Micro-Storage Services for Open Ethernet Drive

WIP Session
PDSW-DISCS 2017

2nd Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems

Hariharan Devarajan, hdevarajan@hawk.iit.edu

Anthony Kougkas, akougkas@hawk.it.edu

Xian-He Sun, sun@iit.edu

ILLINOIS INSTITUTE OF TECHNOLOGY

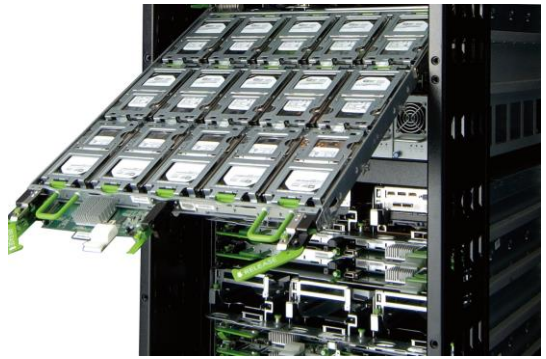
SCALABLE COMPUTING SOFTWARE LABORATORY

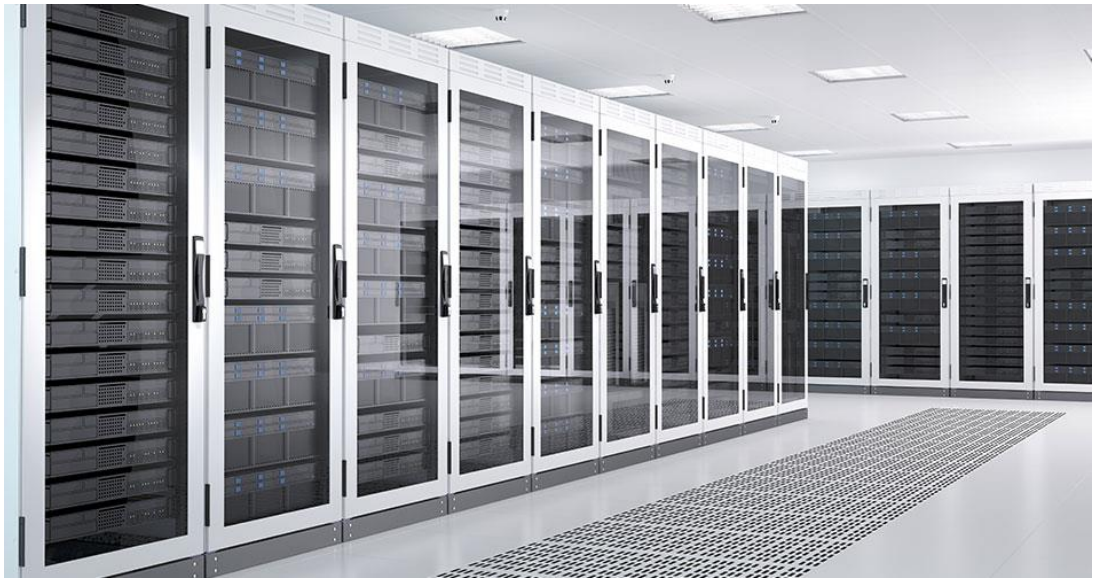
Micro-Storage Services for Open Ethernet Drive
Hariharan Devarajan, PhD Student, hdevarajan@hawk.iit.edu

Introduction

Supercomputer	K	Kaust	Tianhe-2	Trinity
# storage nodes	2000	400	1000	400

- High cost from storage
  - Purchase
  - Real-Estate (physical space)
  - Maintenance
  - Energy
    - Up to 40% of the entire energy footprint
- A very long and complex storage software stack
- Exa-scale will exacerbate this problem

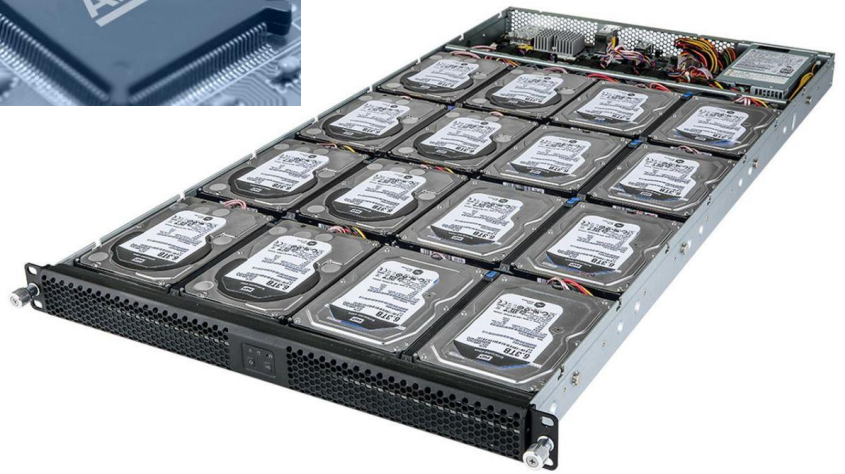




Open Ethernet Drive

- Intelligent drive
  - ARM-powered
  - Fixed sized ram
  - Network card
- Runs full-fledged Linux OS
- Prototype devices by:
  - Seagate Kinetic
  - Western Digital (HGST)
- Presented in enclosures of multiple such drives (JBOD)
- Enclosures have an embedded switched fabric (60Gbit/s)

	OED 1st Gen	OED 2nd Gen
CPU	ARM 32bit 1-core (1GHz)	ARM 32bit 2-core(2.2GHz)
RAM	2GB DDR3 1-Channel 1333Mhz	1GB DDR3 2-Channel 1600MHz
Disk	Megascale DC4000.A 4TB 7200rpm	Megascale DC4000.B 8TB 7200rpm
OS	Debian 8.0	Debian 8.1
Kernel	3.14.3	3.9
Year	2014	2016



Open Ethernet Drive - Initial results

Pros

- OEDs are capable Parallel FS and Object Store servers as well as I/O accelerators (i.e., burst buffers).
- OEDs proved to be 2.2x to 15x more energy efficient than a typical server.
- Can achieve great parallelism for the same power cap

Cons

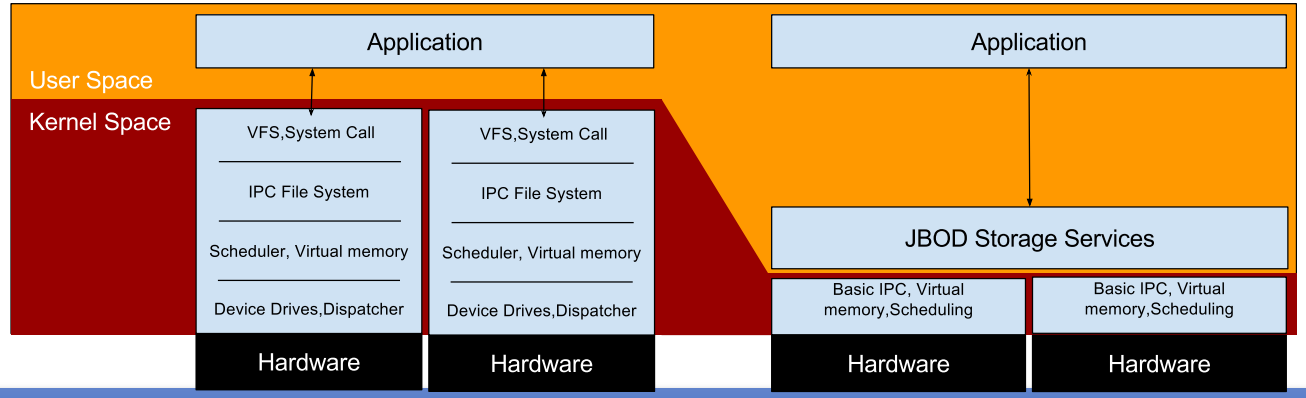
- Computation power is not at par with server nodes
- No API to use JBOD.
- Running a full-fledged Linux OS on OEDs is extremely heavy and poses unnecessary overheads

Published Work

- H. Devarajan, A. Kougkas, and X. H. Sun, "Open Ethernet Drive Evolution of Energy-Efficient Storage Technology." in Proceedings of DataCloud'17, Denver,CO.
- A. Kougkas, A. Fleck, and X. H. Sun, "Towards energy efficient data management in HPC: The open Ethernet drive approach," in Proceedings of PDSW-DISCS'16: 2017, pp. 43–48.

Proposal – Design Objectives

- Micro storage kernel
  - Minimize OS unnecessary overheads.
    - Modules which are not crucial to storage nodes would be removed.
  - Maximize performance
    - Fine-tune the kernel to better suit the needs of the OED technology
- Lightweight API
  - Maximize utilization of JBOD
  - Parallelization of I/O tasks
  - Offload small computation to JBOD
  - JBOD Services:
    - Manager, I/O Scheduler, Load Balancer
  - Provide mount point for application



Our first steps

- BusyBox 1.27.2 Linux
  - As a building block
  - Very small size (i.e., ~5MB)
  - Add XFS file system
- Results
  - Reduced boot time by 1300%
  - Smaller memory footprint leading to more available memory to applications (i.e., from 350MB to only 15MB)
- Next step:
  - Investigate other lightweight Linux distributions for embedded and mobile platforms (e.g., ToyBox)
  - Develop a light-weight parallel file system within the JBOD.

Micro-Storage Services for Open Ethernet Drive

Hariharan Devarajan, hdevarajan@hawk.iit.edu