IN2P3-CC cloud computing (IAAS) status
Réseau SARI - Feb ‘17
Cloud: the big picture

Remote datacenter resource pools:
- storage
- CPU
- memory
- networks

Cloudified IT services:

**IAAS:** I want servers, networks and block devices

**PAAS:** I want MySQL clusters, email servers

**SAAS:** I want an online EDMS, ERP service

Cloud user « The Devops »

IT services end user
Use cases: R&D, HA and computing clouds

R&D: academic labs and experiments desiring to develop and evaluate new solutions.
→ average availability, mainly private networking, average performance

Public cloud: offload IT services on a public IAAS cloud
Main characteristics:
• GPFS backend: horizontal capacity and IOPS scaling
• LACP rr 2x10Gbps NICs
• Live migration
→ highest availability, public networking, requested performance

Computing: designed to handle HTC workloads
Main characteristics:
• ease software deployment and allows specific environment
• specific job management & computing model implementation
→ low availability, private networking with outgoing internet connectivity, huge CPU cycle accesses (high latency & throughput)
Overall usage

- Projects: 50
- Users: 150
- Hosts: 80
- Aggregates: 16
Computing usage

Computing resources:
- ~1k HT cores in regard with 25k on HTC farm
Core services deployment (HA private cloud)

- Physical: -30%
- VMware: -45%
- Openstack: +187%

Last year:
- Physical: 250
- VMware: 150
- Openstack: 100
- Total: 400
Some implementation details
Openstack deployments

Operational components:
- Keystone
- Glance
- Nova
- Neutron
- Horizon
- Cinder
- Ceilometer
- Swift

In deployment/evaluation:
- Heat
- Magnum
- Rally
- Manila
Hardware resources (as of Feb ‘17)

Deployment:
• CentOS 7
• RDO packages
• Puppet configuration

Compute clusters (May ‘16):

<table>
<thead>
<tr>
<th>Deployment</th>
<th>Cores</th>
<th>RAM</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>544</td>
<td>3.6 TB</td>
<td>36 TB</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>288</td>
<td>1.2 TB</td>
<td>24 TB</td>
</tr>
<tr>
<td>Computing</td>
<td>1216</td>
<td>5.2 TB</td>
<td>40 TB</td>
</tr>
</tbody>
</table>

→ 2048 cores in total, 10TB RAM, 100 TB storage

S3 Storage (Swift):
• 24TB DAS

Ceph Storage (Cinder):
• 480TB DAS
Service architecture

- **Openstack clients**
- **Public HA frontends cluster**
- **Failover**
- **Openstack controllers cluster**
  - APIs services (keystone, glance, nova, cinder, neutron...)
  - schedulers
  - glance registry
- **Active MQ (RabbitMQ cluster)**
- **Galera Cluster (persistent data)**
- **Storage resource pools**
  - Cinder / DAS
  - Cinder / Ceph
  - S3 / DAS
- **CPU/RAM resource pools**
- **Compute aggregates (type segregation)**
- **HA frontends (keepalived + haproxy)**
- **Openstack controllers**
- **AMQP RabbitMQ cluster**
- **Galera database cluster**

→ **No SPOF & horizontally scalable**
## Resource pools overview (as of March '17)

### Computing

<table>
<thead>
<tr>
<th>Pool</th>
<th>Cores</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>816</td>
<td>2.4TB</td>
</tr>
<tr>
<td>Bioaster</td>
<td>64</td>
<td>1.5TB</td>
</tr>
<tr>
<td>Batch</td>
<td>168</td>
<td>0.5TB</td>
</tr>
<tr>
<td>Spark</td>
<td>240</td>
<td>0.7TB</td>
</tr>
<tr>
<td>Core services</td>
<td>280</td>
<td>2TB</td>
</tr>
</tbody>
</table>

Grand total: 2,160 cores, 9.9TB RAM

### HA

<table>
<thead>
<tr>
<th>Pool</th>
<th>Cores</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>288</td>
<td>1.2TB RAM</td>
</tr>
<tr>
<td>eTRIKS</td>
<td>208</td>
<td>1.2TB RAM</td>
</tr>
<tr>
<td>Hosted</td>
<td>96</td>
<td>0.4TB RAM</td>
</tr>
</tbody>
</table>

Grand total: 584 cores, 3.6TB RAM

### R&D

<table>
<thead>
<tr>
<th>Pool</th>
<th>Cores</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>288</td>
<td>1.2TB RAM</td>
</tr>
<tr>
<td>eTRIKS</td>
<td>208</td>
<td>1.2TB RAM</td>
</tr>
<tr>
<td>Core services</td>
<td>280</td>
<td>2TB</td>
</tr>
</tbody>
</table>

Grand total: 288 cores, 1.2TB RAM

### Storage

<table>
<thead>
<tr>
<th>Pool</th>
<th>Axes</th>
<th>Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPFS</td>
<td>80</td>
<td>22TB</td>
</tr>
<tr>
<td>GPFS</td>
<td>50</td>
<td>14TB</td>
</tr>
<tr>
<td>GPFS</td>
<td>25</td>
<td>1.6TB</td>
</tr>
</tbody>
</table>

Grand total: 90 axes, 502TB capacity

### Cloud Storage

<table>
<thead>
<tr>
<th>Pool</th>
<th>Axes</th>
<th>Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceph</td>
<td>30</td>
<td>22TB</td>
</tr>
<tr>
<td>LVM/DAS</td>
<td>15</td>
<td>6TB</td>
</tr>
</tbody>
</table>

Grand total: 45 axes, 37.6TB capacity
Neutron migration
From legacy networking to Neutron, what’s the challenge?

- 50 different VLANs served (1 ↔ 1 projects)
- 6 different modes
  - Private isolated
  - Private internally routed
  - Private SNATED to the outside world
  - Floating IPs (SNAT/DNAT)
  - Public
  - Public distributed into existing networks
- Virtualized (non cloudish) production services implemented
- No supported migration process: you just rely on what’s been done by others
- Linux bridges to OVS
- Static network provisioning to SDN
- HA/DVR
Network type 1
- private subnet
- not routed

Use case:
Likely no one
Network type 1 bis
- private addressing
- not routed
- floating ip

Use case:
- webimatics
- LSST/Euclid
- ecole.info/jenkins/...

Allow internet/cc connectivity
Network type 2
- private addressing
- routed

Use case:
- ccin2p3
- htc-atlas

Allow CC connectivity
No NAT for performance
Spare public IPs
Network type 2 bis
- private addressing
- routed
- floating IPs

Use case:
- ccin2p3
- htc-atlas

Allow CC/internet connectivity
Spare public IPs
Network type 3
- private addressing
- routed
- NAT to single IP

Use case:
- htc Wns (atlas, euclid, LSST, bioaster...)

Lower load on nova-network
Spare lots of public IPs
Network type 4
- public addressing
- routed

Use case :
- infra-services
- bioaster
- etriks

Legacy networking mode
No NAT (perf)
Public servicing
Network type 5
- public addressing
- routed
- shared with existing deployments

Use case:
- heberge

Legacy networking mode
No NAT (perf)
Public servicing
Monitoring:

Nagios
ELK
Grafana
### Current Network Status

<table>
<thead>
<tr>
<th>Host</th>
<th>Services</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ckeystone</td>
<td>Check Keystone Service</td>
<td></td>
</tr>
<tr>
<td>ccosd01</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd02</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd03</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd04</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd05</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd06</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd07</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd08</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd09</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd10</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd11</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd12</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd13</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd14</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd15</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd16</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd17</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd18</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd19</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
<tr>
<td>ccosd20</td>
<td>Check Cinder volumes</td>
<td></td>
</tr>
</tbody>
</table>

**Status Grid For Service Group 'openstack-servicegroup'**

**OPENSTACK (openstack-servicegroup)**

- **Up**: 38
- **Down**: 0
- **Unreachable**: 0
- **Pending**: 0

- **All Problems**: 0
- **All Types**: 48

**Service Status Totals**

<table>
<thead>
<tr>
<th>State</th>
<th>Critical</th>
<th>Warning</th>
<th>Unknown</th>
<th>All Problems</th>
<th>All Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Down</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unreachable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pending</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
What's next?
What’s next in the mid-term

- Public cloud platform enhancements (storage & CPU): targeted usage within 3 years: 300 instances, 1.4TB RAM, 800 vCPUs, 50 TB storage
- FSS with Synergy for computing
- Finalize resources orchestration (Heat)
- Container orchestration (Magnum/Kubernetes)
- Cinder refactoring with Ceph backend
- AAI integration
- Interest in shared FS aaS (Manila)
Questions?

(thank you)

Questions?
• Updates : better with puppet
• Cout ETP → taille du déploiement