NATIONAL RESEARCH CENTRE
«KURCHATOV INSTITUTE»
PETERSBURG NUCLEAR PHYSICS INSTITUTE

PIK Data Centre

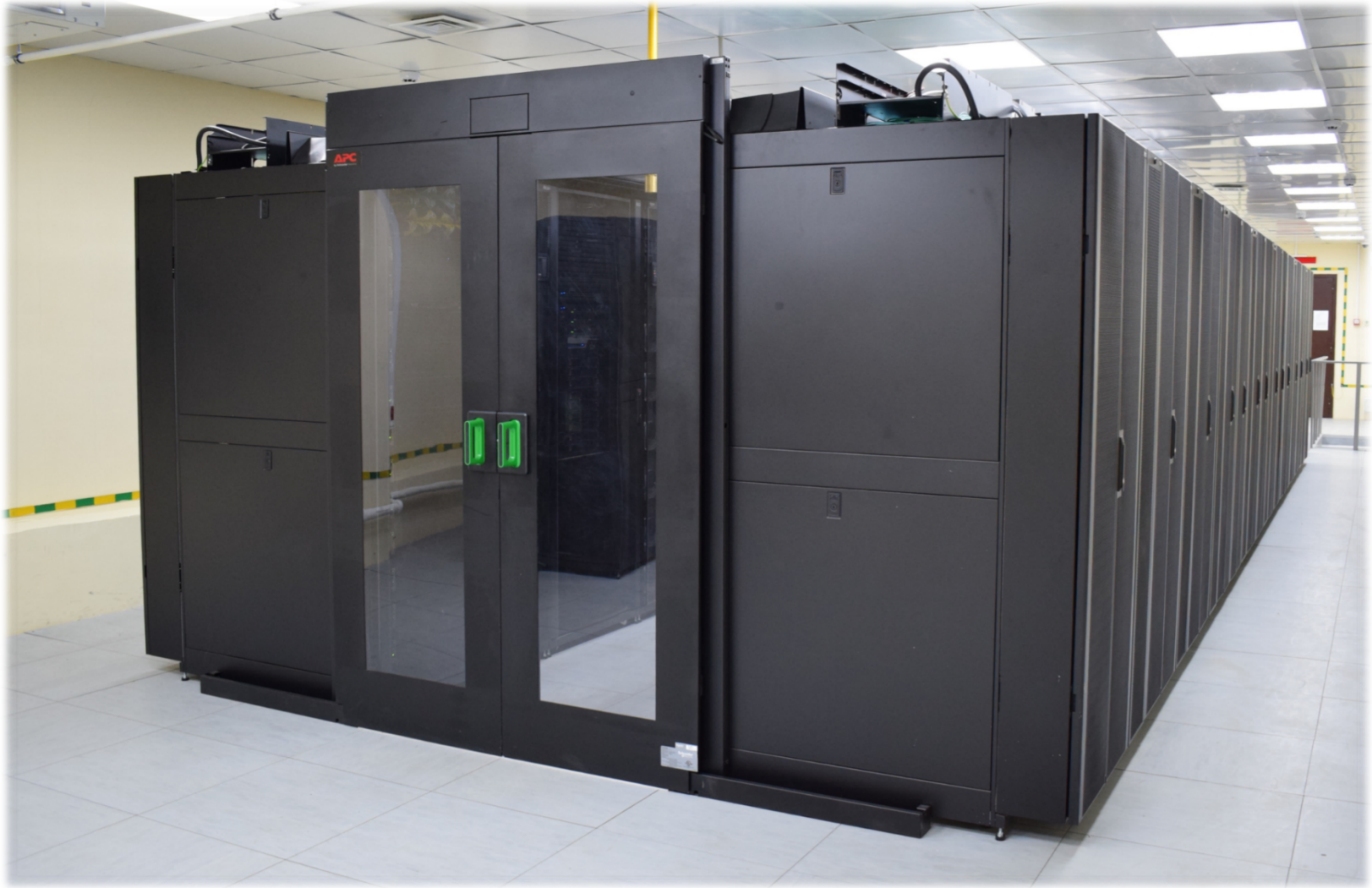
Operations experience and status update

Andrey Kiryanov

Outline

1. Available resources
2. Storage systems
   - Lustre
   - Ceph
3. Virtual infrastructure
4. Network
5. Monitoring
   - Electricity & cooling
   - Nodes
   - Storage
   - Network
   - Accounting

CACS with HACS



Current status

- 4th year of operation
- 24 research groups
- 89 users
- 1.6 PB of user data
- Jobs submitted
  - 416210 in 2018
  - 570533 in 2019
  - 189477 in 2020 ← COVID effect?
  - 166756 in 2021 (first half-year)

Computing equipment

- Peak theoretical performance: 362 Tflops
- Real LINPACK results:
  - 200 Tflops on Xeon CPUs (no AVX-512)
  - 68 Tflops on Xeon Phi (KNL) CPUs (AVX-512)
- Computing equipment:
  - 160 nodes with Xeon CPUs: 2.4 GHz, 28 cores, 128 GB RAM per node (4.5 GB RAM per core) – 4 480 cores
  - 40 nodes with Xeon Phi (KNL) CPUs: 1.4 GHz, 68 cores (272 virtual), 112 GB RAM per node – 10 880 virtual cores
  - 16 nodes with Xeon CPUs: 2.4 GHz, 28 cores, 1 TB RAM + 1.6 TB NVMe SSD – 448 cores
  - 2 nodes with Xeon CPUs: 2.4 GHz, 28 cores, 1.5 TB RAM – 56 cores
- Interconnect: InfiniBand EDR (100 Gbps)

Software

- CentOS 7 as a primary OS
  - Upgraded to 7.8 last fall
  - Kernel version is restricted by MOFED and Lustre releases
- Cluster management: xCAT
  - Patches for VMs and SR-IOV
- Batch system: Slurm
  - Custom made accounting visualization
- User management: FreeIPA
- Monitoring: Zabbix + Grafana, ElasticSearch + Kibana
- MPI: Open MPI (multiple versions), Intel MPI, Platform MPI
- Compilers: GCC (multiple versions), Intel oneAPI

Job scheduling

- Users are organized in workgroups
  - Resources are accounted per workgroup
- At PNPI workgroups normally reflect administrative hierarchy (departments / laboratories)
  - Fair-share tree reflects this hierarchy as well
- Node allocator never shares a node between users from different workgroups
  - Prevents memory consumption surprises
  - Co-allocation requires explicit memory limit
- Accounting information is stored in the database for one year
  - Database size ~10 GB

High Availability

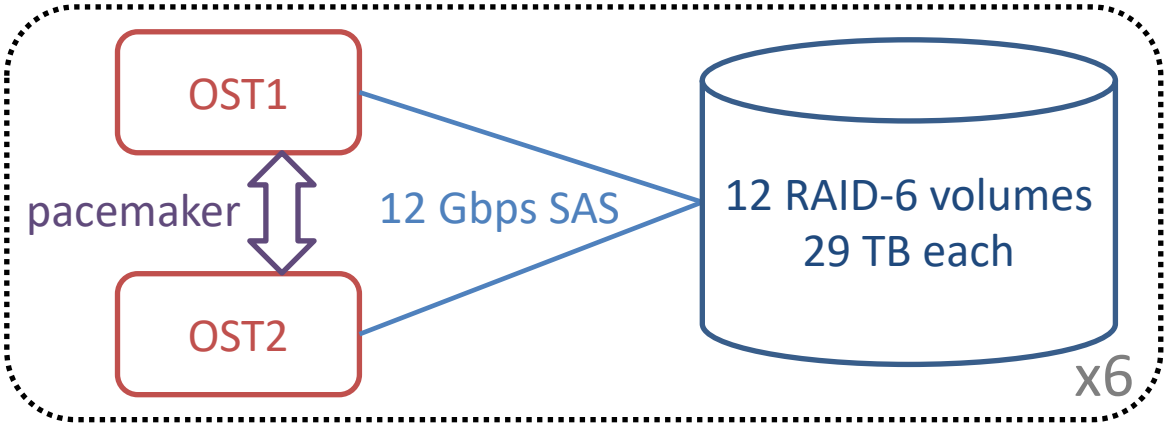
- Server chassis with dual PSUs
- Lustre servers work in pairs
  - Automated recovery with pacemaker
  - Disk shelves with redundant controllers and PSUs
- 2 FreeIPA servers in master-master mode
- 2 Slurm control/database instances
- 2 MariaDB instances for xCAT and Slurm with Galera replication

Outline

1. Available resources
2. Storage systems
   - Lustre
   - Ceph
3. Virtual infrastructure
4. Network
5. Monitoring
   - Electricity & cooling
   - Nodes
   - Storage
   - Network
   - Accounting

Storage: Lustre

- 14 servers: 12 OSTs + 2 MDTs in 7 HA pairs
- 13 high-density MD3460 SAS disk shelves (6 master-slave pairs for data + 1 for metadata), each shelf is connected to both servers in its HA pair
- 2 PB of storage capacity (2.8 PB raw)
- Access via EDR IB

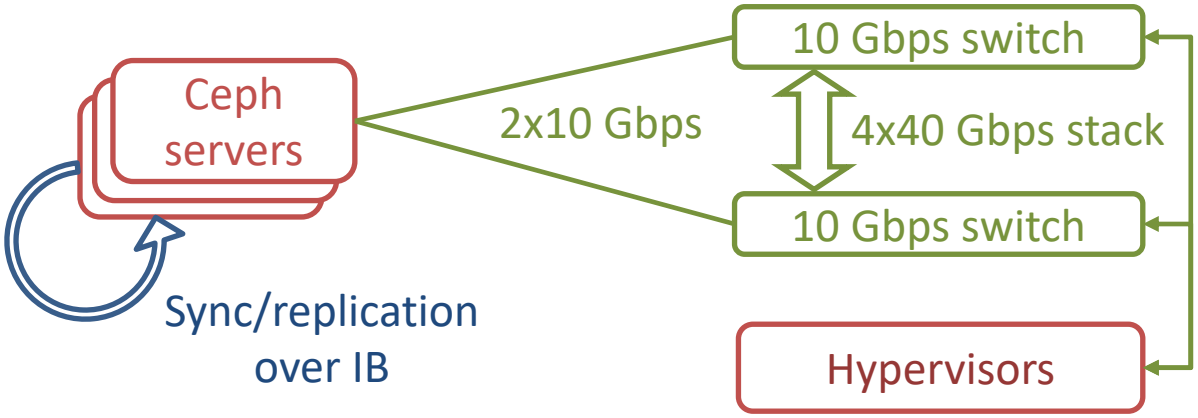


Lustre experience

- Currently on 2.12.5 release
  - Migrated from 2.10 with no issues
  - Will upgrade to 2.12.6 + CentOS 7.9
- ldiskfs on RAID6
  - No ZFS
  - MMP is important in HA scenarios
- Used for:
  - Home directories
  - Shared software area
  - Shared storage areas for workgroups
- Needs constant attention
  - Occasionally becomes unresponsive under load
  - We replace ~7 HDDs every year (1%)

Storage: Ceph

- 20 servers in two racks
- 320 individual OSDs (per disk), 8 GB each
- 2.5 PB of raw storage capacity
  - 2x replicated pools for VM images
  - EC pools (25% redundancy) for data storage
- Access via Ethernet



Ceph experience

- Currently on Octopus (15) release
  - Gradual updates from Luminous (12)
  - Seems to be the last version with CentOS 7 support
  - Lack of Python 3 modules required for full functionality in standard repositories
  - Severe changes to deployment model (*cephadm*, but *ceph-deploy* still works)
- We use the same physical servers for all services
  - *mon*, *mgr* ,*mds*, *osd* live happily together provided there's enough RAM
- We use Ceph RBDs in our virtual infrastructure
  - We also have CephFS, but it is mostly used for backups
- Very stable, no issues apart from HDD failures

Outline

1. Available resources
2. Storage systems
   - Lustre
   - Ceph
3. Virtual infrastructure
4. Network
5. Monitoring
   - Electricity & cooling
   - Nodes
   - Storage
   - Network
   - Accounting

Virtual infrastructure

- We do not use VMs for
  - Compute nodes
  - Storage nodes
- We use VMs as much as possible for everything else
  - Login nodes
  - Monitoring nodes
  - Application-specific services
  - Testbeds
- Infrastructure is based on KVM and managed by xCAT
  - Domain XML generator code required some patching

Virtual infrastructure

- VM images are hosted on Ceph
- ”host-passthrough” mode reveals full CPU capabilities to the guest
- SR-IOV allows low-overhead network access for both Ethernet and InfiniBand
  - Native Lustre and OFED access from VMs
  - Libvirt still cannot manage pools of VFs with non-Ethernet MACs, IB VFs need to be assigned by hand
- All our monitoring services including perfSONAR are deployed as VMs

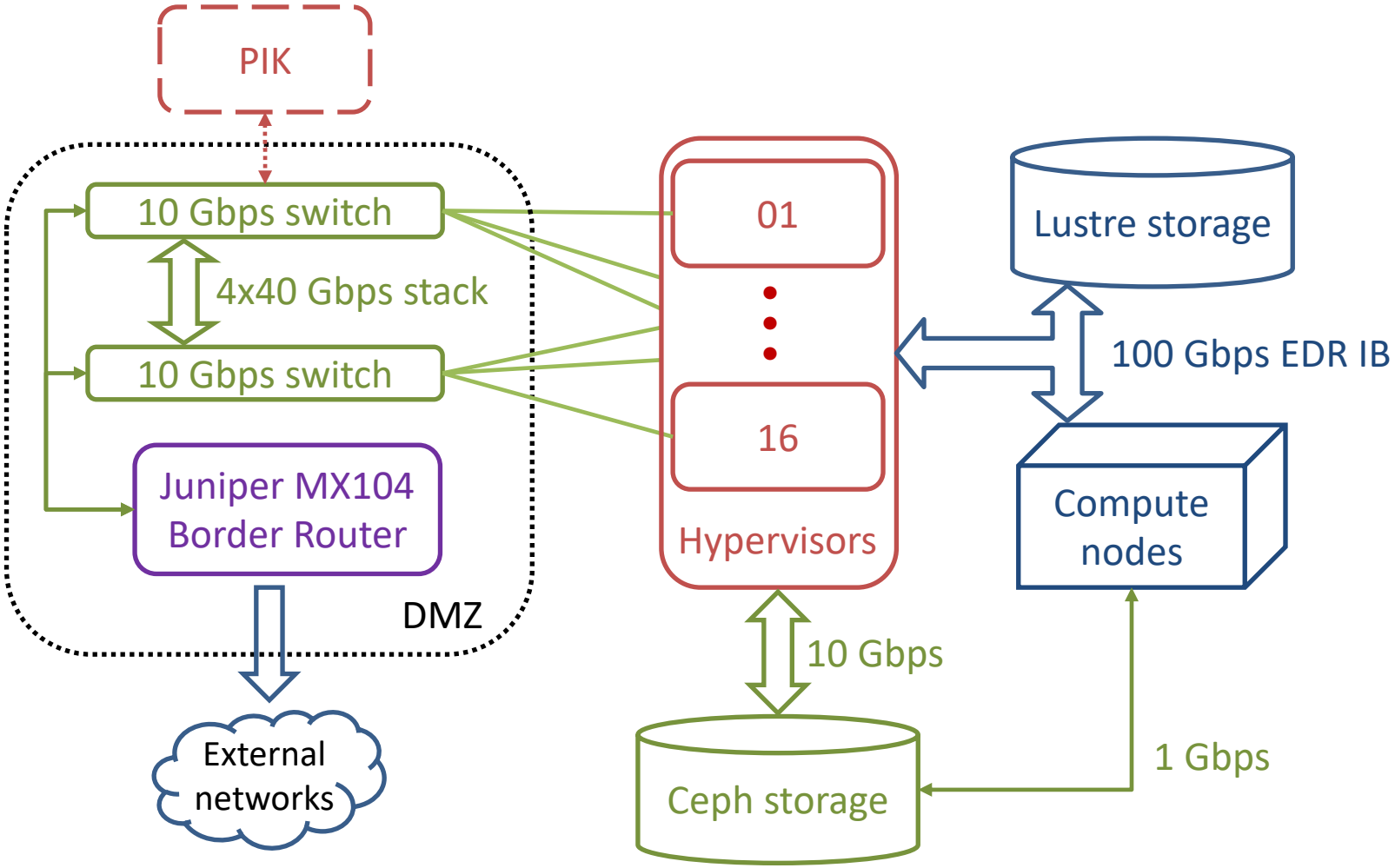
Containers

- Started exploring the containerized HPC workloads in 2020
- We have Singularity deployed on compute nodes
- Running HPL LINPACK inside Ubuntu containers
  - Proof of concept
  - Integration with Slurm
  - 1.5% overhead in comparison to native workload
- The goal is to provide more flexibility to the users
- We can also use free CPU time on our Ceph nodes

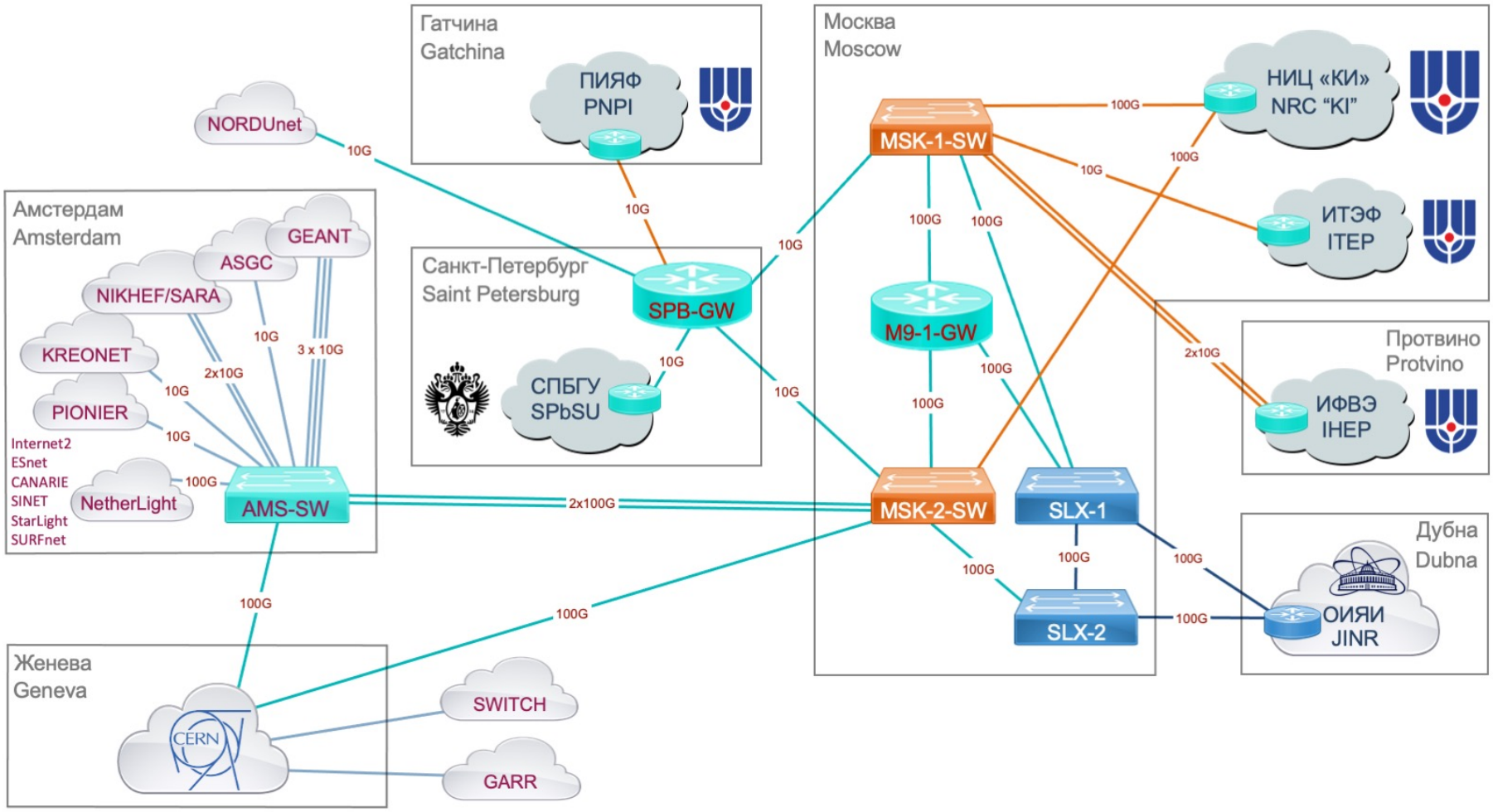
Outline

1. Available resources
2. Storage systems
   - Lustre
   - Ceph
3. Virtual infrastructure
4. Network
5. Monitoring
   - Electricity & cooling
   - Nodes
   - Storage
   - Network
   - Accounting

Network


PIK
10 Gbps switch
4x40 Gbps stack
10 Gbps switch
Juniper MX104
Border Router
DMZ
External
networks
01
16
Hypervisors
Lustre storage
100 Gbps EDR IB
Compute
nodes
10 Gbps
Ceph storage
1 Gbps

External networks



Network security

- We do not use passwords for user accounts
  - Brute-force attacks are impractical
- Users authenticate with their SSH keys
  - Private key can be password-encrypted, but it stays on the user side
- Still we see lots of break-in attempts
  - We have a site-wide fail2ban with a shared database
  - A ban on one login node is propagated to the others

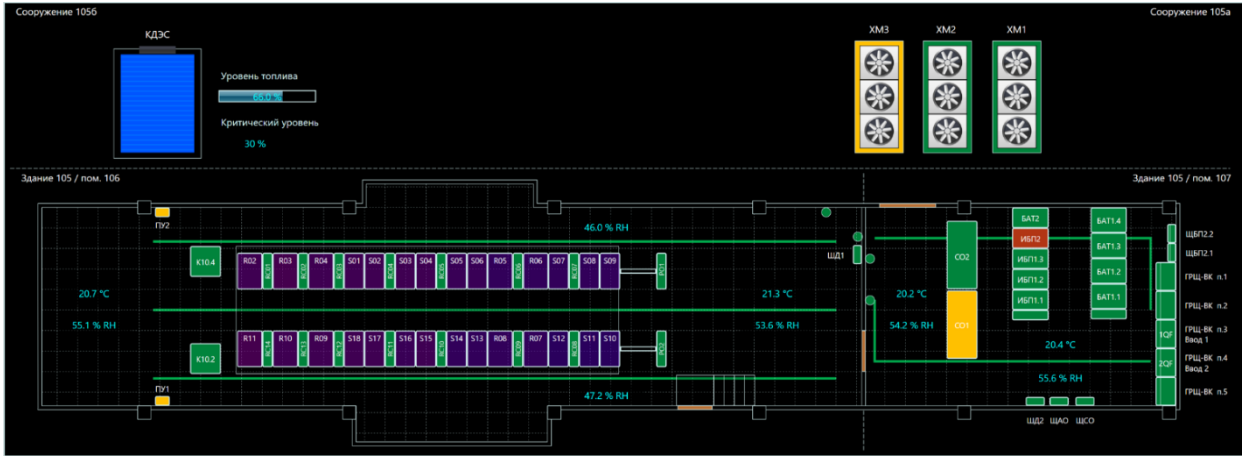
Outline

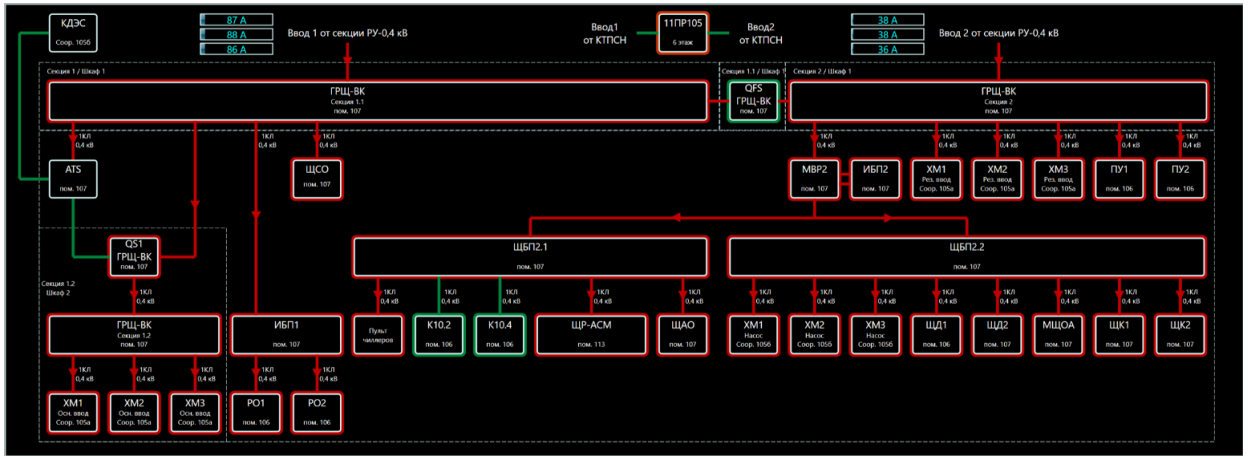
1. Available resources
2. Storage systems
   - Lustre
   - Ceph
3. Virtual infrastructure
4. Network
5. Monitoring
   - Electricity & cooling
   - Nodes
   - Storage
   - Network
   - Accounting

Monitoring of electricity & cooling

Based on GENESIS64 SCADA

- UPSes
- PDUs and circuit breakers
- Chillers
- Air conditioners
- Humidifiers
- Diesel
- Temperature, humidity and leak sensors

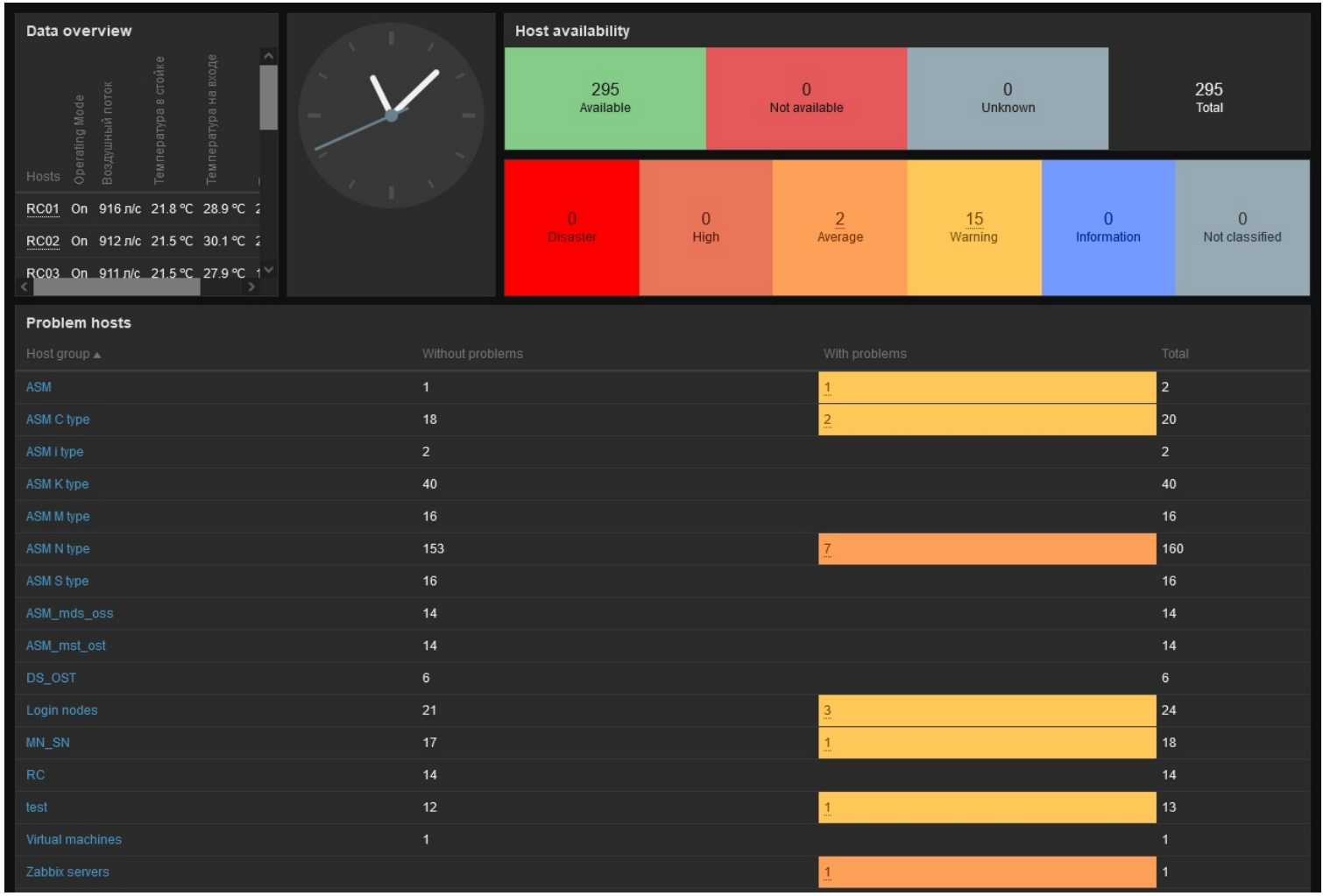




Data Centre Monitoring

Zabbix 5

- Network switches
- Physical nodes
- VMs
- Disk arrays
- Rack °C

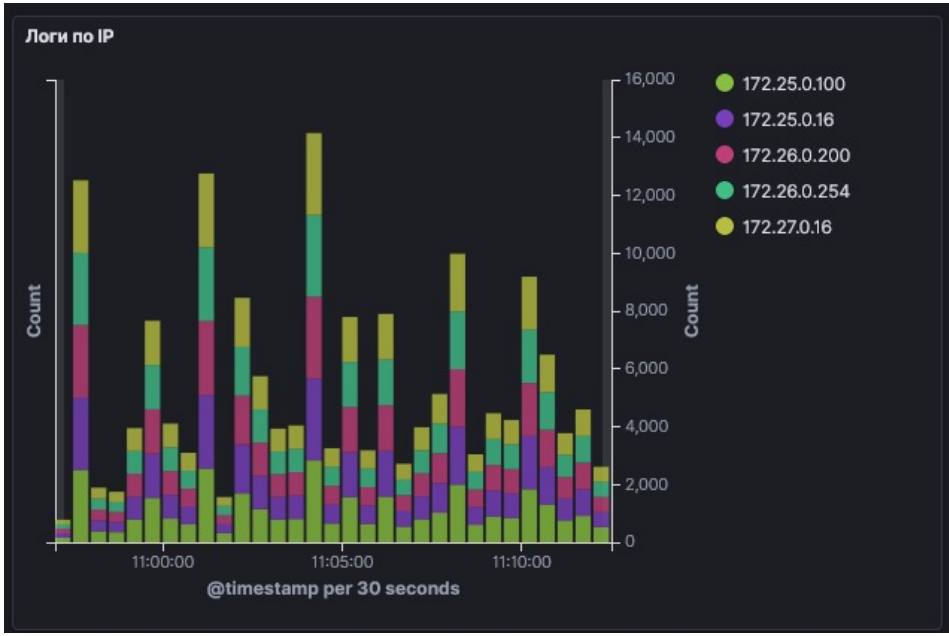


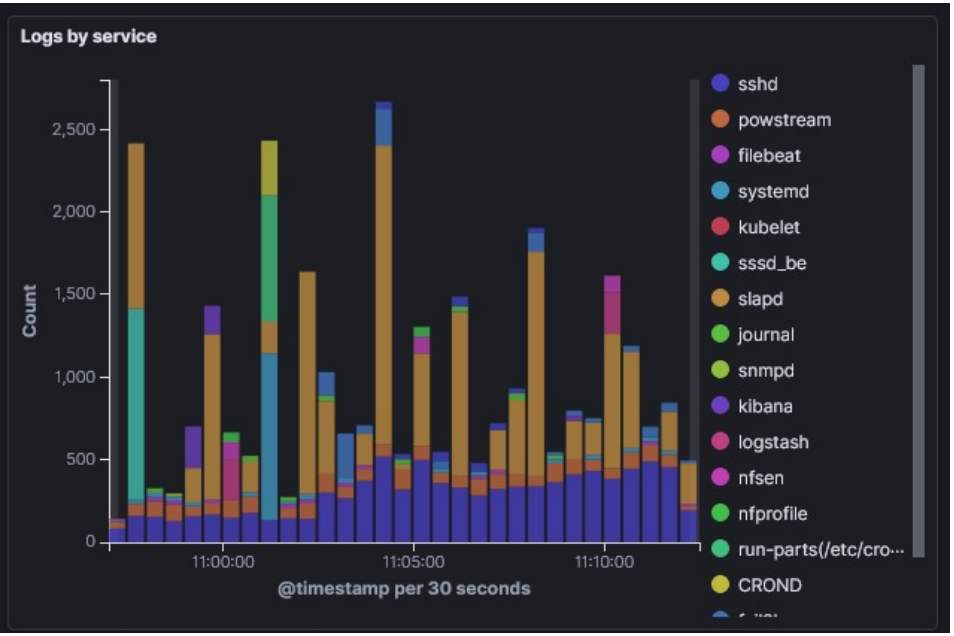
Grafana Heatmap

- Custom made plugin for Grafana
- Data is taken from Zabbix

ElasticSearch & Kibana

- Logs are centrally collected with rsyslog and injected into logstash
  - Covers network equipment as well
  - 2-3 GB of log data daily
- Alerts in Zabbix based on pattern matching

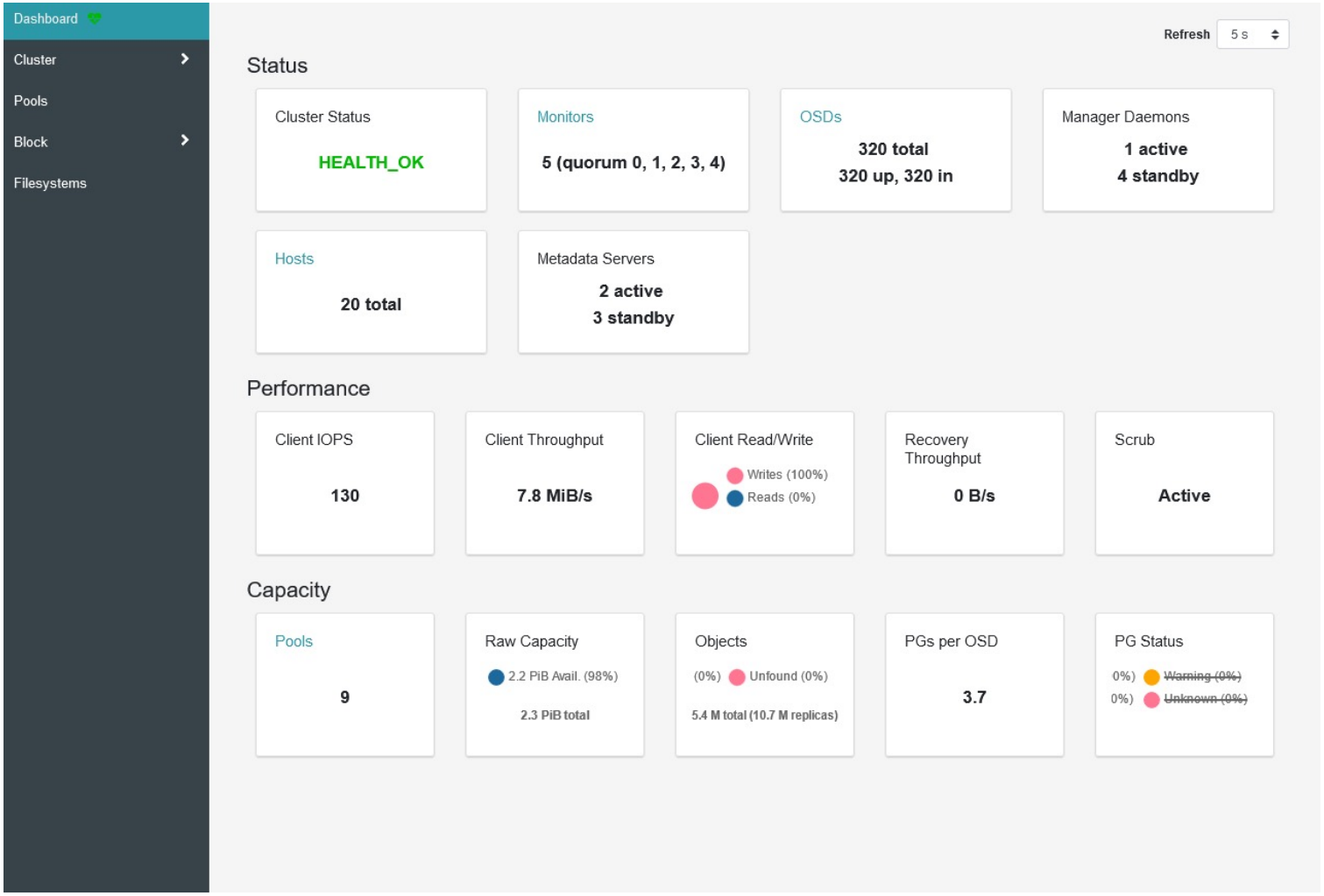




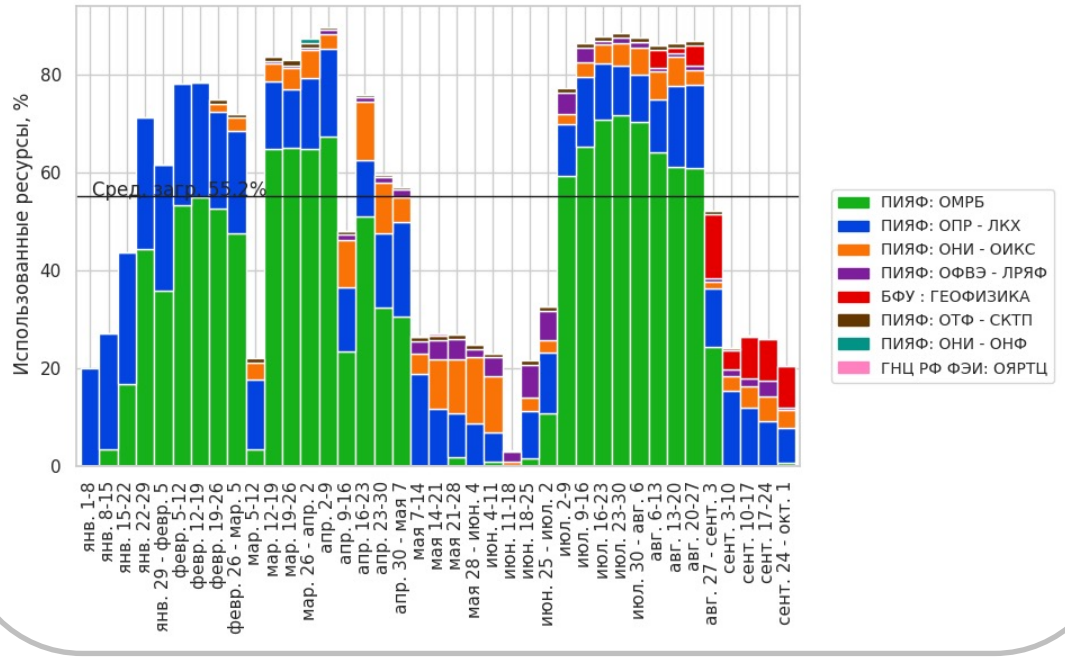
Slurm visual monitoring

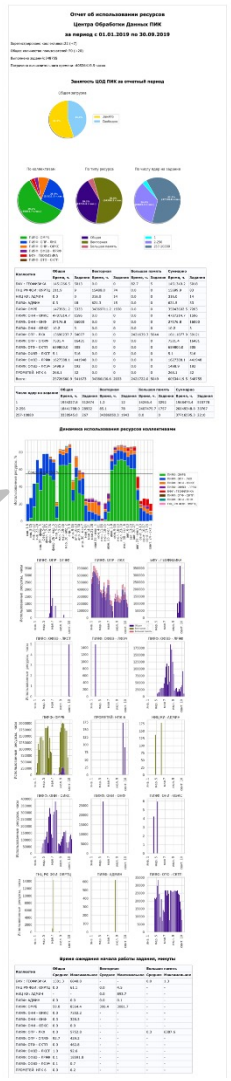
Ceph Monitoring

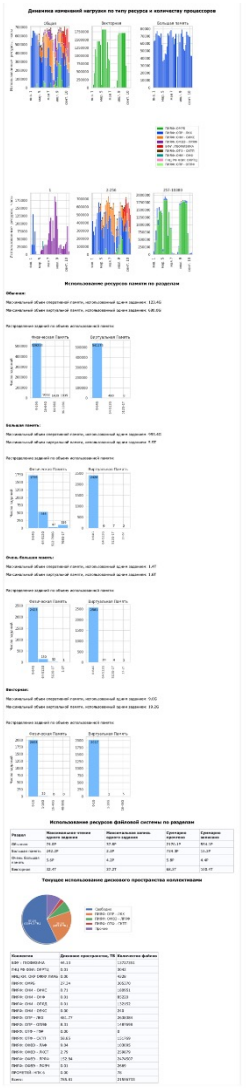
- Built-in
- Python 3
- Requires some manual tweaks on CentOS 7



Accounting







- Mostly Slurm built-in accounting with custom visual reports
- Provides extensive information about user activity and quality of service

Thank you!


APC