Project Fermi - A Highly Available NAS Gateway Built from Open Source Software

Dan Pollack
AOL
Why NAS gateways?

- Decouple storage from NAS server
- Widely deployed SAN infrastructure
- Storage arrays cost less than captured storage in NAS servers (@AOL)
- Multiple tiers of storage in NAS servers
- Heterogeneous storage in NAS servers
What is Fermi?

- Commodity: Industry Standard Servers (ISS) and Open Source Software (OSS)
- Redundancy: Clustered heads for redundancy
- FC backend: Performance and reliability
Why Fermi?

❖ Cost Savings
  ❖ Cost of commercial NAS servers - $$$$  
    ➢ Expensive hardware  
    ➢ Software licenses  
  ❖ Cost of Fermi - $  
    ➢ ISS hardware  
    ➢ No licenses  
❖ More Flexibility (Any storage OK)  
❖ More rapid hardware refresh
What hardware?

- 64-bit x86 CPU server
- Multiple 1Gbps/10Gbps Ethernet links
- Multiple 8Gbps/16Gbps FC links
What software?

- Open Solaris (open Indiana)
- Pacemaker (HA cluster resource manager)
- Heartbeat (HA cluster infrastructure manager)
- ZFS
- Solaris CLI
What did we add?

- FC disk cluster SCSI reservation management
- ZFS zpool management
- Virtual IP address management
- Minor solaris specific bug fixes in pacemaker
What is complete?

- HA Failover
- NFS services
- iSCSI services
- SNMP metrics
What needs doing?

- CIFS services
- Active Directory integration
- Local snapshots
- Remote snapshots (ZFS send/receive)
- Provisioning API
What does it look like?

- Three servers in a stack
- Multipath FC
- Multiple Ethernet interfaces
  - Multi-home
  - LACP/MLAG
- HA services for network and storage
  - Virtual IP for network failover
  - Shared storage for failover
What does it look like?
How does it perform?

Filebench results

<table>
<thead>
<tr>
<th>test</th>
<th>IOPS average</th>
<th>IOPS median</th>
<th>IOPS min</th>
<th>IOPS max</th>
<th>Bandwidth average MB/s</th>
<th>Bandwidth median MB/s</th>
<th>Bandwidth min MB/s</th>
<th>Bandwidth max MB/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>fileserver</td>
<td>7098</td>
<td>7166</td>
<td>4695</td>
<td>7907</td>
<td>169</td>
<td>171</td>
<td>112</td>
<td>188</td>
</tr>
<tr>
<td>webserver</td>
<td>8090</td>
<td>8063</td>
<td>7719</td>
<td>9053</td>
<td>40</td>
<td>40</td>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>randomwrite</td>
<td>8559</td>
<td>8755</td>
<td>5393</td>
<td>9466</td>
<td>66</td>
<td>68</td>
<td>42</td>
<td>73</td>
</tr>
<tr>
<td>openfiles</td>
<td>8518</td>
<td>8471</td>
<td>7916</td>
<td>9397</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>randomrw</td>
<td>9846</td>
<td>10162</td>
<td>3897</td>
<td>12494</td>
<td>76</td>
<td>79</td>
<td>30</td>
<td>97</td>
</tr>
<tr>
<td>mongo</td>
<td>6988</td>
<td>7004</td>
<td>3502</td>
<td>7005</td>
<td>22</td>
<td>23</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>randomread</td>
<td>8920</td>
<td>11259</td>
<td>70</td>
<td>12066</td>
<td>69</td>
<td>87</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>varmail</td>
<td>2931</td>
<td>3296</td>
<td>358</td>
<td>5627</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>
Questions?