Job workflow in the Virtual Laboratory

M. Lawenda[1*], N. Meyer[1], T. Rajtar[1], M. Okoń[1], D. Stokłosa[1], M. Stroiński[1]

[1]Poznań Supercomputing and Networking Center (PSNC)
Z. Noskowskiego 10, 61-704 Poznań, Poland
e-mail: {lawenda, meyer, ritter, hawky, osa, stroins }@man.poznan.pl
*Corresponding author: lawenda@man.poznan.pl

Abstract

In the paper workflow in virtual laboratory (VL) system is discussed. The authors consider also possibilities of the flexible VL system creation to control many laboratory apparatus in a remote way. Depending on laboratory type and measurement scenarios workflow path changing. Virtual laboratory architecture as well as many other system modules have to make possible adapting to various laboratory apparatus and allow realizing diverse workflow scenarios.

1 INTRODUCTION

Virtual laboratories (VL) are now in the scope of activities of research groups and companies, which are working on new solutions for the Grid computational environment [1,2]. The main reason for creating such laboratories is the remote access to the scientific equipment, the facility and acceleration of the education process, exchanging opinions in research groups and the management of the experiment results. This idea is especially attractive for the experimental sciences and technologies, particularly physics, chemistry, structural biology, experimental medicine, radio astronomy and also for widely taken engineering [4,5].

Virtual Laboratory is a heterogeneous, distributed environment, which allows scientists all over the world to work on a common group of projects. This environment should allow conducting experiments with the usage of physical devices, doing simulation using the computational application, and communication between users working on the same topic. Similarly to typical laboratory tools and research techniques depending on the specific field of science, virtual laboratories can benefit from some collaboration techniques as tele-immersion, but they are not mandatory.

In order to meet the virtual laboratory goals, there should be specified some requirements. The most important requirements follow:

- the possibility of experiments performing (real and computational),
- the possibility of comparing the achieved results in real experiment and simulation, which can help adjust the simulation parameters properly,
- reservation of time for experiments execution; especially for experiments that require supervision or can be observed by more people on-line,
- communication with the physical laboratory staff, which can be necessary to help run some specific experiments, e.g. setting up and operating devices to experiments,
- sharing experiments results – the result should be stored in a database and available for other scientists (with the possibility to restrict access rights),

- load balancing – e.g. experiments should be run on the less loaded sites,
- library of electronic papers, reports and books grouped by a specific topic,
- communication with other people working on similar problems – using chat, audio and videoconferences etc.

The Virtual Laboratory system is developed in Poznań Supercomputing and Networking System [11] in collaboration with the Institute of Bioorganic Chemistry [12] and Radioastronomy Department of Nicolaus Copernicus University [13].

Putting into practice two different laboratories will prove the correctness of our conception. One of them is the Virtual Laboratory of Nuclear Magnetic Resonance Spectrometer, the other the Virtual Laboratory of Radioastronomy.

The present state of the art in virtual laboratories development discipline was discussed in [10].

Here, by the workflow we understand flow of the job (different types of job) via whole virtual laboratory system. For example for the Virtual Laboratory of NMR Spectroscopy system workflow begins from preparing the sample and ends on visualization stage.

2 ARCHITECTURE

Workflow in the Virtual Laboratory depends on many aspects, among other things: given laboratory type (e.g. NMR, radioteleskope) and measurement scenario. Virtual laboratory architecture as well as other modules have to make job flow possible for many different situations.

It is a very difficult assignment to define a laboratory framework, which will be appropriate to many types of devices. Finding the solution, which will meet the needs of most research groups and allow performing experiments using different apparatus will consume a lot of time, money and human energy. Nevertheless, achieving this goal seems to be worth a big price and bring a lot of advantages (e.g. it will allow reducing resources which are needed to create new laboratory) in the future.

To meet the conditions mentioned in the introduction, the whole system should have a modular architecture, which allows to increase the system possibilities (e.g. by adding a new modules), easily changing the existing modules into others (e.g. with bigger possibilities) without the recompilation necessity. It can be possible to build a new virtual laboratory only by defining necessary modules and adapting it to the surrounding environment. Creators do not have to care about architecture and many other important aspects of the system. They can only focus their attention on aspects directly connected with the implemented device and organizing work for the users (e.g. defining rights lists and edge conditions, creating laboratory groups, assuring collaboration between apparatus and the existing virtual laboratory modules).

Due to specialization and different nature of some laboratory it is necessary to distinguish two types of modules: general (can be implement in most types of instances of the virtual laboratory) and specific (can be used only with a given type of device).

The conception of the general framework and most important for the virtual laboratory will be discussed next. In the virtual laboratory architecture we can distinguish four layers (Fig.1). The first of them is the Access layer. Originally it consists of tools, which enable access to the laboratory resources and for presentation stored data. The most important modules are:
- Dynamic measurement scenarios – assure control and execution of the operations (experiments) chain, which is submitted by the user,

- Users and laboratories management – tools for creating new laboratories profiles, new users accounts, management of the access rights to devices, digital library and communication media,
- Data presentation tool – presentation experiment results collected in the data management system, on-line visualization tool and accounting data looking through,
- Users communication tools – allow users to communicate with one another

The second is the Grid layer where services, which we can call general, were accumulated. They are required in most laboratories instances. There are also connections with the outside grid services e.g. with Globus services. The most important are:

- Authorization centre – granting (or not) access to many virtual laboratory resources, certificates management,
- Global scheduling – responsible for choosing the appropriate laboratory device and load balancing when possible,
- Digital library – storing experiments results and electronic publications,
- Data transport – responsible for download and upload data from/to destination machine,
- Grid Gateway – communication with the grid broker (e.g. GRMS [15]), sending computational tasks to the grid and receiving experiment tasks from the grid systems

On the supervision layer specific services are gathered. These services have to be implemented taking into consideration a given device and their specification. Usually it consists of:

- Local scheduling – tasks scheduling on a given device taking into consideration parameters and priority
- Resource monitoring – resources using control, tasks actual state control
- Users accounting – information about the used resources

Finally we can define the Resources layer, which consists of devices to experiment execution and also the necessary software:

- Laboratory devices – laboratory apparatus and software for experiment execution,
- Computational servers – also software for pre- and post-processing computation,
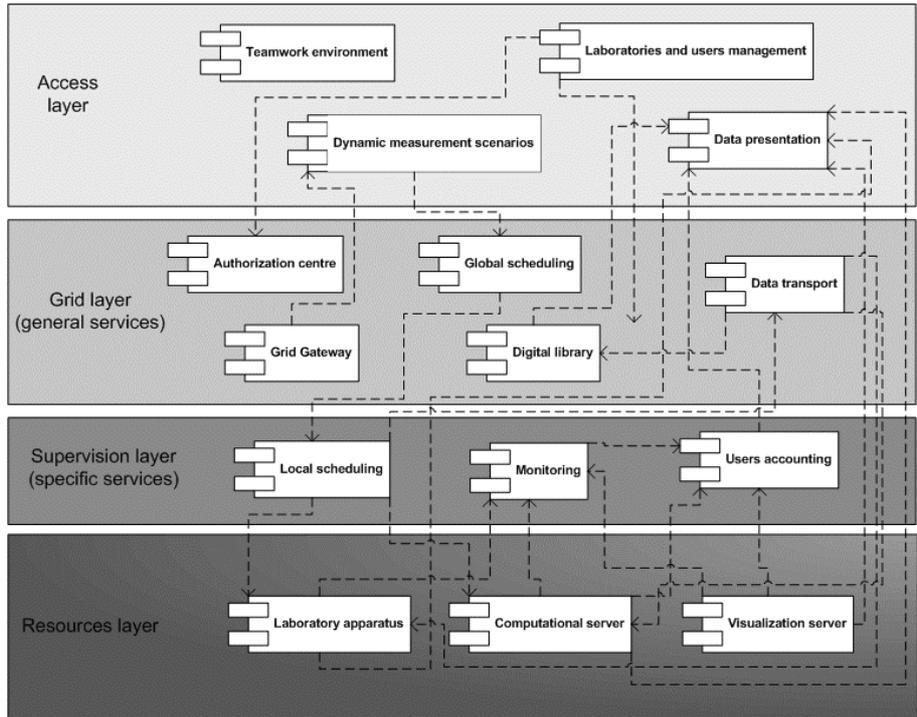- Visualization servers – also visualization software,

Fig. 1 The framework of virtual laboratory architecture

From the connection point of view, laboratory framework can be divided into three levels: client interface, application server and device server. It has a client-agent-server structure. Such layer architecture provides completing tasks, which are assigned to the system. The tasks every level is responsible for are described below.

On the client interface level users have a possibility to submit a task, using the available interface from the WWW page. Some parts of the interface content should be specific for a particular device, in which the distinctive parameters for a given laboratory type should be taken into account. User interface uses a standardized protocol to communicate with the application server.

Most services should be gathered on the application server. On the one hand, it allows to fit the needs to the specific laboratory, and, on the other hand, decrease the load of other servers. Moreover, it allows fast adapting of the laboratory framework to the new device (only few modules have to be installed outside the application server). There, the advanced scheduling process can be executed as well as monitoring, accounting and authorization modules can be running.

The device server is responsible for receiving a task from the application server, job submitting, receiving results of the executed tasks and transporting them to the application server. The module which controls the served device must know its specific. It uses the API device functions to tasks submitting and controlling a device.

3   DEFINING UNIVERSAL MEASUREMENT SCENARIOS

As we mentioned before workflow in the virtual laboratory is strictly connected with measurement scenario. When user characterize measurement scenario, he in fact in this way define the workflow path. Below we want to consider how to

generalize flow of jobs in the virtual laboratory system by defining universal dynamic molecular scenario module.

The experiment process execution in many types of laboratories consists of very similar stages. The given experimental process is often recurrent and must be executed many times by some parameters modification. In this situation the measurement scenarios conception seems to be very useful.

The conception of the dynamic measurement scenarios allows defining the process of an experiment in any way, from pre-processing, through executing the experiment, to the post-processing and visualization tasks. Users are also allowed to add their own module as a part of the scenario. Defining the measurement scenario allows to spare a lot of time during computation. The user does not have to wait for the end of a given process stage to submit another one. It is made automatically.

Initially, we divide the experiment execution process into four steps:
- preprocessing – here the user can prepare data to use as an input source
- experiment – two types of experiments can be distinguished; the first is the real experiment executed on a device available in the Grid; the second one is a computational experiment running on the execution host.
- postprocessing – here we can process the output data from the experiment stage
- visualization – can be treated as a part of the postprocessing stage; we decided to treat it as a separate task due to specific hardware and software needs

This simple task's execution chain we call a static measurement scenario (SMS). It allows connecting particular stages only between themselves. Here, the execution path is specific and the user cannot manipulate it.

To increase the possibility of the jobs scenario the dynamic measurement scenario (DMS) was defined. In the DMS model, besides the definition of the tasks execution sequence, we can also define some extra connections (e.g. loops, parallel connections), conditions on the connections and different lengths of the execution paths. In fact, DMS allows defining many SMS models in one scenario. Thanks to the possibility of the conditions defining on the connections paths the real path is determined during execution and can depend on computational results.

To properly define the dynamic measurement scenario the system has to have knowledge about connections which are enabled. In case of the laboratory type where the processing software is well known, it seems to be easy. The case will be more complex when we want to define the DMS model for a wider range of virtual laboratories.

To solve this problem we defined a special language which determines the following conditions:
- names of the connected applications (source and destination application)
- a condition (or conditions) connected with a given path, which has to be met to pass this path; in case when few conditions are met, the first one is chosen
- an additional condition which has to be met to pass a given path, e.g. when special conversion between application is needed
- a list of output files for source application and a list of input files for a destination application

An expertise from a particular discipline is necessary to define rules for DMS. A laboratory administrator together with a domain expert can do it.

On the figure below (Fig.2) we present an exemplary measurement scenario diagram for the Virtual Laboratory of NMR Spectroscopy:
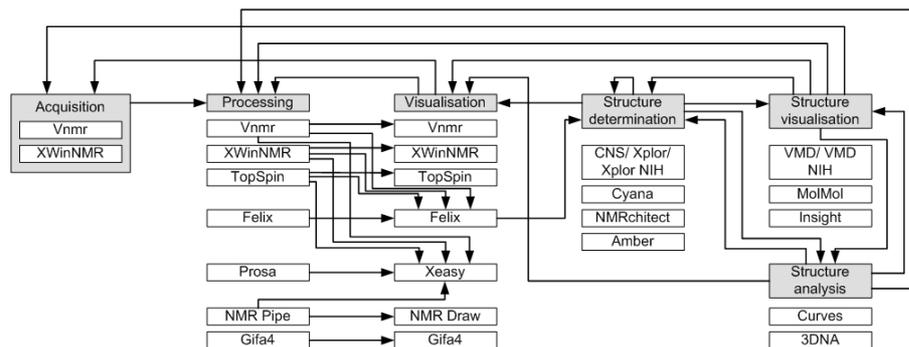
Fig. 2 Exemplary measurement scenario diagram for the Virtual Laboratory of NMR Spectroscopy

4 WORKFLOW DISTURBANCE

Apart from above mentioned aspects there are some conditions which influence on the way of experiment execution. In this section we want to focus on the Virtual Laboratory of NMR Spectroscopy aspects: human factor and preparing the sample.

Many laboratories apparatus need a human service during the experiment execution. A very good example of such situation is an NMR experiment, where the operator has to put a sample into the spectrometer before the acquisition process.

There are a few factors which significantly influence the virtual laboratory idea when sample using is necessary to test. The most important are the following issues: the tested sample and the human factor. Especially the idea of remote access to the laboratory devices and experiment execution "any time any where" is limited by these aspects. They cause the necessity of appropriate experiment planning (e.g. sending a sample to the laboratory several days before tests) and also taking them into account on the scheduling stage.

A frequently tested sample has to be prepared by the user (researcher) who will be executing the experiment. Next it is sent to a given laboratory, the scheduling process has to be discontinued until the sample arrives in place. The user choosing a specific place where he/she wants to execute the experiment makes the devices' load balancing impossible. This problem can be solved in a situation when we have a specialized centre with many laboratory apparatus in one place. Some samples are unstable and can yield to deterioration if they are inappropriately stored during transportation or while waiting for the experiment.

The influence of the human factor on the experiment execution stages is also an important issue. By the human factor concept we understand: device operator accessibility, his/her skills connected with preparing the work environment and many other human disabilities which influence the experiment time. In case of some laboratory apparatus it is necessary to set up particular parameters manually. Moreover, these actions can be done only when the sample is inside the device. These aspects make the batch job unavailable and requires the presence of the device operator.

Due to the specificity of the execution experimental sequence it is very hard to characterize one general execution model. Thus we can state that the above-mentioned aspects make the generalization concept of the virtual laboratories difficult to achieve. In such a situation every virtual laboratory needs a special approach for some number of factors.

5 VIRTUAL LABORATORY WORKFLOW

In this section very simplified workflow in the Virtual Laboratory is described. We want to base our description on NMR static scenario. We should mention that NMR spectrometer is device, which needs human interaction to control, because of its nature. It cause that virtual laboratory workflow depends on human factor.

Firstly user has to define measurements scenario. Usually NMR scenario consists of three stages: experiment, postprocessing and visualization. Next description scenario is sent to the DMS module where is decomposed and every job is sent to modules in proper sequence. Experiment tasks are sent to the Global Scheduling module, computational tasks are sent to the GRMS module via Grid Gateway.

On the Global Scheduling stage appropriate laboratory device is chosen for the job. Additionally user rights and accessible devices are checked.

In next step job goes to Local Scheduling module where is scheduled according to given algorithm and priority. Job is sent to the Job scheduling module only when previous job is finished.

Every NMR experiment can be divided into two parts. First of them needs some operator's activity which includes: preparing the sample, characterizing probe type and mounting it into magnesium, next inserting sample inward and manual probe tuning to appropriate frequency. Second part of this type of experiment is automated and can be executed remotely.

On this stage we can use specialized software to control the experiment: Vnmr for Varian [8] spectrometer or XWIN-NMR for Bruker [9] spectrometer. This software is making available to the user using Virtual Network Computing (VNC) [16]. Commands like: choosing working directory and sequence impulses, modifying sequence parameters can be done remotely, without direct access to the spectrometer. When experiment is finished all results are sent to digital library.

Postprocessing task is sent to Grid Gateway where is transformed into Globus [17] format and sent to GRMS module. GRMS make a decision basing on job description where to run computation. If job needs user interaction VNC software is used again. Before computation spectrometer data are transferred to the server and when postprocessing is finished output data are transferred back.

Execution of visualization task is very similar to postprocessing stage. Only on the scheduling phase the job will be probably directed to other device, better adapt to visualisation needs. Also interactive work will be executed more often.

After every execution stage accounting data are transferred to the Accounting module. Above described workflow in Virtual Laboratory is described below (Fig. 3).
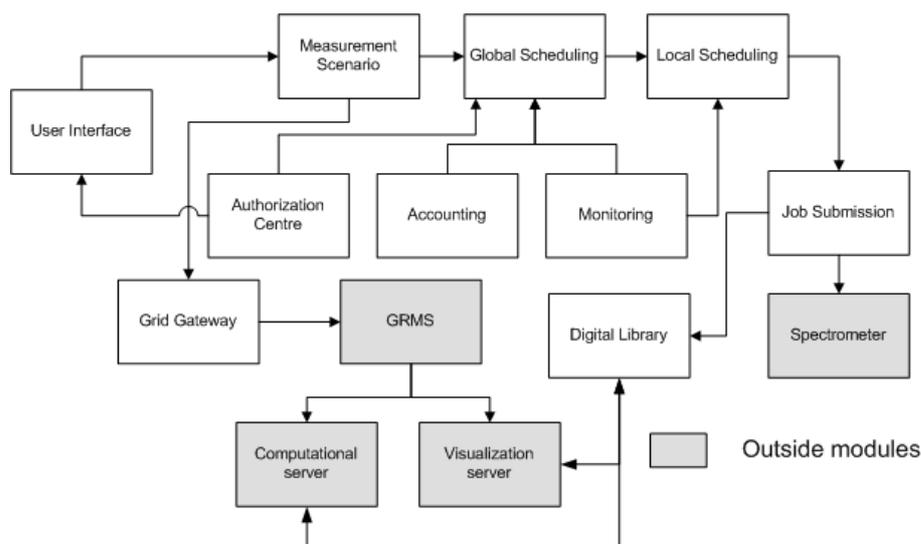
Fig. 3 Simplified workflow in Virtual Laboratory

6 CONCLUSIONS

Virtual Laboratories are undoubtedly the technology of the future. They are perceived as a remedy for problems connected with access or buying costs of expensive and rare devices. Current technologies connected with VL are in their first stadium of development. New created services allow running simple experiments, which can be managed by changes in a limited parameter pool.

Although the virtual laboratory idea is expensive to introduce, it is still the solution of the future for research environments with limited equipment resources. It is the only way to widely open the science for society demands in education and professional specialization in new technologies. The realization of the virtual laboratory environment will allow the scientists and engineers to work on their projects by remote events simulation, interpretation of experimental data and, in some cases, to run real experiments in a customized laboratory. A common platform for this project is an optimal utilization of modern, optical Internet based on the Informational Technology.

The base services for virtual laboratories are currently under development. The hardware (e.g. SGI Reality Center [6]) and software (CAVERN G2 [7]) solutions, which allow remote communication and results presentation, are being developed. However, most of the currently developed laboratories are rather low advanced constructions, created for a small group of specialized tasks.

Workflow aspect is very valuable for the universal virtual laboratory framework, which allows sparing a lot of human energy and money during the development stage. It is possible to build many different types of laboratory systems, which provide and adapt functionality to specified needs. Designers of a new laboratory have to focus their attention only on a few aspects connected with adaptation to a new data set, and a specific scheduling algorithm. VL architecture has to allow define many different workflow paths. It has to also be opened for cooperation with Grid services e.g. GRMS module.

As an exemplary implementation it is planned to make the NMR Spectrometer (Bruker Avance 600) and radiotelescope (diameter 32m) available to the Grid users [14].

REFERENCES

1. M. Lawenda, „Laboratorium Wirtualne i Teleimersja", Poznań Supercomputing and Networking Center, RW nr 34/01
2. M. Bubak, J. Madajczyk, N. Meyer, „Laboratorium wirtualne z wykorzystaniem infrastruktury HPC/HPV"
3. Afsarmanesh, H., Benabdelkader, A., Kaletas, E.C., Garita, C., and Hertzberger, L.O.: "Towards a Multi-layer Architecture for Scientific Virtual Laboratories." In: Bubak, M., Afsarmanesh, H., Williams, R., Hertberger, B., (Eds.), Proc. Int. Conf. High Performance Computing and Networking, Amsterdam, May 8-10, 2000, Lecture Notes in Computer Science 1823, 162-176, Springer, 2000
4. Khettry D, Xian-He Sun. A Windows-NT virtual collaboratory for technical computing. Fifth NASA National Symposium on Large-Scale Analysis, Design and Intelligent Synthesis Environments. Williamsburg, VA, USA. 12-15 Oct. 1999. Elsevier. Advances in Engineering Software, vol.31, no.8-9, Aug.-Sept. 2000
5. Nicholas P, Fushman D, Ruchinsky V, Cowburn D. The Virtual NMR Spectrometer: a computer program for efficient simulation of NMR experiments involving pulsed field gradients.
6. Reality Center Overview - http://www.sgi.com/realitycenter/overview.html
7. Open Channel Foundation – CAVERN G2 - http://www.openchannelsoftware.org/projects/CAVERNsoft_G2/
8. Virtual Reality based Simulation Laboratory – http://www.hlrs.de/organization/vis/projects/vrslab.html
9. M.Lawenda, N.Meyer, T.Rajtar "General framework for Virtual Laboratory" THE 2ND CRACOW GRID WORKSHOP, Cracow, Poland, December 11 - 14, 2002
10. R. W. Adamiak, Z. Gdaniec, M. Lawenda, N. Meyer, Ł. Popenda, M. Stroiński, K. Zieliński, „Laboratorium Wirtualne w Środowisku Gridowym", Polish Optical Internet - PIONIER 2003, 9-11 April 2003
11. Poznan Supercomputing and Networking System WWW pages http://www.psnc.pl/
12. Institute of Bioorganic Chemistry Polish Academy of Sciences WWW pages: http://www.ibch.poznan.pl/
13. Radioastronomy Department of Nicolaus Copernicus University www pages: http://www.astro.uni.torun.pl/
14. Virtual Laboratory, WWW project pages http://vlab.psnc.pl/
15. GRMS http://www.gridlab.org/
16. VNC http://www.realvnc.com/