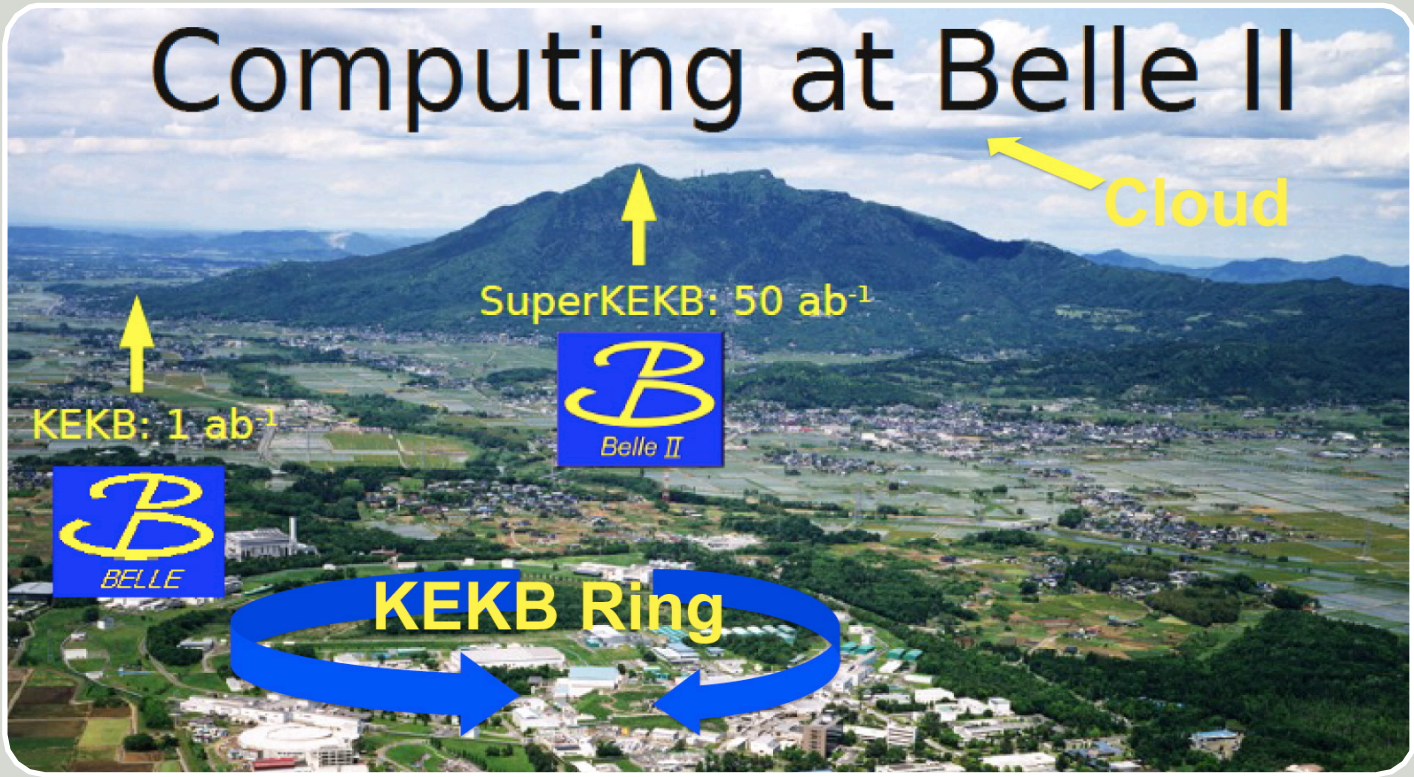
July 28, 2010

# FEATURE STORY




KEK's B Factory experiment, the KEKB-Belle experiment, ended operation on June 30 of this year. The upgrade, called the SuperKEKB-Belle II experiment by the collaboration, will produce electron and positron beams which are 50 times more luminous. This means the computing group expects the upgraded experiment to produce 50 times more data for analysis. Cloud computing may help meet these tremendous computing needs. (Image credit: Dr. Thomas Kuhr)

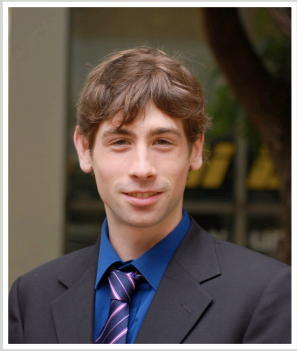
## Computing on demand: the Cloud meets the GRID

**[Belle II, Cloud Computing, DIRAC, GRID]**

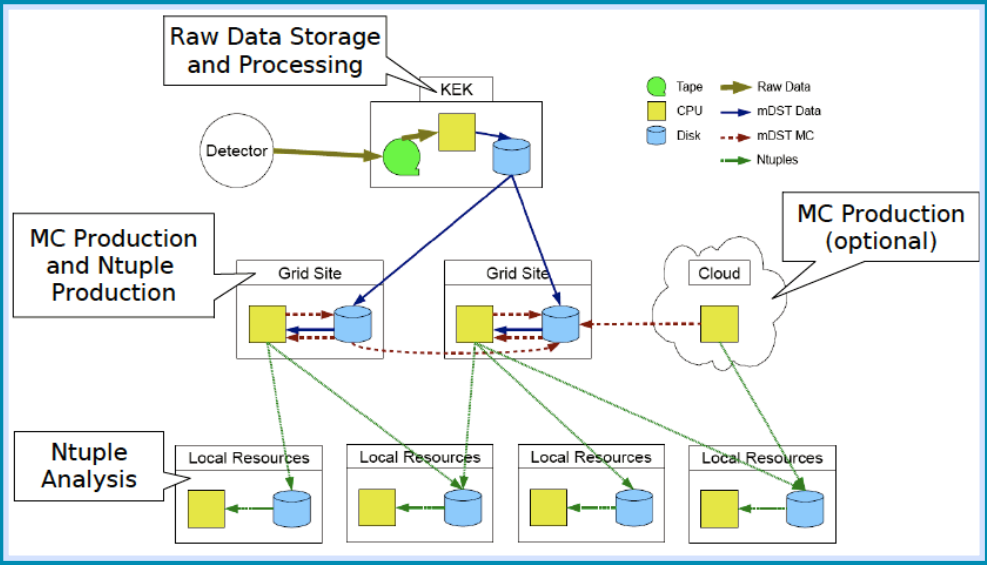
**Collaborators at the Belle/Belle II computing group just completed a series of successful exercises using a combination of Amazon's Cloud computing resources and the Belle II distributed computing GRID resources. DIRAC, a software package used to link the two distributed computing resources, was developed by the LHCb Collaboration and adapted by the Belle/Belle II team in collaboration with scientists at the LHCb computing group from Barcelona. The cost-effective nature of commercial Cloud computing may benefit all parts of the scientific community.**

**Suppose that you need several thousand** computers for, say, the next hour. Your institution likely does not normally require that volume of computing resources, and so you don't have such resources available in-house. Instead, you would need to find outside resources.

Many scientific communities are now forming a worldwide network of computer resources, linking computer systems from multiple institutions in order to accomplish common goals. These networks are called the GRIDs. However, sometimes, even all the resources of the GRIDs combined are still not sufficient to fulfill instantaneous needs. In this case, you could turn to commercial Cloud computing.



Tom Fifield of the University of Melbourne is a cloud computing expert and a member of the Belle/Belle II computing group.




The raw data processing is done using local KEK resources. The reconstructed event data is then placed on the GRID. In parallel with the raw data processing, Monte Carlo (MC) simulations produce synthetic data using resources on the GRID and the cloud. Physics analysis is done on the GRID by comparing the reconstructed data and the results of the simulations. Afterwards, end users can download candidate events to their local systems for analysis.

Cloud computing is the name given to offering computing resources made accessible via the Internet. The Google mail service is one prominent example of Cloud computing. This service provides everything that users need to send and receive emails, from the hardware of the mail servers to the software web applications where users access and manage their mail. “Cloud computing is a convenient, management-free solution that allows you to just pay for what you need,” says Tom Fifield of the University of Melbourne, a Cloud computing expert and member of the Belle/Belle II computing group.

Cloud computing will play an important role in fulfilling the increasing computing demands of future particle physics experiments. During the decade-long run of the KEKB-Belle B Factory, the experiment accumulated 3 petabytes of data. The upgraded version of this experiment, Belle II, will accumulate 50 times more data by 2020. The data processing needs of the upgraded experiment will exceed those of even the current world’s most powerful collider, the Large Hadron Collider (LHC) at CERN.

The advent of Cloud computing has caused a revolution in the field of scientific computing. In this upheaval, members of the Belle II computing group embarked on a project to unite the capabilities of commercial Cloud computing and the resources of the scientific GRID. Their goal was to manage jobs and data flow as efficiently as possible. To achieve this goal, Fifield worked with the three creators of the DIRAC software framework for distributed computing. DIRAC, the distributed infrastructure with remote agent control, is the tool used to manage distributed computing activities for the user community at the LHCb, one of the LHC experiments. According to Fifield, “DIRAC is the first software that can manage and integrate for the end user resources on GRID, Cloud, and local cluster.”

## Computing at Belle II

At any particle physics experiment, the purpose of most computing is to perform complex analyses on previously stored data. At both Belle and Belle II, the data is taken in periods of typically a few months, called an experiment. During each experiment, the data acquisition system (DAQ) collects all the data from all the detector components, and stores the raw data in a buffer system. The computing system then takes over and stores the raw data on tape.

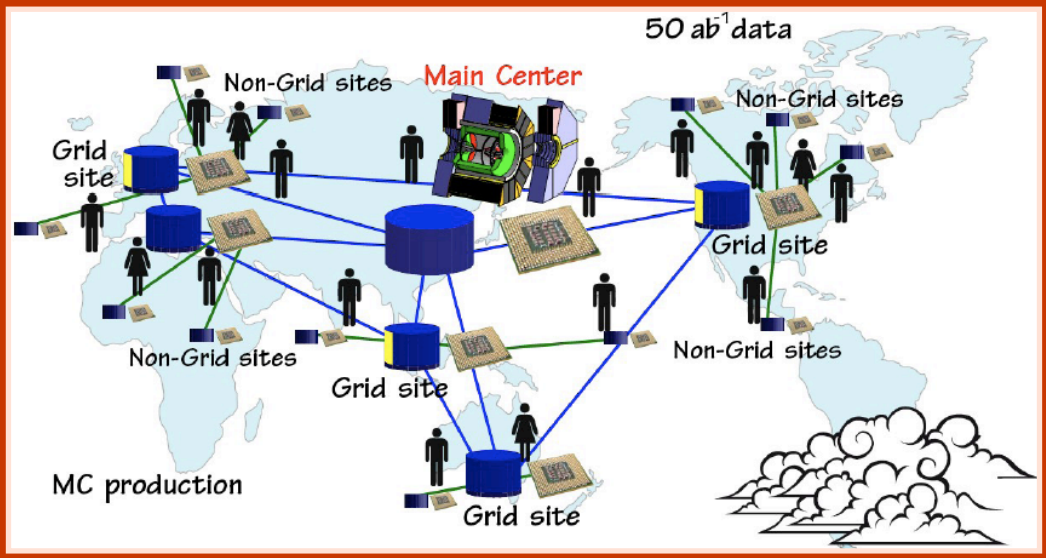
The first step in the analysis of the recorded data is the calibration of the detector. The conditions of the accelerator and the alignment and response of the Belle II detector are described by a set of calibration constants, and those values are used in the calculations that follow. The constants are fed into two separate calculations: raw data processing and Monte Carlo simulation.

During the raw data processing, software reconstructs particle tracks, energy, and momentum, and also identifies the type of each particle. The result is stored in a format called mini data summary tables (mDST). This contains all the information necessary for analysis, but in a much smaller file. The mDST data is then copied onto GRID sites.

In parallel with the raw data processing, other software carries out Monte Carlo simulations using the calibration constants, producing a synthetic data set six times as large as the original data set. By comparing the real data set with the larger, synthetic data set, physicists can cross check their measurements with different model predictions. The simulation is generic, simulating all possible physical processes that could have occurred during the actual run. This includes all decay channels as well as background noise. Unlike a signal Monte Carlo process, which simulates only one specific type of event, the results of a generic Monte Carlo simulation look like those from the real data including all background processes.

“For simplicity, the raw data is processed at KEK. Otherwise, we would need to go through the trouble of copying the entire raw data set to the GRID sites,” says Dr. Thomas Kuhr of Karlsruhe Institute of Technology, one of the two co-coordinators of the Belle II computing group. “For simulations, however, the only inputs needed are the calibration constants and some background information. That

Dr. Thomas Kuhr of Karlsruhe Institute of Technology is one of the two co-coordinators of the Belle II computing group.




Monte Carlo simulations will use computing resources from the GRID. Cloud resources can also be used if the GRID resources are insufficient for peak loads.

information is small in size and easily distributed." Once the inputs are available, simulation only requires good CPUs. This, therefore, is where the GRID and the Cloud come into play.

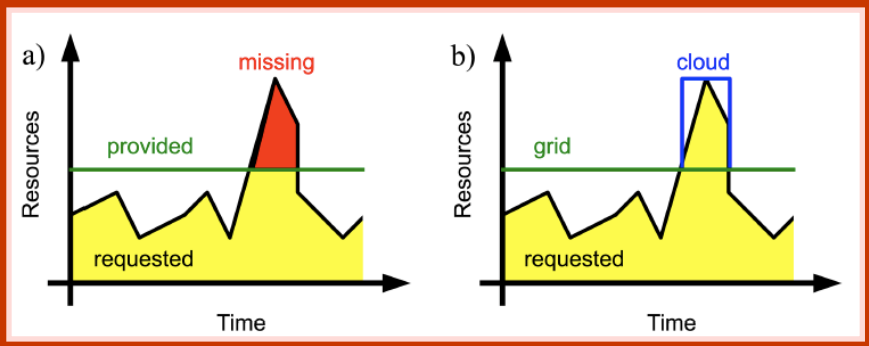
The physics analysis is usually performed on the GRID. Each physicist will be looking at different processes or channels for her/his analysis, and when she/he finds candidate events in the mDST files, they will be stored in an analysis-specific format, called an ntuple. The ntuples can be downloaded to the investigator's local site. An ntuple includes kinematic information about the relevant events, such as mass and momentum.

"One great advantage which the Belle II computing framework has over previous high energy physics experiments, including the original Belle, is the standardized data format. The Belle II framework uses the same data format throughout the DAQ and the initial stages of computing, all the way until the data reaches the end users," says Kuhr. The standardized data format mainly owes to the new programming standard in data processing and simulation in high energy physics, called object-oriented programming. The use of an object-oriented language allows developers to simply and easily develop software for many different detector components. The common Belle II software framework is designed to handle event data in a unified way.

## Ride on a Cloud

Due to the advancement of Internet technology, laboratories around the world now increasingly rely on shared computing resources. For example, LHC experimenters have formed a network of computing resources, the WorldWide LHC Computing GRID (WLCG), in order to cope with their enormous data handling requirements. The WLCG currently consists of computers in 200 member institutions from 60 different countries.

Belle II computing will also be heavily dependent on GRID technology. Already, seven member institutions from seven different countries (Australia, Czech, Germany, Japan, Korea, Poland and Slovenia) are involved in the test runs. The Belle II GRID will eventually include around 20 member institutions. However, in one way, the GRID is no different from ordinary, local computing resources. This is that the number of maximum available resources is ultimately fixed. Researchers are currently working to understand whether the currently planned GRID will be sufficient for the computing needs of the Belle II. If not, they will have to decide whether they want to add additional CPUs to satisfy the demands during peak loads.




When the experiment is running, the large volume of data processing and Monte Carlo simulation place an unusually heavy load on computing resources. The maximum available GRID resources are fixed. If computing needs exceed the available GRID resources, cloud resources can be brought in to fill the gap.

Heavy loads are anticipated at the end of each experiment. At this time, raw data processing is immediately followed by Monte Carlo simulation. However, at other times, most of the GRID CPUs will be at rest. "The idea of Cloud computing is to own only enough CPUs to meet everyday computing requirements, and to rely on outside sources for additional resources during peak loads," explains Prof. Takanori Hara of KEK, the other co-coordinator of the Belle II computing group.

## Virtual Clouds

GRID computers generally work on projects with similar scientific goals, and so the computer resources on a GRID are set up to provide a similar environment to all users. "What is different in Cloud computing is the idea of virtualization," says Kuhr. "In a Cloud, virtual machines simulate hardware, so a person running a virtual machine has full control over their virtual hardware. This way the system configuration is simple."

The concept of virtualization has been around since the 60s. Back then, IBM was the first to embrace the easy setup and the swift switch on/off of the machines, which virtualization allowed. Back then, simple operating system level virtualization enabled one to have multiple virtual machines running at once on the same physical CPU.

The concept of virtualization has long been popular among system engineers in high energy physics as well. The challenge was in the difficulty of virtualizing hardware components. "The recent development of new tools to stabilize hardware abstraction accelerated the adoption of the virtual machine," explains Fifield. Because of this, in 2007, big commercial Cloud computing started to take off.

Cloud computing services come in three different types: infrastructure as a service, platform as a service, and software as a service. The interest of the Belle/Belle II computing team is in the infrastructure as a service, namely CPUs. To avoid the risks associated with vendor lock-in, data cannot be permanently stored on Cloud. The collaboration will rent only the CPUs, for only the precise amount of time they are needed, and pay for only the time that they have used.

In his undergraduate years, Fifield created an open source management program for Cloud resources. When he joined the Belle II collaboration in 2008, he was already considering the possibility of using commercial Cloud computing for Belle/Belle II simulations. Where would the central catalog and the job managing system sit in the Cloud of virtual machines? Where can users configure the jobs and settings dynamically? In which conditions should virtual machines be terminated or started up?

## DIRAC's pilot model and dynamic configuration

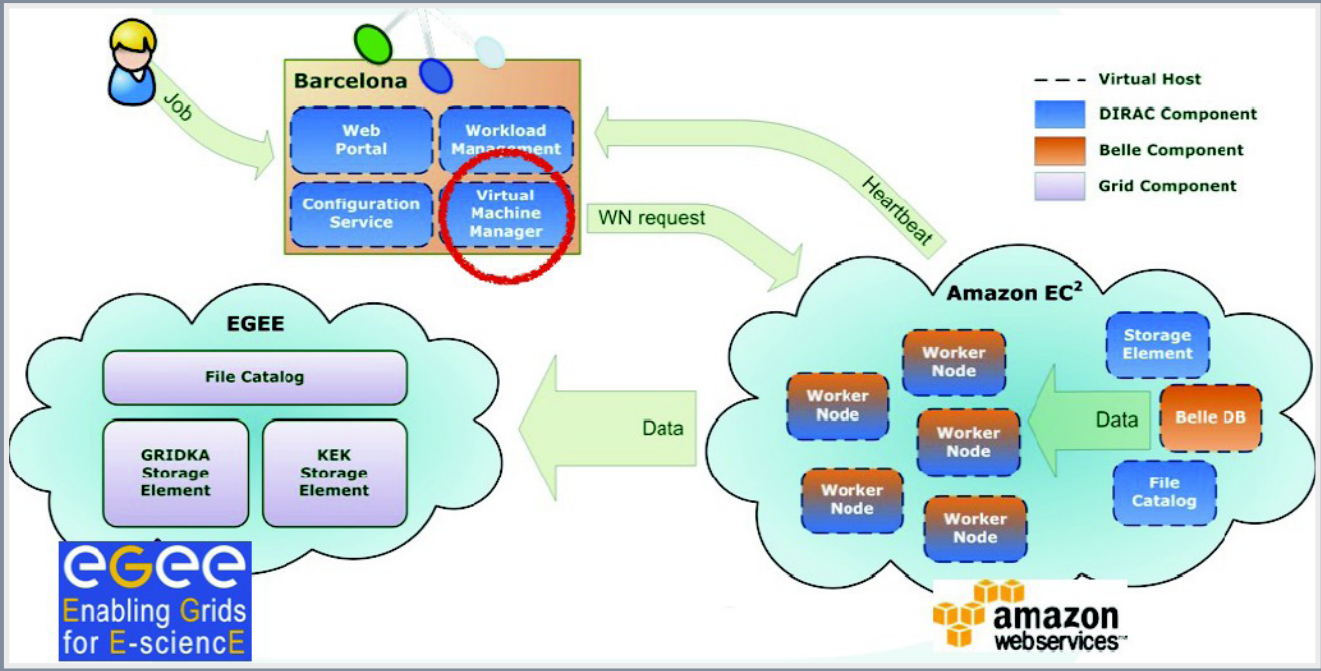
In 2008, Fifield sat in a pub in Prague with his old colleagues from ATLAS. There, he met three of the developers of the DIRAC system, Adria Casajus, Ricardo Graciani and Stuart Patterson. DIRAC is the software solution used by the LHCb to manage all their distributed computing activities. The software started out as a tool to manage large scale Monte Carlo simulations run across multiple computer centers. It eventually evolved into a full, stand-alone GRID solution that could handle raw data processing, data replications, physics analysis, and user data access, as well as the monitoring and supervision of all those activities.

"DIRAC was designed to be a generic solution, rather than experiment specific," says Fifield. "We found out that the job management, catalog management, and software components that we needed were all already in there." In July 2009, Fifield and three DIRAC developers from Barcelona put their heads together and started discussing how to adapt the DIRAC system for Cloud computing.

Prof. Takanori Hara of KEK is one of the two co-coordinators of the Belle II computing group.

Two of DIRAC's capabilities are particularly important for dealing with Cloud-type distributed systems. These capabilities are the pilot model and the ability for dynamic configuration. Cloud computing does not provide the user with a way to find out which specific machines are running a specific job, or where in the world they might actually be found. Instead, virtual machines simply sit around, hidden in Cloud, waiting for users to boot up. No means of dynamic configuration is provided as a service either. DIRAC provides solutions to these limitations of Cloud services.

The pilot model describes how each end work node obtains computing jobs. In a GRID, a central management system knows the status of all computing resources (CPUs) on all contributing sites. This means that, when the system receives new jobs, it knows where to send them. In DIRAC, on the other hand, a software entity called the pilot factory sends out blank job inquiries to all currently contributing sites, asking if any CPU resources are available at each site. Upon receiving the inquiry, contributing sites with free CPUs contact the central management system for more jobs. This is called the pilot model. "In the pilot model, jobs are pulled rather than pushed. In Cloud computing, there is no simple way to push jobs to virtual machines," says Fifield. "In the pilot model, you don't have to know about everything that goes in all contributing sites. You are also guaranteed that, once pulled, the job will be run, which is not always the case in distributed push type job management."

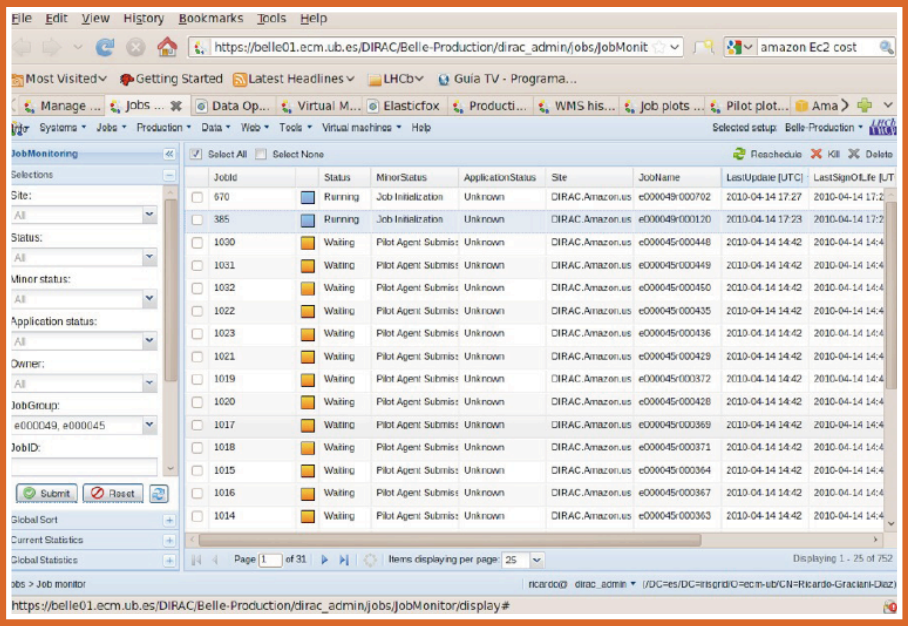



The distributed infrastructure with remote agent control (DIRAC) software is a framework to manage distributed computing activities for the LHCb user community. Fifield and three DIRAC creators recently added virtual machine management capability to allow users to take advantage of cloud computing resources.

Virtual machines in a Cloud also contact to the configuration service provided by DIRAC. The configuration service allows users to dynamically configure various parameters such as the location for job outputs, and the location of task queues. The service allows users to become aware of the existence of virtual machines and everything that is happening in the DIRAC system.

## DIRAC for the Cloud

The Cloud interface for DIRAC was built on top of the existing layers of DIRAC. The goal of the cloud interface was to minimize the number of hours that virtual machines are online, as CPUs are paid for by the hour. If a virtual machine stops responding, it is terminated. If the central task queue says it needs more resources, then more virtual machines will be brought in. In addition to working out these basic functionalities, Fifield and his collaborators exchanged many emails trying to work out optimal data flows between GRID and Cloud.




The virtual machine instance browser allows users to schedule, reschedule, and delete jobs using a web browser.

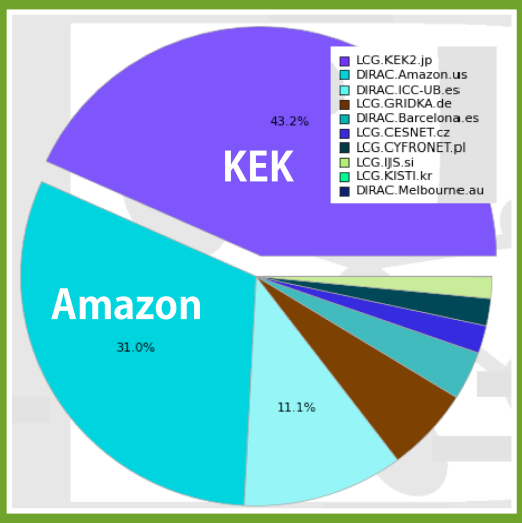
The DIRAC system manages data and jobs from a central location. From the web interface, users can configure, display, schedule, reschedule, and delete jobs. “From a web browser, a user can select the experiment period, the type of Monte Carlo, the data source and destination, and schedule a job,” explains Fifield, with his web browser at hand.

## The cloud: a cost-effective solution

Starting in April of this year, Fifield and his collaborators performed three phases of testing. The team used actual Belle data to simulate events that were used in actual physics analyses by Belle collaborators. The first phase used cloud resources, the second used both cloud and local cluster resources, and then the third used a combination of cloud, local, and GRID resources. “The test was a great success,” says Fifield. “At the peak, we used 2,000 CPU cores for 24 hours. This is the size of all the Belle GRID resources combined.”

The Belle II computing team has estimated the total cost of their system, and found themselves in a surprisingly promising situation. Owning computers is an expensive operation. The total cost includes not only the capital cost of hardware, but also the cost of services such as building space, electricity, cooling, and Internet connectivity, as well as the cost of the manpower to service the hardware and keep the system alive. “Cloud computing costs only twenty US cents for 10,000 simulated events. We estimate that using cloud resources for Monte Carlo simulations will be about 50 percent cheaper than using our own computers,” says Fifield.

Members of the Belle II computing group gather for a group photo.




This image shows the CPU days by site during the successful test run of DIRAC on cloud, GRID and local cluster. Nearly one third of the CPU time was provided by Amazon cloud computing.

## DIRAC's potential and future

DIRAC is now production ready. Anytime that the collaboration needs to run Belle Monte Carlo simulations, they can do so. Looking forward, Fifield says there are a couple of improvements he would like to investigate. First, he would like to verify that DIRAC works with cloud computing providers other than Amazon. Amazon was one of the first companies to offer cloud services, and those services have worked quite well for the Belle. Second, Fifield would like to investigate further uses of cloud computing to improve physics analyses. He plans to investigate the nature of the data flow between GRID and cloud to optimize performance for physics analyses.

Installation of DIRAC also allows computing resources on non-GRID sites—not just cloud resources—to be shared as distributed resources. Currently, each GRID site needs the GRID middleware installed, which can be difficult when an institution has strict firewall restrictions. When the installation of the middleware is not possible, DIRAC allows any computing resources, even laptops, to contribute. In the future, DIRAC, may be a promising option to broaden the computing resources available to Belle II.

DIRAC's management capabilities will also be a great help in the management of local, GRID, and cloud computing resources. For the Belle II computing team, the final user interface will ideally be a single pane on which users can choose specific options for particular physics analyses. Based on the user inputs, the system would then work out the necessary details. "Everything else will be a black box. You don't have to understand file names or anything else. You can simply specify the basic information about your experiment and analyses, such as experiment period and the type of simulation," says Fifield. "If the system ends up using Cloud resources, a local cluster at some institute or the GRID, the users would not know."

The Belle II management will review the cloud computing option, carefully evaluating the pros and cons. The use of commercial computing resources would be a significant change from the way that particle experiments have traditionally been conducted. However, the advantages of cloud computing, including the greatly decreased cost, are major. "This is an exciting new field. With DIRAC, any GRID user community, from high energy physics to computational sciences or medical imaging, can now use the extra CPUs on the cloud," says Fifield.

The Belle II computing group consists of 35 physicists and computer scientists from 8 countries. Aside from the GRID-cloud Monte Carlo simulations, there are many other projects in which members are actively involved. Their software-related activities include a new data processing framework and new analysis tools, which they hope will make raw data processing and event reconstruction more efficient. Also, they are working on a new project client from which users can access computational resources. There is lots of work to be done, and the only limitation is the size of the group. "We welcome new members and new contributions," says Kuhr. "This is an interesting project, where people could make significant contributions to the exciting work of the Belle II computing group."


The current Belle II collaboration consists of around 300 physicists and engineers from 53 institutions in 13 nations.

**Related Link:**
Belle II

**Related Issue:**
The legacy of Belle and BaBar

Designing the ultrafast DAQ for Belle II

New electronics tested for Belle II central drift chamber
Belle II's new logo and new beginning

SuperKEKB making headway toward higher luminosity

Belle II collaboration meets at KEK

HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION (KEK)

Address : 1-1 Oho, Tsukuba, Ibaraki 305-0801 JAPAN
Home : http://www.kek.jp/intra-e/feature/
Mail : misato@post.kek.jp