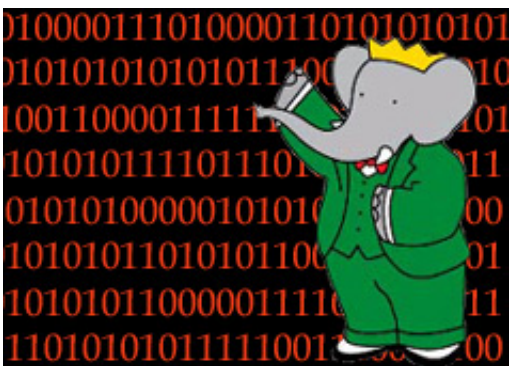


BaBar Collaboration Completes Data Reprocessing

December 18 2008, By Kelen Tuttle



(PhysOrg.com) -- One might think that processing the records of 22 billion electron and positron collisions once would be enough. But not so for the BaBar collaboration, which this week announced the completion of reprocessing for 99.99 percent of its huge coffers of Upsilon(4S) raw data.

Processing is one of the very first steps in data analysis, and involves putting raw data into a more useful form. This requires taking the signal recorded by BaBar's many layers of detectors and reconstructing which types of particles left the signals, while traveling in what directions and at what speeds. These reconstructed data are then compared to simulated data to identify particularly interesting events, and divided into many different streams from which researchers can pluck event types of

interest.

Over the years, the collaboration has again and again reworked the method and programs it uses to process data. By reprocessing the entire dataset with the newest software, the collaboration has now created a standardized dataset across the experiment's eight years of data collection.

"This was a huge effort undertaken by many people," said BaBar Computing Coordinator Homer Neal. "It takes a lot of work to do something like this, but it's worthwhile to create such a uniform and deep dataset."

The reprocessing project began in 2007, when the collaboration decided to invest the time and effort to produce the best software possible for the final phase of data-taking. "And from there, the argument was easy for reprocessing everything with that same software," said Emeritus BaBar Computing Coordinator Gregory Dubois-Felsmann.

The first step was to write the new reconstruction software—no easy task. Taking the signals from the detector and working backward to figure out what actually happened is an extremely complex process. When you make improvements in one area of the software, Dubois-Felsmann said, there is always the chance that you have worsened some other aspect accidentally. Nonetheless, through multiple iterations and by checking the software against large amounts of data, researchers validated the new software last spring.

"We were slowed down a bit by the bad budget news and the decision to take the last few months of data at a lower energy," said Dubois-Felsmann. "We needed to rewrite the software for this lower energy as well, so we didn't finish until about a month later than originally planned."

Even with this late start, the collaboration finished reconstructing BaBar's eight years of data with the new software ahead of schedule. The success, Neal and Dubois-Felsmann agreed, is a result of hard work and the ability to expand computing resources both at SLAC and at the Padova computing center in Italy, where much of the reprocessing took place. "Both SLAC and Padova were wonderfully supportive and made this happen," said Neal.

In addition to improving the event reconstruction, researchers also made improvements to two other areas of the production process: simulation and data skimming.

Although it seems slightly counterintuitive at first, simulated data are integral to the analysis of real data. That's because the only way to understand the output of BaBar's detectors is to simulate the many different types of collisions that could occur—and what those collisions would look like when recorded by the layers of detectors—and then compare the real data to the simulations. "Essentially, you see how theoretical, fundamental physics interacts with your detector," said Neal. "Doing all of these simulations was the biggest challenge to the reconstruction effort."

About 20 computing sites around the world contributed to this effort. Thanks to these sites (including SLAC, where simulation production nearly doubled in 2008), the collaboration simulated about 7.5 billion events. While almost all sites achieved record production levels, two of the sites, the computing center at IN2P3 in France and the Rutherford Appleton Laboratory in the United Kingdom, were at times producing as much as 20% of the total production each. "It was really a great effort from the collaboration's computing centers," said Dubois-Felsmann. "Not only did we expand infrastructure to make this happen, but we also optimized the process, accumulating a lot of one-percent improvements. In this way, we increased the speed by 20 to 30 percent."

The last step in processing is the separation of data into different streams based on the event types apparent in each collision. This process, called skimming, was performed at the computing centers—SLAC, GridKa in Germany, and RAL and the University of Manchester in the U.K. By upgrading the efficiency of the data skimming system, researchers hastened along this time-intensive process.

In all, the reprocessing was so successful that not only is the BaBar data set more accurate, but more of the data are being used than ever before. "Our data set has actually been growing since we turned the detector off," said Homer.

This is possible because in the past, any possibly distorted data were immediately discarded from analysis. "For example, for each hour worth of data we took, we would make a 'rough and ready' decision on the data to decide if it was good enough for analysis," said Dubois-Felsmann. "This time around, we went back and looked at the excluded data to decide if the original decision was too conservative." Researchers also loosened the filters that determined whether specific events were interesting and thus worth looking at in the future, and revived data that had been excluded because they seemed too anomalous at the time.

"Essentially, we were able to fix some problems that we were not able to fix in the past," said Dubois-Felsmann. "As a result, our dataset is about five percent larger and better than ever before."

Provided by SLAC National Accelerator Laboratory

Citation: BaBar Collaboration Completes Data Reprocessing (2008, December 18) retrieved 3 May 2024 from <https://phys.org/news/2008-12-babar-collaboration-reprocessing.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.