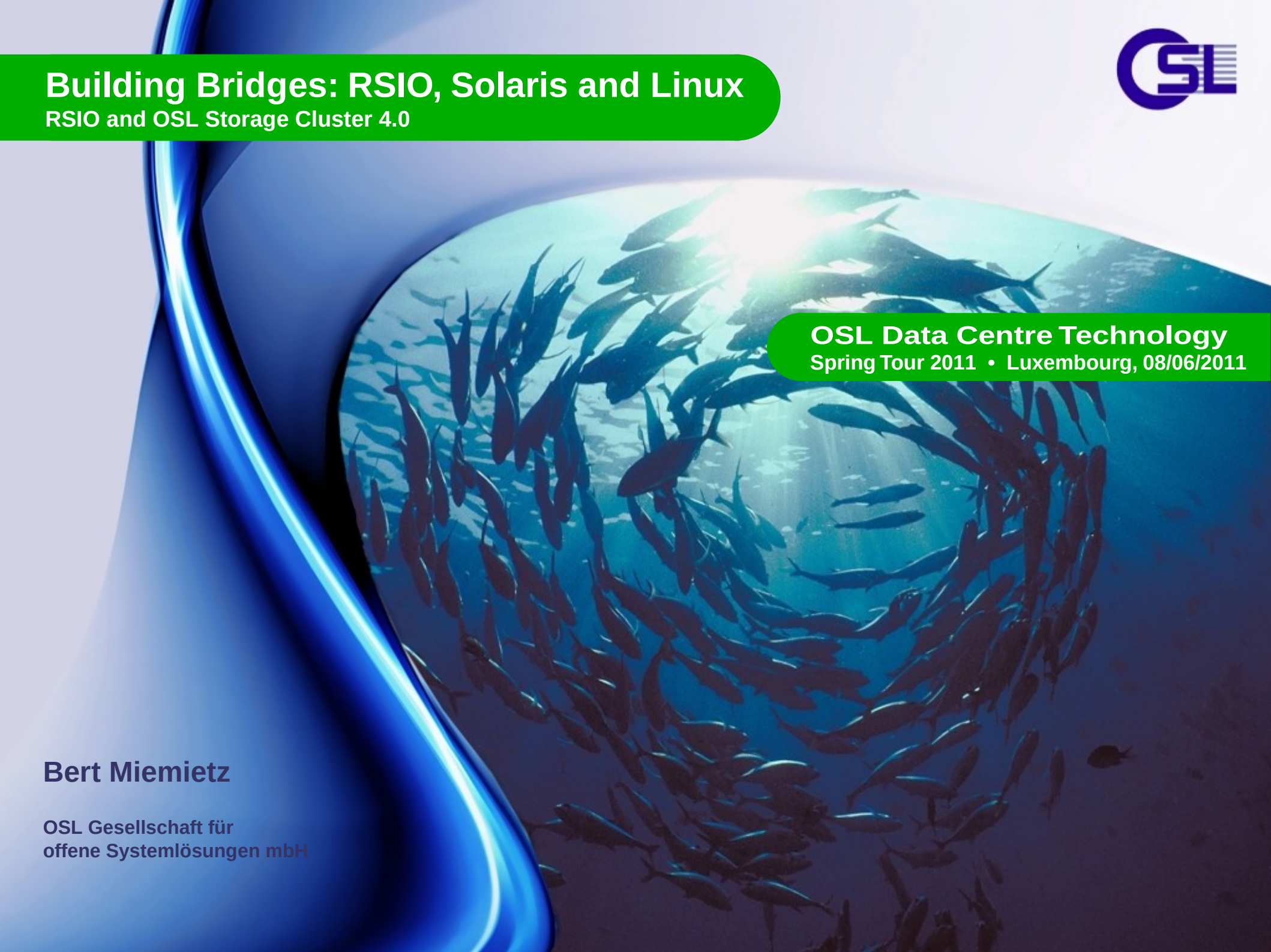
Building Bridges: RSIO, Solaris and Linux

RSIO and OSL Storage Cluster 4.0



OSL Data Centre Technology

Spring Tour 2011 • Luxembourg, 08/06/2011

Bert Miemietz

OSL Gesellschaft für
offene Systemlösungen mbH

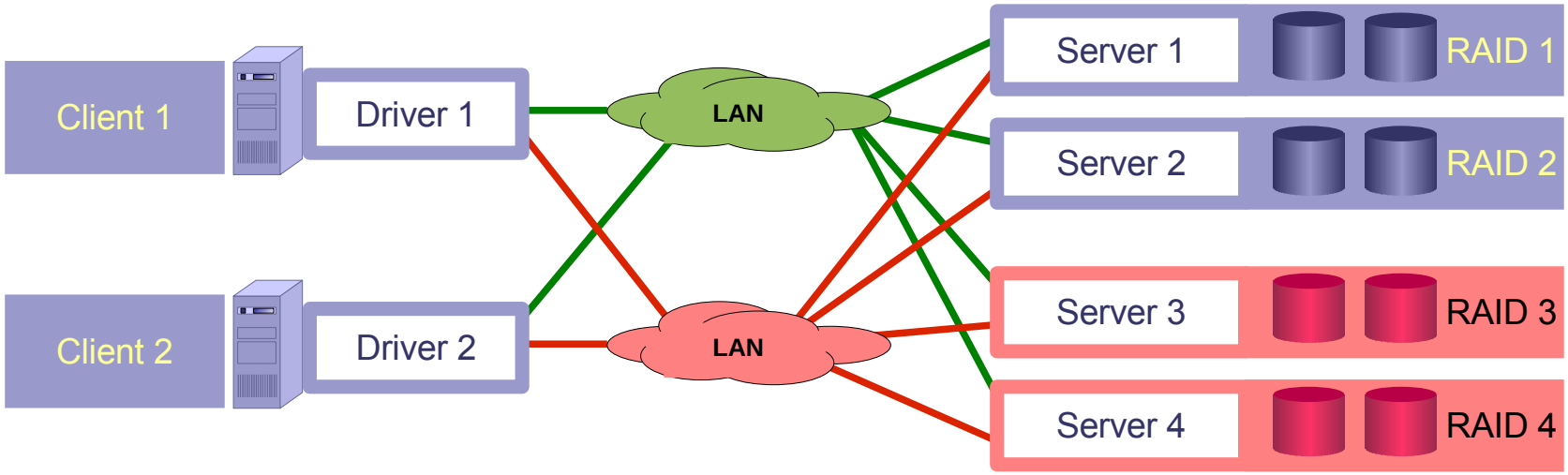
New Platforms and New Infrastructures

Both topics are depent on each other

- *so far only 3 supported operating systems:*
  SunOS, Solaris, OpenSolaris
- *there have been increasingly more request for other platforms:*
  *ubuntu, fedora, openSUSE, **debian**, Mandriva, LinuxMint, PCLinuxOS, **slackware**, gentoo linux, CentOS, **Red Hat**, SLES, SLED*
  ***they all share**: - the name "Linux"*
  - the same kernel development tree
- *reasons to go Linux:*
- *are there disadvantages of the Linux OS?*
- *the question of a proper I/O infrastructure rises:*
  - Fibre Channel -> ?
  - Ethernet -> Why not? But if Ethernet then use block I/O

Block I/O over Ethernet – a Different Approach

Networked infrastructures do require network paradigms



- *send I/O requests read(), write(), ioctl()*
- *suitable encapsulation*
- *connection setup and close down, monitoring*
- *channel multiplexing*

- *process I/O requests read(), write(), ioctl()*
- *suitable encapsulation*
- *connection setup and close down, monitoring*
- *channel multiplexing*

Block I/O over Ethernet – a Different Approach

Networked infrastructures do require network paradigms

Is SCSI up-to-date?

- ***device identification and description can be much easier (address, port, data structures)***
- ***networking issues should be much more reflected***
- ***much of the SCSI data is irrelevant, but many interesting functions are poorly implemented***
- ***without SCSI there is no need for converting data to that low level***
- ***some of the SCSI mechanisms are counter-productive in networks (z. B. Bus-Reset)***
- ***communication overhead can be reduced by far***

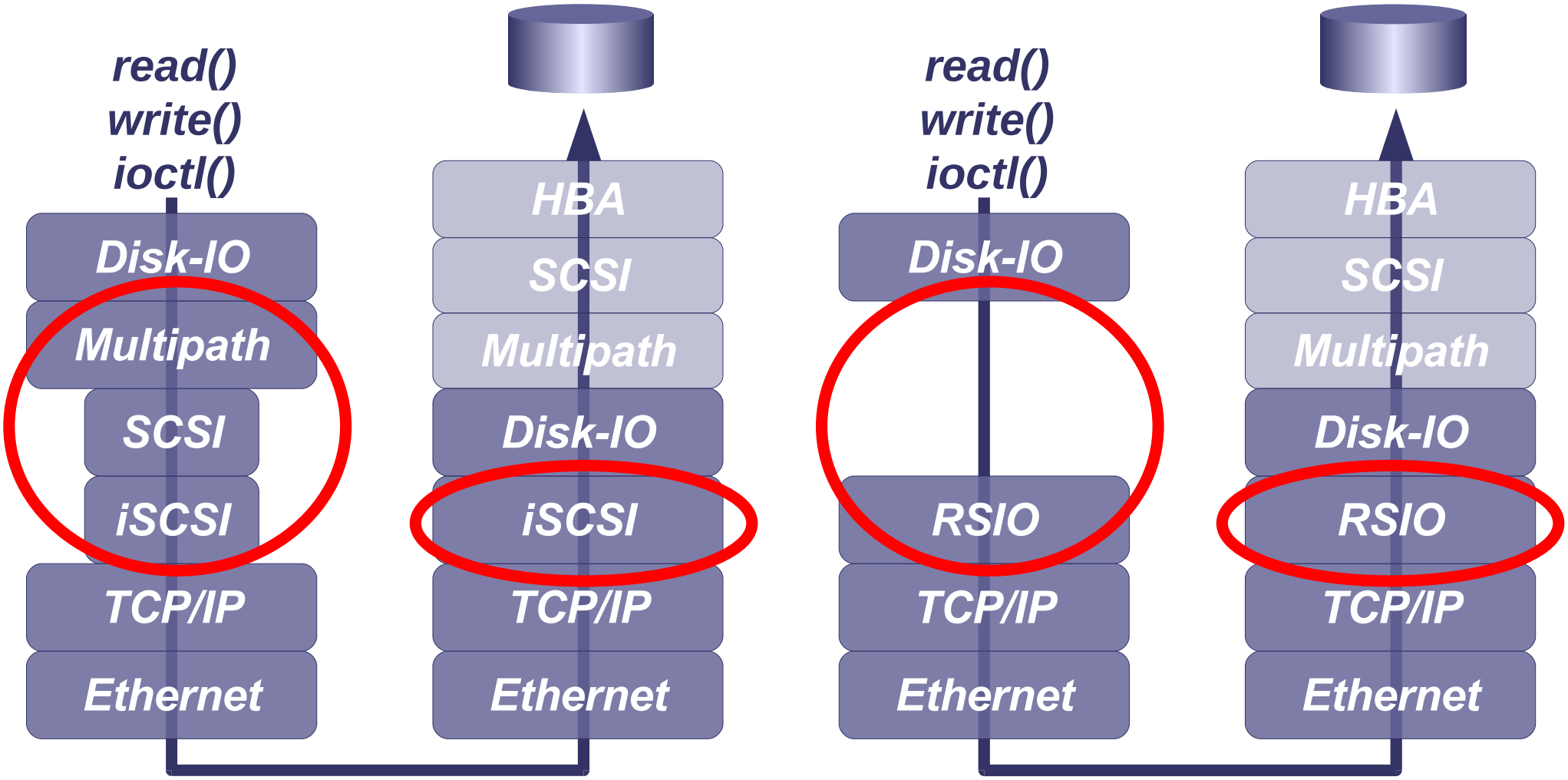
RSIO - Remote Storage I/O

The new technologie for LAN-attached (shared) block devices in brief

- *new protocol developed by OSL*
- *direct transport of all relevant IO requests (read, write, ioctl)*
- *integrated connection setup, monitoring, path multiplexing, trunking*
- *capable of error-recovery and self-configuration*
- *can be used in all modern storage scenarios:*
  - *simple server and clients incl. multipathing*
  - *clustered storage servers (targets)*
  - *clustered storage clients (initiators)*
  - *integrated clusters of servers and clients*
  - *storage server farms*
  - *cloud architectures*
- *special suitability for storage virtualisation*
  - *administrator-friendly names*
  - *fdisk (partitioning) at client side no longer needed*
  - *on-demand allocation and online reconfiguration*
  - *many special add-on features*
  - *can be administered from the client side*

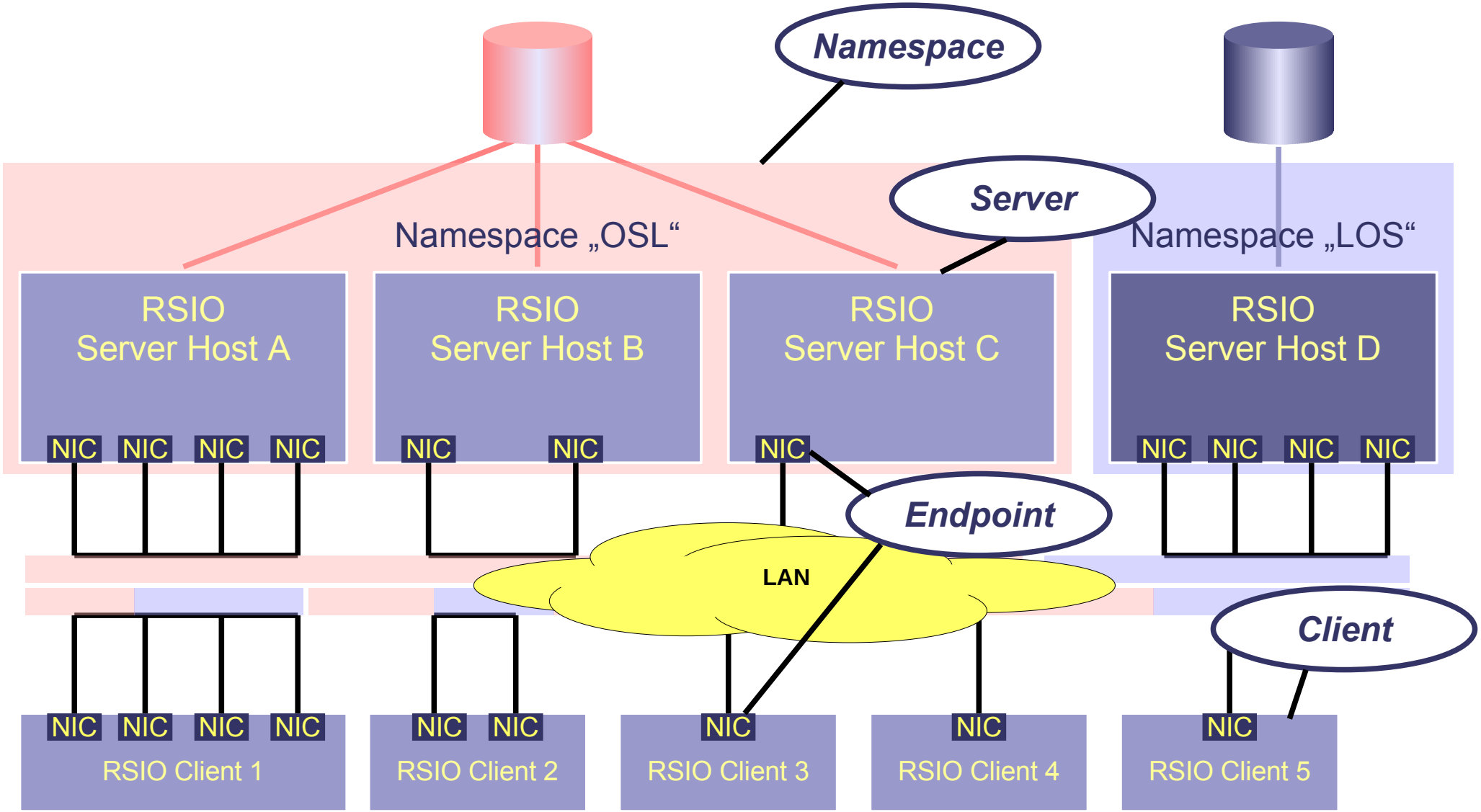
RSIO - Remote Storage I/O

Simplified comparison against the iSCSI protocol stack



RSIO – Architecture Survey

A logical and flexible design



How OSL Has Implemented RSIO

Combination with OSL Storage Cluster

An iSCSI-Lun reports as follows (“format” - Solaris)

```
29. c3t227d0 <DEFAULT cyl 1021 alt 2 hd 64 sec 32>
    /iscsi/disk@0000iqn.1986-03.com.sun%3A02%3A06df3360-bb85-ee33-bf59f2d03474f708.target-00001,0
30. c3t229d0 <DEFAULT cyl 1021 alt 2 hd 64 sec 32>
    /iscsi/disk@0001iqn.1986-03.com.sun%3A02%3A06df3360-bb85-ee33-bf59f2d03474f708.target-00001,0
```

and this is the view of disk resources from an RSIO client

```
# rsconfig -q
000 osl
    clt: big-6
    srv: 000 big-5
         0  tvol1          disk        2097152 blocks of  512 bytes
         0  shadow         disk        2097152 blocks of  512 bytes
         0  ora_db         disk       10485760 blocks of  512 bytes
         0  postgres_db    disk       10485760 blocks of  512 bytes
         0  whole_zone     disk       41943040 blocks of  512 bytes
```

And what about performance?

The protokoll allows high performance and impressive scalability

Server Performance Cache Read / 8k

iSCSI	*10 Clients*	*100 Threads*	*7,6 Cores*	***31.000 IOPS***
iSCSI / comstar	*10 Clients*	*100 Threads*	*10,0 Cores*	***85.000 IOPS***
RSIO	*4 Clients*	*64 Threads*	*5,6 Cores*	***98.000 IOPS***
RSIO	*4 Clients*	*128 Threads*	*6,3 Cores*	***102.000 IOPS***

Client Performance Throughput

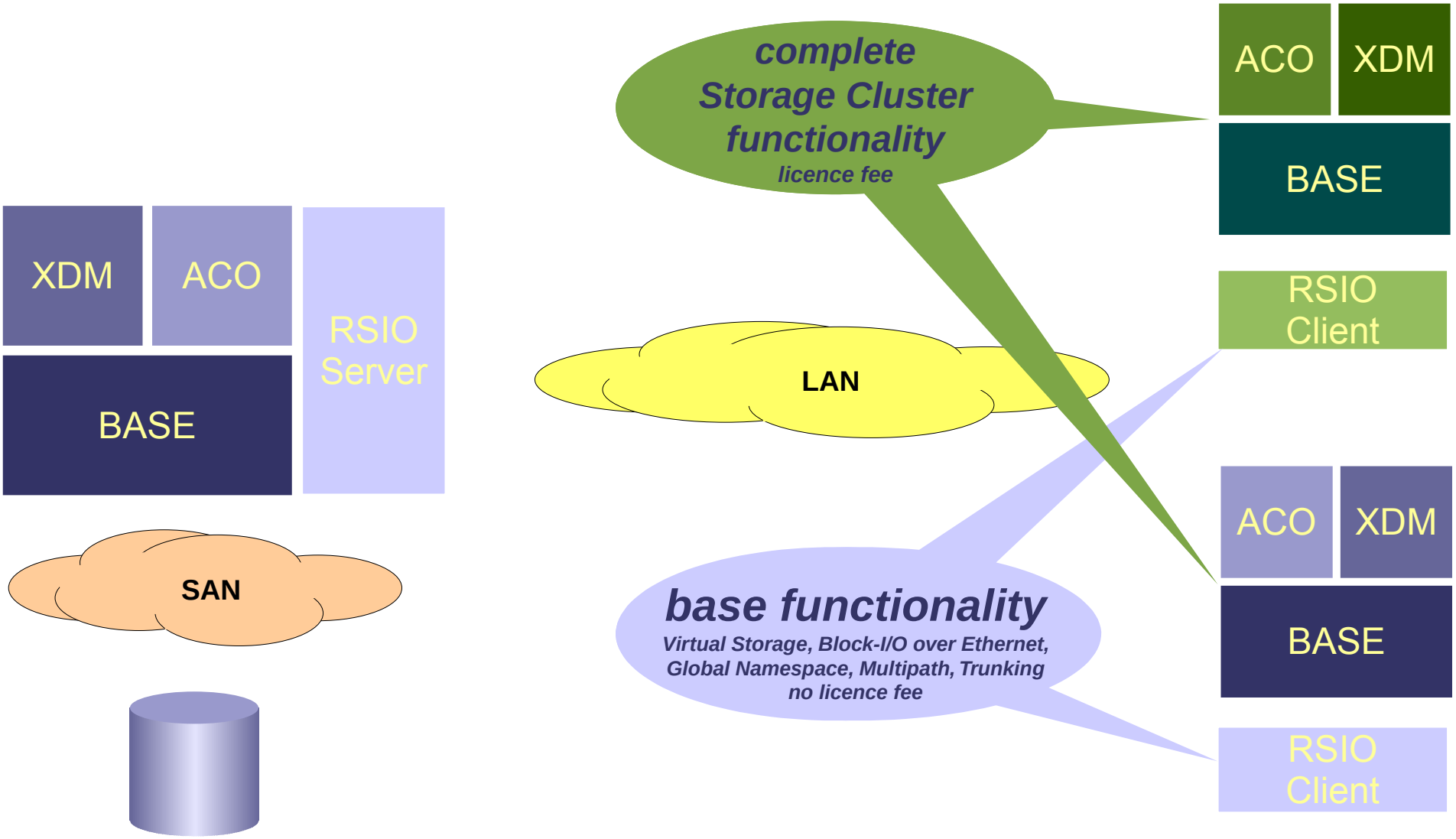
RSIO	*1 x 1 GBit*	*ca. 0,5 Cores*	***> 110 MByte/s***
RSIO	*2 x 1 GBit*	*ca. 1,0 Cores*	***> 220 MByte/s***
RSIO	*4 x 1 GBit*	*ca. 2,0 Cores*	***> 440 MByte/s***
RSIO	*8 x 1 GBit*	*> 4,0 Cores*	***up to > 900 MByte/s***

Use-cases

What Concepts Can Be Implemented Using RSIO?

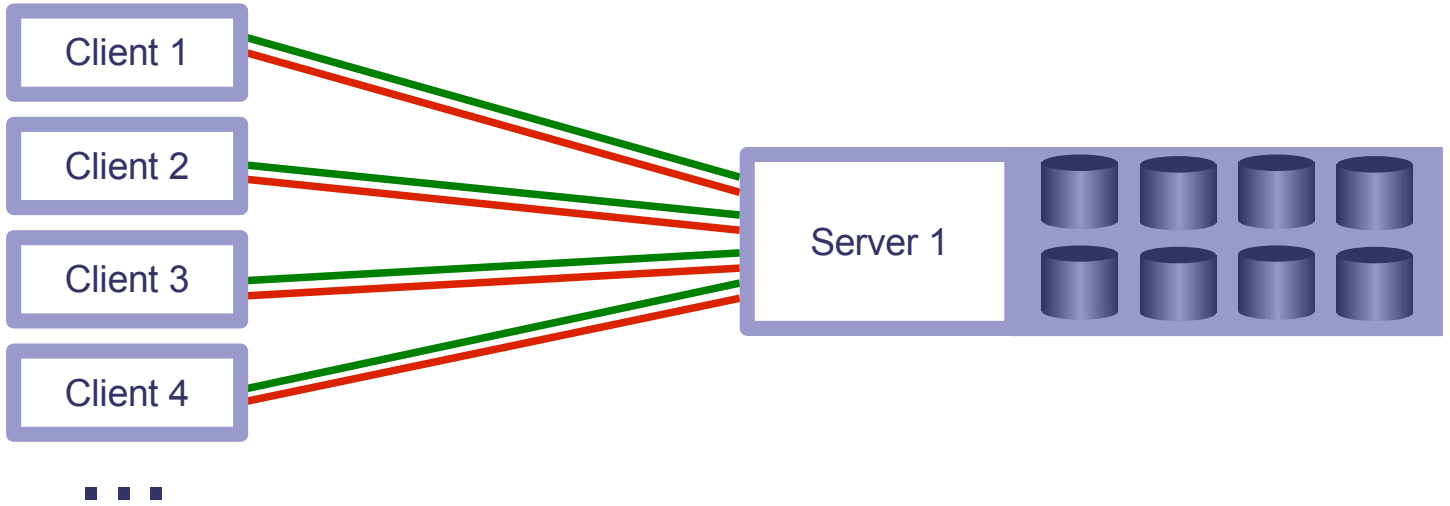
What Concepts Can Be Implemented Using RSIO?

The principle



What Concepts Can Be Implemented Using RSIO?

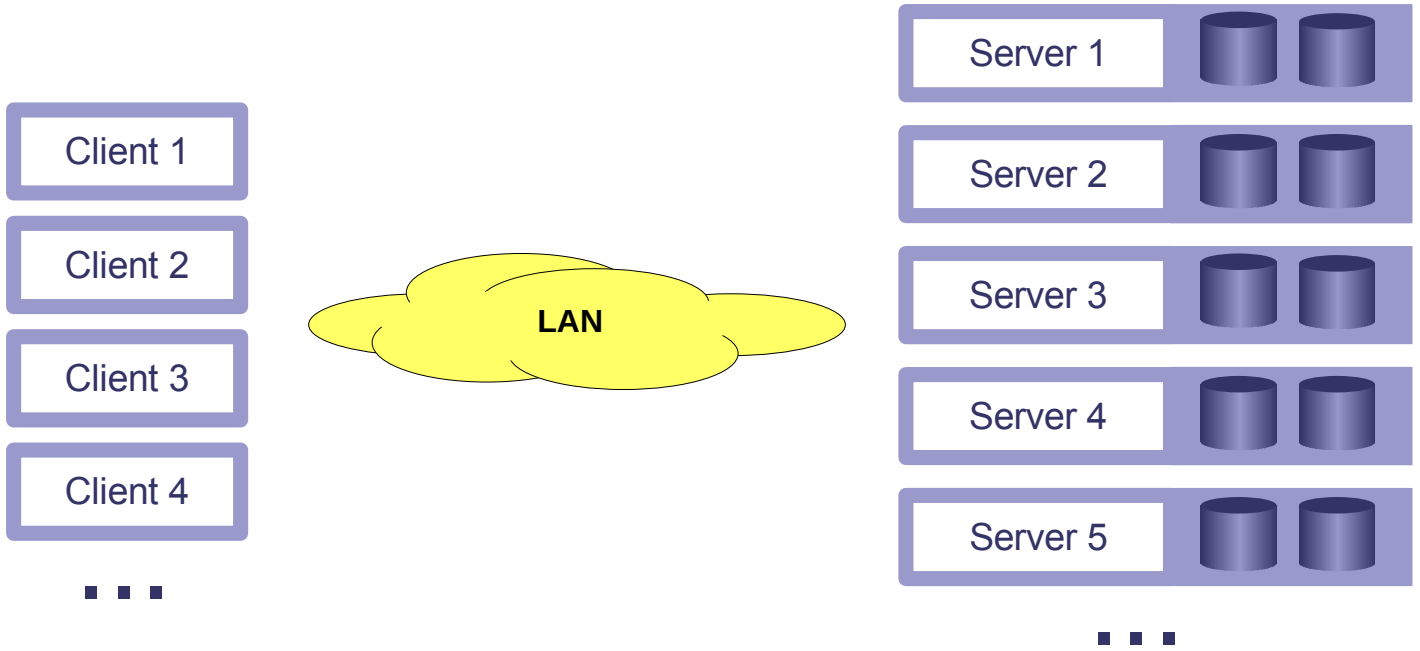
Example 1: Simple access to disk resources over the (IP) network



- *access to central storage system -> global pool, global namespace*
- *virtualisation + clustering (HA) for the clients can be implemented with ease*
- *possibility for centralisation of backups, snapshots ...*
- *low-price storage connectivity combined with good performance*
- *redundant data paths, throughput can be scaled according to real needs*

What Concepts Can Be Implemented Using RSIO?

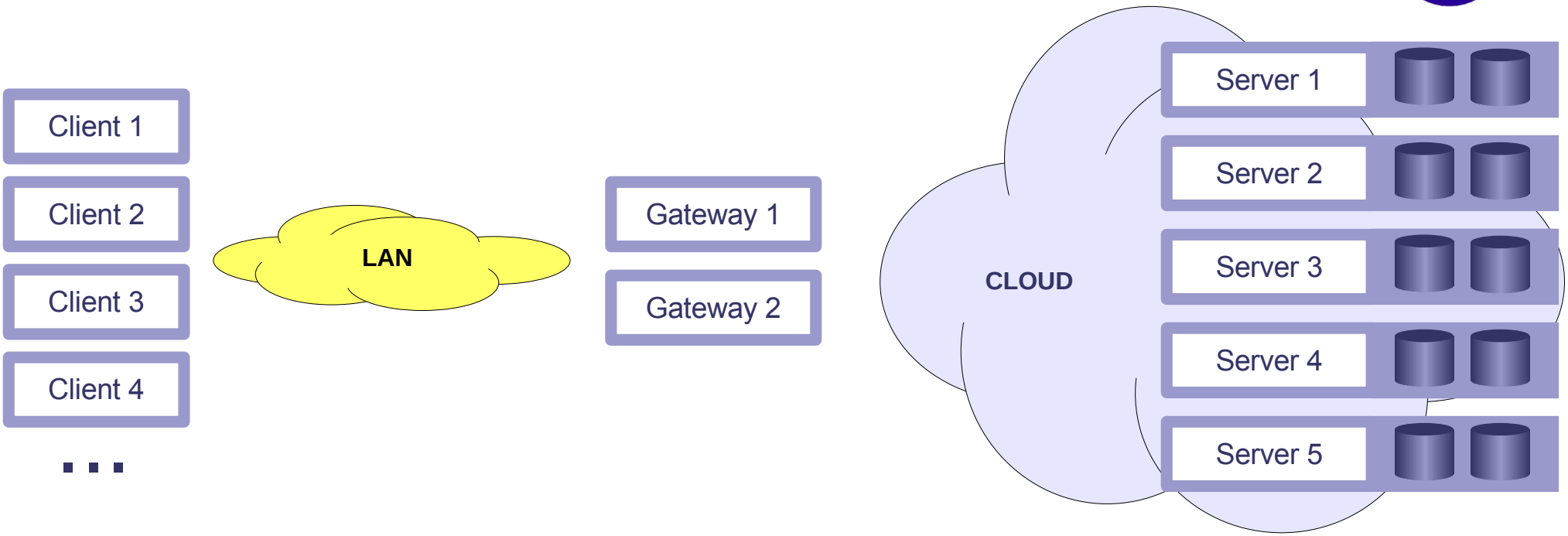
Example 2: Storage server farm



- *scales in terms of capacity as well as performance*
- *each server can run an individual namespace*
- *storage capacities can be “gathered” →*
  huge capacities and bandwidths can be created
- *not to be forgotten: availability of each member of the server farm!*

What Concepts Can Be Implemented Using RSIO?

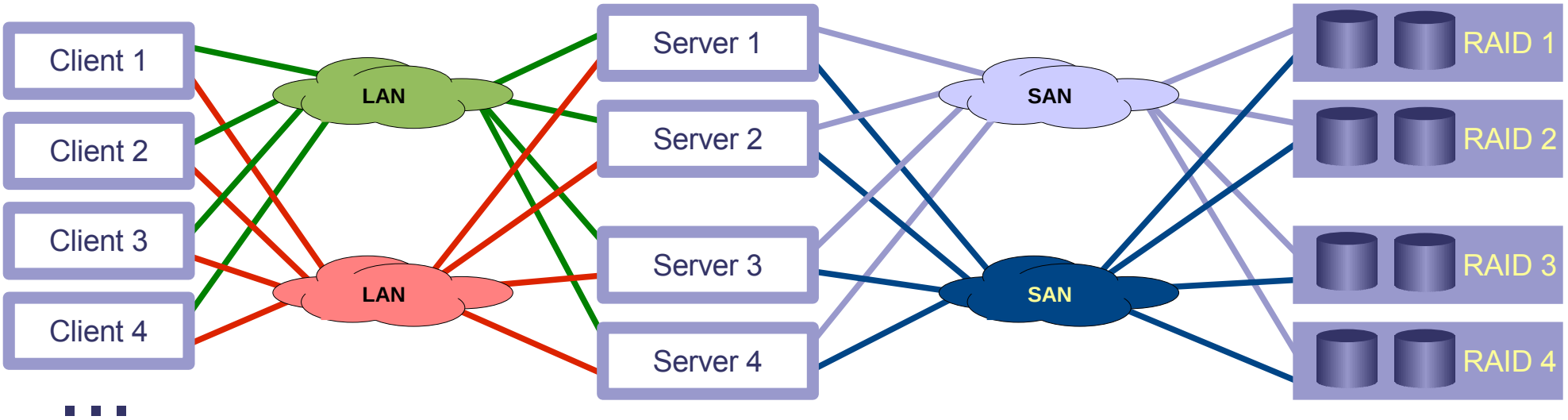
Example 3: Szenario for cloud storage



- *access to storage resources out of the LAN*
- *multipath, bandwidth, performance are of lower importance*
- *administration similar to RSIO in the LAN*
- *facilitated by the IP routing capabilities of RSIO*

What Concepts Can Be Implemented Using RSIO?

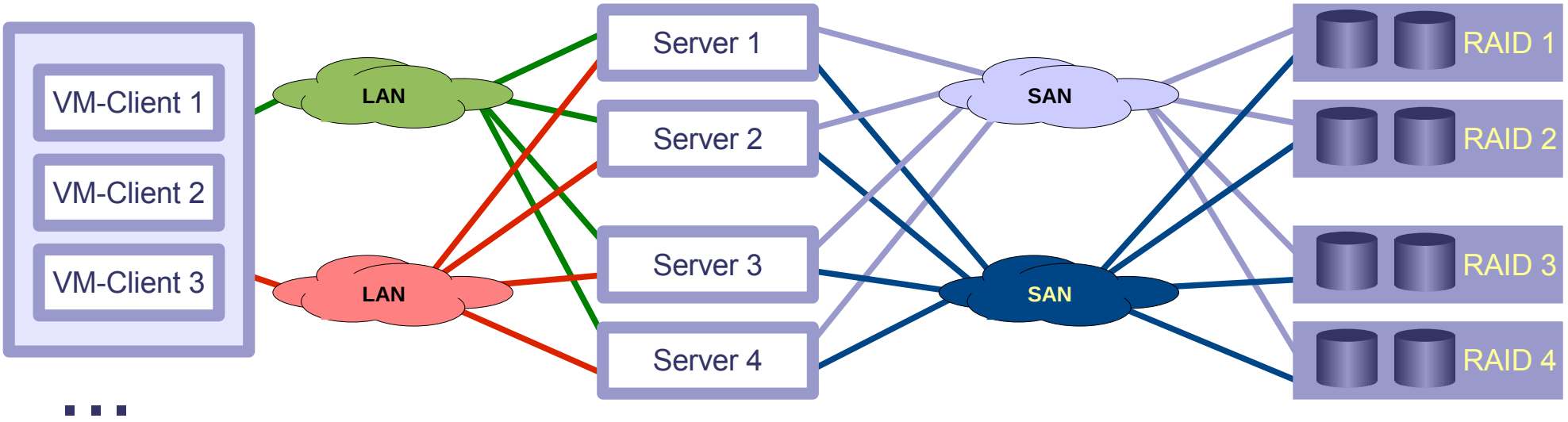
Example 4: SAN-LAN convergence and clustered storage servers



- *SAN is extended into the LAN*
- *SAN-attached servers serve storage resources as a “background task”*
- *improved (=more effective) use of SAN resources, performance rightsizing*
- *high performance, high availability at extremely low costs for RSIO clients*
- *perspective of additional improvements in performance and system capacity exploitation e. g. by using free memory as additional cache*

What Concepts Can Be Implemented Using RSIO?

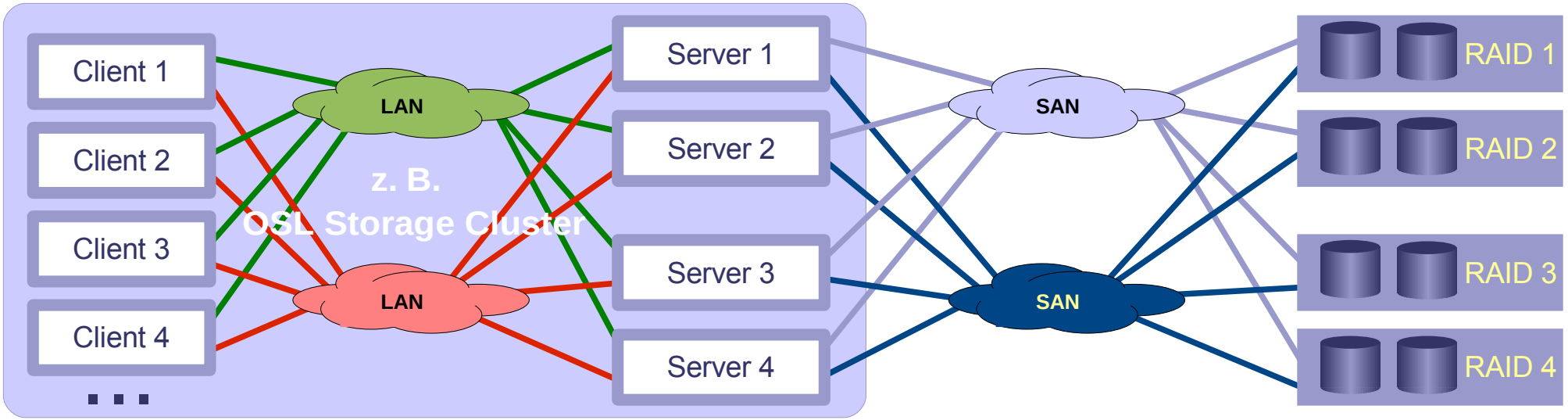
Example 5: Simple access to the entire storage world from a VM



- *SAN is extended into the VM world over the IP interfaces*
- *all devices are accessible, beside this one can use self-configuration ...*
- *aggregation for 10GbE, use of VMDq possible*
- *enormous ease of administration*

What Concepts Can Be Implemented Using RSIO?

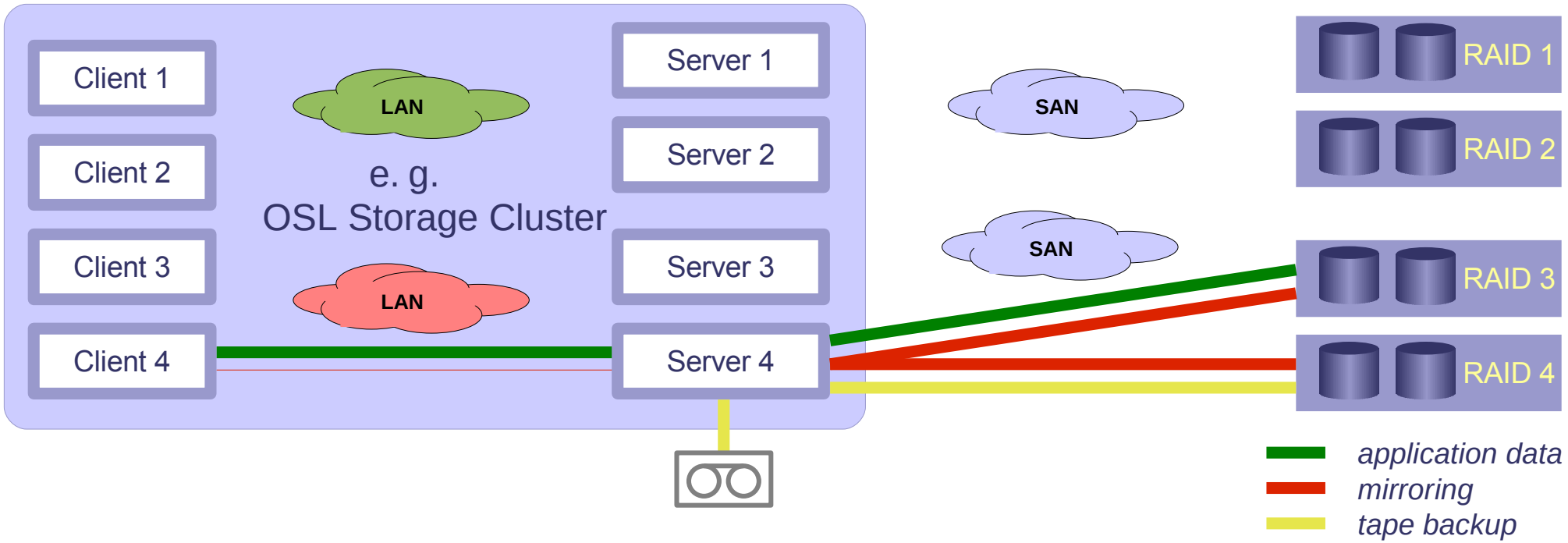
Example 6: Integrate servers and clients in a single cluster



- *alle functions like in example 3 (SAN-LAN-integration)*
- *in addition outstanding storage management functions are available:*
  - *storage allocation and management from the client side*
  - *application aware storage virtualisation can be used to its full extent on the client*
  - *transparent use of data mirroring, backup to disk etc.*
- *clients and servers merge to homogeneous cluster*
- *run applications everywhere*

What Concepts Can Be Implemented Using RSIO?

Example 7: High-speed backup for LAN-attached block devices



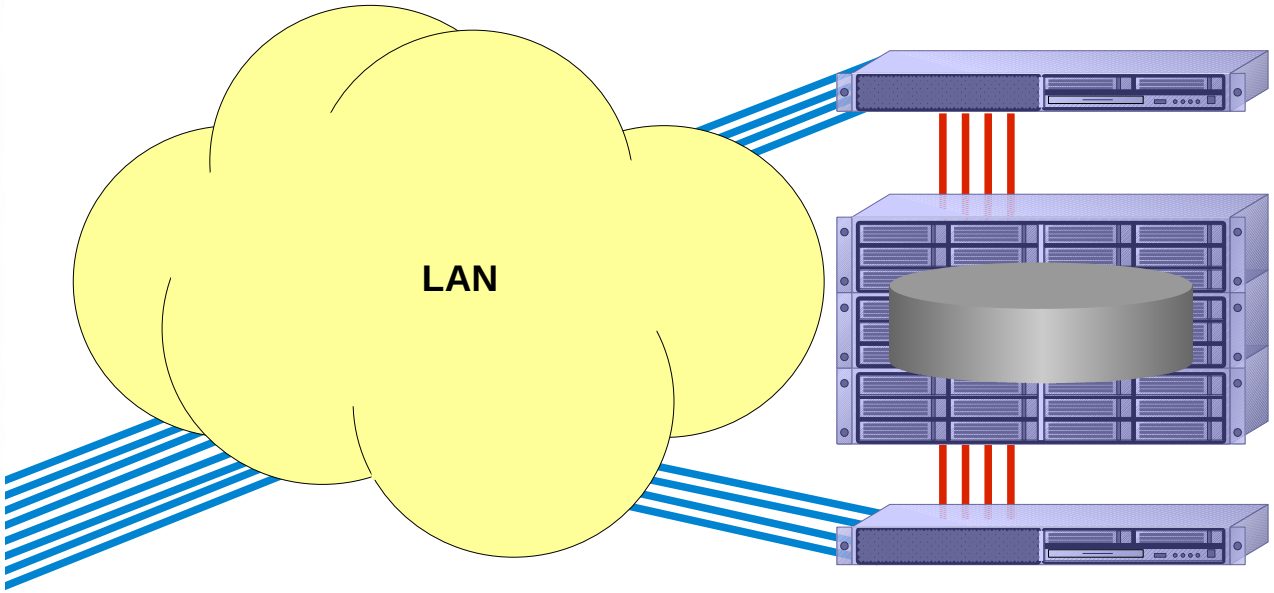
- *the LAN is only used for application data and control messages*
- *LAN-less backup:*
  - *high speed*
  - *full control from the client side*
  - *application aware commands*

What Concepts Can Be Implemented Using RSIO?

*Example 8: Storage, Management and HA for Cheap Server Farms**

CX-1000: 38 x ca. 300 RIP -> ca. 11.500 RIP*

compare yourself: M9000 32x SPARC64 VI 2400MHz ca. 1200 RIP



RSIO used for storage connectivity
OSL SC provides the framework
*for management, backup/DR, HA**

** comceptual draft*

Just use it!

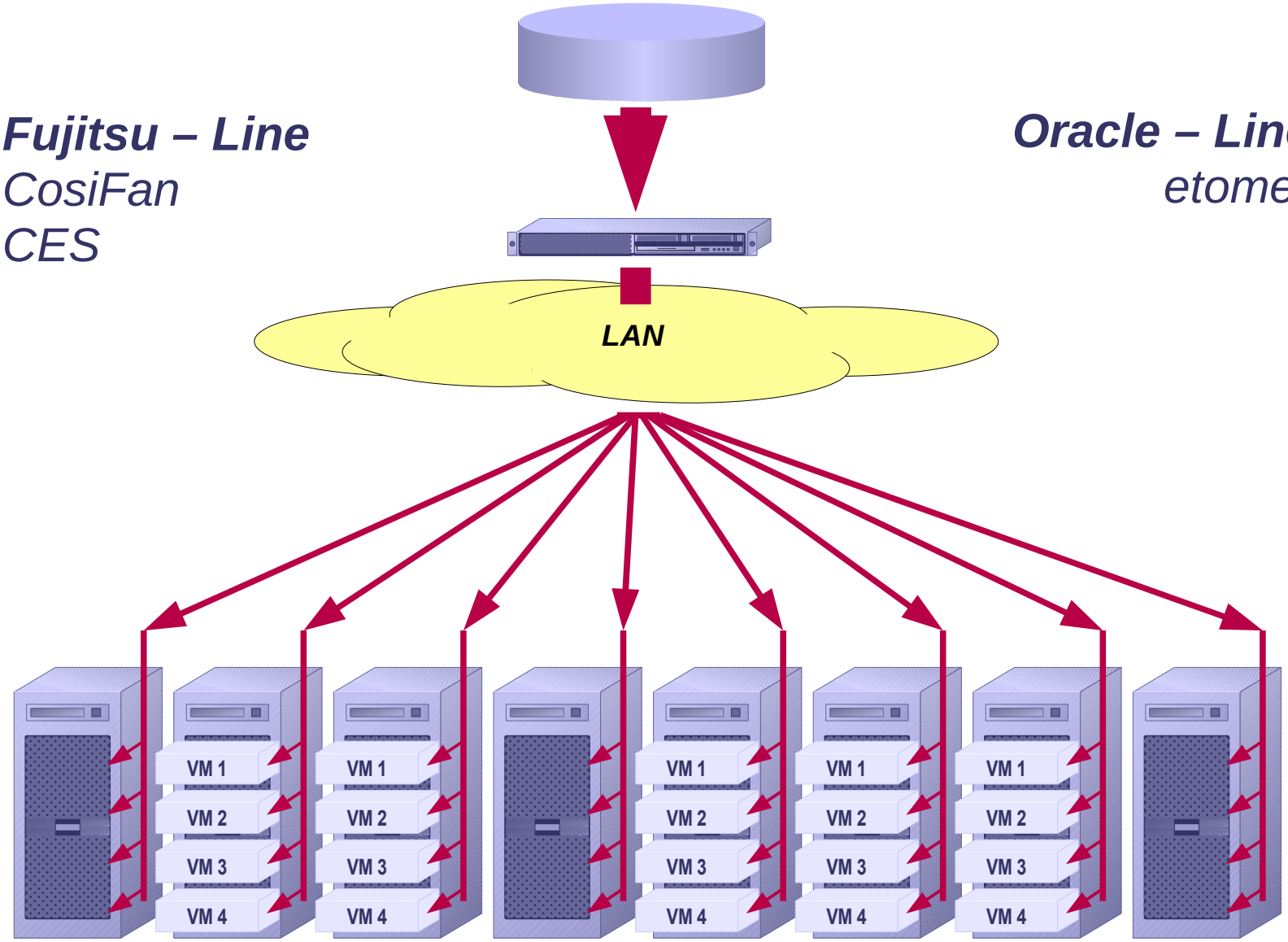
Pre-configured packages

Is there a Ready-to-Run Solution with RSIO?

The principle of pre-configured packages

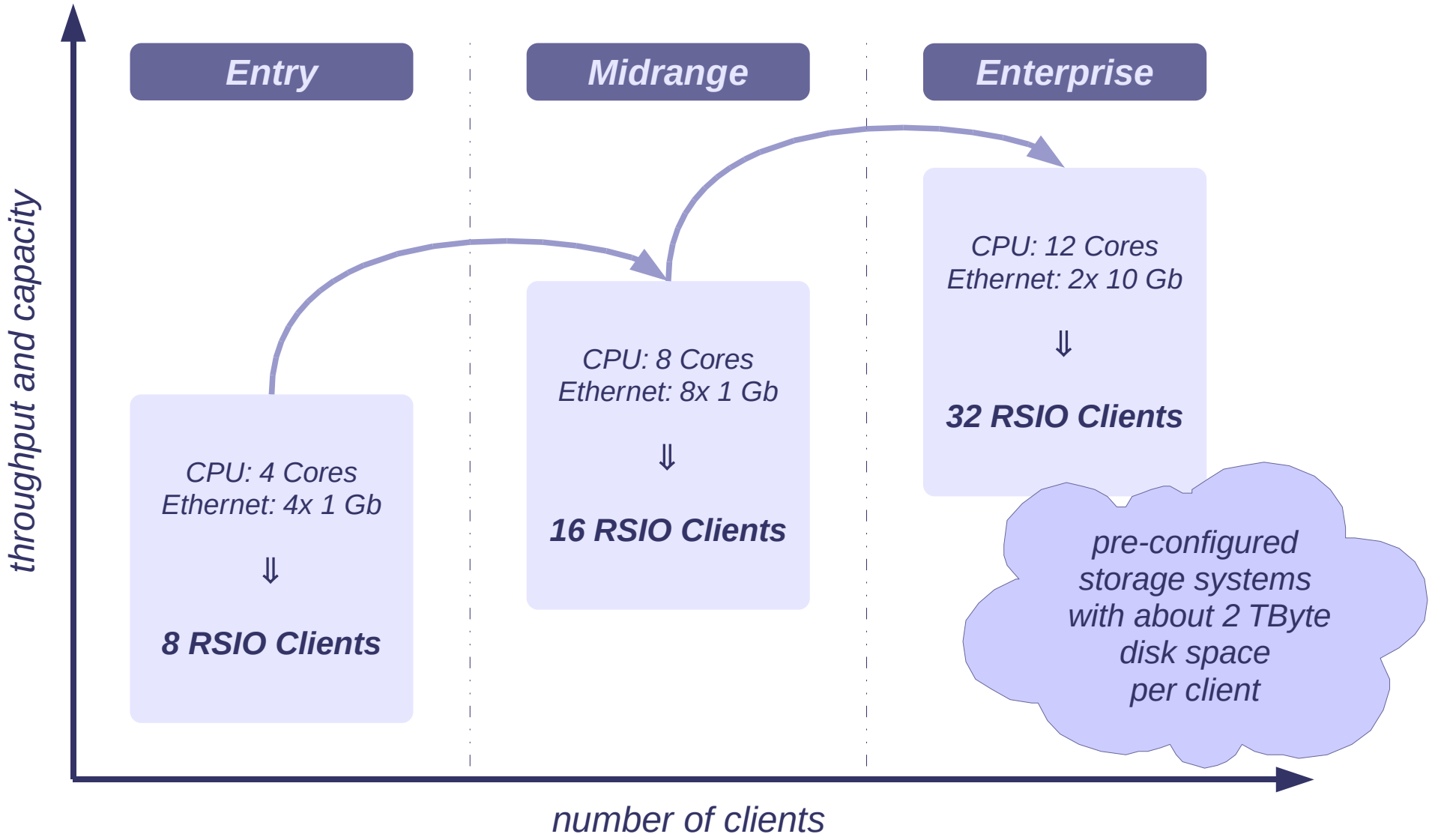
Fujitsu – Line
CosiFan
CES

Oracle – Line
etomer



RSIO – Pre-Configured Packages

Three expandable configurations



RSIO – “Entry Level” Package

Good value – targeted for up to 8 Clients

Fujitsu

Eternus DX80 (Single-Controller)

8 Gb Fibre Channel

Primergy RX100 S6
CPU: 1x Xeon X3430 (4 Cores)
LAN: 1 Karte 4x 1 Gb Ethernet

1 Gb Ethernet for RSIO
(8 Clients)

Oracle

ES-8200D (Dual-Controller)

8 Gb Fibre Channel

Sun Fire X4170 M2
CPU: 1x Xeon E5620 (4 Cores)
LAN: 4x 1 Gb Ethernet (internal)

1 Gb Ethernet for RSIO
(8 Clients)

OSL Gesellschaft für offene Systemlösungen mbH
www.osl.eu

RSIO – “Midrange” Package

Stronger systems for up to 16 clients

Fujitsu

Eternus DX80 (Dual-Controller)

8 Gb Fibre Channel

Primergy RX300 S6
CPU: 2x Xeon E5620 (2x 4 Cores)
LAN: 2 Karten 4x 1 Gb Ethernet

1 Gb Ethernet for RSIO
(16 Clients)

Oracle

ES-8200D (Dual-Controller)

8 Gb Fibre Channel

Sun Fire X4170 M2
CPU: 2x Xeon E5620 (2x 4 Cores)
LAN: 1 Karte 4x 1 Gb Ethernet
4x 1 Gb Ethernet (internal)

1 Gb Ethernet for RSIO
(16 Clients)

OSL Gesellschaft für offene Systemlösungen mbH
www.osl.eu

RSIO – “Enterprise” Package

High performance and high capacity – suitable for up to 32 clients

Fujitsu

Eternus DX80 (Dual-Controller)

8 Gb Fibre Channel

Primergy RX300 S6
CPU: 2x Xeon E5645 (2x 6 Cores)
LAN: 1 Karte 2x 10 Gb Ethernet

10 Gb Ethernet for RSIO
(32 Clients)

Oracle

ES-6600 (Dual-Controller)

8 Gb Fibre Channel

Sun Fire X4270 M2
CPU: 2x Xeon X5675 (2x 6 Cores)
LAN: 1 Karte 2x 10 Gb Ethernet

10 Gb Ethernet for RSIO
(32 Clients)

OSL Gesellschaft für offene Systemlösungen mbH
www.osl.eu

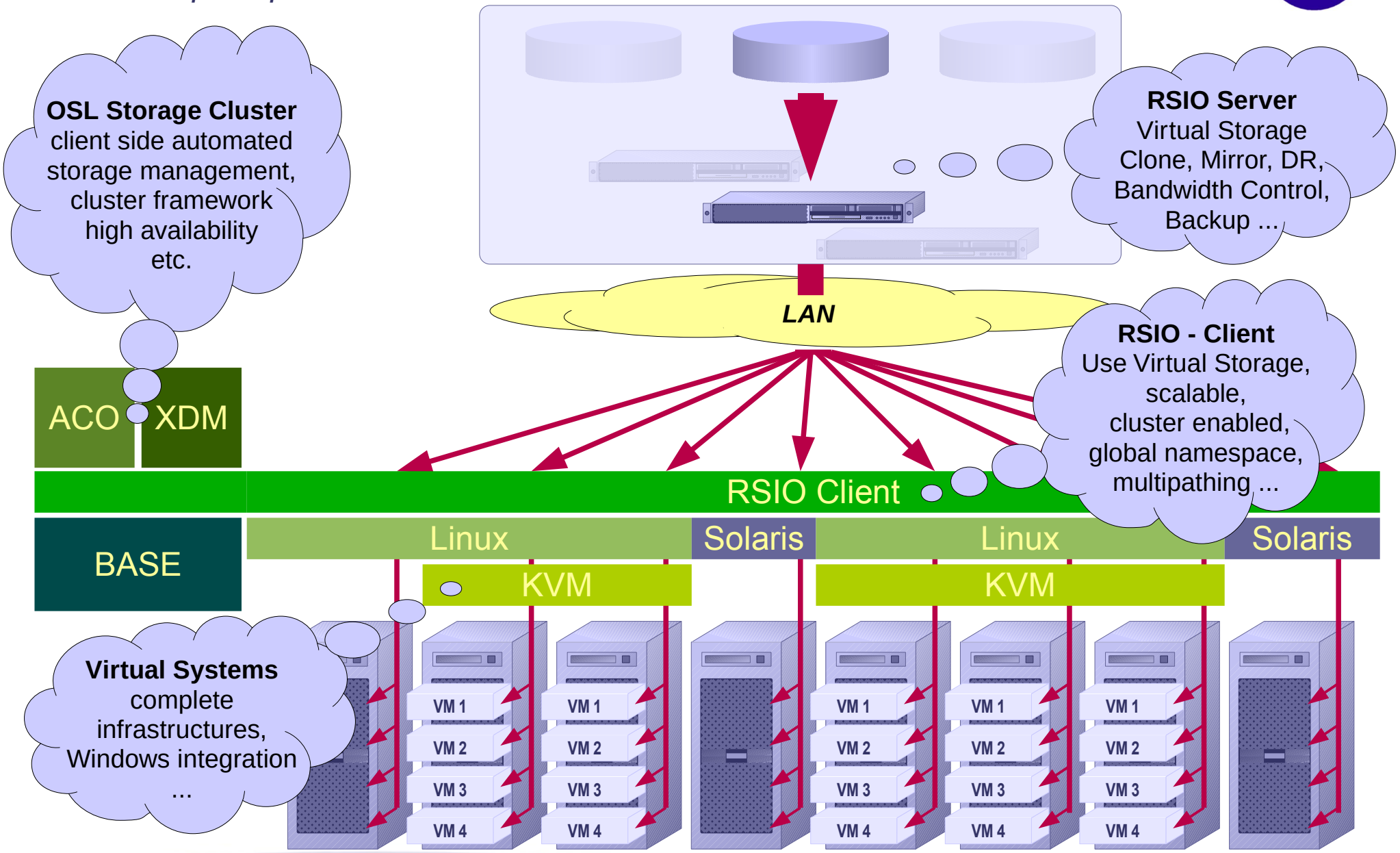
RSIO – Packages: The Costs

Infrastructure costs for SAN (4GB FC) and RSIO / Ethernet (without storage system)

	SAN	RSIO
Entry (8 Clients)	ca. 12.700	ca. 5.000
Midrange (16 Clients)	ca. 29.300	ca. 8.500
Enterprise (32 Clients)	ca. 70.000	ca. 18.800

Pre-Configured Packages

Back to the principle



OSL Gesellschaft für offene Systemlösungen mbH

www.osl.eu

How Far Is the Implementation?

Running in parallel: development, Piloting, Linux support, preparing general release

- *RSIO- pilot phase has started in february 2011 (Solaris / Linux)*
- *Announcement-Tour: Mai/Juni*
- *general release together with Storage Cluster 4.0 (exp. Oct. 2011)*
- *Technology Days 2011: 14./15. 9. 2011*
- *User Association / release workshop: exp. 29./30. 9. 2011*

OSL SC 4.0 – Further Topics

Seamless integration of virtual machines into the cluster-wide administration

- *introduction of Virtual Nodes*
  - *- can be “created” by means of software*
  - *- requires modification of node administration*
  - *- administration shares “look and feel” with physical node administration*
  - *- HA scenarios can be run/tested on a single physical node*
  - *- full access to storage virtualisation*
  - *- integration of resource management*
- *maximum number of nodes will be extended to 64*
  - *- the use of VMs brings about a considerable increase in the number of nodes*
  - *- an average company should be able of running all machines in a single cluster*
- *zones are a special challenge*
  - *- not an independent OS instance -> no own drivers*
  - *- new release shall grant full access to all SC services to zones (virtual nodes), e. g. application aware automatic backup*
  - *- special programs for zones will be eliminated*
- *improved application management*
  - *- virtualised applications can be moved easily between physical nodes and virtual nodes*

Summary OSL SC and RSIO

- *leading technology in storage virtualisation and clustering*
- *convincing roadmap – clear communication*
- *open for new systems (e.g. Linux) and integration of virtual machines*
- *considerable advantages of RSIO for Linux/Solaris*
  - *it has never been easier to turn standard servers into powerful storage servers*
  - *it has never been easier and cheaper to implement a powerful storage connectivity for servers*
  - *never before storage servers and clients have been able to share the same highly integrated cluster*
  - *never before all things could be run over the same cable (virtual storage, HA cluster, admin)*
  - *never before there was such a degree of application awareness and integration of HA and DR with the storage administration*
  - *never before there was such a deep integration of storage virtualisation and backup to disk/tape*
  - *never before there was such an integration of Solaris and Linux in a single cluster*
  - *never before there was a system that could integrate several concepts of virtual machines*
  - *never before there was such a wide variety of possible combinations of standard modules*
  - *never before there was such a comfort for several platforms*