STORAGE AS A SERVICE AND CLOUD COMPUTING FOR BIOINFORMATICS COMPUTING ENVIRONMENT

Arbex W[1,2], Tagliatti R F[1], Andrade L G D[1], Muniz M N M[2], Guedes E[1], Silva M V B D[1,2]

[1]*Laboratório de Bioinformática e Genômica Animal - Embrapa Gado de Leite*
[2]*Núcleo de Bioinformática de Juiz de Fora - Universidade Federal de Juiz de Fora/Embrapa Gado de Leite*

Cloud computing is a new business model of IT, but it is also a new way of organization of computing resources, which may be available available software, platform, infrastructure, storage area, among others, as well as the end-user aplications, such as services, or better, providing and making these resources to users and clients from simple access procedures. Under the technical aspects the cloud computing is structured from a subset of the distributed systems and operating systems concepts and tools, e. g., transparency of distribution and virtualization. Storage systems are excellent solutions for bioinformatics computing environments, which, in general, the nature of the activity difficult to predict the volume of data to be treated. Furthermore, storage systems and are expensive solutions that require changes in the computing environment to allow the use of this type of storage system. This paper describes a storage solution built for the exclusive use of free software, and within the concept of cloud computing, provides storage as a service (StaaS), implementing transparency of access, location, migration, replication, and concurrence. This storage system was developed using four CPUs and eight HDs under Linux operating system, creating a disk array with capacity of over 4 TB of raw space. To implement the disk array was used MHDDFS and NFS applications, available for Linux and other operating systems. The MHDDFS allows the union of several mount points in a single space, assembling the disk array, and the NFS provides disk array over the network. Integrated into the storage system was implemented a backup/restore system that runs copies all of the contents of storage, four times by day, in external hard drives with network access. The backup/restore system was also implemented with native Linux tools - such as CRON and RSYNC - and has its operation without the intervention or even the user's knowledge. This storage system was implemented with software free features and low cost hardware, but, despite this, the logical organization and the working structure of the cloud computing allows allows scalability and security in storage to bioinformatics computing environments and scientific computing in general.