

1. Introduction (*Jae, Patrice)

In general: what time scale are we talking about? Two months, two years,???. Obviously we need a consensus. This is also important to have a consistent planning of the hardware, software and maybe even organization of data. My feeling is that these are not necessarily guaranteed. Just to mention two points:

Obviously the planned scheme comes at least close to the GRID ideas for LHC etc. Do we have the software for this already? A planning should take this into account and thus it appears very important to also include a discussion about the plans and realistic time scales for GRID computing within D0 and define the planning of the hardware architecture in a consistent manner.

Another point is that there might be too little bandwidth between Europe and the Fermilab. I suggest to perform some studies. The outcome will certainly have important implications for an optimal distribution of data.

It is mandatory to have that an optimal access to the important data for the analysis at every institute at ANY time. My feeling is that the current draft aims at an ideal situation which may take some time to develop. What happens until then? Let me remind you that during the plenary meeting we discussed the need for an approach in several phases. I do not see this reflected in this draft

2. Operating Assumptions (*Jae, Lee)

Instead of a monthly plan it might be more useful to state something about the expectations for the years ahead and the corresponding amount of data.

3. Motivation for Regional Analysis Centers and Use Cases(*Chip, Jae)

4. DØ Remote Analysis Model (DØRAM) Architecture(*Jae)

5. Services Provided by the Regional Analysis centers (*Frank, Patrice, Lee, Iain, Chip)

a. Data Delivery Services(*Lee)

b. Data Reprocessing Services(*Frank)

calibration runs, or run calibration programs (quasi) on-line.

c. Database access services(*Chip)

d. MC Production and Processing (*Iain)

e. MC Data Storage Services(*Iain)

6. Data Characteristics (*Meena, Peter)

To define the optimal distribution of the data we need more and better estimates about the number of users at the various places (e.g. on both sides of the Atlantic) and the need of data transfer. Based on such information we should then perform realistic studies on data transfer etc.. Also simulation tools like MONARC might be of help (we in Wuppertal did make MONARC studies in some other context, if requested we could try and apply it to the D0 needs and constraints).

One may think of various schemes depending on the available resources and band width: World wide distribution without replications, all data available on each continent (some replication), specific data sets distributed according to the main needs of the various main RAC users (high level of replication).

As I mentioned before, the solution depends not only on how many people do what and where but also on the development of the proper software: which of these solutions should be realized at which stage depends also on the progress of the GRID infrastructure.

a. Raw data

Some significant (?) fraction of raw data must be transferred to these sites and permanently stored in their cache system for reprocessing.

How often do we expect reprocessing once stable running is achieved?

Note at LEP: about once a year – do we really need data in the cache at any time?

b. DSTs

These are the ones needed to develop new analysis algorithms at the cluster and track level and are therefore of high importance for the physics analyses.

c. Thumbnail

The thumbnail that constitutes to full data set statistics must be kept at these sites. These thumbnail data should be sufficient to provide significant data set for higher statistics refinement of analyses. Maybe ?

d. Specific Data Streams

The data volume can be significantly reduced if groups are focusing on specific physics topics. In this case RACs may limit the amount of reconstructed data to a well defined subset taking into account, however, a broad enough selection to reliably estimate the backgrounds. This requires some pre-studies as to:

- which kind of data should be safely stored
- what advantage in terms of speed and reliability one expects from such a focused analysis.

It should be up to the main users of a specific RAC to decide on which data they intend to store. This refers to both data and Monte Carlo simulation.

e. MC Data Tier

f. Data Replication

In case of data replication it is of utmost importance to guarantee that the same reconstruction versions are used for all RACs and that the results are identical (this requires some kind of certification procedure e.g. on a well defined subset). Also this implies that reprocessing has to be done in a centrally organized manner. No RAC by itself is allowed to replace an official data set with a privately (even if it is improvement) reprocessed version.

In case of an official reprocessing of the data sets it should be clearly defined by a central institution which RAC is to reprocess which data set. More than one reprocessing on any event has to be avoided.

7. Requirements of Regional Analysis Centers(**Bruno, Christian, Patrice, Frank*)

8. Recommended Sites and Justification(**Christian + all for relevant parts*)

Could we get standardized informations (a table) for the various potential RACS on the existing and forseen infrastructure until 2007 **allocated to D0**

CPU (in terms of Specint95 e.g.), disk space, tapes, bandwidth to Fermilab, and within the region to be mainly served, number of users in this region.

Also it should be outlined if the center is used by other experiments and how much D0 specific person power is needed/available.