

## PBS Professional at Toulouse Genopole: A Faster Path to Genomics Research Results



### Key Highlights

#### Industry

Life Sciences

#### Challenge

More processors & intensive utilization needed for increasing number of complete genome sequences.

#### Altair Solution

Powerful workload manager with Job Arrays to increase power & user-friendliness of Toulouse Genopole's systems.

#### Benefits

- Enables scientists to write genomic sequence analysis workflows much more easily.

### Customer Profile

The Toulouse Midi-Pyrénées Genopole, a research program set up in 1999 in southern France by the French Minister of Research, is part of the massive French research initiative known as the National Genopole Network. The genopole name is metaphoric. It brings to mind a magnetic pole that attracts researchers and entrepreneurs to the potential of genomic research. And it is working: The initiative has drawn together, at seven locations across France, a symbiotic mix of public laboratories, biotech companies, and educational institutions. The genopoles stimulate genomics research and may provide an incubator for biotech enterprises.

Thirty academic laboratories in and around the city of Toulouse make up the Toulouse Genopole. Some 3,000 researchers are working here in five basic fields: Microbiolog, plants and animals, health, ethics, and bioinformatics. The Toulouse Genopole Bioinformatics platform supplies computational resources, data, and bioinformatics-specific software for work by Genopole scientists working in sequence analysis, gene expression, and gene mapping. Some are biologists or bioanalysts using a bioinformatics environment proposed by the platform for large- and medium-scale sequence analysis. Others are doing research in bioinformatics fields, focusing on methodological development.

# Toulouse Genopole Success Story

**“In the bioinformatics world, we run a lot of small, uncorrelated jobs. With other workload management technologies, handling these huge numbers of jobs can quickly become a nightmare. PBS Professional with Job Arrays makes all these problems disappear.”**

**David Allouche,**

Operations Manager, Bioinformatics Computing Platform, Toulouse Genopole

The Genopole’s bioinformatics resources are hosted by the Toulouse INRA (National Agronomic Research Institute) computing center. The startup compute platform for bioinformatics at the Toulouse Genopole in 2000 was a four-processor Dell server, supplemented later by a four-CPU IBM system. Today Altair's PBS Professional® workload management software is helping INRA to achieve maximum utilization of its compute resources and accelerate the pace of genomic studies.

## **Basic Needs: More Processors and Intensive Utilization**

Four processors were not enough for an ambitious bioinformatics program and the hungry researchers developing bioinformatics tools in three Genopole-connected laboratories. They were also coping with a swiftly increasing number of complete genome sequences. An 11-node, 22-CPU LNXI Opteron cluster, set up in 2004, has now been superseded by a substantial production compute infrastructure.

“When we were planning the implementation of the first cluster in 2003 I saw right away that queuing would be very important,” says David Allouche, Operations Manager for the Bioinformatics Computing Platform. “I investigated the available workload management systems, including OpenPBS. We wanted maximum utilization of our processors, and we knew we were going to upgrade to many more processors.”

“We ultimately selected PBS Professional for two reasons: We liked their dynamic approach and willingness to work with us on bioinformatics-specific technology, and we knew that they were developing a new feature named Job Arrays that had the potential to help us. We wanted to work with a leading-edge vendor to develop the tools we needed.”

Job Arrays, which was released in PBS Professional, allows users to submit and manage a group of jobs — for example, a large number of bioinformatics searches — as a single job. Researchers can submit, query,

modify and display this group of jobs as a unit. This is of great significance in large-scale bioinformatics studies.

“In the bioinformatics world, we run a lot of small, uncorrelated jobs,” says Allouche.

“In gene annotation, for example, a researcher may run 2,000 or more jobs. In one case, when we were doing gene annotation on a new bacterium, we ran two million jobs. The situation is similar with comparative genomic runs, in which we compare a genome from one species with a genome from another. With other workload management technologies, handling these huge numbers of jobs can quickly become a nightmare. PBS Professional with Job Arrays makes all these problems disappear.”

Today the INRA computing center serves four kinds of users:

- Biologists who access bioinformatics tools through the platform’s open web site (15,000 visits in a typical month)
- UNIX-literate biologists who run

### Part of the National Genopole Network

Thirty labs in and around the city of Toulouse make up the Toulouse Genopole. The functionality of Job Arrays significantly increases the power and user-friendliness of our systems.



bioinformatics tools in console mode (about 100 accounts)

- Software engineers developing new interfaces and workflow systems for large-scale genomics or bioinformatics programs (about 30 users in the Toulouse INRA center and neighbouring Genopole laboratories)
- Scientists developing new algorithms and methodologies for gene detection, comparative genomics, microarray analysis, sequence searches and other fields (10-15 users)

There are probably 250 applications running on the cluster, and the demand for computational resources is steadily increasing.

### The New Cluster: An Efficiency Breakthrough with Job Arrays

The French Altair team worked with Allouche to integrate their bioinformatics applications with the Job Arrays function within PBS Professional. It was installed, along with PBS Professional, on INRA's LNXI Opteron

cluster when it was upgraded to 96 processors and came online in January 2006. The upgraded cluster went immediately into beta operation with three users.

"I did several benchmarking tests after we brought the cluster on line," says Allouche. "The first thing I learned is that PBS Professional is very powerful. It is really interesting to have a software layer that virtualizes computational resources for the scheduling process and is transparent to our users."

"I also found that the functionality of Job Arrays significantly increases the power and user-friendliness of our systems. It enables us to write genomic sequence analysis workflows much more easily. Where we had thousands of jobs we now have one. In many cases, we want to take the same set of jobs through a series of computational steps. Job Arrays enables us to do this very easily."

### Ahead: More Compute Power and Extended Access

"We plan to continue expanding our computational power," says Allouche, "and to continue using PBS Professional with Job Arrays to handle the special requirements of our bioinformatics workload. "

"We decided to work with Altair in the first place because we knew they wanted to be at the leading edge of scientific innovation. Job Arrays is very important for our bioinformatics work. It can be applied in numerous other fields, so I can imagine that Altair is working on other things for us, too."

INRA has implemented web-based services into its HPC platform at Toulouse Genopole and is attracting new visitors. With help from Altair, the web interface will soon be integrated with PBS Professional to manage job submission from remote users, including researchers with other bioinformatics research platforms in France.

Visit the PBS Works library of  
**Success Stories**  
at [www.pbsworks.com](http://www.pbsworks.com)

## About Altair

Altair empowers client innovation and decision-making through technology that optimizes the analysis, management and visualization of business and engineering information. Privately held with more than 1,800 employees, Altair has offices throughout North America, South America, Europe and Asia/Pacific. With a 27-year-plus track record for high-end software and consulting services for engineering, computing and enterprise analytics, Altair consistently delivers a competitive advantage to customers in a broad range of industries. Altair has more than 3,000 corporate clients representing the automotive, aerospace, government and defense, and consumer products verticals. Altair also has a growing client presence in the electronics, architecture engineering and construction, and energy markets.

---

## About PBS Works

PBS Works™, Altair's suite of on-demand cloud computing technologies, allows enterprises to maximize ROI on existing infrastructure assets. PBS Works is the most widely implemented software environment for managing grid, cloud, and cluster computing resources worldwide. The suite's flagship product, PBS Professional®, allows enterprises to easily share distributed computing resources across geographic boundaries. With additional tools for portal-based submission, analytics, and data management, the PBS Works suite is a comprehensive solution for optimizing HPC environments. Leveraging a revolutionary "pay-for-use" unit-based business model, PBS Works delivers increased value and flexibility over conventional software-licensing models.

[www.pbsworks.com](http://www.pbsworks.com)



**Altair Engineering, Inc., World Headquarters:** 1820 E. Big Beaver Rd., Troy, MI 48083-2031 USA  
Phone: +1.248.614.2400 • Fax: +1.248.614.2411 • [www.altair.com](http://www.altair.com) • [info@altair.com](mailto:info@altair.com)