Stockage distribué

Pourquoi maintenant ?

Romaric David & Pierre Rolland

david@unistra.fr, pierre.rolland@univ-rennes1.fr

Unistra / Direction Informatique

Institut des Sciences Chimiques de Rennes

4 Mai 2016

UNIVERSITÉ DE STRASBOURG

Plan

- Introduction
- Une évolution du stockage
- Pourquoi est-ce possible ?
- Conclusion

Introduction

- Cette journée se veut un ensemble de retours d'expériences pratiques sur la mise en place d'architectures de stockage(s) distribué(s)
- En provenance de la communauté Enseignement-Supérieur Recherche
- Stockage distribué : un thème actif en ce moment :
  - Journée du réseau Argos 24/03/2016
  - Journées Succes 2015 :
    Table ronde Stockage distribué
  - Ce tutoriel
  - et.... (cf fin de la présentation)

Introduction

- Pourquoi ce Tuto maintenant ?
  - L'organisation a commencé mi 2015 (initialement : Cloud)
  - Explosition du nombre de logiciels et de filesystems :
    - Ceph, HdFS, MooseFS, BeeGFs, RozoFS, Scality, Spectrum Scale, Lustre, GlusterFS...
  - Certains bénéficient de support commercial
  - Des réalisations de plus en plus nombreuses dans les laboratoires, les DSI, …
- Pourquoi un tel choix ?
- Faisons un point d'étape avec vous !

Plan

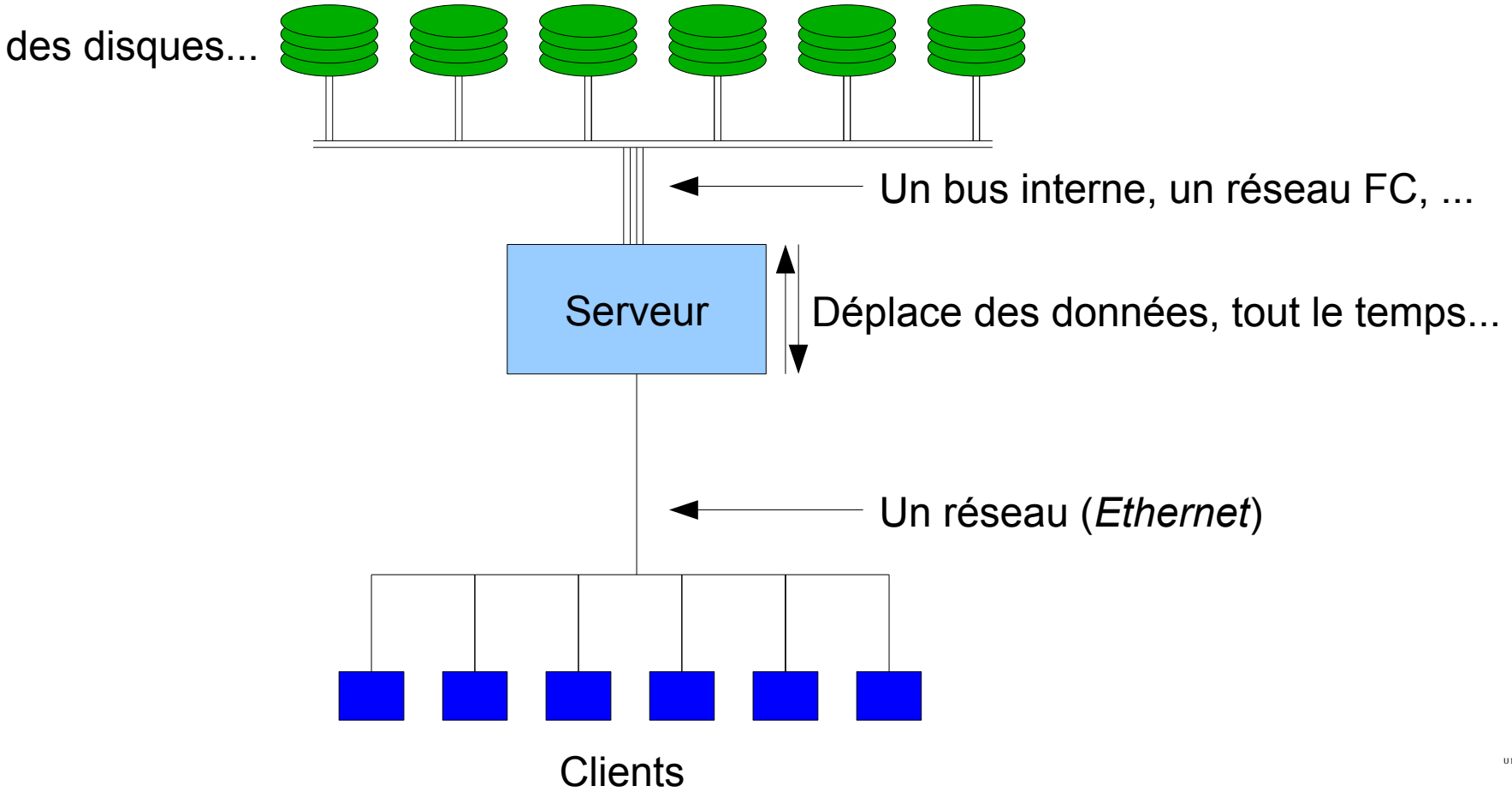
- Introduction
- Une évolution du stockage
- Pourquoi est-ce possible ?
- Conclusion

Une évolution du stockage

- La combinaison stockage-réseau existe depuis longtemps !
  - NFS a été introduit en 1985 !
    - Quelques composants : UDP, TCP, Scsi, PATA, Ethernet....
  - Les réseaux de stockage ont été introduits vers 1995
    - Un réseau dédié au stockage : Fiber Channel
- Et puis on a commencé à tout mélanger :
  - iSCSI (2001), FcoE (2007)
  - Un réseau holistique : Ethernet, des switchs duaux
- Dans tous les cas, une architecture *Nord-Sud* : le stockage à un endroit, le traitement ailleurs !

Une évolution du stockage

- Le stockage ressemblait donc à cela :



Et au niveau logiciel ?

- Sur les clients : des couches simples : NFS, CIFS, Samba, Apple File System, …
- Sur les serveurs : des drivers de carte raid, des drivers HBA FC, un accès en mode bloc
- Passage à l'échelle : quelques optimisations :
  - Daemons NFS en mode noyau, paramétrages réseau, TCP-Offloading, …
  - Répartition de charge ???
  - Pas de changement d'architecture majeur

Quelques logiciels disruptifs

- Raid logiciel, LVM (2001), ZFS (2005)
  - Redondance et tolérance aux pannes
  - Fonctionnalités évoluées (snapshots, compression, ...)
  - **Pour matériel banalisé !**
- Fuse (2005 dans le kernel Linux)
  - Filesystem in UserSpacE
  - Simplifie le développement de systèmes de fichiers !
    No more kernel panic
  - À entraîné une explosion du nombre de filesystems

Quelques logiciels disruptifs

- Dès 2001, on peut construire du stockage fiable sur du matériel générique.
- Et les systèmes de fichiers distribués ?
  - Avant 2000, principalement Andrew File System (commercial, reprise incomplète dans le noyau Linux) et GPFS
  - Année bascule : 2004 : explosion de l'offre

Plan

- Introduction
- Dans les machines
- Pourquoi est-ce possible ?
- Conclusion

Accéder aux données en réseau ?

- Est-ce que le réseau provoque des pertes de performances lors de l'accès aux données ?
- Comparons l'évolution des bandes passantes des disques durs et des réseaux

Année	Techno. Disque	Débit (MB/s)	Ethernet	Débit (MB/s)	Ratio
2002	Pata (IDE)	133	1gb	120	1,11
2010	Sata / Sas	600	10gb	1200	0,5
2012	SSD, Nvme	3000	100Gb	12000	0,25

Accéder aux données en réseau ?

- Il faut de plus en plus de disques durs en parallèle pour saturer un lien réseau !
- Disques et liens réseaux peuvent être mis en parallèle
- Hormis les points liés à la latence, ce n'est pas le réseau qui ralentira l'accès aux données situées sur les disques
- L'algorithmique permet de supporter la perte d'un ou plusieurs serveur (codes à effacement, réplication)

Accéder aux données en réseau ?

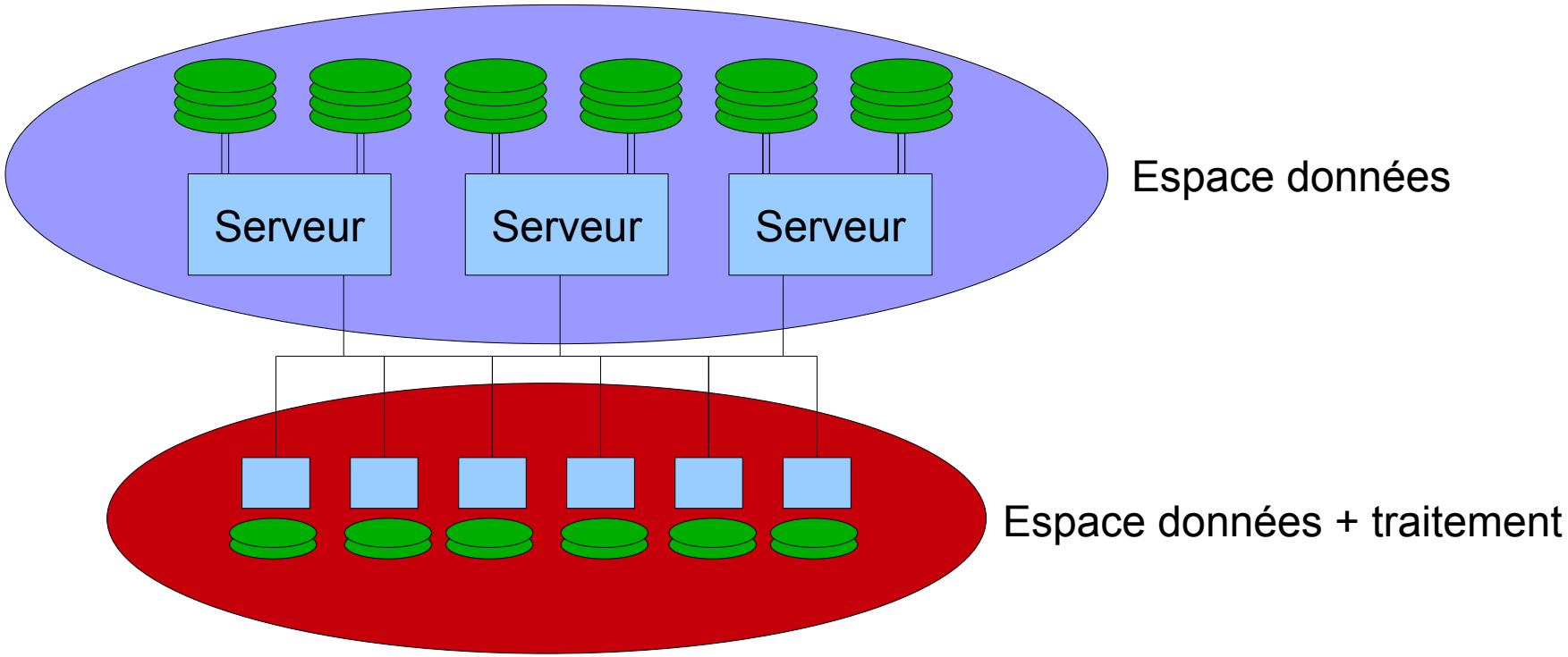
- La multiplication des serveurs d'entrée-sortie permet d'augmenter les débits et le nombre d'accès simultanés (scale-out)
- Éventuellement chaque serveur peut voir sa capacité de stockage augmenter (scale-up)
- Tendance : scale-up et scale-out en standard, idéalement illimités
- Le couplage des serveurs est l'oeuvre du logiciel (Software Defined Storage)

Les utilisations du stockage en réseau

- En termes d'usages , on peut observer la spécialisation des systèmes de stockage réseau,
  - les espaces *rapides*
  - les espaces sécurisés (« coffre-forts numériques »)
  - les espaces de type objet
- L'utilisateur doit choisir l'espace le plus approprié en fonction de la typologie de ses données (**les** big data)
- Rapprocher traitement et données
- 1 usage = 1 stockage ?

Une évolution du stockage

- Le stockage d'aujourd'hui ressemblerait-il à cela ?



- La fin du système de stockage unique ?

Plan

- Introduction
- Dans les machines
- Pourquoi est-ce possible ?
- Conclusion

Conclusion

- Aujourd'hui, chaque système SDS définit à sa façon la répartition des données entre serveurs
- À l'avenir, pourra-t-on changer de SDS sans transférer des données, en travaillant sur les méta-données ?
  - Certains projets en font la demande
  - Un autre travail algorithmique !
- Comment se prémunit-on des pannes du logiciel ?
- Migration automatique des données en fonction de leur température ? Tiering de SDS ? Un équivalent de lessfs/btier en réseau ?

Conclusion

- La formation sur les systèmes de stockage distribués continue
- Décembre 2016 : École CNRS de 5 jours organisée à Gif-sur-Yvette

  Des données au BigData: exploitez le stockage distribué
- Inscription à partir de Juin sur https://indico.mathrice.fr/event/5/overview