DEMO: *introduction to managing a HPC cluster*

{ know it; manage it

Maria Verina

Abdus Salam International Center for Theoretical Physics (ICTP)
Trieste, Italy

The Information and Communication Technology Section (**ICTS**)

Member of HPC team

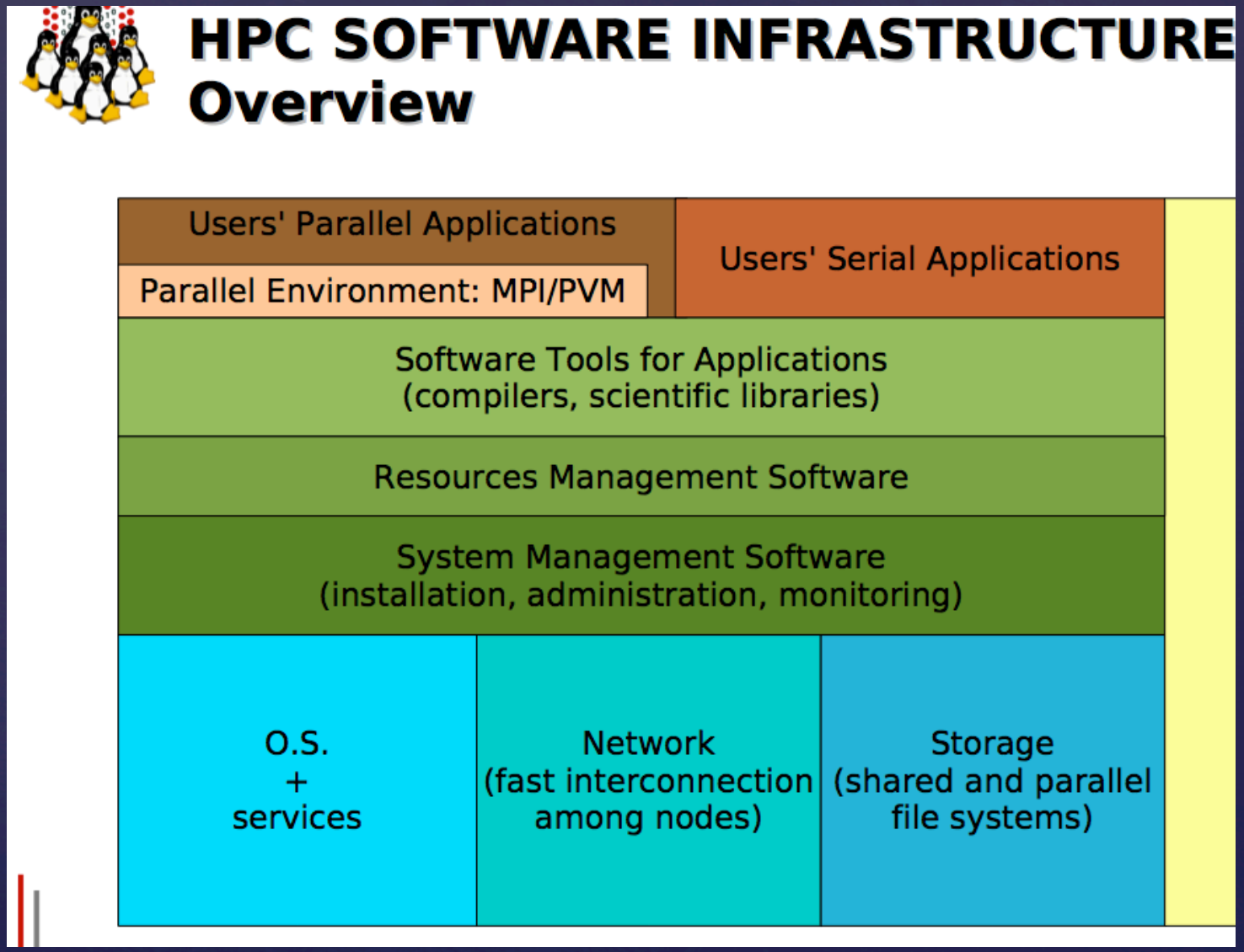
Managing HPC Cluster

- "Rocks: make clusters easy" (for users)
- system admin grows to HPC-sysadmin role:

  initial effort to install/config;
  Production (monitor and troubleshoot);
  forget about it (almost)

- new concepts/terminology: batch job, queue, scheduler, MPI, IB,... and the overall architecture
- complexity: number of elements (e.g. cables)

The Big Picture

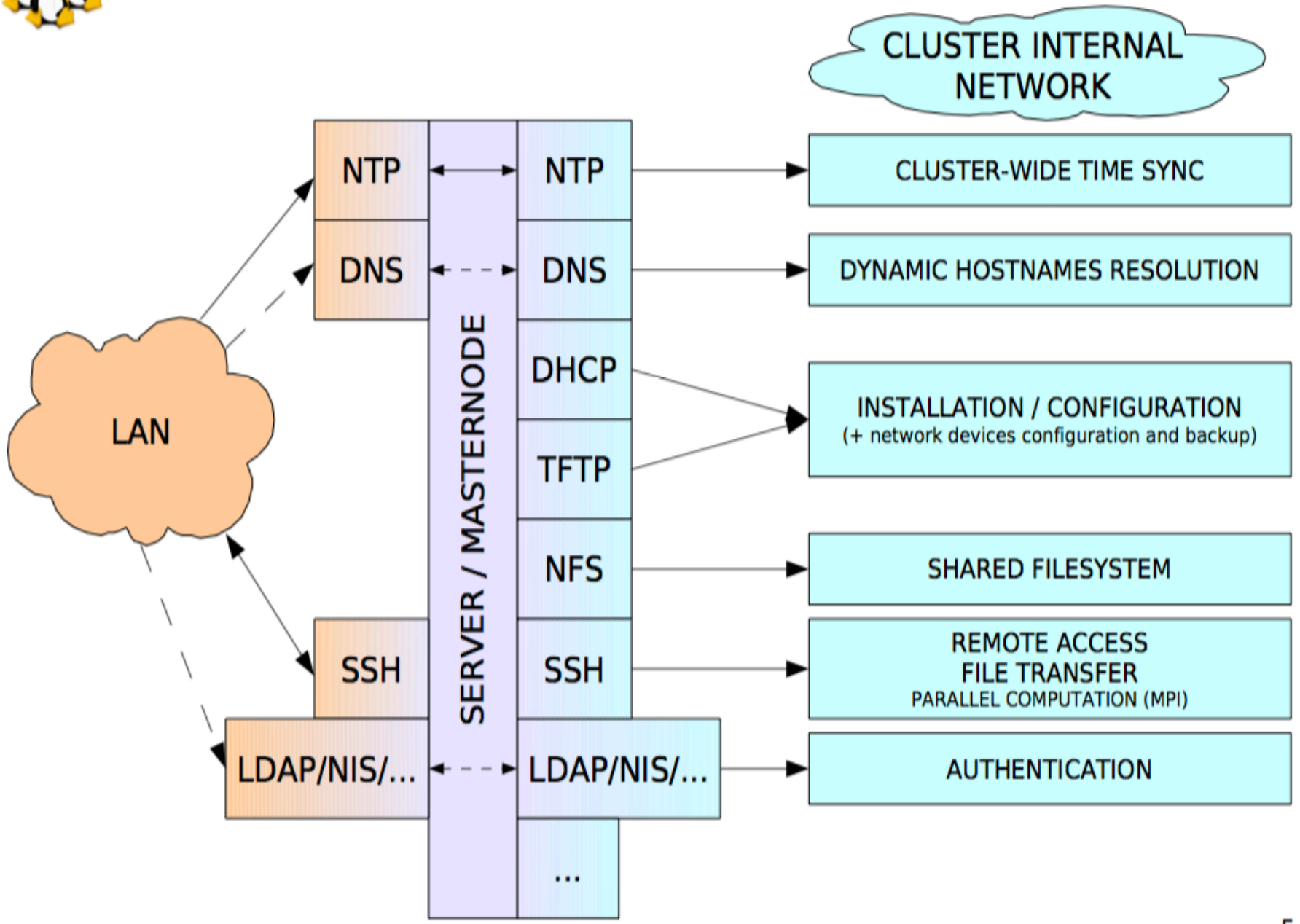


credit: Moreno B.

Recap: What runs where?

- On Master node:
  - OS
  - Resource manager/ Scheduler
  - Installation services
  - Compilations
  - MPI ?
  - …
- on Compute node:
  - OS
  - Resource manager client
  - User apps
  - MPI
  - ganglia monitoring client

CLUSTER SERVICES



5

credit: Moreno B.

Services running on a typical Master node

- DNS(/etc/hosts)
- DHCP
- TFTP
- NTP
- NFS (opt)

- sshd
- Torque and Maui
- Ganglia
- Syslog
- http (opt)
- …

Cluster views and common language

- User view (need and like limited structure view)
  - "Power" user can obtain more performance

Common language:
job, jobID,
node name,
"why the job is not running?",
"why the job is not producing output",
job script, working directory

- HPC sysadmin view: all the architectural details, known issues

Some useful commands

- during node installation, you can ssh there and see logs in /tmp of the progress. If not, attach real console, and see problems

```
# rocks set host boot compute-0-1 action=install

# rocks list host boot
HOST        ACTION
compute-0-0: os
compute-0-1: install
```

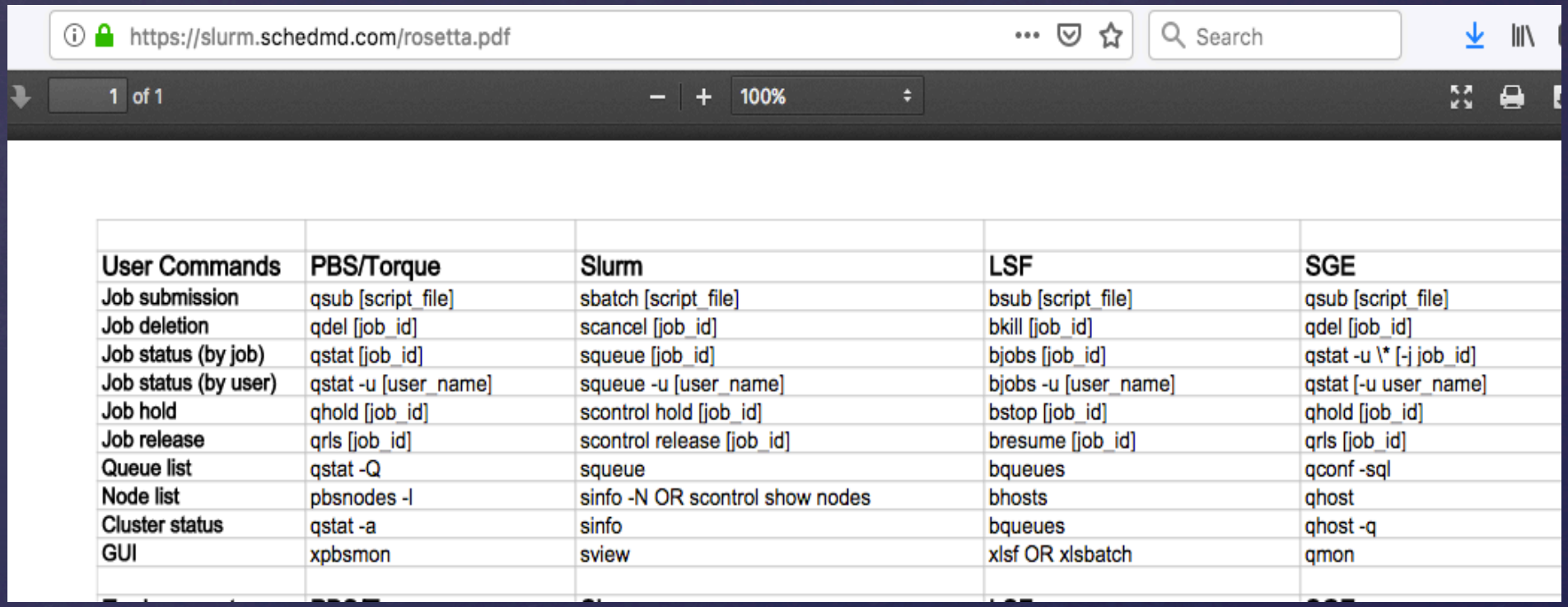
Commands (cont)

```
# rocks list network
NETWORK  SUBNET       NETMASK       MTU   DNSZONE SERVEDNS
private: 10.1.85.0    255.255.255.0 1500  local   True
public:  140.105.32.0 255.255.255.0 1500  ictp.it False

# rocks list host
HOST         MEMBERSHIP CPUS RACK RANK RUNACTION INSTALLACTION
hpcdemo:   Frontend   12   0    0    os       install
compute-0-0: Compute    40   0    0    os       install
compute-0-1: Compute    40   0    1    os       install

# rocks list roll

# useradd testuser
# passwd testuser
# rocks sync users

# rocks run host compu  "shutdown -r now"  or ssh compu "shutdown –r now”
```

Commands (cont.)

- describe hosts (node list): # qhost
- cluster status: # qhost -q
- see all queues: # qstat -g c
- see jobs for all users: qstat –u “*”
- queue list: # qconf -sql
- List exe nodes: qconf -sel
- subset running jobs: [user1@master ~]$ qstat -s r
- for problems use explain $ qstat -s r -explain [a|c|A|E];
  # qstat -u user1   -explain A

“Rosetta stone”



User Commands	PBS/Torque	Slurm	LSF	SGE
Job submission	qsub [script_file]	sbatch [script_file]	bsub [script_file]	qsub [script_file]
Job deletion	qdel [job_id]	scancel [job_id]	bkill [job_id]	qdel [job_id]
Job status (by job)	qstat [job_id]	squeue [job_id]	bjobs [job_id]	qstat -u * [-j job_id]
Job status (by user)	qstat -u [user_name]	squeue -u [user_name]	bjobs -u [user_name]	qstat [-u user_name]
Job hold	qhold [job_id]	scontrol hold [job_id]	bstop [job_id]	qhold [job_id]
Job release	qrls [job_id]	scontrol release [job_id]	bresume [job_id]	qrls [job_id]
Queue list	qstat -Q	squeue	bqueues	qconf -sql
Node list	pbsnodes -l	sinfo -N OR scontrol show nodes	bhosts	qhost
Cluster status	qstat -a	sinfo	bqueues	qhost -q
GUI	xpbsmon	sview	xlsf OR xlsbatch	qmon

Let’s see some demo!

Conclusion:

HPC sysadmin's life is ritch with chalenges

- Initial effort (new concepts, architecture)
- Problems become Known Problems
- Operate a Production cluster

Thank You!

Maria Verina

Questions ?

Common files view cluster-wide

- user accounts (/etc/passwd)
- hostnames (/etc/hosts)
- mount points (/etc/auto.home)
- ...

- Pushed to nodes at regular intervals
- and at boot time