

Fermi National Accelerator Laboratory

FERMILAB-TM-2076

The Fermilab Computing Farms in 1998

Marina Albert et al.

*Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510*

April 1999

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Distribution

Approved for public release; further dissemination unlimited.

Copyright Notification

This manuscript has been authored by Universities Research Association, Inc. under contract No. DE-AC02-76CHO3000 with the U.S. Department of Energy. The United States Government and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government Purposes.

The Fermilab Computing Farms in 1998

Marina Albert, Mark Breitung, Jim Fromm, Lisa Giacchetti, Terry Jones
Tanya Levshina, Igor Mandrichenko, Ray Pasetes, Marilyn Schweitzer
Karen Shepelak, Dane Skow, Stephen Wolbers

March 15, 1999
modified April 5, 1999

Introduction

The farms in 1998 continued to change, and the change accelerated from previous years. The remaining old SGI and IBM farms, purchased in 1991, 1992 and 1993, were all retired in 1998. PC farms were purchased and installed, one farm primarily for fixed-target data reconstruction and the other for Run 2 and Theory prototype studies. A large amount of processing was successfully accomplished in 1998. 1999 will continue to be a busy year on the farms, both for reconstruction of data and for preparing for new initiatives at the laboratory.

The year in review

The two largest changes to the farms in 1998 were the retirement of all the “old” UNIX farms (SGI 4D/35 and R3000 Indigo, IBM RS6000 320, 320H and 220) and the installation and commissioning of the first production PC farms. These two changes are related – the new capacity of the PC farms (a dual Pentium 333 MHz is rated at 376 MIPS, whereas the old IBM or SGI nodes were approximately 25 MIPS each) made the retirement of the old nodes possible. The other reasons for the node retirement were space and management. The space on the computer room floor was needed for the new EMASS tape robot and library. The old farms, though they did not require massive amounts of care, did take time and energy away from other more productive and necessary tasks. The old farms served us well.

The PC farms provided a big increment of computing power and they allowed us to learn how to run LINUX farms. There were many issues to work through, such as I/O node configuration, porting software, and process accounting. Some issues were non-trivial and did (and still do) require attention. The model of data and process

control was (and is) one such issue which led to using the Farms Batch System (FBS) rather than CPS and CPS batch. FBS is Run II prototype software, but since it is file, rather than event driven, it helped isolate some issues related to running PC worker nodes attached to a non-PC I/O node. Currently, fnsfh, a Challenge XL, is used as the main I/O node for the PC Farms.

During 1998 the farms were utilized much more fully, as the demand in some months exceeded the supply of computing available. E831 went into full production (rather suddenly) in January of 1998, and it took real effort to provide support for their reconstruction. One big change that was made for E831 was input and output tape staging to and from disk. This was not unique to E831, and essentially all experiments used tape staging to buffer themselves from Exabyte tape drive problems and to allow smoother processing. This was not planned for when the farms were designed, but has been very useful. They succeeded in reconstructing their entire data sample by November, and used the largest amount of CPU power yet of any experiment on the farms. This was a big accomplishment, and was due in part to the E831 organization of effort and very much to the support that was given to them by the Farms and Central Systems Support Groups of the Computing Division. E781 finished their first pass on the farms during 1998. E871 finished a 10% reconstruction early in the year, and threatened to go into full production after that. However, due to various problems and issues (almost all on the E871 side) the full production did not begin until late in the year. The adjustments of their reconstruction model during the year led to reconfigurations of their farms (primarily disk space reorganization and I/O node changes). In addition, the move from cps and cps_batch to FBS on the PC farm required a great deal of effort, but the effort and tests taught everyone a great deal about how the PC farms and the FBS system work.

Experiment E835 came to the farms with a sizable splitting task. A plan was devised and executed during the summer using the O2000 node fnio, and this was very successful. E872 used the farms occasionally during the entire year. Their needs were not large, nor were they very predictable, but the farms were adequate for their needs.

Other users of the farms came from many parts of the lab and many areas of physics. The Auger project continued to use the farms to generate substantial shower libraries for their detector studies. Each shower took a substantial amount of CPU time so their use of the fastest processors on the farm (the O200's and fnio) was a good match. Another user of the farms was the NUMI project, specifically for calculations of radiation in the ground and caverns of the beam-line of NUMI using the MARS

program. These calculations are very CPU intensive, and are best accomplished on the fastest processors available. As in the Auger case, the O200's and fnio were the main machines used for this task.

At the beginning of the year the node allocations were given to E831, E781, E871, E872, Auger and there were smaller users who were using small parts of the farms. As noted earlier during the year E831 ramped up to become a huge user of the farms. E871 also ramped up substantially early in the year as they completed their 10% run. E781 stayed more or less constant while they finished off their first pass. Auger rose and fell to fill in gaps, as did NUMI. Other smaller experiments and projects came and went as required. When E831 finished their main processing the nodes were reallocated, mostly to E871. E871 also received all of the PC farms. NUMI and Auger filled in the E781 farms after E781 finished.

The remaining R3000 machines were all decommissioned during 1998 as were the old IBM machines. At the end of 1998 the farm consists of about 45 SGI R5000 Challenge S, 5 SGI O200 (4 R10000 each), 20 IBM RS6000/43P (133 MHz), 14 IBM RS6000/43P (200 MHz), 29 333 MHz dual PC's, 8 333 MHz single PC's, and a Challenge XL with 24 150 MHz processors. All of these are used as compute nodes and sum to a total of about 30,000 MIPS.

CPU utilization

Table 1 provides the summary of CPU time (in VUP-equivalent units) for the whole farm in 1998. A plot of the CPU utilization (including all previous years of the UNIX farms) is shown in Figure 1. The utilization during 1998 was certainly the highest ever, and as usual was driven primarily by the user demand. There are issues of how much CPU can be delivered, given the architecture of the farms, but it is expected that any part of the farms should be able to deliver 85% or more of its total CPU capacity when the demand is high enough.

Table 2 and Figure 2 show the utilization for each of the many experiments that have used the farm during 1998. E831 was clearly the largest user of the farms. E871, E781, Auger and NUMI used significant amounts of the farms as well in 1998. Other projects and experiments used lesser amounts.

Table 3 is a sum of all the CPU time used by all the experiments that have used significant CPU time on the farms during the last 8 years, along with the totals used by each. E831 is now the largest farm user, with D0 offline second. E871 is expected to move up rapidly on this list when 1999 is totaled, especially given the large requirements of their 1999 fixed target reconstruction.

Table 1 – Total CPU use on the Farms – 1998

<u>Month</u>	<u>CPU delivered</u> (Vax-Months/month)
January	6907
February	10161
March	10845
April	11745
May	13814
June	13382
July	13379
August	12651
September	11965
October	8153
November	6819
December	8661

Table 2 – CPU use by experiment – 1998

<u>Experiment</u>	<u>CPU time</u> (Vax-years)
E831	5988
E781	1566
E871	1349
Auger	1023
NUMI	412
Beams	112
E872	60
E835	52
recycler	19
E866	8
TOTAL	10707

Table 3

Integrated Farm Use
(In units of MIP-years)
Through December, 1998

Experiment	1991	1992	1993	1994	1995	1996	1997	1998	Total
E831							234	5988	6222
D0 offline		59	570	1072	1184	1743	96		4724
E706	28	82	732	992	795	686	19		3335
E791		100	1232	1249	214	11			2806
CDF		110	320	438	752	956			2576
Auger							880	1023	1903
E781							302	1566	1868
E871							191	1349	1540
D0 MC	101	162	396	197	108	23			987
E665	14	105	733	91	2	10			955
E771		94	211	339	219				863
E866						55	409	8	472
NUMI								412	412
E789		156	247						403
E687	99	235	29	30	2				395
Theory							309		309
Minos							221		221
Beams							105	112	217
Recycler						61	130	19	210
E872							21	60	81
E760		54							54
E835								52	52
E731	38								38
Magnet							28		28
Total	267	1156	4541	4408	3276	3545	3217	10707	32117

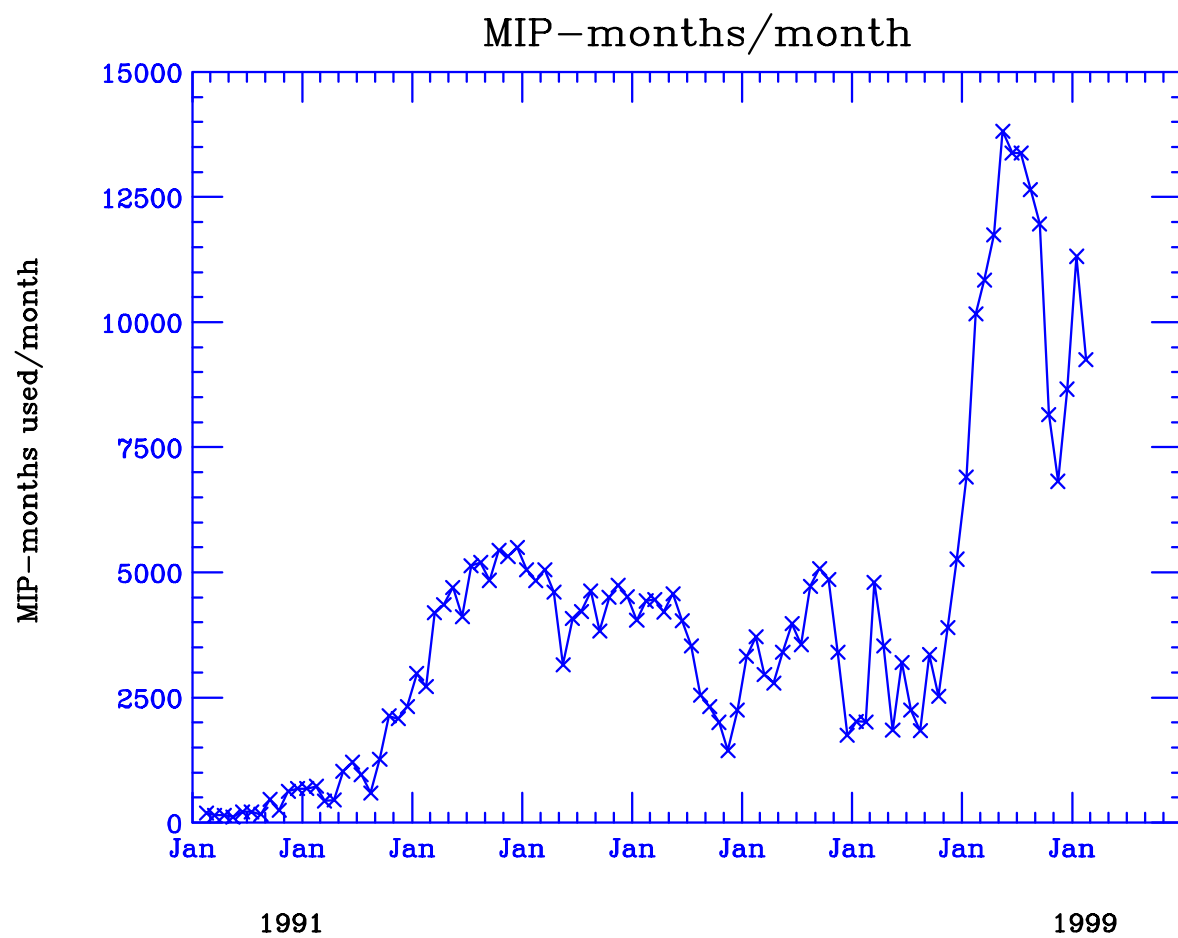


Figure 1.

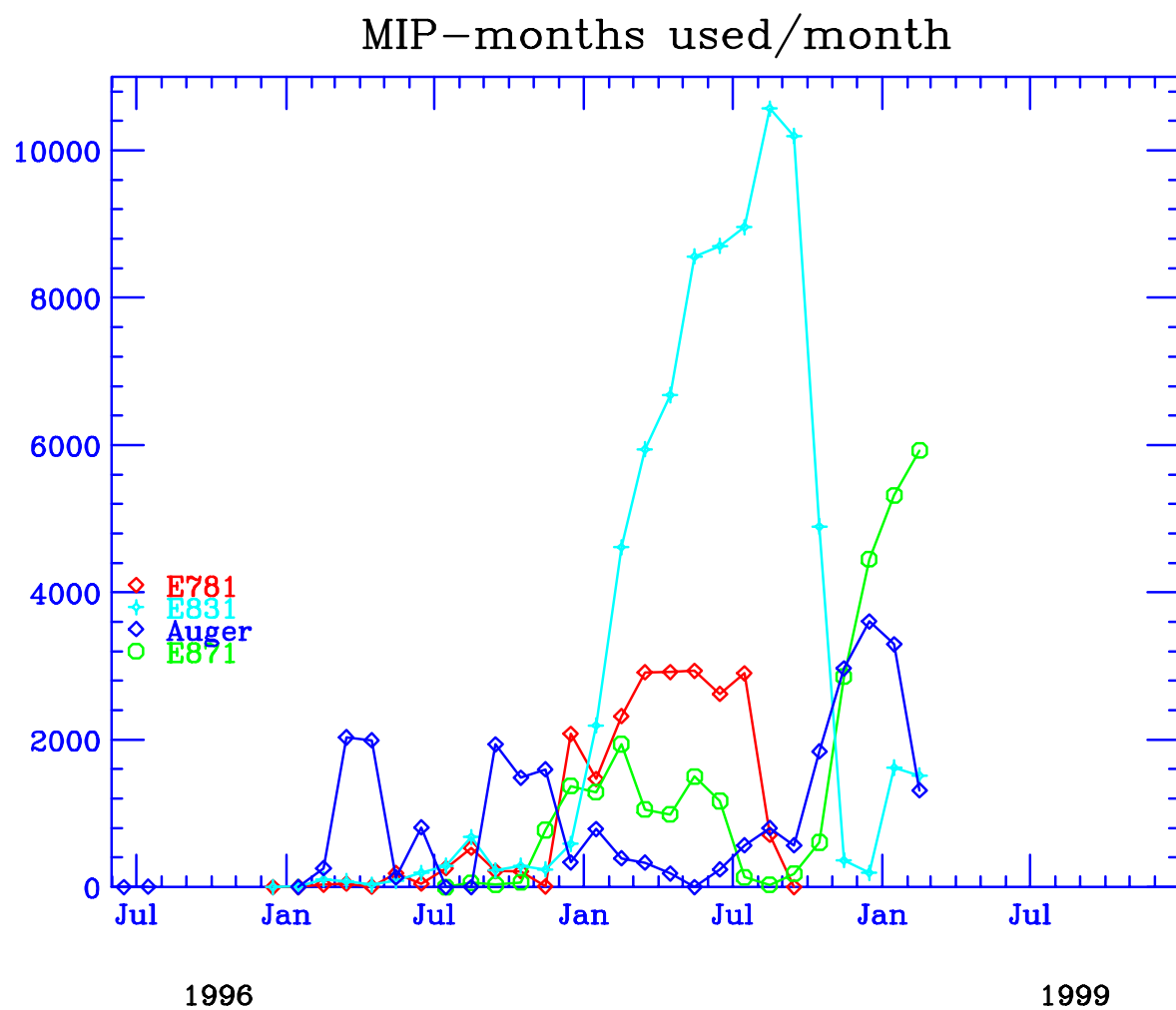


Figure 2(a).

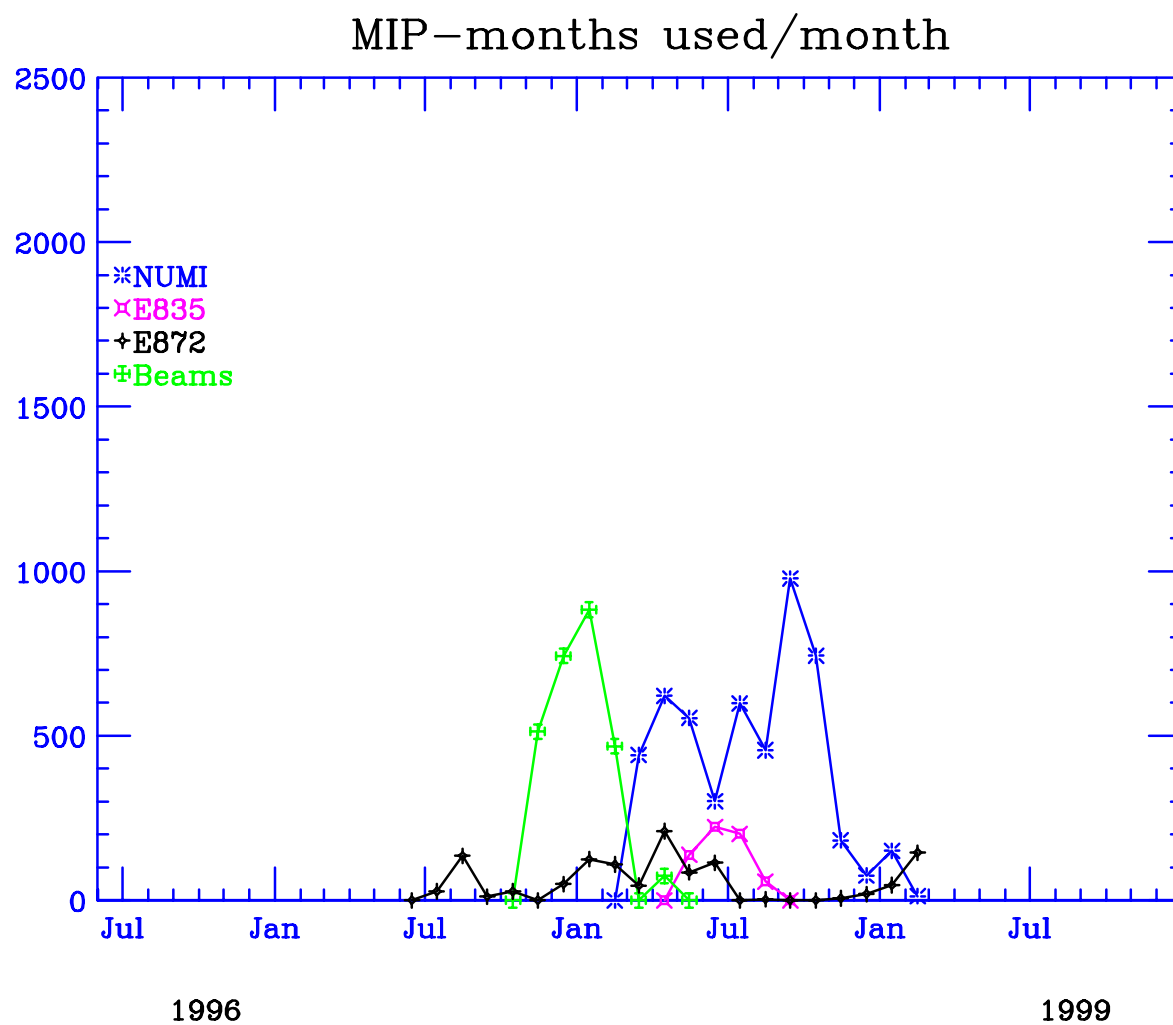


Figure 2(b).

Plans

During 1999 there will be very large demands for CPU time on the farms, almost all coming from E871, E781, and the Auger and NUMI projects. The largest obvious demand for CPU power will be to reconstruct the large datasets that were collected during the fixed target run that ended in September 1997 and the upcoming fixed target run in 1999.

During 1999 it is expected that more of the old farms will be decommissioned, as will the `cps` and `cps_batch` software. The goal is to simplify the support of the farms, while allowing more effort to go into the preparations for Run II.

The next large increase in compute power of the farms will occur in 1999, and this will include the first large set of nodes for Run II as well as a set of nodes to augment the non-Run II farms. This will allow the fixed-target experiments enough compute power to finish their processing in a reasonable amount of time. It will also allow for other potential large CPU users (NLC calculations, NUMI, Muon Collider, etc.). The Run II farms will be the first substantial (≈ 100 nodes) computing resource necessary for the beginning of data taking in the year 2000.