IHEP Site Status

Jingyan Shi , shijy@ihep.ac.cn

Computing Center, IHEP

2015 Spring HEPiX Workshop

Outline

- Infrastructure update
- Site status
- Errors we've suffered
- Next plan

Infrastructure Update -- Since Mar. 2014

- Local cluster
  - Cpu Cores
    - New added: 1416
    - The amount cpu cores of local cluster: 10936
  - Storage
    - New added: 810TB
      - New device: DELL MD3860F - DDP disk arrays
        - » With the expectation to decrease disk rebuild time
    - The amount of storage: 4PB
  - Core switch
    - Old Switch “Force 10” was replaced by the one borrowed from vender temporarily

Infrastructure Update -- Since Mar. 2014 (cont.)

- EGI site
  - All grid services migrated to VM on new machines
  - All disks replaced by 4TB*24 array.
  - All storage servers replaced by new machines
  - Total disk capacity increased to 940TB
  - All old work nodes (1088 cpu cores) will be replaced by the new ones

Outline

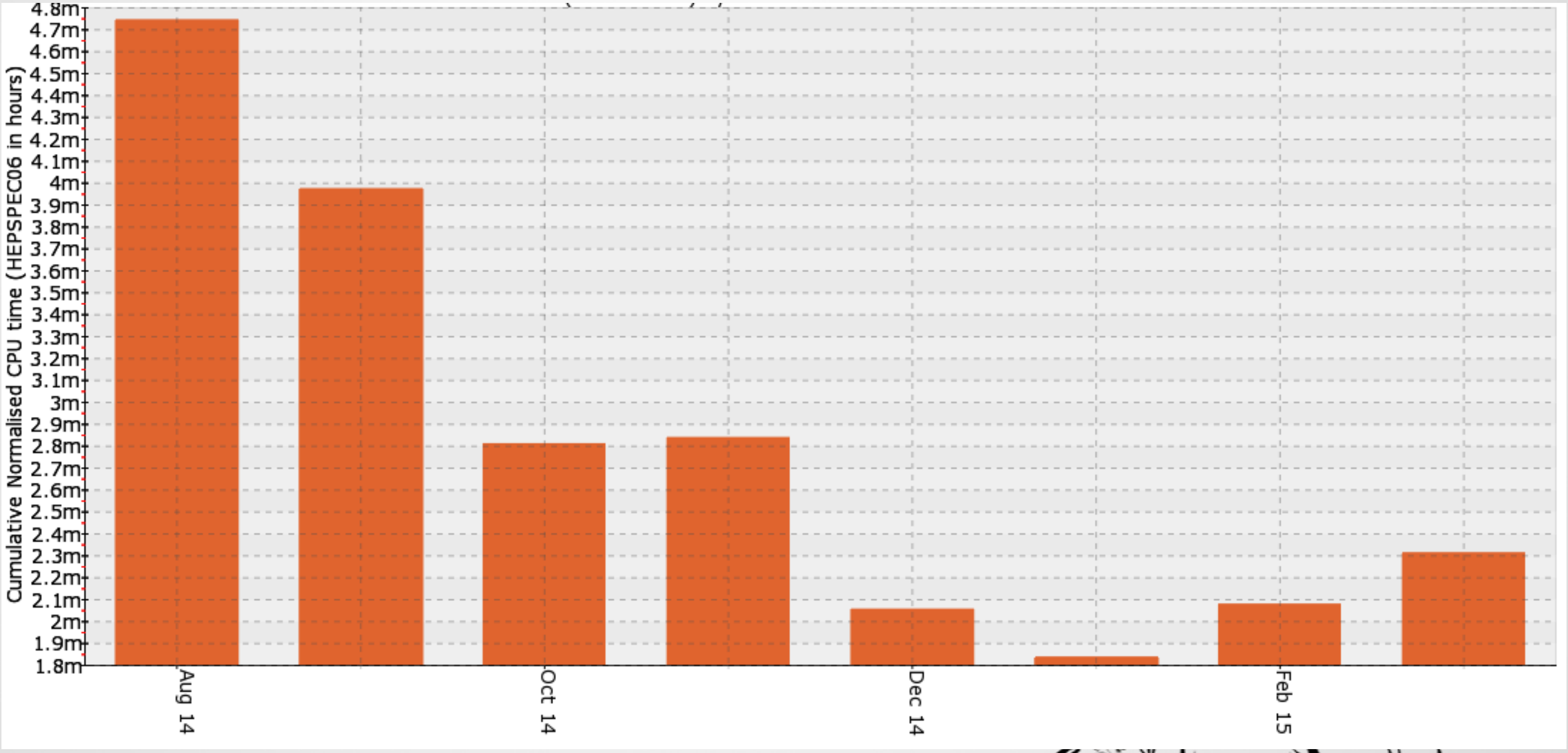
- Infrastructure update
- Site Status
- Errors we suffered
- Next Plan

Scheduler -- HTCordor

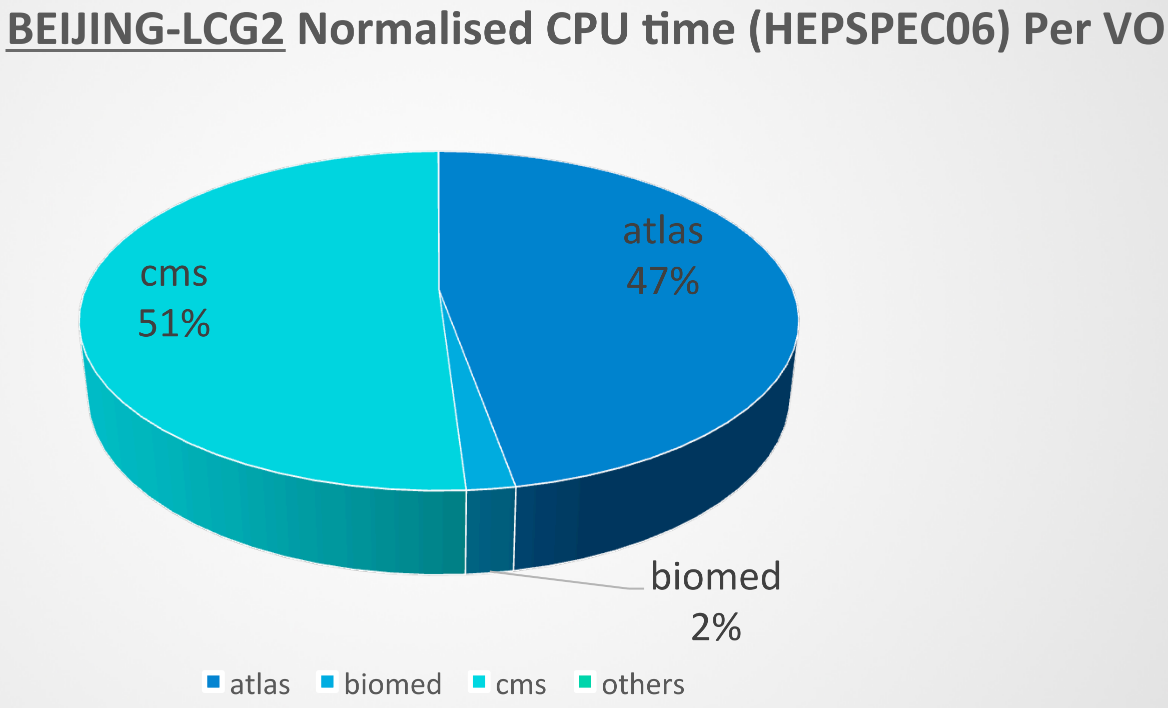
- Small cluster created and managed by HTCondor last month
  - 400 cpu/cores
  - One experiment (JUNO) supported
  - Running well

EGI site

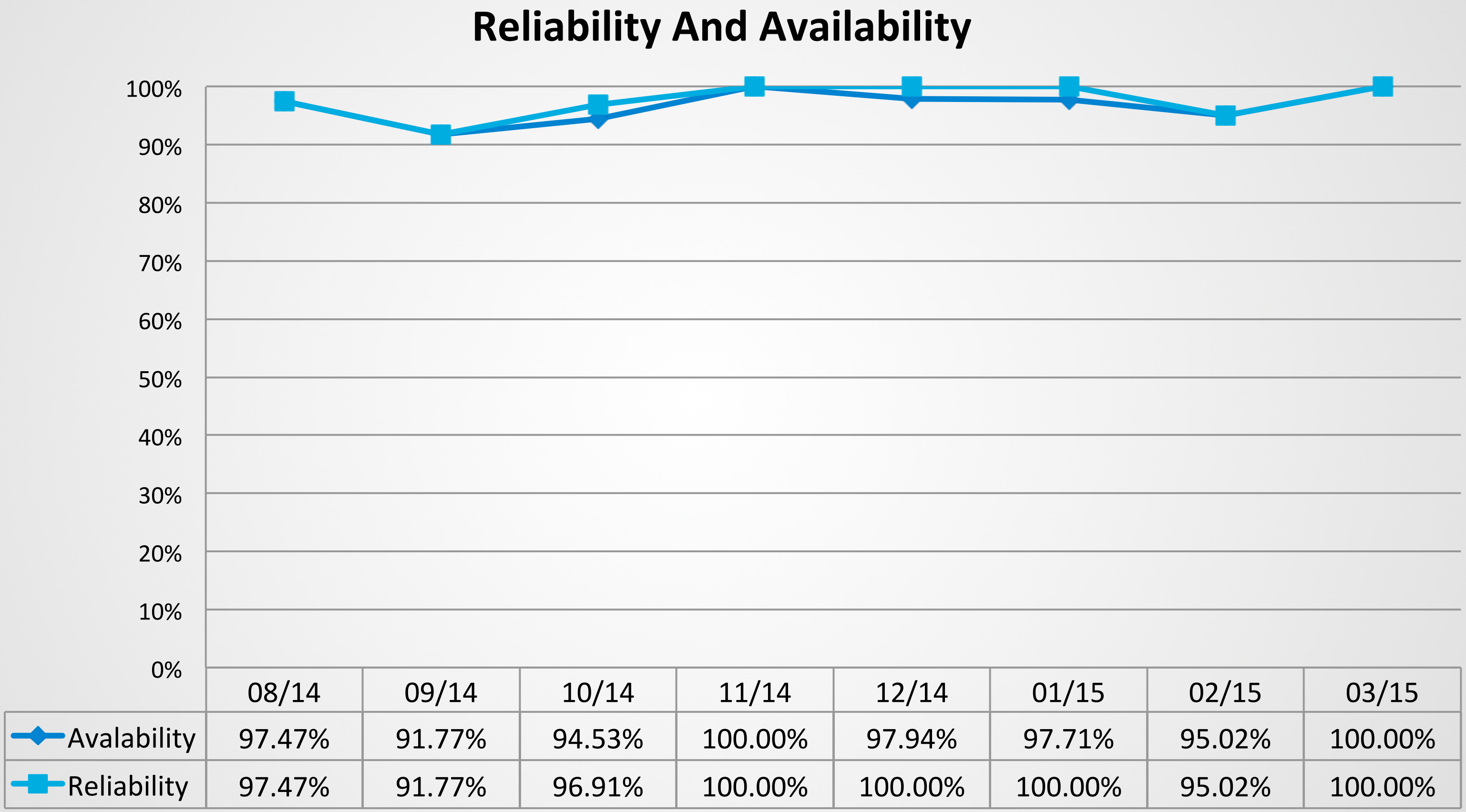
BEIJING-LCG2 Cumulative CPU Time



EGI site (cont.)



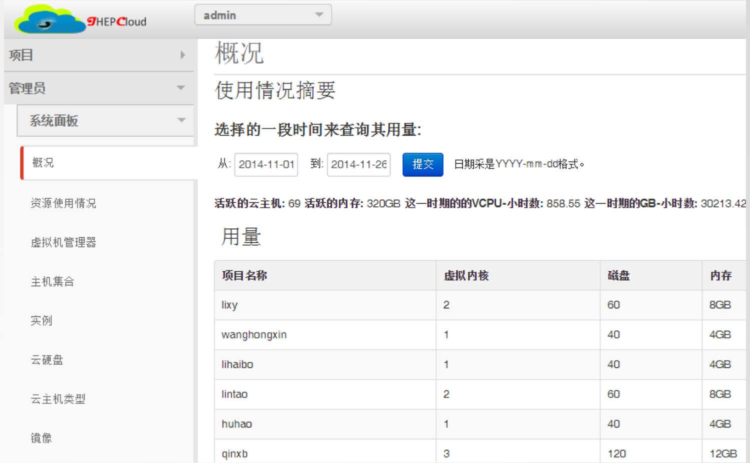
Reliability and Availability



	08/14	09/14	10/14	11/14	12/14	01/15	02/15	03/15
◆ Avalability	97.47%	91.77%	94.53%	100.00%	97.94%	97.71%	95.02%	100.00%
■ Reliability	97.47%	91.77%	96.91%	100.00%	100.00%	100.00%	95.02%	100.00%

IHEP Cloud established

- Released on 18th. November, 2014
- Built on openstack Icehouse
- 8 physical machines: 224 vm capacity
  - 1 control node, 7 computing nodes
  - User applies and gets the VM on line



概况

使用情况摘要

选择的一段时间来查询其用量:

从: 2014-11-01 到: 2014-11-26 日期采是YYYY-mm-dd格式。

活跃的云主机: 69 活跃的内存: 320GB 这一时期的的VCPU-小时数: 858.55 这一时期的GB-小时数: 30213.42

用量

项目名称	虚拟内核	磁盘	内存
lixy	2	60	8GB
wanghongxin	1	40	4GB
lihaibo	1	40	4GB
lintao	2	60	8GB
huhao	1	40	4GB
qinxb	3	120	12GB

- Three types of VM provided
  - Provide user VM that same as login node
    - DNS and IP management , Email and AFS account, puppet, NMS, ganglia …
  - Provide user VM with root right and no public IP
  - Provide administrator VM with root right and public IP
- Current Status
  - Active 172 VM, 628GB memory and 4.7TB disk

Outline

- Infrastructure update
- Site Status
- Errors we suffered
- Next Plan

Errors we’ve suffered – core switch

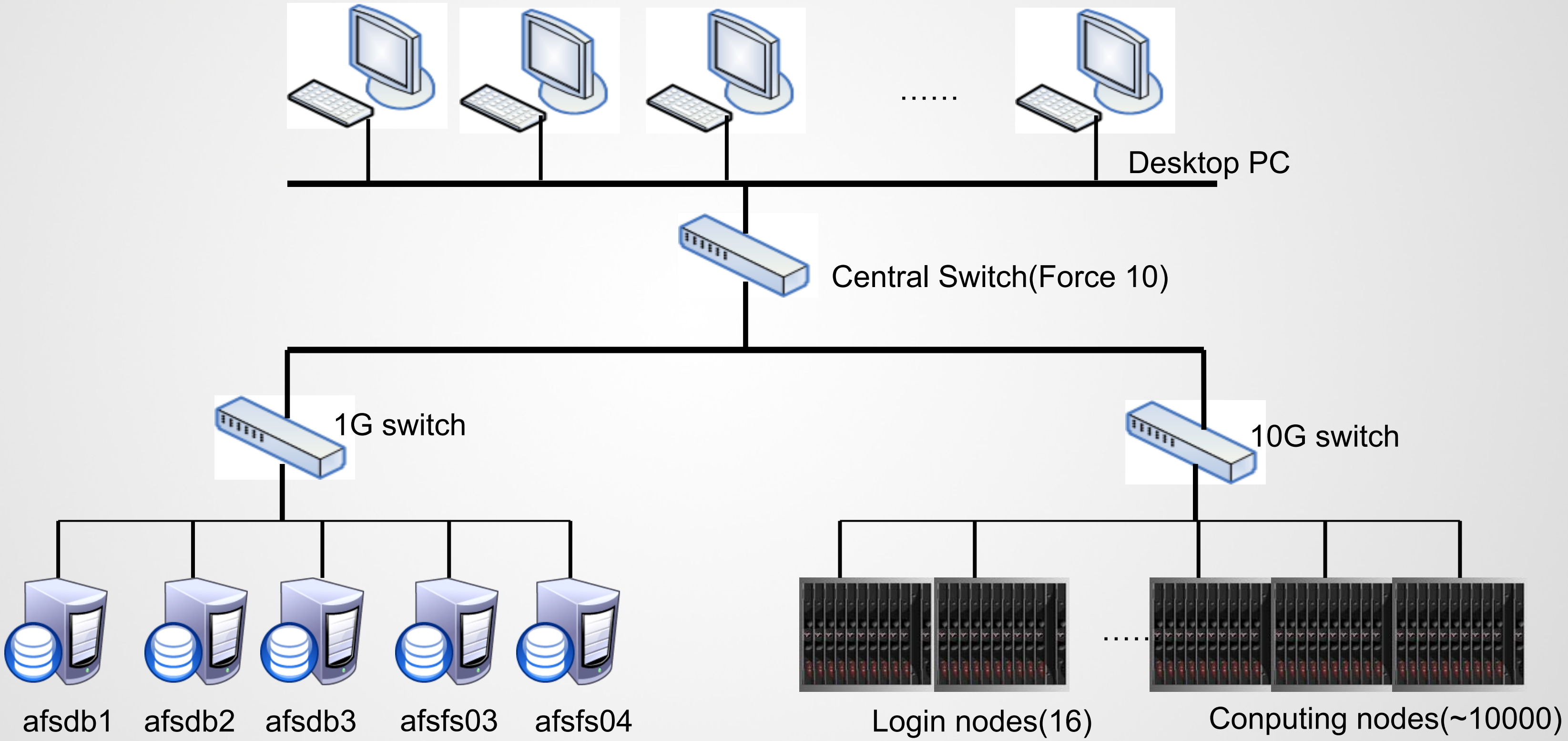
- Service/Device Online: 2006 ~ 2014.10, So Long Time…
- Capability problems
  - Port density
  - Backplane switching capability
- Despite of the problems, it had been running stable
- Suffered Reliability problems in 16*10G cards by the end of last year
  - No Switching
  - Loss Packets
  - No backup cards from the vendor

Errors we've suffered -- core switch (cont.)

- Why ?
  - Force10 E1200 should be ended
  - Unstable Network →Unstable Computing
- Choices
  - Too many choices for vendors
  - Least expensive technically acceptable
  - Production Environment Evolution
- Production & Evolution vendors and devices
  - Ruijie Networks: Core Switch / RJ18010K,TOR/RJ6200
  - Huawei Networks: Core Switch / HUAWEI12810
  - Brocade
  - DELL-Force10
- Ruijie won the bid

Errors we've suffered -- AFS

- AFS deployment



1. Master database: afsdb1
2. Slave database:afsdb2,afsdb3
3. Fileserver:afsdb2,afsdb3,afsfs03,afsfs04
4. Total size: 7.8TB

1. AFS client installed in all login nodes
2. Tokens when login using PAM
3. UID and GID stored in /etc/passwd file, no password in /etc/shadow

1. AFS cache set 10GB
2. Jobs scheduled to computing nodes access software lib in AFS

Errors we've suffered – AFS (cont.)

- AFS status
  - Used as home directory and software barn
- All HEP libraries in AFS have copies to make sure data availability
- Errors we suffered
  - AFS file service crashed down irregularly (Due to its old version?)
    - Inconsistent replica leaded jobs failure when job read the wrong replica
    - Failed to release replica volume when the fileserver service crashed
- Solutions
  - Add monitoring to put AFS file service under surveillance
  - Plan to upgrade Openafs during summer maintenance
    - Upgrade testing is being ongoing

Errors we’ve suffered - network card

- Most of work nodes had been upgraded from SL5.5 to SL6.5 last year
  - The drivers of some network cards didn’t work properly
    - Jobs failed due to un-stability of those network cards
  - Errors disappeared after driver upgrades

Outline

- Infrastructure update
- Site status
- Errors we’ve suffered
- Next plan

Next plan -- HTCondor

- Monitoring and optimization
  - Integrated with our job monitoring and accounting tool
  - Optimization need to be done
- Scale would be expanded

Next plan -- Storage

- Hardware
  - 5 years old disk array (740TB) will be replaced by a new set of servers connected with DELL DDP disk arrays

Next plan -- Storage (cont.)

- Software
  - 3 years old disk array(780TB) will be reconfigured to a dual-replication file system powered by gLusterfs
    - Test has been done
    - More stable
  - To avoid predictable incompatibility with newly purchased hardware, all the file system severs and clients will be upgraded to Lustre 2.x over SL 6 this year.
    - Currently Lustre 1.8 over SL5,

    fall behind the Linux trend

Next Plan -- Monitoring

- Flume + Kibanna deployed to collect logs generated by servers and work nodes
  - Logs collected from 1000+ devices on line
  - Gave a lot of help when NIC error happened
  - Need to be optimized
- Nagios has been used as the main monitoring tool
  - One Nagios server is not enough to show the errors and their recoveries in time
  - New monitor plan is under going

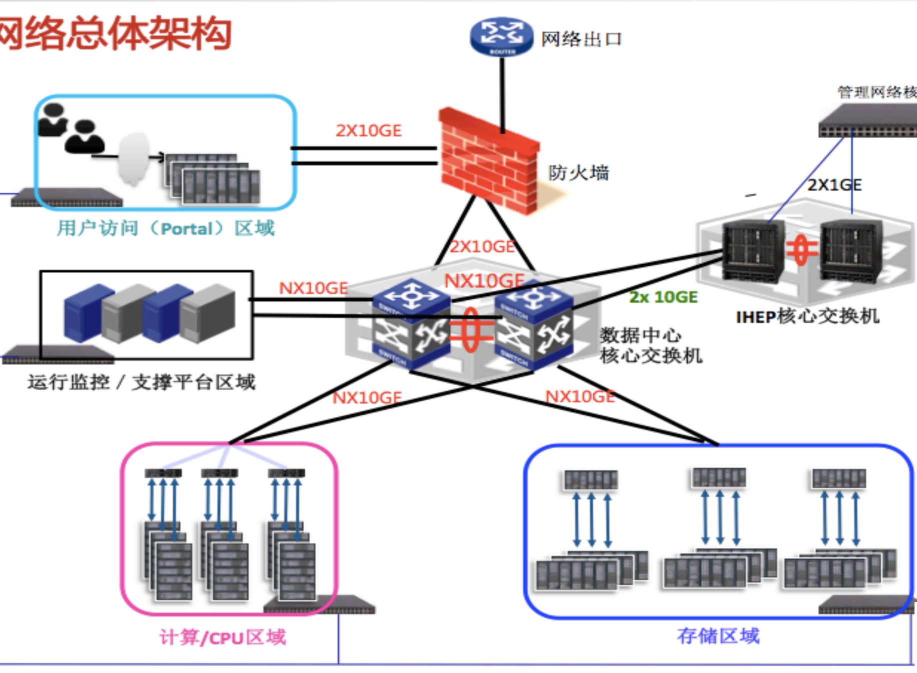
Next plan -- puppet

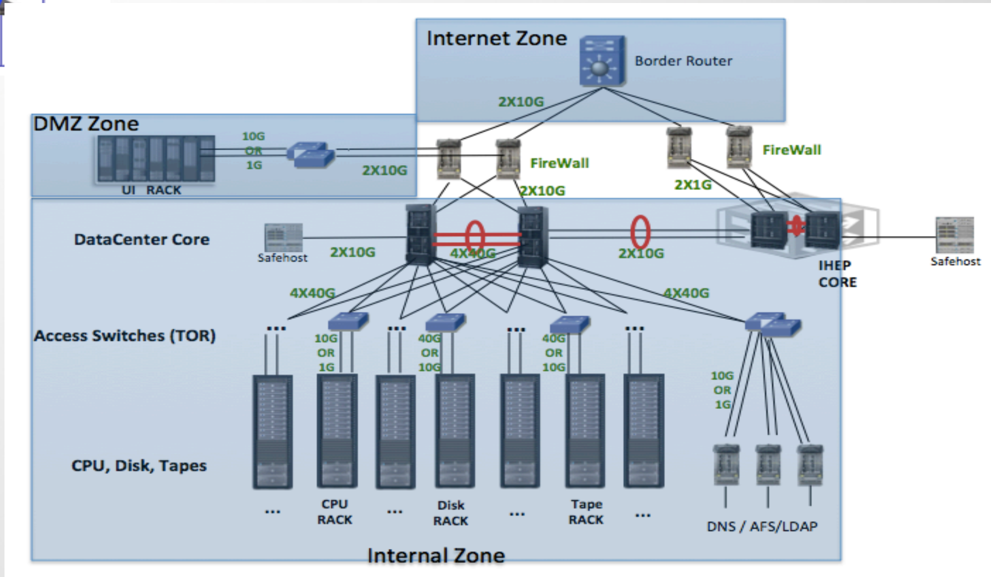
- OS and software running on most of devices are installed and managed by puppet
- Performance problem
  - Long waiting time to upgrade software of 1000+ servers
- Optimization has been done
  - Less than 40 min to upgrade 1000+ machines
  - Need more optimization

Next plan -- network

- Double Core switches & Firewalls
  - High Capability , Stability, Easy Management…
- Better Backbone Bandwidth
  - 160Gbps(4X40Gbps) for Layer2 Switches
  - 2X10Gbps for Storage Nodes
- Clear Zones
  - Internal: Computing/Storage/AFS/DNS/Monitoring
  - DMZ: Public Servers/Login Nodes/…
  - Internet: Network Performance Measurement nodes
- Two networks
  - Data Network
    - High Stability & Capability
  - Management Network
    - High Stability

Next Plan -- network (cont.)





Next plan -- IHEP Cloud

- Integrated with local cluster
- Jobs submitted to dedicated queue of local cluster would be migrated and run at IHEP Cloud automatically
  - Keep the way same as that of pbs to run job
  - No any extra modification requested to user
  - VM could support short peak requirement from the experiment
- Resource of VM would be expanded or decreased depending on the amount of local cluster jobs
- Under development and will be released next month

Thank you!

Question?