Grid Based Distributed Remote Computing at DØ

T. Nunnemann, O. Biebel, M. Binder, T. Christiansen, J. Elmsheuser, M. Erlebach, F. Fiedler, B. Leonhardt, D. Schaile, P. Schieferdecker, and R. Ströhmer

The enormous amount of experimental data collected with the DØ detector at Fermilab and the international composition of the DØ collaboration, both demand a global distribution of its computing resources. The simulation of high-energetic $p\bar{p}$ collisions and the processing of the experimental data require massive CPU resources which cannot be provided centrally at the experiment in a sufficient amount. On the other hand, distributed remote computing enables collaborating physicist to perform data analysis with their local or regional computing resources.

We integrated the national high-energy physics computing center GridKa [1] as a regional center within DØ's computing environment, which has been demonstrated by its contribution to the processing of experimental data. In addition experiment specific software and Grid middle-ware have been deployed at LMU and LRZ [2] to enable local development and smaller scale data analysis.

We have established a data-grid within DØ based on the data management and distribution system SAM (*Sequential Access through Meta-data* [3]), which incorporates a replica catalog, an optimization of input file delivery and book-keeping for user jobs. Based on meta-data which describe the content of stored files a user can define and request datasets. SAM translates a data set into a list of files which then can be automatically be imported from the most convenient (i.e. usually closest) location of a copy as defined in the replica catalog. The file transfer itself is done via GridFtp based on Globus authentication [4]. On average an effective transfer rate of about 200 GB per day could be sustained between Fermilab and GridKa.

In addition to event data which are stored in data files deliverable through SAM, calibration and configuration constants are queried from data base servers. Since the compute nodes at GridKa (as well as many other sites) are only connected to local networks without access to and from outside networks, the central data base servers at Fermilab cannot be queried directly but data base proxy servers are deployed on a local head node. Data base queries from running applications are directed to local proxy servers which themselves communicate with the central servers and cache retrieved data for subsequent queries. The communication between the application and the multiple levels of data base servers is done using CORBA [5]. The local caching of calibration constants also significantly increases both reliability through minimized dependence on network communication, as well as performance, e.g. the time to obtain a calibration set decreases by approximately a factor 15 compared to a remote query.

DØ's code distribution is regularly been updated and adopted at GridKa to offer the same functionalities for production and user analysis computing applications as on the central Fermilab systems.

We have used GridKa to reprocess a significant fraction of DØ's data, which contributed to the production of a homogeneous data set with improved efficiencies and resolution and which corresponds to an integrated luminosity of approximately 200 pb^{-1} in total.

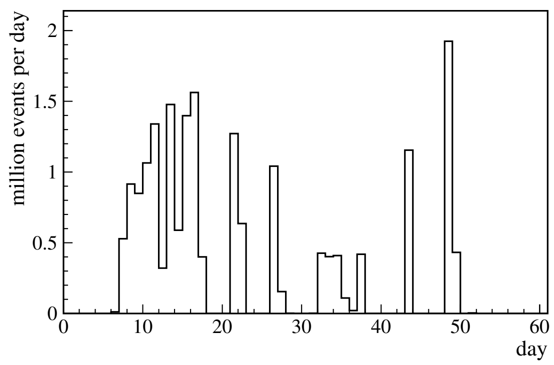

Fig. 1: DØ data reprocessing at GridKa (daily rate).

About 21 million events have been processed over a period of about 7 weeks at GridKa, which corresponds to about 11,000 compute jobs with approximately 5 h processing time each (CPUs correspond to 2.3 GHz Xeon on average). During stable system conditions a daily production rate of more than one million events could be achieved (cf. Fig. 1).

In total, 5 TB of data processed with a previous version of the reconstruction program was imported to GridKa and dynamically stored using SAM. Job submission was performed semi-automatically based on a work-flow planner [6] developed for the DØ and CMS experiments and on SAM providing dynamic file delivery. A script-based book-keeping system was set up which kept track of succeeding and failing production jobs and which defined sets of SAM data sets to be resubmitted.

The data production demanded an exceptional performance of the file-servers. It could be demonstrated that IBM's GPFS file-system [7], which has been recently deployed at GridKa, was able to serve the in- and out-put of more than 300 parallel running production jobs without apparent limitations.

Besides data processing, GridKa is extensively used by the German groups participating in DØ for Monte Carlo production and for computing intensive data analysis. Based on our developments and experiences made during the data production phase, we are planning the next large reprocessing in summer 2004.

References

[1] Grid Computing Center Karlsruhe (GridKa), `http://www.gridka.de`.

[2] Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, `http://www.lrz-muenchen.de`.

[3] A. Baranovski *et al.*, FERMILAB-TM-2175.

[4] Globus Alliance, `http://www.globus.org`.

[5] Common Object Request Broker Architecture (CORBA), `http://www.corba.org`.

[6] G. E. Graham, D. Evans, and I. Bertram, eConf **C0303241** (2003) TUCT007 [arXiv:cs.dc/0305063].

[7] F. Schmuck and R. Haskin, Proceedings of the Conference on File and Storage Technologies, Monterey, 1002.