

Data Reprocessing on Worldwide Distributed Systems

Mike Diesburg* and Daniel Wicke[‡]
FNAL, Batavia, IL 60510, USA
for the DØ-Collaboration

Abstract

The DØ experiment faces many challenges enabling access to large datasets for physicists on 4 continents. The strategy followed is to make use of worldwide distributed computing clusters.

Since the begin of Tevatron Run-II (March 2001) all Monte-Carlo simulations are produced outside of Fermilab on remote systems. For analyses a system of regional analysis centres (RACs) was established which supply the associated institutes with the data. This structure which is similar to the Tier structure foreseen for LHC was used in autumn 2003 to reprocess all DØ-data with the up-to-date and much improved reconstruction software.

With these achievements DØ is the first high energy physics experiment to have implemented and operated all essential computing tasks on non-dedicated, worldwide distributed systems. The experiences gained in DØ can be applied to evaluate the LHC computing model.

INTRODUCTION

The improved understanding of the DØ detector lead to significant improvements of the physics reconstruction software. Its procedures and parameters are no longer based on design specification, but on the actual detector layout and its performance.

To benefit from these new developments 100 pb⁻¹ of data processed with previous software releases had to be reprocessed with this new reconstruction software. As the on-site resources weren't sufficient to perform the task, computing resources of 7 institutes from 6 countries were included in the effort to perform the first HEP data reprocessing on distributed, non-dedicated systems [1, 2, 3, 4].

This effort built on the experiences DØ made in remote MC production and automatic on-demand data distribution for remote analysis. The former has been part of the DØ computing model since the beginning of Run-II [5]. The latter was established in 2002 by implementing a hierarchical structure for distributing data from the central repository through regional analysis centres to the individual institutes [6].

In the following first the reprocessing task is described from the computing perspective. Then the resources available to DØ are recorded before the work-flow of the operation is detailed. Finally the problems observed during the effort are discussed and put into perspective to LHC efforts.

THE COMPUTING TASK

The DØ reconstruction code aims to reconstruct physics objects such as tracks, calorimeter clusters or vertices from the recorded digitised output of the detector, the so called RAW data. Besides this RAW data, reconstruction (in principle) needs access to calibration constants which are held in a calibration database. The result of this reconstruction is stored in two formats. The DST keeps most of the RAW information for detailed studies, while the TMB is a condensed format aimed at physics analyses.

In addition to the actual results, metadata describing each of these files are created. These contain information about the data format, the release version of the software with which they were created as well as the corresponding input files. This work-flow is shown in the left branch of Fig. 1.

The TMBs created from each of the input files are too small to efficiently access them from the tape system. Therefore around 10 of them need to be merged into a single file. This complicates the work-flow.

At the time, when the new software was deployed to the processing farm, about 300M events were already recorded by DØ and reconstructed with older software versions. These sum to a total size of 75TB of RAW (input) data and an expected size of 45TB of reconstructed output data (DSTs and TMBs).

Given a reconstruction speed of around 50s/Event the estimated need for CPU time was 4 million hours or 5800 months on a 1GHz PIII machine. With a desired completion time of 3 months nearly 2000 CPUs were required.

As the DØ reconstruction farm is dimensioned to keep pace with data-taking it could only be used during the fore-

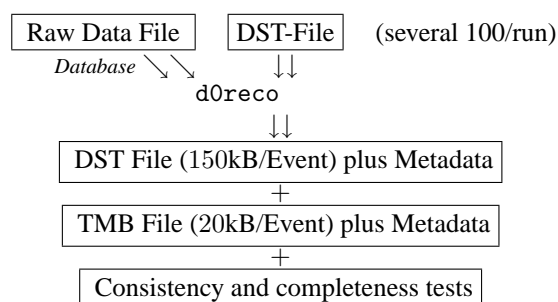


Figure 1: Application flow of a reconstruction job. Depending on the type of input data, database access is required.

* diesburg@fnal.gov

[‡] wicke@fnal.gov, on leave from Bergische Universität Wuppertal

seen shutdown of 8 weeks. With its size equivalent to 1000 1GHz PIII processors this farm was expected to handle only around one third of the events.

Thus other resources had to be used to achieve a timely completion of the task.

AVAILABLE RESOURCES

Besides the central farm DØ has access to additional resource through the participating institutes. The sites finally participating in the effort are listed in Tab. 1. With these centres the computing power available to the reprocessing effort could be more than doubled, however, most of these resources are shared with other experiments. Coordination with their projects was therefore required.

In general remote clusters have a reduced network bandwidth and an increased IP round trip time to the central systems serving the databases and input files. In addition each of these computing clusters has a different hardware and software configuration. To exploit these resources the procedures and the software needed to be setup to suit the reduced network connectivity and the many different cluster setups.

In first tests the reduced network connectivity appeared to be a catastrophic performance issue. Processing individual events would sometimes take up to an hour instead of minutes. The problem arose from huge number of database accesses done for each event which are used to access the calibration constants. A medium to long term solution to this problem is to install database proxies at each site which cache the information so that only the very first event of a processing run would be affected.

For this round of reprocessing, however, DØ opted not to use database access at all. This was possible because the improved tracking didn't rely on such constants while calorimeter constants hadn't changed much since the initial processing.

The differences between the clusters setups occur on different levels: most importantly some clusters use a common file space (e.g. nfs distributed) for all worker nodes, while others have a more localised setup. The data distribution software SAM [7] used in DØ can be configured to work in all these situations, however, the optimal way of

Centre	Size				
GridKa (Karlsruhe, Germany)	300				
IN2P3 (Lyon, France)	220				
Nikhef (Amsterdam, Netherlands)	<table border="0"> <tr> <td>{ local</td> <td>80</td> </tr> <tr> <td>{ EDG</td> <td>320</td> </tr> </table>	{ local	80	{ EDG	320
{ local	80				
{ EDG	320				
SAR (Texas, USA)	130				
UK (London, Manchester, Rutherford)	340				
WestGrid (Canada)	300				

Table 1: Computing centres participating. The computer power estimated to be available to DØ is given for each site in numbers of CPUs equivalent to a 1GHz Pentium-III processor.

using it from within the jobs may vary. Therefore also the Job Management software Runjob [8, 9] needed manual site by site configuration.

A special challenge came in for the resources that didn't allow for any DØ-specific installation, i.e. the EDG resources entered through Nikhef. The effort to allow running the reconstruction software on such naked environments is described in another contribution [10]. As this effort was only successful shortly before the effort finished we will concentrate on DØ-aware operation in the following.

OPERATION

Following the requirements imposed on the software by the available resources the plan for operation was tuned not to use database access. Due to limited person-power the merging of TMB results was done centrally, where corresponding scripts including sufficient cross-checking to avoid duplication existed.

When not using database constants the input to the reconstruction are the previously reconstructed (old) DSTs, from where the previously used constants could be taken. Besides avoiding the DB access this reduced the size of input data that needed to be transported over wide area networks by 50%.

To distribute the inputs to the various sites, SAM datasets were created manually, assigned to the sites and published on a web page. Each dataset contained at most one "physics run". The actual number of files per dataset was adapted to the size of the various sites such that it was appropriate to be processed in parallel. This had to take into account the actual number of available CPUs as well as the available disk storage for the output files.

An operator at each site would start the productions for the assigned datasets usually one after another. Most sites decided to access the files of these datasets before the actual processing jobs were started. This effectively pre-stages the files into the sites SAM cache and was reported to be less error prone.

Once the jobs that process a given dataset finished, the output TMBs created, the corresponding metadata and a defined set of log-files were copied to a central location on `d0mino.fnal.gov`. The script responsible for the transport was based on the `sam_cp` interface provided by SAM and provided rudimentary file based book-keeping that avoids repeated transport and indicates successful completion to the central site. Depending on the configuration of SAM `sam_cp` uses GridFTP, `bbftp` or other methods for the transport.

After automatic checking of the log-files and metadata the TMBs were then merged and stored into SAM at the central site. From this point on SAM could be queried for results missing from a input dataset that was believed to be finalised.

The larger DST output was for most of the sites not transported back to the Fermilab. Instead each site stored them

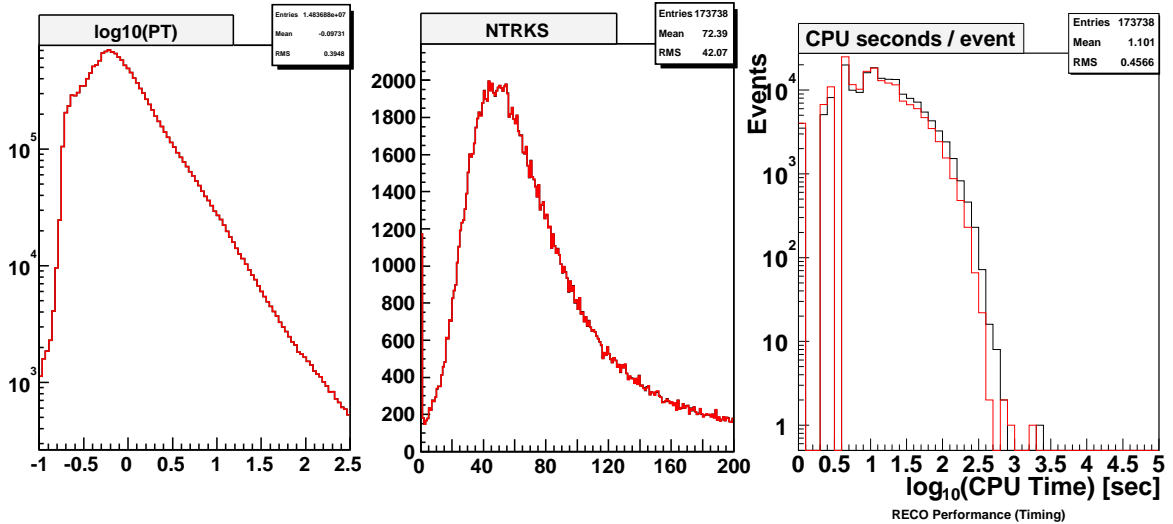


Figure 2: Comparison of results produced in Lyon and Manchester show differences only in a large magnification or for variables which depend on the speed of the CPU.

into a SAM location local to the corresponding site. We rely thereby on SAM to bring them to the analyser in the future on request.

Beside the manual job submission at individual sites the 3 british sites were operated from a single entry point using the Globus [11] and Condor [12] based tool JIM [13, 14, 15] to distribute the jobs among the 3 sites.

CERTIFICATION

Before a site was allowed to join the official production running, it had to be certified. The certification process aimed to cross-check the correct setup at each site and to spot any numerical deviations due to varying operating system versions or configuration problems.

To get certified each site processed one well defined dataset of 66 files and stored the results back to the central location in the defined way. From these results a large number of plots were computed and compared between the sites as well as to a production done from RAW.

The comparison to a database enabled production from RAW was done to verify the special procedure of reprocessing from DSTs without database access. These plots are expected to show some differences and were judged by reconstruction experts.

Also between different sites both processing from DSTs tiny differences were observed on results which in principle should be identical. These were traced to arise from differences in the precision of the FPU's on Intel Pentium and AMD Athlon processors, which were considered to be insignificant for physics results. Fig. 2 shows example plots comparing results produced in Lyon with those done in Manchester.

OBSERVED PROBLEMS

In projects like this reprocessing, which are likely to be repeated in the future, an analysis of the observed problems is as important as the actual results are for physics analysis. Problems fall into two big categories: Administrative problems and technical problems.

Organisational problems

Organisational problems occurred mainly during the planning phase of the undertaking. The release of a software version not requiring DB access was very much delayed by waiting for physics related bug fixes. This in turn delayed the tests of running remotely. Some implicit dependencies on the Fermilab environment were therefore found late.

These delays caused discussions on how to keep a given date to finalise the production. At an intermediate stage reprocessing was foreseen to be performed on selected events only, which made the development of additional code necessary. In parallel code optimisation was performed, which finally improved the speed sufficiently to stick to the original plan. These changes between different modes of operation caused additional delays.

Planning the reprocessing effort should have taken the flexibility of such code which is in constant development into account and avoided large changes in philosophy.

Technical problems

Technical problems occurred in all major components used for the reprocessing.

Data distribution started extremely slowly so that the remote farms could process events faster than they could receive input data. In contrast to first naïve ideas the reason wasn't the network bandwidth. Instead other applica-

tions were heavily accessing central tape drives. Setting up an additional router for reprocessing, allowing to prefetch more files, and increased priority solved the problem without banishing the other application.

The largest problem occurred during running the actual jobs. On large clusters technical problems due to software or hardware failures are quite common and usually followed by job crashes. GridKa observed failure rates of up to 20%. To avoid production duplication, completed results should be kept, while crashed jobs had to be restarted. At this intermediate state SAM couldn't be used to query for which jobs a restart was necessary.

Each of the participating sites developed their own set of scripts or databases to follow the production in this detailed level and ease the recovery of such problems.

Finally the transport back to the central file-server showed corrupted files at a rate of 1 in 1000, which were spotted by the sanity check programs run at the merging step due to inconsistent data-structures.

CONCLUSIONS AND OUTLOOK

DØ performed the first HEP-data reconstruction on globally distributed, non-dedicated clusters. 300 million events or 45TB of input data were processed at 7 centres. Around 30% or 15TB were processed remote of Fermilab using Grid concepts for data- and job-distribution as they are foreseen for the LHC Computing Grid (LCG).

The successful completion of the DØ reprocessing has proven these concepts to be functional, but also revealed problems that need to be addressed for a fully automated production. The most important of these being the automatic detection and recovery from job failures which don't have their cause in the program itself, but arise from a temporary failure in the environment.

With the experience gained in the first round of reprocessing, DØ foresees a second round of reprocessing starting around the turn of the year. As the physics improvements this time mainly contain an improved understanding of the calorimeter and its new electronics, access to the calibration database is required. Full application of SAMGrid throughout all steps of reprocessing will allow to bring merging to the remote sites and to implement a coherent book-keeping to recover from externally triggered job crashes.

REFERENCES

- [1] Daniel Wicke and Mike Diesburg.
<http://www-d0.fnal.gov/computing/reprocessing/>, 2004.
- [2] K. Riesselmann. *Ferminews* **27**, 2(2004) 2.
<http://www.fnal.gov/pub/ferminews/ferminews04-02-01/p1.html>.
- [3] Daniel Wicke. Proceedings to DPF2004, Riverside, CA; to be published in IJMPA, 2004.
- [4] K. Riesselmann. *CERN Courier* **44**, 7(2004) 16.
<http://www.cerncourier.com/main/article/44/7/15>.
- [5] G. Garzoglio et al. Contribution #38 in *these proceedings*, 2004.
- [6] I. Bertram et al. DØ-Note 3984, 2002.
- [7] Mike Diesburg et al. DØ-Note 3464, 1998.
- [8] G. Graham and D. Evans. *CHEP Proceedings* (2001) .
- [9] G. Graham, D. Evans, and I. Bertram. *CHEP Proceedings* (2003) . ArXiv:cs.dc/0305063.
- [10] T. Harenberg et al. Contribution #441 in *these proceedings*, 2004.
- [11] I. Foster and C. Kesselman. *Intl J. Supercomputer Applications* **11**(2)(1997) 115–128.
- [12] D.H.J. Epema, M. Livny, R. van Dantzig, X. Evers, and J. Pruyne. *Future Generation Computer Systems* **12**(1996) 53–65.
- [13] A. Baranovski et al. *Nucl. Phys. Proc. Suppl.* **120**(2003) 119–125.
- [14] A. Baranovski et al. *Nucl. Instrum. Meth.* **A502**(2003) 423–425.
- [15] A. Baranovski et al. (2003) . ArXiv:cs.dc/0307007.