Interstate Data Moving and the Last Block Problem: Lessons Learned in the CAPS Spring Experiment 2014

Keith A. Brewster, Ph.D.
Center for Analysis and Prediction of Storms
University of Oklahoma
SPC/NSSL Spring Program in the Hazardous Weather Testbed

- Testing and calibration of new forecasting methods in a simulated operational setting
- 6 weeks in spring season
- Collaboration among
  - NOAA research units
  - NOAA operational units
  - Universities
  - Private sector
- Testbed located between the NOAA Storm Prediction Center and Norman National Weather Service Forecast Office
CAPS Spring Experiment

- Part of NOAA/SPC Spring Experiment at the Hazardous Weather Testbed
- Run Large Ensemble of Convection-Allowing NWP Forecasts for 6-weeks in Spring
- New methods for severe weather prediction in 1-2 days time frame
- 25 NWP models run at XSEDE Centers
- 2013-2014 Darter at NICS (UTenn @ Oak Ridge)
Goal: “Real-time” 4D Data Visualization

Procedure Since 2007

• Run models at PSC or NICS
• Bring 2D files and images back
• 2D fields and levels pre-selected

Issues

• Does not allow full 3D visualization
• May want to examine other fields and levels

Need to move 3D Data from NICS at UTenn CAPS at OU
Scoping the Task

- CONUS Domain at 4-km Resolution
  \[1163 \times 723 \times 53\]
  Output for one time: **4.2 GB**

- Domain decomposition onto 384 (6x64) processors results in **384** split files per output time, each file = **11 MB**

- For smooth animations, 10-minute output is generated for 5 members covering the afternoon and evening, forecast hours 18-30.

- Complete Forecast:
  - 60-h forecast, hourly output + 10 minute output 18h-30h: **508 GB**

- Day-1 Afternoon and Evening for Animation
  - Forecast 18h-30h with 10-minute output: **307 GB**
Plan “A” Workflow

WRF Model → wrfout split files

Daily Region of Interest → joinWRF Join and subset

WDSS-II files → WDSS-II Converter

wdfcreate → VAPoR vdf files

WDSS-II files → VAPoR

WDSS-II files → GridFTP or bbcp

sftp or scp

NICS Darter

CAPS Server

WDSS-II → VAPoR

WDSS-II → NSSL or SPC Machine in HWT
Data Route

NICS/U Tenn' in Oak Ridge, Tennessee to CAPS/OU in Norman, Oklahoma

845 miles 13 hours via MichelinNet
Internet Route

Darter → 10G → Kraken → 10G → Beacon → 10G → Datamovers → 10G → Keeneland → 10G → Nautilus

Medusa site-wide file system
Internet Route “Last Mile”

• OneNet Tulsa to OU – Norman (4PP)
• 4PP to National Weather Center
  • Across the parking lot
• National Weather Center to
  • CAPS Switch to File Server System
  • to CAPS Office Workstation via Firewall
Recent Networking Initiatives

1. University of Tennessee BLAST
2. OneOklahoma Friction Free Network (OFFN)
3. National Weather Center Upgrade

- University of Tennessee BLAST
  - 100 Gps Upgrade of Research Network
  - Includes connections to HPC at NICS
Recent Networking Initiatives

- OneOklahoma Friction Free Network (OFFN) NSF Campus Cyberinfrastructure-Network Infrastructure and Engineering Program (CC-NIE)
  - Establish 10 Gbps Network Ring
  - OU-OSU-Langston-Tandy Supercomputing Ctr
Testing revealed a “last block” problem, actually within the building itself, mostly due to a slow firewall.
Packing and Compression?

Try creating compressed tar file before sending?  
Sending large files is faster ~100 MB/s vs ~10 MB/s

BUT! Creating a compressed tar file takes time

Test for 1-hour of Full Domain Split Files

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ tar –zcvf</td>
<td>22 min</td>
</tr>
<tr>
<td>bbcp</td>
<td>2.5 min</td>
</tr>
<tr>
<td>Total</td>
<td>24.5 min</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ tar -cvf</td>
<td>15 min</td>
</tr>
<tr>
<td>bbcp</td>
<td>5 min</td>
</tr>
<tr>
<td>Total</td>
<td>20 min</td>
</tr>
</tbody>
</table>
Plan “D” Workflow

**WRF Model** → wrfout split files → *joinWRF* Join and subset → wrfout joined files

**Daily Region of Interest**

**SPC Workstation**

- WDSS-II files
- WDSS-II Converter

**WDSS-II**

- sftp

**CAPS Stratus**

- wrfout joined files
- vdfcreate

**VAPoR**

- vdf files
- WDSS-II files
- bbcp

**SPC Workstation**

- sftp

**WDSS-II**

- SPC Workstation

**VAPoR**

- WDSS-II files
- bbcp

**CAPS Laptop**

- SPC Workstation

- VAPoR

**NICS Darter**
Re-Scoping the Task

- **Selected Subdomain of the Day**
  - $203 \times 203 \times 53$
  - Output for one time: 111.5 MB

- Single file for each time = **111.5 MB vs 4.2 GB**

- Day-1 Afternoon and Evening for Animation
  - Forecast 18h-30h with 10-minute output:
    - 73 output times,
    - **8.2 GB vs 307 GB**

- Processing and Transfer: ~**20 min**
Fruits of Labor
Lessons Learned

• Involve networking pros early
• Your mileage (throughput) may vary
• Be flexible with workflows
• Evaluate overhead of all steps
• Find ways to fund equipment upgrades where needed – slowest link sets your rate
• Software Programmable Networks/Science DMZ’s may be needed for the largest jobs
Future Plans

• National Weather Center (September, 2014)
  • Upgrading Network to 10 Gps Switches
  • Software Programmable Networking Enabled
  • Replacing Slow Firewall

• CAPS (September-October, 2014)
  • Upgrading to Two 10 Gps Gateway Servers with Virtual Router Redundancy Protocol (VRRP)
  • Software Programmable Networking Enabled
  • Upgrading servers to 10 Gps network interfaces

• For 2015
  • Set Jumbo Size Maximum Transmission Unit (MTU) for route
  • Explore use of Science DMZ within OneNet
Questions?

Thanks to:
Chris Cook, CAPS
Kevin W. Thomas, CAPS
Victor Hazlewood and his networking team, UTenn
Matt Runion & James Deaton, OneNet
Henry Neeman & OSCER Team
Mike Coniglio, NOAA/NSSL

Contact Info:
Keith Brewster
CAPS/University of Oklahoma
kbrewster@ou.edu