

# Data protocols issue

M. David

[david@lip.pt](mailto:david@lip.pt)

- The usage of the file protocol on the grid context:
  - **With the lcg\_utils:**
  - `lcg-cp file:/localdir/file1 file:/your_Lustre_Dir/atlas/spacetoken/published/file1`
  - **With GFAL lib:**
  - `gfal_open (file:/your_Lustre_Dir/atlas/spacetoken/published/file1, ...)`

But if you have dCache or DPM or Castor you may well have instead of **file:/**

**gsidcap:/**  
**rfio:/**

- My understanding was that
  - the gridftp protocol would only be used to transfer files between sites. (Not to be used inside the site).
  - That's why jobs end up in the sites which have the data.
  - The site publishes the protocols which they support in the info system. If there is some incoherence, then it should be solved by the site admin (GGUS tickets).
  - For whatever storage system you use, the gftp protocol is only efficient if the total bandwidth (BW) of all your gftp servers balance the total BW of ALL your WN's??
    - The gridftp dCache cell can be configured so that the data connection goes directly from the pool to the WN...

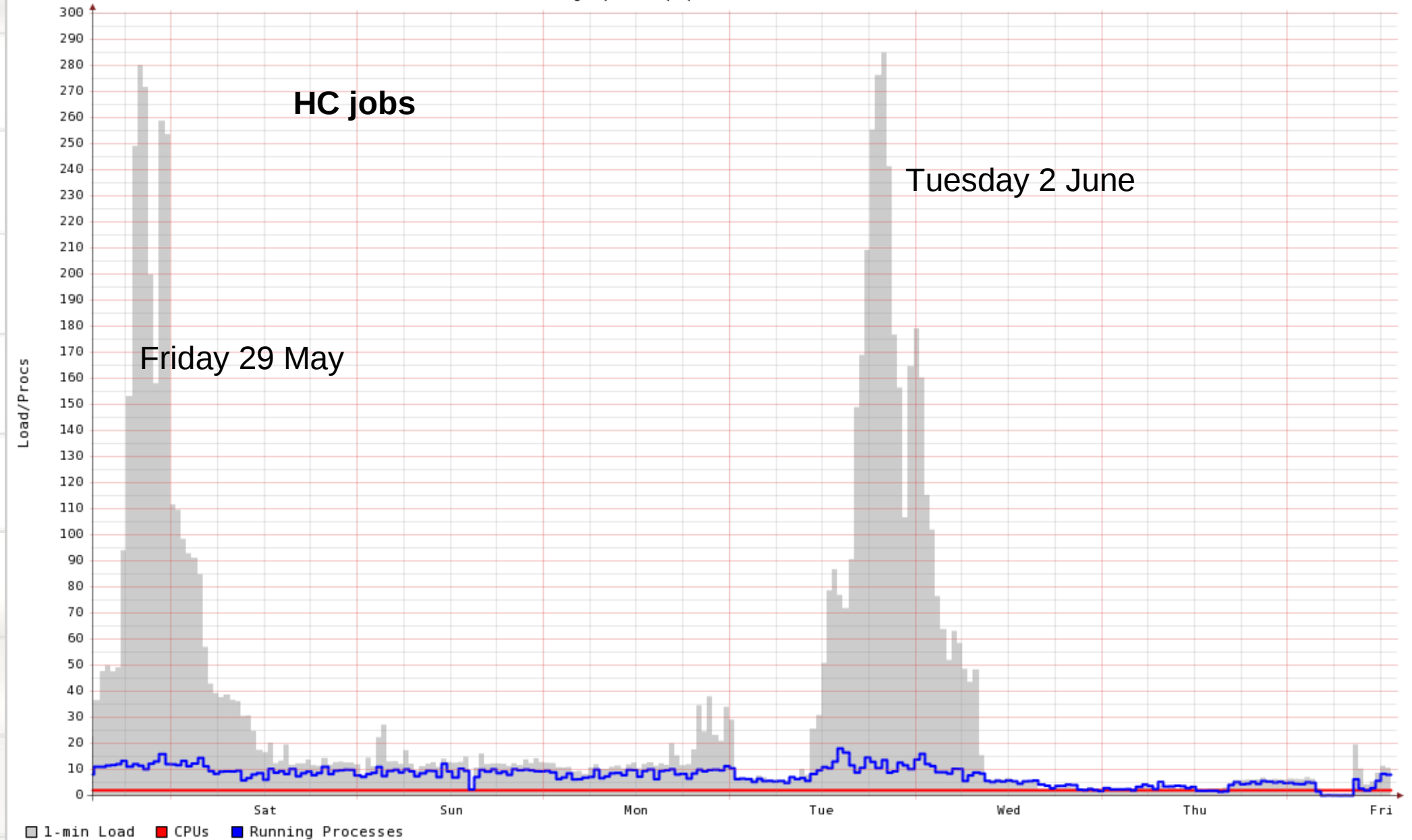
# Distributed Analysis: HammerCloud

- 1 file is needed by all jobs (~200MB size)
  - **lcg-cp'ied by every job.**
    - Solution: replication (dCache has this).
- File stager: input file is first lcg-cp'ied to the WN before job
  - **Quite bad for sites with 1 or 2 gftp servers, as probably the case of most storm T2's.**

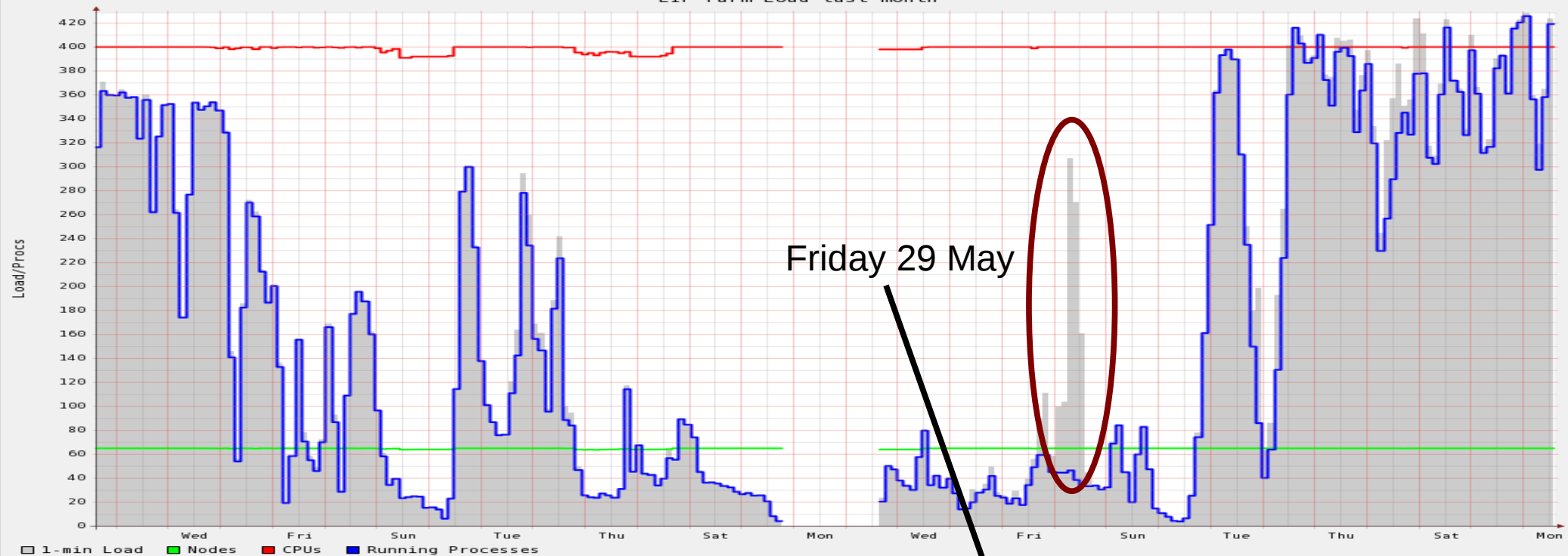
HammerCloud		
DQ2_LOCAL		
Test#	# jobs	Effic. CPU/Wal
383	1345	35.8
385	1382	43.6
387	1762	34.8
389	1532	43.3
402	3031	23.9
404	3943	29.6
406	5682	31.2
408	4094	38.5
FILESTAGER		
Test#	# jobs	Effic. CPU/Wal
384	1459	79.2
386	1216	36.1
388	1734	90.1
390	1350	34.5
403	2685	65.5
405	2902	27.6
407	4837	62.5
409	2938	29

gftp01.lip.pt Load last week

PROTOCOL / T / DETIKER



LIP-Farm Load last month



LIP-Farm CPU last month

