Choose Lustre

Stephen Simms

Manager, High Performance File Systems

Indiana University

https://www.flickr.com/photos/kubina/521327318

Lustre is scalable – 55 PB at LLNL

http://www.flickr.com/photos/shadowstorm/404158384/


https://www.flickr.com/photos/davesag/18735941

Lustre is fast – 1 TB/s at ORNL


https://www.flickr.com/photos/johngarghan/5688711425

Lustre can support thousands of clients

Lustre is Open

Lustre is Open source software under GPLv2

That means it’s
“Free Like Beer”
right?

Actually, more like a free puppy…

It Takes Lots of Work to Maintain Lustre

- Bug Fixes
- Rigorous Testing
- Feature Development
- Maintaining Documentation
- Tree Hosting
- Code Reviews by Peers

Many Hands Make Lighter Work

Lustre Needs You to Join the Party!

https://www.flickr.com/photos/alvi2047/3688993279/

Lustre is moving forward

- Hiccup when Lustre moved from Oracle
  - Lustre 2.0 – Fall 2010
  - Lustre 2.1 – Fall 2011
- Since then Lustre has accelerated
  - Goal of 2 major releases a year
  - Spring / Fall

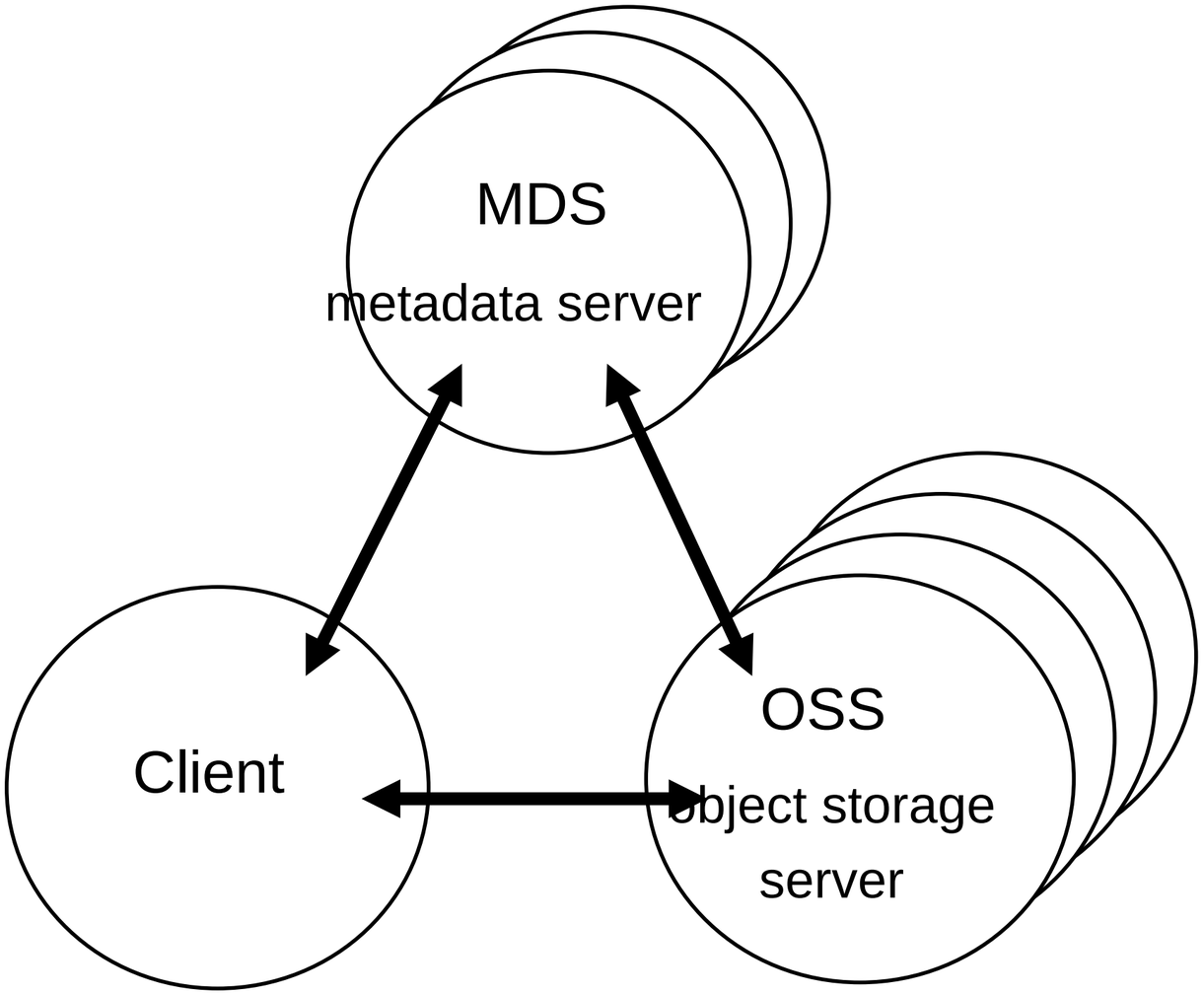
Alright, alright, stop the marketing

Lustre = Linux +Cluster

Lustre is a *parallel* distributed file system

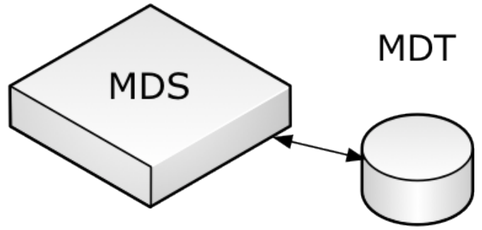
- High performance filesystem used by >60 of the top 100 supercomputers in the world
- POSIX compliant – behaves like other file systems

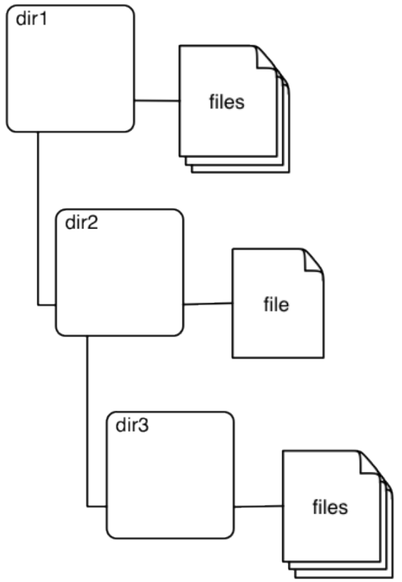
Lustre: The Players



Lustre Architecture - MDS

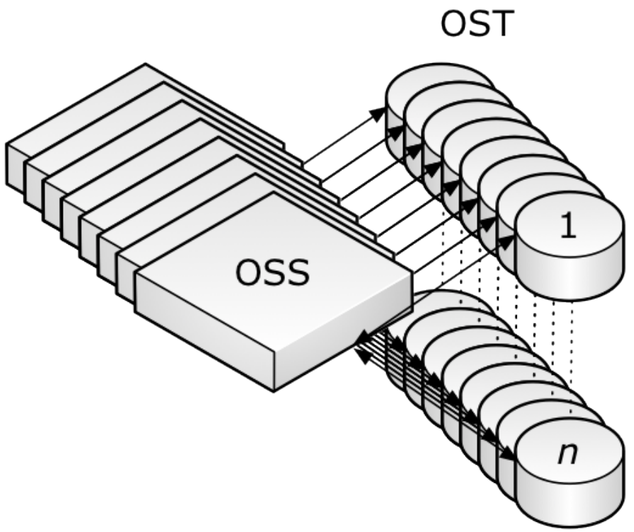
- **Metadata Server (MDS)**
  - Node(s) that manage namespace, file creation and layout, and locking.
    - Directory operations
    - File open/close
    - File status
    - File creation
    - Map of file object location
  - Relatively expensive serial atomic transactions to maintain consistency
- **Metadata Target (MDT)**
  - Block device that stores metadata

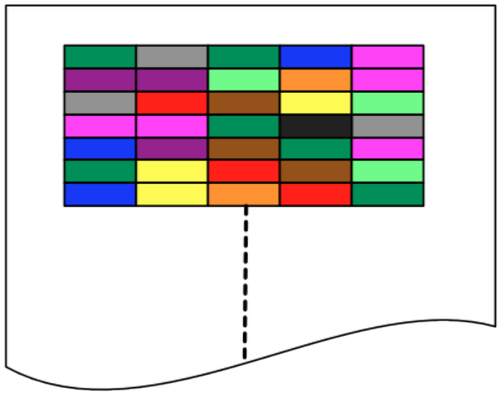




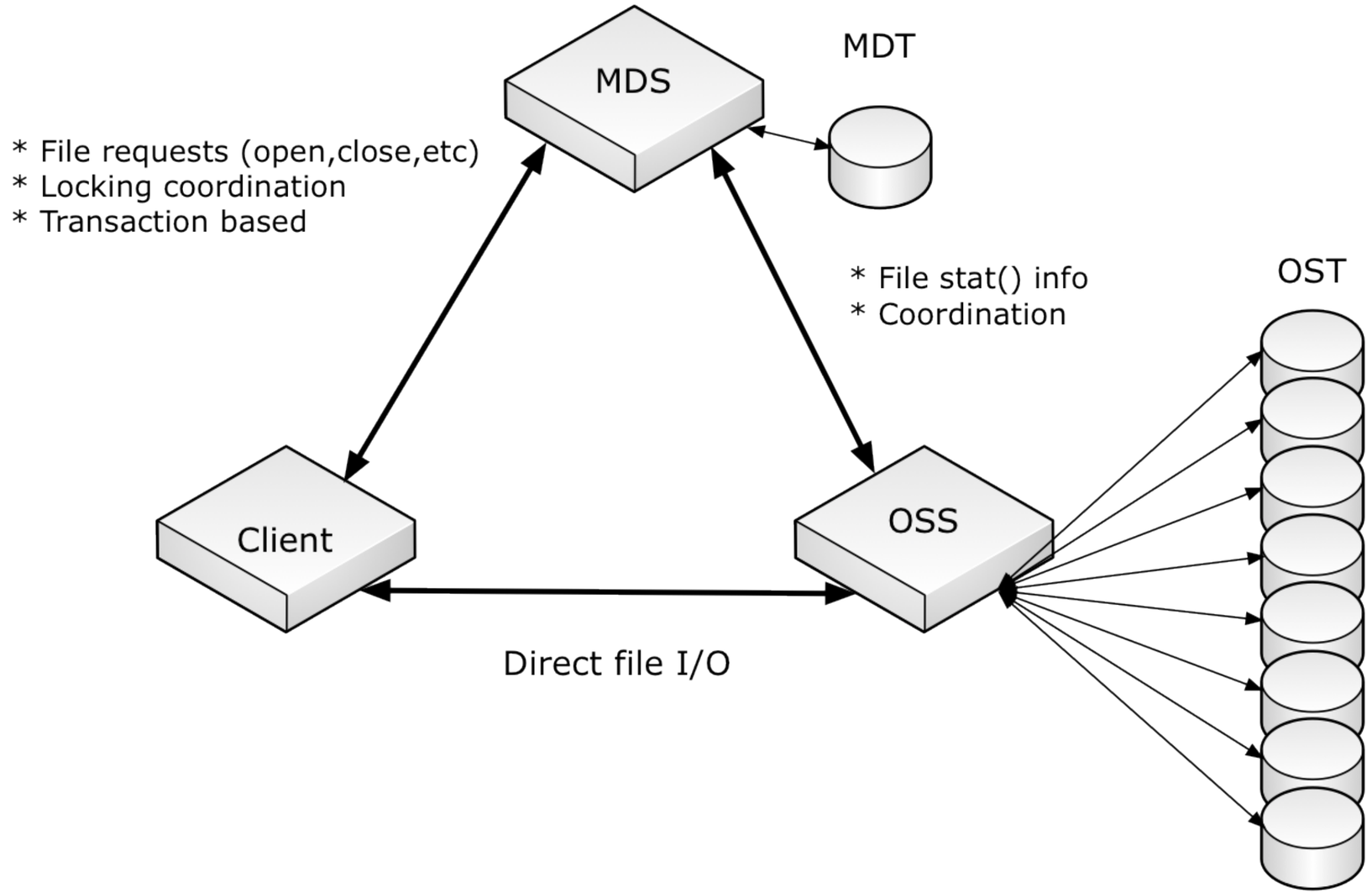
Lustre Architecture - OSS

- **Object Storage Server (OSS)**
  - Multiple nodes that manage network requests for file objects on disk.
- **Object Storage Target (OST)**
  - Block device that stores file objects

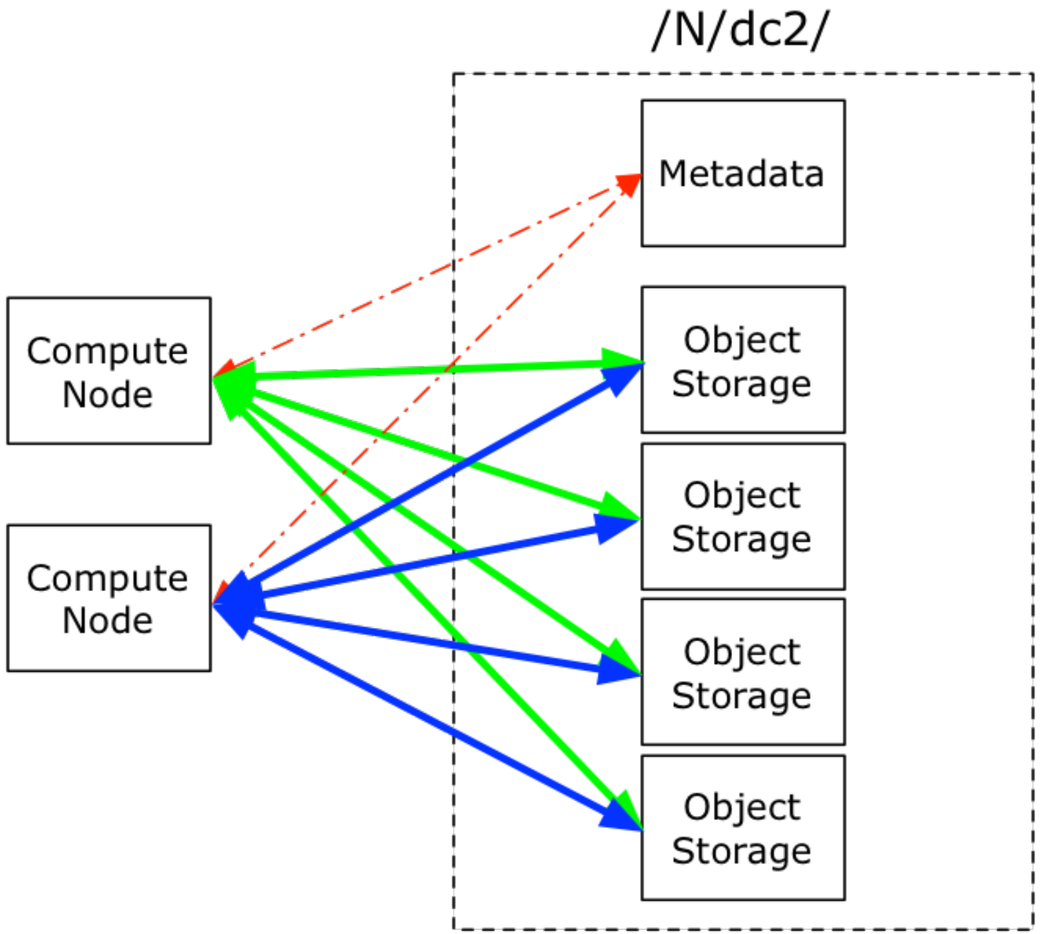




Simplest Lustre System

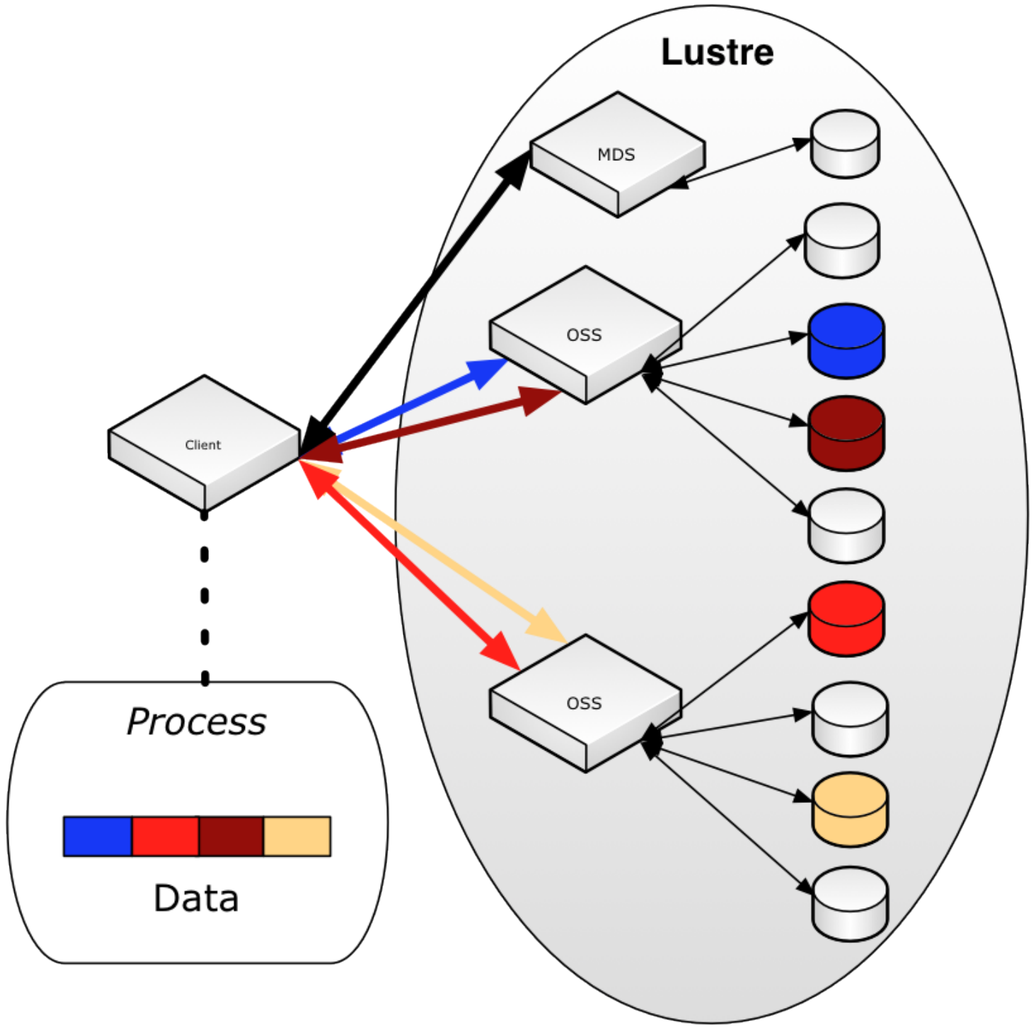


Lustre Parallel I/O



Striping Data

- Lustre allows you to control *how* data is written, if you want
  - *Stripe* data across multiple OSTs
    - can stripe files OR directories
  - Can increase I/O performance with reading and writing
  - With DNE2 metadata can be *Striped* across multiple MDTs
- Striping analogous to RAID 0
- Default striping set by sysadmin



Take Home Message

Choose Lustre!

It scales – size, speed, clients

It’s open, growing, and needs your help

It gives users powerful options

Tools available to help with installation

Filets, Chops, Removes household odors

Act now and no salesman will visit your home


CHOOSE
LUSTRE

Thank You for Your Time and Attention

www.opensfs.org

Open Scalable File Systems, Inc.
3855 SW 153rd Drive
Beaverton, OR 97006
Ph: 503-619-0561
Fax: 503-644-6708
admin@opensfs.org