McFarm
(DØ Monte Carlo Farm Controller)

Drew Meyer
The University of Texas at Arlington
April 18, 2003
Reasons for Using McFarm

• Simplifies Monte Carlo Production
• Manages the Cluster with Minimum Labor
• Manages the Cluster Efficiently
• Minimizes Impact of Changes to SAM, mc_runjob, other DØ software
• User-Oriented
Life Cycle of a Request

Researcher

Farmer

M.C. Requests WebSite

Finished

Finished

Jobs

Monitor

McFarm Server

D. Meyer - UTA Farm Team
Automated Request System

Researcher → Finished

Farmer → Monitor

McFarm Server

Mc_runjob Dispatcher

Finished

Other Farms

D. Meyer - UTA Farm Team
Original McFarm Design Criteria

• Keep Processors Maximally Occupied
• Minimize Farmer’s Labor
• Robust in Spite of Problems (nodes down, network failures, SAM down, binaries fail)
• Manage Dedicated Resources (cache, CPU)
• Minimize Network Traffic (locality, locality, locality!)
Evolving Design

• Share the Resources (Batch Queues, Quotas)
• Remote Monitoring
• Automate Request Handling
• Grid-Friendly (Less Node Prep, Fewer Daemons)
• Relax Locality Preferences
• Easier Installation, Re-Configuration
McFarm Software Integration

• **DØ Binaries** - minitars or full release
• **SAM** - declaration, storage, retrieval
• **mc_runjob** - job and metadata construction
• **NFS** - access to binaries, minbias database
• **NIS** - account management
• **ssh** - intra-cluster monitoring and control
• **Batch Queues** - PBS and Condor
McFarm Job Life-Cycle

Job Server
  Ready Jobs

Prod Node
  Local Cache

Prod Node
  Local Cache

File Server
  Global Cache

Global Monitor

Local Monitor

Distribute Daemon

Local Monitor

SAM Store

Cache Store

Farmer

SAM

D. Meyer - UTA Farm Team
Sample McFarm Alerts

The following alert(s) are being issued by the farm monitor:

* Resolve these problem jobs:
  UTA-DSRT-Req5386-bbh-bbbb-03084233103 on hepfm010

* Bring these nodes back on-line:
  hepfm011
  hepfm012
  hepfm013

* Restart the "execute" daemon on:
  hepfm014

D. Meyer - UTA Farm Team
Sample McFarm Alerts

The execute-daemon on ouhep001 has a problem with this job:
OU-D-Req5257-minbias-cdf-03092152359

Job was apparently Looping in D0G
It is removed from this node's execute-queue.
Job was moved to the error_queue
Sample McFarm Alerts

SAM declare failed after 6+ hours and 6 tries:
import_pythia-gen_UTA-P-Req5502-zz-znunu+zjj-03083175022.py

***Detail SAM Response follows***

Code: person work group not found in db (Category User)
Additional information: Person 'mcfarm' w/ Id '383' not found in Working Group 'higgs sensitivity' w/ Id '61'.

***End of Detail Response***
Sample McFarm Alerts

Excessive Swap Usage on Node hepfm000 at Sat Apr 5 06:21:42 2003
(Swap MB:570.0 Pct:71.3%)

Job UTA-DSRT-Req5386-bbh-bbbb-03084221448
Phase D0RECO/D0RECO Status IN PROCESS

Will suspend new exec until swap space becomes available.

Highest swap MB encountered in this event: 570.0
Highest percent encountered in this event: 71.3%
New Job Execution resumed

D. Meyer - UTA Farm Team
# McFarm Utilization Metrics

<table>
<thead>
<tr>
<th></th>
<th>Files</th>
<th>Events</th>
<th>OutputMB</th>
<th>CPU Time</th>
<th>ClockTime</th>
<th>%Util</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>6</td>
<td>6000</td>
<td>6362.41</td>
<td>155:08</td>
<td>156:13</td>
<td>99.3</td>
</tr>
<tr>
<td><strong>Per File</strong></td>
<td>1</td>
<td>1000</td>
<td>1060.40</td>
<td>25:51:28</td>
<td>26:02:14</td>
<td></td>
</tr>
<tr>
<td><strong>Per Event</strong></td>
<td>0</td>
<td>1</td>
<td>1.06</td>
<td>0:01:33</td>
<td>0:01:33</td>
<td></td>
</tr>
</tbody>
</table>

**CPU Time In Period:** 2975.5

**Period ClockTime:** 167:59  CPUs:46  Extend Clock: 7727:59

**Less InProcess:** 3507:34 Available ClockTime: 4220:24

**Farm CPU Utilization:** 70.5%
Summary

• McFarm is a mature, robust, flexible cluster controller for Monte Carlo production
• It protects the user from software changes, production problems, operating expenses
• Ongoing effort to evolve its capabilities, improve behavior in non-dedicated environments (Grid)

D. Meyer - UTA Farm Team