



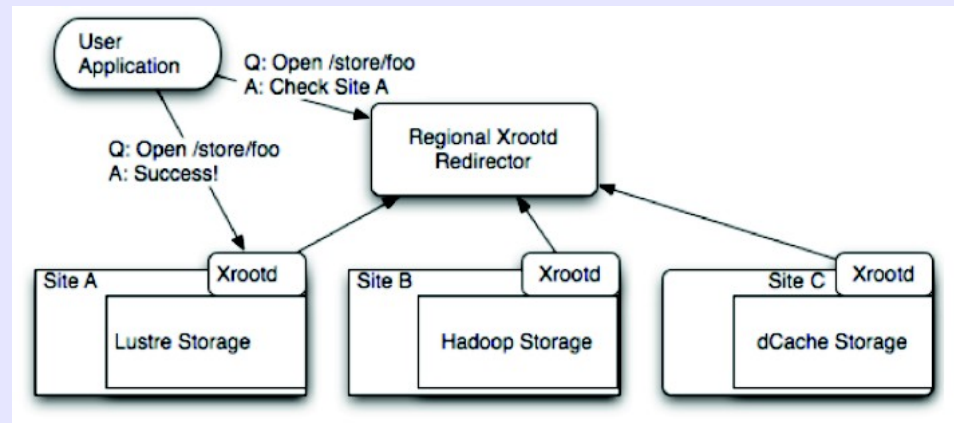
XrootD Scale Testing for AAA

Carl Vuosalo
University of Wisconsin-Madison



Any Data, Anytime, Anywhere

- AAA makes CMS data available **transparently** at **any** CMS site
- Utilizes **XrootD** to provide uniform interface for multiple storage systems (dCache, Hadoop, etc.)
- Applications query XrootD redirector to find files
 - Redirector then queries sites to find the files and caches results for future use





AAA Scale Testing

- Scale testing measures ability of CMS T2 sites to handle predicted peak loads for AAA
- Tests emulate CMS jobs running at CMS sites
- Two measurements performed:
 - › Rate to open files
 - › Rate of reading data from files
- Six US T2 sites successfully tested:
 - › Caltech, Florida, MIT, Nebraska, UCSD, Wisconsin
- T2_US_Purdue and T2_US_Vanderbilt working on improving performance
- Testing started on European T2 sites



Scale Testing: File Opening

- File-opening test measures rate files at site can be opened via redirector
- Test runs up to 100 jobs simultaneously that open files at rate of 2 Hz each, so highest total rate is 200 Hz
- Projected maximum site load is 10^5 jobs opening files at a rate of 10^{-3} Hz each
 - Gives maximum total rate at a site of **100 Hz**, which becomes **target rate** for the test
 - Higher rates not expected under real conditions



TFC Change for Scale Testing

- Need a way to ensure scale tests are accessing files local to the tested site
- Solution: Sites use Trivial File Catalog (TFC) trick* to allow file access by names with the form
 - */store/test/xrootd/SITENAME/LFN*
- This TFC change can be implemented on various storage systems
 - Tested sites use dCache, DPM, Hadoop, Lustre, or StoRM
- Tests always access files via a redirector:
 - Nebraska for US sites
 - Bari for European sites

*<https://twiki.cern.ch/twiki/bin/view/Main/XrootdTfcChanges>



XrootD Configuration for Performance



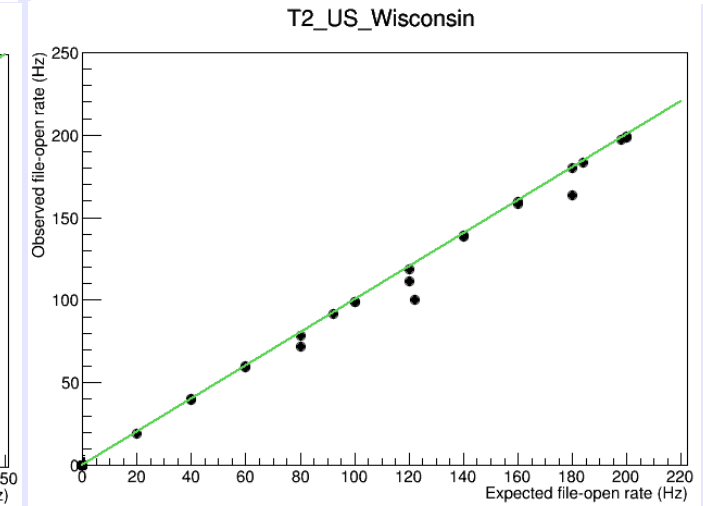
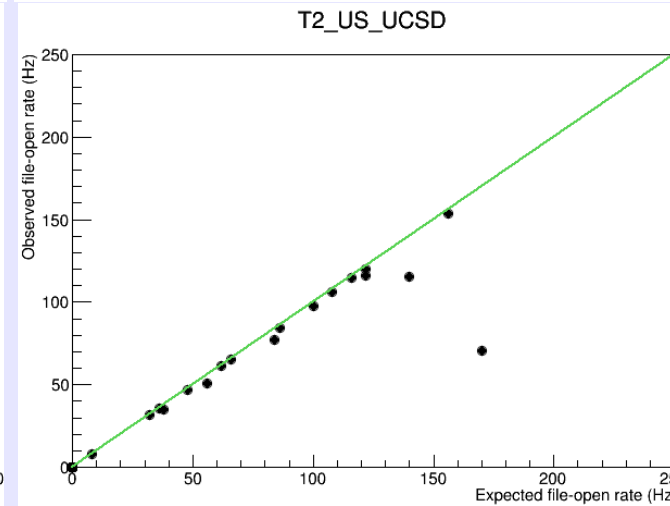
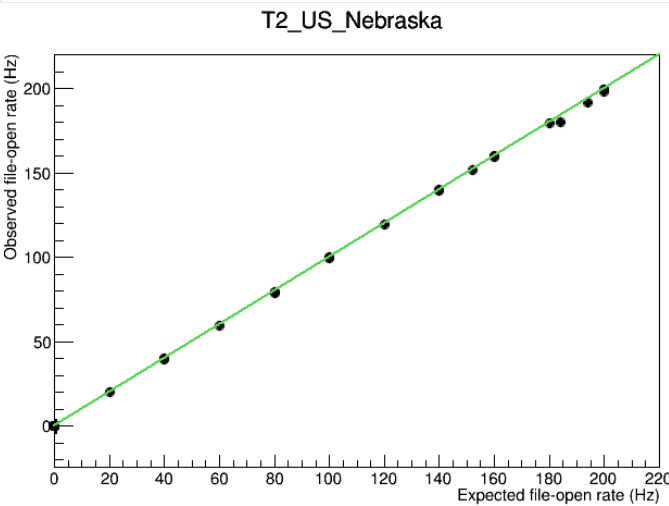
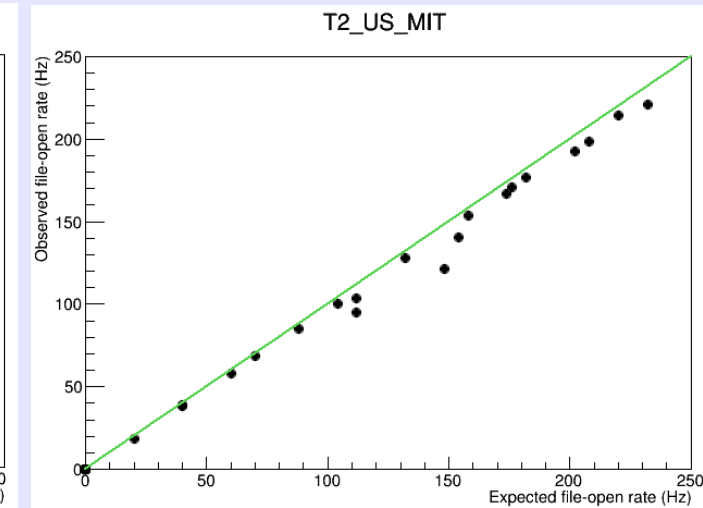
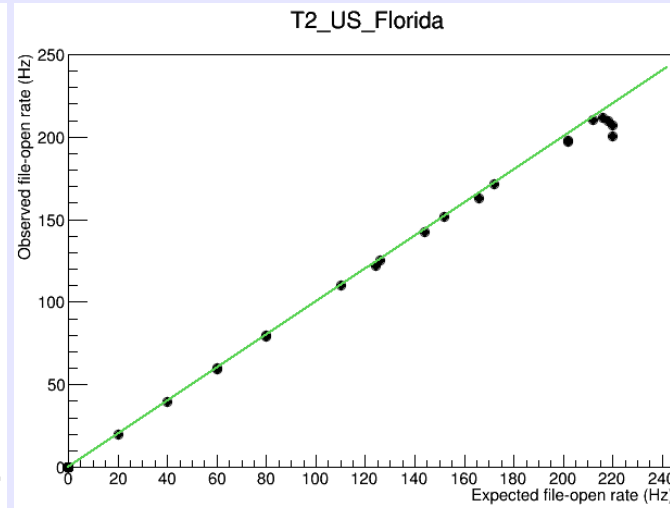
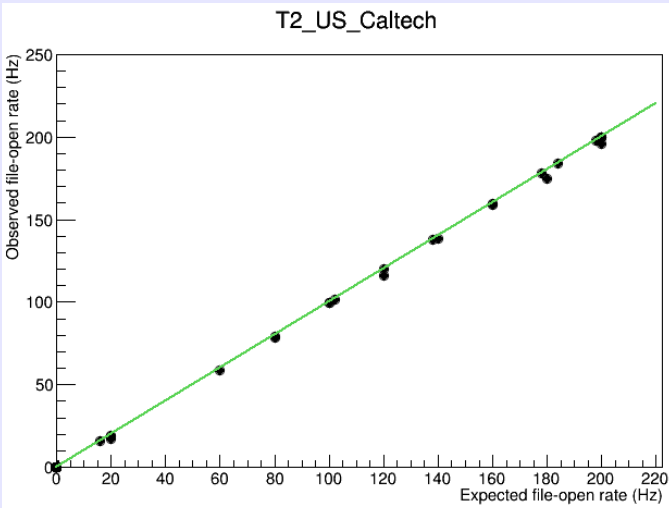
- xrootd.cfg has configuration directive `cms.dfs` for distributed file system handling
- Performance on file-open test greatly affected by this directive
- `cms.dfs lookup central` gives very poor performance
- Change to `cms.dfs lookup distrib` to get good performance
- `distrib` means file existence checked by data server nodes
- `central` means it's checked by the manager node



File-opening Results (US)



Plots show attempted file-open rate vs. observed rate. Ideal is observed = attempted (green line)



All six sites achieve 100 Hz target

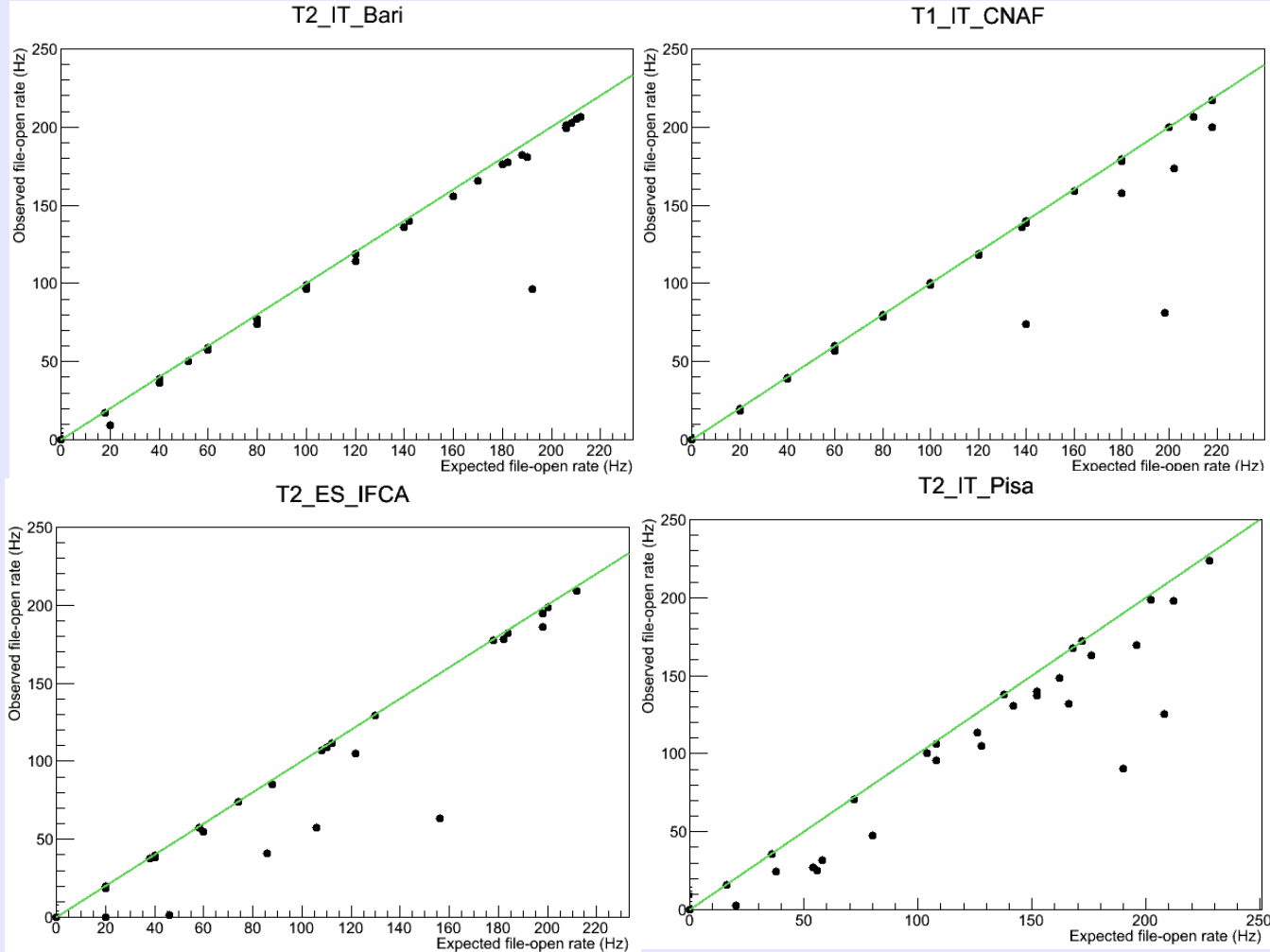
Carl Vuosalo



File-opening Results for Europe (1)



Plots show attempted file-open rate vs. observed rate. Ideal is observed = attempted (green line)



These sites use StoRM

Thanks to Federica Fanzago for plots

Pisa plot has many stray points -- should be re-tested

These sites achieve 100 Hz target

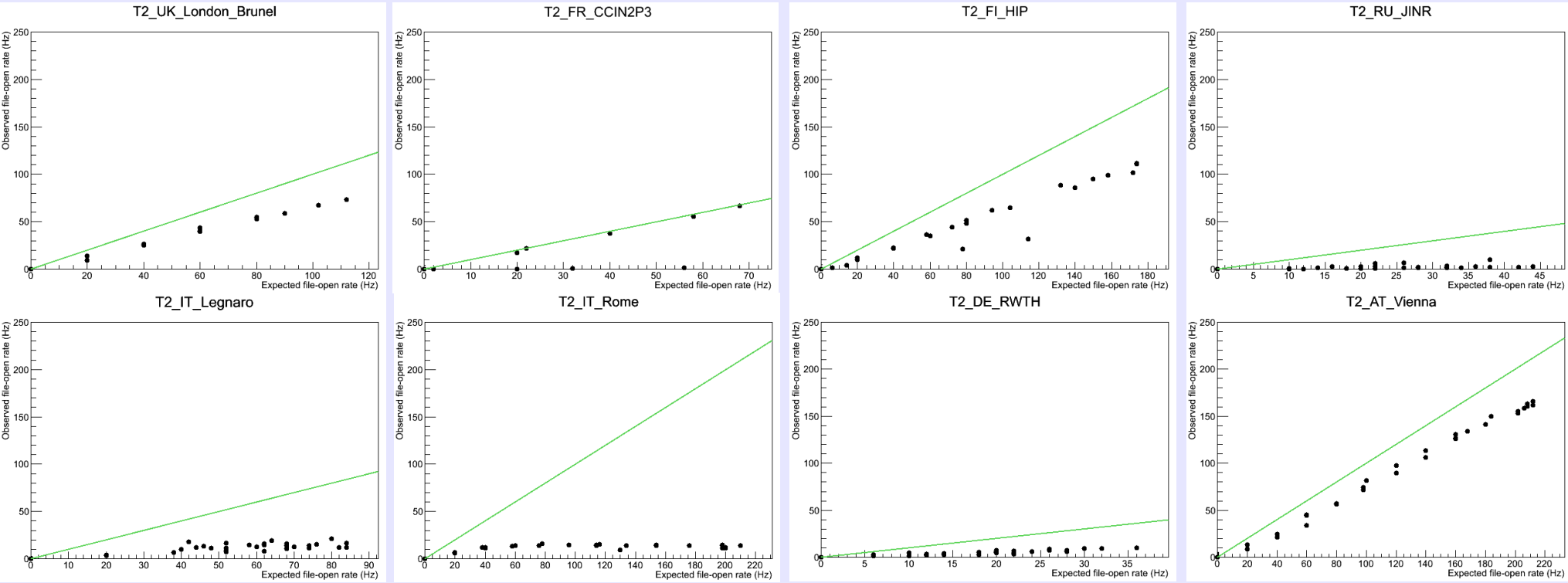
Carl Vuosalo



File-opening Results for Europe (2)



Plots show attempted file-open rate vs. observed rate. Ideal is observed = attempted (green line)



Still investigating why these sites don't achieve target
These sites use dCache or DPM -- related to bad performance?

Thanks to
Federica
Fanzago
for plots



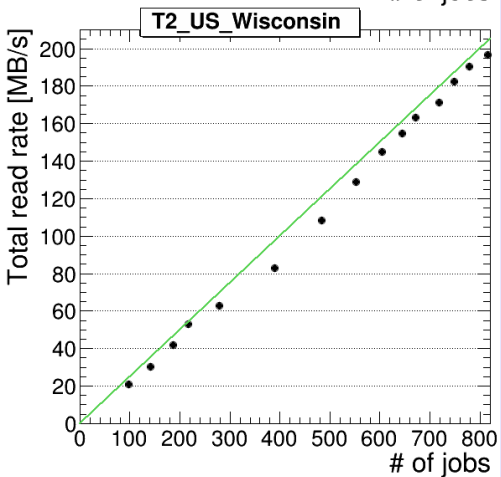
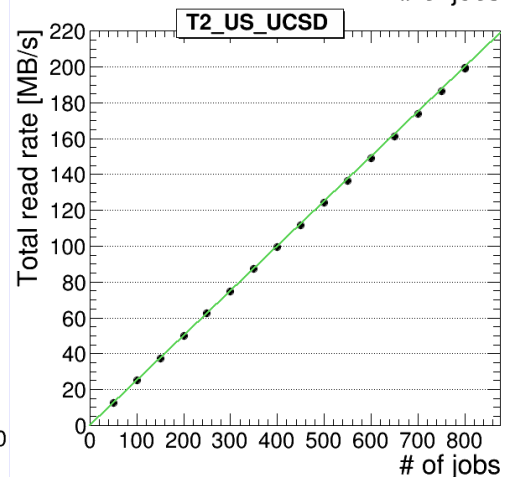
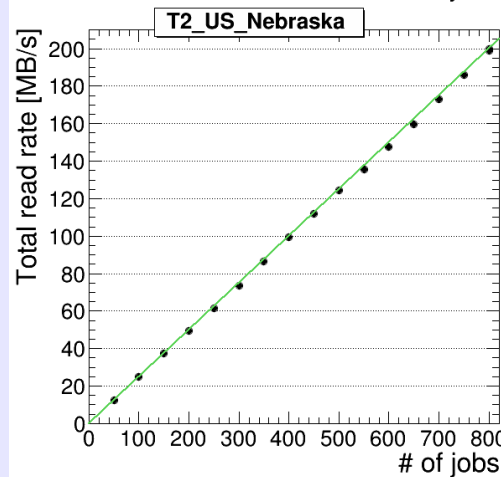
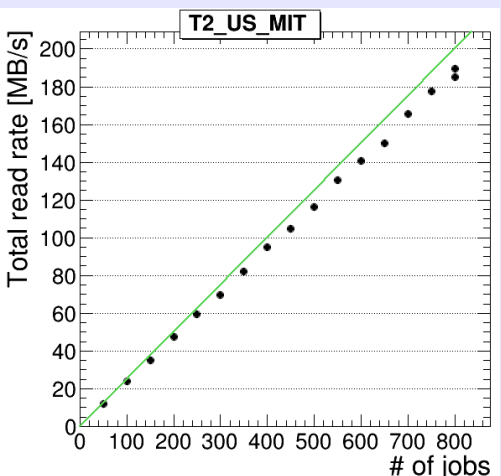
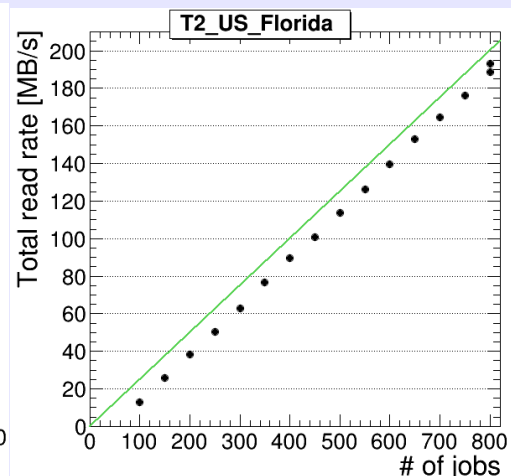
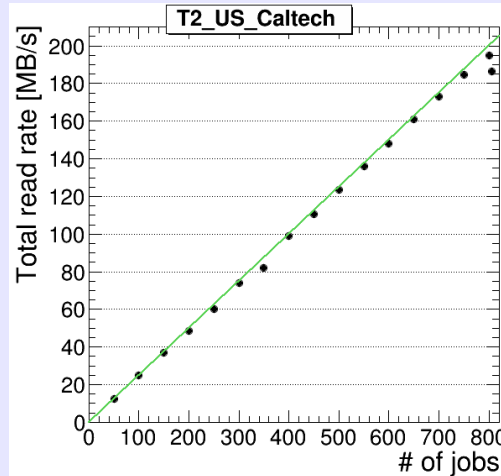
Scale Testing: File Reading

- File-reading test measures rate data can be read from files at site opened via Nebraska redirector
- Test **emulates real CMS jobs**, which show average read rate of 2.5 MB every 10 seconds
- Target performance is 600 jobs reading at this average rate
- Test runs up to 800 jobs that sleep between reads so each job maintains constant read rate of 2.5 MB per 10 seconds
- Tests run from Wisconsin except for test on Wisconsin files that was run at Nebraska

File-read Test - Total Rate

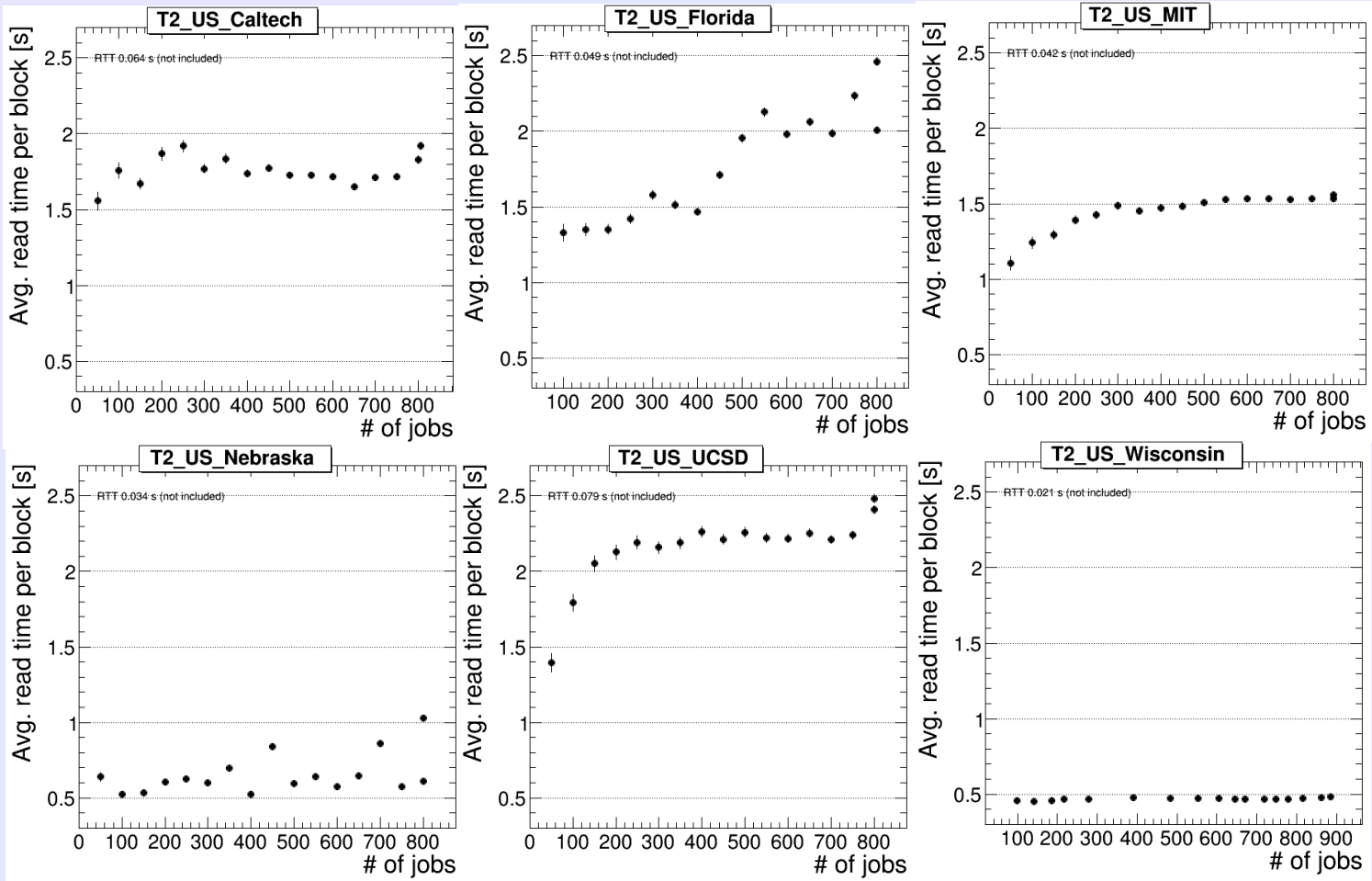


- Plots show **total read rate** for all jobs – should follow **green line**
- All sites show **good performance**
- Deviations from line probably due to high machine loads and Unix job scheduling effects during tests





File-read Test - Avg. Read Time



- Plots show average read time per 2.5 MB block (lower is better)
- Read time ranges from 0.47 to 2.2 s for different sites
- Round-trip time is not included in the read time



Improved File-read Test

- Planning new file-read test that will perform vector reads
- Real CMS jobs perform random-access reads throughout file
 - Current file-read test only performs consecutive block reads
- New file-read test will emulate this random-access read behavior
- Preliminary results very similar to block-read test results



Daily Site Monitoring

- Low-rate file-opening and file-reading tests performed **automatically every night** on six US T2 sites
- Output logs found at <http://www.hep.wisc.edu/cms/aaa/sitemonitoring>
- Log reports for each site number of successfully opened files, number failed, and average read time per 2.5 MB block
- Site problems indicated by:
 - File-open failures $> 6\%$ of successes
 - Block read time > 3 s



Daily Test Results To Date



Site	24-3	25-3	26-3	28-3	29-3	30-3	31-3	1-4	2-4	3-4	4-4	5-4	6-4	7-4	8-4
Caltech	N/A	N/A	N/A	N/A	W	G	G	G	F	F	W	G	G	G	G
Florida	W	W	W	G	G	W	G	G	W	G	W	G	F	G	G
MIT	W	W	G	G	F	F	F	G	W	F	W	W	F	F	G
Nebraska	G	G	G	G	G	G	G	G	G	W	G	G	G	G	G
UCSD	G	G	G	G	G	G	G	G	G	W	W	G	G	G	G
Wisconsin	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G

Key	
F	Fail -- no files could be opened
G	Good performance
W	Warning – very poor performance



Scale Testing: Plans

- Work with local experts to improve results from T2_US_Purdue and T2_US_Vanderbilt
- **European** site tests underway now in Italy
- Expanding testing to **T1** sites in April
- Start **client-hosting** tests in April
 - › Measure # of jobs using remote access that a site can run
 - › Similar to file-reading test



Scale Testing: More Plans

- Total chaos test (multiple sites together) during **CSA14**
- In later phase of scale testing, may use CMS analysis jobs for tests rather than programs that emulate CMS jobs
- Scale test non-CMS sites that provide opportunistic use of computing resources
- Include daily test results in **Site Status Board (SSB)**



Summary

- AAA scale tests assess capability of sites to handle predicted loads
- Tests measure **file-opening** and **file-reading** rates
- Six US T2 sites performed well on tests:
 - Caltech, Florida, MIT, Nebraska, UCSD, Wisconsin
- Tests performed **daily** to monitor site status
- Expansion of tests to **Europe** and **T1** sites in progress
- **Additional** types of tests planned