Simulation approach for improving the computing network topology and performance of the China IHEP Data Center

Andrey Nechaevskiy, Gennady Ososkov, Darya Pryahina, Vladimir Trofimov, Weidong Li, Li Wang, Fazhi Qi
symsim@jinr.ru
LIT, Joint Institute of Nuclear Research, Dubna, Russia
Computing Center, Institute of High Energy Physics Chinese Academy of Sciences, Beijing, China

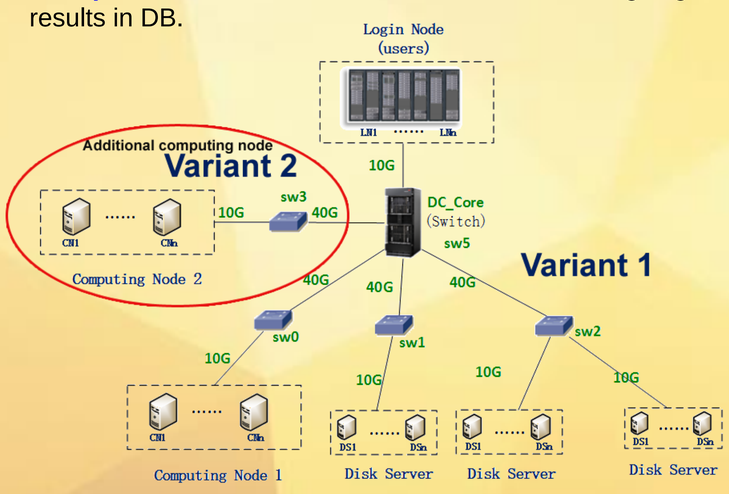
Dynamically developing IHEP experiments are expecting to deal with exabyte data scale and need corresponding means of distributed computing. The development of sophisticate grid-cloud systems intended to store, distribute, and process super-big volumes of experimental data inevitably demands a substantial study of their optimality by detailed simulation of these systems.
Simulation program SyMSim (Synthesis of Monitoring and SIMulation) was developed at LIT JINR and modified for IHEP simulation.

Basic concepts of simulations

- ✓ The goal of basic concepts of simulations of a modern computer center is to satisfy some **optimality criterion** which minimizes the equipment cost under unconditional fulfilment of **SLA** (Service Level Agreement).
- ✓ The best way to evaluate dynamically the system functioning quality is using its **monitoring tools**;
- ✓ The simulation program is to be combined with a real monitoring system of the grid/cloud service through a special **database** (DB);
- ✓ To ensure a developer from writing the simulation program from zero on each development stage it is more feasible to accept a **twofold model structure**, which consists of
  - ✓ **a core – its stable main part** independent on simulated object and
  - ✓ a declarative **module for input of model parameters** defining a concrete distributed computing center, - its setup and parameters obtained from monitoring information, as dataflow, job stream, etc;
- ✓ DB intention is just to realize this declarative module work and provide means for output of simulation results;
- ✓ **Web-portal** is needed to communicate with DB assigning concrete simulation parameters and storing results in DB.

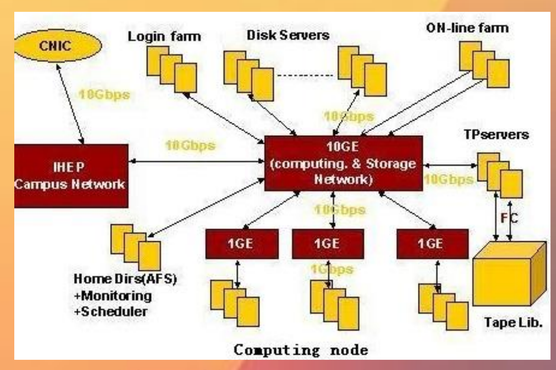
Beijing Electron Positron Collider II (BEPCII) with BES III experimental setup

High Energy Photon Source (HEPS) project



The simulation experience with a simplified IHEP computing scheme

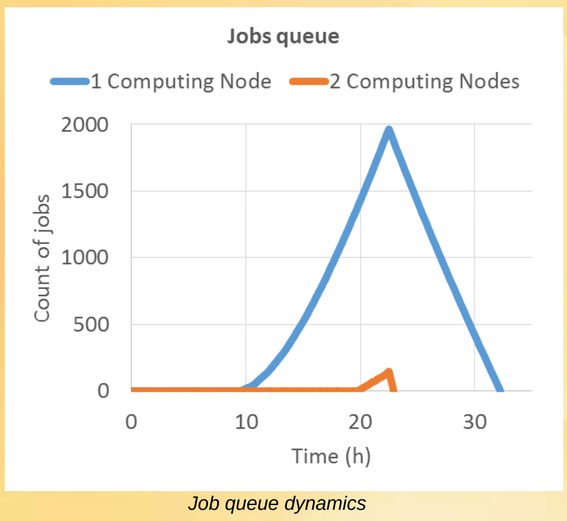
After obtaining simulation results one can compare the usage level of various variants of computing node extensions by analyzing the intensity of the data and job flow and the load of communication equipment. Based on these results one can identify problems, confirmed the quantitative characteristics that arise in the process of data processing.
For the first simulation experience with IHEP computing we choose two simplified schemes.
Let us consider a typical process of job flow in one of Computing Nodes with 500 PC.
A file needed to perform one or more jobs must be available on the remote Disk Server and require downloading to a local pool.
The time, when CPUs are busy by idling jobs, because they cannot start waiting for a file, can be considered as the important characteristic of the computing system loss.



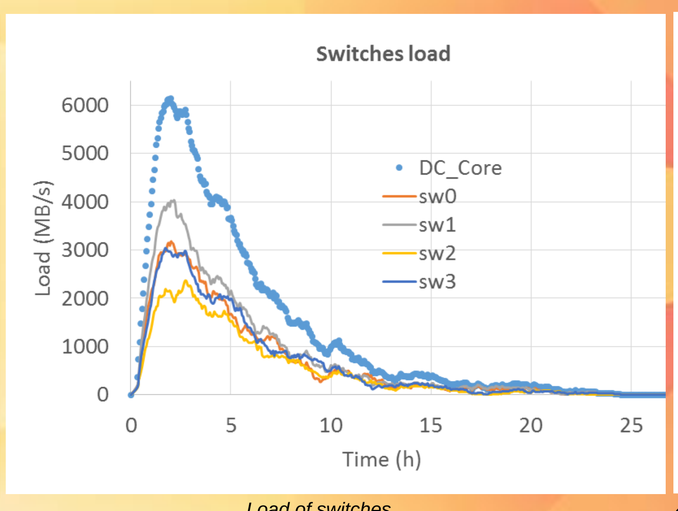
Scheme of the IHEP Computing Center

Examples of computing process characteristics obtained by simulation

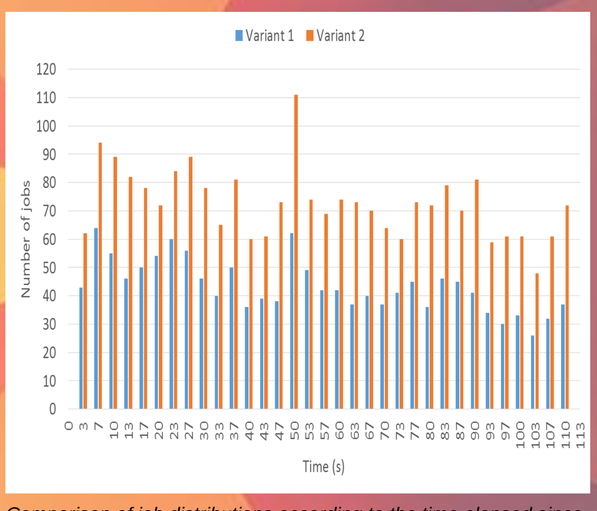
Among events occurring at the system during a simulation run one can compare for considered variants such characteristics of the computing process, as the job queue dynamics, the load of switches, or cases of the system loss since CPUs are busy by idling jobs. We decided to base the comparison of considered variants on the system loss due to CPU occupations by idling jobs waiting for a file.
As our simulations show, the system for variant1 looses 8%, but after adding the second computing node (Variant 2) the loss increases up to 15% of system time.
However, if we choose the different way of increasing the computing power and add not a computing node, but extra 500 cores to existing computing node, then on the node with 1000 cores the losses stays on the level 8%.



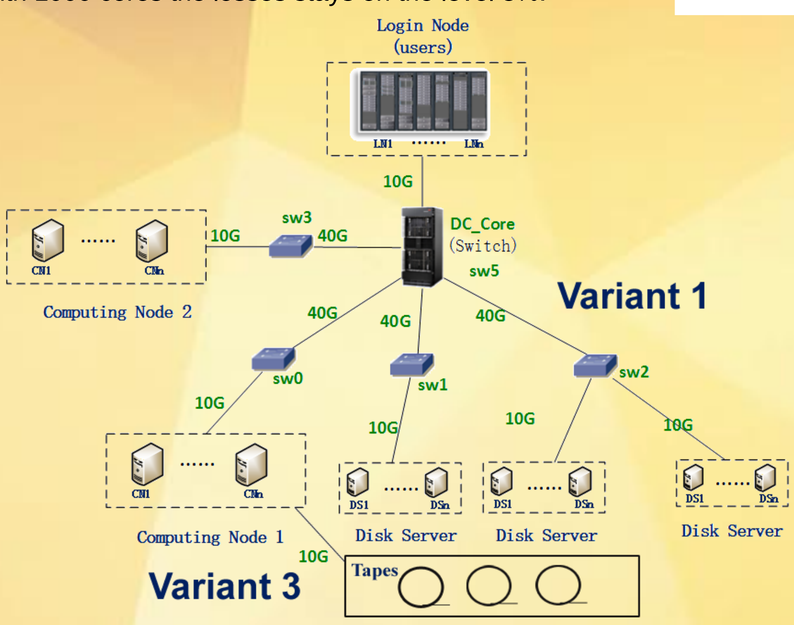
Job queue dynamics



Load of switches



Comparison of job distributions according to the time elapsed since sending the job to the CPU to start calculations for two variants



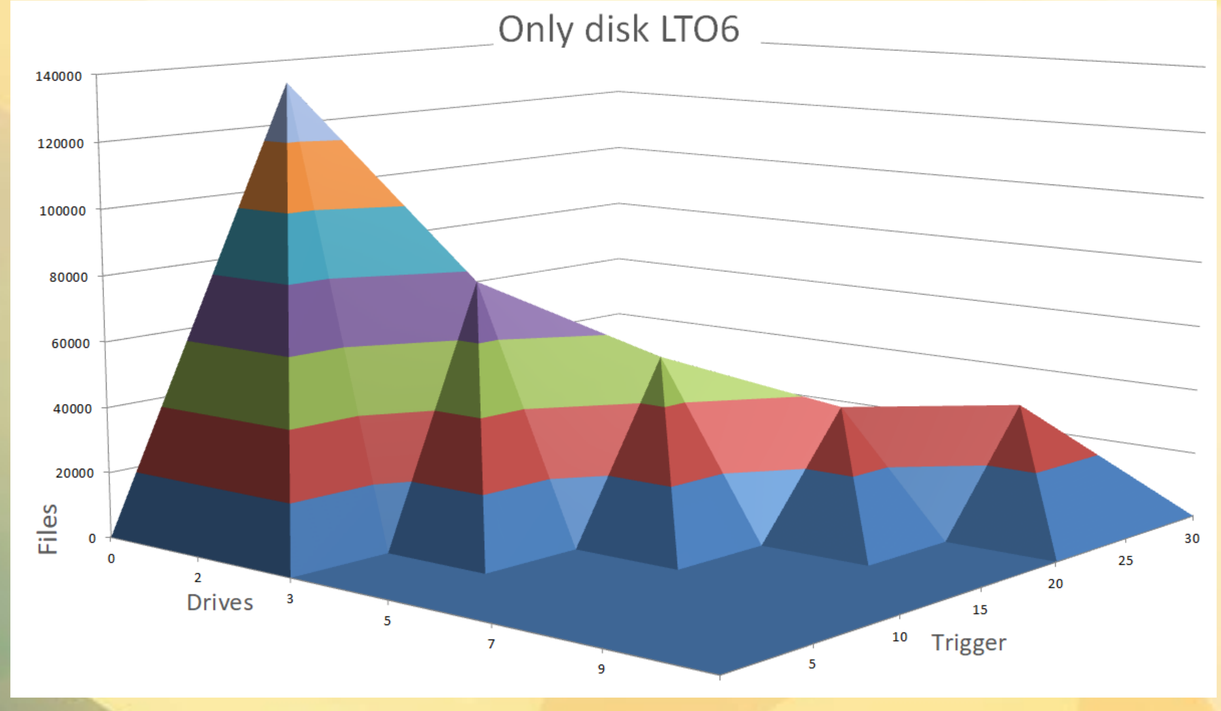
Possible improvements of the job flow process

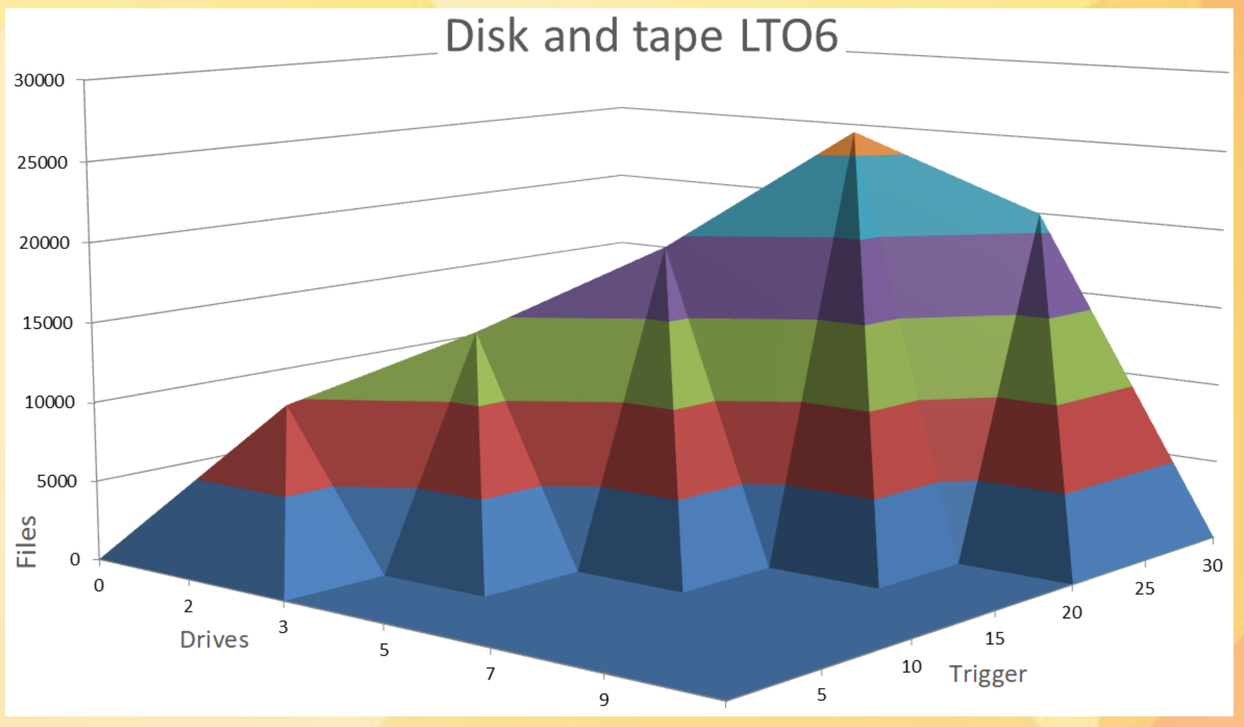
Thus it is shown that the program SyMSim is successfully adopted and allows to obtain a number of important quantitative characteristics of job flow and dataflow processes needed to see how to optimize the system.
In particular, simulations shows that attempts to increase the power of computer system by enlarging the number of computer nodes leads to increasing system losses due to idle processors. However, you can keep losses at the same level, if you would increase the computing power by enlarging the number of cores in one node.
There are, at the same time, technologically different solutions to speed up the workflow process and improve CPU usage:

1. *The use of cloud infrastructures and virtualization.*
2. *Develop a scheduler that will load the job to execution, taking into account the availability of the needed file(s).*
3. *Launch procedures of pre-load files.*

The choice of the solution depends on the architect of the data processing system, which he can take based on the simulation results.

As the first step, the LIT team extended the SyMSim algorithm to include such its important parts as data stream from data acquisition infrastructure to be stored on robotized tape library.





Our first experience with simulating the IHEP computing is very preliminary and intended just to try to adapt an existing simulation program to the IHEP specifics.
The new program version was already installed in the CC IHEP, adapted to the CC parameters and tested.
Success of this experience has demonstrated the applicability of the simulation program, so we are going to extend the IHEP Computing Center model to be simulated gradually approaching to its present and then planned structure.