



Lecture 21: Parallel Filesystems

Abhinav Bhatele, Department of Computer Science



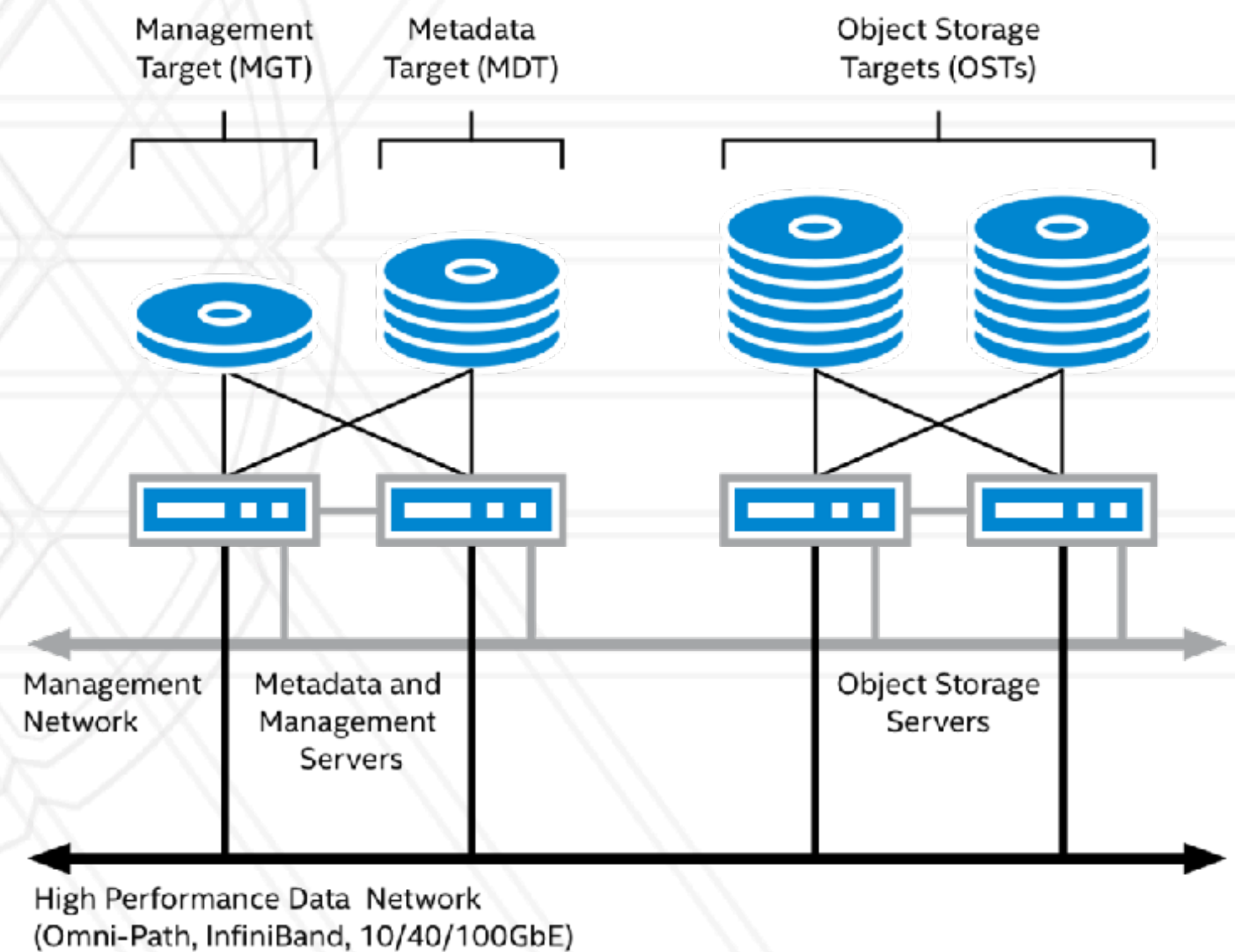
UNIVERSITY OF
MARYLAND

Announcements

- Interim report was due yesterday
- Assignment 4 is due next week

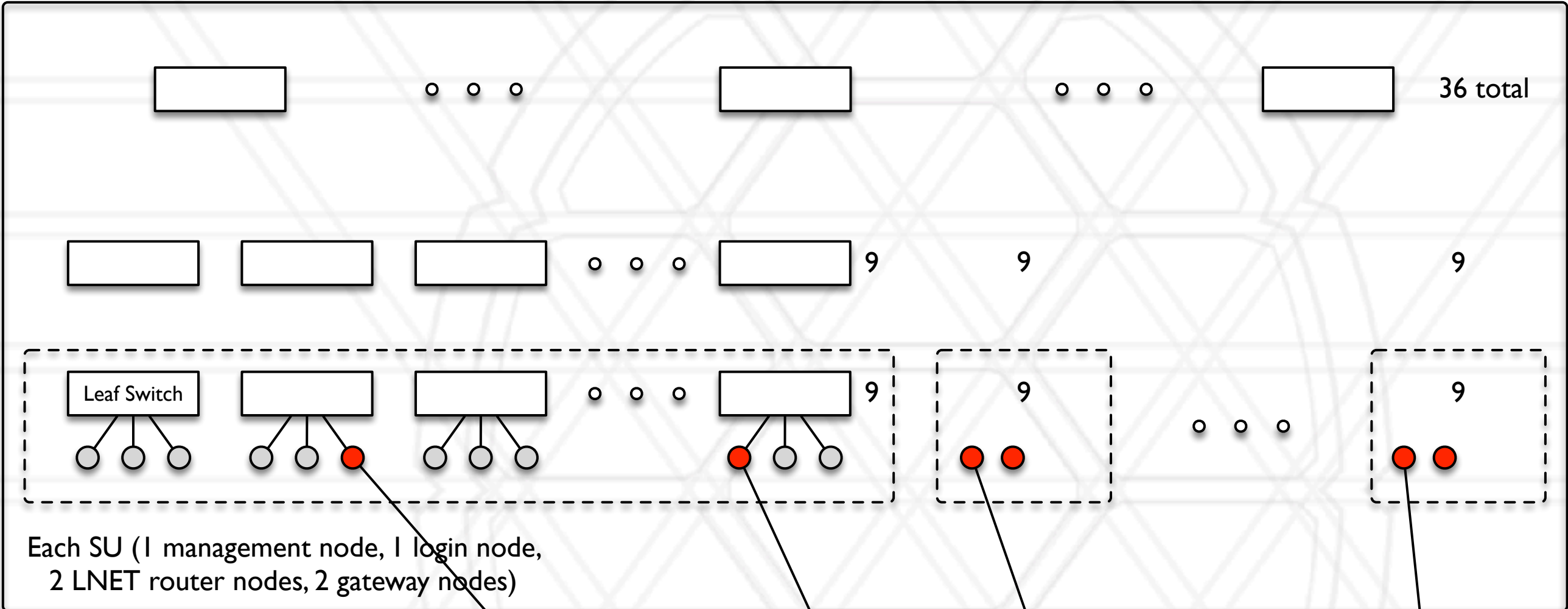
Parallel filesystem

- Home directories and scratch space typically on a parallel file system
- Mounted on all login and compute nodes
- Also referred to as I/O sub-system

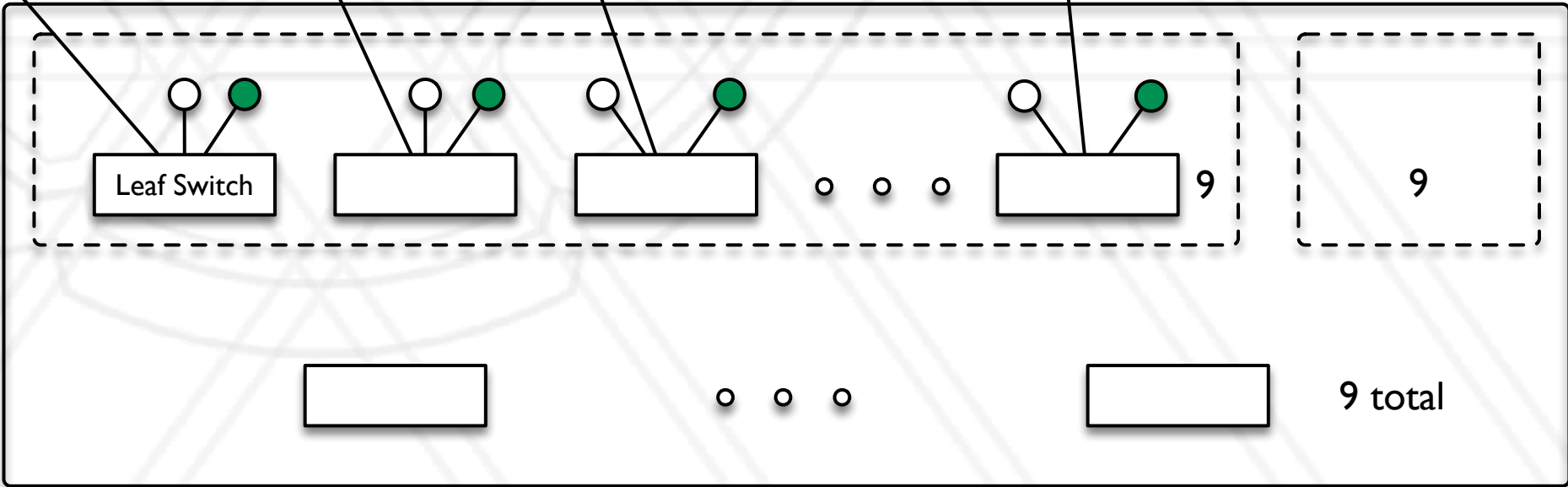


http://wiki.lustre.org/Introduction_to_Lustre

Links between cluster and filesystem



- Compute node
- LNET router node
- Object storage server (OSS)



Different parallel filesystems

- Lustre: open-source (lustre.org)
- GPFS: General Parallel File System from IBM, now called Spectrum Scale
- PVFS: Parallel Virtual File System

Tape drive

- Store data on magnetic tapes
- Used for archiving data
- Use robotic arms to access the right tape: <https://www.youtube.com/watch?v=d-eWDuEo-3Q>

Burst buffer

- Fast, intermediate storage between compute nodes and the parallel filesystem
- Two designs:
 - Node-local burst buffer
 - Remote (shared) burst buffer

I/O libraries

- High-level libraries: HDF5, NetCDF
- Middleware: MPI-IO
- Low-level: POSIX IO

Different I/O patterns

- One process reading/writing all the data
- Multiple processes reading/writing data from/to shared file
- Multiple processes reading/writing data from/to different files
- Different performance depending upon number of readers/writers, file sizes, filesystem etc.

I/O profiling tools

- Darshan
 - Lightweight profiling tool from Argonne National Lab
- Recorder
 - Research prototype from UIUC



UNIVERSITY OF
MARYLAND

Abhinav Bhatele

5218 Brendan Iribe Center (IRB) / College Park, MD 20742

phone: 301.405.4507 / e-mail: bhatele@cs.umd.edu