T1 risk assessment

Scenario 1 – "Data loss at a T1"
Scenario 2 – "Partial loss of a T1"
Scenario 3 – "Procurement failure at a T1"
Scenario 4 – "Extended T1 outage"

PRELIMINARY – Daniele Bonacorsi – Last update: 22 Feb 2010

Actions:

1. Verify the incident without quantifying the damage (yet). Talk to the CMS T1 contact(s), and confirm that an incident occurred – beyond most obvious misinterpretations of the observed symptoms. Acnowledge the emergency state. Start the recovery process as from the following bullets. Consider to schedule daily calls to assess the recovery progress, among the CMS T1 contacts and {Data,Fac}Ops.

2. Communicate the incident qualitatively to WLCG (if they were already informed, share the info and check you are in-sync). Inform WLCG that CMS will come back with quantitative details on the impact of the incident on the CMS activity as soon as possible.

3. Depending on the nature of the incident and its environmental conditions – i.e. at what time it occurs (e.g. data-taking or not, computing system idle or not, etc.) – the Computing Coordination together with {Data,Fac}Ops L2's must clarify who will be responsible to overview and manage all recovery operations (eventually just one operator to avoid misbehaviours and interferences).

4. The central Transfer Operator(s) centrally suspend all transfers in progress to/from the affected T1 (so to avoid complicating the namespace even more)

5. Quantify the damage. Identify the list of lost files/datasets. Confirm that they cannot be recovered, and they have to be considered as lost forever. The central Transfer Operator(s) runs `FileDeleteTMDB` in 'preview'-mode, and gives to DataOps the lists of lost files/datasets.

6. Communicate the incident quantitatively to WLCG. Quantify the amount of lost data (files/datasets, and the overall size).

7. If and only if the affected T1 is FNAL: (same as Scenario 2)

   - Each T1 hosting a non-custodial copy of the currently-not-accessible custodial data at FNAL gets this data 'promoted' to custodial. In any case

the data will need to be moved back to FNAL when it comes back since probably such a big processing cannot be done at other T1's in a timely manner.

- ✦ Depending on the details of the crisis assessment, the Computing Coordination will consider the option to use (a set of) T2's as a tapeless T1 for some period[1].

8. DataOps starts to discuss with Physics what needs to be reproduced/ reprocessed/retransferred.

9. The central Transfer Operator(s) runs `FileDeleteTMDB` optimized in 'deletion/invalidation'-mode, and posts the resulting logs to {Data,Fac}Ops.

10. The central Transfer Operator(s) runs the command to change in DBS the state of the lost datasets (to '`INVALID`' state, or to whatever state DataOps prefers).

11. Depending on the results of the discussion with the Physics, DataOps publishes the updated priority list for unsuspension, and the central Transfer Operator(s) starts to un-suspend following the aforementioned priority list.

    - ✦ Proposal of priorities (from higher to lower):
      - SAM/JobRobot data (immediate action);
      - current custodial RAW/RECO data (immediate action);
      - old custodial RAW data;
      - old custodial RECO data or MC data (depends on the discussion with Physics);
      - non-custodial data (maybe even not retransfer).

12. The central Transfer Operator(s) runs a global `BlockDownloadVerify` to verify the post-recovery status.

[1] A number of T2(/T3) have "special" access to 'their' T1. E.g. RALPP T2, CCIN2P3 T2, FNAL T3. In a case where CPUs at the remaining T1's are a problem (rather than disk or tape), these resources could be used to help 'their' T1 to sustain the load.