



Technical Trend of Object-Based Storage Systems

- (Y.K. Kim)
- (H.Y. Kim)
- (Y.C. Kim)
- (K.S. Jin)
- (M.Y. Lee)
- (M.J. Kim)
- (C.K. Kim)

(storage) 가 가

I. 가 .  
 e- , e-commerce, .  
 가 가 , IT ,  
 가 가 가 가  
 가 가 가 가  
 , Fibre Channel  
 가 [1]-[3].  
 , IT trade-off가 , 가  
 IT ,  
 DAS(Direct Attached Storage)  
 가 ,

Ethernet, Fibre Channel  
 Storage Area Network(SAN) 가  
 Network Attached Storage(NAS) DAS DAS 가

## 2. SAN

SAN DAS가

## II.

switched fabric  
 가가 SAN DAS 가  
 가 (

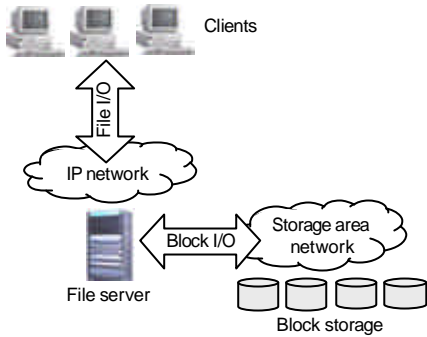
가 , SAN zoning fabric  
 trade-off가

DAS, SAN, NAS, SAN File System [2].

## 3. NAS

### 1. DAS

DAS (block - based) ( )  
 ATA/IDE SCSI 1) NAS



( 1) NAS

가 . ,  
가 가

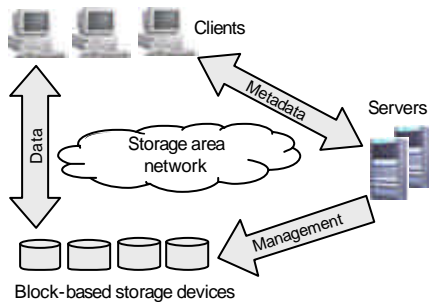
#### 4. SAN File System

SAN File System NAS가

( 2)

가 SAN

SAN



( 2) SAN File System

### III.

trade-off가

가 . ,  
가

가 .

(Object-based Storage Device: OSD)

#### 1. OSD

OSD Carnegie Mellon Parallel  
Data Lab. NASD(Network Attached  
Secure Disk) [4] Active Disk  
[5] , HP, Intel,  
Segate, IBM

OSD

NASD

SNIA(Storage Networking Industry  
Association) OSD technical work group

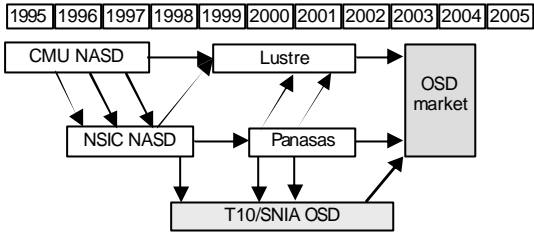
SCSI 가 SCSI OSD

[6]

( 3) NASD OSD

가

, Panasas Lustre



( 3) OSD

[7],[8].

2.

(container)

3가

가

QoS

가 ( , inode)

OSD

3.

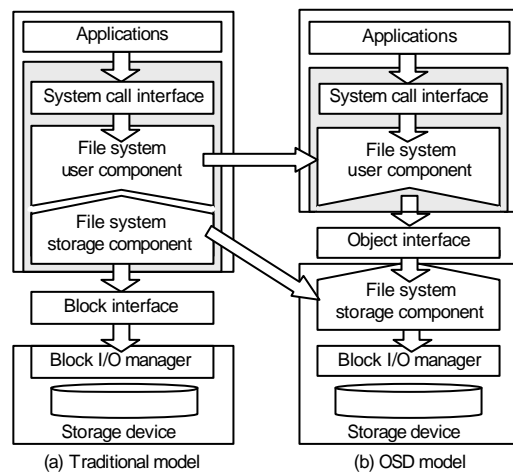
OSD

OSD

(offload)

, ( 4(a))

( )  
 ( )  
 ( )  
 ( )  
 4(b)) 가  
 OSD



( 4) OSD

OSD (self-management) 가

4.

( 5)

OSD

IV.

NASD, Lustre, Panasas가 /

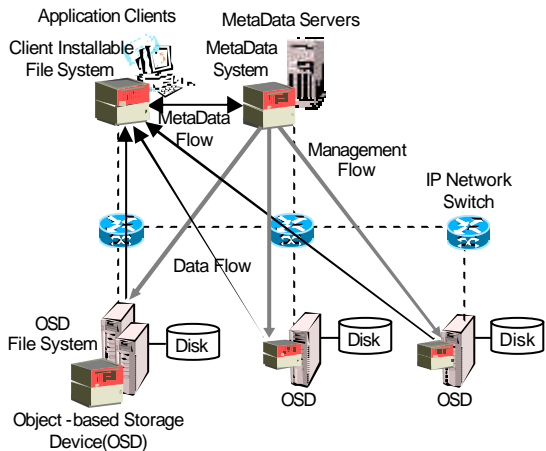
1. NASD

NASD[5] Carnegie Mellon 1998

NASD 가

OSD

가 가



( 5)

NASD 가

(striping) 가

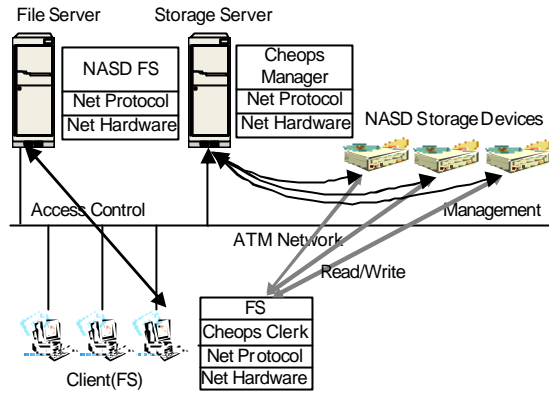
가

가. NASD

NASD SCSI

가

- read/write
- namespace



( 6 ) NASD

ABACUS FS

. Cheops Clerk

NASD

가

가

가

NASD

DVD

(Derived Virtual Device)

namespace export

Cheops Manager가

NASD

NASD  
NASD

migration  
가

가

. NASD

NASD ( 6)

4가

. NASD

NASD

Cheops Clerk

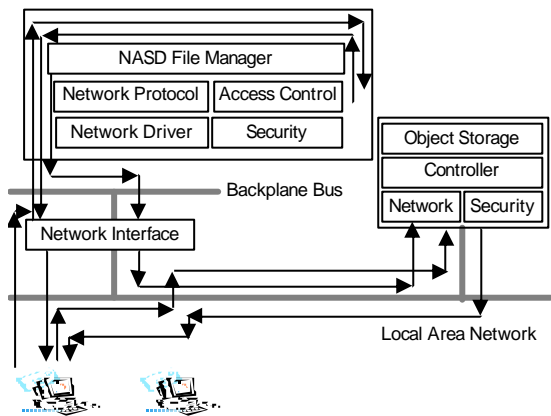
AFS NFS

가

, 가

가 NASD

• capability ( ), capability / NASD / capability read , capability NASD capability ( ), 가 , capability ( ). • NASD ( , ). , 가 capability ( ) , prefetching, 가 가 capability가 , . NASD 가



( 7) NASD • Cheops NASD read/write , NASD 가 NASD 가 NASD Cheops NASD RAID controller , striping disk array . Cheops 가 - NASD capability NASD capability

- 
- NASD

- StorageBlade
- StorageBlade
- RAID

2. Panasas

StorageBlade가 가

Panasas[7]

- DirectorBlade
- DirectorBlade

가

ActiveScale Storage Cluster

DirectorBlade StorageBlade

가. ActiveScale

- ActiveScale
- ActiveScale DirectorBlade
- StorageBlade

-

-

SCSI/OSD

가

, StorageBlade, DirectorBlade  
Gigabit Ethernet  
DirectorBlade RPC  
StorageBlade

. ActiveScale

iSCSI/OSD DirectFlow

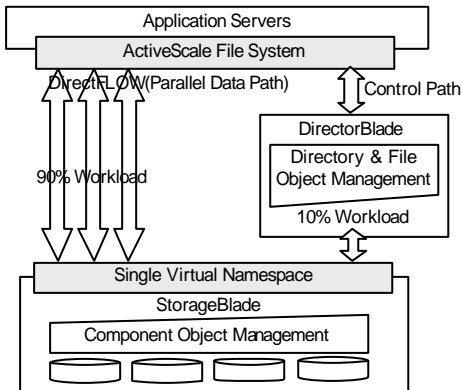
ActiveScale ( 8)

3가

, ActiveScale , StorageBlade  
, DirectorBlade

. ActiveScale

ActiveScale



( 8) ActiveScale

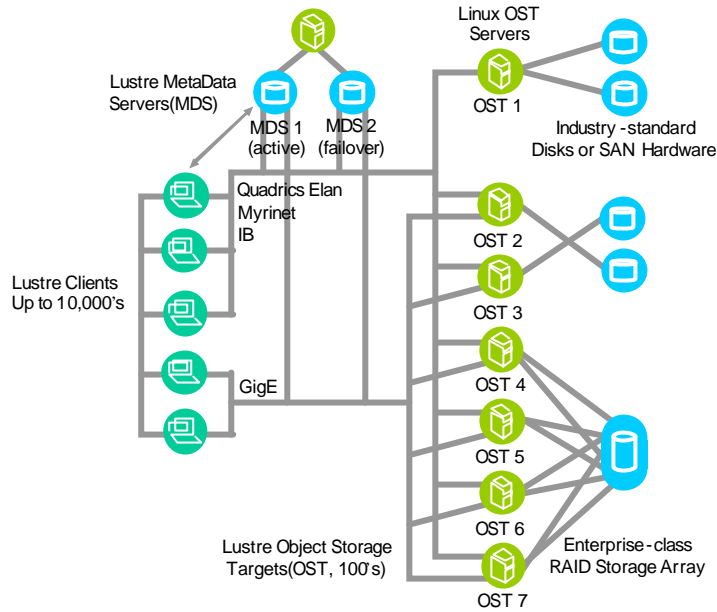
- ActiveScale
- StorageBlade DirectorBlade

- ZeroCopy
- StorageBlade TOE



---

• Active RAID RAID  
 RAID 0, 1, 5  
 RAID Active RAID  
 StorageBlade  
 가 Active Spares Lustre  
 • NFS/CIFS OSD Linux  
 ActiveScale NFS CIFS 가 , ANSI T10 SCSI/  
 Blade NFS/CIFS Director- 가 Lustre I/O  
 CIFS 가  
 • - Kerberos 5, PKI, capability-based I/O, encryption  
 • - Linux 가 ,  
 , GPL  
 • Lustre  
 Lustre ( 9) Lustre  
 (MDS)  
 (OST)  
 • MDS  
 PanActive Manager PanActive MDS  
 Manager Linux  
 ( )  
 3. Lustre  
 Lustre[8] Cluster File System • OST  
 OST



( 9) Lustre

Linux	DAS, SAN				
		OST	RAID	• NAL	
				Lustre	가
• Lustre					Potal API
Lustre			Linux VFS	Quadrics Elan, Myrinet GM, Scali, SDP, Infini-Band	API TCP,
			, MDS		
OST	Posix			Network Abstraction Layer	
API				(NAL)	API
• Lustre				• OBD	OST
				Lustre	Object Based Device(OBD) driver
					native OBD driver Linux
	Lustre				, ext3, xfs
				OST	ext3obd, xfsobd
•					OST
MDS				SCSI/OSD	Lustre
				• LOVM	
					OST OST
	MDS			가	logical object volume manager

RAID  
RAID 0

- 가

Lustre MDS OST MDS

failed-OST bypass

- MDS

Lustre MDS MDS

가

MDS

- Lustre

AFS

- 

GSS-API  
Kerberos 5 PKI , Posix ACL  
, SFS counter mode  
, capability-based OST

- Intent

Lustre MDS

intent

MDS 가

MDS 가

< 1> OSD

	NASD	ActiveScale	Lustre
platform	Digital Unix 3.2g	Linux IA32 Linux IA64	Linux 2.4/2.6
File System	Abacus FS NFS AFS	ActiveScale File System	Lustre File System
name space	Global NS	Global NS	Global NS
cache coherence	Timestamp ordering	Callback	Opportunistic Lock
RAID level	RAID 0/1/5	RAID 0/1/5 Active RAID Active Spares	RAID 0
storage access interface	native object interface (similar to SCSI OSD)	SCSL/OSD	native object interface
network	155Mbps ATM	Gigabit Ethernet	Ethernet Quadrics Myrinet Elan3 or 4 Scali
MDS protocol	DCE RPC over UDP/IP	DirectFLOW SCSI RPC	Portal with NAL
availability	N/A	clustering	clustering
access control	Posix ACL (capability)	Posix ACL (capability)	Posix ACL (capability)
encryption	key-based HMAC-SH1	IPSec	Counter mode encryption

가 /  
< 1>

V.

4.

가 ,  
가 ,  
WAN  
가

[1] G.A. Gibson and R.V. Meter, "Network Attached Storage Architecture," Communications of the ACM, Nov. 2000, pp.37-45.  
[2] M. Mesnier, G.R. Ganger, and E. Riedel, "Object - Based Storage," IEEE Communications Magazine, Aug. 2003, pp.84-90.  
[3] Intel White Paper, "Object -based Storage," <http://www.intel.com/labs/storage/>.  
[4] <http://www.pdl.cmu.edu/NASD>  
[5] <http://www.pdl.cmu.edu/Active>  
[6] <http://www.snia.org>  
[7] <http://www.panasas.com>  
[8] <http://www.lustre.org>