

# Sistemas de computación masiva (Clusters) y software libre

<http://gabriel.verdejo.alvarez.googlepages.com/cluster>

Iván Couto Vivas (ivanc@lsi.upc.edu) / Gabriel Verdejo Álvarez (gabriel@lsi.upc.edu)

Julio 2008

---

---

# *Clusters y software Libre*

## ÍNDICE

1. Introducción
2. Entornos propietarios vs Software libre
3. LCLSI: Clusters OpenMosix
4. LCLSI: Arquitectura de clustering v2.0
5. LCLSI: SunGrid + DRBD + Lustre
  - 5.1 Limitaciones GlusterFS
  - 5.1 Lustre + DRBD
  - 5.2 Rembo
6. Benchmarking
7. Sugerencias y preguntas

# 1. Introducción

\* Desarrollo tecnológico-problemas reales

1965 – Gordon Moore

VS

2008 – Google, MySpace

\* Paradigmas de computación

MainFrames (IBM, HP, SUN)

VS

Personal Computers (Universidades)

# 1. Introducción II

Redes de comunicaciones (2008):

LAN

Segmento doméstico Gigabit

Segmento profesional 10Gbit

WAN

Segmento doméstico 20Mbit

Segmento profesional Gigabit

## ***2. Entornos propietarios vs libres***

Hasta la década de los 90:

- \* Soluciones soft+hard propietarias (HP, IBM, SUN)
- \* Primer intento serio de plataforma libre de supercomputación: Beowulf (1994)

## ***2. Entornos propietarios vs libres II***

Desarrollo de múltiples opciones

- \* Beowulf
- \* Condor
- \* Linux Virtual System
- \* ....

Proyectos propietarios pasan a ser libres

- \* OpenMosix
- \* SunGrid Engine

## *LCLSI: Cluster OpenMosix*

Actualmente en LSI tenemos 3 clusters con OpenMosix que suman unos 40 nodos.

- \* Nozomi
- \* Eixam
- \* Tenada

Cada cluster tiene su zona de disco común (nfs) y sus nodos en una red interna a Gigabit.

# *LCLSI: Cluster OpenMosix II*

Sistema de Clustering tipo SSI

OpenMosix migra los procesos entre los diferentes nodos balanceando por la carga

- \* Parte residente del proceso
- \* Parte de ejecución remota



## *LCLSI: Cluster OpenMosix III*

### Inconvenientes del paradigma OpenMosix

- \* No hay limitadores de proceso
- \* Problemas con memoria compartida
- \* Problemas con threads (JAVA...)
- \* Estrechamente ligado a kernel (old)
- \* Proyecto cancelado el 1/4/2008

# *LCLSI: Arquitectura de Clustering 2.0*

Lista de deseos (wishlist):

- \* Un modelo único para los nodos
- \* Un modelo escalable y sostenible
- \* Idiosincrasia de los grupos de investigación
- \* Control de usuarios estricto
- \* Modelo flexible a cambios
- \* Tolerancia a fallos....**¡ESTABILIDAD!**

# ***LCLSI: Arquitectura de Clustering 2.0***

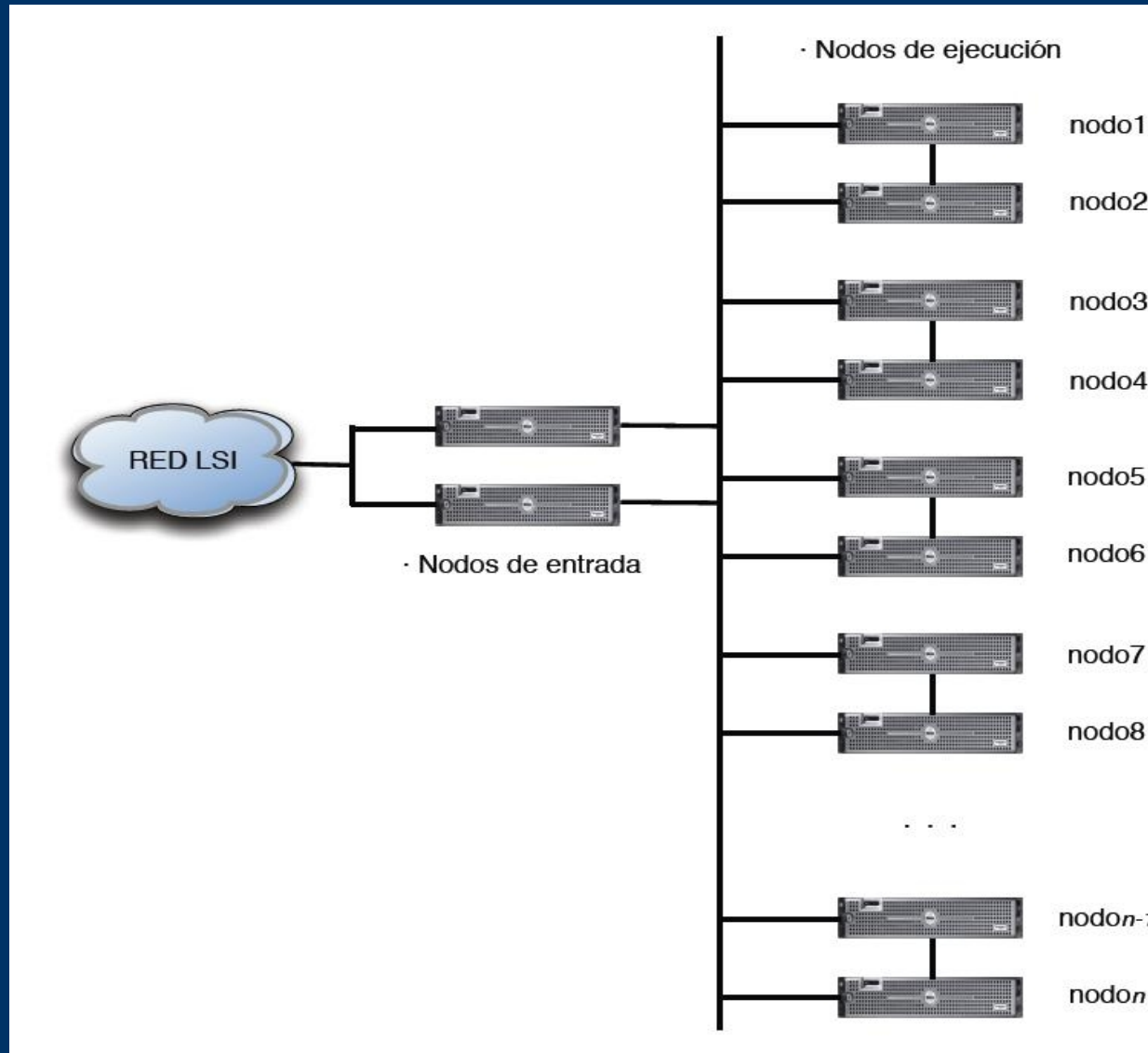
## Características básicas (hardware):

- \* 2 nodos de entrada
- \* N nodos de ejecución de procesos
- \* Dos discos en Raid1 por nodo
- \* Dos tarjetas de red por nodo

## Características básicas (software):

- \* Gestión de trabajos mediante colas
- \* Sistema de ficheros distribuido en red

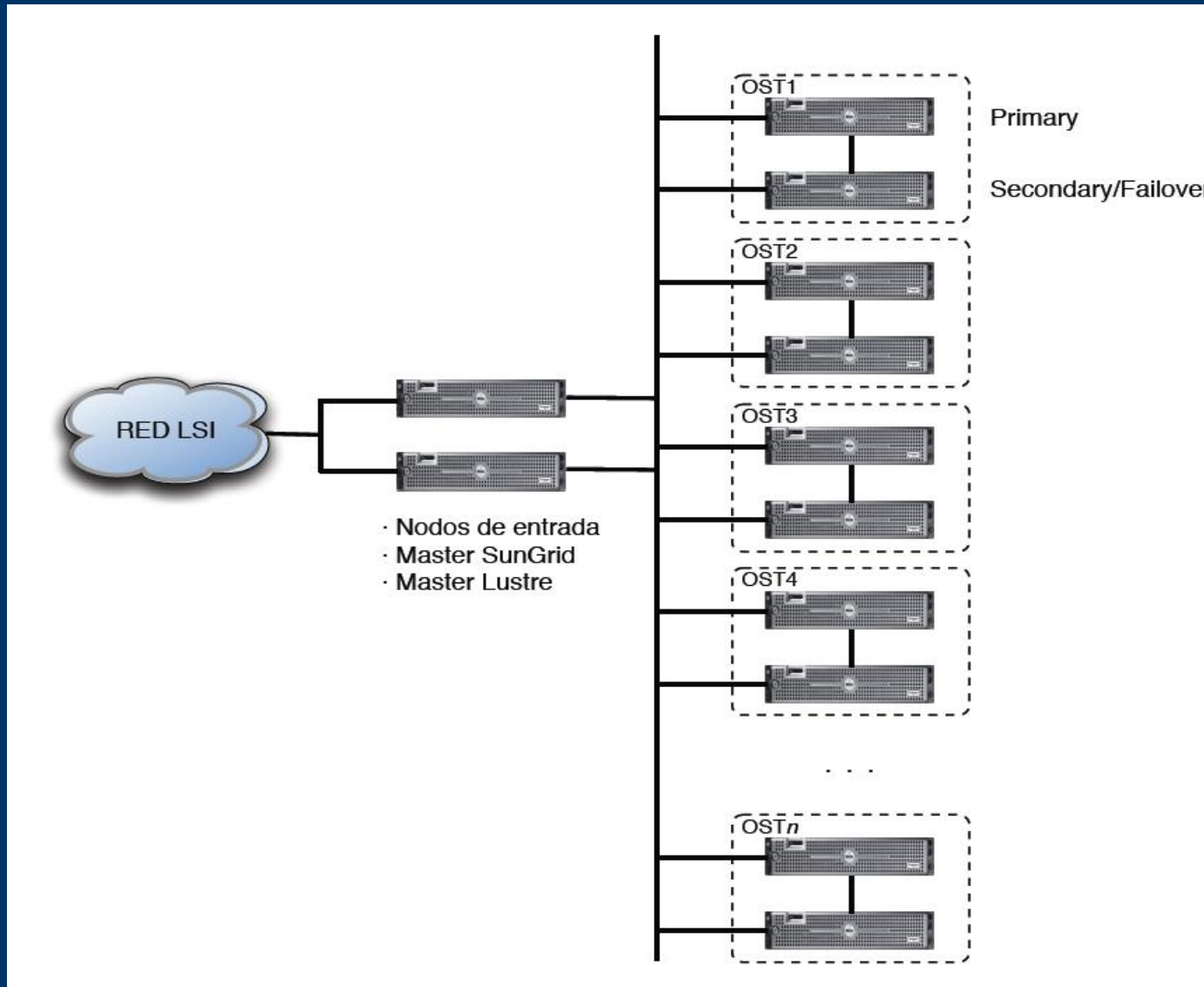
# LCLSI: Arquitectura de Clustering 2.0



## ***LCLSI: SunGrid+DRBD+Lustre***

- Licencias abiertas al uso no comercial
- Proyecto estable y maduro
- Flexible
- \* Gestión de colas: SunGrid Engine
- \* Sistema de ficheros en red: Lustre

# LCLSI: Arquitectura de Clustering 2.0



# *LCLSI: Limitaciones de otras opciones*

Problemas con GlusterFS:

- \* Soporte de bloqueos Posix (Fuse)
- \* Pruebas de stress con AFR
- \* Soporte de los desarrolladores

**NOTA: Nueva versión de Gluster en Julio'08**

NFS y otros sistemas de ficheros quedaron descartados por rendimiento y escalabilidad

# *Benchmarking FileSystem*

\* Se han realizado cientos de pruebas de estabilidad:

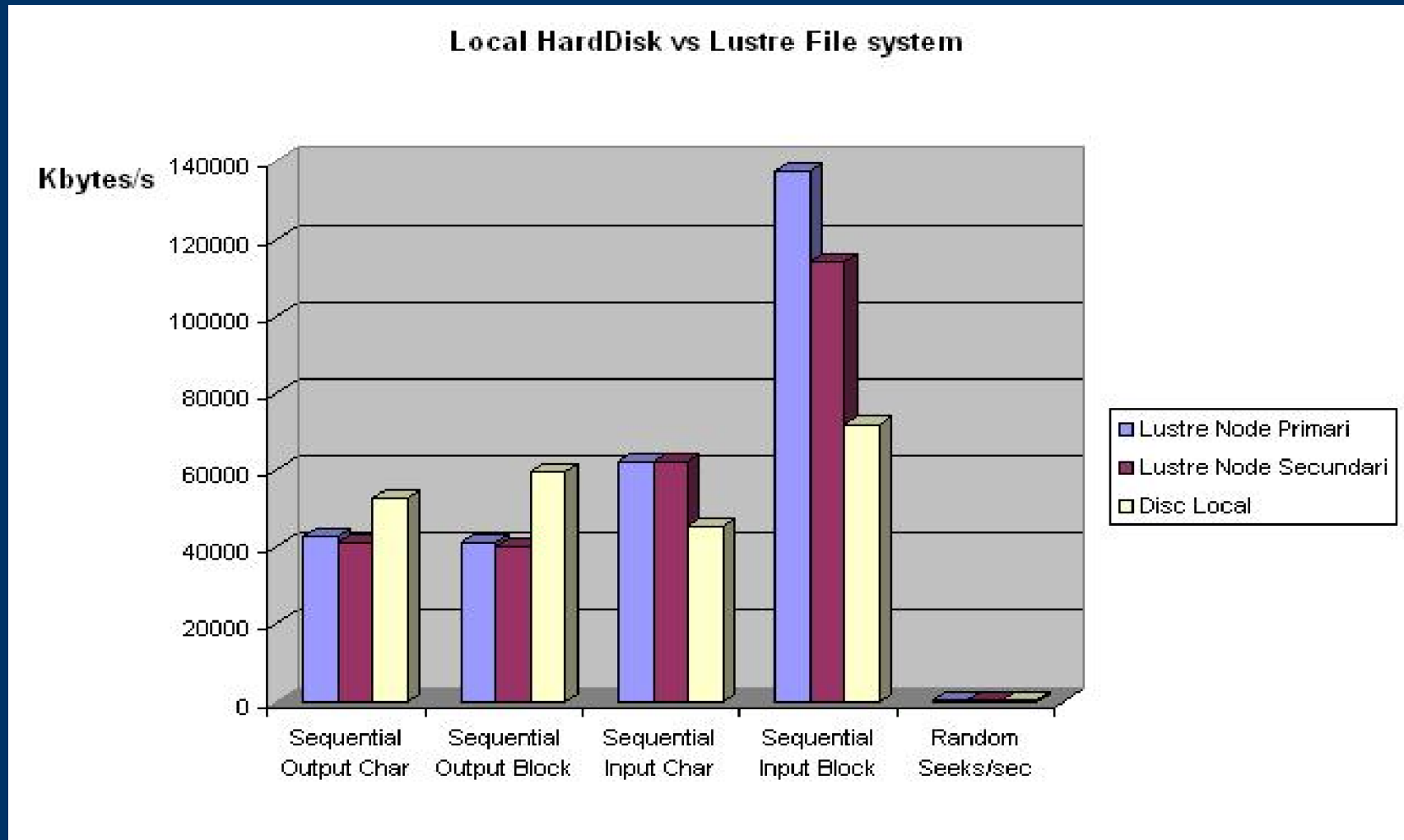
- Ficheros de 1Kb..2Gbytes => 52Gb
- 1024 Directorios por tipo de archivo

\* Bonnie++

```
bonnie++ -s 16386 -n 1:4096:4096:1024 -d <dir> -u <usr>
```

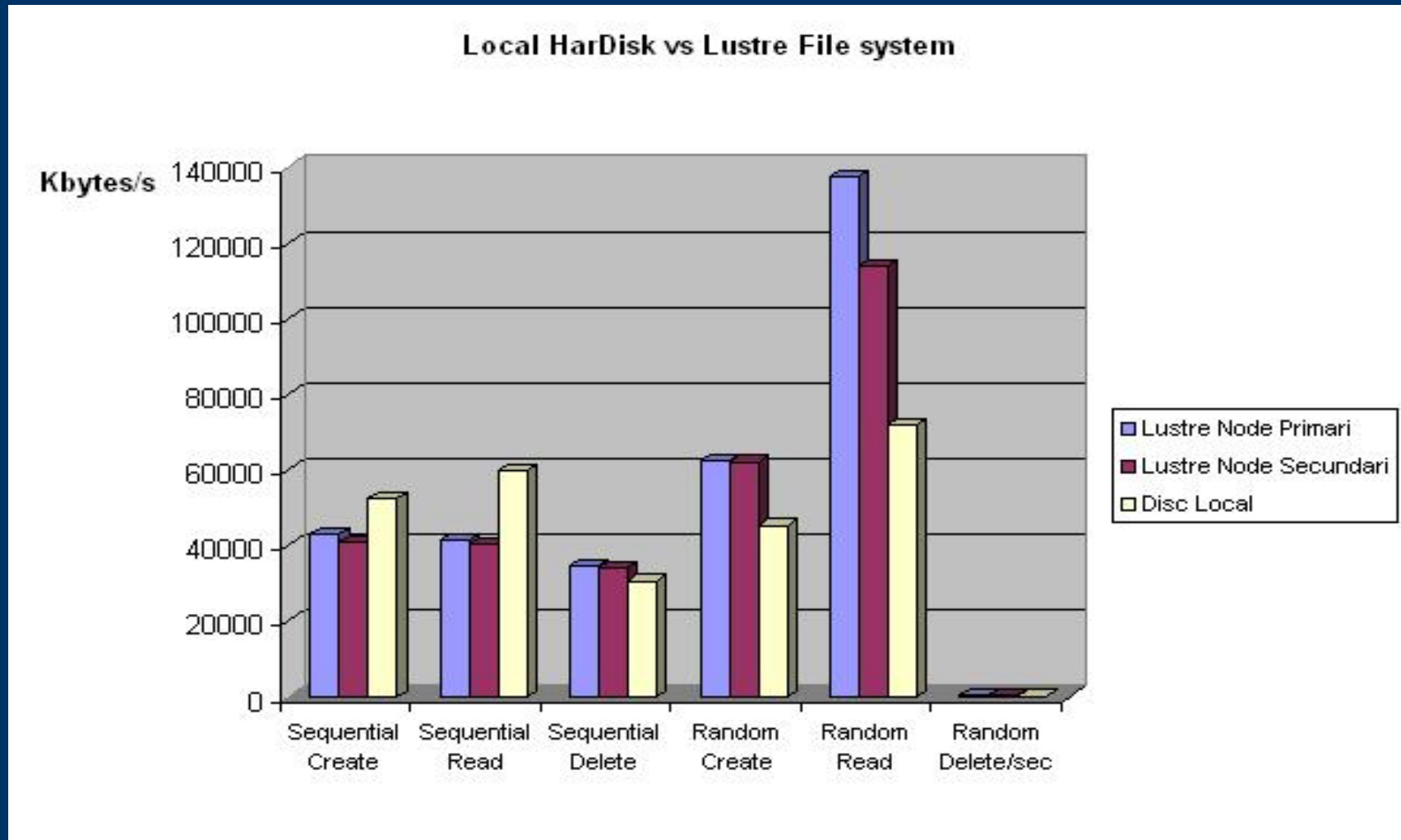


# Benchmarking FileSystem II



Acceso secuencial con un fichero de 16Gb

# Benchmarking FileSystem III



Acceso a directorios y ficheros.

## *LCLSI: Rembo*

Imagen de sistema en los nodos

- \* Misma imagen para todos los nodos
- \* DHCP -> Rembo

NOTA: Actualmente NO existe ningún proyecto libre con la funcionalidad y **estabilidad** necesaria

¿Virtualización y uso de imágenes prefabricadas?

# *Sugerencias y preguntas*

Este proyecto formará parte del PFC de Iván Couto Vivas.

Esta presentación y otra documentación relacionada se encuentra disponibles en:

<http://gabriel.verdejo.alvarez.googlepages.com/cluster>

¡Gracias por vuestra asistencia!