

SAN AoE (ATA OVER ETHERNET)

PARIS - 1 DÉCEMBRE 2010

Gabriel Moreau

LEGI - Laboratoire des Ecoulements Geophysiques et Industriels

CNRS / UJF / G-INP - Grenoble, France

<http://www.legi.grenoble-inp.fr/>

1 décembre 2010

- 1 Généralités sur l'AoE
- 2 Fabricant de baie SAN AoE
- 3 Monter son propre serveur SAN (target) AoE
- 4 Pourquoi j'ai choisi AoE

- 1 Généralités sur l'AoE
- 2 Fabricant de baie SAN AoE
- 3 Monter son propre serveur SAN (target) AoE
- 4 Pourquoi j'ai choisi AoE

Généralités sur l'AoE

Un p'tit dessin pour bien commencer

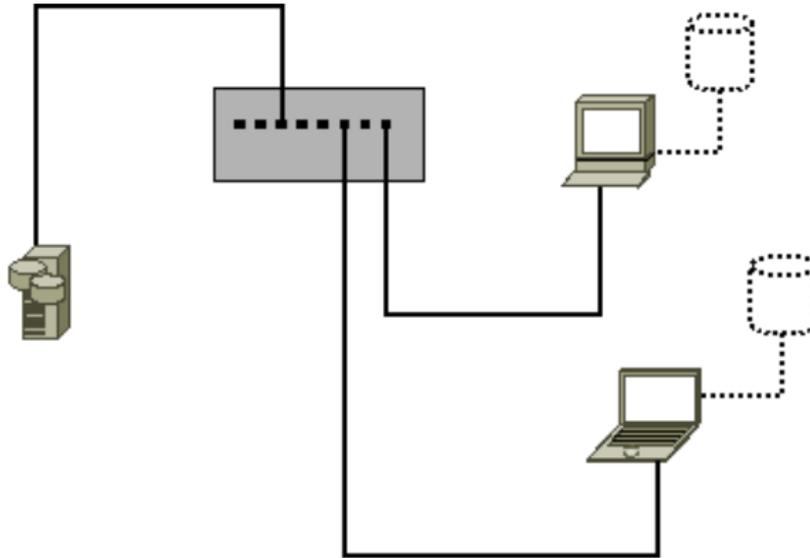
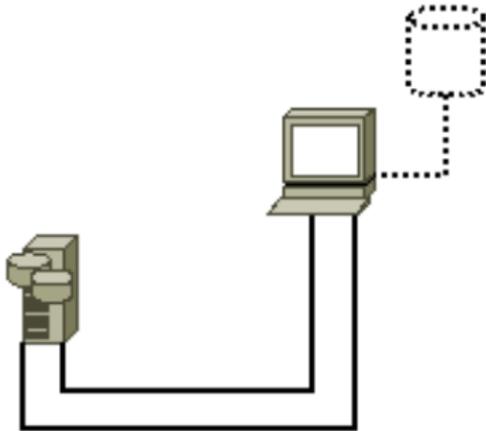


Schéma général d'un SAN

Généralités sur l'AoE

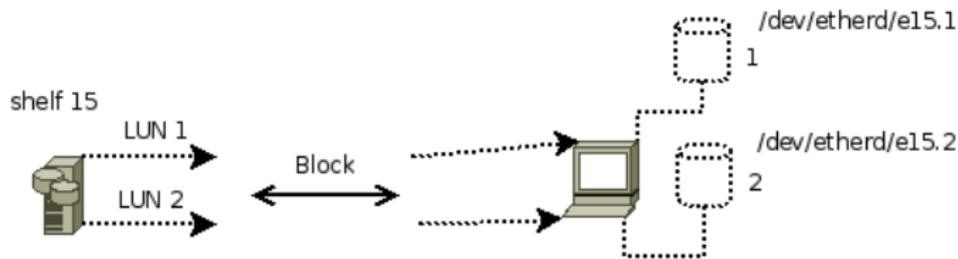
SAN avec attachement direct



- Attachement simple ou double. . . ou plus.
- DAS comme une baie SAS, SCSI ou eSATA. . .

Généralités sur l'AoE

SAN : Comment ça marche ?



- Chaque baie a un numéro unique (ex shelf 15)
- Une baie exporte un ou plusieurs LUN (disques virtuels)
- Un ou des clients montent ces LUN (ex `/dev/etherd/e15.1`)
- Montage local (ex `mount /dev/etherd/e15.1 /mnt/data`)
- Dialogue client (initiator) / serveur (target) en mode block
- Les blocks transitent sur le réseau Ethernet

Généralités sur l'AoE

Une baie AoE, à quoi cela ressemble



Au dos

- Ports Ethernet
- Port console série

...à une autre baie de stockage !

Généralités sur l'AoE

Comparaisons des protocoles SAN

Protocoles SAN : couches protocolaires

FC	iSCSI	FCoE	AoE
File System	File System	File System	File System
SCSI	SCSI	SCSI	ATA
(Multi-Path Driver)			
FC	TCP IP	FCoE	AoE
FC	Ethernet	DCB	Ethernet

- DCB : Data Center Bridging

Généralités sur l'AoE

Qu'est-ce que c'est ?

- Protocole créé en 2004 par la société Coraid
- Spécification libre
- Conçu pour être simple
- Conçu pour être performant dans les accès disques SATA
- Fine couche au dessus d'Ethernet
- Basé sur des composants réseaux standard
- Driver AoE intégré dans le noyau Linux depuis 2005
- Spécification iSCSI : 257 pages
- Spécification AoE : 12 pages !

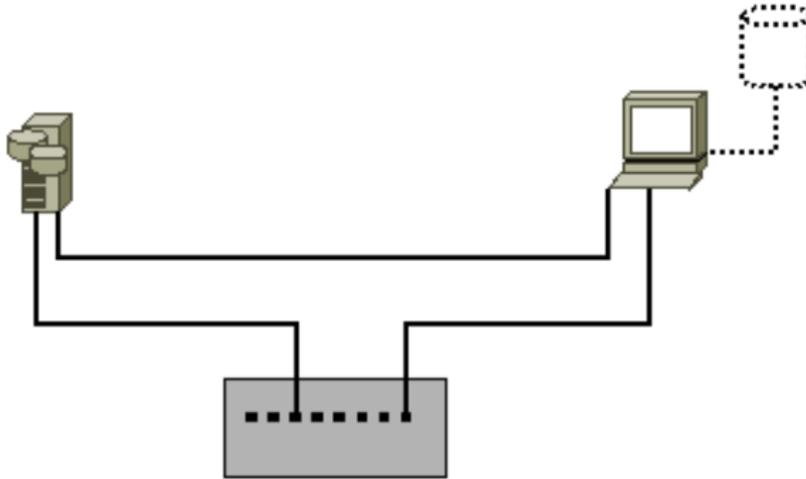
Généralités sur l'AoE

Avantages

- Pas de couche IP (donc pas de TCP)
- Pas routable (sauf tunnel)
- Plus proche du FCoE que de l'iSCSI
- Protocole léger
- Limite la charge réseau dû au protocole
- Limite la charge sur les serveurs
- L'implémentation est plus facile
- Moins de risque de bogue (aspect sécurité)

Généralités sur l'AoE

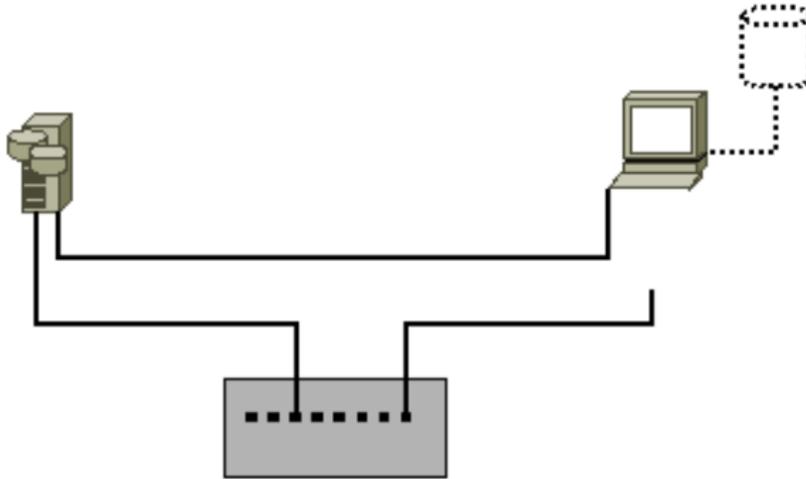
Avantage : Multi-path natif



Multi-path natif - rien à configurer

Généralités sur l'AoE

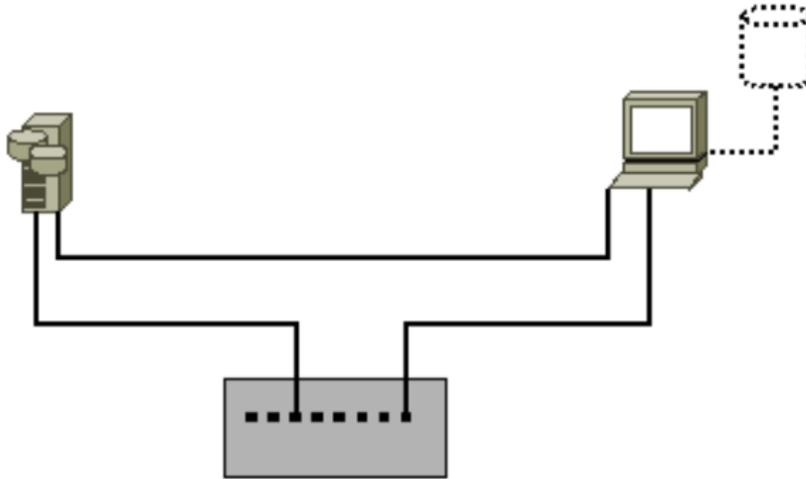
Avantage : Multi-path natif



Fonctionne même si un chemin est coupé...

Généralités sur l'AoE

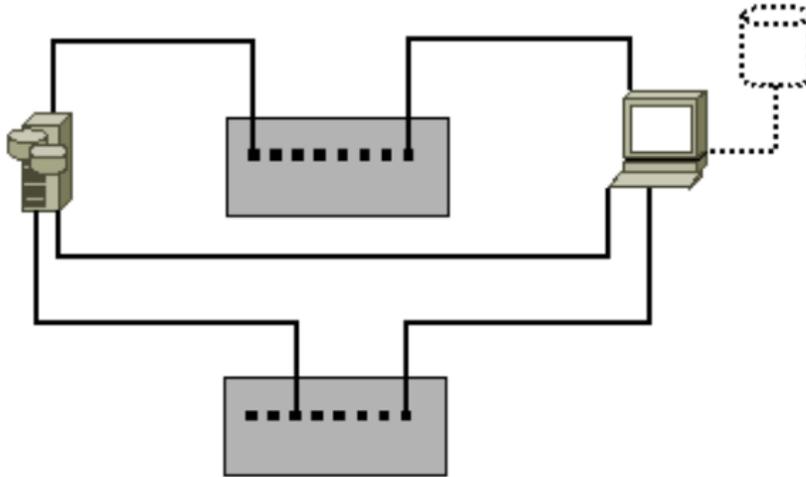
Avantage : Multi-path natif



Fonctionne même si un chemin est coupé... puis remis

Généralités sur l'AoE

Avantage : Multi-path natif



Ajout de chemin à chaud (aoe-discover aoe-stat aoe-revalidate)

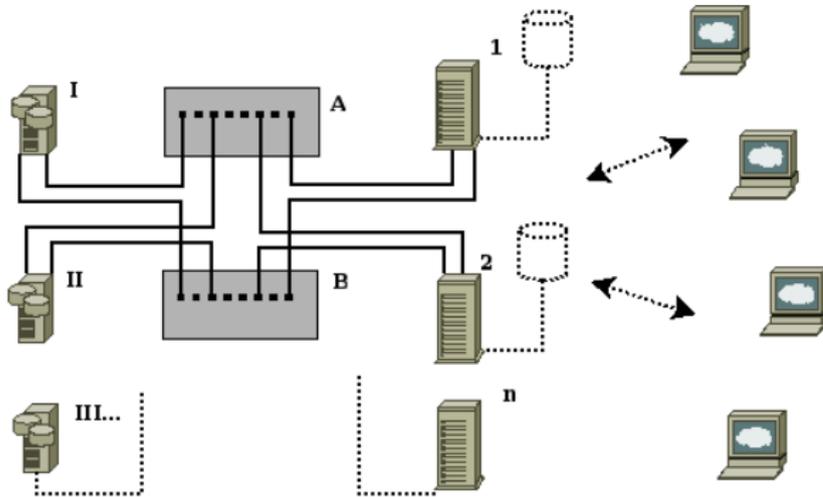
Généralités sur l'AoE

Avantages

- Multi-path natif
- Commutateur standard (jumbo frame conseillé)
- Maximum de la bande passante pour les données
- A qualité égale, c'est plus performant
- Permet de concevoir une architecture à base de SAN à bas prix

Généralités sur l'AoE

Exemple : Haute disponibilité



Exemple
Typologie
d'architecture
haute disponibilité

Généralités sur l'AoE

Comparaisons des protocoles SAN

Protocoles SAN : couches protocolaires

FC	iSCSI	FCoE	AoE
File System	File System	File System	File System
SCSI	SCSI	SCSI	ATA
(Multi-Path Driver)			
FC	TCP IP	FCoE	AoE
FC	Ethernet	DCB	Ethernet

- TRILL : Transparent Interconnection of Lots of Links

Généralités sur l'AoE

Sécurité

Sécurité simple et basique

- Filtrage des accès par les adresses physiques (MAC)
- VLAN dédié
- Sécurité des données → chiffrement des disques par le client

- 1 Généralités sur l'AoE
- 2 Fabricant de baie SAN AoE
- 3 Monter son propre serveur SAN (target) AoE
- 4 Pourquoi j'ai choisi AoE

Fabricant de baie SAN AoE

Principaux fabricants

- Coraid <http://www.coraid.com/>



Mais encore...

- Du côté de la chine...
- LayerWalker
<http://www.layerwalker.com/>
- ...
- Tout cela semble très flou !

Fabricant de baie SAN AoE

Revendeur en France

- Alyséo <http://www.alyseo.com/>
- Procédure d'achat simple
- Évite les marchés (choix de la technologie AoE)
- Prix dépendant de la politique de Coraid
- Prix en forte hausse ces derniers temps. . .

Fabricant de baie SAN AoE

SAN haut de gamme



SRX4200 - 36 Disk High Performance Ethernet

- 4U Ethernet SAN storage appliance
- 72 TB Capacity (with 2TB disks)
- Supports 36 SAS, SSD or SATA hot swap disks
- Up to four 10 GigE or six GigE interfaces
- Supports RAID 0,1,5,6,10 or JBOD
- Access speed > 1800 MB/s
- Redundant hot swap power supply and fans

Fabricant de baie SAN AoE

SAN entrée de gamme



SR821 - 8 Disk Cost-Effective Ethernet

- 2U Ethernet SAN storage appliance
- 16 TB Capacity (with 2 TB disks)
- Supports 8 hot swap SATA disks
- Dual GigE interfaces
- Supports RAID 0,1,5,6,10 or JBOD
- Access speed > 200 MB/s
- Redundant hot swap power supply and fans

Fabricant de baie SAN AoE

Autour du SAN

Une gamme complète

- Management - Appliance de monitoring et contrôle d'un parc de baie SAN
 - Gateway - Appliance passerelle vers IP pour liaison WLAN (backup distant)
 - Mirroring - Appliance de gestion de la haute disponibilité
 - HBA - Carte réseau avec le protocole AoE câblé en dur (driver non libre)
-
- Voir site web Coraid <http://www.coraid.com/>
 - Non testé car hors de portée de mon budget !

- 1 Généralités sur l'AoE
- 2 Fabricant de baie SAN AoE
- 3 Monter son propre serveur SAN (target) AoE**
- 4 Pourquoi j'ai choisi AoE

Monter son propre serveur SAN (target) AoE

AoE Targets (serveur) disponibles

- vblade <http://aoetools.sourceforge.net/>
- kvbade (mort ?)
- ggaoed <http://code.google.com/p/ggaoed/>
- aoedriver (mort ?)
<http://code.google.com/p/aoedriver/>
- ...

Monter son propre serveur SAN (target) AoE

vblade par la pratique

- Liminaire : `apt-get install vblade`
- Commande : `vblade[d] [-m mac[,mac...]] shelf slot netif filename`
- Exemple : `vblade 15 1 eth0 /dev/vg15/part1`
- C'est tout !

Coté client (initiator) AoE

AoE par la pratique

- Liminaire : `apt-get install aoetools`
- Chargement du module noyau : `modprobe aoe`
- Découverte : `aoe-discover`
- Formater : `mkfs.xfs /dev/etherd/e15.1`
- Montage : `mount /dev/etherd/e15.1 /mnt/data`
- C'est tout !

Coté client (initiator) AoE

- Compiler la dernière version de aoe (aoetools)
- Compiler cec (console des baies Coraid sur AoE)
- Limiter AoE sur certaines interfaces

```
modprobe aoe aoe_iflist="eth1"
```

- Monitorer : aoe-stat

```
e15.1 10002.048GB eth1 up
```

En savoir plus

- <http://blog.linuxing.org/2010/01/how-to-build-a-low-cost-san/>
- <http://www.debian-administration.org/articles/553>

- 1 Généralités sur l'AoE
- 2 Fabricant de baie SAN AoE
- 3 Monter son propre serveur SAN (target) AoE
- 4 Pourquoi j'ai choisi AoE

Pourquoi j'ai choisi AoE

Retour en 2007

- Raz le bol des baies SCSI
- Attachement direct → déplacement très difficile
- Multi-attachement ?

Essai Coraid : SR421

- AoE et Coraid avait la côte dans certains milieux
- Tout sur Ethernet
- Simple, basique, efficace, pas de chichi
- Client Linux libre (GPL) intégré dans debian
- Prix tout à fait correct
- Coraid SR421 - 1U avec 4 disques de 1To

Pourquoi j'ai choisi AoE

A ce jour

- 9 baies Coraid peu puissantes dont 2 petites
- Environ 200 To en RAID 5 ou 6
- Aucune panne matérielle
- Des disques qui cassent mais aucun RAID

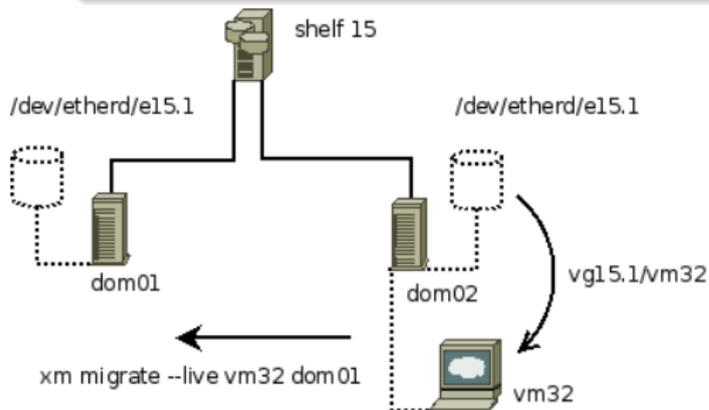
Configuration des dernières baies

- 4U - 24 disques de 2To
- Privilégie les données devant les performances
- Deux RAID 6 sur 11 disques, deux disques de spare
- Assemblage avec LVM
- Formatage XFS de volume taillé dans les LVM

Pourquoi j'ai choisi AoE

Utilisation

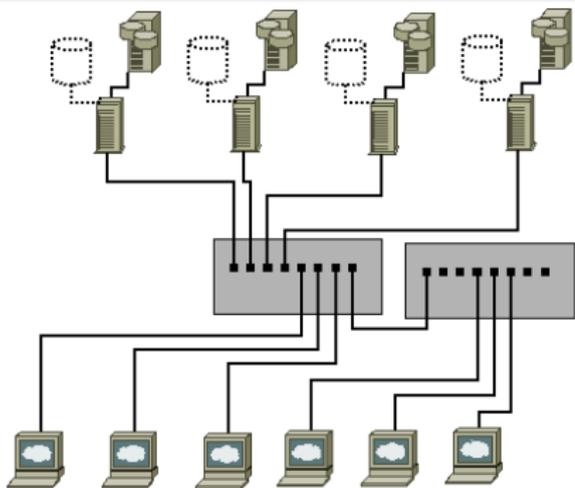
- Stockage calcul : attachement direct - serveur NFS
- Stockage dossiers partagés - serveur samba + NFS
- Recyclage backup



Stockage machine domU Xen migration possible

- Interruption du ping < 2s
- Attention LVM en cluster, `lvscan` avant chaque commande

Pourquoi j'ai choisi AoE



Le futur

- Encore plus de stockage
- Encore plus de clients

Consolidation du stockage et de son utilisation

- NFS + AutoFS + cfengine / pNFS ?
- Samba + DFS + cfengine
- **Glusterfs + cfengine**