THE PROTOTYPE FARMS

- Introduction
- The 14-node test
- The ROOT I/O test
- Test code in the prototype farms

THE 14-NODE TEST

1) Performed throughput and reliability tests of the prototype farms using all 14 worker nodes during a 3-day period.

2) Goals were to gauge the performance of the system, to understand the limitations of the hardware and to help us design, test and optimize the production control software.

3) The data transfer rate between the I/O node and the worker nodes is about 10 $\sim$ 15 MB/CPU-s with the current hardware.

4) Overall CPU utilization rate when running the CDF ProductionExe on the prototype farms is $\sim$ 95%, including the time needed by FBS to schedule, start and end the job.

5) A more detailed account of the 14-node test can be found in CDF Note 4941.

THE ROOT I/O TEST

1) Rob Kennedy and I performed some ROOT I/O tests using a prototype EDM and compared the performance of this ROOT-based, stand-alone (no AC++ framework) EDM to the requirements of the CDF Production Farm for Run II.

2) The tests were performed on fnpcc (Pentium-II 400 MHz, rated by "tiny" at 223 VUP) and cdfsga (rated at 192 VUP) with optimized EDM code.

3) We extrapolated the results to the target machine (assumed to be $\sim$ 50% faster than cdfsga) to obtain 17 MB/CPU-s (Read) and 15 MB/CPU-s (Write) with compression factor = 0.

4) The Production Farm system design specifies a steady-state I/O rate of 11 MB/s for splitting and merging on the I/O node.

5) A detailed account of the ROOT I/O tests can be found in CDF Note 5052.

TESTS IN THE PROTOTYPE FARMS

1) Optimized ProductionExe on a Pentium-II 400 MHz LINUX machine gives ~ 3.5 CPU-s/event. Assuming an event size of 250 KB and CPU utilization rate of $\sim 95\%$, we can process a 1 GB file in ~ 3.5 hours, if we extrapolate to Pentium-III 500 MHz.

2) Assuming a TCP/IP transfer rate of 10 MB/s with the present system, we can transfer a 1 GB file in 100 sec in real time.

3) Using the ROOT I/O test results and assuming CPU utilization rate of 95%, we estimate that we will need $\sim$ 325 MIPS (93 for Read and 232 for Write) for splitting/merging.

4) The SGI I/O node (4 R12000 CPU's at 300 MHz) is rated at 300 MIPS (estimate). Using the above estimates, a single CPU will be needed for splitting/merging the inclusive and physics datasets.

5) Another CPU would be dedicated to reading in the inclusive PAD datasets.

6) A third CPU would become the "input" node.

7) The control software design assumes that the I/O performance of the ROOT-based EDM will allow for the splitting/merging to be done on a single CPU. Life would become very complicated otherwise.

FARM BATCH SYSTEM

1) The Farm Batch System (FBS) is a software package built on top of the LSF batch system. It is designed for a farm architecture with all the traditional tools such as job submission, job queuing, load balancing, etc.

2) FBS needs just two LSF licenses running on the I/O node (dual CPU). Thus, it is a cost-effective way to address the need for a batch system in a computing farm environment.

3) FBS was first released (v1.0) in Nov. 98 for evaluation by the users. Based on these evaluations, FBS v2.0 was released in April 99. FBS v2.1 was released in July 99 with further improvements.

4) CDF, D0 and fixed target experiments have adopted FBS as their batch software system. It has become an official FNAL product.

5) More detailed FBS documentation can be found in http://www-hppc.fnal.gov/fbs/fbs.html.

FARM INTER-PROCESS COMMUNICATIONS

1) The Farms Inter-Process Communications (FIPC) software package was first released in May 99, as a tool to address the need for IPC in a computing farm environment.

2) Both CDF and D0 have been evaluating FIPC. The general reaction has been positive.

3) Documentation for FIPC will be available soon. Further improvements from user feedback are on-going.

4) There are plans to make FIPC an official FNAL product.

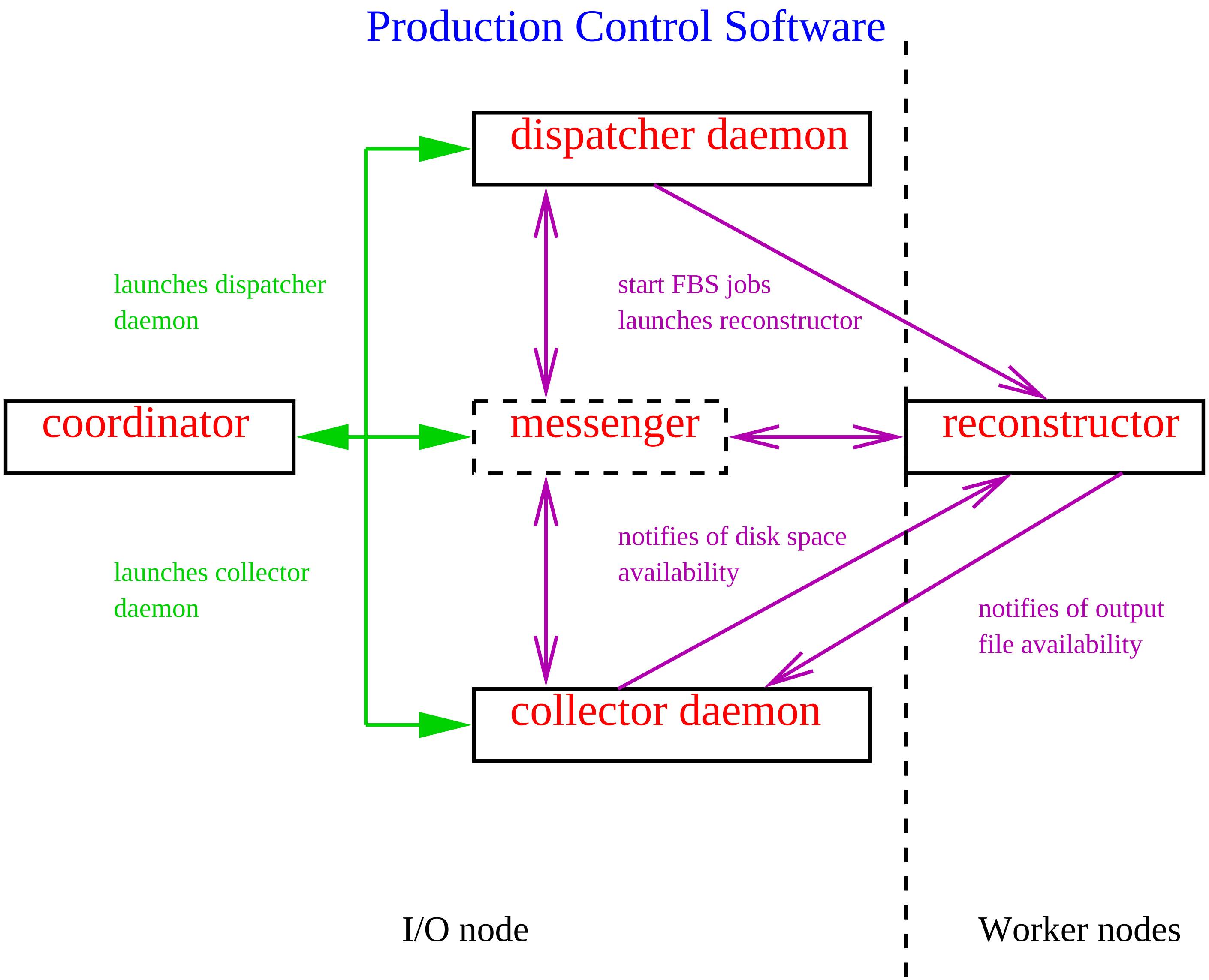
PROTOTYPE CONTROL SOFTWARE

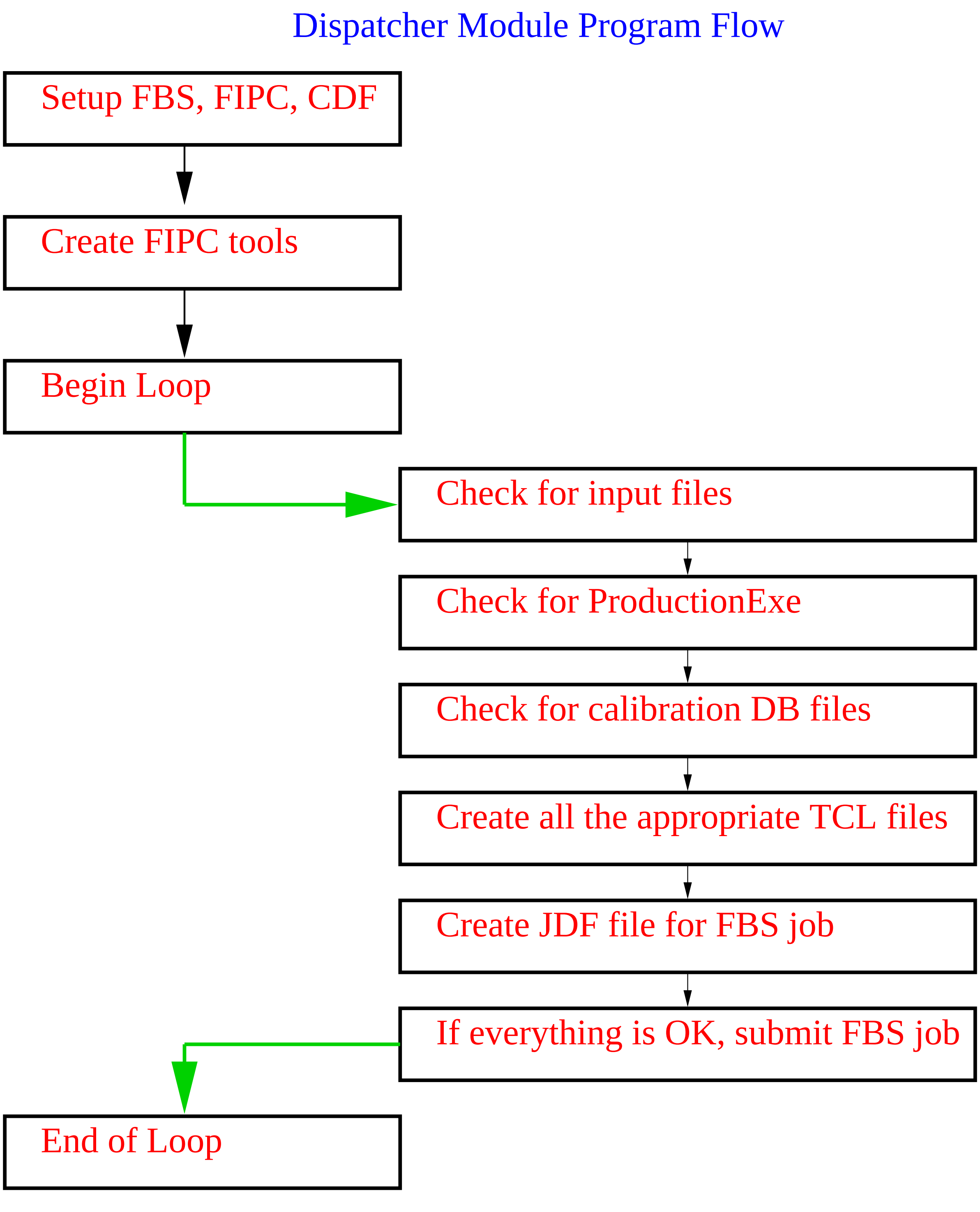
1) The first version of the integrated production control software already exists and has been running on the prototype farms since July.

2) The existing production control software makes use of FBS and FIPC to operate.

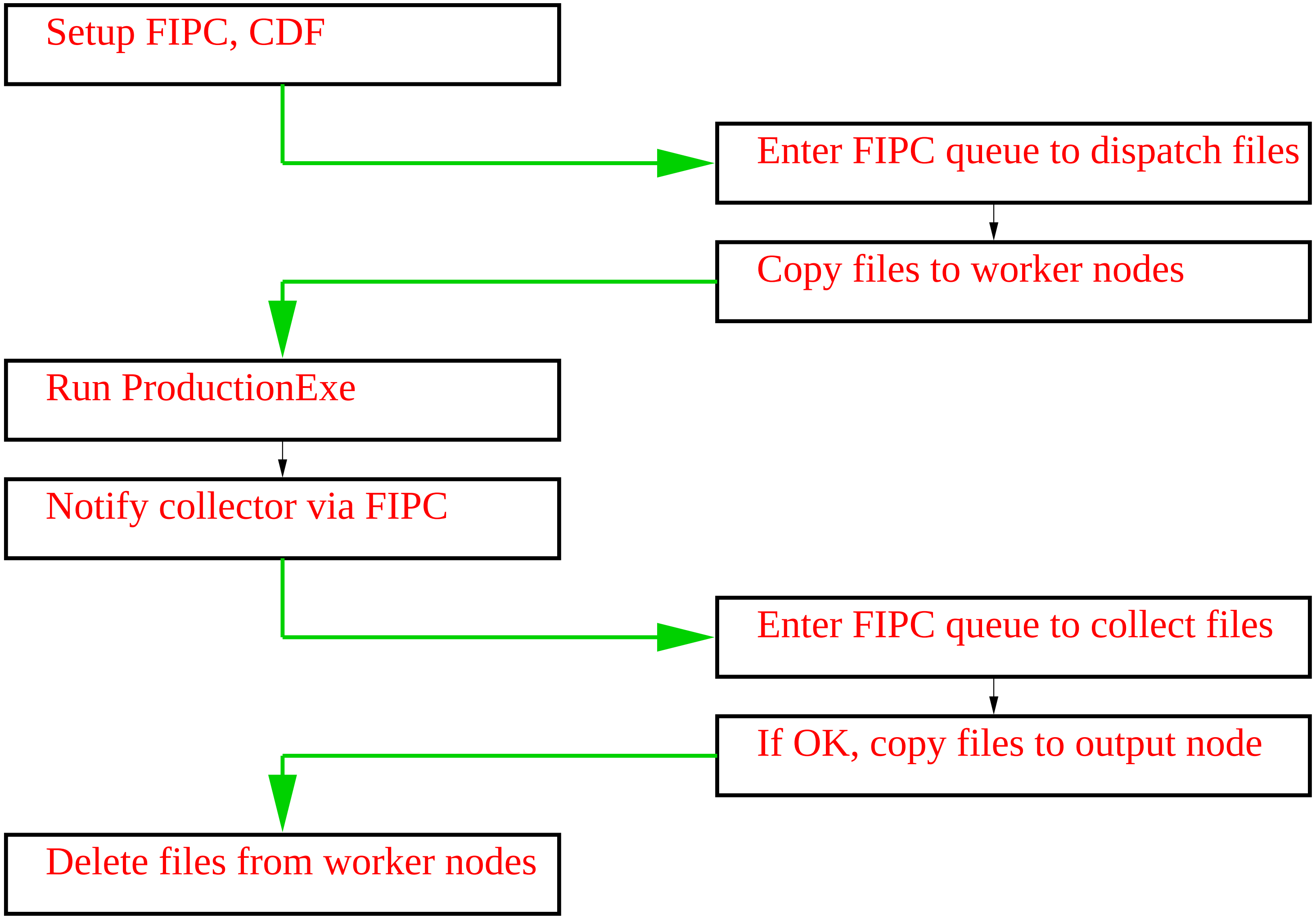
3) There is on-going work to integrate the remaining internal control modules (Coordinator, Bookkeeper and Messenger) into the existing system.

4) The production control code will continue to be developed and integrated with external packages.


Production Control Software
dispatcher daemon
launches dispatcher
daemon
start FBS jobs
launches reconstructor
coordinator
messenger
reconstructor
notifies of disk space
availability
launches collector
daemon
notifies of output
file availability
collector daemon
I/O node
Worker nodes


Dispatcher Module Program Flow
Setup FBS, FIPC, CDF
Create FIPC tools
Begin Loop
Check for input files
Check for ProductionExe
Check for calibration DB files
Create all the appropriate TCL files
Create JDF file for FBS job
If everything is OK, submit FBS job
End of Loop

Reconstructor module program flow



Collector module program flow