Contribution ID: **108**   Type: **Presentation**

Bioinformatics simulations by means of running virtual machines as a grid jobs in the MolDynGrid Virtual Laboratory

Wednesday, 11 November 2015 11:40 (20 minutes)

There are number of software for bioinformatics simulations used in MolDynGrid virtual laboratory (VL) [1] for in silico calculations of molecular dynamics, including GROMACS, NAMD, Autodock, etc [2-3]. Computational resources for such simulations are provided by a number of HPC clusters that are mostly the part of Ukrainian National Grid (UNG) infrastructure powered by Nordugrid ARC [4] and few clusters of European Grid Infrastructure.

In the heterogeneous grid environment ensuring that every resource provider has required build of software, particular version of software with its dependencies, and moreover handling software updates is a non-trivial task. When number of software and build flavors grows, like in MolDynGrid case, the software management across dozens of clusters becomes almost impossible.

The classical approaches to software maintenance includes building software on the fly within grid-job execution cycle and relying on VO-managed common filesystem like CVMFS with pre-built software. Both approaches works well in case of similar resource providers environment, but in case of completely heterogeneous hardware and software, including different OS distributions you should handle software builds for every of this platform.

To efficiently handle software in such environments for MolDynGrid researches another approach has been introduced - running hardware accelerated virtual machines (VM) as a grid jobs. This approach eliminates the necessity to build software on every resource provider and introduce a single point for software updates. Software should be build for one virtual platform only. Moreover this approach also allows to use software for Windows.

Thus adding virtualization layer will drop performance, the first thing that had been analyzed is the amount of such drop. On the UA-IMBG cluster, molecular dynamics in GROMACS for the same biological object had been computed on the same hardware with and without virtualization. The software environment was cloned from the host to the guest VM. GROMACS was chosen as a main software used by MolDynGrid VL in terms of CPU time consumption.

To run VMs as grid jobs on grid-site worker nodes, there are several helpers running with root priveleges needed to setup virtual hardware and transfer job data to VM. The framework of components that support VM execution cycle as a grid job has been originally developed as a part of Ukrainian Medgrid VO project [5] and called Rainbow (ARC in the Cloud) [6]. Rainbow start with providing interactive access to Windows VMs running on UNG resources for analyses of medical data stored in grid for telemedicine [7]. For MolDynGrid VL several components had been added to Rainbow framework that implements data staging to VM and allows to add VM layer to grid-job processing cycle.

Both CLI and Web MolDynGrid VRE interfaces had been extended to support VM submission with Rainbow. This approach allows to involve more resources for computations with particular software builds.

Further ongoing developments of Rainbow for MolDynGrid includes support of Docker containers in addition to KVM VMs and GPGPU computations by means of GPU device pass-through.

Links, references, publications, etc.

[1] MolDynGrid Virtual Laboratory web site: http://moldyngrid.org/

[2] Savytskyi OV et al. J Mol Recognit. 2013;26(2):113-20.

[3] Mykuliak VV et al. European Biophysics Journal. 2014;43(12):613-622.

[4] Salnikov A.O. et al. MolDynGrid virtual laboratory as a part of Ukrainian Academic Grid infrastructure. 5-th International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Rende (Cosenza), Italy September 21-23, 2009. pp. 237-240

[5] Ukrainian MedGrid VO web site: http://medgrid.immsp.kiev.ua/

[6] Andrii Salnikov. Rainbow (ⵥARC in the Cloudⵥ) at a Glance, https://goo.gl/ifC6cl

[7] Salnikov A., Vishnevskey V., Boretskyi O. Platform as a service for in1077;ercative analysis of medical data. Mathematical Machines and Systems. 2015 #1 pp. 53-64

Additional information

Rainbow usage for interactive access to VM running on grid resources YouTube demo: https://youtu.be/-OgeQkUI2LQ

Primary author: SALNIKOV, Andrii (Taras Shevchenko National University of Kyiv)

Presenter: SALNIKOV, Andrii (Taras Shevchenko National University of Kyiv)

Session Classification: Federated accelerated computing