Implementing SAN & NAS with Linux
by Mark Manoukian & Roy Koh

I invited Silent Gary to help us decide on a technology direction.

We think he's a genius because he has a beard and he never speaks.

Gary, do you think we should use open source software for our support platform?

Here it comes.

He's rubbing his beard and giving me creepy eye contact.

I detect a slight hint of disgust. It means Gary hates the idea!

Yes, it's all so obvious now. This is the worst idea in the history of mankind.

The meeting is over. Silent Gary has spoken.

You're actually a moron, aren't you?

Don't ruin this for me.
Housekeeping

• Evaluations
  – OSS2
  – Be Honest
  – Comments

• Questions? Raise your hand, wait for the Microphone.

• Check the on-line version of this presentation for updates.
North Central

Roy

Roy Koh
Network Administrator
Kegler, Brown, Hill & Ritter
Columbus, OH

318100531
• Nerd
• Geek
• 2\textsuperscript{nd} time ILTA attendee
• 1\textsuperscript{st} time speaker
“Babysitters direct to you in 20 minutes!”
STORAGE!

- **1956** - IBM Model 350, the first-ever hard disk drive.
- **About 4.4 MegaBytes!**
The Good News?

- Storage is getting cheaper!

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost per MegaByte</th>
</tr>
</thead>
<tbody>
<tr>
<td>1956</td>
<td>$10,000.00</td>
</tr>
<tr>
<td>1980</td>
<td>$192.31</td>
</tr>
<tr>
<td>1984</td>
<td>$87.86</td>
</tr>
<tr>
<td>1989</td>
<td>$12.00</td>
</tr>
<tr>
<td>1994</td>
<td>$0.95</td>
</tr>
<tr>
<td>1999</td>
<td>$0.02</td>
</tr>
<tr>
<td>2004</td>
<td>$0.00115</td>
</tr>
</tbody>
</table>
Cheap, Cheap, Cheap!

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost per GigaByte</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>$0.16</td>
</tr>
</tbody>
</table>

I AM MORDAC THE PREVENTER OF INFORMATION SERVICES. YOU HAVE EXCEEDED YOUR SERVER STORAGE LIMIT.

HERE’S 25 CENTS SO YOU CAN AFFORD TO DOUBLE MY STORAGE SPACE.

I THINK MY MYSTIQUE JUST TOOK A HIT.
Why are you here?

• Learn to build your own SAN or NAS (DIY)
• SAVE $$$
• Know your systems better – hardware, software
• Supplement your PRIVATE CLOUD

Sept 2009 BackBlaze built a 67 TB storage unit for $7,867
http://bit.ly/s0Emg
COST OF A PETABYTE

- RAW DRIVES: $81,000
- BACKBLAZE: $117,000
- DELL MD1000: $826,000
- Sun X4550: $1,000,000
- NetApp FAS-6000: $1,714,000
- Amazon S3: $2,806,000
- EMC NS-960: $2,860,000

* Amazon S3 Storage over three years (minus electricity, co-location and administration).
SAN & NAS in Linux

- SAN - AoE: Target & Initiator config
- SAN - iSCSI: Target & Initiator config
- NAS – Samba/CIFS
- Openfiler
- DRBD – Block level replication
What is SAN?

• Stands for Storage Area Network
• Wikipedia: an architecture to attach remote computer storage devices to servers in such a way that the devices appear as locally attached to the operating system
• **Block-level operations**
• Protocols: Fibre Channel, iSCSI, ATA over Ethernet (AoE)
• **SAN is not NAS**
What is NAS?

• Stands for Network Attached Storage.
• Wikipedia: A NAS unit is a computer connected to a network that only provides file-based data storage services to other devices on the network.
• Protocols used: NFS, SMB/CIFS, AFP.
What is Block Level?

- A **block** is a sequence of **bytes** or **bits**, having a nominal length (a block size).

- Wikipedia:
  Blocking is used to facilitate the handling of the data-stream by the computer program receiving the data. **Blocked data are normally read a whole block at a time.**
• SAN = Block Level
• NAS = File Level
The System

MOORE’S LAW: Computers will get exponentially faster.

MURPHY’S LAW: Anything that can go wrong will go wrong.

ROSSCOTT’S LAW: The faster the computer, the faster it can go wrong.

What do you MEAN “Unexpectedly Quit”??
SAN: AoE on Linux

• network protocol developed by the Brantley Coile Company
• light weight, does not rely on network layers (TCP or IP)
• designed for simple, high-performance access of SATA storage devices over Ethernet networks.
• like FCoE not like iSCSI
• AoE exists on Linux as a kernel module
AoE: Why use it?

• Ethernet: is getting faster
• and cheaper!
• and Ethernet can be multiported (switches).
• Enterprise ATA (or SATA) disks are also getting cheaper.
• **ATA over Ethernet (AoE) — a SAN at a fraction of the cost of Fibre Channel.**
AoE on Linux

SAN:
AoE TARGET
SERVING DRIVE SPACE

NETWORK

SERVER:
AoE INITIATOR
AoE Target setup on Linux SAN

• Hardware RAID – make sure you have Linux drivers
• Or you can use Software RAID
• Check AoE is enabled in kernel – modprobe aoe
• AoE Target setup:
  a. Install aoetools (sourceforge)
  b. export block devices
  vblade 0 1 eth0 /dev/sdb
AoE Target setup on Linux SAN

vblade 0 1 eth0 /dev/sdb

What does this mean?

vblade = command

1st number 0 = shelf
2nd number 1 = slot

eth0 = ethernet device at address 0
/dev/sdb = name of exported block device
AoE Target setup on Linux SAN

IMPORTANT!

If you are planning to have many AoE Target, make sure there are uniqueness to your shelf and slot #s!

```
vblade 0 1 eth0 /dev/sdb
```
AoE Initiator (Linux client)

On Linux server:
1. Make sure AoE exists in kernel as a module # modprobe aoe
2. # aoe-stat
3. # ls -l /dev/etherd
4. # mkfs.ext3 /dev/etherd/e0.0
5. # mount /dev/etherd/e0.0 /mnt
AoE Initiator (Windows client)

On Windows server:

## Screenshots of Starport Initiator

This software interface displays a list of devices categorized by type. The list includes:
- **Local RAM disk devices**
- **Remote iSCSI devices**
- **Virtual DVD devices**
- **Image File devices**
- **AoE devices**
- **FCoE devices**
- **RAID-1 (Mirror) devices**
- **RAID-0 (Strip) devices**
- **JBOD devices**

### Device Details:

<table>
<thead>
<tr>
<th></th>
<th>Type</th>
<th>Vendor</th>
<th>Product</th>
<th>Revision</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATA</td>
<td>Coraid EtherDriver</td>
<td>V1.6</td>
<td>Disk</td>
<td>Connected</td>
</tr>
<tr>
<td>2</td>
<td>ATA</td>
<td>Coraid EtherDriver</td>
<td>V2.0</td>
<td>Disk</td>
<td>Connecting</td>
</tr>
</tbody>
</table>
SAN: iSCSI on Linux

• iSCSI = Internet Small Computer Systems Interface
• IP-based storage networking standard for linking data storage facilities
• Uses TCP/IP Ports 860 and 3260
iSCSI on Linux

SAN:
iSCSI TARGET
SERVING DRIVE SPACE

NETWORK

SERVER:
iSCSI INITIATOR
iSCSI Target setup on Linux SAN

- Make sure hardware RAID is supported by Linux or use software RAID
- Or in Debian
  ```bash
  # apt-get install iscsitarget
  ```
- Edit config files:
  ```bash
  # nano /etc/default/iscsitarget
  # nano /etc/ietsd.conf
  # nano /etc/initiators.allow
  ```
- Start the service...
iSCSI Initiator (Linux client)

On Linux:

1. Install open-iscsi
   http://www.open-iscsi.org/
   or apt-get install open-iscsi in Debian

2. Edit config file
   # nano /etc/iscsi/iscsid.conf

SERVER:
iSCSI INITIATOR
iSCSI Initiator (Windows client)

On Windows:

1. Windows 7 & Windows 2008 is built-in

2. Windows XP & Windows 2003: Download MS iSCSI software initiator v2.08
   http://bit.ly/2F7AC1

3. Or use Starport’s Initiator
SAN Best Practices

• Design and Plan your network
• Network isolation (security)
  - Logical network isolation – VLAN
  - Physical network isolation
• Replication (Block level replication)
QUESTIONS

• When setting up your Linux SAN

THE HIGHLY PRODUCTIVE BUT USELESS GUY

HERE’S A COPY OF MY WHITE PAPER.

IT’S A STATISTICAL ANALYSIS OF THE CORRELATION BETWEEN DISK STORAGE AND EMPLOYEE ABSENTEEISM.

I DON’T KNOW HOW TO DO STATISTICS BUT IT DOESN’T MATTER BECAUSE I DIDN’T HAVE DATA.

• rkoh@keglerbrown.com
NAS on Linux

• Centralized “file-based” storage
• Wikipedia- “NAS unit is a computer connected to a network that only provides file-based data storage services to other devices on the network.”
• “NAS uses file-based protocols such as NFS (popular on UNIX systems), SMB/CIFS (used with MS Windows systems), or AFP (used with Apple Macintosh computers). NAS units rarely limit clients to a single protocol.”
NAS SERVER:
SERVING SHARED
FOLDERS – NFS OR SMB
OR AFP, etc

NETWORK

NAS CLIENT:
SERVER OR WORKSTATION
Simply said:

• NAS is file sharing
• NAS can be a file share on a server
• NAS can also be a dedicated hardware device
• NAS can also be an embedded device
The easy path...

How to get things working.

1. You know this metal rectangle full of little lights?
   - Yeah.

2. I spend most of my life pressing buttons to make the pattern of lights change however I want. Sounds good.

3. But today, the pattern of lights is all wrong! Oh god! Try pressing more buttons! It's not helping!
Linux Distro: Openfiler NSA

• Web-based GUI
• Multiple NICs (NIC bonding)
• Protocol rich:
  1. Block level: iSCSI, Fibre Channel
  2. File level: NFS, SMB/CIFS, Http/DAV, FTP, rsync
• Does not support AoE out of the box BUT…
Openfiler NSA (SAN)

1. iSCSI is built-in-GUI-web-configure-goodness.

2. AOE is not built-in.
   aoetools needs compiling and it works.
   a. before compiling, install gcc
      # conary update gcc
   b. download aoetools from Sourceforge
      # make && make install
Openfiler NSA (NAS)

- Industry-standard protocol suite — SMB, NFS, FTP, WebDAV
- Share management
- Quota / resource allocation — group, user, guest
- Accounts management — PAM, NIS, LDAP, Active Directory
Openfiler NSA (NAS screenshots)

SMB settings

Server string: Openfiler NAS
NetBIOS name: OPENFILER
WINS server: 
Passwords: Use encrypted passwords
Winbind Policy: Use default domain
LDAP User Suffix: ou=People
LDAP Group Suffix: ou=Group

Services section
- Manage Services
- SMB/CIFS Setup
- LDAP Setup
- UPS Setup
- Rsync Setup
- iSCSI Target Setup
- FTP Setup

Support resources
- Report bug
- Get support
- Forums
- Admin Guide
Other Distros

• FreeNAS (FreeBSD) – SAN and NAS
  SMB/CIFS, FTP, NFS, TFTP, AFP, RSYNC, Unison, iSCSI (initiator and target)

• Other Linux distros –
  a. Debian-like distros, Ubuntu, etc. – SAN and NAS = command line
  b. Other Linux distros
REPLICATE!

WHAT THE...?

DON'T BE SURPRISED.

IN ANY LARGE COMPANY THERE IS AT LEAST ONE EMPLOYEE WHO IS YOUR EXACT REPLICA AND HAS THE SAME ASSIGNMENT.

WHY DIDN'T I KNOW THAT?

I'M NOT YOUR REPLICA. I'M A LOOK-ALIKE THAT IS MUCH SMARTER.

www.dilbert.com
Why Replicate Data?

• Fail safe (backup)
• HA (high availability)
• Wikipedia: “Replication is the process of sharing information so as to ensure consistency between redundant resources, such as software or hardware components, to improve reliability, fault-tolerance, or accessibility.”
Software Replication

• SAN Block Level Replication – DRBD (*Distributed Replicated Block Device*)
On to DRBD and beyond...

Richard Stallman!
Your viral open source licenses have grown too powerful.
The GPL must be stopped.
At the source.
You.

Zzzz

Hah! Microsoft lackeys!
So it has come to this.

Shing!

A night of blood I've long awaited.
But be this my death or yours.
Free software will carry on!
For a GNU dawn! For freedom!

...Hey, where are you going?

Man, you're right,
That never gets old.

Let's do Eric S.
Raymond next.

Or Linus Torvalds. I hear
He sleeps with nunchucks.
62 Attorneys
Columbus
4/8 TB

Offices

4/8 TB
Marion
1 Attorney

2 Attorneys
Cleveland
Solutions: Pieces and Parts

- AoE
- DRBD
- OpenFiler
- iSCSI
Distributed Replicated Device = DRBD Block Device
Replicated Block Device

• RAID 1

• Between servers, instead of hard drives.
What’s it mean?

- Raid 1
- Across Servers, Instead of Drives
Runs on Linux

- RedHat/CentOS
- Suse
- Debian
- Ubuntu
I Use Ubuntu

• Create identically, sized unmounted partitions.
  – One partition for DRBD metadata.
  – One partition for user data.
• Give servers static IPs.
• Make sure time on servers is in sync.
Implementing DRBD

- All of this is command line...
- Install DRBD from repository.
- Configure drbd.conf file.
  - Complementary
  - Primary or Secondary
- Load DRBD: modprobe DRBD
- Install and configure Heartbeat from repository.
- Set virtual IP address.
- Start DRBD.
What does it look like?

<table>
<thead>
<tr>
<th>Filesystem</th>
<th>Size</th>
<th>Used</th>
<th>Avail</th>
<th>Use%</th>
<th>Mounted on</th>
</tr>
</thead>
<tbody>
<tr>
<td>/dev/sda5</td>
<td>4.6G</td>
<td>430M</td>
<td>4.0G</td>
<td>10%</td>
<td>/</td>
</tr>
<tr>
<td>tmpfs</td>
<td>126M</td>
<td>0</td>
<td>126M</td>
<td>0%</td>
<td>/dev/shm</td>
</tr>
<tr>
<td>/dev/sda1</td>
<td>89M</td>
<td>11M</td>
<td>74M</td>
<td>13%</td>
<td>/boot</td>
</tr>
<tr>
<td>/dev/drbd0</td>
<td>24G</td>
<td>33M</td>
<td>23G</td>
<td>1%</td>
<td>/data</td>
</tr>
</tbody>
</table>
File Systems

- NFS
- CIFS\SMB