CNAF flooding: effects on ALICE

GDB, May 9, 2018

L. Betev, M. Litmaath

v1.0

CPU and disk data

- No large effect from the loss of CPU
  - Compensated by additional CPU offered at KIT and CERN
- The unavailability of 10% of unique files (single copy) resulted in reduction of statistics for certain analyses
- After the re-activation of the CNAF storage, the files were found to be intact
- The files which had replicas elsewhere were replicated to other storages, in the order of importance for analysis
  - To preserve the turnaround speed of the organized analysis and to avoid bottlenecks

Tape data

- Data on damaged tapes has one replica at CERN
  - Mindful of the high cost of tape restoration, we have signaled to CNAF that no attempt should be made to recover those tapes
  - ALICE will replicate again the lost data files once these are scheduled for reprocessing
    - To avoid re-spinning the tapes twice
  - In addition, about 1 PB of 2017 raw data still due to be replicated at CNAF will be copied similarly

Lessons & conclusions

- The negative effects were largely mitigated by two factors:
  - Compensation of CPU provided by KIT and CERN
  - The ALICE data placement rules - *randomly* distributing data sets across sites
    - No large dataset is ever stored on a single SE in its entirety
    - This incident strongly validates the rule
- If there was sufficient space for a second replica of all production datasets, the 10% statistics penalty could have been avoided
- About 2 FTE weeks are estimated to have been spent by ALICE experts on the recovery from the incident so far
  - Some work still needed for the re-replication of tape data