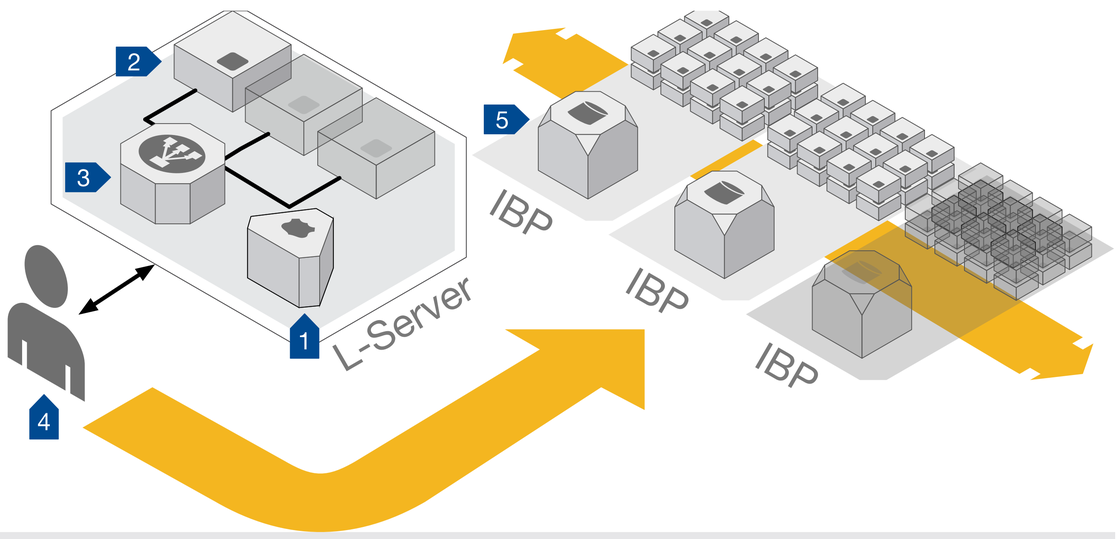
[LSTORE] Logistical Storage

http://www.lstore.org

L-Store provides a flexible logistical storage framework for distributed and scalable access to data for a wide spectrum of users. It contains:

- Virtually unlimited scalability in raw storage
- Support for arbitrary metadata associated with each file
- User controlled fault tolerance and data reliability on a file and directory level
- Scalable performance in raw data movement
- A FUSE-based file system interface with both a native mount in Linux (exportable via NFS and CIFS to other platforms)
- High performance command line interface
- Support for the geographical distribution and migration of data

These features are accomplished by segregating directory and metadata services from data transfer. L-Store clients use the **L-Server** only for metadata operations, removing an important bottleneck.



L-Store Overview

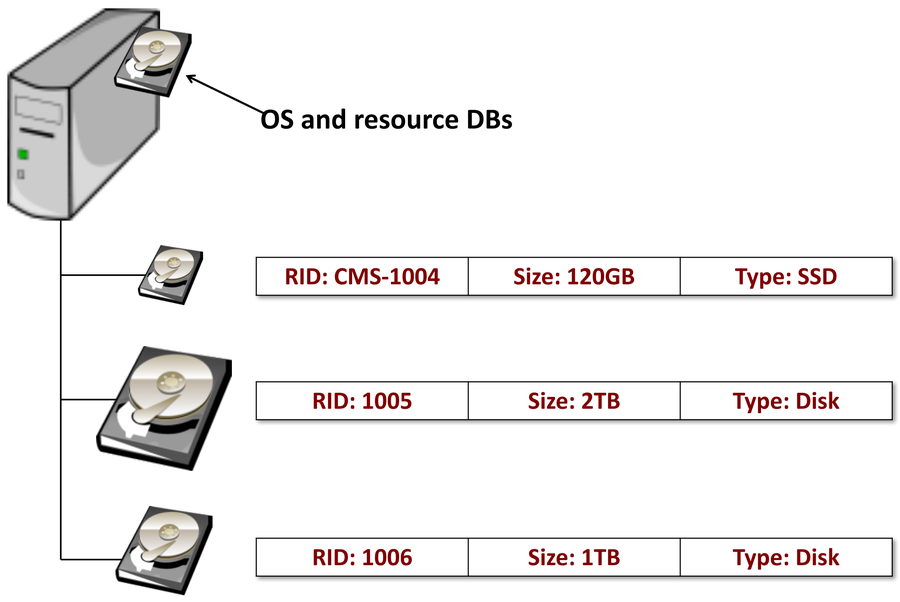
1 L-Store includes an extensible AuthN/AuthZ framework. This easily allows adding additional functionality to support local needs

2 The L-Server serves all of the relevant metadata to clients, after they pass authentication. It also implements distributed cache coherency for connected clients

3 High-availability can be provided with a load balancer-backed active/passive solution.

4 Once the client has retrieved the appropriate metadata, it can read/write data directly from the IBP depots, bypassing the L-Server itself.

5 Scaling capacity and performance of L-Store is simply a matter of adding more depots



Traditional RAID arrays completely reconstruct a single failed drive on a single replacement drive.

L-Store uses distributed RAID arrays which are designed to overcome these limitations. Instead of using the whole disk the disk is broken up into many smaller blocks. These blocks are combined with blocks on other disks creating many small logical RAID arrays utilizing a large subset of the available drives. The free space on each drive can be used to store the newly reconstructed data. This allows for a large number of drives being read and written to simultaneously providing significantly faster rebuild times.

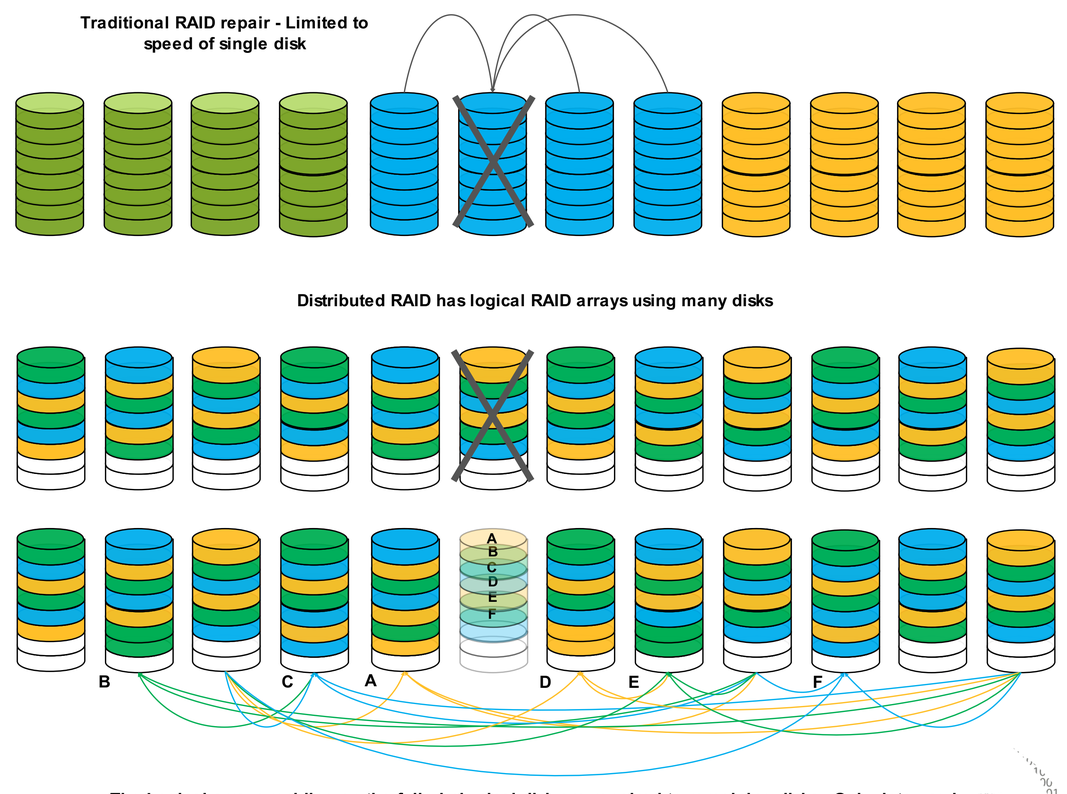
The largest L-Store installation is currently used by the CMS Tier-2 at Vanderbilt. This installation has grown from 100s of Terabytes to nearly 5PB of CMS data divided among millions of files without any appreciable scaling issues. This installation has proven to scale to thousands of concurrent clients and bulk transfer rates in excess of 150Gbit/sec.

IBP
internet backplane protocol

L-Store builds on a highly generic, best effort storage service, called the Internet Backplane Protocol (IBP)

IBP was designed at the University of Tennessee, Knoxville to be the data analogue of the network functionality provided by the Internet Protocol (IP). Similar to the relationship between TCP and IP, L-Store extends the relatively simple semantics of IBP to provide features like fault-tolerance and scalability.

An **IBP depot** is a server with one or many drives in a JBOD configuration. This host runs an **IBP daemon**, which exports each drive individually to interested clients. This allows a client to implement, for instance, RAID-1 by performing all writes to two separate drives. This can be done independent of any IBP-specific knowledge of the implemented algorithm.



VANDERBILT UNIVERSITY

ACCRE Advanced Computing Center for Research & Education

Enabling Researcher-Driven Innovation and Exploration

LStore Highlights:

Home-grown distributed filesystem
Built to be flexible, extensible and able to
  use storage with simple semantics
Used at the multi-PB scale with CMS