THE COMPUTING MODEL OF THE EXPERIMENTS AT PETRA III

Melvin Alfaro, Martin Flemming, Julia Grabitz, Thorsten Kracht*, Birgit Lewendel, Teresa Núñez, Peter van der Reest, André Rothkirch, Frank Schlünzen, Eugen Wintersberger
DESY, Hamburg, Germany

Abstract

The PETRA storage ring at DESY in Hamburg has been refurbished to become a highly brilliant synchrotron radiation source (now named PETRA III [1]). In comparison with the DORIS beamlines, the PETRA III experiments have larger complexity, higher data rates and require an integrated system for data storage and archiving, data processing and data distribution. Tango [2] and Sardana [3] are the main components of our online control system. Tango serves as the backbone to operate all beamline components, certain storage ring devices and equipment from our users. Sardana is an abstraction layer on top of Tango. It standardizes the hardware access, organizes experimental procedures, has a command line interface and provides widgets for graphical user interfaces.

The high brilliance together with the rapid read-out of modern 2D detectors dramatically increase the data rates and volumes. At PETRA III all data are transfered to an online file server which is hosted by the DESY computer center. Near real time analysis and reconstruction steps are executed on a CPU farm. A portal for remote data access is in preparation. Data archiving is done by dCache [4]. An offline file server has been installed for further analysis and inhouse data storage.

INTRODUCTION

By the year 2009 the reconstruction work to turn the storage ring PETRA into a highly-brilliant 3rd generation synchrotron radiation source was completed. The new facility, PETRA III, consists of 14 beamlines which are operated by DESY, the EMBL and HZG. The EMBL stations are independent of DESY as far as IT is concerned. By now, most of the beamlines are in user operation, the rest are currently commissioned.

DESY has a long-lasting experience in the field of synchrotron radiation experiment control which comes from the beamlines at DORIS, FLASH and PETRA. PETRA III creates new challenges. The beamlines have longer extent, consist of more elements and use a greater variety of electronic devices. In addition, the requirements of user interfaces are increasingly demanding. This applies to the scripting language, the command line and the graphical user interfaces. Consequently the online control system had to be newly designed. The original solution of controlling the experiments by a single program, Online [5], was replaced by a distributed client-server system.

The new 2D detectors which are used at PETRA III generate very high data rates. Therefore a concept for data acquisition, data management, reconstruction, analysis and data transfer had to be developed.

This note describes the layout of the computing model which has been implemented to meet the above mentioned requirements.

Figure 1: The PETRA III Hall.

EXPERIMENT CONTROL

The following issues have to be considered when designing an experiment control system:

- **Sustainability** A modular system with defined interfaces for hardware, communication and application layers can be extended and maintained by a group of software developers.
- **Flexibility** An efficient customization of the system is important for the best use of the beamtime. This customization is required due to the frequent change of user groups at the beamlines.
- **Platform independence** Linux and Windows have to be supported. Bindings to scripting and programming languages are mandatory.
- **Performance** An online control system needs sufficient bandwidth to handle motors, counters, timer, etc. However 2D detectors, creating by far the highest data rates, do not use the control system for data transfer but access storage media directly.

An online control system has to be based on a product that is supported by an international community. One reason is to save resources by benefiting from the work of others. At least as important is the communication among the members of the collaboration about new ideas and experiences. And finally, collaborations help to develop common user interfaces.

* Thorsten.Kracht@desy.de

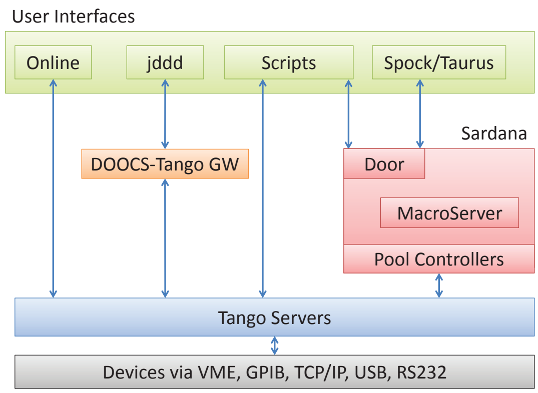

Figure 2: The main components of the experiment control system.

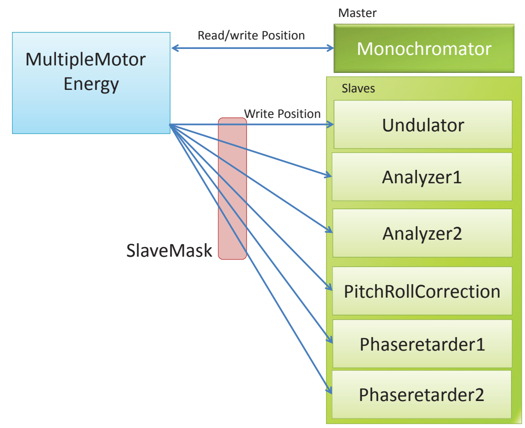

Figure 3: An example for a MultipleMotor server: the beamline energy.

Figure 2 gives a schematic view of the experiment control system. Tango serves as the hardware access layer. Clients can communicate directly to the Tango device servers or make use of the additional functionality that is provided by the Sardana framework.

The Tango Implementation

In order to limit the effect of corrupted database servers at least one Tango instance is installed for every beamline. Another choice was to group devices into classes with identical Tango interfaces. For cameras the LIMA [6] framework has been imported. Currently, area detectors available at PETRA III beamlines are being adopted.

The experiments at PETRA III use a variety of motors with many different parameters. It has been decided to implement a minimal interface in all motor classes, the TANGO_MOTOR. It executes the commands Calibrate, StopMove, ResetMotor and has the attributes Position, UnitLimitMin, UnitLimitMax. At DESY several motor servers of common interest have been developed. All of them implement the TANGO_MOTOR interface:

- **MultipleMotor** This server moves several TANGO_MOTORs concurrently. One of them is assigned to be the master which defines the position of the compound device. The slave motors, which follow the motion of the master, are enabled by a mask attribute, see figure 3.
- **AttributeMotor** A server that converts an attribute of a arbitrary device into a TANGO_MOTOR.
- **TcpIpMotor** This server uses an ASCII protocol via TCP/IP to control arbitraty motors that are temporarily installed by user groups.

Other general purpose servers that have been developed for PETRA III are: the CollisionSensor which avoids hardware damage due to unintentional movements and the ClipBoard server which facilitates a client-client communication by exporting several string attributes.

Tango Clients

These clients are currently used at the PETRA III experiments:

- **Online:** For a long time many DESY beamlines have been operated by this program. It has a command line interface, a binding to Perl and a graphical user interface.
- **Python scripts** Scientists started to write small python applications using PyTango [7] shortly after the PETRA III operation began. PyTango is a python module which exports the complete c++ Tango Api to Python.
- **Diffractometer Server** At PETRA III two beamlines have Eulerian 6-circle diffractometer. The diffractometer server, written by F. Picca of SOLEIL, is used for crystal orientation and direct-reciprocal space transformations imposing certain conditions.
- **jddd** jddd [8] applications exist for every beamline. They display selected variables for a quick overview of the setup, figure 4. jddd has a hierarchical structure with several expansion levels.

Although PETRA III started operation successfully using the above mentioned clients, it was planned to upgrade our user environment. We have chosen the Sardana framework to be the backbone of our future online control system.

Sardana at PETRA III

Sardana was created at ALBA. Important contributions came from the ESRF. Sardana is a framework that structures the domain of experiment control programs and user interfaces. It has several components: the device pool is an additional hardware standardization layer, the MacroServer organizes the execution of procedures that are needed for

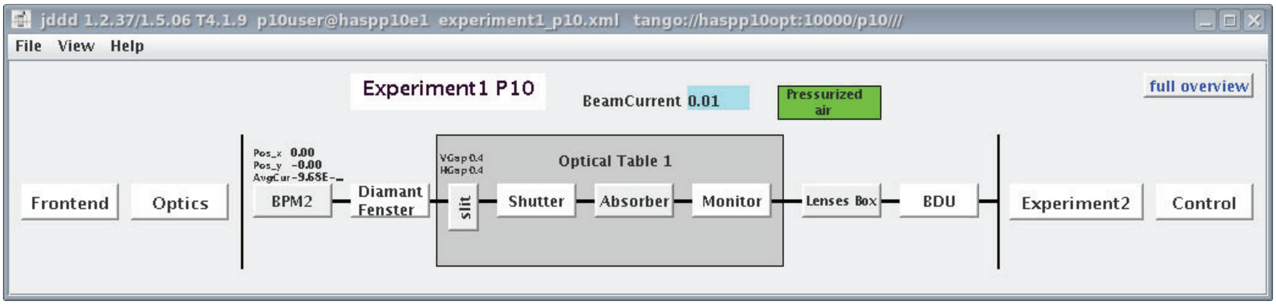

Figure 4: jddd application of a PETRA III beamline, partially expanded.

the measurements, Spock serves as the command line interface and Taurus builds GUIs. These features make Sardana well suited for PETRA III. In the meantime DESY participated in this project by adding some extensions. It has been rolled-out recently to the experiment PCs. First test measurements have been performed.

Data Format

The current practice to produce in a single experiment myriads of small files in a large variety of partially proprietary formats has a bad impact on the performance of the storage and archive system, tremendously increases the complexity of the data management tasks, hinders application development, prevents usage of data across disciplines as well as facilities and makes the long-term preservation of scientific data almost impossible.

To enable users to work with data collected at different beamlines at the same or different facilities, the PNI-HDRI [9] initiative (in close co-operation with the PaN-data project [10]) aims to develop and implement a standard data format, which also permits to aggregate data files from different sources into a performant, portable and self-describing form.

Nexus [11] has been selected as the underlying file format since it's based itself on widely accepted standards (HDF5) and because it has a growing user community in the photon, ion and neutron science world. It is part of the HDRI project to demonstrate experiment specific implementations and to provide guidelines how the application specific information is organized. The experiments at PETRA III will follow the developments of HDRI.

It cannot be expected that the Nexus will generally be accepted by the international photon science community. SOLEIL and ANSTO proposed to add an abstraction layer, CDMA [12], to help application developers to uniformly access files. CDMA uses a dictionary mechanism to translate application notions to data file entities. In order to benefit from these developments, it is planned to write a CDMA plugin for Nexus.

DATA MANAGEMENT

So far it was discussed how an experimental setup is controlled to perform measurements and collect data. This section describes how this data is treated afterwards.

The PETRA III beamlines send their output directly to the online file server. It has the purpose to buffer the data before they are archived and to make them available for near real time analysis, figure 5. All data are copied to the dCache tape pool for archiving. Data to be analyzed is staged to the dCache disk pool. In addition there is an offline file server for storing analysis results and temporary data. These services are provided by the DESY computer center.

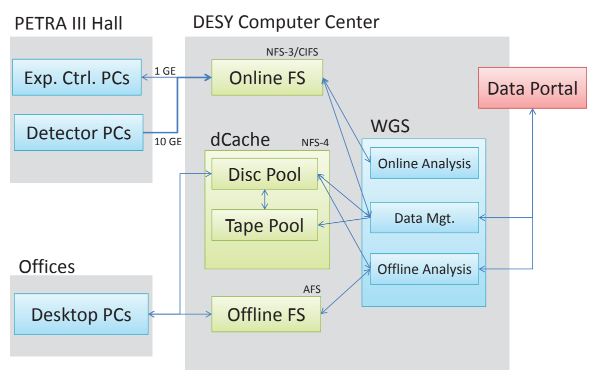

Figure 5: Main data and control flow paths.

Network

Internally the PETRA III experiment hall has a 1 GE Ethernet infrastructure. The hall itself is connected to the DESY computer center by 10 GE Ethernet. Detectors which generate high data rates are connected by 10 GE lines. The effective speed for writing data to disk via the 10 GE lines is about 300 MB/s (from Linux PCs).

File Servers

The online file server storing the data during measurements can be accessed from either Linux or Windows™ systems. Typically, the Linux based experiment PCs and the detector PCs mount this server by NFS (version 3). It offers CIFS-shares for Windows™ based systems.

The online file server consists of 2 file server systems, each managing 11 drive trays of 14 disks each. It is set up as a raid 6 storage system, has takeover options for the 2 heads in a single system and is operated with snapshot option of 20% of total storage. The total capacity of the online file server is 145 TB for all beamlines (excluding snapshot and spare disks).

If an experiment needs a fast reconstruction step to optimize the measurement, it is done with the data on the online file server.

The offline file server stores analysis results, no raw data. It is accessible from the experiment PCs and the workgroup server via AFS. The capacity of this file server is about 40 TB in a raid 6 setup.

dCache

The dCache is a storage system consisting of a disk pool and a tape pool. It was developed by DESY and FERMILAB for storing high energy physics data. The dCache becomes increasingly important for the photon science department of DESY.

Data are migrated from the online file server to the dCache. During this migration the ownership changes from the beamline account to a user-specific account. The beamline scientists in charge keep full access rights.

Data are directly accessible (via NFS4.1 exports) by analysis programs as long as they are stored on the disk pool. Data that are already on the tape pool need to be staged before they can be read. Data stay on the disk pool for at most 180 days. The residence time in the tape pool is not limited so far.

Data Portal

It is expected that certain guest groups do not have the sufficient IT infrastructure at their home institutes to store and process large data volumes. DESY plans to provide these scientists with storage and compute services. A data portal is currently being prepared that facilities access to the dCache and permits the full remote management of the data, which includes the search for data and meta-data and a user controlled rights management.

Compute Server

Currently PETRA III computing is performed on a blade center which is divided into 16 work group servers (WGS). They have a dual-quad-core cpu with 24 GB RAM and 16 GB swap space running 64bit Scientific Linux. Part of the WGS are organized in a server pool, the remaining ones are addressed directly and can be allocated to specific beamlines. Computing can also be performed on Nvidia™ GPUs. WGS and GPUs are used for reconstruction, analysis and simulation as well as for software compatibility tests or data migration.

CONCLUSIONS

PETRA III computing integrates products that originate from activities of international and national partners: Tango, Sardana, CDMA, dCache. This approach saves resources and enhances the communication on the relevant topics which is crucial for the fast and continous incorporation of new developments. Furthermore, collaborations among synchrotron radiation facilities are the precondition for establishing common user interfaces. A feat that would ease the burden of our scientific users in performing measurements at different sources.

It has been demonstrated that the experiment control system consists of several layers with well defined interfaces making it flexible enough to fulfill the current requirements and to cope with future challenges.

ACKNOWLEDGMENTS

We would like to thank all who made contributions to the Tango control system and our colleagues from ALBA for having created Sardana, in particular T. Coutinho, C. Pascual and J. Klora who gave us so much support. We thank F. Picca for allowing us to use his diffractometer server. We also acknowledge the work of DESY-MCS on jddd.

REFERENCES

[1] The New Synchrotron Radiation Source at DESY, `http://petra3-project.desy.de`.

[2] Tango Control System `http://www.tango-controls.org`.

[3] J. Klora, T. Coutinho et al., The architecture of the ALBA Control System, Proceedings NOBUGS 2008

[4] `http://www-dcache.desy.de`.

[5] Online, Data Acquisition and Beamline Control, `http://hasylab.desy.de/online`.

[6] LIMA, Library for Image Acquisition, `http://forge.epn-campus.eu/Lima`.

[7] PyTango, `http://www.tango-controls.org/static/PyTango/latest/doc/html/index.html`.

[8] jddd, A Java DOOCS Data Display, `http://jddd.desy.de`

[9] High Data Rate Initiative for Photons, Neutron and Ions, `http://www.pni-hdri.de/`.

[10] PANDATA – Photon and Neutron Data Infrastructure, `http://www.pan-data.eu/`.

[11] NeXus, `http://www.nexusformat.org`.

[12] N. Hauser et al., CommonDataModel. A Unified Layer to Access Data from Data Analysis Applications Point of View, proceedings, ICALEPCS 2011