

R. A.  
patv

IV Encontro Acadêmico de Modelagem Computacional do LNCC  
Petrópolis, Rio de Janeiro, Brasil  
11-13 de Janeiro de 2011

## Storage as a service and utility computing for bioinformatics computing environment: aspects of cloud computing applied to scientific computing

**Wagner Arbex<sup>1</sup>, Marcos Vinícius Barbosa da Silva<sup>1</sup>, Marta Fonseca Martins Guimarães,  
Ricardo Ferreira Tagliatti, Leonardo Gerheim de Andrade, Elizangela Guedes**  
Embrapa Gado de Leite, Rua Eugênio do Nascimento, 610, 36038-330, Juiz de Fora, MG  
E-mail: {arbex, marcos, mmartins}@cnpqgl.embrapa.br, {rictagliatti, general.andrade, elizgguedes}@gmail.com

**Marcelle Nayane Marques Muniz**  
NuBio - UFJF/Embrapa Gado de Leite, Rua Eugênio do Nascimento, 610, 36038-330, Juiz de Fora, MG  
E-mail: marcelleufjf@gmail.com

**Luiz Alfredo Vidal de Carvalho**  
Universidade Federal do Rio de Janeiro, Centro de Tecnologia, Bloco H-319, 21941-972, Rio de Janeiro, RJ  
E-mail: alfredo@cos.ufrj.br

Cloud computing is a new business model of IT, but also a new way of organization of computing resources where can be made available software, platform, infrastructure, storage area, among others, as well as the end-user desktop, such as services, or better, providing and making these resources to users and clients from simple access procedures. Both technical and operational aspects of cloud computing are based on utility computing, which, by its turn, is structured by 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 with 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 system that runs copies all of the contents of storage, four times a day, in external hard drives with network access. The backup system also has 4 TB of workspace, it 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.

<sup>1</sup>Membro do NuBio - UFJF/Embrapa Gado de Leite

SP 5287  
P 167

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.

On its next release, this storage system, will receive improvements in backup mechanism and its storage capacity will be tripled its capacity, so should be expected up to 12 TB, for each of the storage and backup areas. Another feature that should be modified is the access way for users, since new interface and access mechanism will be developed.

**Keywords:** *Bioinformatics computing environment, Cloud computing and Storage system*