Lecture 4: parallel computing, batch systems, cluster monitoring

https://sites.google.com/site/clustergateorg/

- Parallel performance
- Job — task to perform data processing
- Batch processing

- Batch processing systems

Andrey.Shevel@pnpi.spb.ru

Parallel computing

- The cluster (and even one server) consists of many independent components. It gives idea that a range of operations in computing (in general data conversion) might be divided on some independent steps which can be performed at the same time (in parallel).
- It seems parallel operation performance will decrease total time of computing. Quite often it takes place. But not each time.
- In reality parallel computing could decrease the total computing time in that degree in which we can divide our task in separate stages.

Types of parallelism

- **Algorithm pallelism - parallel (almost at the same time) execution of different part of the program (program system).**
- **Data parallelism - parallel (almost at the same time) execution of one program with different data**
  - Statistics values (mean, sigma, etc) in different data sets.

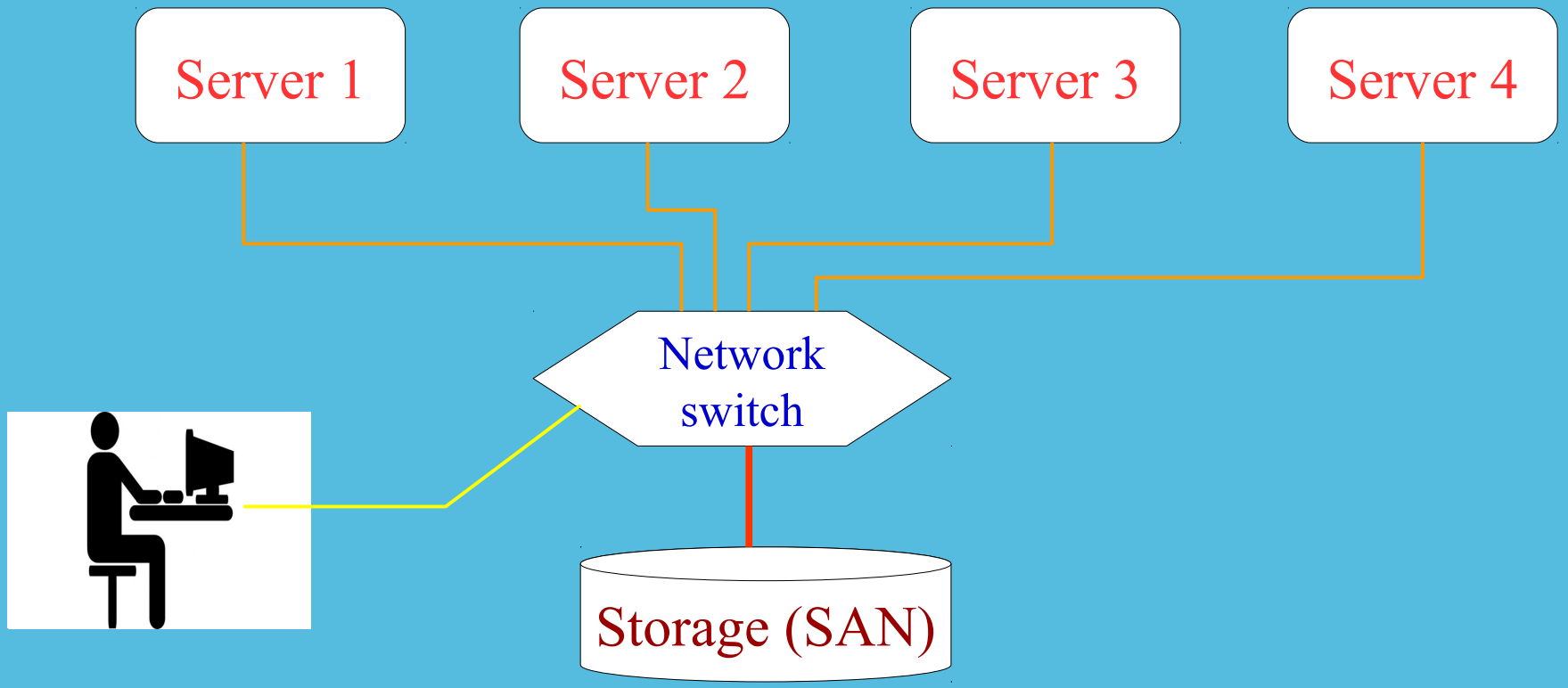
Amdal's law (~1960)

- If *alpha* is share of algorithm (or programs, or task), which can be performed in parallel on N CPUs, than maximum speedup can be achieved on level *1/alpha* even N becomes unlimited. (http://en.wikipedia.org/wiki/Parallel_computinghttp://en.wikipedia.org/wiki/Parallel_computing).

Which situation in computer infrastructure might prevent parallel performance?

What is possible speedup for data processing if we upgrade all servers to more powerful CPU?

CPU load 100% for all servers



Storage bandwidth load is 100%

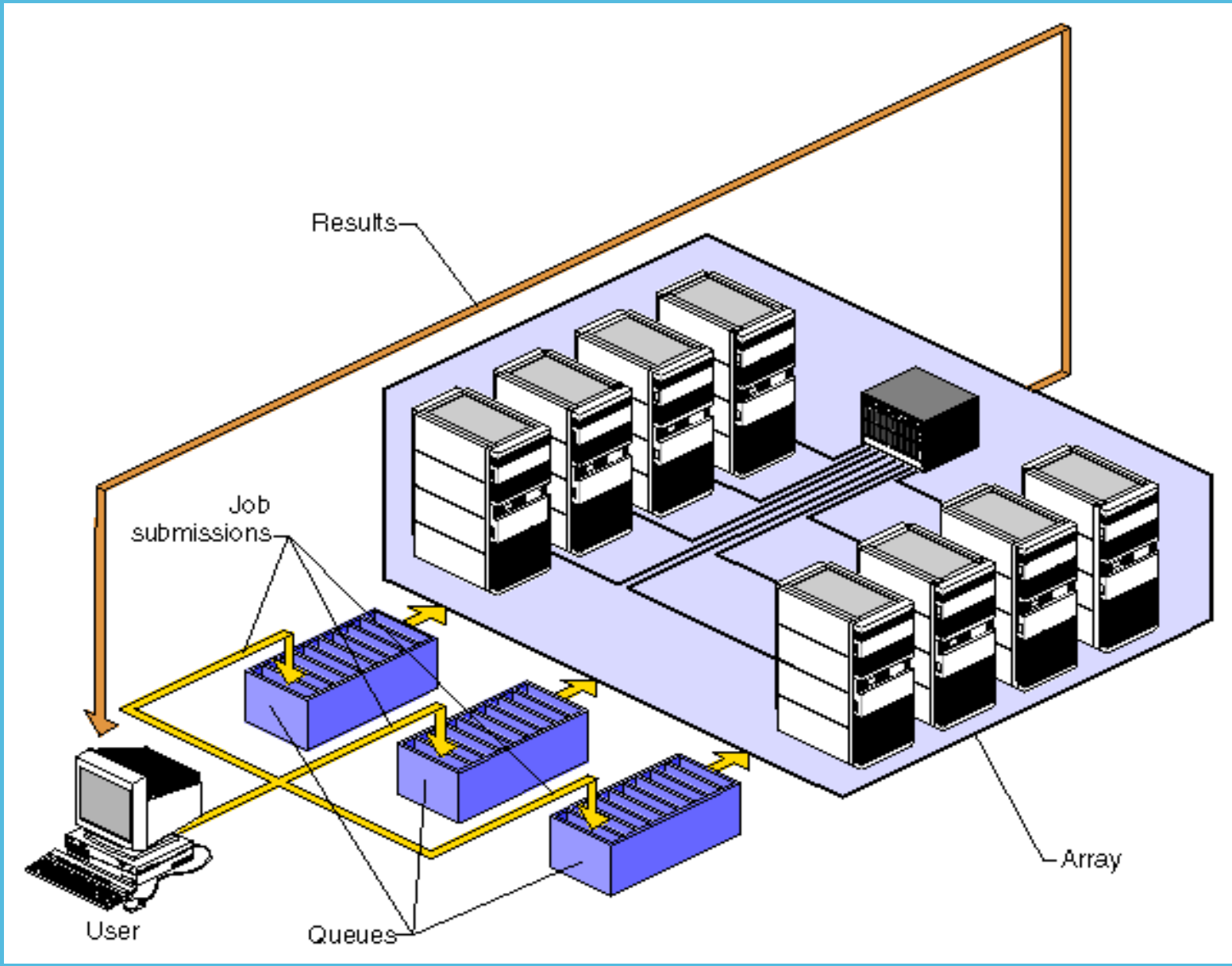
Parallel performance of computing jobs

- **Job** (request to perform something) – the description, which is interpreted by ***batch processing system***. Job might be interpreted and performed on one server or on the cluster. One job is accomplished in finite time. Often one job is part of large computing process.
- On the cluster there are many jobs from a number of users (customers) at the same time. So we can say about stream of jobs (or **batch**).
- To maximize the cluster usage (maximize the number of accomplished jobs per unit of time) the ***batch processing system*** is used.

Job queues

- If you have more jobs than it is possible to perform at the same time you need to put them into ***input queue, the batch system will get jobs from input queue when a cluster node becomes free***.
- **Usually jobs enter into batch processing system in one of the possible input queues.**
- **It has to be the rule how the concrete job is entered into concrete queue.**
- **Also it has to be the rule which input queue is more preferable in concrete time.**

Batch processing system functionality



Andrey.Shevel@pnpi.spb.ru

Examples of batch processing systems

- **PBS/Torque** (http://en.wikipedia.org/wiki/Portable_Batch_System, http://en.wikipedia.org/wiki/TORQUE_Resource_Manager)
- **Condor** (http://en.wikipedia.org/wiki/Condor_High-Throughput_Computing_System)
- **LSF** (http://en.wikipedia.org/wiki/Platform_LSF)
- **SGE** (http://en.wikipedia.org/wiki/Oracle_Grid_Engine)
- **SLURM** - https://computing.llnl.gov/linux/slurm/
- https://en.wikipedia.org/wiki/Comparison_of_cluster_software

Several API for batch processing

- Qsub – submit the job
- Qstat – get the jobs status
- Qdel – delete the job from any queue.

Message Passing Interface (MPI)

- **It is possible to use some library to create the program with parallel parts? YES, it is.**
  - http://www.open-mpi.org/
- **http://cluster.linux-ekb.info/mpi2b.php**

Cluster monitoring

- **Why the monitoring is required?**
  - Ganglia - http://ganglia.sourceforge.net/
  - Zabbix - http://www.zabbix.com/
  - Nagios - https://www.nagios.org/
  -

Ganglia view

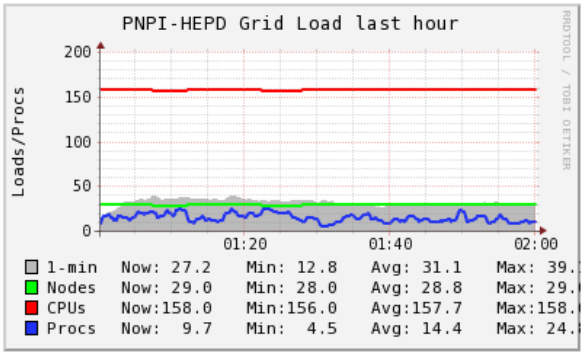
PNPI-HEPD Grid (12 sources) (tree view)

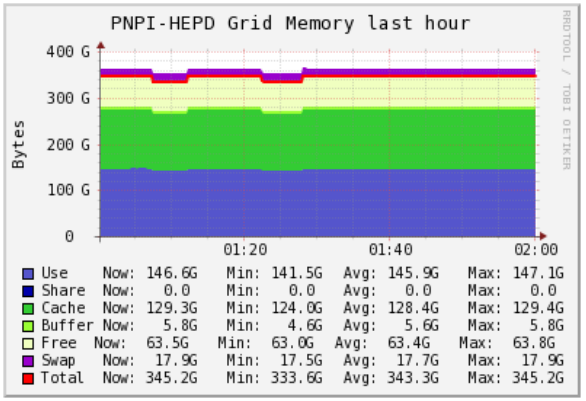
CPUs Total:	**158**
Hosts up:	**31**
Hosts down:	**0**

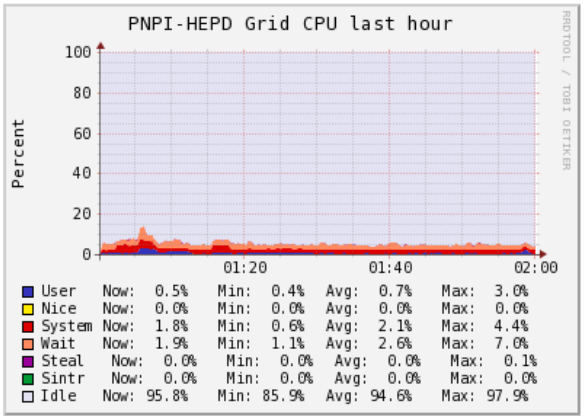
Current Load Avg (15, 5, 1m):
18%, 18%, 17%

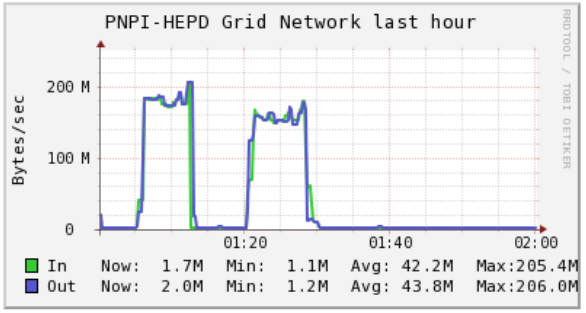
Avg Utilization (last hour):
20%

Localtime:
2017-05-07 02:00









End of lecture