Contribution ID: **675** Contribution code: **contribution ID 675**   Type: **Poster**

DIRAC at JINR as a general purpose system for massive computations

Joint Institute for Nuclear Research has several large computing facilities: Tier1 and Tier2 grid clusters, Govorun supercomputer, cloud, and LHEP computing cluster. Each of them has different access protocols, authentication and authorization procedures, data access methods. With the help of the DIRAC Interware, we were able to integrate all these resources to provide a uniform access to all these facilities. Right now, it is possible to perform basic workflows on all resources. The main use-cases covered by the DIRAC service in JINR: centralized Monte-Carlo simulation for the MPD experiment, Monte-Carlo for the Baikal-GVD neutrino telescope, as well as running jobs for the Folding@HOME project.

During the pre-production stage, it is important to estimate the characteristics of user jobs. That information is crucial for planning of the execution process on different resources. An approach was elaborated to collect data about RAM, CPU, and network consumption by each user job. This helps to tune the mass production algorithms before the production actually starts.

Since the system processes tens of thousands of similar jobs during one particular production, it appeared to be possible to collect from DIRAC data about these jobs and perform analysis of their execution parameters. We collect data related to the CPU model, wall-time, CPU benchmark DB12, hostname, username, and resource name. An approach was elaborated to extract meta-data about job execution and visualize it. With this visualization, it became possible to compare different computing resources, study the CPU and worker node performance. This approach does not require submitting special jobs, so the resources are not wasted for this analysis.

In this contribution, a detailed explanation of the used techniques and methods will be presented. Results of the performed analysis of the computing facilities at JINR will be discussed.

Significance

We propose a new approach in estimating the performance of heterogeneous resources on real workload. That allows performing "passive" estimations with real user workload and not only artificial benchmarks. That is probably the most interesting and unique result of this talk. But, it is just one side. From another side, we propose approaches to study not only resources but workload itself. It is the second result. All this became possible with the use of DIRAC as a service not only by one particular experiment but several of them.

References

https://indico.jinr.ru/event/1469/contributions/9985/
https://indico.jinr.ru/event/1086/contributions/13121/
https://indico.jinr.ru/event/1119/contributions/10558/

Speaker time zone

Compatible with Europe

Authors: TSAREGORODTSEV, Andrei (Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France);   PELE-VANYUK, Igor (Joint Institute for Nuclear Research (RU))

Presenter: PELEVANYUK, Igor (Joint Institute for Nuclear Research (RU))

Session Classification: Posters: Orange

Track Classification: Track 1: Computing Technology for Physics Research