

DB2 Information Management Software

IFCA promotes collaborative Grid research with IBM computing solution.

Overview

■ **Application**

Grid computing network for scientific research in astrophysics, particle physics, meteorology and biology

■ **Business Benefits**

Access to high-power distributed computing resources for collaborative research; ability to support unexpected surges in computing demand; greater system resilience to run 24x7, compute-intensive applications

■ **Software**

IBM Informix® Dynamic Server, Version 9.3; Red Hat Linux

■ **Hardware**

IBM@server xSeries®



As part of Spain's largest research institution, IFCA serves a diverse set of practitioners, from medical professionals to meteorologists monitoring critical weather patterns.

The genius of Albert Einstein may have relied on no more than a pencil and a flat surface to prove breakthrough theories, but today, computing power is what catapults scientific research to new frontiers. In the not-too-distant future, it will be the norm for numerous researchers to collaborate across continents, running calculations and simulations on high-powered computers. Together, they will generate trillions of terabytes of data that will be stored in electronic databases and shared over high-speed networks—creating a worldwide virtual research organization. This will be the age of Grid computing.

“IFCA is one of the most active centers participating in Grid research, with IBM and Linux technology being two of the key components.”

— Jesús Marco, eScience and Grid Leader, IFCA



IBM@server xSeries provides the hardware foundation for IFCA's Grid computing solution.

At the Instituto de Física de Cantabria (IFCA), or the Physics Institute of Cantabria, in Spain, Grid computing is a reality now. Having recently implemented its own Grid resource, it now participates in a Grid testbed that provides dozens of researchers access to the combined computing resources of hundreds of computers, spread across ten different European countries, providing the equivalent of hundreds of gigaflops of processing power. For these researchers, engaged in solving complex problems in fields such as particle physics, astrophysics, meteorology and biology, life without Grid computing would now be unimaginably slow.

Funded by the University of Cantabria and a member of the CSIC (Spanish Highest Center for Scientific Research), the largest non-profit research institution in Spain, the IFCA is a very

active center in Grid research. From its base in Cantabria, the IFCA provides a cluster of 80 IBM @server xSeries 220 systems with dual Pentium III processors, running Red Hat Linux, IBM Informix Dynamic Server 9.3 and a variety of scientific research software. "IFCA is one of the most active centers participating in Grid research, with IBM and Linux technology being two of the key components," says Jesús Marco, eScience and Grid Leader, IFCA.

Research data will be exchanged over the Grid at a rate of more than 100 megabytes per second. To process this volume of data without a Grid would otherwise require a single computer center managing thousands of latest-generation PCs. Instead, the data is now distributed and processed over the Grid, allowing all the researchers to expand their computing resources with minimal investment. "Though we have not done a cost-benefit analysis, the benefits are quite obvious to us," says Marco. "We can ensure high availability operations, handle more complex research requests and support unexpected demands for computing resources in a Grid environment."

"We've gained tremendously from IBM's knowledge of Grid computing and like its technology roadmap for this environment."

—David Rodríguez, Software Developer, IFCA

Open, interconnected and cost-effective

Being part of a Grid network, it is essential for the IFCA to deploy open standards-based technology to connect seamlessly with diverse systems and applications. It is also important to deploy cost effective solutions and maximize every euro of funding received. Thus, when the IFCA began its search for a technology solution, it leaned toward the Linux operating system and open source software. "At the same time, we also wanted an established name in the IT industry. So when we heard that IBM is deeply involved in Grid technology, we were keen to test its solution," notes Marco.

It proved to be an excellent match. "IBM's commitment to open standards and its extensive range of hardware and software solutions for Linux was an unbeatable combination for us," says Marco.

Managing a data avalanche

The IFCA participates in several prestigious international projects in collaboration with laboratories like the European Laboratory for Particle Physics (CERN) and the European Space Agency (ESA). In each project, the interactive use of very large geographically distributed databases and system resources, as well as the application of data mining technologies, is an essential requirement.



IBM technology enables IFCA and other Grid collaborators to perform distributed database queries and then apply data mining on the results.

For example, the DataGrid project, created in 2000 under the European Union's Information Society Technologies Program and coordinated by CERN, is building a computational infrastructure to allow intensive calculation and analysis of large-scale shared databases among widely distributed scientific communities. Complementing it, the CrossGrid European project, initiated in 2001, aims to support interactive applications in a Grid environment, and IFCA is responsible for the distributed data mining.

"For the data in our Grid, we needed a very powerful database engine," explains Marco. "After working with Informix Dynamic Server 9.3 in the Grid for a year as a beta tester, we were impressed with its performance. It met all the functional requirements

needed in physics and meteorological projects—splendid stability and scalability, a strong replication scheme, interoperability among useful distributed tools like the Object Translator, and XML support."

At any given time, the IFCA stores several terabytes of data in its data repositories. "That's a lot of load to put on the system," notes David Rodríguez, IFCA software developer. "But Informix Dynamic Server has very useful replication techniques that allow us to continue working even if some of our servers are offline, say, for routine maintenance."

"For some of our applications we have a three-tier architecture providing a Web interface to query data. We tested Informix Object Translator for this purpose. The Informix Object Translator makes it possible to extend the data to three-tier Web applications, using XML as the data interchange format," explains Rodríguez.

The Globus Toolkit is the middleware used as the basis for the Grid system. It includes software services and libraries for resource monitoring, discovery and management, plus security and file management. The data access layer is built on top of it.

A new version of the Toolkit (GT3) has just been released, implementing Open Grid Services Architecture (OGSA) standards. IBM is an active partner of the Globus Alliance, which will continue developing the Toolkit and enforcing the standards.

Robust, scalable platform

Another significant benefit of the IBM platform was the ease with which the IFCA was able to deploy and configure its system. The installation process was automated, incorporating a Preboot Execution Environment (PXE) and Kickstart procedure, a Linux-based industry standard for booting over a network. This allowed an entire 80-node cluster to be configured in a few minutes. "This was a remarkable achievement," notes Marco.

IBM Business Partner Consulting Informático Cantabria (CIC) provided the IBM servers for the Grid solution.

"Considering the rate at which demand for computing resources grows in a scientific environment, we need a highly scalable hardware infrastructure," notes Marco. "The IBM xSeries servers provide not only the most scalable environment, but also a resilient, highly available one for our mission-critical applications."

As demand for processing resources increases and data volume surges, the IFCA is seeking even greater scalability and performance. Rodríguez adds: "We've gained tremendously from IBM's knowledge of Grid computing—and we like its technology roadmap for this environment."

For more information

Please contact your IBM marketing representative, IBM Business Partner or IBM Direct at: 1 800 IBM-CALL.

Visit our Web site at ibm.com/software/data

For more information about IFCA, visit: www.ifca.unican.es



© Copyright IBM Corporation 2003

IBM Corporation
Silicon Valley Laboratory
555 Bailey Avenue
San Jose, CA 95141
U.S.A.

Produced in the United States of America
09-03
All Rights Reserved

@server, IBM, the IBM logo, Informix and xSeries are trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product or service names may be trademarks or service marks of others.

This case study is an example of how one customer uses IBM products. There is no guarantee of comparable results.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.