

## IPMC Accelerates Genomic Sequencing With ProActive Parallel Suite ®

The Institute of Molecular and Cellular Pharmacology in Sophia Antipolis (France) is supported by several French Research Agencies (CNRS, INSERM, ANR), and by the Nice - Sophia Antipolis University. With a 20-year old dedication to outstanding science in Molecular Medicine, it covers disciplines from Integrated Neurosciences, Molecular and Cellular Biology, Molecular Pharmacology, Cardiovascular Sciences, Bio-Informatics, Genomics, and Biophysics.

### Customer challenge

To develop its activity in genomics sequencing, the IPMC acquired in 2008 a SOLiD™ system, a next-generation sequencing platform developed by Applied Biosystems for genomic analysis.

The volume and the size of data currently generated in genomic analysis grow steadily. In parallel, algorithms to process and consolidate these data are always more complex. The need for an increased computing power is then pressing. But hardware resources are a limited and costly asset: to mutualize available hardware resources and enable a better workload management is the solution to overcome the lack in computing power.

### Challenge

- Process an increasing number of genomic sequences using SOLiD™ platform analysis software

### Solution

- Integrate ProActive Parallel Suite in SOLiD™ software, Transcriptom pipeline
- Use available hardware resources, especially lab desktops, to increase the computing power

### Results & Benefits

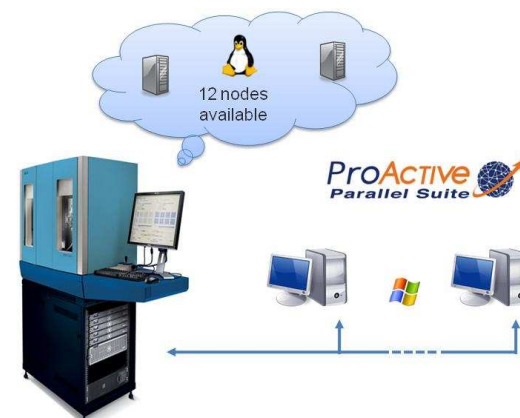
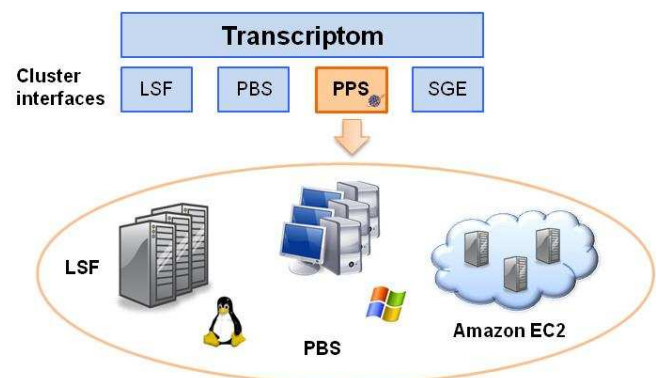
- Accelerated by 4 times the data processing
- Added great scalability to process always larger data and increase the speedup
- Increased flexibility by adding the ability to run SOLiD™ software everywhere

### Solved with ProActive!

In the frame of a collaboration between IPMC and ActiveEon, the SOLiD™ software and hardware configuration have been upgraded:

#### SOLiD™ Software

The transcriptom pipeline, designed in Java, is part of SOLiD™ software and includes compute intensive processing. Support for LSF, PBS and SGE is already provided in SOLiD™ but none of them allow to easily add various external nodes like ProActive can do. ProActive was the solution to allow aggregating all available computers at IPMC to SOLiD system and increase its computing power. As a fully Java solution, ProActive integration has been really simple.



#### Hardware configuration

The SOLiD™ sequencer includes 4 blades, each made of 4 cores, leading to a total of 16 cores available. However four of them are dedicated to the primary biological analysis, not to the Transcriptom pipeline.

In addition to the internal nodes, PCs already available in

ActiveEon is a Professional Open Source Software company, co-developing ProActive Parallel Suite with the development team OASIS (a joint team between INRIA, CNRS, I3S and the University of Nice – Sophia Antipolis).

ActiveEon provides a full range of services for ProActive Parallel Suite:

- Training and certification
- Consulting
- Technical support
- Integration and development

## Contact information

ActiveEon  
2004, route des Lucioles  
BP 93  
06902 Sophia Antipolis Cedex  
France

Tel: +33 (0) 4 92 38 50 00  
[info@activeeon.com](mailto:info@activeeon.com)

For more information, visit:  
[www.activeeon.com](http://www.activeeon.com)

the lab have been added to the pool to increase the computing power. The hardware resources are connected via the Ethernet network. Both internal nodes and external PCs are controlled by ProActive. Furthermore, ProActive is fully compatible with Torque, LSF and SGE and can then aggregate clusters managed with one of these job schedulers to SOLiD™ internal resources and Windows PCs.

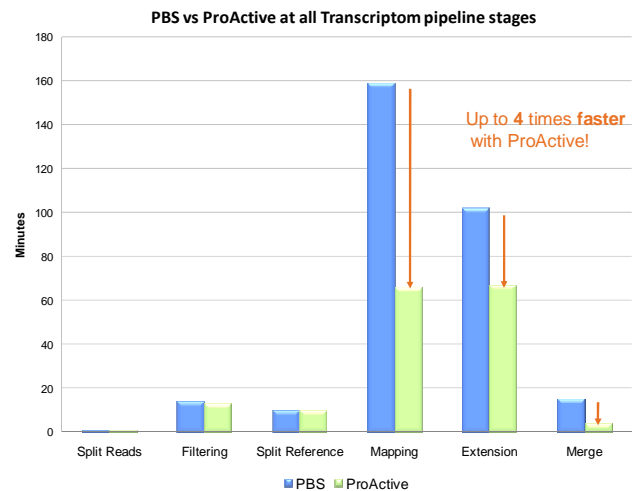
## Results

This upgraded version of SOLiD™ software has been quickly setup thanks to the fully Java solutions, SOLiD software and ProActive form.

After SOLiD™ software code analysis, it only took a couple of days to add a new cluster interface for ProActive.

Testing has been done in three steps:

- First, the proper integration of ProActive has been checked. The output file generated by the SOLiD™ version distributed by ProActive has been compared to the one generated by the standard configuration using only internal nodes: files were identical. Thus, the correct behavior of the upgraded software is assumed.
- In a second step, external resources have been added to assess the speed up that can be achieved. In total, 46 nodes have been made available. The diagram above highlights the speed up achieved with the upgraded SOLiD™ software: up to 4 times faster!
- As a last test, SOLiD™ software has been successfully deployed over Amazon EC2 nodes. This allows any organization to accelerate SOLiD™ execution based on its needs.



## Solution benefits

The IPMC team has experienced several benefits including:

- **Increased productivity** through a reduced run time: the higher throughput allows testing more genomic sequences.
- **Improved scalability**: depending on the input and reference data size, the user can chose to increase or reduce the number of extra resources used. The solution is then ready for next generation reads file.
- **More flexibility**: additional nodes can be easily acquired even when the program execution is on-going. Also, the run can be paused or resumed by the user when needed and priorities between jobs can be set, giving this way a higher control of the users on the processing.
- **Simplified maintenance**: ProActive directly supports common schedulers like PBS, LSF or SGE. Any hardware resource can then easily be added to the pool dedicated to the SOLiD™ sequencer data processing.
- **Substantial costs saving**: thanks to an optimized utilization of existing resources there is no need anymore for the acquisition of additional computing power to meet users and applications requirements.