Curso Grid y e-Ciencia 2010

ID de la contribución : **18**

Tipo : **no especificado**

Applications: Biomed

viernes, 9 de julio de 2010 10:00 (60)

The development of communication networks has produced a breakthrough in most areas of life. In science, the increase of bandwidth and coverage of research networks have developed a new way of doing science. E-Science, as enhanced science, is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it. E-Science is the way to deal with the interdisciplinary approach for tackling the Data Deluge that new biomedical disciplines are posing.

Biomedicine, as the storage, management and processing of data related with the physiology and structure of living beings, comprise many different disciplines, such as Bioinformatics, Medical Imaging, Epidemiology and Simulation. Those technologies are posing many challenges regarding computing, data storage and collaborative work. Many of these challenges are analyzed in the SHARE Roadmap published by the HealthGrid Association.

Medical Imaging is becoming one of the largest consumer of global storage. This huge need in storage is also leading to a large need on computing resources. This large availability of data constitutes a great opportunity for creating knowledge repositories to drive research and training. In this sense, the middleware TRENCADIS, developed by the UPV, enables creating virtual repositories of medical imaging, organized by content through the use of DICOM Structured Reports, from data coming from different centres. TRENCADIS uses a model based on virtual communities to organize authorization models. This federated model is also being developed by major industrial providers.

In bioinformatics, genomics and metagenomics are requesting a large amount of computing resources. Many bioinformatics tools, such as alignment tools, can be executed in parallel, exploiting the potential of current e-Infrastructures. In the experience of the GRyCAP in several experiments, it is crucial to minimize job error failure rates. Error failure rates are reduced by achieving a good replication of data sources and by selecting the rightmost partition strategy. By planning carefully the data splitting, it is possible to define execution models that predict the response time and guide the experiments in deciding how to deal with slow jobs. By using those approaches on large infrastructures, the GRyCAP has achieved performance rates of 18000 sequences of 363 nucleotids in average, using BLAST. But this will not be sufficient when new sequencing techniques will be generalized, achieving ratios of more than 20 thousand millions of sequences per week, per machine.

In conclusion, many tools for bioinformatics and medical imaging require an enormous processing time. Data storage and management is becoming a problem that cannot be faced by the individual scientists, requiring the consolidation of multidisciplinary collaborative teams that could exploit current e-Science e-Infrastructures.

Summary

Primary author(s) : Dr. BLANQUER ESPERT, Ignacio (UPV)

Presenter(s) : Dr. BLANQUER ESPERT, Ignacio (UPV)