

# DZero

## BREAKS

## NEW

## GROUND

## in global computing efforts

### First steps toward Grid application with 'real data'

#### ON THE WEB:

Reprocessing of  
DZero Run II data:  
[www-d0.fnal.gov/computing/reprocessing/](http://www-d0.fnal.gov/computing/reprocessing/)

by Kurt Riesselmann

**S**earching for subatomic particles very much resembles the often-cited search for the needle in the haystack. Since the beginning of Collider Run II in March 2001, DZero scientists have collected more than 550 million particle collisions. The data fill five stacks of CDs as high as the Eiffel tower—storage cases not included. And the (hay)stacks are growing every day.

"The Fermilab farms can process four million events per day," said Mike Diesburg, who manages a cluster of 600 PCs for the DZero experiment at Fermilab. "That's enough to handle the daily flow of incoming events."

Yet when the DZero collaboration decided to re-examine the entire set of collision data, encompassing more than 500 terabytes, scientists had to look for computing power beyond Fermilab. For the first time ever, DZero scientists had to send actual collision data—the crown jewels of their experiment—off site.

"In the past, DZero and other particle physics collaborations have used remote computing sites to carry out Monte Carlo simulations of their experiments," said DZero scientist Daniel Wicke, University of Wuppertal, Germany. "We are now one of the first experiments to process real collision data at remote sites. The effort has opened up many new computing resources for our collaboration. The evaluation of our experience will provide valuable input to the worldwide development of computer grids."

The reprocessing of the DZero collision data, coordinated by Diesburg and Wicke, so far involves computing resources in six countries: Canada, France, Germany, the Netherlands, the United Kingdom and the United States. (Many other countries contribute to the computing of simulated DZero data and the analysis of processed data.) From November to January, DZero groups in each of the six countries had access to local PC clusters and Grid networks, ranging from one hundred to more than one thousand PCs.

"In the UK, the software installation, submission and monitoring of jobs was done centrally for all participating UK sites in a grid-like manner," said Gavin Davies at Imperial College London. "The machines at Imperial College, for example, are shared across the whole College, so it takes grid software to keep it all running smoothly."

The largest amount of off-site computing took place at the Centre de Calcul in Lyon, France, which reprocessed 36 million collisions.

"Reprocessing involves large volumes of data to be transferred in both directions on a scale that was simply unthinkable a few years ago," said Patrice Lebrun, IPN Lyon. "It will open new possibilities that we are only beginning to explore."

To provide participating computer systems with collision data, the DZero collaboration relied on the SAM software developed at Fermilab. The Sequential Access Manager is essentially a catalog of all the DZero data, and it transfers data on demand. Wyatt Merritt, who is a co-leader of the SAMGrid project at Fermilab, explained the process.

"If a DZero scientist submits a job to the computer system in Karlsruhe, Germany, it may need a particular set of data files," she said. "If those files are not in the local system, the SAM software will automatically determine where they are and retrieve them. With the SAM software, a user doesn't need to know whether the data is stored on tape or on disk, whether it is located at Fermilab or at Karlsruhe."

#### Western Canada Research Grid (WestGrid):

- 3,000 processors\* in beta test shared by DZero, a chemistry group and a second subatomic physics group.
- Processors\* used by DZero: 1,000
- DZero events reprocessed: 12 million
- URL: [www.westgrid.ca](http://www.westgrid.ca)
- Funding: Canada Foundation for Innovation; Natural Sciences and Engineering Research Council

#### Feynman Computing Center at Fermilab:

- About 1,000 processors\* strictly dedicated to DZero experiment
- Events reprocessed: 400 million
- URL: [www-d0.fnal.gov/computing/reprocessing/](http://www-d0.fnal.gov/computing/reprocessing/)
- Funding: U.S. Department of Energy

#### Centre de Calcul de l'IN2P3:

- 1,070 processors\* used for particle and nuclear physics, astrophysics, and biology
- Processors\* used by DZero: 160
- DZero events reprocessed: 36 million
- URL: [lyinfo.in2p3.fr/d0/](http://lyinfo.in2p3.fr/d0/)
- Funding: Institut National de Physique Nucleaire et de Physique des Particules (IN2P3)

#### Grid for UK Particle Physics (GridPP)

- Imperial College London, Manchester University and Rutherford Appleton Laboratory provide more than 550 processors\*
- Processors\* used by DZero: 270
- Events reprocessed: 23 million
- URL: [www.gridpp.ac.uk/dzero/](http://www.gridpp.ac.uk/dzero/)
- Funding: Particle Physics and Astronomy Research Council, and other organizations

#### NIKHEF DataGrid:

- Local installation of the LHC Computing Grid with 500 processors\*, of which 100 are always reserved for DZero
- Processors\* used by DZero: 400
- DZero events reprocessed: 7 million
- URL: [www.dutchgrid.nl/Org/Nikhef/](http://www.dutchgrid.nl/Org/Nikhef/)
- Funding: Foundation for Fundamental Research on Matter

#### Grid Computing Centre Karlsruhe (GridKa):

- Forschungszentrum Karlsruhe provides 900 processors\* for several particle physics experiments
- Processors\* used by DZero: 200
- Events reprocessed: 21 million
- URL: [www.gridka.de/D0/](http://www.gridka.de/D0/)
- Funding: German Federal Ministry of Education and Research

\*computing power equivalent to 1 GHz Pentium III processors

Although the DZero collaboration has automated the global tracking and transfer of data, the reprocessing of data does not yet represent a full, global Grid. So far, DZero scientists manually assign computing jobs to specific clusters and local grids. However, scientists at the NIKHEF laboratory in Amsterdam made great progress.

"We have been able to show that we can really use the LHC [Large Hadron Collider] Computing Grid for DZero processing," said Kors Bos, who leads the Dutch computing efforts. "We saw jobs submitted from Wuppertal being executed on our CPUs, and we executed jobs in Karlsruhe, at Rutherford Appleton Laboratory and a few more places."

Wuppertal's Wicke praised these efforts.

"The group at NIKHEF has pushed the Grid concept the most," he said. "They have devoted themselves to running DZero computing jobs on generic computers that have no prior knowledge of DZero programs and data bases. When their efforts pay off, then we can run our DZero jobs on any computer cluster in the world."

The DZero collaboration conducted the reprocessing of all Run II data to improve, among other things, the identification of particle tracks. Raw data contain track information in the form of a vast collection of disconnected points. To connect the right dots, scientists use sophisticated track reconstruction programs. Until recently these

programs relied on the theoretical design of the DZero detector rather than its real-world performance.

"The new algorithm is based on our knowledge of how well we put the detector together," said Dugan O'Neil, one of the DZero scientists working with the WestGrid in Vancouver, Canada. "This has dramatically improved our efficiency of finding particle tracks."

The collaboration also has adopted the new algorithm to process all new experimental data. Yet the collaboration expects to carry out another reprocessing of all Run II data, old and new, in less than a year, applying further refined analysis tools to the raw data. The new round of reprocessing will require even more off-site computing power, providing ample of opportunity to further develop the Grid system.

"You can't make the Grid work without motivation," said O'Neil. "It's one thing to have a vision, and it is another thing to stay up to three in the morning to make things work because they need to get done. DZero is a real application. We need to get the physics results out." ☛

# DZero



SFU campus on Burnaby Mountain, Vancouver



*"You can't make the Grid work without motivation. It's one thing to have a vision, and it is another thing to stay up to three in the morning to make things work because they need to get done. DZero is a real application. We need to get the physics results out."*

– Dugan O'Neil, Simon Fraser University, Canada



Wuppertal's landmark, the suspension railway



*"In the past, particle physics collaborations have used remote computing sites to carry out Monte Carlo simulations. We are now one of the first experiments to process real data at remote sites. The effort has opened up many new computing resources. The evaluation of our experience will provide valuable input to the Grid development."*

– Daniel Wicke, University of Wuppertal, Germany



Tower Bridge, London



Street scene in Lyon



*"With the SAM software developed by the Fermilab Computing Division and DZero, a user doesn't know whether the data is stored on tape or on disk, whether it is located at Fermilab or at Karlsruhe."*

– Wyatt Merritt (left), with Mike Diesburg and Amber Boehnlein, Fermilab, U.S.A.



Chicago skyline



*"The machines at Imperial College, for example, are shared across the whole college, so it takes grid software to keep it all running smoothly."*

– Gavin Davies, Imperial College London, UK

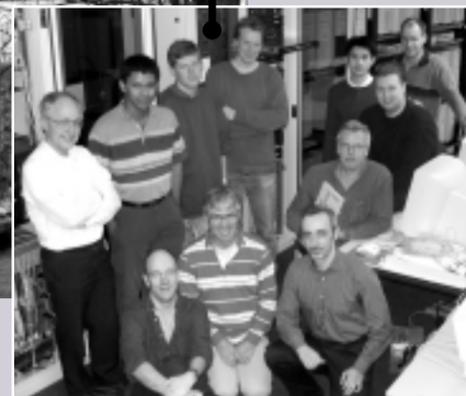


Amsterdam, famous for its canals



*"We've participated in large-scale Monte Carlo production in the past, but data reprocessing involves large volumes of data to be transferred in both directions on a scale that was simply unthinkable a few years ago. It will open new possibilities that we are only beginning to explore."*

– Patrice Lebrun (right), with Tibor Kurca, CCIN2P3, Lyon, France



*"The re-processing was a major milestone for DZero. For us it is also important that we have been able to show that we can really use the LHC Computing Grid for DZero processing. We saw jobs submitted from Wuppertal being executed on our CPUs, and we executed jobs in Karlsruhe, at Rutherford Appleton Laboratory and a few more places."*

– Kors Bos (front row, second from left) and the Scientific Computing team at NIKHEF, Amsterdam, Netherlands