H-791
SELEX Reconstruction and Analysis Computing Plan

Peter Cooper, Jürgen Engelfried, Michael Procario, Dane Skow

August 21, 1997

Abstract

This document sets out the postdatataking computing needs of SELEX.

1 Introduction

This document describes the reconstruction and analysis model to be used by SELEX after the completion of data taking. It includes the hardware requirements needed to implement that model.

2 Pass 1 and 2 CPU requirements

Pass 1 is the full reconstruction of physics quantities from raw data. Pass 2 selects samples of interesting events from the pass 1 output data and writes these samples to separate data streams for further analysis. It may be possible to condense pass 1 and pass 2 into a single executable, but the outputs of pass 1 and pass 2 are still logically and physically separate.

Both pass 1 and pass 2 are most appropriate to be done on the farms. We expect that we can perform a pass in ~ 4 months with 32 farm nodes. This is based on a dataset of 1 billion events which require 250 msec/event on a 200 MHz R4400 processor to analyze and the assumption that we can utilize 70 % of the CPU capacity. This CPU requirement is consistent with our previously stated CPU needs.

It is most likely that pass 1 will need to be repeated at least once. We want to start the first iteration in the fall of 1997.

3 Pass 2 IO and data storage requirements

Pass 2 will reduce the data from 6.5 TB of input to 1 TB of output. The output contains approximately 10 % of the input events, but they are larger due to storing calculated quantities. During our stripping jobs on the farms, we have been able to maintain an adequate input data rate by mounting tapes directly on the farm system. We have had difficulties maintaining our input data rate when we attempted to use needfile. This would be the preferred solution to avoid having to replicate error recovery techniques built into needfile.

We currently have a 3 TB quota in the mass storage system, and have used approximately 1.5 TB of it. We do not expect to need more quota in the mass storage system. It will be critical that the mass storage system perform better for us than we have experienced so far.

4 Pass 3

Pass 3 is production of ntuple or other highly compress data sets from the pass 2 output stored in the mass storage system. This stage does not require the full CPU resources of a farm, and needs to be done on a system open to a larger fraction of the experiment. Fn781f, a 12 processor SGI Challenge is adequate to serve as a pass 3 analysis engine, if its I/O, data storage and batch capabilities are improved.

The data must be read from the mass storage system into this system. If we attempt to go through the entire 1 TB in 1 week then we will need to average 1.7 MB/sec of I/O. The ethernet has not been able to sustain rates above 600 MB/sec, so we will request an FDDI interface installed on fn781f. The mass storage sytem must be able to deliver to us an average of 1.7 Mb/sec while in regular use and hit peaks rates that are higher.

The output of pass 3 needs to be stored on disk. If we reduce our 1 TB by a factor of twenty, then we need 50 GB of local storage on fn781f. Individual jobs will obtain much higher compression ratios but there will be many such jobs. We would like and additional 50 GB of working space, in order to make the most efficient use of the mass storage system (provide cache for data for the pass 3 processing). Discussions with Krzysztof Genser of the mass storage group lead us to believe that *some* of our current problems with the mass store would be mitigated if we had more disk storage available than we have on fsgi02, so that we could make fewer requests for larger datasets. We will request more for fsgi02, which will be addressed later, but it also motivates us to ask for 50 GB of working space on fn781f.

The 100 GB of storage is an increase of 60 GB over what is on there now.

5 Development and analysis

Fn781a will still be needed at the level of two processors for constants production and storage. Our constants, database, code management and web server all reside on fn781a and we propose to keep this system under administration by the experiment directly. The equipment borrowed to outfit this machine as a second filter machine will be returned at the end of the 97 FT run. Whether this includes a full swap of the machine chassis we leave to CD discretion.

Fsgi02 will continue to be our main development and final analysis platform. This will require more local disk. We are willing to trade in most of our space in the /afs/fnal/files/data/e781 tree for locally attached disk. We need to increase our local disk space on fsgi02 by at least 30 GB.

Many of our Xterminals are black and white and/or limited in memory. These should be upgraded with 6 new large color Xterminals (which may be direct replacements of the existing older Xterminals). Members of the collaboration have almost completely taken over the 10th floor computer clusters xterminals. The need for color is driven by the strong reliance on color for event displays and other diagnostics critical to the development of analysis strategies and code. The Xterminals need to be NCD Xterminals to match the existing infrastructure and reduce the maintenance demands.

6 Summary and Requests

- Two passes through the farms at 4 months each using 32 processors. Consistent with previous requests.

- 1 TB of space in the mass storage system for Pass 2 output. Already available in 3 TB total allocation for E781.

- Use of fn781f as our pass 3 computer.

  - Upgrade from ethernet to FDDI so that 2.0 MB/sec can be sustained.
  - Increase of disk space to 100 GB from 40 GB for storage of pass 3 output and working space for pass 2 data.
  - Installation of LSF/fbatch on fn781f and help configuring queues for processing Pass 2 output and using resultant ntuple sets.
  - Assurance that the mass store can provide 2.0 MB/sec steady state to E781 on combination of fn781f and fsgi02.

- Continue use of fn781a for code management server, constants server, web server. Return of all equipment borrowed for use as 2nd filter machine.

- Continued use of fsgi02 for development and analysis. Need to increase of local disk there to 50 GB. Anticipated CPU use of this machine may increase to up to twice current levels over next year.

- Addition of 6 20 inch color Xterminals to replace old small memory black and white Xterminals.