July 5th - 6th 2006, Dwingeloo

PSNC and JIVE Meeting

EXPReS Project
JRA1: FABRIC

Meeting Report

Table of Content

Document History 4
1 Meeting information 5
1.1 Location and duration 5
1.2 Participants 5
First day (5th of July 2006) 6
2 Introduction to the correlation process (Huib) 6
2.1 VLBI 6
2.2 Distributed correlation 7
2.3 Explanation of the broadband correlator (Ruud) 7
2.4 How to control and get access to the data stream from the telescope (Arpad) 8
3 Live Virtual Laboratory (VLab) demo (Dominik, Marcin) 11
3.1 General information about Virtual Laboratory System 11
3.2 Sample experiment using NMR Spectrometer device 13
3.3 Digital Science Library 15
Second day (6th of July 2006) 18
4 Presentation of different file formats (Huib) 18
5 Discussion on system design and definition of aims 19
5.1 General aims 19
5.2 Short term aims 20
6 Live demo of the SFXC application (Ruud) 22
7 Agreed todo list 22

Table of Figures

Figure 1. VLBI correlation process 6
Figure 2. The current software correlator architecture 8
Figure 3. Possible ways for outcoming data stream from Mk5 system 9
Figure 4. The overview of the solution with file servers 10
Figure 5. Virtual Laboratory architecture 11
Figure 6. Interaction between the VLab modules 12
Figure 7. Task submission in GRID environment 12
Figure 8. Remote control of the NMR spectrometer – Varian 300 MHz 13
Figure 9. Sample workflow 13
Figure 10. The main pane of the SSA application 14
Figure 11. The properties dialog of the SSA application 15
Figure 12. VLab Digital Library 16
Figure 13. Digital Library – sample visualization 17
Figure 14. System architecture for distributed broad band correlation 19
Figure 15. System architecture for single channel correlation 21

Document History

File name	Date	Remarks
psnc_jive_meeting_notes_05_06_july_2006_v_0.1.doc	10.07.2006	Draft
psnc_jive_meeting_notes_05_06_july_2006_v_0.3.doc	13.07.2006	Draft
psnc_jive_meeting_notes_05_06_july_2006_v_0.4.doc	18.07.2006	Draft
psnc_jive_meeting_notes_05_06_july_2006_v_0.5.doc	25.07.2006	Draft
psnc_jive_meeting_notes_05_06_july_2006_v_0.6.doc	02.08.2006	Final

1 Meeting information

1.1 Location and duration

The meeting was hosted by JIVE in Dwingeloo, The Netherlands 5th – 6th of July 2006.

1.2 Participants

The following table summarizes the list of participants.

JIVE		
Name	**Position**	**Email**
Huib Jan van Langevelde	Head of software development	langevelde@jive.nl
Ruud Oerlemans	Software developer	oerlemans@jive.nl
Sergei Pogrebenko	Development engineer	pogrebenko@jive.nl
Arpad Szomoru	Head of Data Processor Research and Development	szomoru@jive.nl

PSNC		
Marcin Okoń	System analyst and developer	hawky@man.poznan.pl
Dominik Stokłosa	System analyst and developer	osa@man.poznan.pl

First day (5th of July 2006)

2 Introduction to the correlation process (Huib)

2.1 VLBI

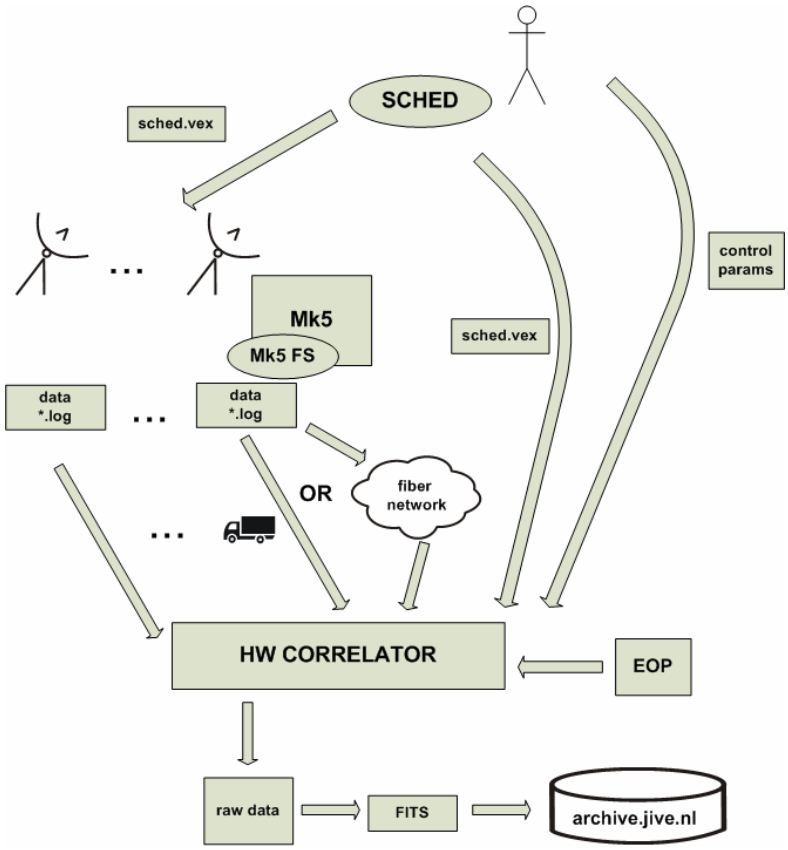

Figure 1. VLBI correlation process

The above diagram describes the VLBI principle as for the current moment. It depicts the following procedure:

- The radio astronomer (user) is given a research grant for performing the observation. The exact time slot is scheduled and the information is returned to the user.
- The user uses SCHED program to produce a single VEX file, which contains the complete description of the VLBI experiment. The file is of the ASCII text format. SCHED uses a text file as an input, and has both command line and graphical interface.
- The VEX file is transferred to the radio telescopes, via FTP system. The telescope operators are responsible for loading the file and setting up the observation.
- During the observation the Field System controls the telescope using settings obtained from the VEX file. Data is recorded in the Mk4 data format on Mk5A disks. These disks are inside a Mark5 computer and can only be accessed through special software. More information on these subjects can be found at the Haystack website: http://www.haystack.mit.edu/tech/vlbi/mark5/index.html.
- After the observation the collected data is sent to the central JIVE processor at Dwingeloo,

via fibre channel connection or still in most cases by physical transport of disks or tapes.

- In order to perform the correlation, the JIVE hardware correlator needs the following:
  - Data from all the telescopes
  - User VEX file
  - EOP (earth orientation parameters)
  - Additional control parameters, supplied by the user
- After the correlation, the raw output data is verified by the JIVE staff, and if no errors are detected the data is converted into the FITS format. This file format can be read by standard astronomical data processing software.
- Finally the converted correlation results are put in the JIVE archive.

2.2 Distributed correlation

- white board discussion on the role of workflow manager
- diagram of the distributed correlation process – starting from the preparation of the observation, through the correlation process and finally data archivization

The discussion came to first conclusions and agreements on the initial system design, and the role of workflow manager in the process, but afterwards it was postponed until next day. After that the created drawings became obsolete, therefore they are not put into this section. (For the final solution see section. 2.4 and 5).

2.3 Explanation of the broadband correlator (Ruud)

Ruud explains briefly the current functionality and operation of broadband software correlator .

In January 2005 the descent of the Huygens probe in the Titan atmosphere was successfully tracked using a specially developed software correlator. These correlator was transformed into a much simpler broadband correlator mimicking the EVN MarkIV hardware correlator operated bij JIVE.

This software correlator, further called SFXC, has grown organically and is depicted in the next diagram. The data is processing according to the pipeline principle using files as intermediate storage buffer between applications. All the processing steps will be described briefly:

1. Mark5A: transfer data (Mk4 formatted) from Mk5A disks to linux type disk using the Haystack proprietary software
2. cxmk4: extract channel pairs from the Mk4 formatted data files. The extraction is necessary because the data in the Mk4 file it not contiguous.
3. cx2fl: data is still packed, 2 channels per byte, in order to reduce the data volume. However floating point numbers are needed by the next step. Therefore cx2fl converts a single cx file containing data from two channels into two separate file each containing single channel data in float format. The data volume by now is increased by a factor of 8. cx2fl optionally can filter down the original bandwidth.
4. delpha: during a VLBI session the participating telescopes receive the signals at different times. Because the VLBI technique requires the correlation of the same wave fronts, delay corrections have to be applied to the original signal. delpha performs the delay corrections on a single fl file based on a pre-calculated delay file.
5. When the data from all participating telescopes has gone through the previous processing steps, data is correlated into a correlator product.

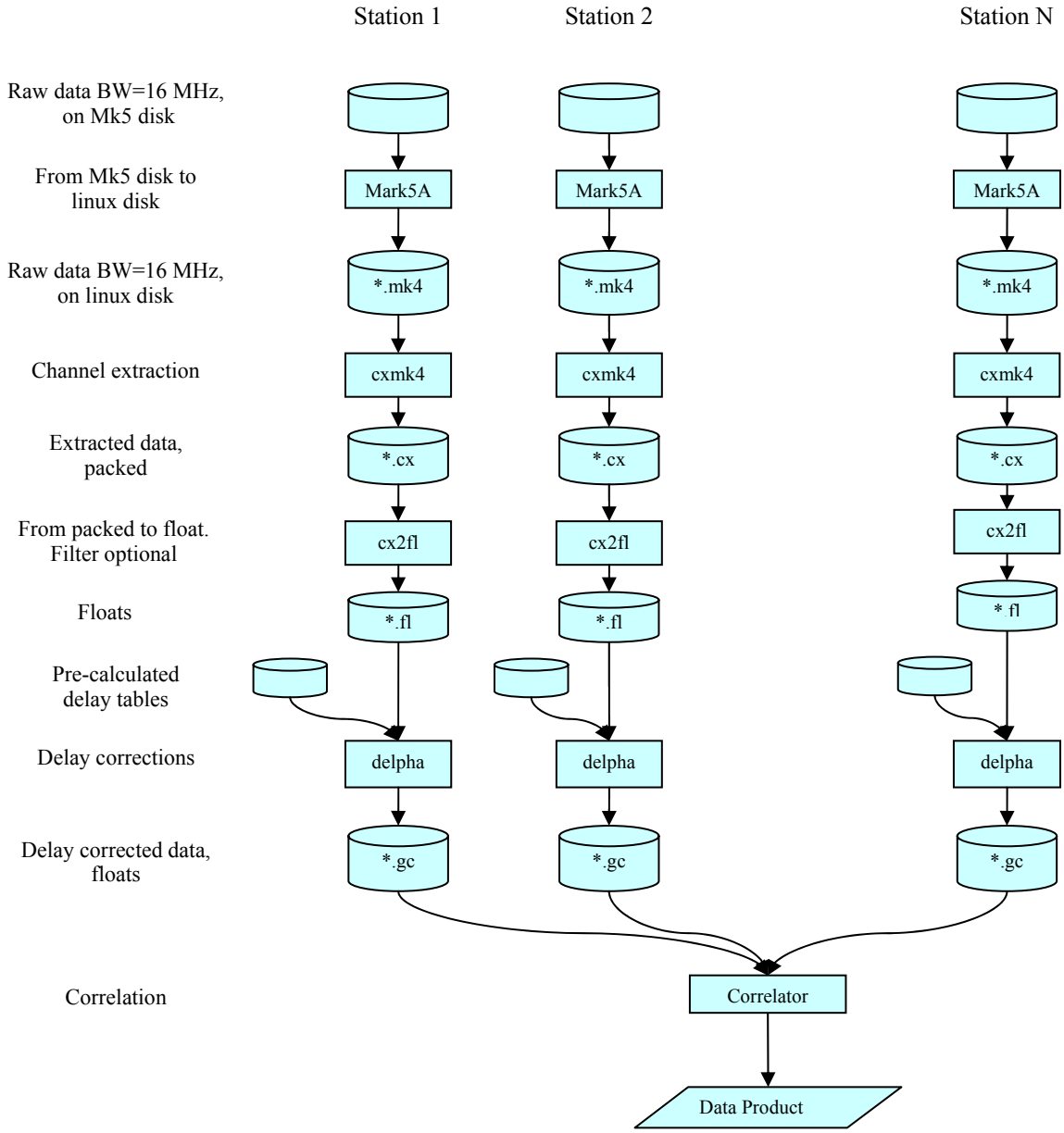

Figure 2. The current software correlator architecture

Each application is controlled by a control file with settings, file names, processing options, etc. This file is of the keyword-value type. All keywords are unique so the control files for the different applications can be merged into one.

2.4 How to control and get access to the data stream from the telescope (Arpad)

Arpad Szomoru was asked how to get a real-time access to the Mk5 data collected during observation and how to effectively implement routing the data stream for the distributed correlation. At this stage the problem was identified being unable to convert the data from internal Mk4 file format into the UNIX file system in real-time, before sending them over the network to the Grid environment. Another problem is effective switching the routing path from the telescopes. (there is no automated way to do that)

The following figure shows the capabilities of the Mark5 system of sending the Mk4 formatted data stream over the network. The transport of Mk4 files from Mark5 to Linux/UNIX file system can be done at the telescope site or at a remote site. Which option will be optimal or possible depends on the available resources at the different sites. At the Mark5 system the application puts a Mk4 data stream on a network (local or wide) and at the Linux/UNIX system the application Net2file gets the data from the network and puts it in a Mk4 file.

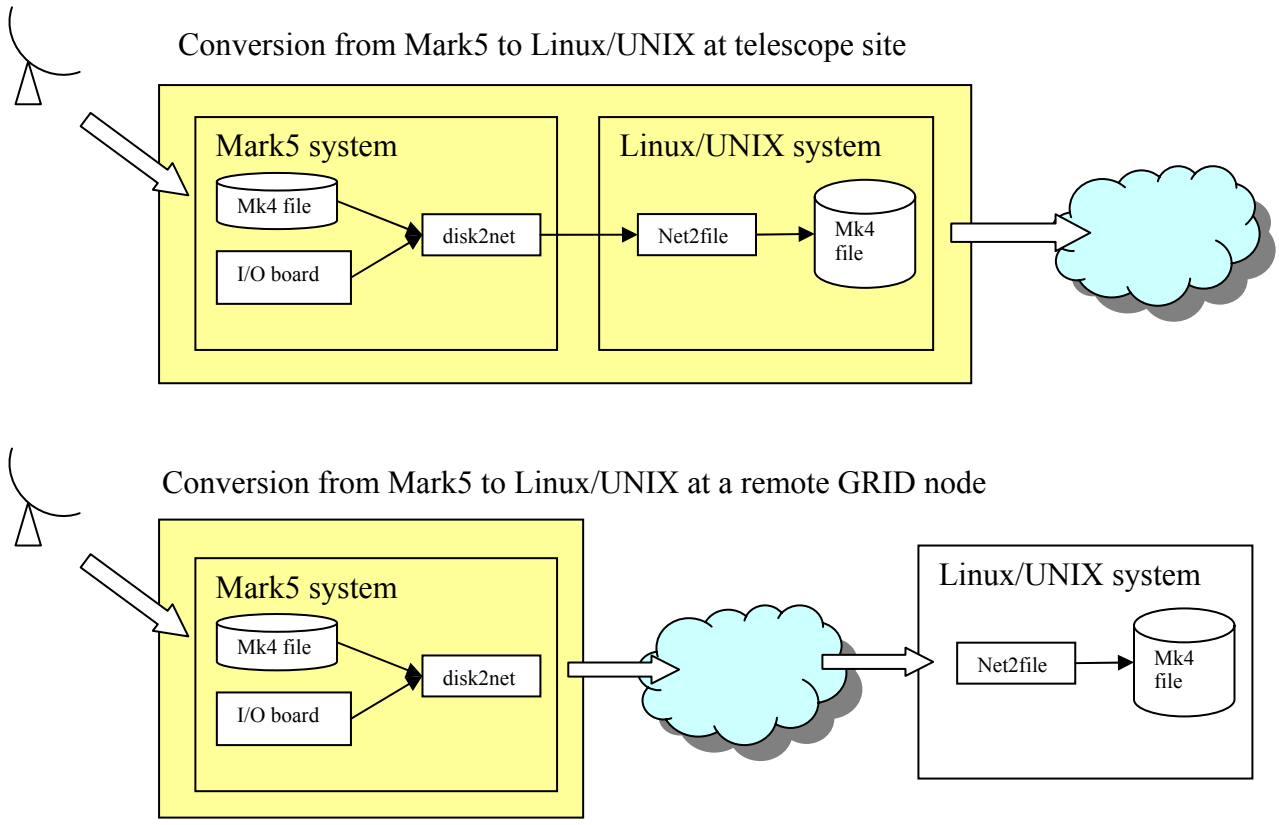

Figure 3. Possible ways for outcoming data stream from Mk5 system

In the first case a powerful file server at the telescope site is required, this functions as a buffer. The data can be sent sooner or later over the internet to a compute node. The second case requires a point to point internet connection between the Mark5 system and a remote file server. Buffering is can be done at the telescope site on the Mark5 system it self.

The file server can also do the time slicing of the incoming data into e.g. chunks of 10 seconds. This time-slicing can be done after the conversion of all the data or it is done simultaneously with a small delay after the reception of the data. Both solutions are technically feasible, but the second one requires more resources. After this, the workflow manager will have the possibility to send the data to the computational nodes for the distributed correlation.

The whole process is presented in a more general view in the figure below:

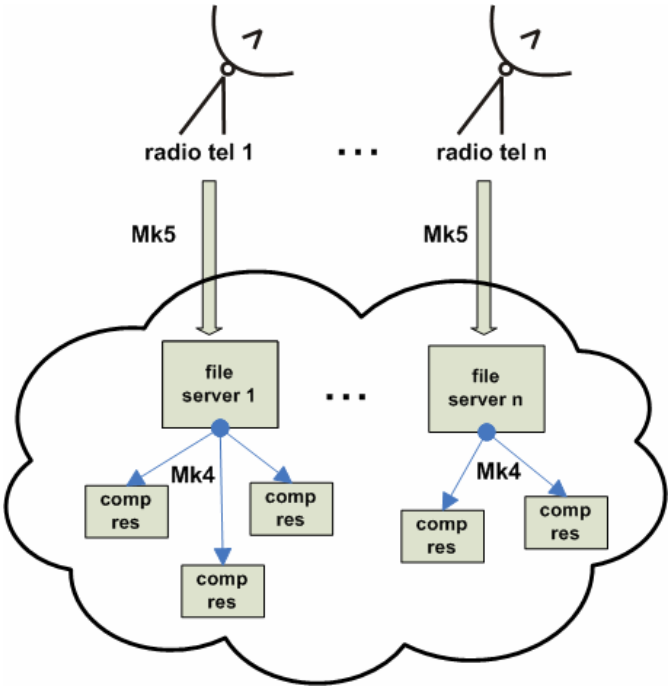

Figure 4. The overview of the solution with file servers

It was agreed that JIVE will be responsible for delivering, installing and testing of the software that will receive the data in Mk4 format, it and place it in the easily accessible UNIX file system, on the file servers.

3 Live Virtual Laboratory (VLab) demo (Dominik, Marcin)

3.1 General information about Virtual Laboratory System

Marcin Okoń was asked to present the general idea of the Virtual Laboratory System. The topics covered are as follows:
The architecture of the Virtual Laboratory (see figure below) was also discussed.

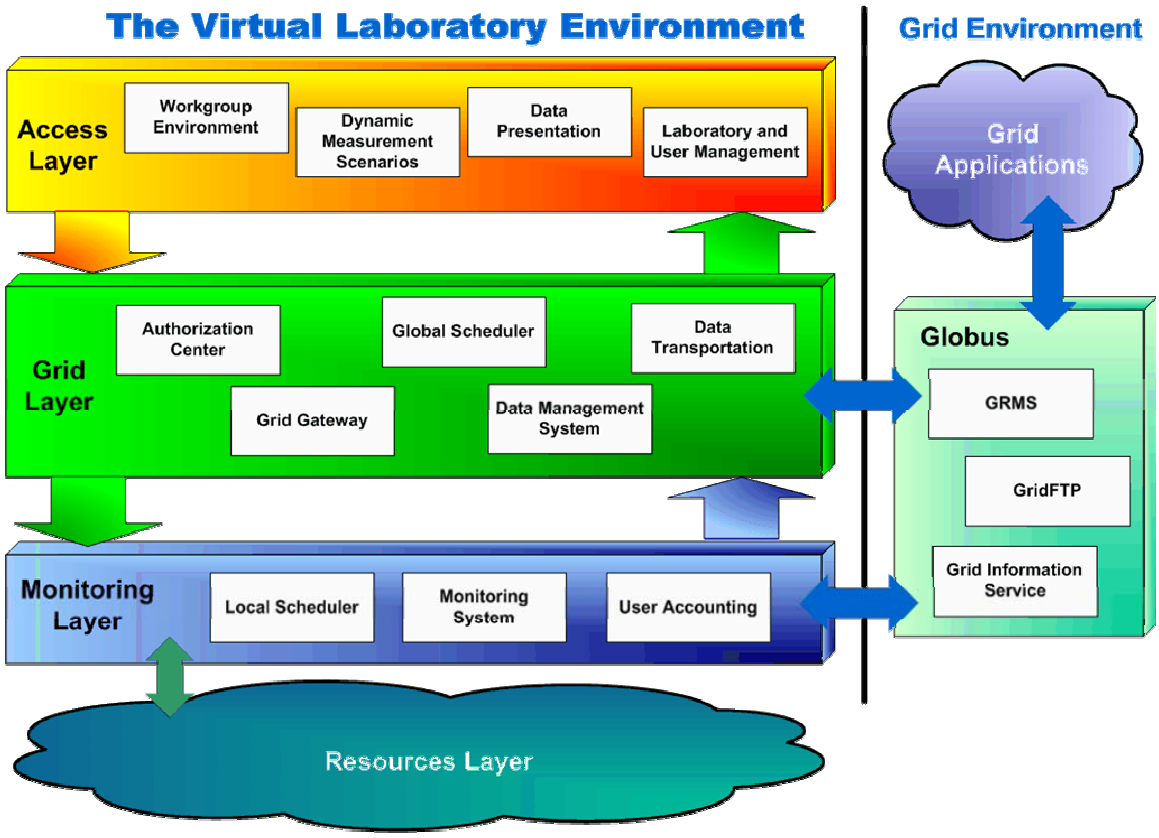

Figure 5. Virtual Laboratory architecture

The detailed description of the VLab architecture is not a scope of this document.
each VLab module was explained, together with the interaction with other components (see figure below)

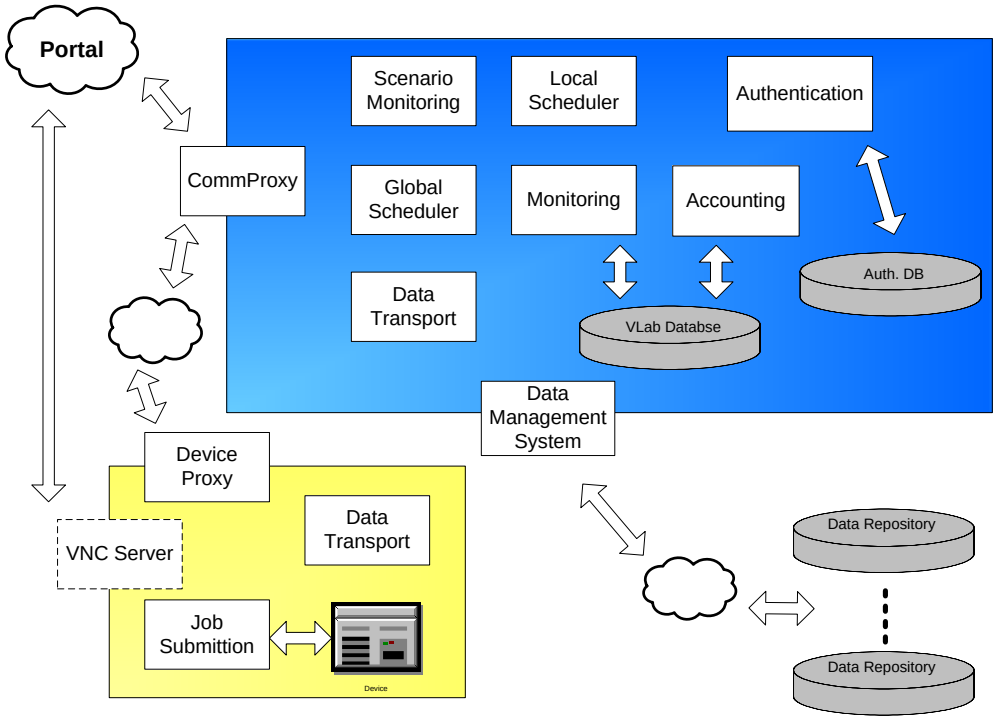

Figure 6. Interaction between the VLab modules

Integration with GRID environment was presented in detail.

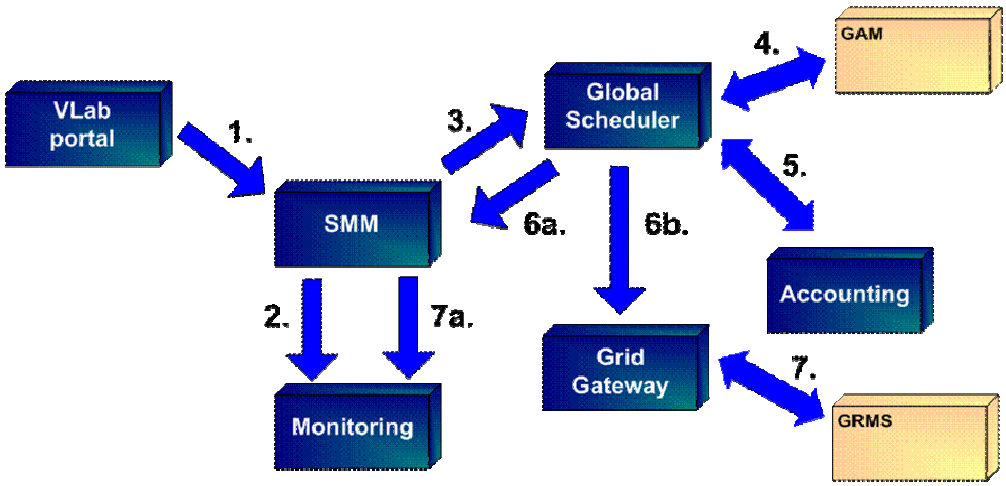

Figure 7. Task submission in GRID environment

Diagram description:

1. Task is sent to SMM module
2. Task is added to the DB
3. Task sent to Global Scheduler
4. Grid authorization in GAM
5. Accounting verification
6. Task submitted to GRMS (via Gateway)

3.2 Sample experiment using NMR Spectrometer device

Dominik Stokłosa was asked to give a tour of Virtual Laboratory usage. In order to demonstrate how the VLab works and how the modules discussed above interact with each other sample Nuclear magnetic resonance (NMR) experiment was demonstrated.

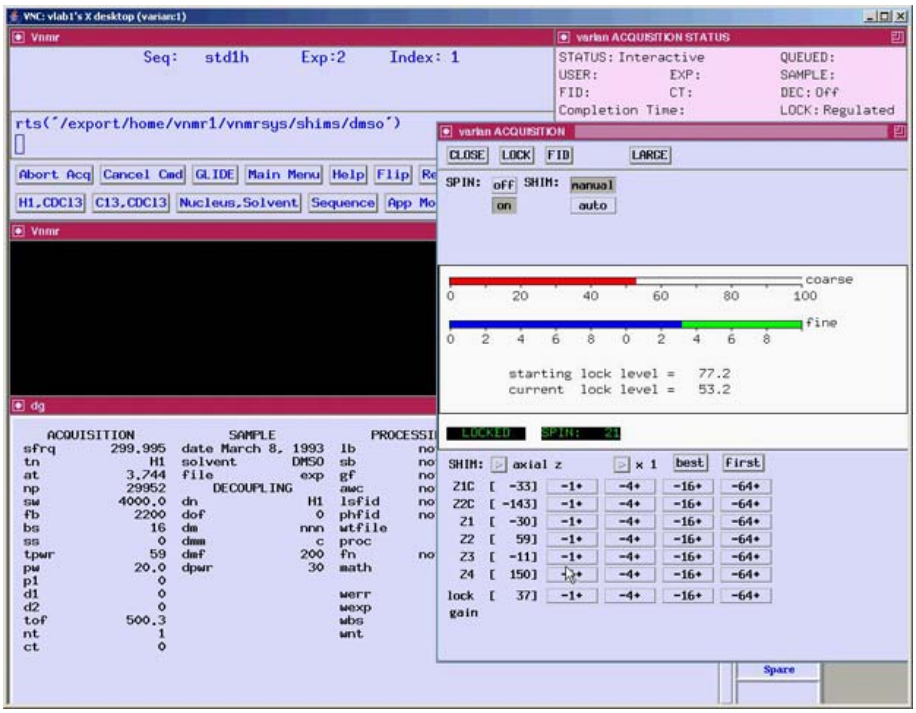

Figure 8. Remote control of the NMR spectrometer – Varian 300 MHz

One of the steps required to conduct an experiment is the creation of so called Dynamic Measurement Scenario (DMS) graph. The purpose of the DMS is to control the jobs execution and control of data flows between jobs. The visualized sample of the workflow graph is presented on the figure below.

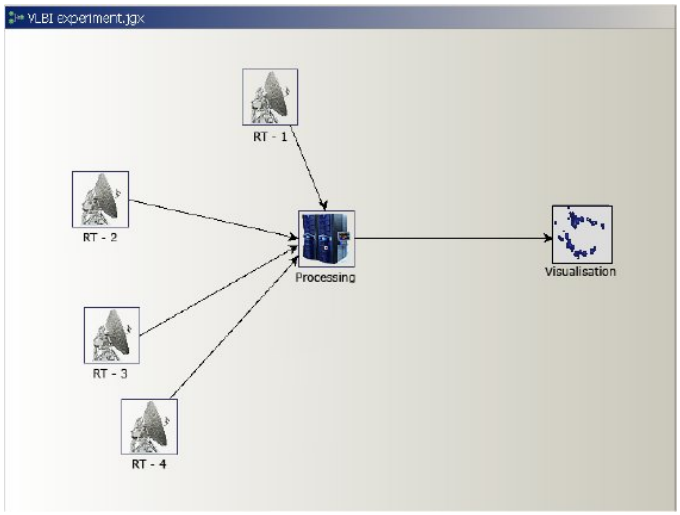

Figure 9. Sample workflow

PSNC has designed the special application for designing dynamic measurement scenario (DMS) and submitting such a scenario to the VLab System. The application is called Scenario Submission Application (SSA). It was agreed that the application will get redesigned in order to create a prototype of the Workflow Manager Application. The sample screenshots of the SSA application are presented on the figures below:

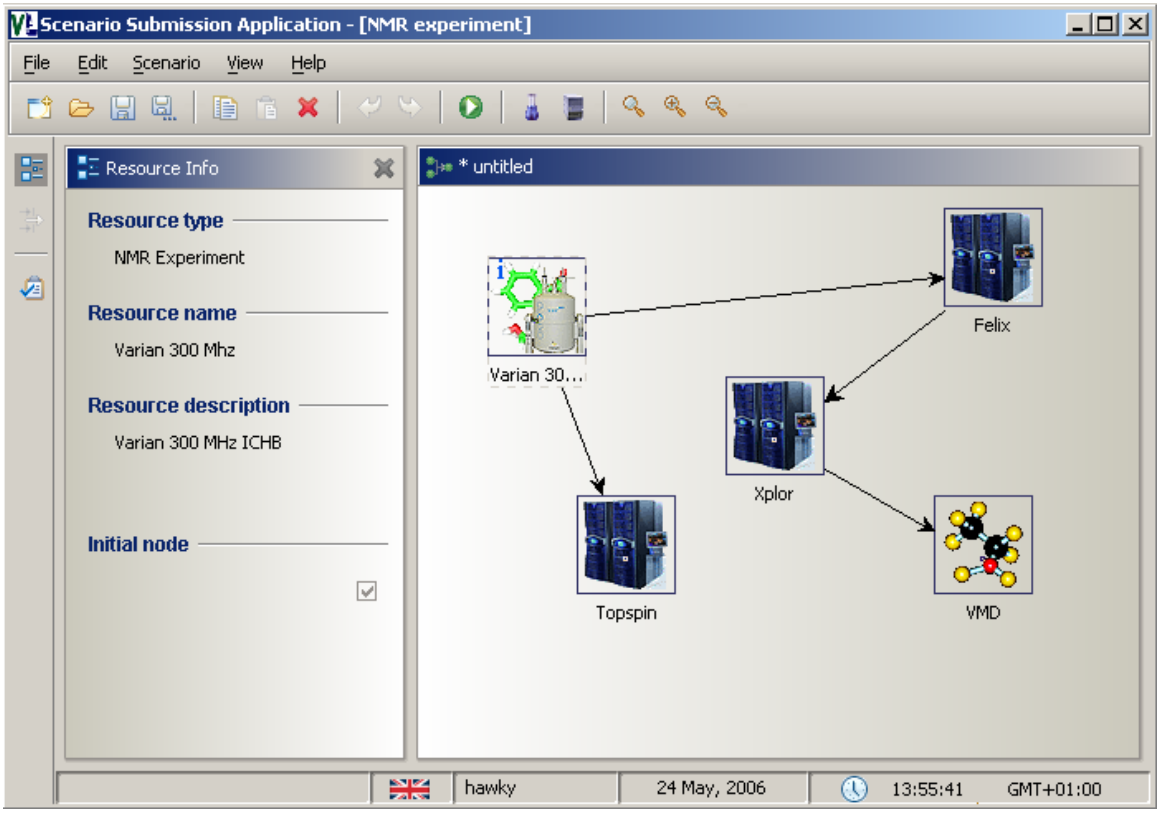

Figure 10. The main pane of the SSA application

The figure presents main pane of the Scenario Submission Application, which is divided into two sections:

- Resource Information Pane – dynamic pane, which displays certain details according to the given state of the application
- DMS Design Pane – this central pane is used during workflow creation process.

The next figure shows *Resource Properties dialog.* This dialog allows user to set up all the parameters required by the scientific device or computational node. This parameters can be easily customized for different purposes. The sample figure presents parameters from the field of NMR spectroscopy.

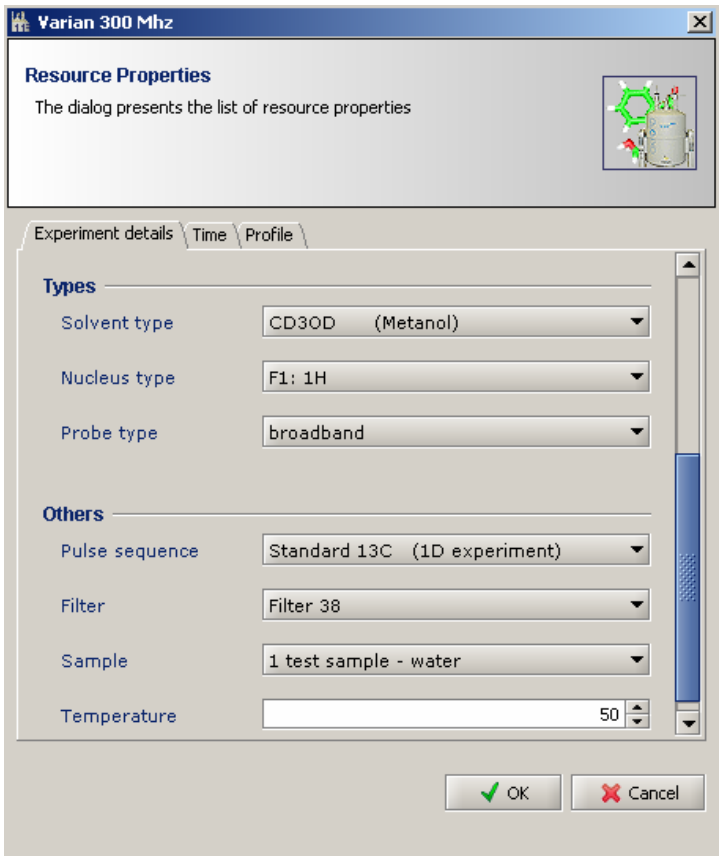

Figure 11. The properties dialog of the SSA application

The detailed description of the SSA application is not a scope of this document. However, the are several issues concerning SSA application that were agreed during the meeting:

- The design and architecture of the Scenario Submission Application will be used as a base for the design of the Workflow Manager
- The Workflow Manager will be able to read and visualize the VLBI experiment specification file. It was not agreed yet what parameters will get visualized first.

3.3 Digital Science Library

The Digital Science Library (DSL) was created on the base of the Data Management System (DMS http://progress.psnc.pl/English/index.html, which was developed for the PROGRESS project (http://progress.psnc.pl). Its main functionality, which is storing and presenting data in grid environments, was extended with the functions specific to the requirements of the Virtual Laboratory.

The general idea behind the Digital Science Library was also presented. Presentation was focused on the aspect of data distribution, meta data and GUI. The demonstration was performed using the NMR data as an example.

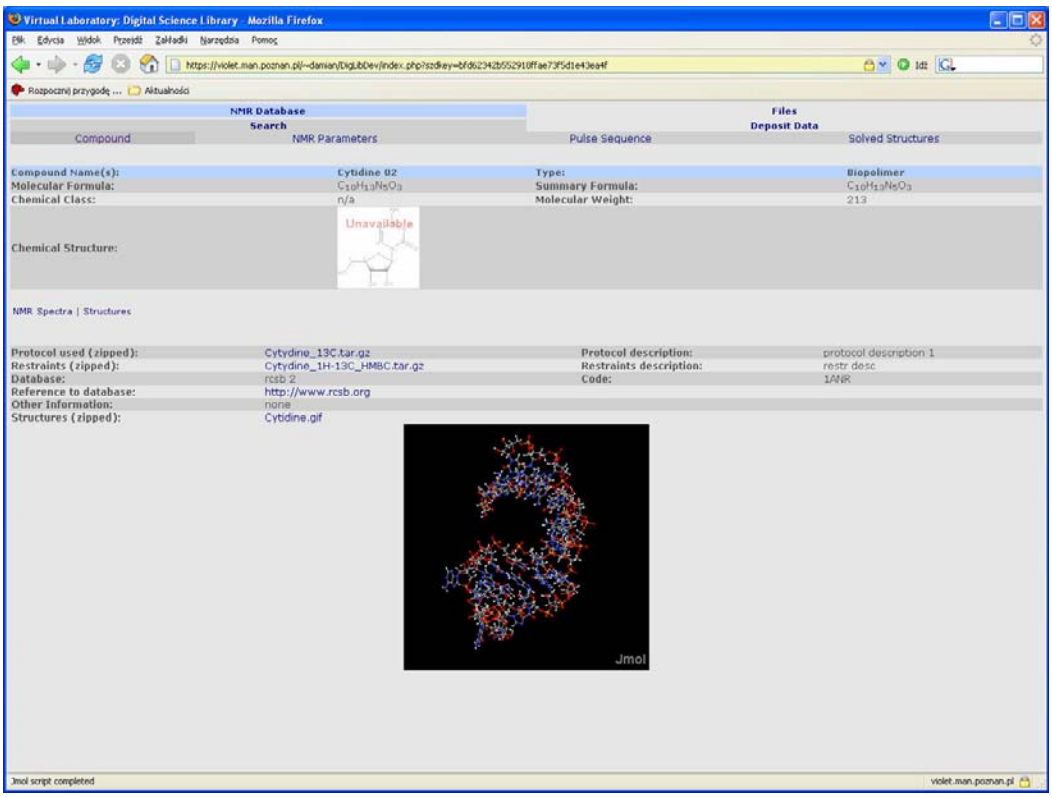

Figure 12. VLab Digital Library

The main idea of the Digital Science Library for the purpose of Nuclear Magnetic Resonance Spectroscopy (NMR) is to provide storing and data presentation used by the virtual laboratories. This data can be input information used for scientific experiments as well as the results of performed experiments. Another important aspect is the capability of storing various types of publications and documents, which are often created in the scientific process. This type of functionality, which is well known to all digital library users, is also provided by the DSL.

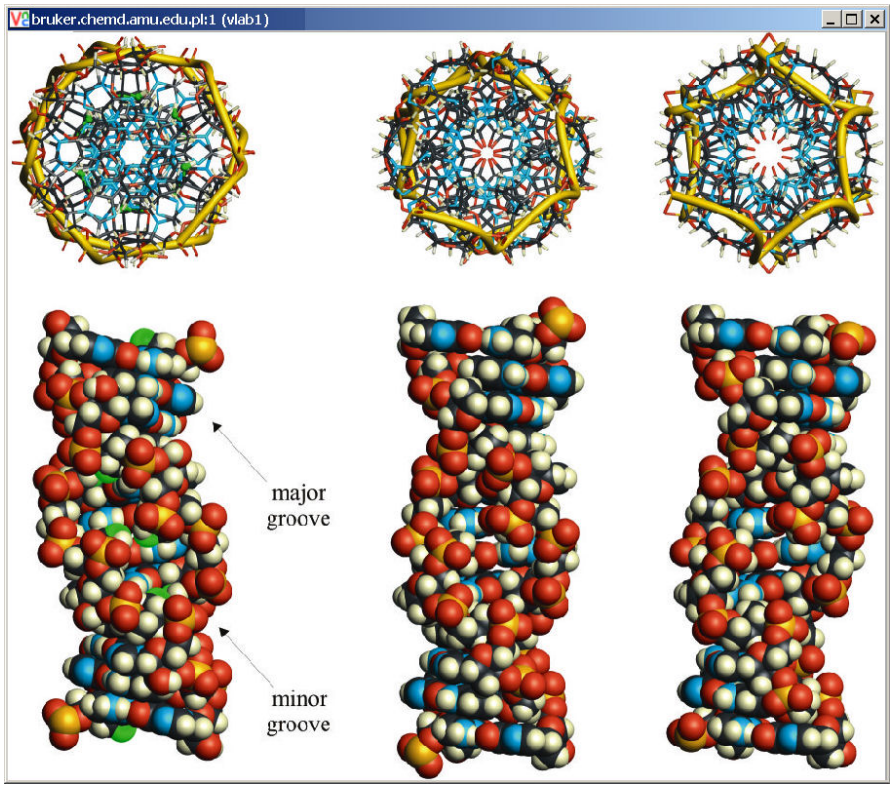

Figure 13. Digital Library – sample visualization

The implementation of the Digital Library for VLBI is beyond the scope of FABRIC, but this system can be a part of future projects and collaboration.

Second day (6th of July 2006)

4 Presentation of different file formats (Huib)

In the morning Huib shows some typical files and applications used in the definition of a VLBI experiment.

SCHED is a program for planning and scheduling VLBA, Global VLBI, EVN, and some VLA observations.

Scheduling with SCHED involves creating an input file with a text editor, and then running the program on that file. SCHED creates a variety of output files that provide summary and detailed information to the scheduler and telescope control information for the VLBI stations. Creation of the relevant SCHED output files (e.g. VEX files for EVN) remains the responsibility of the Principle Investigator (PI).

The main schedule input file is the file that contains the details of the particular project. It can have the most of the other files imbedded in it. This file must be created using a text editor by the user. This file should be given a name like bv016.key for project BV016.

The most important output file is the VEX formatted file, usually with the extension *.skd. These are the files needed for stations under control of the Goddard "Field System", e.g. MkIV telescopes. Writing of such files is available in SCHED They will be named, for example, bv016.skd. A single such file describes the observations for all antennas.

Extensive documentation on SCHED and its file formats can be found on the website: http://www.aoc.nrao.edu/~cwalker/sched/sched/sched.html. However for the FABRIC project only the VEX formatted file is important because this one is read by the workflow manager to control the whole correlation software correlation process on the GRID. A more detailed VEX file description can be found on the website: http://www.haystack.mit.edu/tech/vlbi/mark5/vex.html.

In order to read the VEX file a Fortran parser is available. Also some wrappers around this parser are available but these wrappers are not officially released and not supported.

5 Discussion on system design and definition of aims

5.1 General aims

During the discussion, and based on the issues addressed in sections 2.2 and 2.4 the system architecture shown in the diagram below came out as a possible end goal for the FABRIC project.

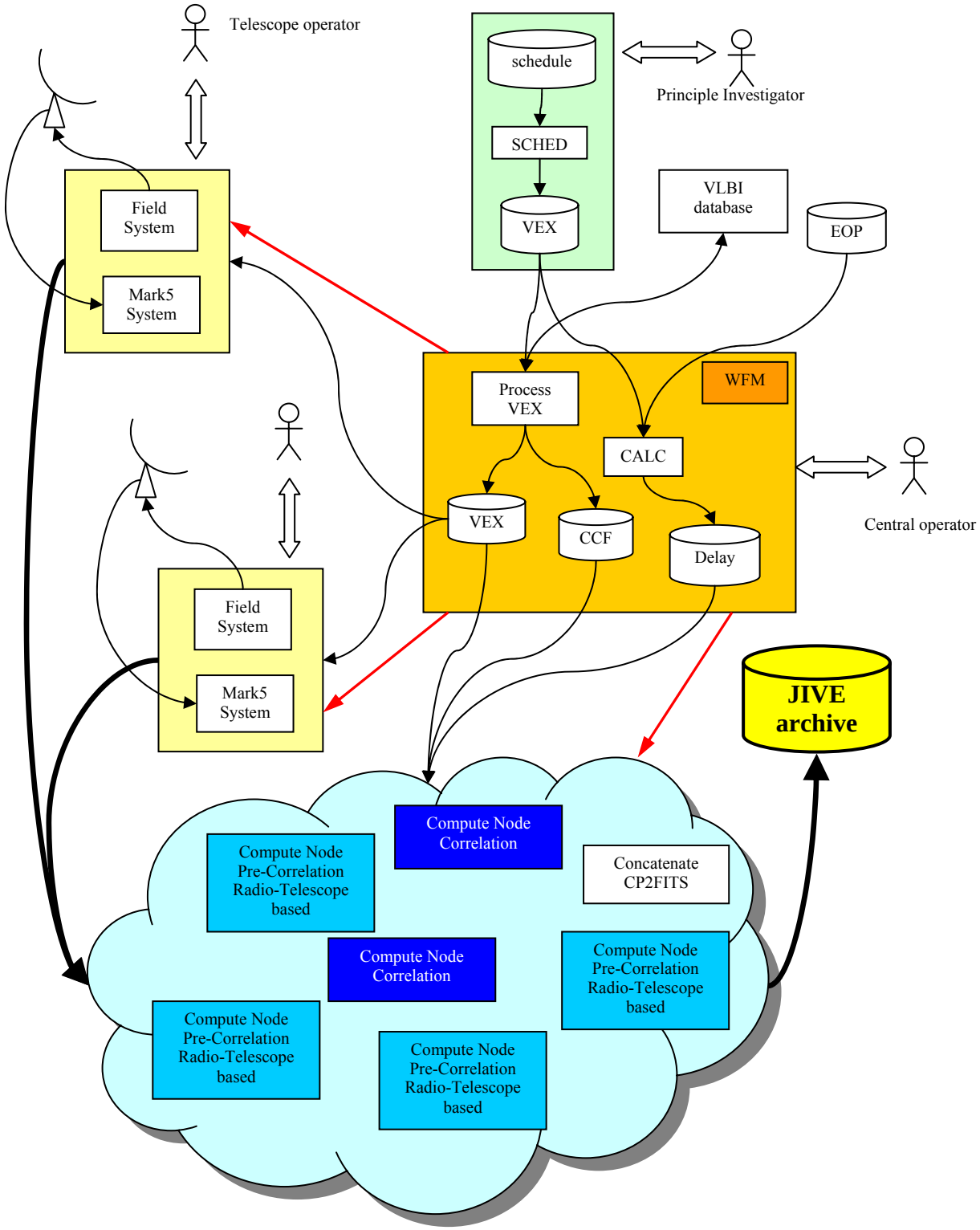

Figure 14. System architecture for distributed broad band correlation

A Principle Investigator (PI) has been granted observation time on various VLBI telescopes. Before the actual observation can start the PI creates a schedule file containing all the details of the observation, like what source to observe, which telescopes to use and when to observe. This schedule file is used to create a VEX file with the SCHED application. A central operator will be notified of the experiment. Now the PI has to wait until the data are in the JIVE archive.

The central operator uses the Work Flow Manager (WFM) to process the VEX file into a new VEX file and then sends these files to the telescopes participating in the observations. He also notifies the telescope operators a new experiment is scheduled. The WFM sends routing information to the telescopes and allocates the necessary compute nodes and broad band internet connections.

The telescope operator loads the VEX file in the Field System which controls the telescope and in the Mark5 system which records the data. At the start of the experiment the allocated compute nodes should be ready and waiting for data. The WFM has already send the relevant VEX file, Correlator Control File (CCF) and Delay file to the compute nodes. The data recorded by the Mark5 system is automatically send to the allocated compute nodes doing the processing as the data arrives.

The WFM also calculates the necessary delay tables before the correlation takes place using CALC and Earth Orientation Parameters (EOP). CALC is a standard application developed by geodesists to accurately determine positions on earth. Information on CALC can be found on http://gemini.gsfc.nasa.gov/solve/.

The Pre-Correlation nodes do the telescope based processing and other nodes do the correlation. A separate node concatenates the correlation products and converts it into the FITS file format. Finally the data is send to the JIVE archive and the PI is notified he can get his data. After a year the data becomes public.

In the previous case near real time distributed processing of the recorded data is assumed. This however puts heavy demands on the connections and on the compute nodes in terms of availability and performance. As an alternative data can be recorded first at the telescopes and processed e.g. in the week after the experiment.

The previous diagram assumes that the Pre-Correlator nodes are somewhere on the grid receiving the raw Mk4 formatted data files. However the processing done by these nodes is telescope based and it seems logical to do this processing at the telescope site itself and sending the processed data to the correlator nodes directly. This requires a heavy file server at the telescope site and more band width to get the pre-correlated data to the correlator. Both options were discussed in section 2.4.

The feasibility of the end goal described in this section is not clear yet, there are still many questions to be answered and uncertainties to be solved. However it is worth while investigating and on the way to it achievable goals have to be defined and pursued. In the end the architecture might look different.

5.2 Short term aims

Despite the general goal, which is to create a system based on the architecture described in section 5.1, shorter term-goals have to be established. This will allow to validate the proposed approach and initial assumptions, and will help to discover more issues and potential problems to solve.

The very first goal is to make the current software correlator as described in section 2.3 cluster friendly. To achieve this data have to be passed from one application to the other using MPI and

each CPU in the cluster will run one application. The following diagram shows the architecture:

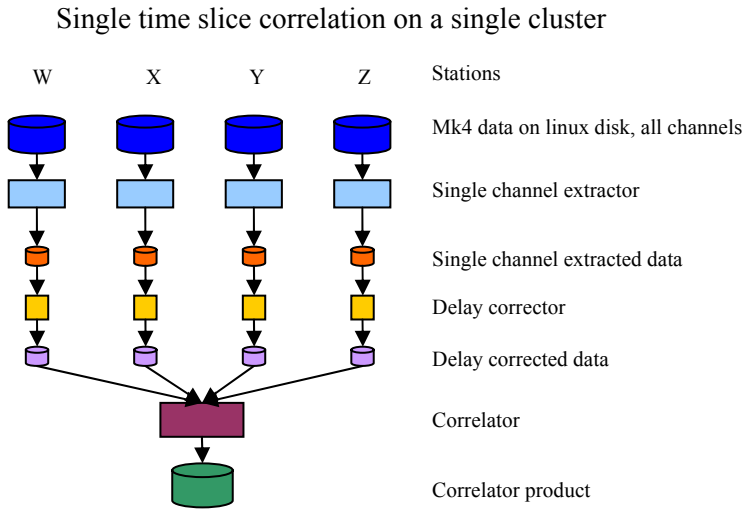

Figure 15. System architecture for single channel correlation

Mk4 formatted data from all telescopes is available on linux disks at a single cluster. For all telescopes only a single channel is extracted and correlated. All applications produce files as output and accept files as input. Each application runs on a single CPU. On the cluster scheduling the various applications has to be coordinated centrally. In the next development step intermediate files have to be replaced by a more direct way of data transfer, e.g. by using MPI. The correlation process can be accelerated by time slicing the correlation process. Finally all time slices have to be concatenated.

The advantage of this architecture is that soon hands on experience can be gained. Issues like data flows, computational performance, inter program communication, coordination of all applications etc. can be addressed in a relatively simple environment.

In a next stage recorded data is played back in a remote place e.g. at a lower rate then the original recording rate. One or more compute clusters have to process the incoming data

Other intermediate goals on the way to the final architecture have to be defined in a software correlator design document

6 Live demo of the SFXC application (Ruud)

Ruud demonstrates the current broad band correlator software as it is described in section 2.3. The demonstration was very useful because

- It gave Marcin and Dominik an idea which processing steps are needed in the correlation process and how to operate the applications so they can use it themselves.
- Marcin and Dominik indicated what changes had to be made to make the software "cluster friendly". They proposed to use MPI for inter application communication
- Opportunities for improvement and optimization of the code were identified.

7 Agreed todo list

JIVE

- Deliver correlator code and sample input data
- Deliver sample VEX file
- Documentation of the VEX format
- Decide what information from the VEX file should be displayed in the Workflow Manager
- VEX parsers (yacc, lex, perl, any others ?)
- Correlator control files
- User manual of the correlator
- Creation of the general control files validator
- Software Correlator design document

PSNC

- MPI information (external links, reading material)
- Intel Trace Collector and Trace Analyzer license information
- This document