

Published on *CERN openlab* (<http://test-static-05.web.cern.ch>)

[Home](#) > Integrity Checking and Monitoring of Files on the CASTOR Disk Servers

---

## Integrity Checking and Monitoring of Files on the CASTOR Disk Servers <sup>[1]</sup>

**Date published:**

Wednesday, 17 August, 2011

**Document type:**


Summer student report

**Author(s):**

H. Lien

We proposed to review the code that check the integrity of the data we keep on disk against the corresponding metadata (checksum). The project was called "Error Scanning of the Petabyte Disk Store for LHC" and it had as starting point an existing procedure. The drawbacks of the existing procedures were on the little efficiency (in order not to interfere with normal I/O the number of files checked per day was extremely low). In addition the output (logging information) were not sufficient to study systematic effects (like correlations between host, model and failures (these failures being extremely rare). As a first step he took the old code and drastically improved performances by detecting disk activities in an efficient way. He improved the logging and within few weeks he delivered a new version which entered production (and it is still in production now). The program is now part of the standard disk server set of services and as such is deployed across more than 15000 disk serves. After that he developed a Django portal to allow visualization of the data and basic analysis (failures vs time, host name and host model). This is a nice prototype which he eventually extended to browse across other log files of CASTOR in an efficient way. The data from the checksum verification are routinely mined in our log repository and we regularly scan them for detecting pathological case and to compute error statistics (this is part of the activity of a technical student in the group).

**Technical document file:**

 [Lien Hallgeir's Report.pdf](#) <sup>[2]</sup>

- [Visit Us](#)
- [RSS Feeds](#)

DISCLAIMER: This Web page contains pointers to material related to the management of CERN openlab in the Information Technology Department at the European Organization for Nuclear Research (CERN). Their use and distribution are regulated by the CERN copyright notice.



---

**Source URL:** [http://test-static-05.web.cern.ch/publications/technical\\_documents/integrity-checking-and-monitoring-files-castor-disk-servers](http://test-static-05.web.cern.ch/publications/technical_documents/integrity-checking-and-monitoring-files-castor-disk-servers)

**Links**

[1] [http://test-static-05.web.cern.ch/publications/technical\\_documents/integrity-checking-and-monitoring-files-castor-disk-servers](http://test-static-05.web.cern.ch/publications/technical_documents/integrity-checking-and-monitoring-files-castor-disk-servers)

[2] [http://test-static-05.web.cern.ch/sites/test-static-05.web.cern.ch/files/technical\\_documents/Lien%20Hallgeir%27s%20Report.pdf](http://test-static-05.web.cern.ch/sites/test-static-05.web.cern.ch/files/technical_documents/Lien%20Hallgeir%27s%20Report.pdf)