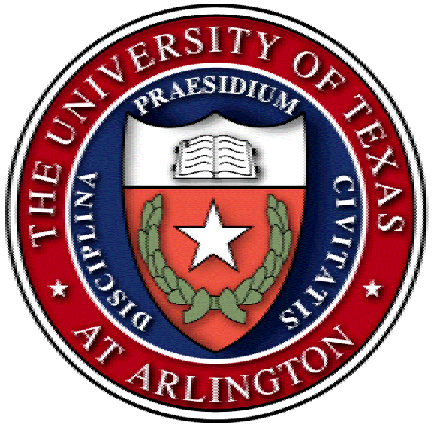
Status of UTA MC production farm and Its Software

Karthik Gopalratnam

Drew Meyer

Tomasz Wlodek

Jae Yu



UTA D0 Monte Carlo Farm

- **UTA operates 2 Linux farms: HEP and CSE**

  HEP farm: *6-566 , 36-866 MHz processors, 3 file servers, (250 GB) one job server, 8mm tape drive.*

  CSE farm: *10 866 MHz processors, 1 file server (20 GB), 1 job server*

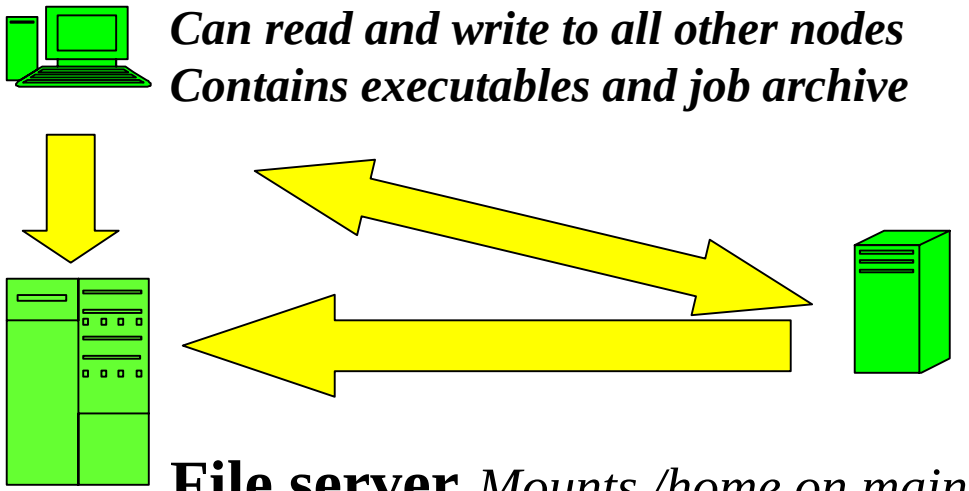
- **Control software** (job submission, load balancing, archiving, bookkeeping, job execution control etc) **developed entirely in UTA** by former UTA student, Drew Meyer, currently *Meyer Systems*
- **Scalable**:*started with 7 nodes, then 25, now 52 processors, soon few hundred distributed nodes*

  (More about expansion plans later)

http://wwwhep.uta.edu/~mcfarm/mcfarm/main.html

MCFARM – UTA farm control system

- MCFARM is a specialized batch system *for:*
  Pythia, Isajet, D0g, D0sim, D0reco, recoanalyze
- Can be adapted *for ATLAS, CDF*
- It is intelligent: *it knows how to handle and – in most cases – recover typical error conditions.*
- Hard to break – *even if several nodes crash the production can continue for a few hours*
- Interfaced to SAM and bookkeeping package, *easily exports production status to WWW page*
  http://www-hep.uta.edu/~mcfarm/mcfarm/main.html



Main server (Job Manager)
Can read and write to all other nodes
Contains executables and job archive

Execution node
(The worker)
Mounts its home directory on main server
Can read and write to file server disk

File server *Mounts /home on main server*
Its disk stores min bias and generator files
bnd is rw by everybody

- **Easy to install on other farms**

If any new institution wishes to build a farm –

we will give the system to you! It can be running

on your farm within hours – URL soon available

(Provided that your farm satisfies some system

requirements, ask me for details)

Production Status

- Delivered 1.6M mcp06 events.
- Details: see http://www-hep.uta.edu/~mcfarm/mcfarm/main.html
- **Mcp07 installed**

  There was a problem with reco, patched by David Evans.
- Recoanalyze crashes (as of Sep 10,2001)
- Mcp07 production is now under way

Data transfer to FNAL

- **In the past we have been using tapes to send results to FNAL**
- **At present we are switching to network transfer to SAM**
- **SAM station has been installed, but not yet properly configured**
- **Hope that with the help of FNAL experts it can be configured easily. I would be very grateful if someone from the experts could sit with me on the phone and guide me through the procedure.**

Issues and Difficulties

- **Authors of SAM documentation grossly overestimate my IQ**

Example:

The min-delivery flag works as follows. The SAM optimizer (global resource manager) groups requests for file deliveries by the file's tape, in order to minimize the tape mounts. Normally, the station will deliver all or none of the files in the group. If the entire group cannot be fit on disk and the min-delivery flag is set to a number in KBytes, then the station will attempt to deliver a fraction of the group but at least as much as the value of the flag. If the value of min-delivery is zero, there is no minimum unit of delivery (i.e., a single file may be delivered).

This can be written in Greek!

Difficulties – continued:

- Instructions on how to install SAM station are not entirely accurate

Instructions : ***(how to start SAM station)***

sam start station [--name=<name>][--quiet|--verbose] [--nofork] **…**

Expert: (***correct answer)***

you used the option "--name=<name of your station>". This should always be "--station=<name of your station>".

Good News:

- I received enthusiastic help from many people in CD: Carmenita Moore, Lee Lueking and Lauri Carpenter who did sit with me on the phone and guided me through various stages of installation.

  (and I apologize to those whom I have missed)

- Without their help installing SAM software would have been much harder

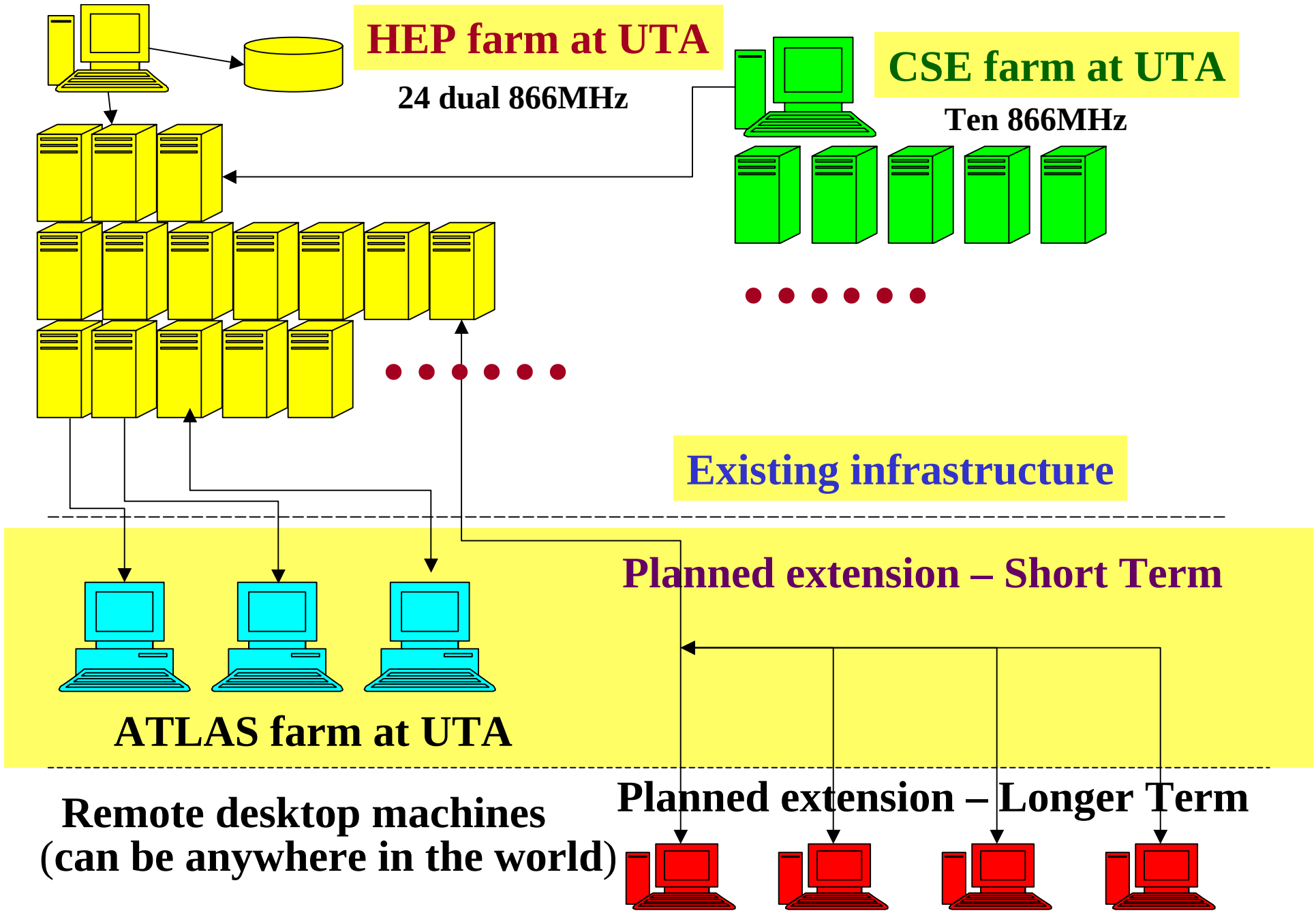
Problems - continued

- Files disappear in SAM.
- Example: file reco_mcp06_p08.10.00_UTA_isajet_glga-jmET-PtGt5.0+PtLt9999.0_mb-poisson-1.0_228160638_2001 has been written to tape PRH289. It sits on this tape, its metadata has been declared to SAM. It is mentioned in the tape metadata file.
- The file cannot be located by SAM! Other files on this tape can be located by SAM
- There are more files/tapes like that.
- I have notified helpdesk about this problem

  (Ticket # 20722)

UTA farm expansion

- **We run generator,DØg, sim, reco and recoanalyze jobs in a chain**
- **DØg is CPU intensive, but uses little IO**
- **DØsim+reco use less CPU but a lot of IO**
- **Why not run DØg part of the chain on idling desktop PC's?**


HEP farm at UTA
24 dual 866MHz
CSE farm at UTA
Ten 866MHz
Existing infrastructure
Planned extension – Short Term
ATLAS farm at UTA
Planned extension – Longer Term
Remote desktop machines
(can be anywhere in the world)

External nodes:

- They will run d0gstar simulation in background, invisible to user, non invasive. Only minimal resources will be used. Data transfers will be done only at night, if node owner requests so.
- Only minimal node requirements will be needed: ssh, mcfarm account and few GB of scratch space
- Will be user friendly, idiot-proof installation. Download a tar file from www, un-tar it and go!
- A chance for small institutions with only a handful of PC's to contribute!

Plans

- **Within a month** –

  we will have a “Proof of concept” prototype network with 3 external nodes

- **At next DØ meeting**

  we will submit a proposal for the use of DØ remote desktops and be looking for volunteers

- **Goal:**

  we would like to connect large number of external nodes in the near future

Conclusions:

- **The mcp06 run was very successful**
- **Mcp07 is running**
- **We are switching, with pain, to network data transfer to SAM to replace tapes**
- **UTA MCFARM software is solid and robust ➔ Will be made available to interested parties**
- **We plan to expand the UTA farm cluster into a globally distributed MC farm (DØMCGRID??)**

  (Which means: we would like to build a ~100k$ farm for free)