  - utilisation over time, comparison of HW generations, global FS capacity

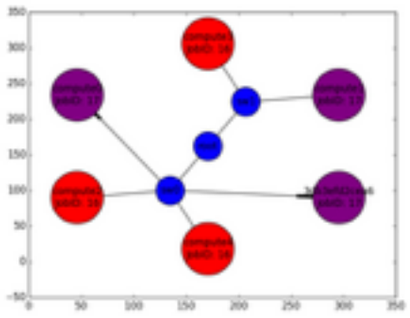
SLURM Dashboard

Slurm Dashboard

Zoom Out May 3, 2015 14:35:02 to May 3, 2015 14:43:52

slurmd

Job Map



Click picture for a bigger version.

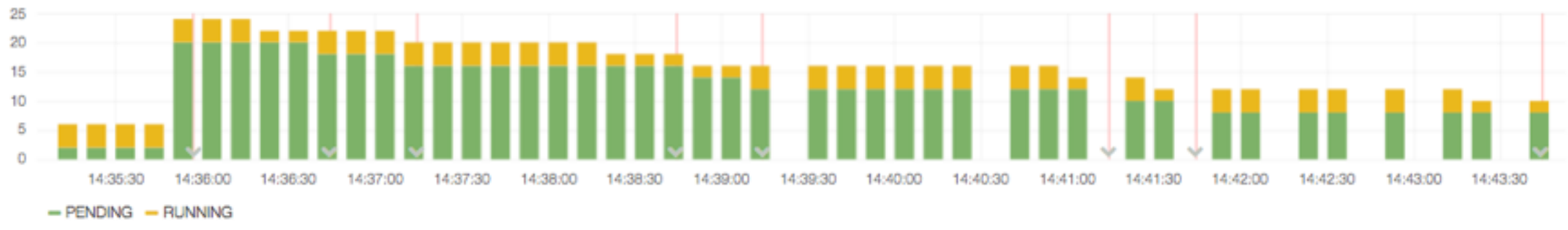
running jobs

JobID 17
2015-05-03 14:41:44 PingPong (john)

5 most recent finished jobs

JobID 16 (2.0) EC:0
2015-05-03 14:41:14 PingPong (carol)
JobID 15 (2.0) EC:0
2015-05-03 14:39:14 PingPong (jane)
JobID 14 (2.0) EC:0
2015-05-03 14:38:44 PingPong (eve)
JobID 13 (1.0) EC:0
2015-05-03 14:37:14 PingPong (bob)
JobID 12 (2.0) EC:0
2015-05-03 14:36:44 PingPong (john)

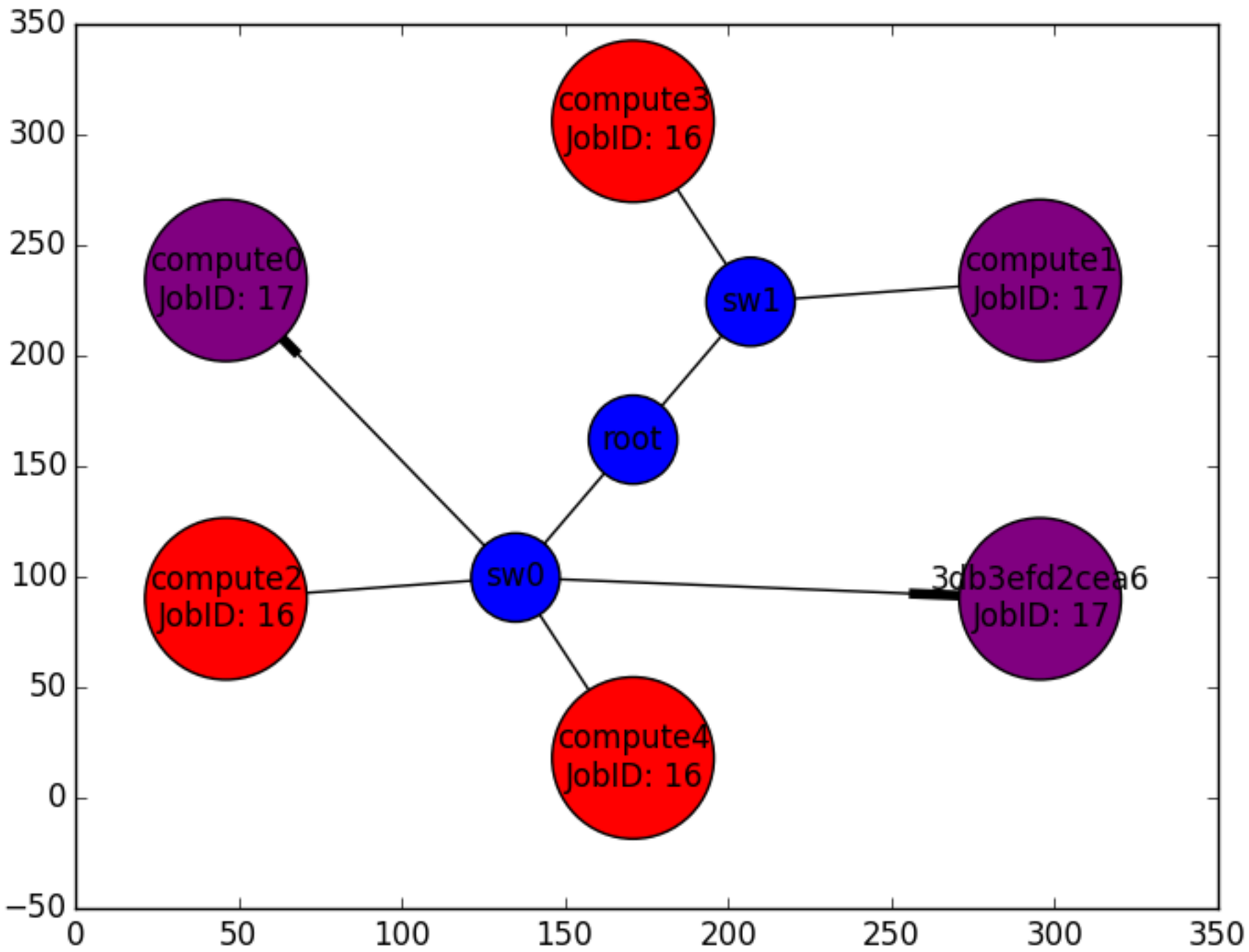
Job within QUEUE



Nodes use within RUNNING jobs by user

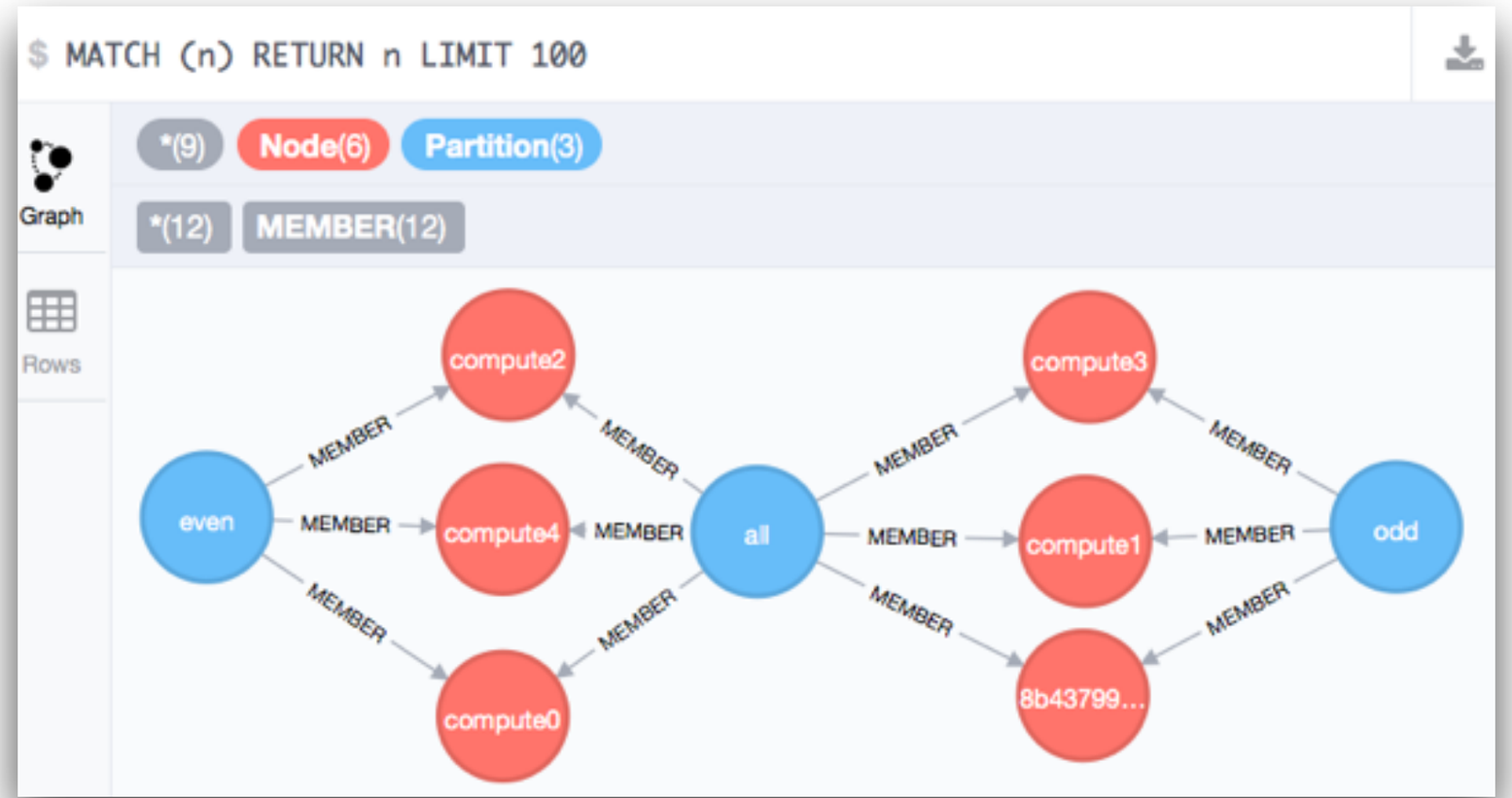
Nodes use within RUNNING jobs by group

SLURM Dashboard



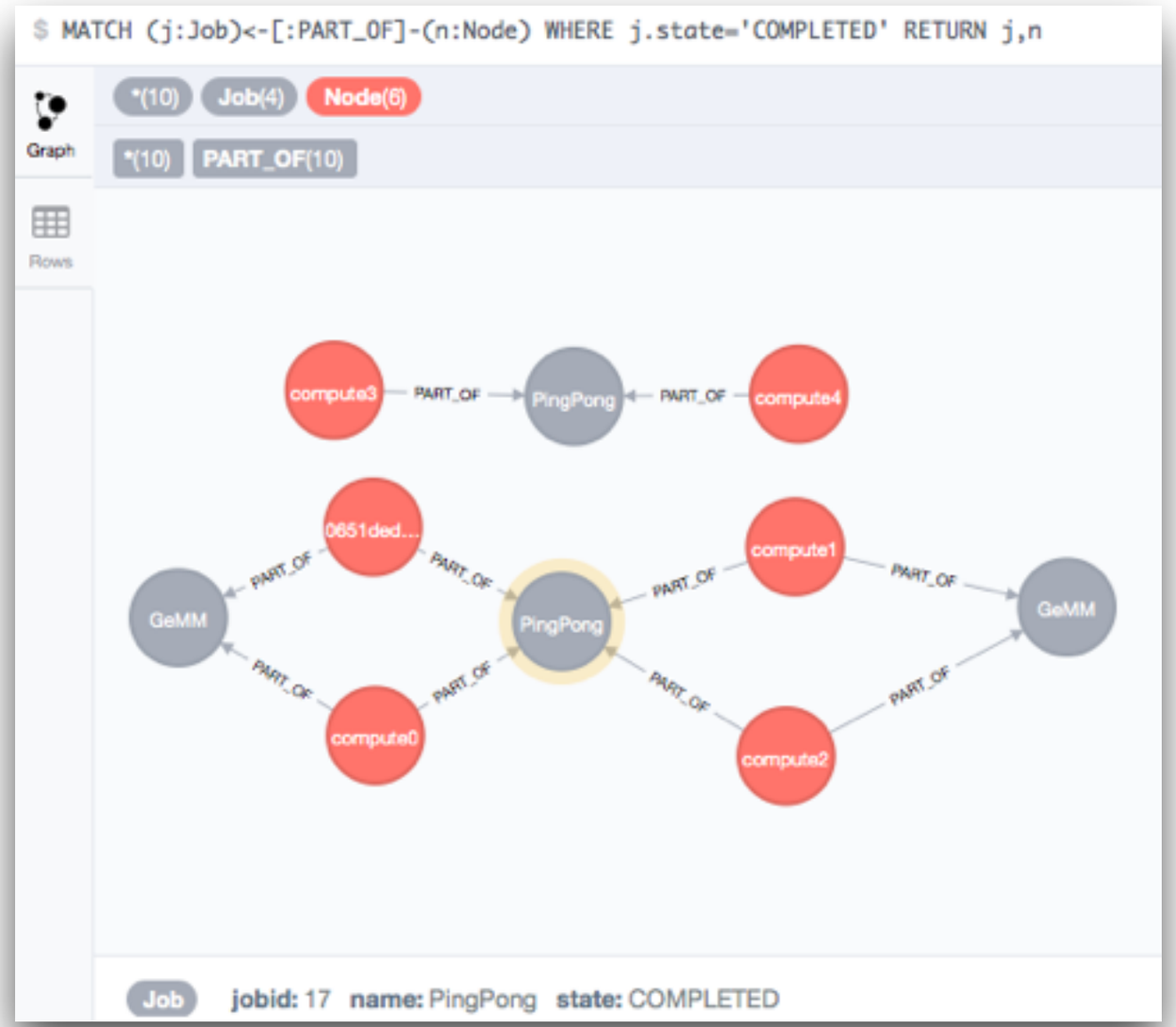
SLURM Inventar

- Nodes are connected to Partitions



SLURM Inventar

- Nodes are connected to Partitions
- Jobs are connected to both

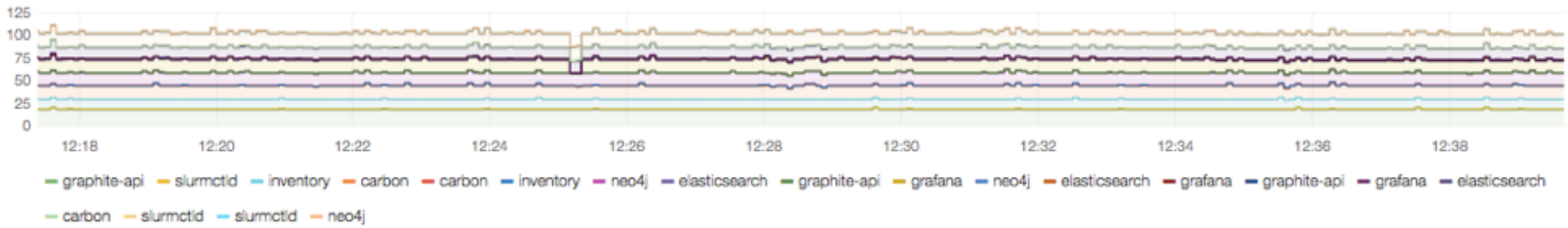


SLURM Dashboard

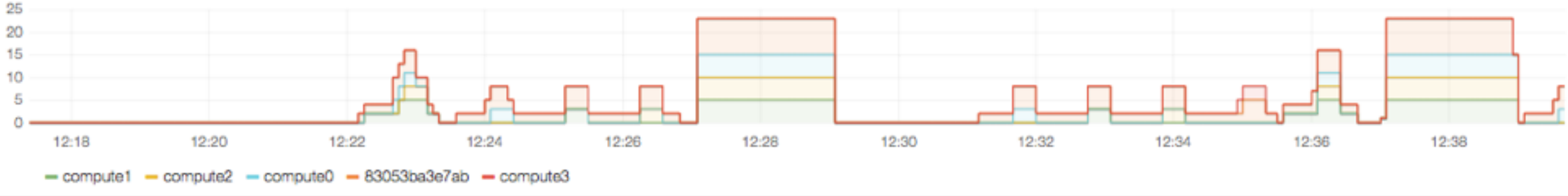
Process

Zoom Out 22 minutes ago to a few seconds ago

Processes on Service nodes



non-root processes on compute nodes by node



NonRoot Processes on Compute nodes by User

ADD A ROW

Enduser Context

- Live progress of SLURM job
  - Monitor iteration speed to estimate workload behaviour
  - Get to know job while it's running (instead of postmortem)
  - Introduce application profiling / log events (enhance feedback)

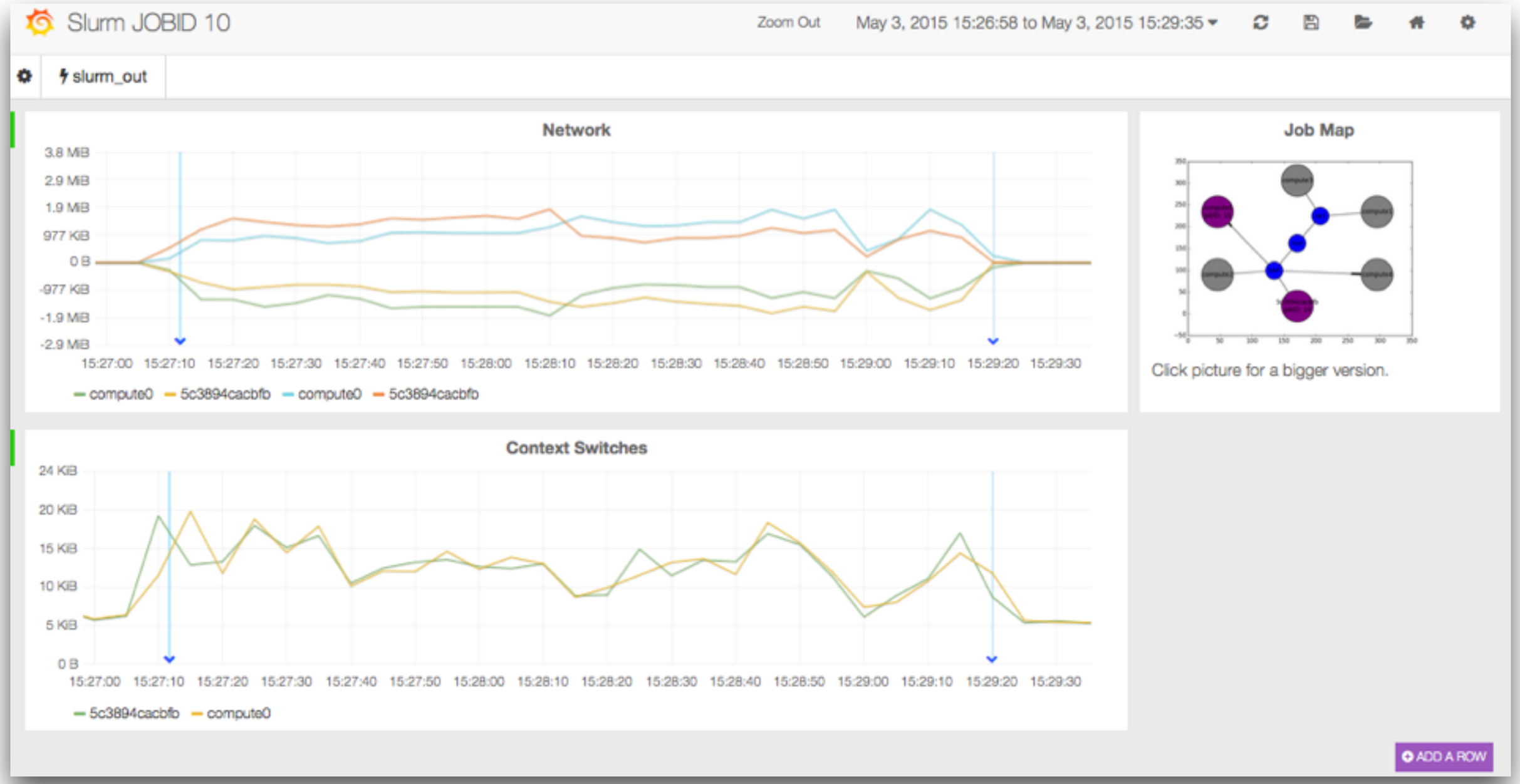
Enduser Context

- Live progress of SLURM job
  - Monitor iteration speed to estimate workload behaviour
  - Get to know job while it's running (instead of postmortem)
  - Introduce application profiling / log events (enhance feedback)
- Post Mortem
  - Get detailed report after job has finished

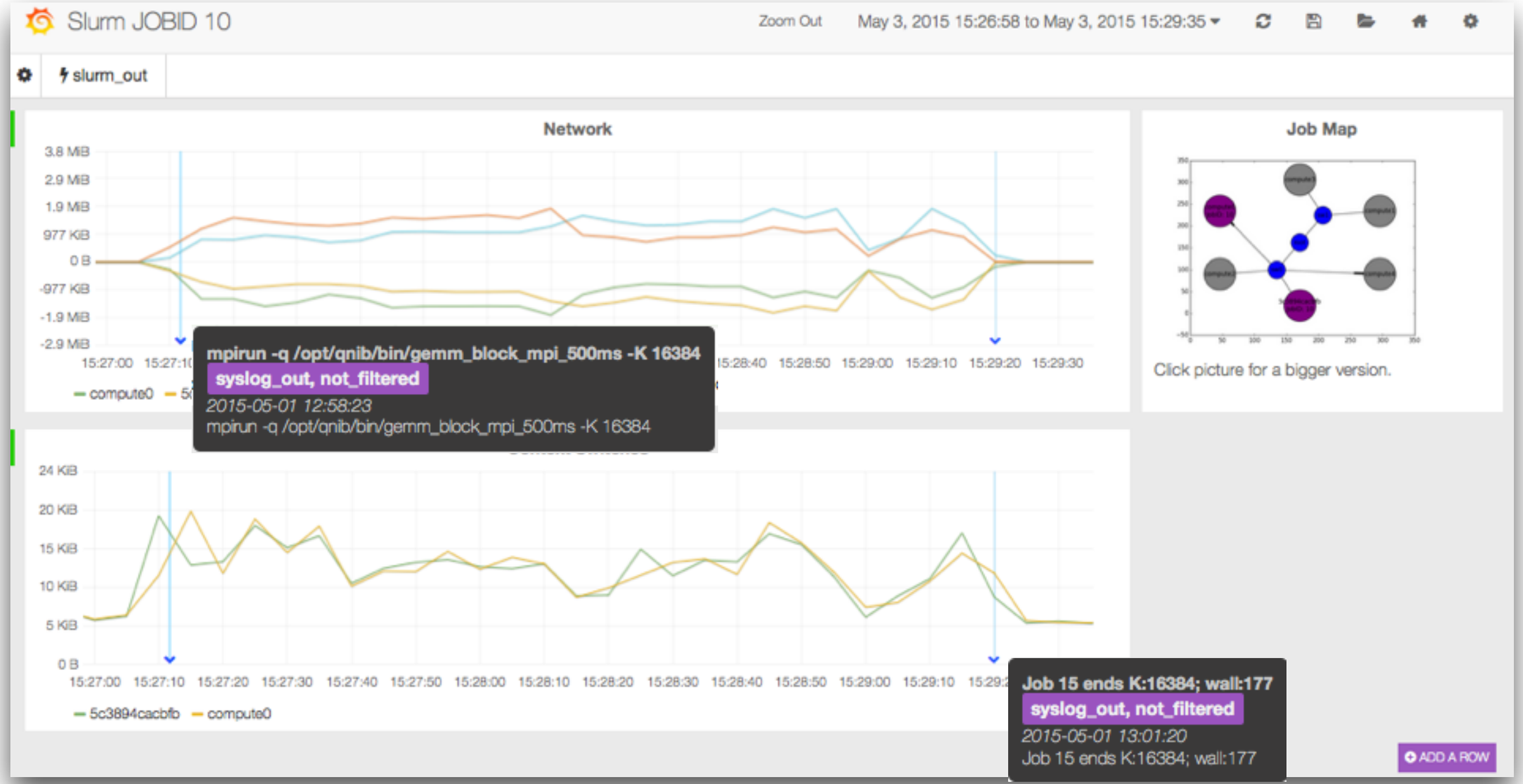
Enduser Context

- Live progress of SLURM job
  - Monitor iteration speed to estimate workload behaviour
  - Get to know job while it's running (instead of postmortem)
  - Introduce application profiling / log events (enhance feedback)
- Post Mortem
  - Get detailed report after job has finished
- MDO jobs
  - depending on outcome and progression submit next iteration(s)

SLURM Dashboard



SLURM Dashboard



SysOps Context

- Live cluster Status
  - USE method overviews (Utilisation/Saturation/Errors)
  - Anomaly detection (w/ and w/o humans)
  - Spotting abnormal behaviour

SysOps Context

- Live cluster Status
  - USE method overviews (Utilisation/Saturation/Errors)
  - Anomaly detection (w/ and w/o humans)
  - Spotting abnormal behaviour
- Drill into monitoring
  - verify hypothesis about incidents/problems
  - correlate events, metrics and inventory

SysOps Context

- Live cluster Status
  - USE method overviews (Utilisation/Saturation/Errors)
  - Anomaly detection (w/ and w/o humans)
  - Spotting abnormal behaviour
- Drill into monitoring
  - verify hypothesis about incidents/problems
  - correlate events, metrics and inventory
- Guid through ‘known problems’
  - close feedback loops provide confidence

Central Logging

Logstash Search

15 minutes ago to a few seconds ago

QUERY | FILTERING

LOG EVENTS

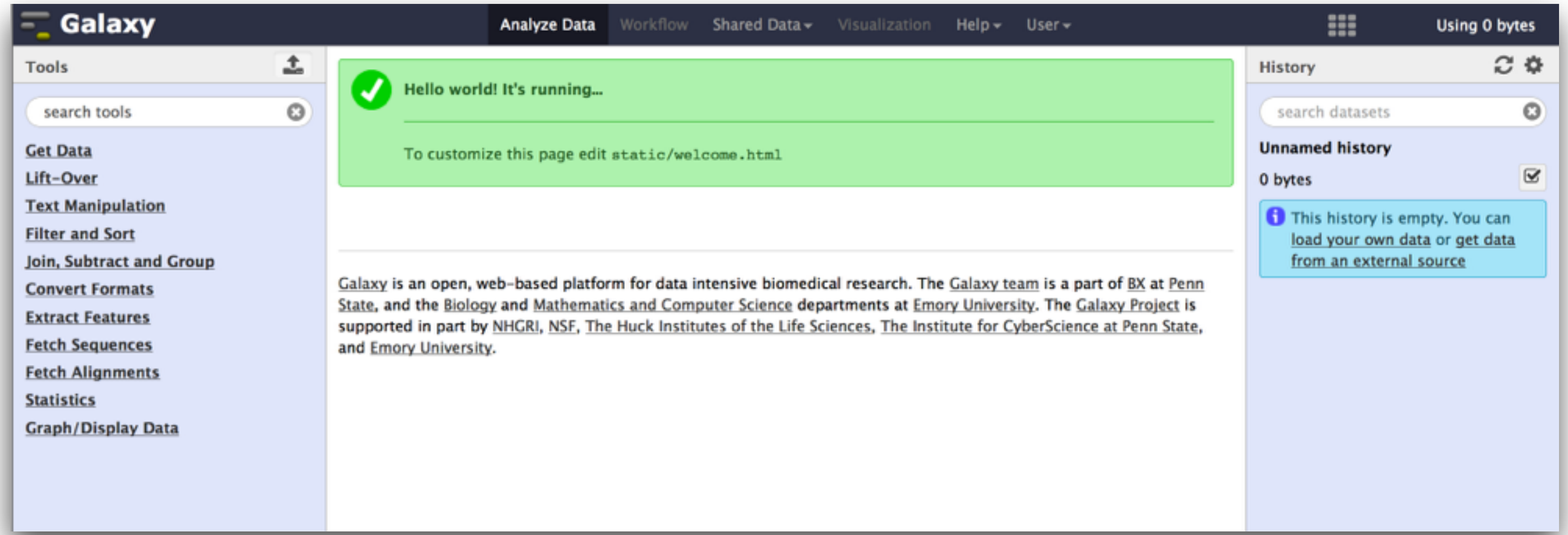
View ▸ | Zoom Out | * (101) count per **10s** | (**101** hits)

ALL EVENTS

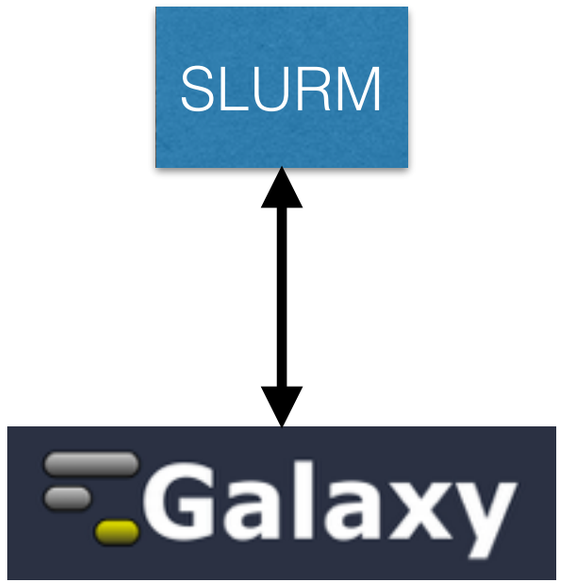
0 to **100** of 101 available for paging

@timestamp	logsource	program	message
2015-05-03T14:28:29.699+02:00	compute4	sshd	pam_unix(sshd:session): session opened for user alice by (uid=0)
2015-05-03T14:28:29.697+02:00	compute4	sshd	pam_systemd(sshd:session): Failed to connect to system bus: Failed to connect to socket /var/run/dbus/system_bus...
2015-05-03T14:28:29.657+02:00	compute4	sshd	Accepted publickey for alice from 172.17.0.124 port 33110 ssh2: RSA 5f:0b:88:e8:f0:be:a4:26:23:9b:9e:5a:35:07:ee...
2015-05-03T14:28:29.516+02:00	compute1	sshd	pam_unix(sshd:session): session opened for user dave by (uid=0)
2015-05-03T14:28:29.514+02:00	compute1	sshd	pam_systemd(sshd:session): Failed to connect to system bus: Failed to connect to socket /var/run/dbus/system_bus...
2015-05-03T14:28:29.481+02:00	compute1	sshd	Accepted publickey for dave from 172.17.0.121 port 57603 ssh2: RSA 5f:0b:88:e8:f0:be:a4:26:23:9b:9e:5a:35:07:ee:...
2015-05-03T14:28:29.469+02:00	compute2	slurmd	launch task 3.4 request from 3001.3001@172.17.0.124 (port 59851)
2015-05-03T14:28:29.454+02:00	compute2	slurm_out	rsync --bwlimit=3000 -aP /tmp/test_3.dd compute4.node.consul:/scratch/3/

Galaxy

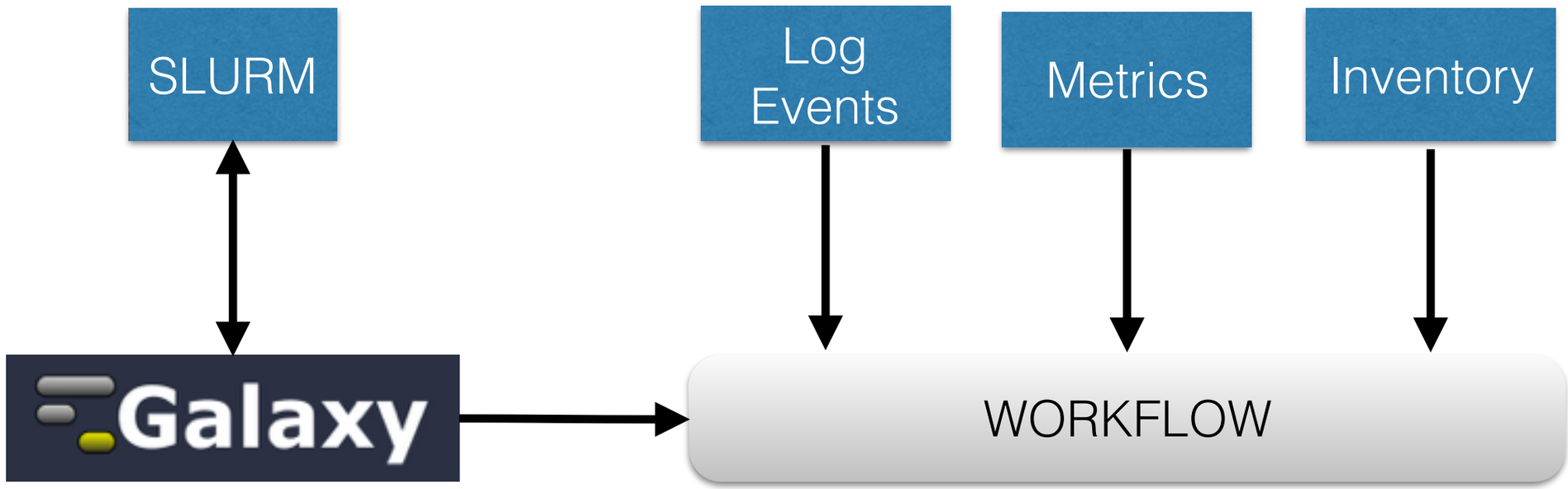

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 0 bytes
Tools
search tools
Get Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Statistics
Graph/Display Data
Hello world! It's running...
To customize this page edit static/welcome.html
Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.
History
search datasets
Unnamed history
0 bytes
This history is empty. You can load your own data or get data from an external source

Galaxy Use-Cases

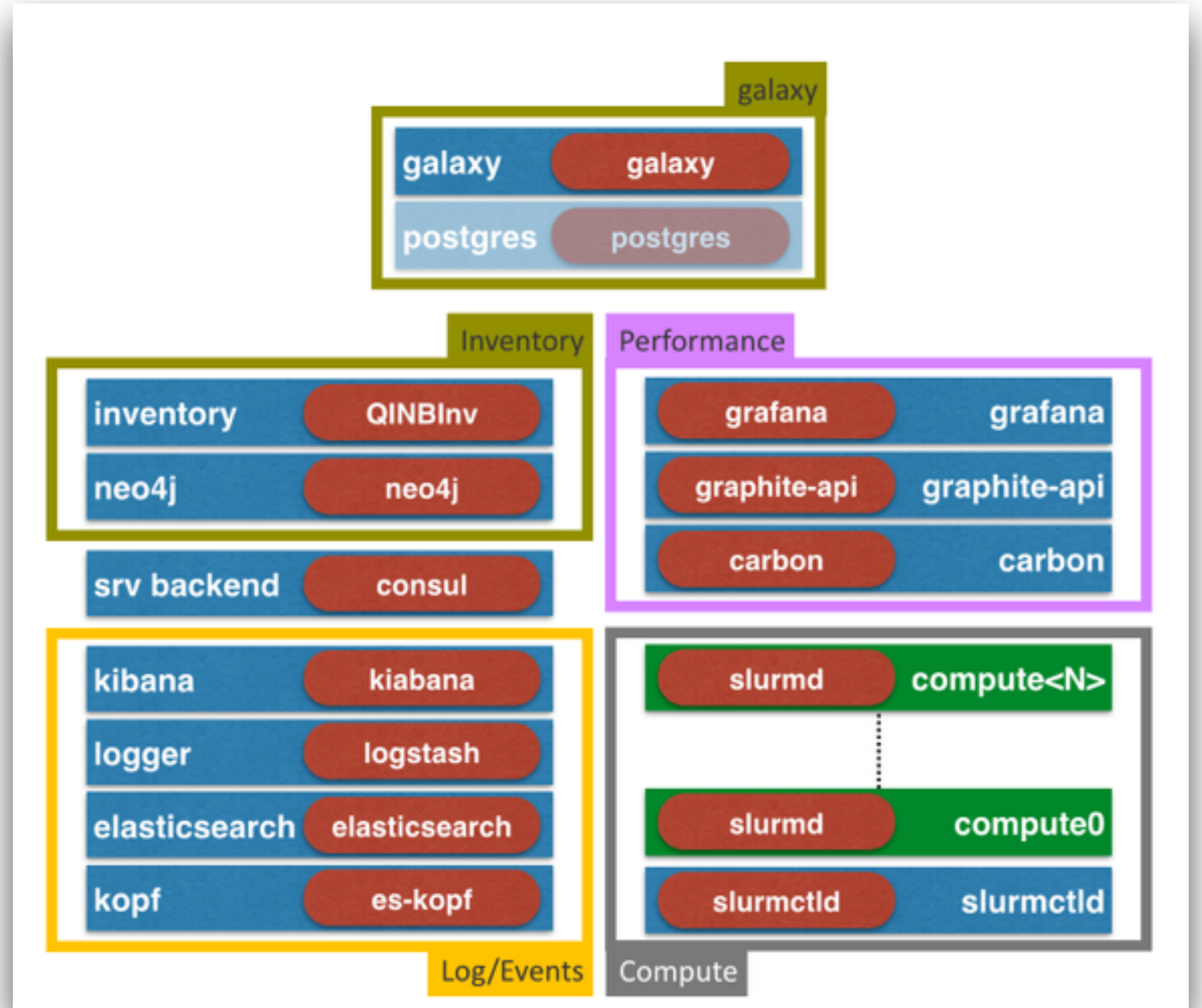


Galaxy Use-Cases

- Model Assess Workflow in Galaxy
  - Easy to grasp (in contrast to Hadoop, Spark, …)
  - Event triggered, Cronjob?
  - Using idle compute resources



Thank you!



- Contact
  - christian@qnib.org
  - @CQnib, @_qnib
- Web
  - www.qnib.org (blog)
  - doc.qnib.org (Paper)
- Feel free…
  - …ask questions (now / later)
  - …ask for a Demo