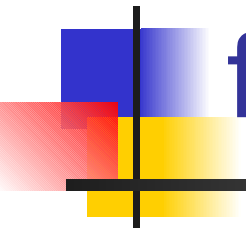
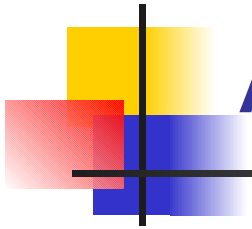


Proboscis : Flexible support for cluster storage services

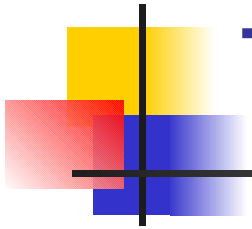
Sardes Project (IMAG-LSR / INRIA)

DistLab-DIKU



Aim

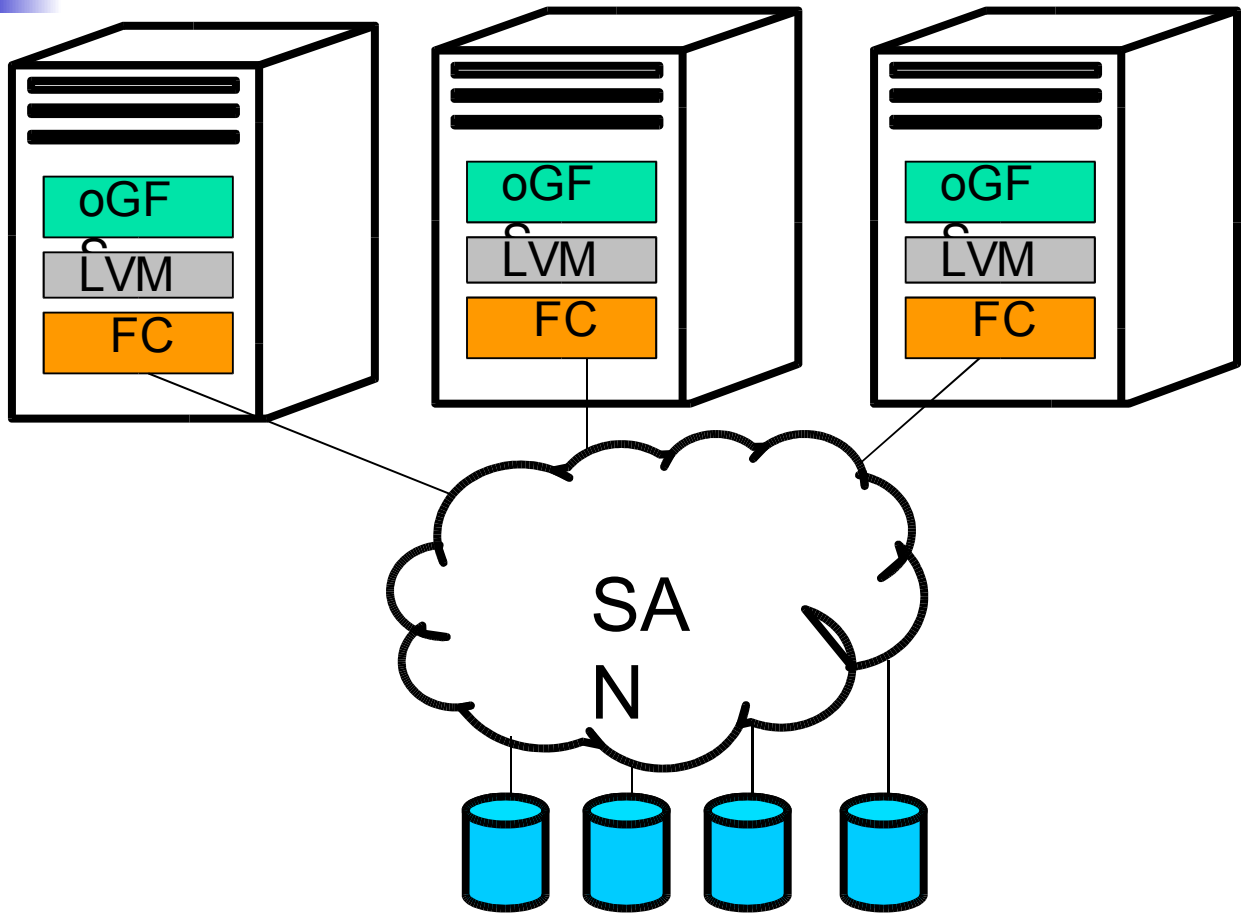
- Avoid I/O bottlenecks and expensive SAN hardware
  - Efficient sharing of disks locally attached to cluster nodes
  - Use existing system area network
  - Cost-effective
  - I/O bandwidth can scale with the number of nodes
- Each node of the cluster acts as a storage node
  - And also as a compute node
- Limited overhead with Remote DMA interconnects
  - SCI
  - Myrinet



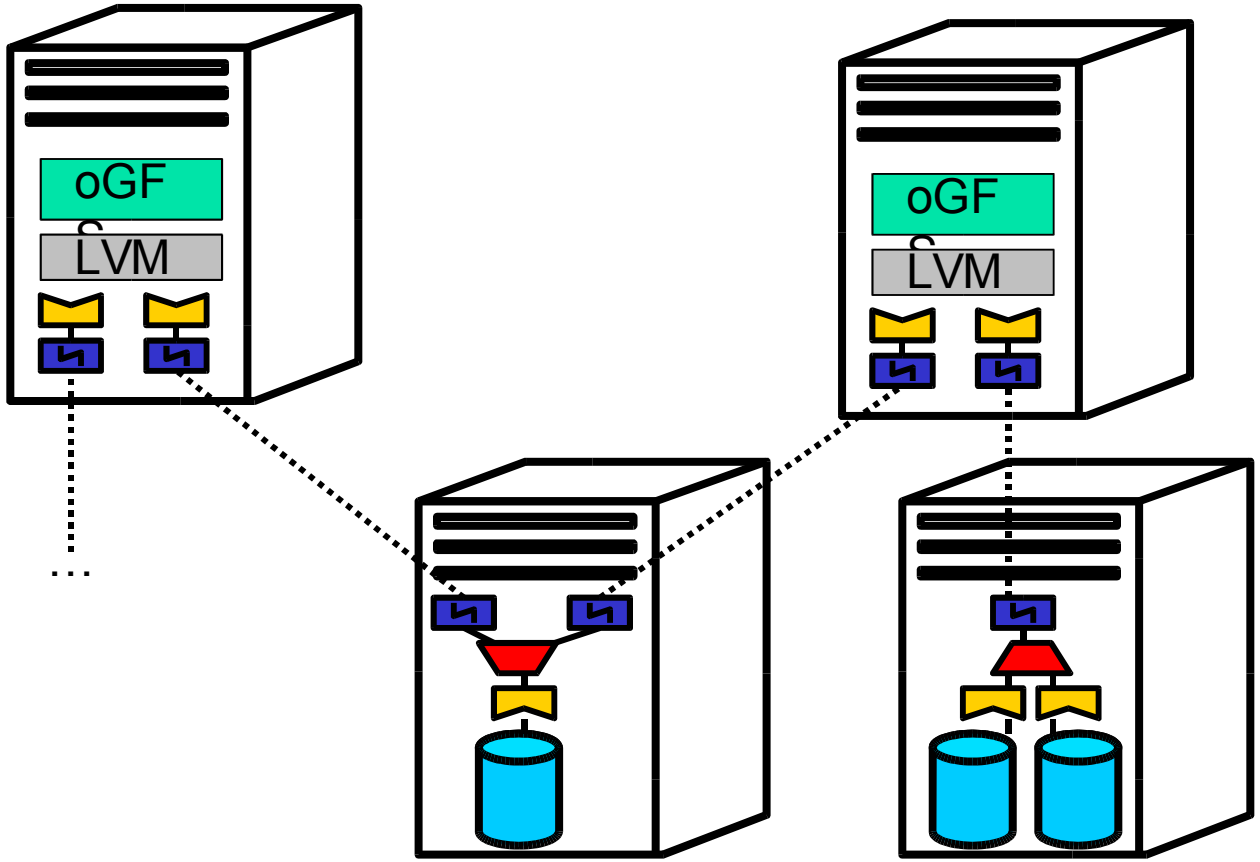
The Proboscis Framework

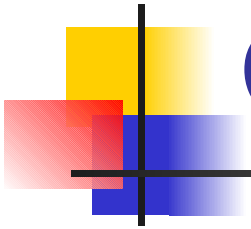
- A modular software infrastructure for remote disk access construction and administration
  - Basic building blocks
    - Software RAID, disk scheduling
  - Flexibility
    - Reconfiguration
- *Not* a Parallel File System
  - Block-device level
  - Can be combined with a Parallel FS
    - (open)GFS

Comparison : (1) openGFS / SAN



Comparison :
(2) openGFS / Proboscis





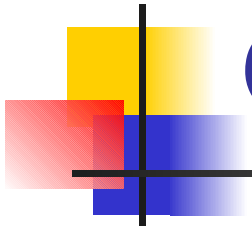
Current status

Stable version for Linux 2.4 on IA32

SCI and Ethernet (TCP/IP)

Remote disk access performance penalty:

	No CPU load	CPU under load low load / high load
SCI	= local	5-10% / 10-20%
Gigabit Ethernet	= local	40% / 50-60%
Fast Ethernet	10 MB/s	20-30% / 30-40%

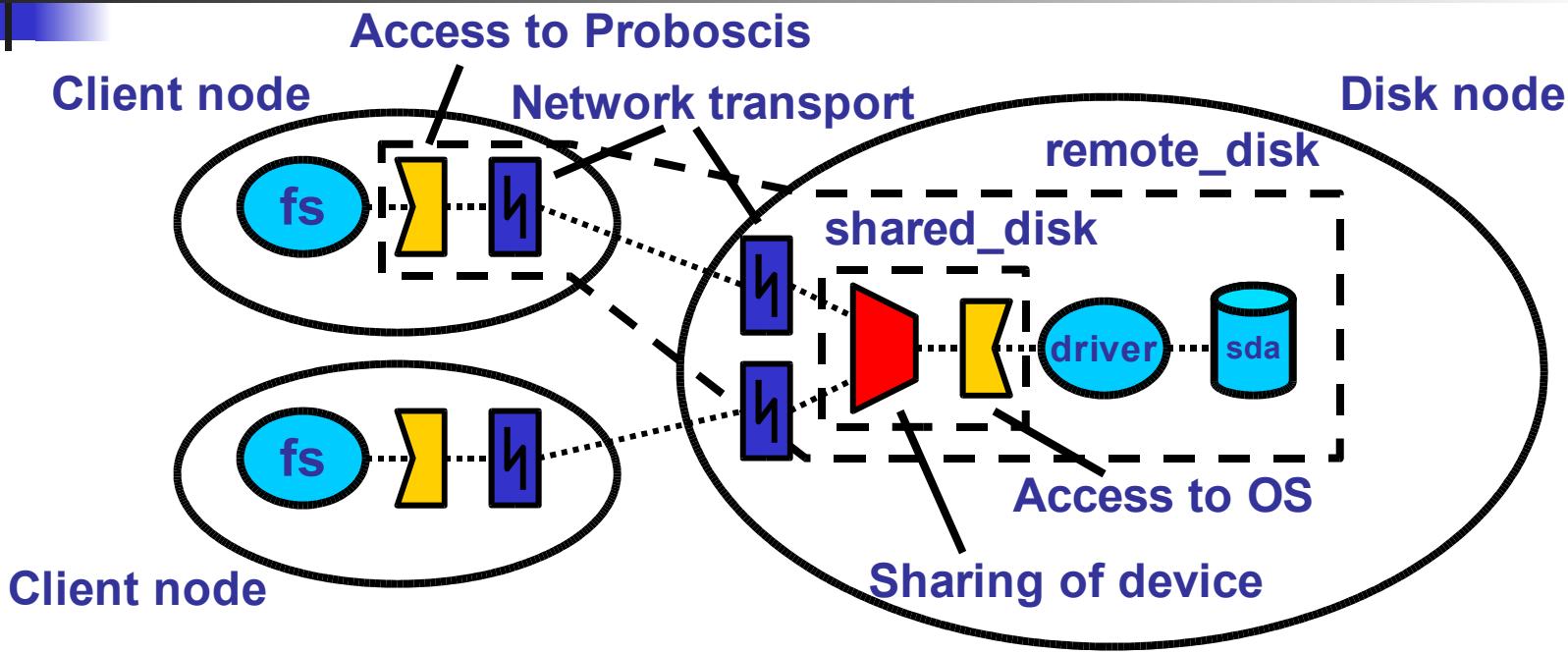


Ongoing & future work

- Dynamic reconfiguration
  - Load balancing
  - Fault-tolerance
- Myrinet support
- IA64 port

More info at :
http://sardes.inrialpes.fr/~rlachaiz/proboscis.shtml

Example



shared_disk = “/share/prob2os dev=sda” [bound]

remote_disk = “/os2prob dev=prob1/{shared_disk}”