# LARGE SCALE DATA FACILITY FOR DATA INTENSIVE SYNCHROTRON BEAMLINES

R. Stotzka, W. Mexner, T. dos Santos Rolo, H. Pasic, J. van Wezel, V. Hartmann, T. Jejkal, A. Garcia, D. Haas, A. Streit, Karlsruhe Institute of Technology, Karlsruhe, Germany

*Abstract*

ANKA is a synchrotron light source located at the Karlsruhe Institute of Technology, providing light from hard X-rays to the far-infrared for research and technology. It serves as a user facility for the national and international scientific community, currently producing 100 TB of data per year. Within the next two years a couple of additional data intensive beamlines will be operational producing up to 1.6 PB per year. These amounts of data have to be stored and provided on demand to the users.

The Large Scale Data Facility LSDF is located on the same campus as ANKA. It is a data service facility dedicated for data intensive scientific experiments. Currently, storage of 4 PB for unstructured as well as structured data and a HADOOP cluster, used as a computing resource for data intensive applications, are available. The campus experiments and the main large data producing facilities are connected via 10 GE network links. An additional 10 GE link exists to the internet. Tools for an easy and transparent access allow scientists to use the LSDF without having to bother about the internal structures and technologies. Open interfaces and APIs support a variety of access methods to the highly available services for high throughput data applications. In close cooperation with ANKA, the LSDF provides assistance to efficiently organize data and meta data structures, and develops and deploys community specific software running on the directly connected computing infrastructure.

## DATA INTENSIVE SCIENCE AT ANKA

ANKA was designed and constructed by Forschungszentrum Karlsruhe in 2003 and is now operated by the Karlsruhe Institute of Technology (KIT). As a large scale research facility, ANKA provides beamtime for fundamental and application–oriented research to users from Germany, Europe and beyond [1]. Fourteen beamlines at ANKA are operational. The beamlines for analytic service cover techniques from spectroscopy to diffraction over a large spectral range from far infrared to hard X–rays. At the moment a high–resolution X–ray diffraction beamline (NANO) and an imaging beamline (IMAGE) are under commissioning and construction.

### *The Next Generation of Ultra Fast Tomography Beamlines*

The substantial progress made in recent years, in fields like mechanics, X–ray optics and detector systems as well as the tremendous increase of processing speed and I/O bandwidth gives rise to a paradigm shift in the design and execution of tomography experiments. A new type of smart experimental station becomes possible using the vast computational power of massively parallel computation units. The process under study as well as the measurement procedure itself can be actively controlled by data driven feedback loops in direct dependence on image dynamics and spatio–temporal contrast.

Digital area detectors are an essential component for high–speed X–ray imaging experiments. A widely used X–ray detector type is the so–called indirect detector, which X–rays converts into visible light by a thin scintillating crystal. The resulting image is then magnified onto the CCD (or CMOS) camera by a diffraction limited optical system. The use of highly absorbing scintillation crystals with high light yield and optimized detector optics that are able to withstand the intense broad–band radiation produced by modern synchrotron sources has permitted a significant reduction of exposure times. High frame rates are achieved by employing column–parallel or block readout active pixel detectors, which reduce the read–out time for a full frame substantially, compared to classical high-dynamic range CCDs.

Commercially available, scientific grade cameras produce projection data at a rate of up to 750 MB/s. A detailed description of the reconstruction algorithm and its speed–up in comparison to a CPU implementation can be found in [2].

A smart experimental station for high-speed and high–throughput tomography is currently designed for the IMAGE beamline, which is constructed at the ANKA storage ring. The photon source employed for tomography will be a superconducting wiggler, which covers an energy range up to 100 keV. Compared to the bending magnet beamline Topo–Tomo [3], the flux density will be increased by more than one order of magnitude, yielding a reduction of exposure time on the same scale.

### *The Need for High Performance Data Management*

Processing and storing the data of high–speed and high–throughput imaging beamlines are challenging tasks. Using robotic sample changers, the sample throughput is increased substantially. For instance, raw data rates presently produced at the Topo–Tomo bending magnet beam line, with a sample changing time of 15 seconds and a total scan time of 30 seconds, are 350 MB/s averaged over a daily run.

A raw reconstructed volume has a size of 50 GB, which needs to be stored alongside the raw projection data.

Data analysis is often a laborious manual process that includes several steps that cannot be automated easily, for instance finding optimal parameters of segmentation routines or even selecting pre–processing algorithms to present suitable data for post–processing. Thus, large amounts of data need to be accessible for long periods of time with high bandwidth and low latency for data processing and visualization.

The management of large scale data, its secure storage and archiving, fast and reliable access as well as high performance processing are essential for cutting edge beamline research. Usually, costs caused by growing infrastructures, reliability and sustainability of large scale data are limiting the extension of current beamline experiments.

## STATE-OF-THE-ART IN DATA MANAGEMENT FOR DATA INTENSIVE BEAMLINES

### *The Topo–Tomo Beamline*

The data management of the Topo–Tomo beamline has evolved over the years with its construction and each technical extension. Most data storage and access operations are directly steered by the user.

Samples are mounted on a rotary stage that allows a precise rotation with typical 5–10 kHz depending on the used camera. The whole detector and sample positioning setup is controlled via TANGO [4, 5] and SPEC [6, 7]. In a first step, the image data are temporarily stored on a local high performance 30 TB SAN. To check the quality of the acquired data, radiograms are monitored. Then the acquired data are processed with a dedicated four Nvidia Tesla GPU server. The final visualization for tomography is performed by VG Studio MAX [8]. It runs on a dedicated visualization server with 196 GB RAM, which allows a high performance processing of the whole tomography in memory.

Finally, the 3D tomogram plus the raw data are stored as unstructured data at the Large Scale Data Facility (LSDF). Most of the sample specific meta data that could not registered via the control application is still archived via the old fashioned laboratory book.

### *Data Management Systems for Beamlines*

The maturity level of the data management seems to be very diverse across a variety of institutions and beamlines, and this heterogeneity is partly caused by different problems and requirements. The crucial point in beamline data management is indeed the amount of generated data and the desired throughput. In case of very data–intensive experiments the data management becomes challenging [9]. Nevertheless, beamlines with a primitive data management, where the user seems to be the main data management component and is responsible for all kinds of data storage, processing, transport and archiving, are still widespread.

An example of a data intensive beamline with complete data management during the whole data life cycle can be found at the National Ignition Facility (NIF), a research device located at the Lawrence Livermore National Laboratory [10]. The most important highlights are automatic data ingest, automatic triggering of analysis upon new data arrival, data provisioning from various data sources, policy–based automated hierarchical storage management, long term storage architecture and employment of HDF5. It is designed for beamlines with a fixed configuration that rarely needs adaptions for new scientific experiments.

Another data management system was developed and employed at SOLEIL [11]. The main idea is to produce individual data sets in form of a single and well defined atomic package containing all the relevant contextual information (including informations about the experiment itself, instrumentation, sample, user, etc.). The logical file format NeXus [12] with HDF5 as underlying physical data format is used.

The software architecture reflects the natural separation of different concerns, including hardware–layer and software–layer components for DAQ on device level, data collections and transformations and the data storage. It employs a TANGO based modular control system, so the data management components for data collection, transformation and storage, are mainly treated and integrated as TANGO devices.

Due to the required flexibility for continuously adapted beamline experiments, we think the SOLEIL Experimental Data Storage Management could be a great foundation for a sophisticated data management system, coupled with the Large Scale Data Facility infrastructure at KIT.

## THE LARGE SCALE DATA FACILITY

The Large Scale Data Facility (LSDF) comprises on–line storage, data analysis computing cores, archival capacity and data management services, initially funded by the Helmholtz Association and the state of Baden–Württemberg in order to facilitate data–intensive science. Based on the experience in designing and continuously enhancing and the documented excellence of the WLCG Tier–1 centre GridKa [13] at KIT for the high-energy physics community, the LSDF is not limited to single scientific communities — instead it aims to be open for all scientific disciplines requiring large scale data storage, analysis and archiving on a regional, national and international level. Detailed plans and budget allocations exist to enhance the LSDF, in accordance with the requirements of the scientific user community.

The project has currently 4 PB of storage available, together with a compute cluster for data processing and a high speed dedicated network infrastructure, Figure 1. Access protocols, like GridFTP and the data management system iRods [14], are provided by dedicated servers attached through GPFS to the storage systems. For internal use CIFS and NFS access are offered for unstructured data.

An 10 Gigabit Ethernet network is spanned between all servers and the internet. Through dedicated high–speed network links the LSDF is connected to the research network and data sources, like ANKA inside KIT, allowing a rapid data taking. It enables scientific communities to work with and share the data in collaborations of global scale.

Based on the described infrastructure, several high level services for structured data are provided. The KIT Data Manager KDM consists of data and meta data management and repository components especially designed for large scale scientific data. The KDM DataBrowser [15] allows meta/data ingest and access using a graphical user interface that can be easily adapted for special needs. Additional services for organizing the data life cycle and data support are offered to the user communities. Open interfaces and APIs support a variety of access methods to the highly available services for high throughput data applications.

To process the experimental data a compute cluster with 60 nodes equipped with Hadoop [16], an implementation of Google's MapReduce [17] programming paradigm is directly attached to the storage. This allows for data intensive computing with excellent scalability, but requires adapted applications which run in a predefined environment. LAMBDA [18] – the LSDF Execution Framework for Data Intensive Applications — simplifies large scale data processing for scientific users by reducing complexity, responsibility and error–proneness. The description of an execution is realized as part of LAMBDA administration in the background via meta data and can be steered by the KDM DataBrowser.

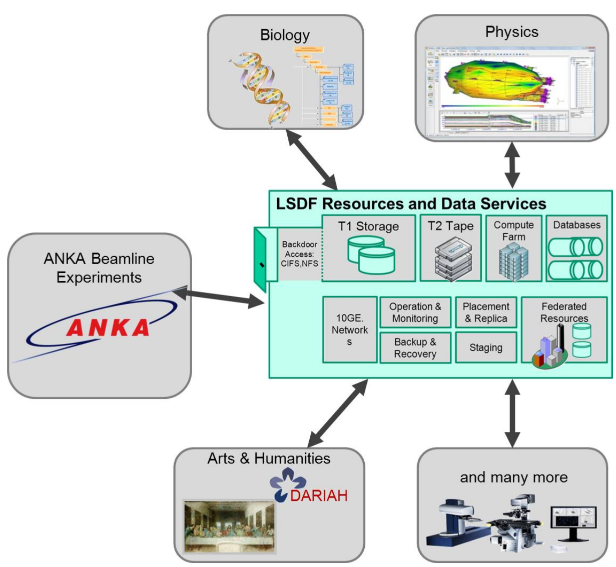


Figure 1: The Large Scale Data Facility and connected scientific communities.

If processing of experimental data requires a dedicated software environment, (for instance Windows, or specific software versions or licences) the OpenNebula [19] Cloud environment is foreseen.

In close cooperation with the user communitites, the LSDF provides assistance to efficiently organize data and meta data structures, and develops and deploys community specific software running on the directly connected computing infrastructure.

## COMPONENTS OF A NOVEL DATA MANAGEMENT SYSTEM

Figure 2 depicts the conceptual data flow within a imaging beamline (left) to the storing and processing components within the LSDF. All beamline components are steered by a Control System with various user interfaces (UI). Control & monitoring data are exchanged with the components as well as user information administrated by the "User Office" for e.g. authentification and authorization.

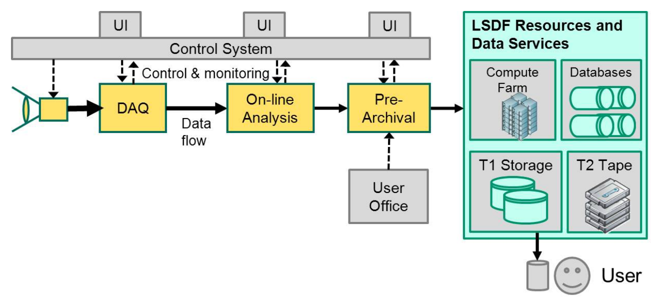


Figure 2: Data flow from the camera to the LSDF.

Starting from the detector or camera, a massive flow of data is piped to the data acquisition system "DAQ". In "DAQ" relevant measurement data are selected and annotated for further processing and storage.

"On–line Analysis" processes the data in various preprocessing steps, reconstructs tomographic images and volumes using a GPU–accelerated system and visualizes the results. Based on the visualization, the user will select promising data sets for further processing and archiving.

In the "Pre–Archival" step the date is prepared by adding meta-data for the storage, archival and processing in the Large Scale Data Facility LSDF. Additional meta data about the experiment and user data will be added from the "User Office".

Within the LSDF the data will be stored, processed and archived. The data ingest process is mostly standardized, including data transfer and meta data storage, via APIs or the KDM DataBrowser. Since the LSDF has also provided computing resources, a variety of sophisticated and computationally expensive processing algorithms can be applied automatically. For tomography beamlines algebraic reconstruction will provide high–quality images and volumes for further scientific analysis. The results will be enriched with additional meta data for data provenance and stored with the corresponding measurement data sets. Users will access their data sets on the LSDF directly via the KDM DataBrowser or a Web interface. External users will be provided with a temporary user account, according to ANKA's policies.

It is planned to use NeXus as a primary file format for all scientific data sets written to persistent storage. The CDM technologies [20] will provide an abstract interface for accessing NeXus.

The whole architecture is designed to fulfill the requirements for continuously adapted experiments and enhanced hardware components.

## CONCLUSION

The ANKA Synchrotron facility is part of KIT, a world-leading engineering institution with top ranking Faculties of Informatics, Electrical Engineering and Information Technology, and Mechanical Engineering. KIT supports a broad research palette ranging from astro–particle physics over biology to supercomputing with superior infrastructures. The modern data storage and processing infrastructures enable cutting edge research with high measurement data rates. Because the infrastructure is shared among several institutes investments are also shared. This takes some pressure off the financial budgets of ANKA and the beamline experiments.

Very economically the large scale data facility offers nearly unlimited storage, cluster computing, high throughput networks, transparent yet secure access services, and is still easily adapted to the changing requirements of the users. The unique possibilities of this facility for data intensive computing opens a new dimension for data intensive beam-ine experiments and subsequent data analysis.

## ACKNOWLEDGEMENTS

Many thanks to the fruitful discussions within the High Data Rate Initiative HDRI of the German Helmholtz Association research program PNI, and with Alain Buteau and his crew from SOLEIL. Special thanks Francesca Rindone for revising and improving the texts.

## REFERENCES

[1] Y.L. Mathis, B. Gasharova, and D. Moss. Terahertz radiation at anka, the new synchrotron light source in karlsruhe. *Journal of Biological Physics*, 29(2):313–318, 2003.

[2] S. Chilingaryan, A. Mirone, A. Hammersley, C. Ferrero, L. Helfen, A. Kopmann, T. dos Santos Rolo, and P. Vagovic. A gpu-based architecture for real-time data assessment at synchrotron experiments. *Nuclear Science, IEEE Transactions on*, (99):1–1.

[3] A. Rack, T. Weitkamp, S. Bauer Trabelsi, P. Modregger, A. Cecilia, T. dos Santos Rolo, T. Rack, D. Haas, R. Simon, R. Heldele, et al. The micro-imaging station of the topotomo beamline at the anka synchrotron light source. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 267(11):1978–1988, 2009.

[4] J-M. Chaize, A. Goetz, W-D. Klotz, J. Meyer, M. Perez, and E. Taurel. TANGO — an object oriented control system based on corba. In *International Conference on Accelerator and Large Experimental Physics Control Systems*, pages 475 – 479, 1999.

[5] T. Spangenberg, K. Cerff, and W. Mexner. Macro package based enhancement of SPEC controlled experimental setups. In *PCaPAC*, 2010.

[6] SPEC, http://www.certif.com/.

[7] T. Spangenberg, K. Cerff, W. Mexner, and W. Kaiser. MTANGO integration of a SIMATIC WinCC open architecture SCADA system at ANKA. In *ICALEPCS*, 2011.

[8] C. Reinhart. Industrial computer tomography — a universal inspection tool. In *17th World Conference on Nondestructive Testing*, 2008.

[9] S. Esenov, K. Wrona, and C. Youngman. European xfel daq and dm computing technical design report. Technical report, European XFEL facility, 2009.

[10] T. M. Frazier P. Adams, S. G. Azevedo, R. G. Beeler, R. C. Bettenhausen, R. W. Carey, C. B. Foxworthy, M. Hutton, L. J. Lagin, and S. L. Townsend. The national ignition facility data repository. In *Proceedings of ICALEPCS2009, Kobe, Japan*, 2009.

[11] S. G. Azevedo, R. G. Beeler, R. C. Bettenhausen, E. J. Bond, P. W. Edwards, S. M. Glenn, J. A. Liebman, J. D. Tappero, and W. H. Williams A. L. Warrick. Experimental data storage management in nexus format at synchrotron soleil. In *Proceedings of ICALEPCS2009, Kobe, Japan*, pages 75–77, 2009.

[12] Nexus, http://www.nexusformat.org.

[13] H. Marten and T. Koenig. ITIL and grid services at gridka. *Journal of Physics: Conference Series*, 219(6):062018, 2010.

[14] M. Hedges, A. Hasan, and T. Blanke. Management and preservation of research data with irods. In *Proceedings of the ACM first workshop on CyberInfrastructure: information management in eScience*, pages 17–22. ACM, 2007.

[15] R. Stotzka, V. Hartmann, T. Jejkal, M. Sutter, J. van Wezel, M. Hardt, A. Garcia, R. Kupsch, and S. Bourov. Perspective of the large scale data facility (lsdf) supporting nuclear fusion applications. In *Parallel, Distributed and Network-Based Processing (PDP), 2011 19th Euromicro International Conference on*, pages 373–379. IEEE, 2011.

[16] The Apache Software Foundation. Welcome to Apache Hadoop!, July 2010.

[17] J. Dean and S. Ghemawat. Mapreduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

[18] T. Jejkal, J. Otte, A. Garcia, V. Hartmann, R. Stotzka, J. van Wezel, and A. Streit. Lambda – the lsdf execution framework for data intensive applications. In *Parallel, Distributed and Network-Based Processing (PDP), 20th Euromicro International Conference on*. IEEE, 2012. submitted.

[19] B. Sotomayor, R.S. Montero, I.M. Llorente, and I. Foster. Virtual infrastructure management in private and hybrid clouds. *Internet Computing, IEEE*, 13(5):14–22, 2009.

[20] Common data model, http://www.docstoc.com/docs/72729634/soleilcommondatamodelapi_1_.