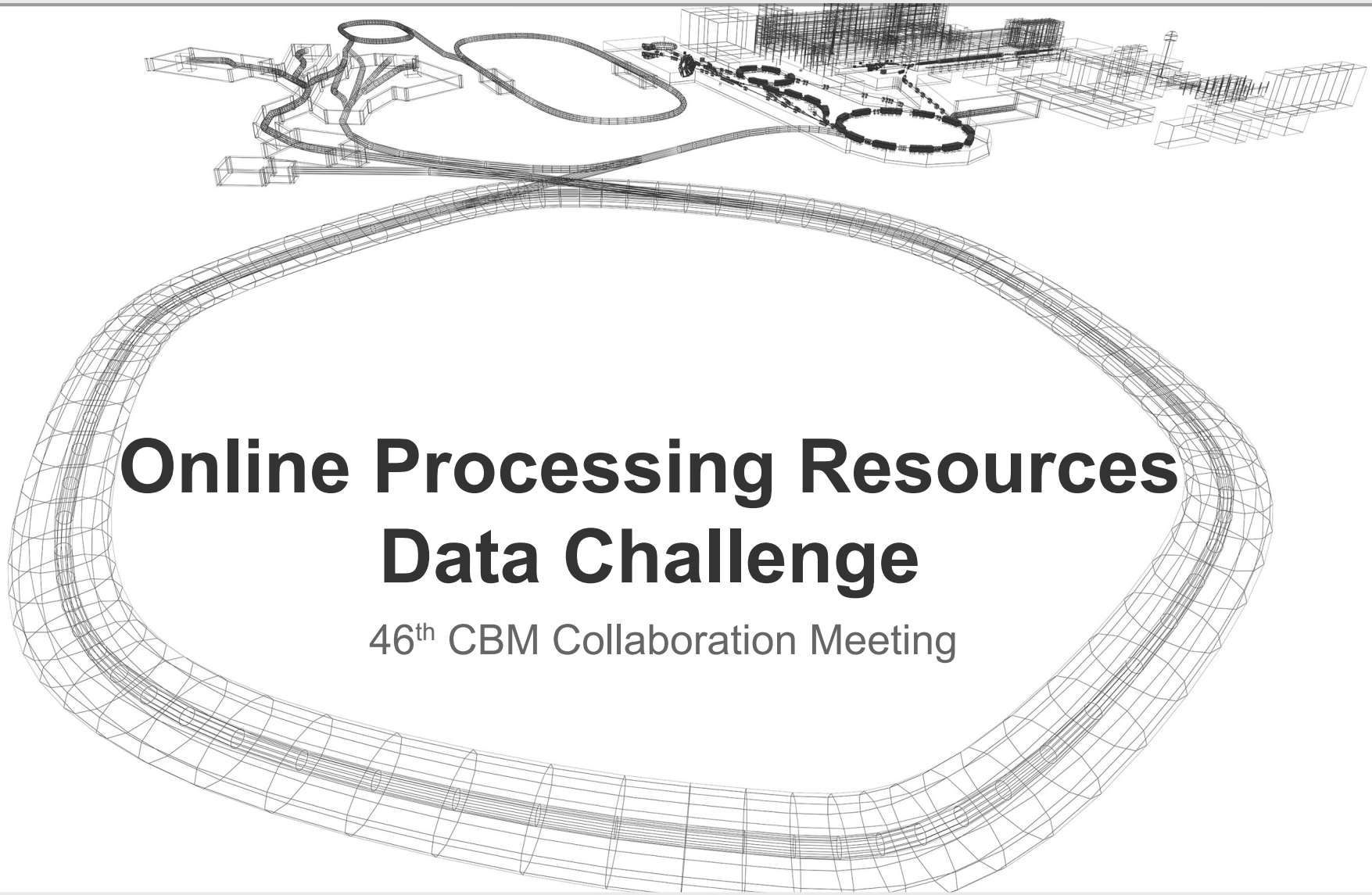
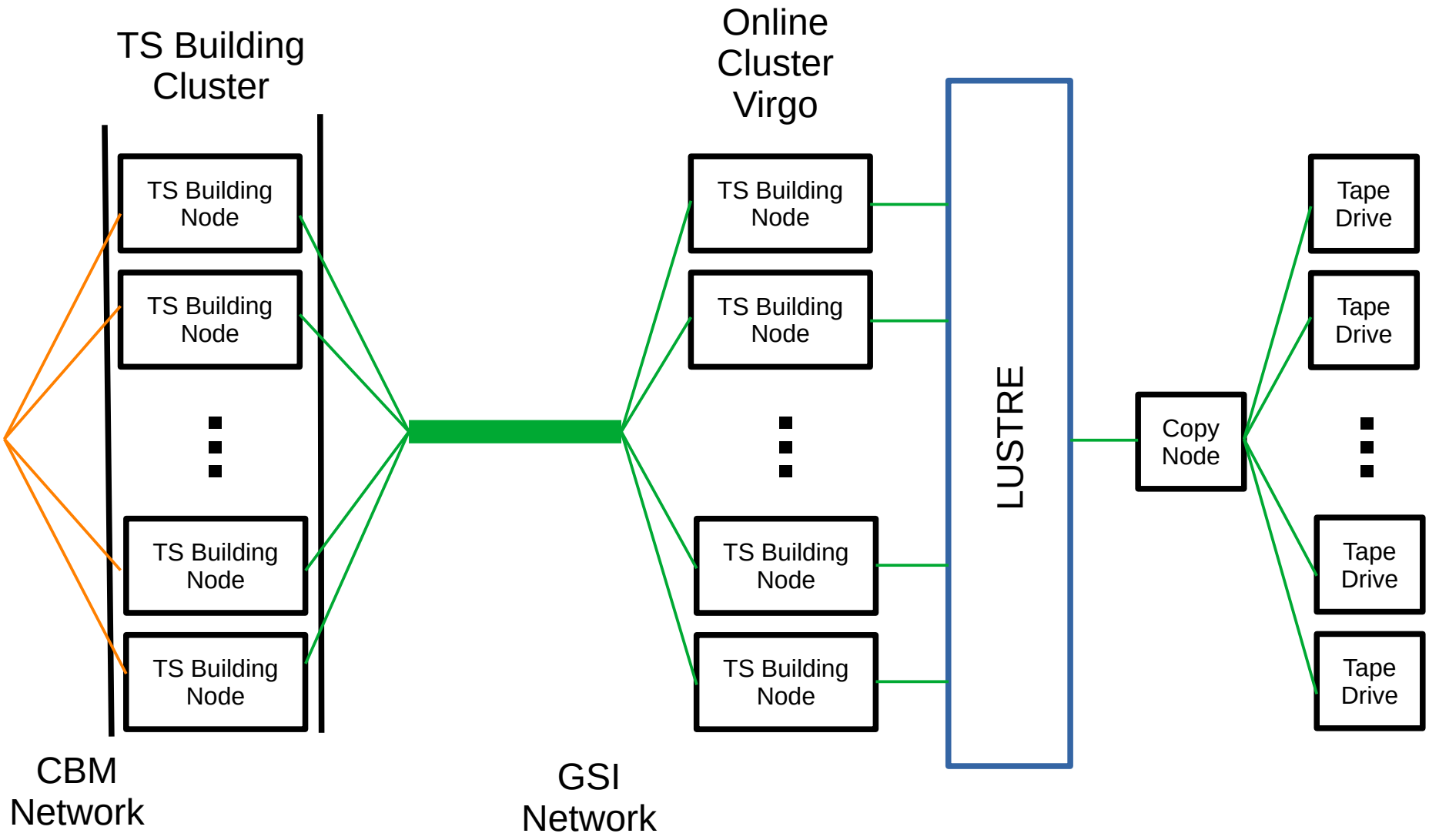


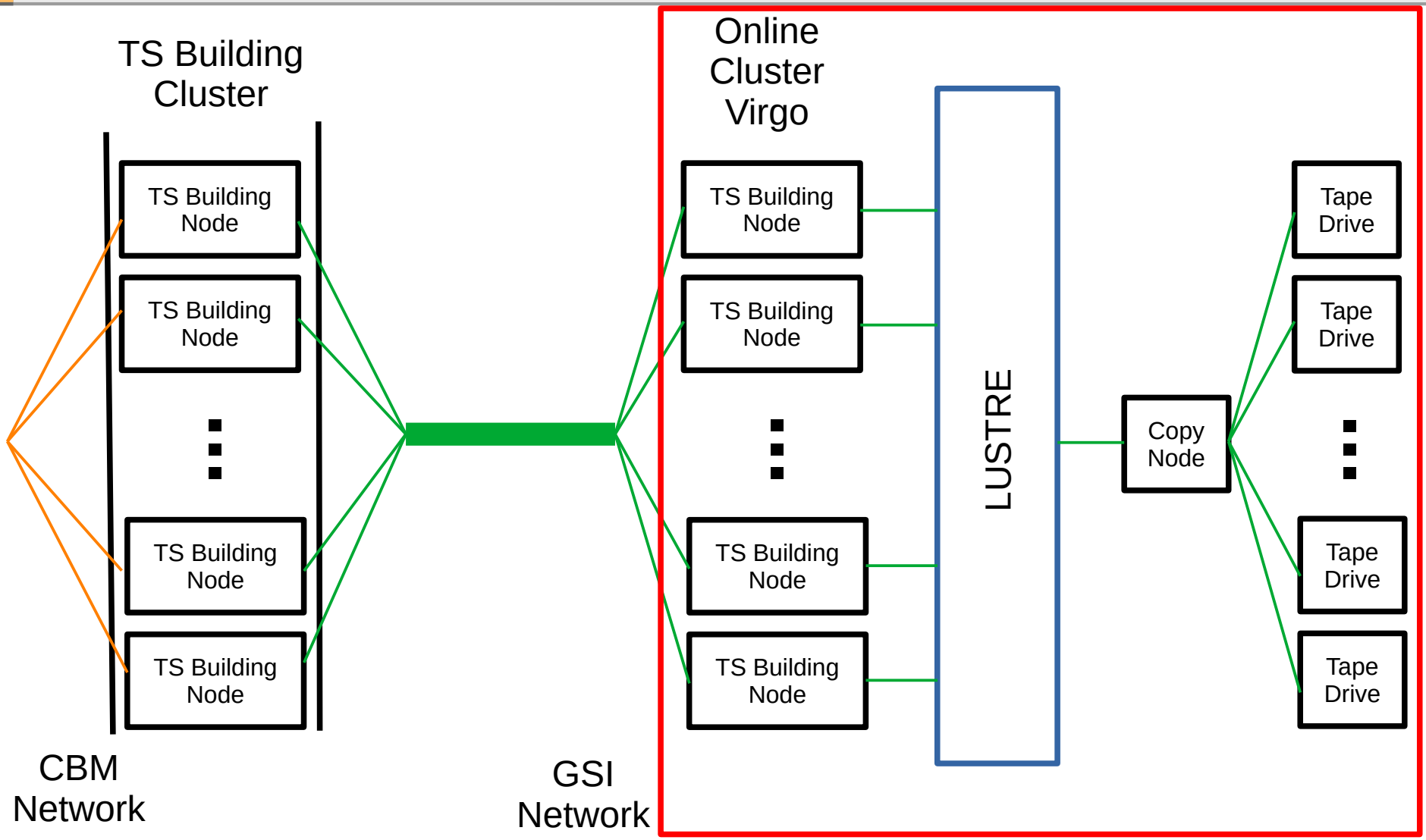
Online Processing Resources Data Challenge

46th CBM Collaboration Meeting

CBM Data Flow Model



CBM Data Flow Model



Goals

- Test writing data to Lustre and to archive those data
  - Use realistic data rates
  - Don’t use realistic data in the beginning
    - Later use real or simulated data
  - Don’t use realistic tasks in the beginning
- Check for potential bottlenecks
- Prepare the needed monitoring
- Test the workflows within GSI IT
  - Several separate groups in GSI IT involved

Data Challenge Storage

- Simulate the following chain
  - Data
  - Data production on Virgo
  - Storage of the produced data on Lustre
  - Archiving of data from Lustre to tape
- 3 different sub challenges
  - CBM default data rate for a long time
  - CBM peak data rate for short time
  - Restore of some archived data from tape to Lustre

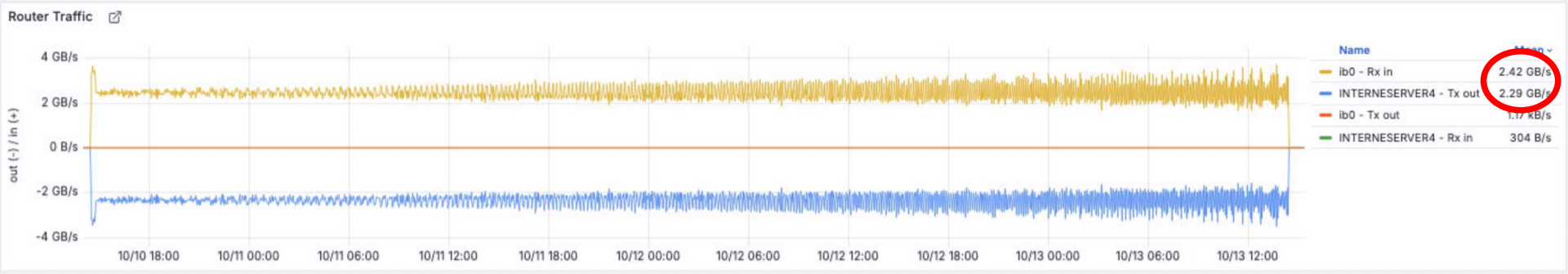
DC: From Virgo to tape

- Produce random data on several Virgo nodes
  - Data produced with fio
    - Tool dedicated for complex IO tests
  - Tool can be started on any number of nodes in parallel with defined write bandwidth
    - Scale data rate per node
    - Scale total data rate to Lustre
- Archive data files on Lustre automatically to tape
  - Tools developed at GSI
  - Currently only very simple setup
    - Copytool is started repeatedly by a cronjob
    - Simple but allows to test the tools as well as the workflow

DC: Requirements for average case

- Realistic estimate from CBM requirements in FAIR Computing TDR
  - Hadron Setup
    - Runtime per year $2.2*10^6$ s
    - 4PB of data
      - 2 PB triggered
      - 2 PB minimum bias downscaled
- Data rate: 2 GB/s
  - 1,9 GB/s + small safety margin
- Test time: 72 h
  - 10 % of the CBM run time
- First archive copy on tape within 1 h

Results



- Reliable data production with ~2,3 GB/s for 72 h
- Data rate from Lustre to Copy Node and from Copy Node to Tape is also 2,3 GB/s
- At startup even with ~3,4 GB/s archived to tape
  - Start of data production and archiving divers by roughly 10 min
  - Coping with the backlog of already created data

Results


Worker Node RDMA Traffic (All)
Router Traffic
Worker Node RDMA Traffic (All)
10 m
Router Traffic
3,4 GB/s

Results

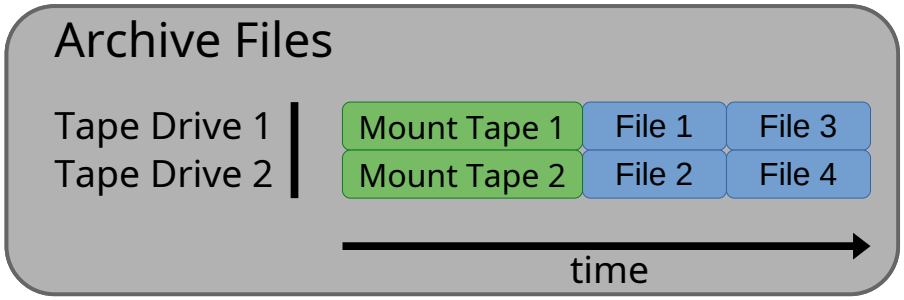
- Data is on tape less than 5 minutes after the file on Lustre was closed
  - Average time is below 90 s
  - Effect of our cronjob which starts each 2 minutes
  - Time to write file to tape is below 30 s

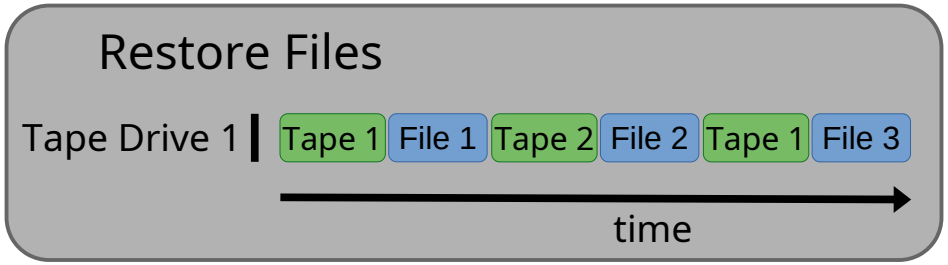
DC: Requirements for maximum case

- Use data rate far beyond the CBM requirements
- Data rate: 25 GB/s
- Test time: 1 h
- First archive copy on tape within 12 h
- Data production was achieved
  - Scale rate per node and number of nodes to achieve 25 GB/s
- Data archiving was not tested yet
  - Calculating time for full archive with previous results
  - ~7 h within the requirement of 12 h

DC: Requirements for Restore test

- Average bandwidth to Lustre: 5 GB/s
- Tape drive efficiency: > 80%
- Restore works but not yet with the required parameters
  - Bandwidth per tape drive ~ 350 MB/s
    - 15 tape drives needed for test
  - Optimisation of the file list needed
    - Access to tape library happens with low level API
    - Need to optimise file access pattern
      - Otherwise tapes are changed nearly for each file





Summary

- Goals achieved
  - Archive at default data rate ✓
  - Archive at peak data rate ✓
  - Restore data ✓