



Contribution ID: 123

Type: **Plenary**

## A Survey on Distributed File System Technology

*Tuesday 2 September 2014 09:00 (35 minutes)*

Distributed file systems provide a fundamental abstraction to location-transparent, permanent storage. They allow distributed processes to co-operate on hierarchically organized data beyond the life-time of each individual process. The great power of the file system interface lies in the fact that applications do not need to be modified in order to use distributed storage. On the other hand, the general and simple file system interface makes it notoriously difficult for a distributed file system to perform well under a variety of different workloads. This has led to today's landscape with a number of popular distributed file systems, each tailored to a specific use case.

This contribution provides a survey of distributed file systems and key ideas of their internal mechanics. Early implementations merely execute file system calls on a remote server, which limits scalability and resilience to failures. Such limitations have been greatly reduced by modern techniques such as distributed hash tables, content-addressable storage, distributed consensus algorithms, or erasure codes. In the light of upcoming scientific data volumes at the exabyte scale, two trends are emerging. First, the previously monolithic design of distributed file systems is decomposed into services that independently provide a hierarchical namespace, data access, and distributed coordination.

Secondly, the segregation of storage and computing resources yields to a storage architecture in which every compute node also participates in providing persistent storage.

**Primary author:** BLOMER, Jakob (CERN)

**Presenter:** BLOMER, Jakob (CERN)

**Session Classification:** Plenary

**Track Classification:** Computing Technology for Physics Research