

OFED Storage Protocols



OPEN**FABRICS**
A L L I A N C E

R. Pearson
System Fabric Works, Inc.

Agenda

- Why OFED Storage
- Introduction to OFED Storage Protocols
- OFED Storage Protocol Update

Why OFED Storage

Goals of I/O Consolidation

Cluster of Servers



Multiple Fabrics
High CapEx and High TCO



Communications

Computing

Management

Storage

Single InfiniBand Fabric
Low CapEx and Optimal TCO



Communications

Computing

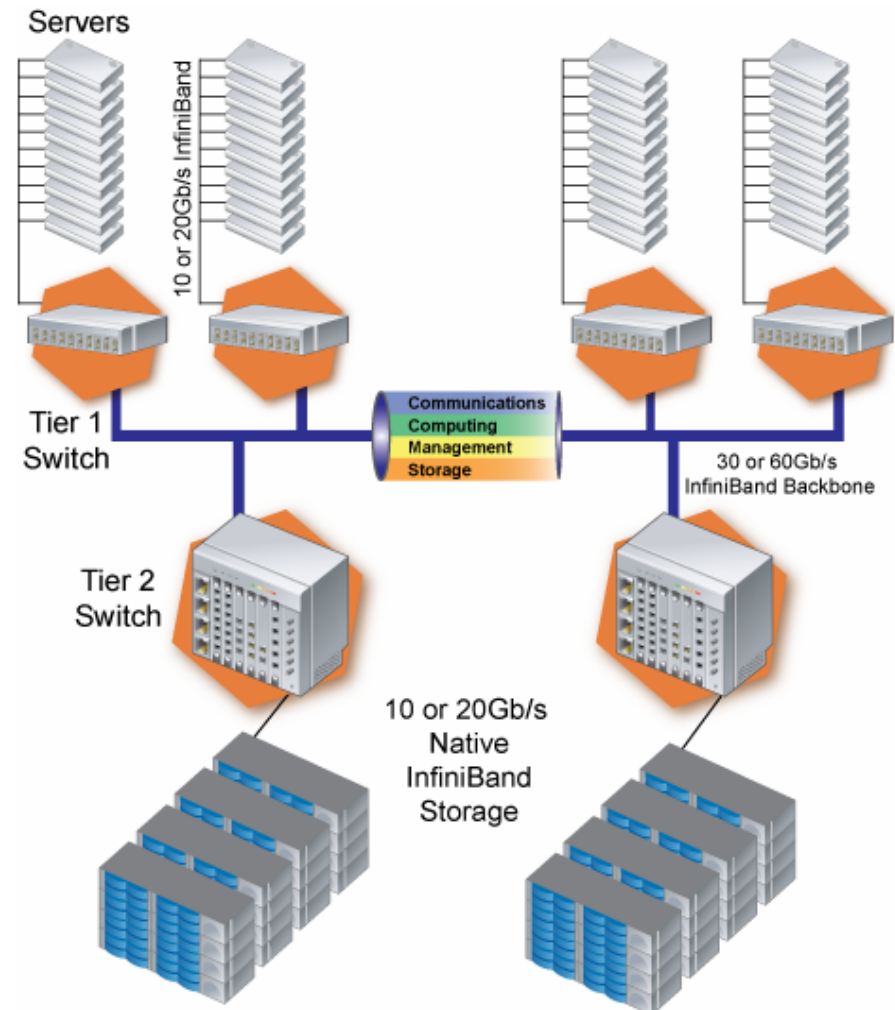
Management

Storage

“One Wire”

Unified InfiniBand Fabric

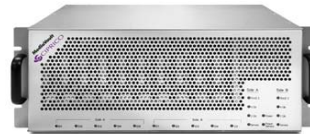
- Performance
- Ultimate scalability
- Cost of ownership
- Dedicated I/O channels enable consolidation



Industry Support for Native InfiniBand Storage



HP StorageWorks
Scalable File Share (HP SFS)

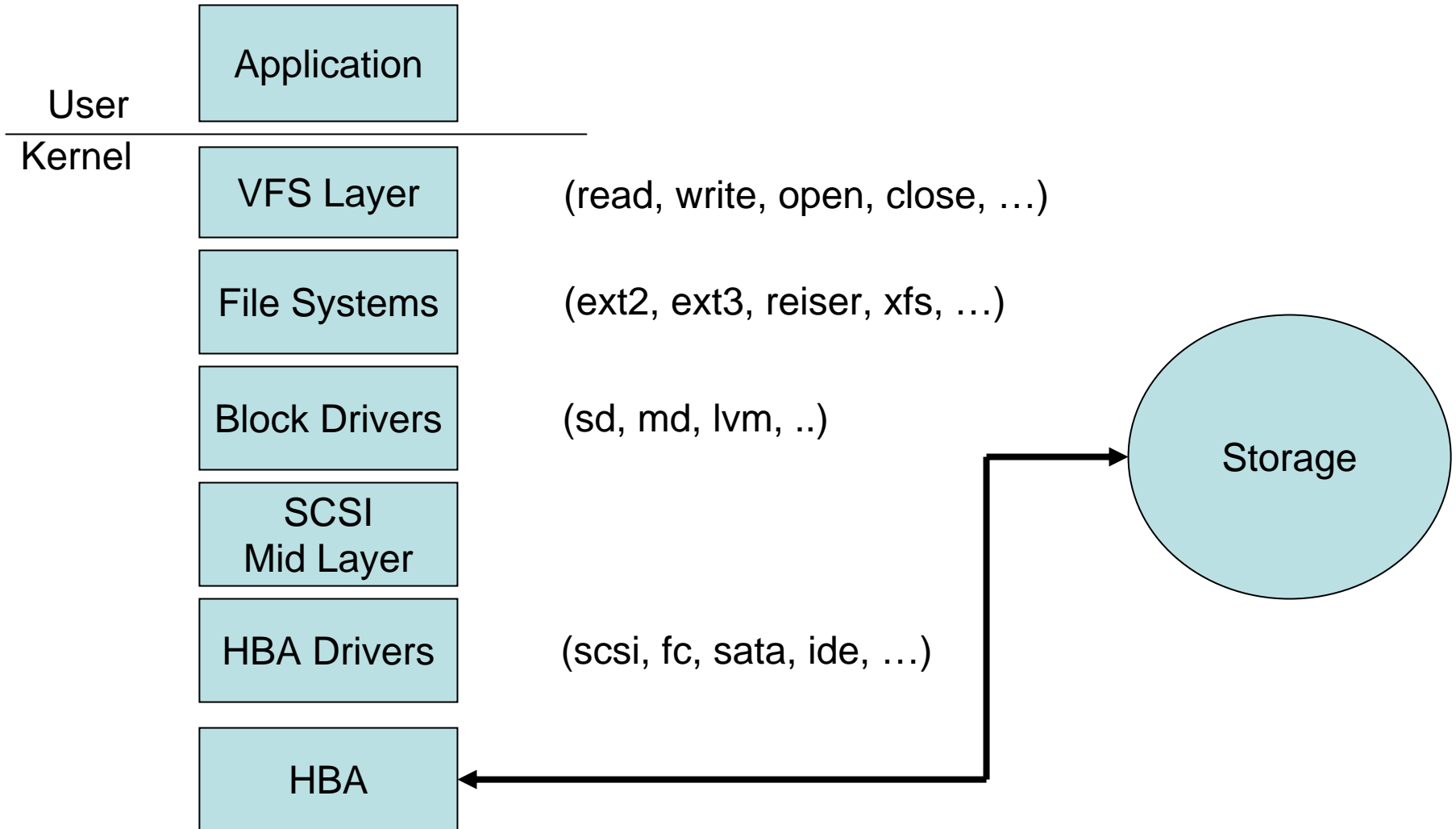


Storage Interconnects

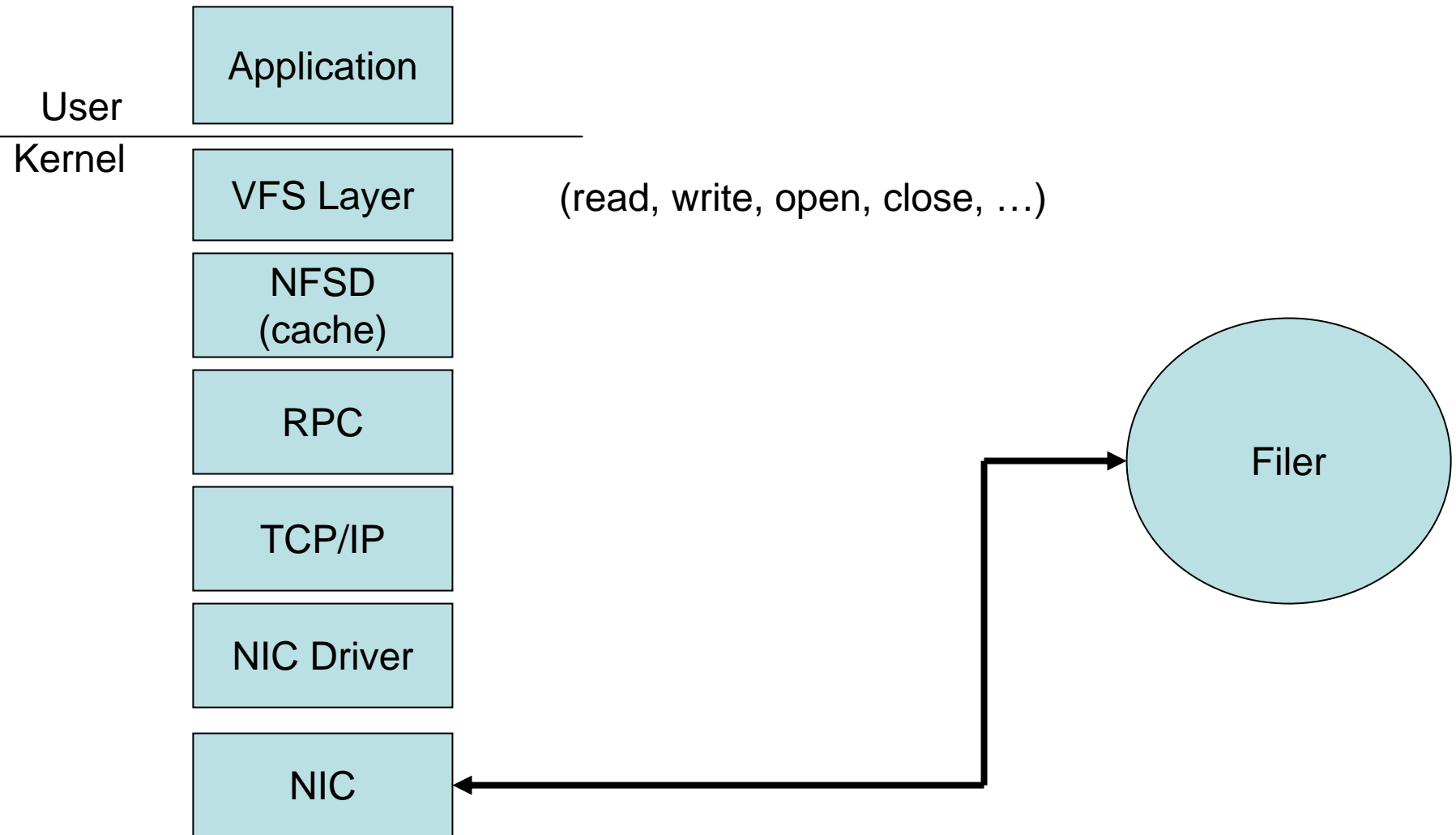
| Type | Max length | Peak data rate | Max devices |
|---------------------|------------|----------------|-------------|
| PATA (IDE) | 46cm | 133MB/s | 2 |
| SATA/eSATA | 1m/2m | 300MB/s | 1 |
| SCSI | 12m | 320MB/s | 16 |
| SAS | 6m | 300MB/s | 16K |
| FC | 50km | 400MB/s | 16M |
| iSCSI (Ethernet) | ? | ? | ? |
| IB Cu (DDR) | 15m/5m | 1gbps/2gbps | 16M(48K?) |
| IB optical (mm/WAN) | ~300m/? | 1GB/s | 16M |

Introduction to OFED Storage Protocols

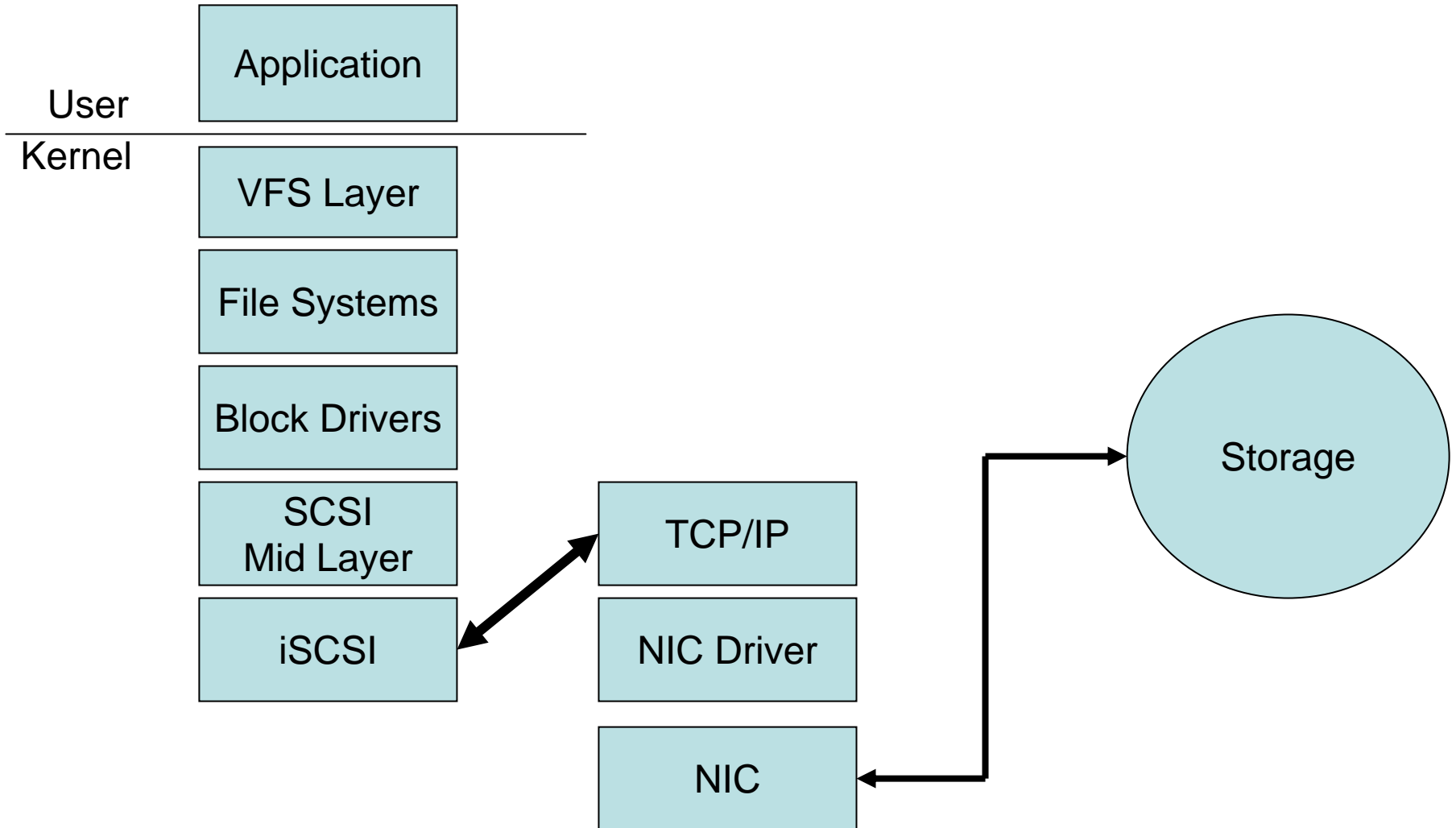
Linux Storage Stack (das/san)



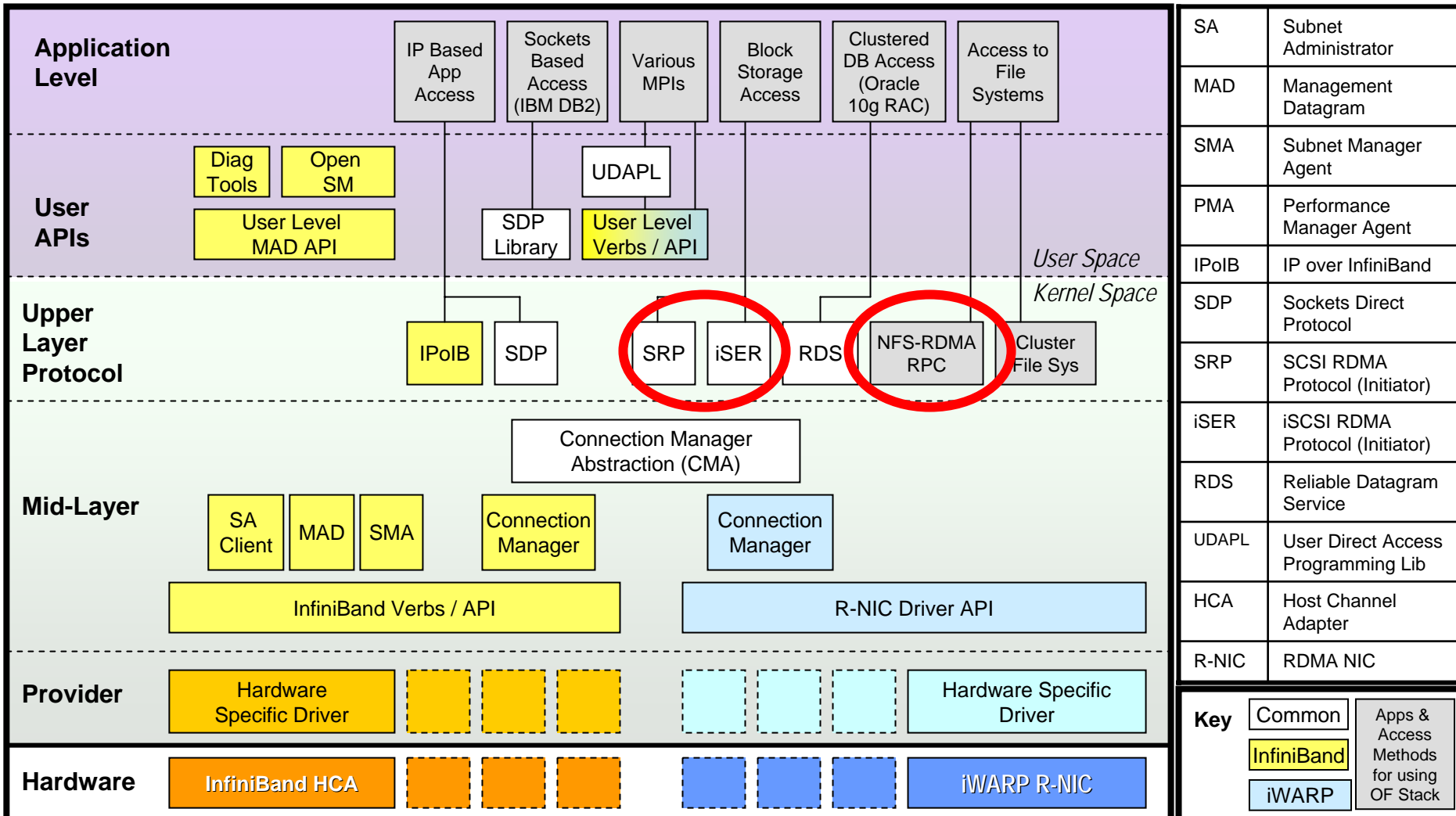
Linux Storage Stack (nas)



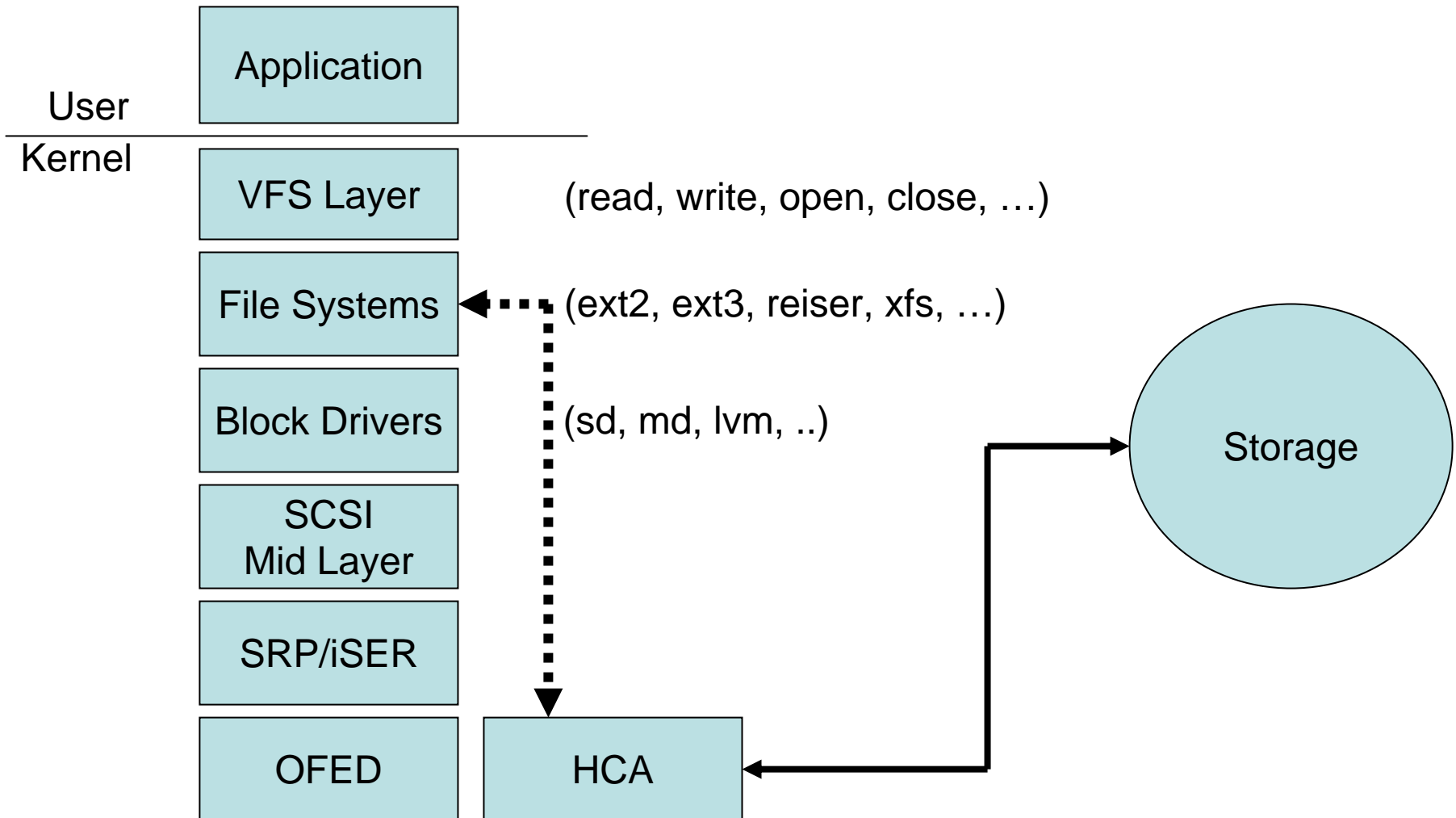
Linux Storage Stack (iSCSI)



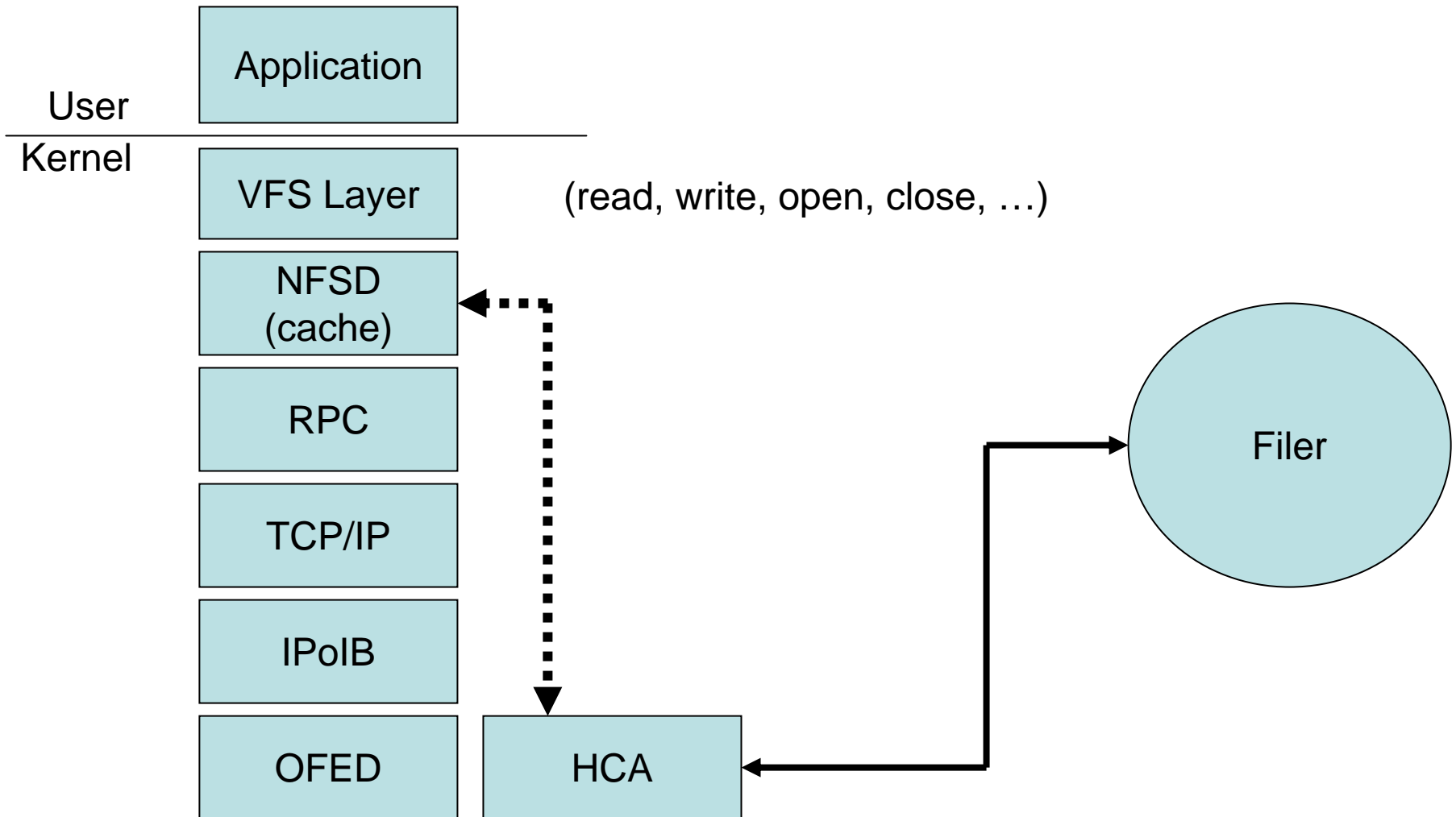
OpenFabrics Software Stack



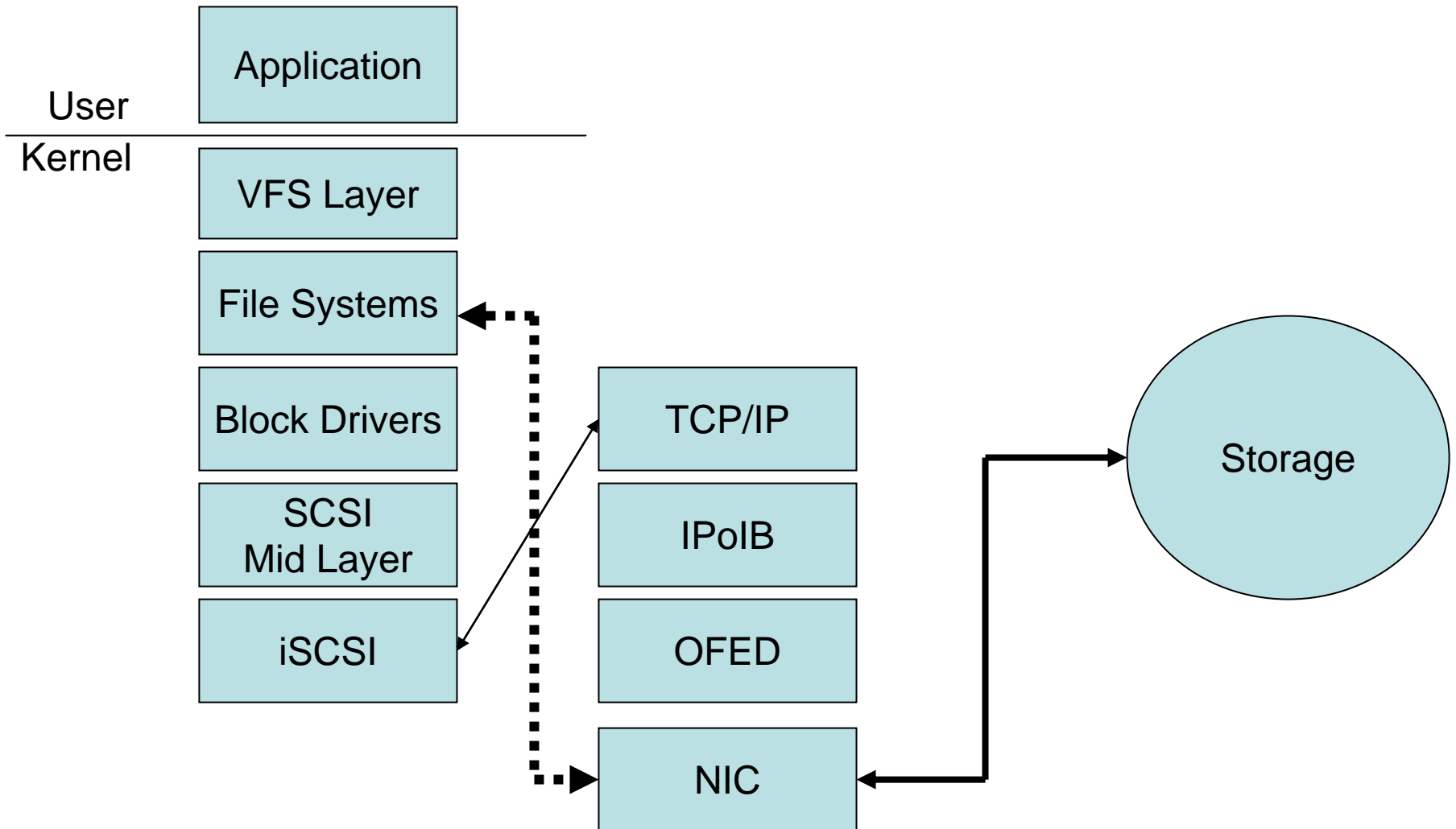
Linux Storage Stack (das/san)



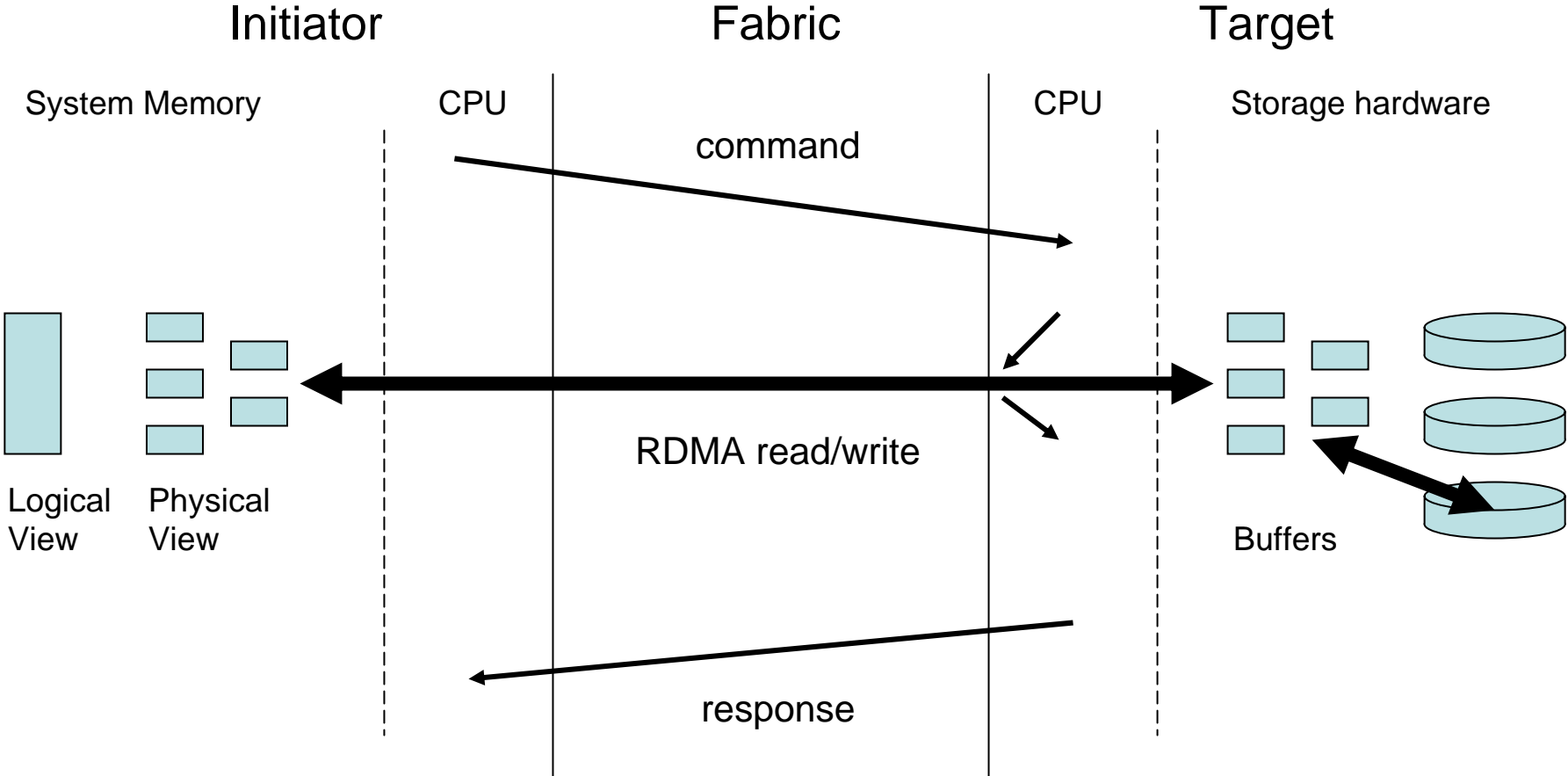
Linux Storage Stack (nas)



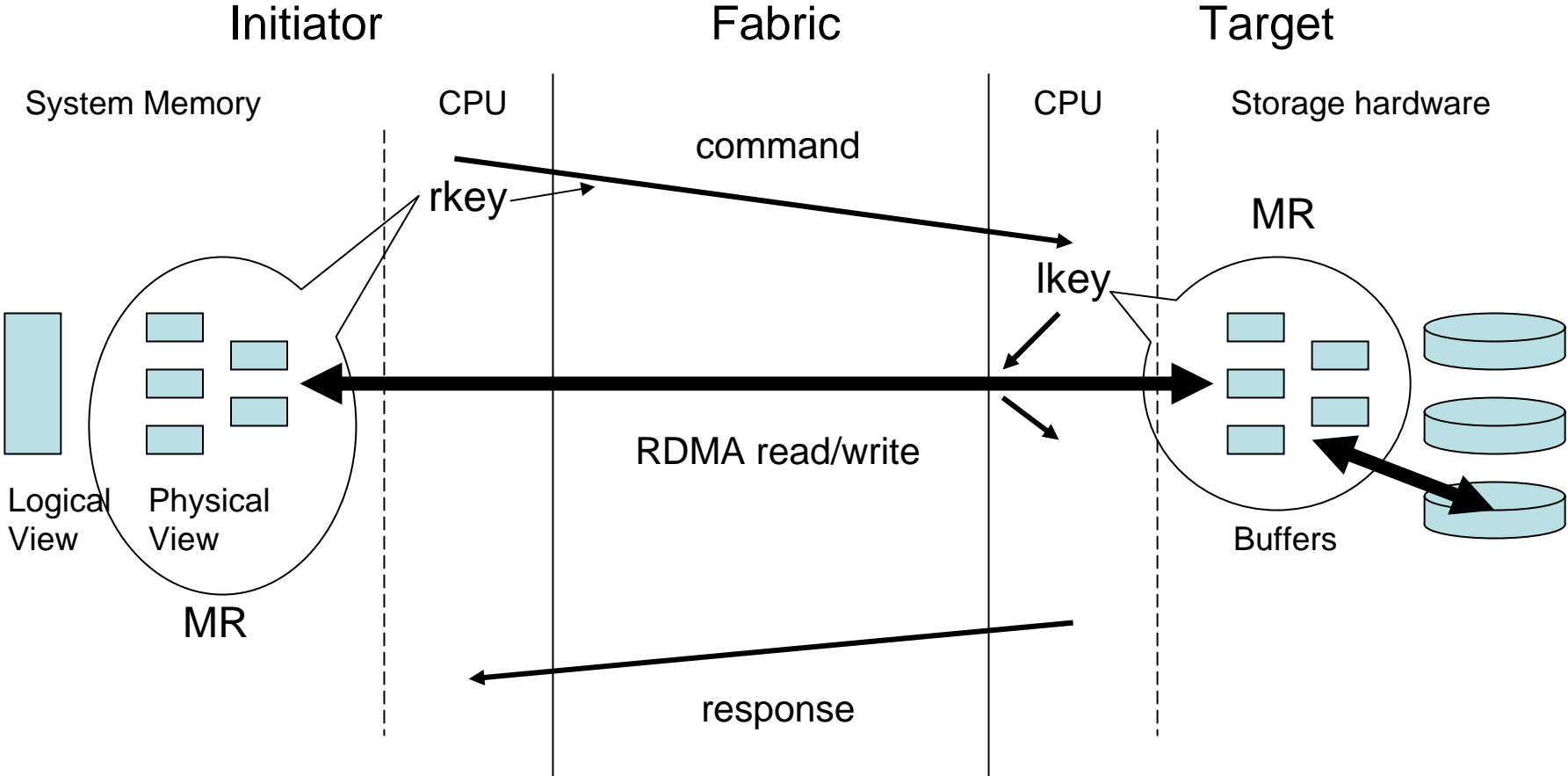
Linux Storage Stack (iSCSI)



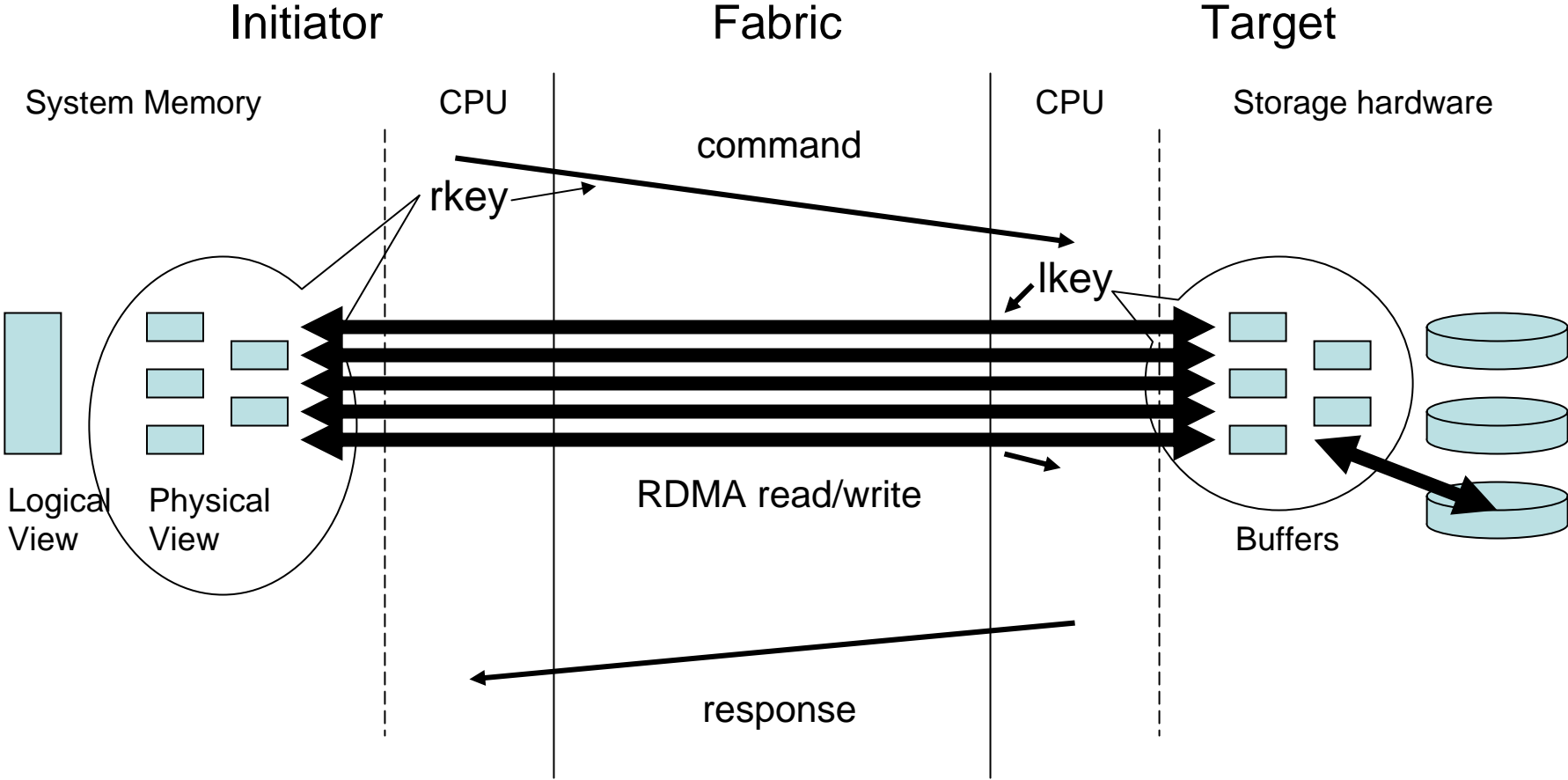
Common I/O Scheme



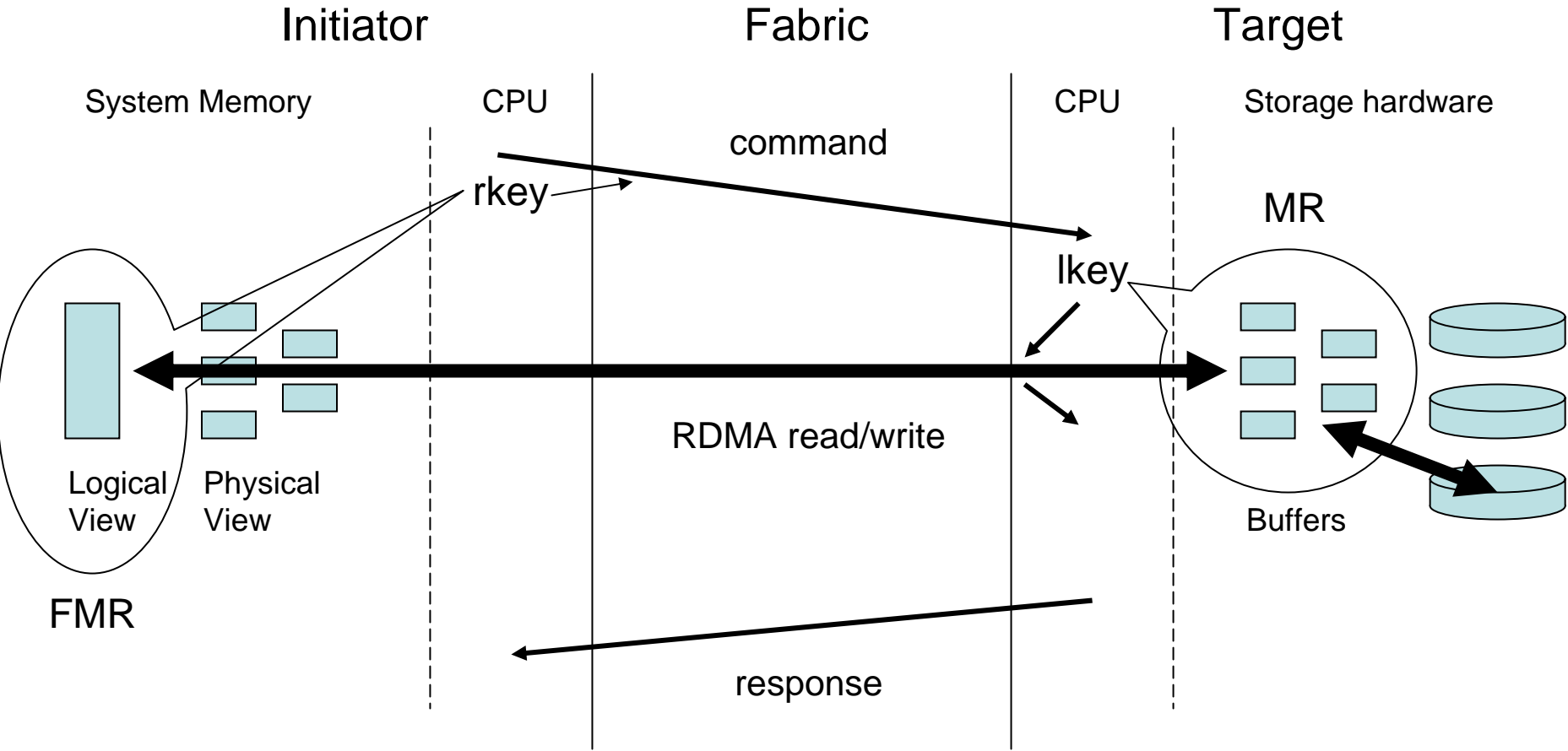
Common I/O Scheme



Common I/O Scheme



Common I/O Scheme



OFED Storage Protocol Update

SRP Update

- **SCSI RDMA Protocol (SCSI over IB)**
 - T10.org specification, similar to FCP (SCSI over FC)
 - All data transfers use InfiniBand RDMA
- **Initiator**
 - Linux proprietary and open source “Gen1”-based in production
 - OFED 1.0 SRP Initiator in Beta (still in development)
 - Windows Beta release in July
- **Target**
 - Linux open source “Gen1”-based target on OFA
 - Vendors in production or in final qualification
 - Linux and other RTOSs

SRP Next Steps

➤ Initiator

- Host failover/multipathing
- Boot support for Linux and Windows

➤ Target

- Call for interest to port target to OFED
 - Mixed file system / block storage Linux targets require this capability

iSER Update

- iSCSI Extensions for RDMA
 - IETF IP storage specification, Leverages iSCSI management and discovery protocols
 - iSER eliminates iSCSI/TCP bottlenecks
 - RDMA, hardware transport
- Initiator
 - Production supported in OFED 1.0
 - Tested against Voltaire iSER target running in IB/FC router
- Target
 - Linux open source “Gen1”-based target on OFA
 - Vendors porting to OFED

- Release 5 NFS/RDMA client and server for Linux 2.6.16.16 on SourceForge
 - Supports OFA CM and RDMA verbs API
 - Client maintainer is Tom Talpey, Netapp
 - Server maintainer is Tom Tucker, Open Grid Computing
- Linux NFS RDMA client and server testing
 - Passes the Connectathon test suite, IOzone and network stress tests
 - Tested with several Mellanox-based IB HCAs and standard Linux x86 servers
 - Initial test data shows 800MB/s+ read to cache
- Release 5 ready for evaluation by users, system integrators and OEMs

OFED Storage Summary



- Benefits of consolidated InfiniBand I/O are clear and being realized by real-world deployment today
- Vendors in production or in final qualification stages of block storage solutions
 - SRP and iSER will both continue development in OF
- NFSoRDMA ready for community trials
 - Stay in lock step with OFED
- Vendors porting clustered file systems to OFED

System Fabric Works

System Fabric Works

- Software/architecture/consulting services
- Vendor neutral
- Experienced team
- Customers include:
 - DDN/LSI/Chelsio/QLogic/Cisco/
Silver Storm/Trilabs/Agilent/USDoD/GE/Panta/
Themis...

Embedded Products

| Vendor | Part Number | Description |
|------------------------|--------------|------------------------------------|
| Diversified Technology | ATC5232 | IB ATCA SBC |
| | ATC2148 | ATCA IB Switch |
| | AMS0508 | 5 slot ATCA chassis IB compatible |
| | AMS1412 | 14 slot ATCA chassis IB compatible |
| SBS/GE Fanuc | IB4X-CPCI-2A | 2 port HCA cPCI 3U/6U |
| | IB4X-PMC-2A | 2 port HCA PMC module |
| | TDB | VITA41 IB switch |
| | TDB | VITA41 IB SBC |
| | | VxWorks 5.5 IB stack |
| Sky Computers | Various | IB based systems |