CRG

Centre
de Regulació
Genòmica

BioInfo Good Practices (Storage)

Rodny Hernández

rodny.hernandez@crg.eu

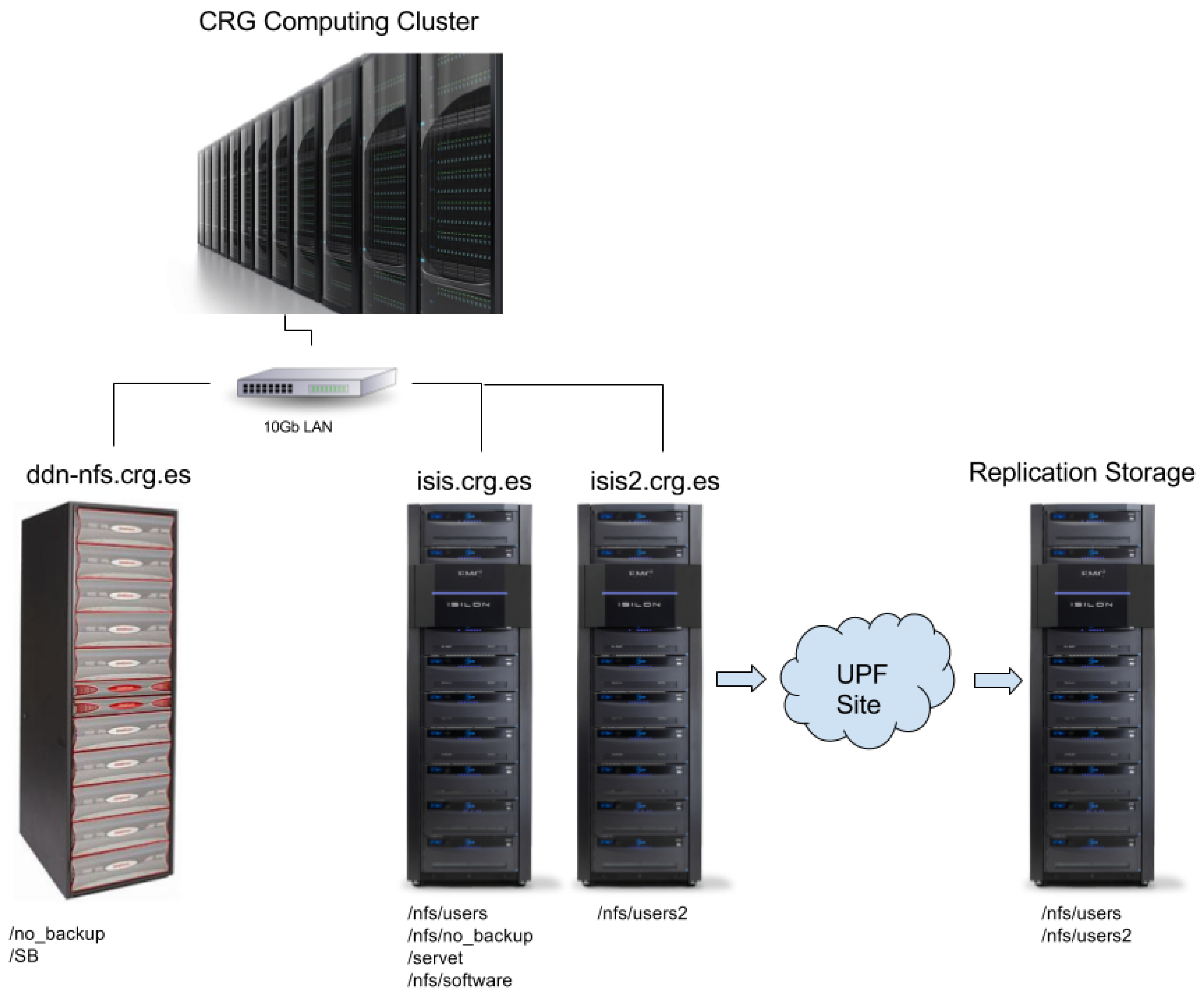
EXCELENCIA
SEVERO
OCHOA

Objetives

- CRG Storage infrastructure description.
- How to recover files.
- File system limits.
- File size and types (Tools).
- Physical vs Logical file size.
- Archiving files.

www.linux.crg.es

CRG Storage Infraestructure.



CRG Computing Cluster

10Gb LAN

ddn-nfs.crg.es

isis.crg.es

isis2.crg.es

UPF Site

Replication Storage

/no_backup
/SB

/nfs/users
/nfs/no_backup
/servet
/nfs/software

/nfs/users2

/nfs/users
/nfs/users2

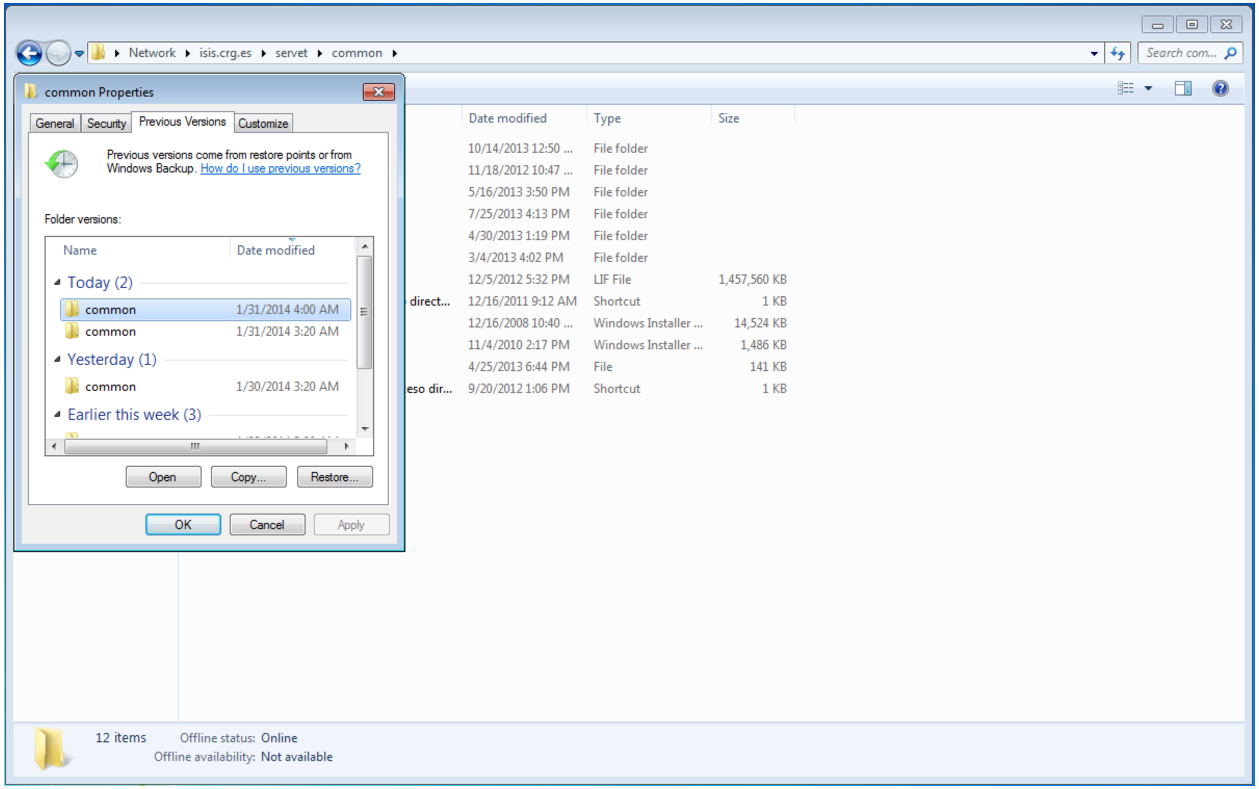
How to recover files (Snapshots)

If you want to capture a moment in time with a camera, you snap a picture. When you want to capture the data on your cluster at a moment, you take a snapshot.

Linux: ‘cd .snapshot/’ (Demo)

How to recover files (Snapshots)

Windows: Right Click, Previuous Versions (Demo)



Previous versions come from restore points or from Windows Backup. How do I use previous versions?

Folder versions:

Name	Date modified
Today (2)	
common	1/31/2014 4:00 AM
common	1/31/2014 3:20 AM
Yesterday (1)	
common	1/30/2014 3:20 AM
Earlier this week (3)	

	Date modified	Type	Size
	10/14/2013 12:50 ...	File folder	
	11/18/2012 10:47 ...	File folder	
	5/16/2013 3:50 PM	File folder	
	7/25/2013 4:13 PM	File folder	
	4/30/2013 1:19 PM	File folder	
	3/4/2013 4:02 PM	File folder	
	12/5/2012 5:32 PM	LIF File	1,457,560 KB
direct...	12/16/2011 9:12 AM	Shortcut	1 KB
	12/16/2008 10:40 ...	Windows Installer ...	14,524 KB
	11/4/2010 2:17 PM	Windows Installer ...	1,486 KB
	4/25/2013 6:44 PM	File	141 KB
eso dir...	9/20/2012 1:06 PM	Shortcut	1 KB

12 items Offline status: Online
Offline availability: Not available

Limits of the file system

Maximum 25,000 files per directory, recommended 2000.

Performance issues due to :

- Large directories and scripts accessing it too often (e.g: looking for new/deleted files).
- Too many concurrent accesses to files in the same directory. Splitting files across multiple directories allows storage requests to be split across multiple nodes in the storage cluster and improve the performance of your jobs.
- Too many jobs doing heavy reads and/or writes at the same time. Even 20 jobs doing a lot of reads/writes to large files can affect performance quite dramatically.

Tools

- Command: ‘du’
- Command: ‘file’
- Web : https://accounting.linux.crg.es/addons/storage/insight/login.php (Demo)

Physical and Logical file size.

- Logical: File size.

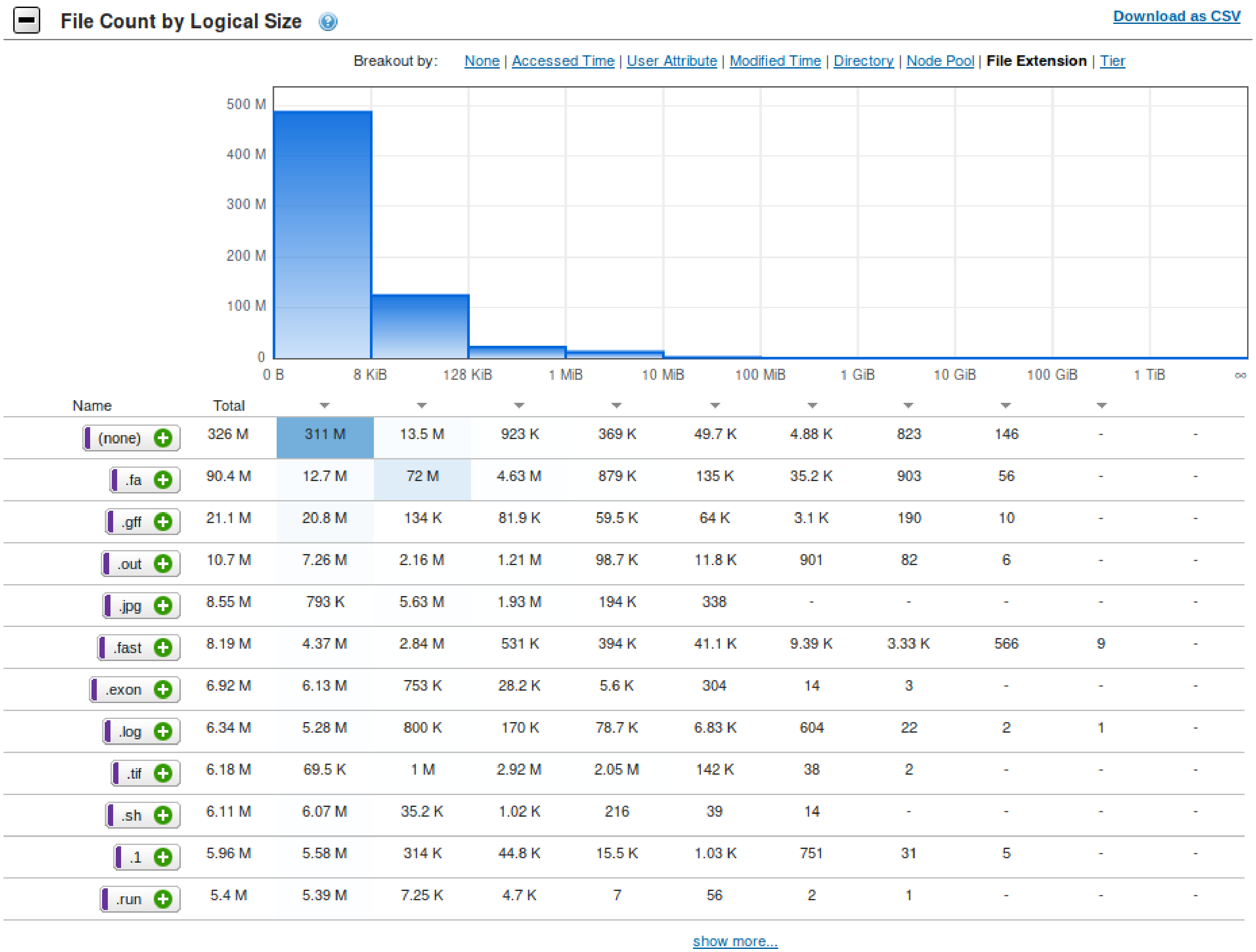
- Physical: Disk size.

  Protection overhead.

  Snapshots.

  MetaData.

Physical and Logical file size.



File Count by Logical Size

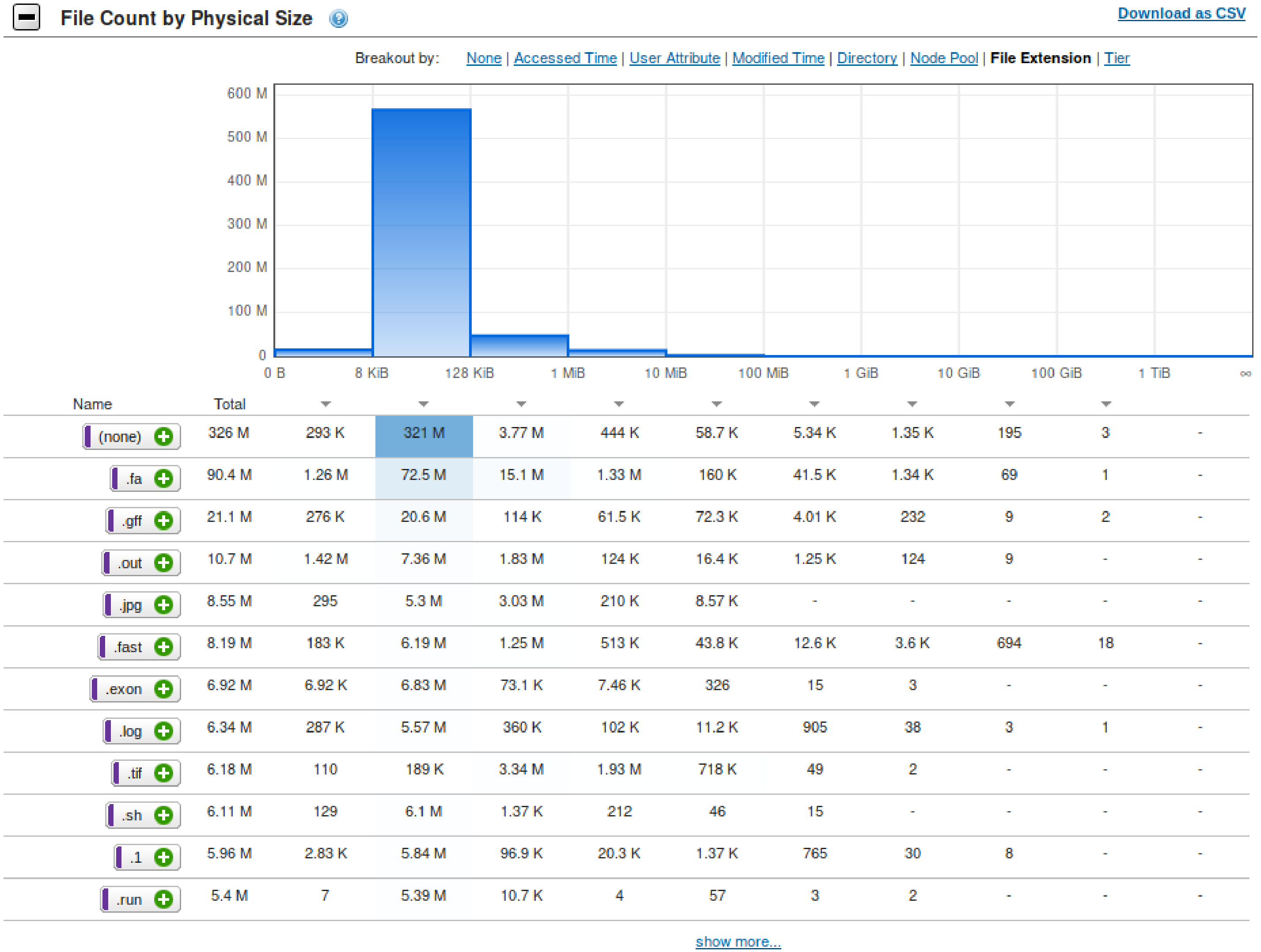
Download as CSV

Breakout by: None | Accessed Time | User Attribute | Modified Time | Directory | Node Pool | **File Extension** | Tier

Name	Total										
(none)	326 M	311 M	13.5 M	923 K	369 K	49.7 K	4.88 K	823	146	-	-
.fa	90.4 M	12.7 M	72 M	4.63 M	879 K	135 K	35.2 K	903	56	-	-
.gff	21.1 M	20.8 M	134 K	81.9 K	59.5 K	64 K	3.1 K	190	10	-	-
.out	10.7 M	7.26 M	2.16 M	1.21 M	98.7 K	11.8 K	901	82	6	-	-
.jpg	8.55 M	793 K	5.63 M	1.93 M	194 K	338	-	-	-	-	-
.fast	8.19 M	4.37 M	2.84 M	531 K	394 K	41.1 K	9.39 K	3.33 K	566	9	-
.exon	6.92 M	6.13 M	753 K	28.2 K	5.6 K	304	14	3	-	-	-
.log	6.34 M	5.28 M	800 K	170 K	78.7 K	6.83 K	604	22	2	1	-
.tif	6.18 M	69.5 K	1 M	2.92 M	2.05 M	142 K	38	2	-	-	-
.sh	6.11 M	6.07 M	35.2 K	1.02 K	216	39	14	-	-	-	-
.1	5.96 M	5.58 M	314 K	44.8 K	15.5 K	1.03 K	751	31	5	-	-
.run	5.4 M	5.39 M	7.25 K	4.7 K	7	56	2	1	-	-	-

show more...

Physical and Logical file size.



⊟ **File Count by Physical Size**

Download as CSV

Breakout by: None | Accessed Time | User Attribute | Modified Time | Directory | Node Pool | **File Extension** | Tier

Name	Total	▾	▾	▾	▾	▾	▾	▾	▾	▾	
(none)	326 M	293 K	321 M	3.77 M	444 K	58.7 K	5.34 K	1.35 K	195	3	-
.fa	90.4 M	1.26 M	72.5 M	15.1 M	1.33 M	160 K	41.5 K	1.34 K	69	1	-
.gff	21.1 M	276 K	20.6 M	114 K	61.5 K	72.3 K	4.01 K	232	9	2	-
.out	10.7 M	1.42 M	7.36 M	1.83 M	124 K	16.4 K	1.25 K	124	9	-	-
.jpg	8.55 M	295	5.3 M	3.03 M	210 K	8.57 K	-	-	-	-	-
.fast	8.19 M	183 K	6.19 M	1.25 M	513 K	43.8 K	12.6 K	3.6 K	694	18	-
.exon	6.92 M	6.92 K	6.83 M	73.1 K	7.46 K	326	15	3	-	-	-
.log	6.34 M	287 K	5.57 M	360 K	102 K	11.2 K	905	38	3	1	-
.tif	6.18 M	110	189 K	3.34 M	1.93 M	718 K	49	2	-	-	-
.sh	6.11 M	129	6.1 M	1.37 K	212	46	15	-	-	-	-
.1	5.96 M	2.83 K	5.84 M	96.9 K	20.3 K	1.37 K	765	30	8	-	-
.run	5.4 M	7	5.39 M	10.7 K	4	57	3	2	-	-	-

show more...

Create and deal with files.

- Small vs Large files.
- Locale and special characters.
- Avoid ‘ls -l’.(size, owner, permissions)
- Command: ‘tar’. (Demo)
- Use relational or not databases.