



SCALABLE GLOBAL GRID CATALOGUE FOR RUN3 AND BEYOND

CURRENT ALIEN FILE CATALOGUE

The AliEn file catalogue provides mapping between a UNIX-like logical name structure and the physical files stored in 80+ storage elements worldwide. In production since 2005, the size today is more than 2 billion logical file names. The ALICE upgrade in Run3[5] will bring 20 fold increase of namespace and access frequency. To anticipate to this growth, we are investigating new DB technologies as replacement of the current set of classical relational databases.

NEW DB SCHEMA AND TECHNOLOGIES

- **Database selection:** Cassandra is a free and open-source distributed database for managing large amounts of structured data across many servers. It provides the crucial key points needed for our use-case: fault-tolerant and linear scalability, high availability and operational simplicity and tunable consistency.
- **User interface:** CVMFS (CERN virtual filesystem) is used across the Grid as a software distribution service. The project will re-purpose CVMFS to present the file catalogue to the user with a familiar view of a standard filesystem.

BENCHMARK RESULTS AND FUTURE WORK

- We have compared MySQL and Cassandra performance both inserting and retrieving the information of logical and physical file names, and using different consistency levels. Besides, we also show how Cassandra scales operations by increasing the query load.
- We will use the query workflow analysis to develop optimal Cassandra schema design. The AliEn code will be refactored to work with Cassandra. The full catalogue information will be inserted in the new database followed by extensive performance and stability tests.