

GlusterFS

Étudiant : Khaled Basbous
Professeur : Gérald Litzistorf

23 avril 2012

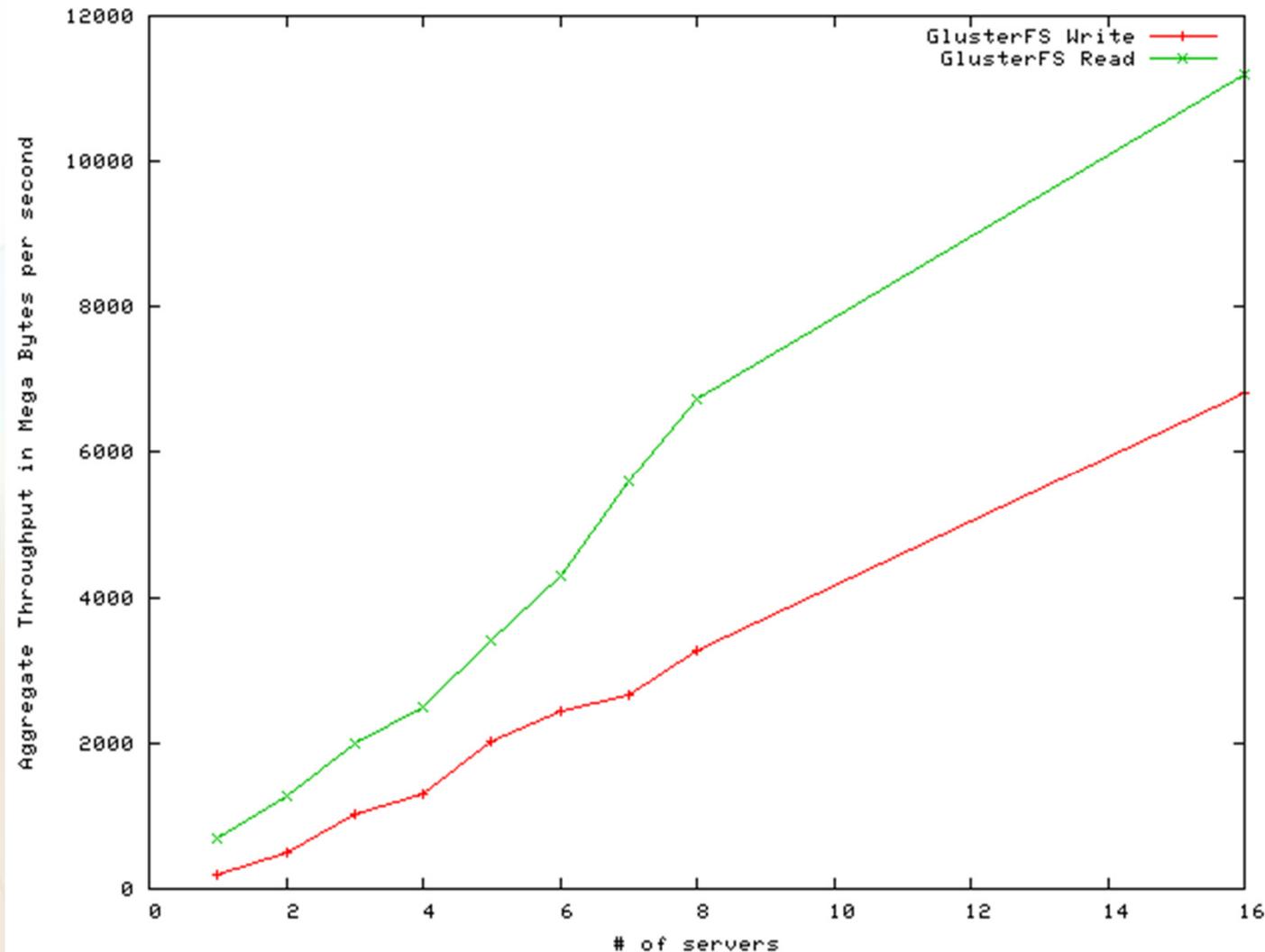
GlusterFS

- Logiciel sous licence GPLv3
- Racheté par Red Hat pour 136 Mio \$ en 2011
- Système de fichier distribué, répliqué
- Cluster de stockage
- Espace de stockage jusqu'à n pétaoctets
- Sur matériels usuels
- Performances linéaires !

Points forts

- Espace de stockage
- Performances linéaires
- Divers architecture (Distribué, répliqué, découpé)
- Géo-Réplication
- Access Control List
- Augmentation de l'espace de stockage à la volé
- Facilité de déploiement
- Compatibilité avec d'autres protocoles
- Quotas

Performance



Ext2/3

Directory

Name	Inode
Name	Inode
...	...

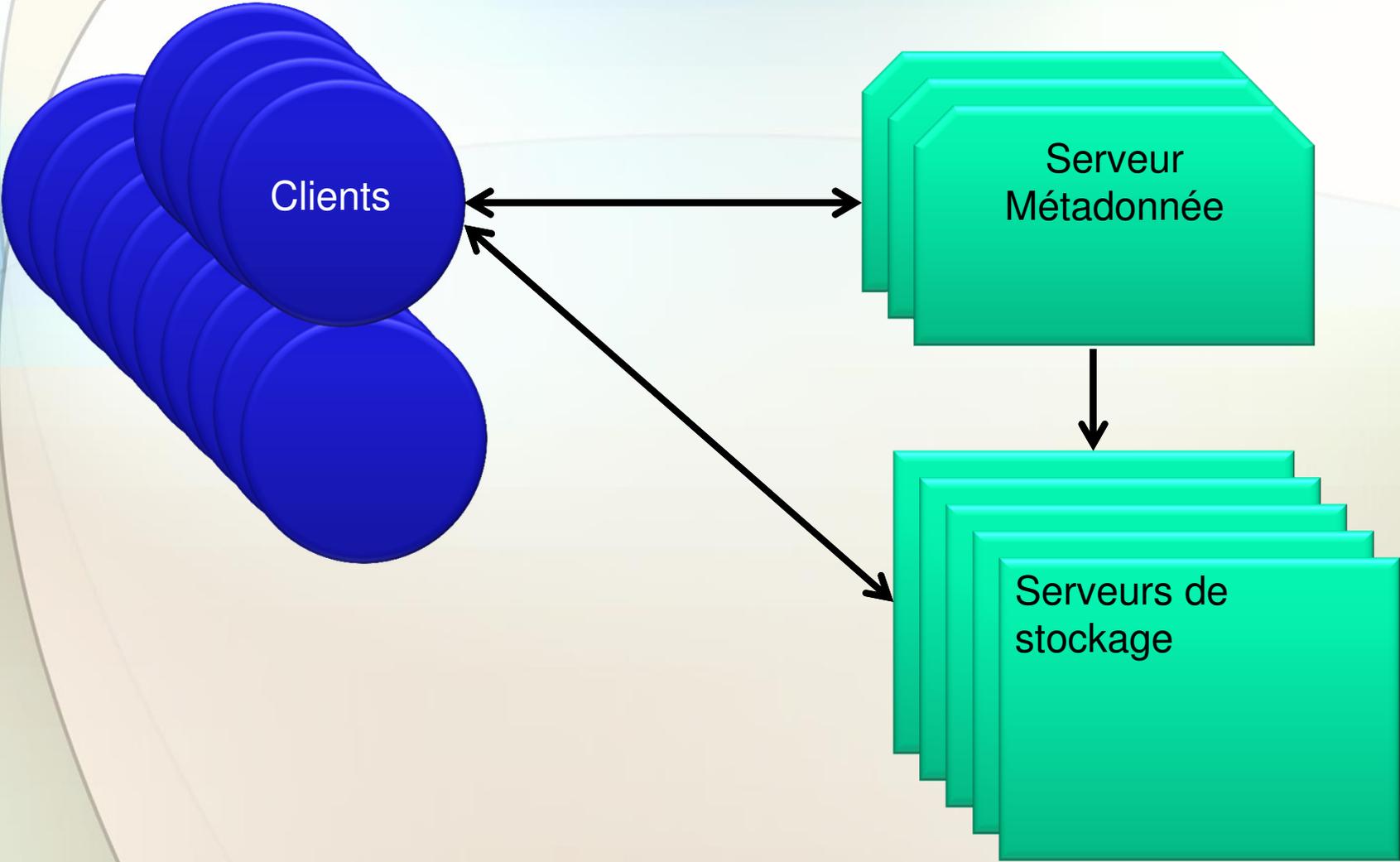
Inode

File type
Access rights
Owner
Size
Access time
Group
Number of links
Number of sectors
Data block 1
Data block 2
...
Data block 15
ACL
...

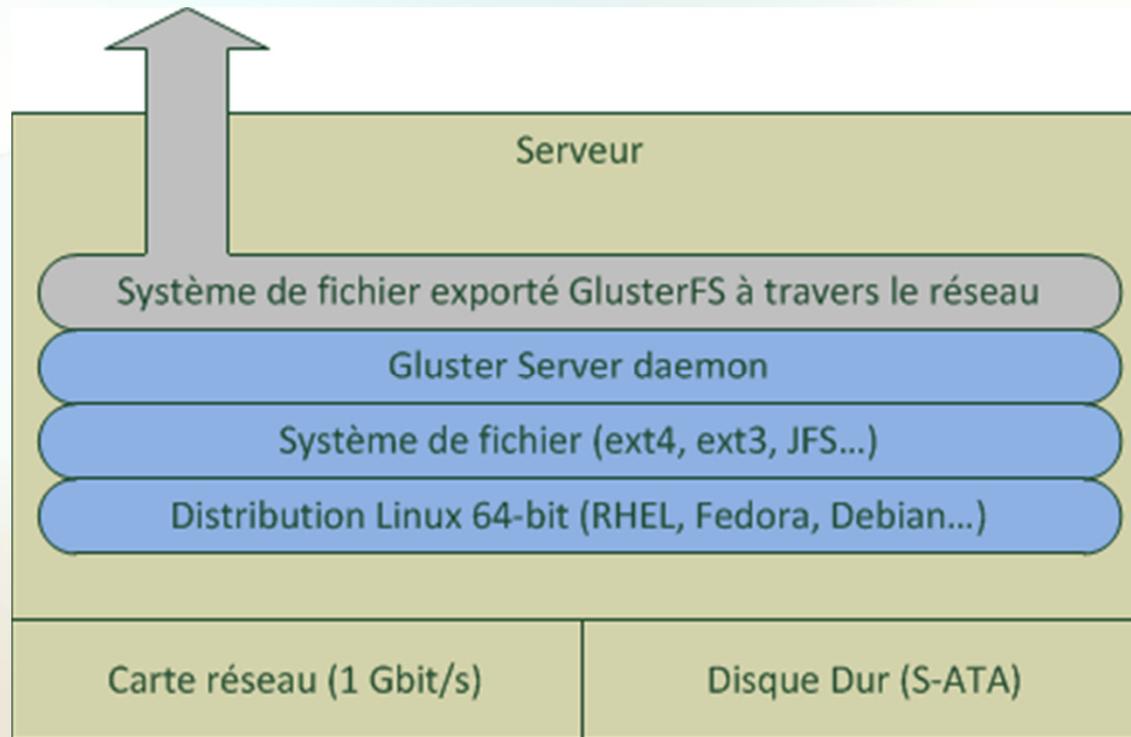
Data blocks

Data
Data

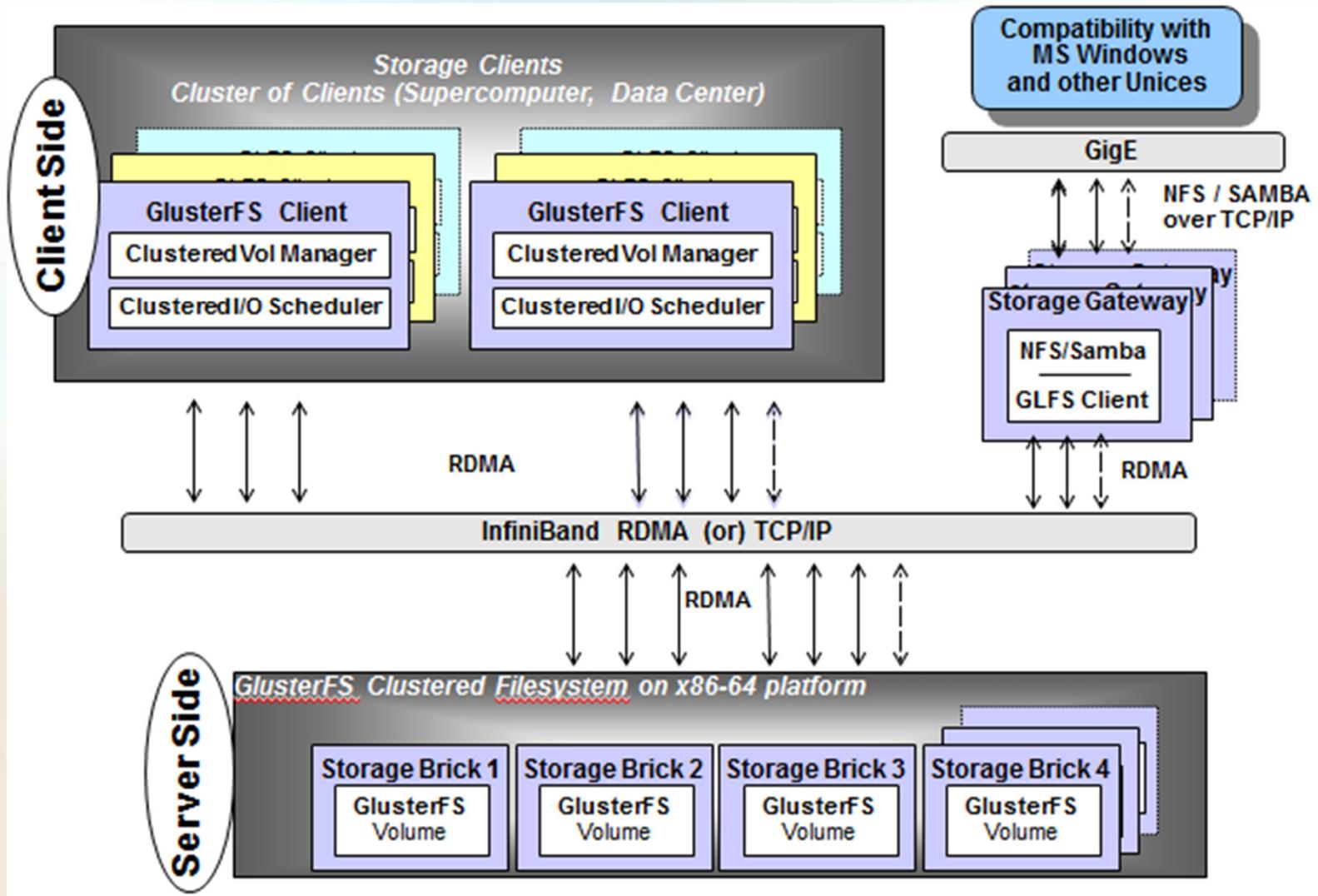
Exemple Ceph



Serveur GlusterFS



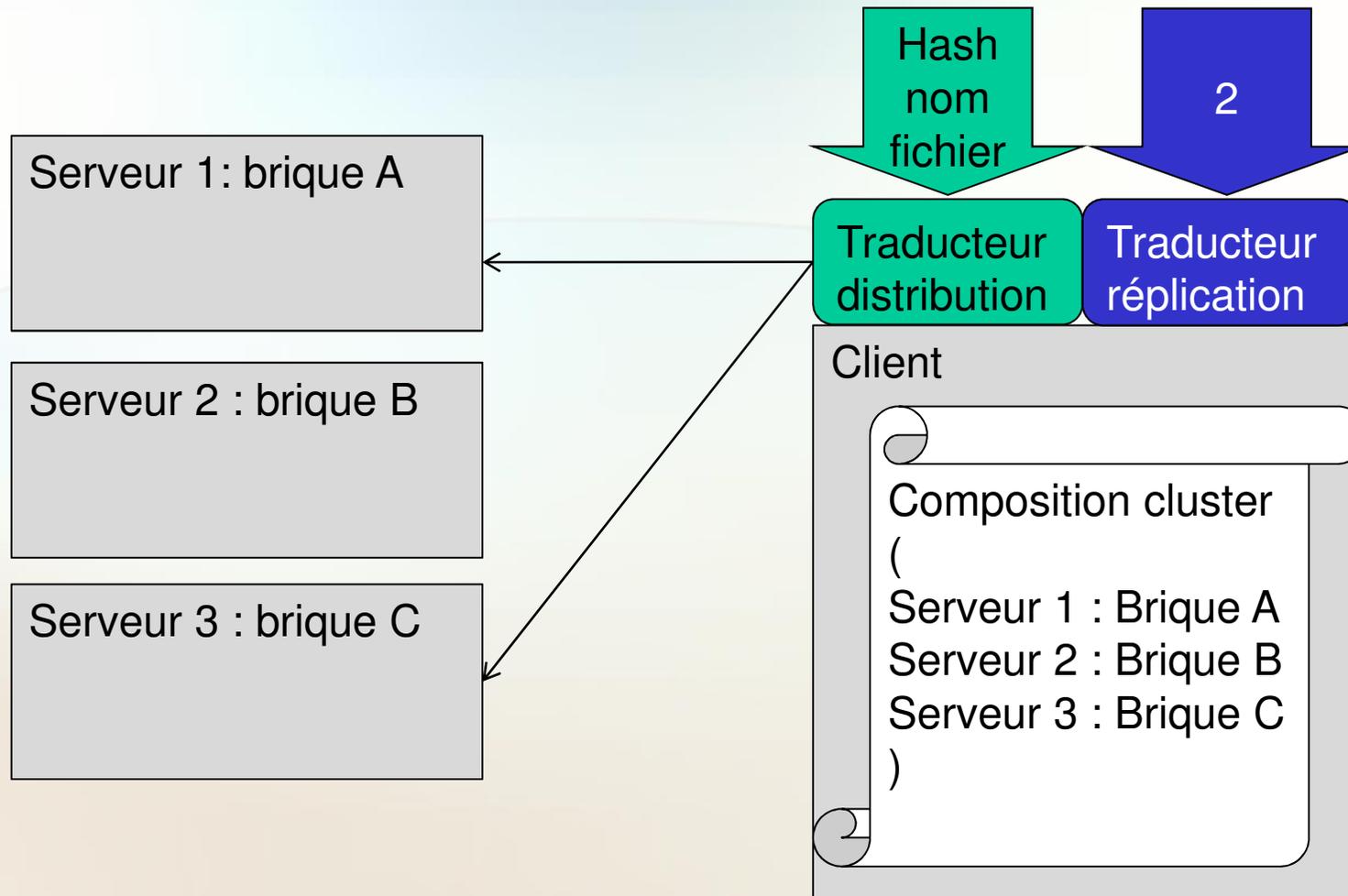
Vue d'ensemble



Ce qui change avec GlusterFS

- Pas de serveur(s) de métadonnées
 - Emplacement fichier calculé
- Sur filesystem de la brique où est stocké le fichier
 - Block données fichiers
 - Date de création, dernier accès,...
 - Propriétaire
 - Quel droit d'accès et à qui
 - Taille

Client Natif



Exemple

- Sur serveur 1 :
glusterd start
mkdir /gluster/brique1
- Sur serveur 2 :
gluster peer probe serveur1
gluster volume create replica 2 vol_deux_brique
serveur1:/gluster/brique1
serveur2:/gluster/brique2
gluster volume start vol_deux_brique
- Sur client n :
mount -t glusterfs serveur1:/vol_deux_brique /mnt/vdb

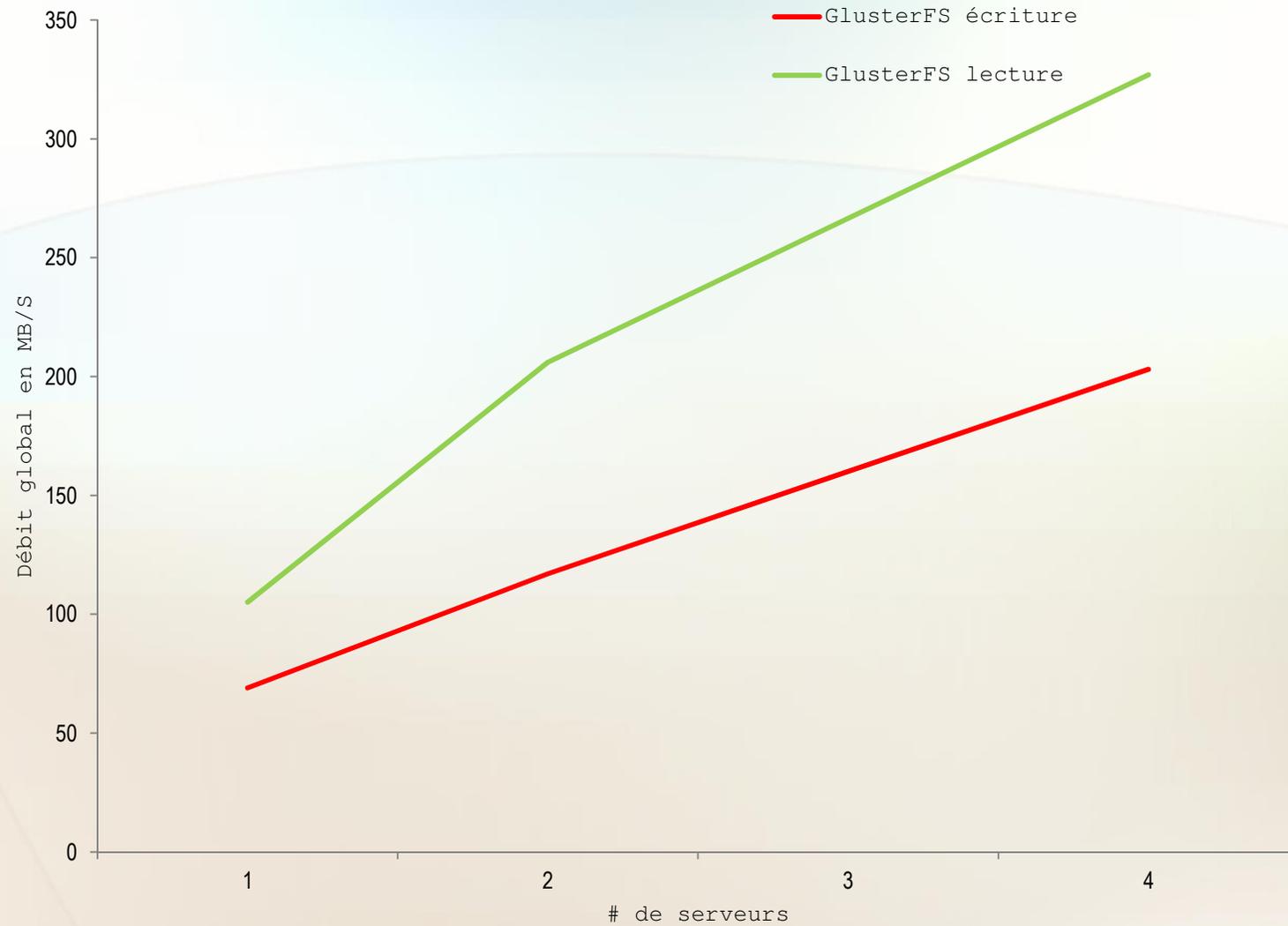
Méthodologie de teste

- Benchmark basique [time,dd]
- Benchmark disque local [IOzone]
- Débit pratique du réseau Gb [iperf]
- Taille fichier et débit utile sur le réseau [wireshark]
- Analyse du protocole GlusterFS [wireshark]
- Analyse débit différent scénario [IOzone]

Problèmes rencontrés

- Le daemon GlusterFS ne démarre pas
- Différence non négligeable des benchmarking
- Décodage Wireshark

Résultats obtenu



Wireshark sans module GlusterFS

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	10.1.40.101	10.1.40.103	TFTP	290	Acknowledgement, Blo
2	0.000231	10.1.40.103	10.1.40.101	TFTP	410	Acknowledgement, Blo
3	0.016103	10.1.40.101	10.1.40.103	TFTP	1234	Acknowledgement, Blo
4	0.016305	10.1.40.103	10.1.40.101	TFTP	302	Acknowledgement, Blo
5	0.016351	10.1.40.101	10.1.40.103	TFTP	222	Acknowledgement, Blo
6	0.016506	10.1.40.103	10.1.40.101	TFTP	102	Acknowledgement, Blo
7	0.016543	10.1.40.101	10.1.40.103	TFTP	222	Acknowledgement, Blo
8	0.016689	10.1.40.103	10.1.40.101	TFTP	102	Acknowledgement, Blo

Frame 1: 290 bytes on wire (2320 bits), 290 bytes captured (2320 bits)
Ethernet II, Src: AsustekC_91:f6:a5 (90:e6:ba:91:f6:a5), Dst: AsustekC_25:2f:34 (e0:cb:4e:25:2f:34)
Internet Protocol Version 4, Src: 10.1.40.101 (10.1.40.101), Dst: 10.1.40.103 (10.1.40.103)
Transmission Control Protocol, Src Port: 1020 (1020), Dst Port: 24009 (24009), Seq: 1, Ack: 1, Len: 224
Session Traversal Utilities for NAT, TURN ChannelData Message
Channel Number: 0x8000
Message Length: 220
Trivial File Transfer Protocol
Opcode: Acknowledgement (4)
Block: 57606

Wireshark avec le module GlusterFS

Protocol	Length	Info
GlusterFS	290	V310 CREATE Call (Reply In 2)
GlusterFS	410	V310 CREATE Reply (Call In 1)
GlusterFS	1234	V310 WRITE Call (Reply In 4)
GlusterFS	302	V310 WRITE Reply (Call In 3)
GlusterFS	222	V310 FLUSH Call (Reply In 6)
GlusterFS	102	V310 FLUSH Reply (Call In 5)
GlusterFS	222	V310 RELEASE Call (Reply In 8)
GlusterFS	102	V310 RELEASE Reply (Call In 7)

Merci pour votre écoute