

Linux Plumbers Conference

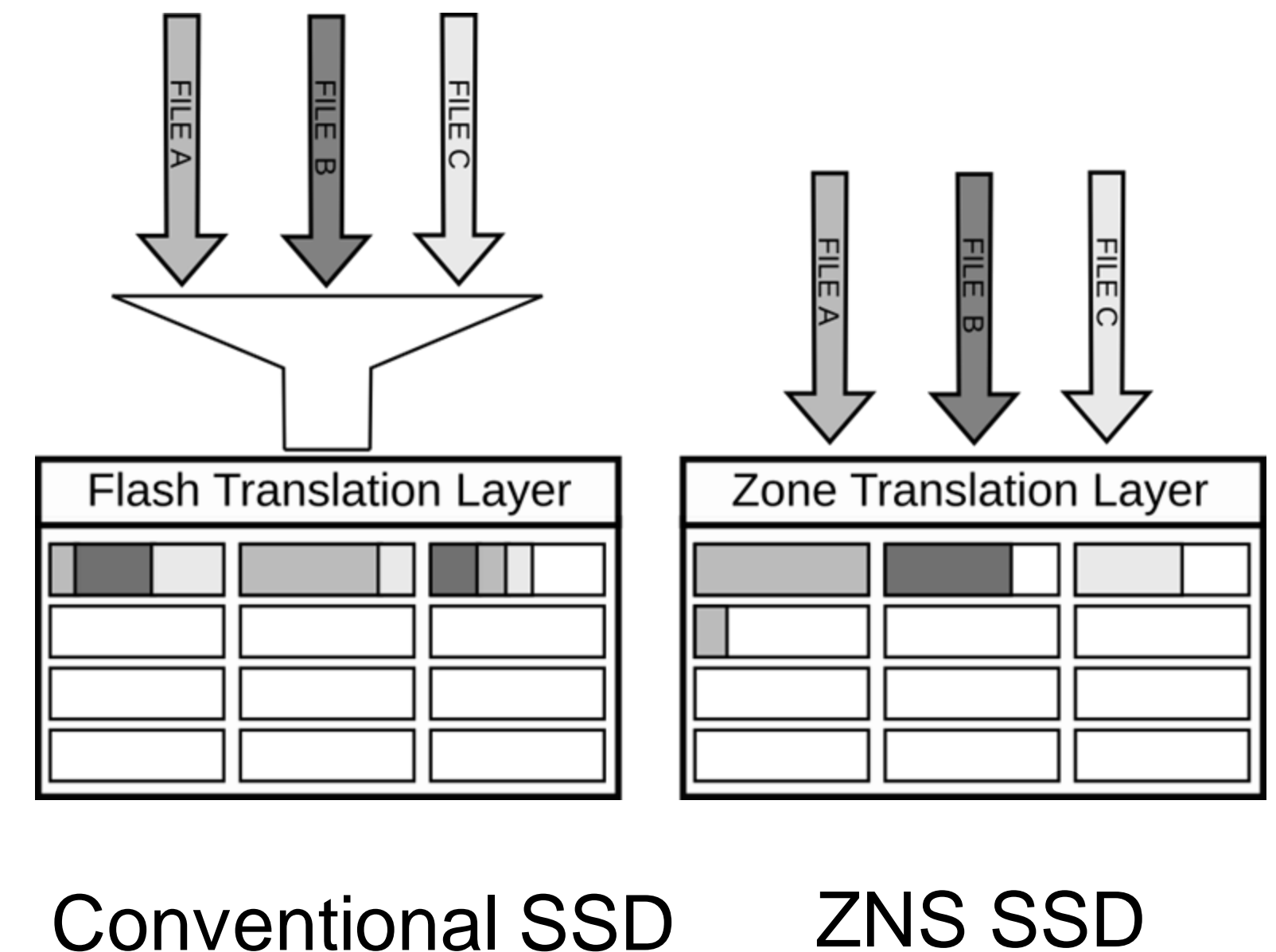
Dublin, Ireland September 12-14, 2022

Improving data placement for Zoned Linux File systems

Hans Holmberg, Western Digital Corporation

Problem description

- Zoned storage enables
 - Better media capacity usage
 - Improved performance
 - ..through **less Garbage Collection**
 - **Less GC** requires smart data placement

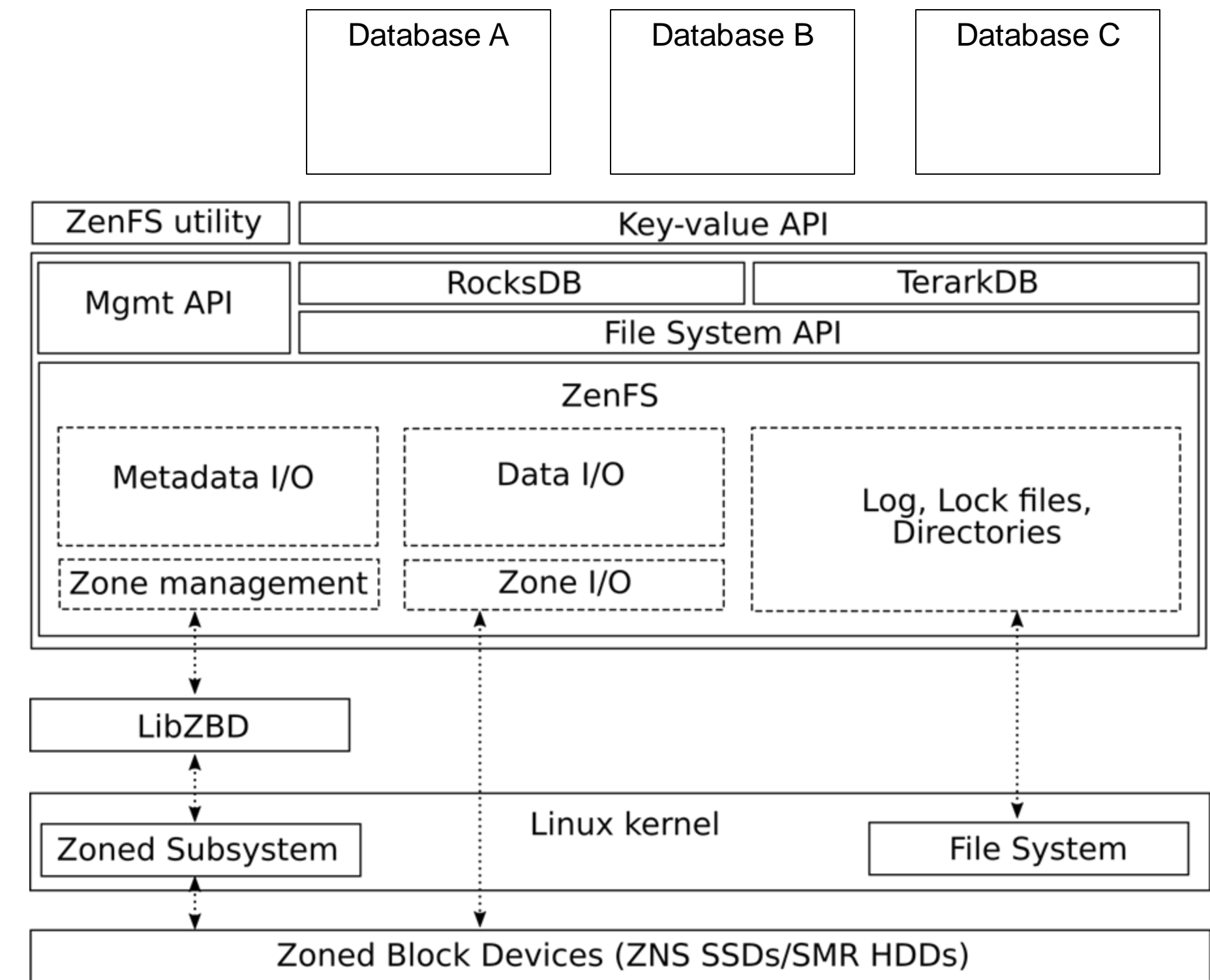


Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

ZenFS

- ❑ Not a POSIX file system
- ❑ Built for RocksDB/TerarkDB
- ❑ Separates files into zones
- ❑ Uses write life time hints for colocation when needed



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

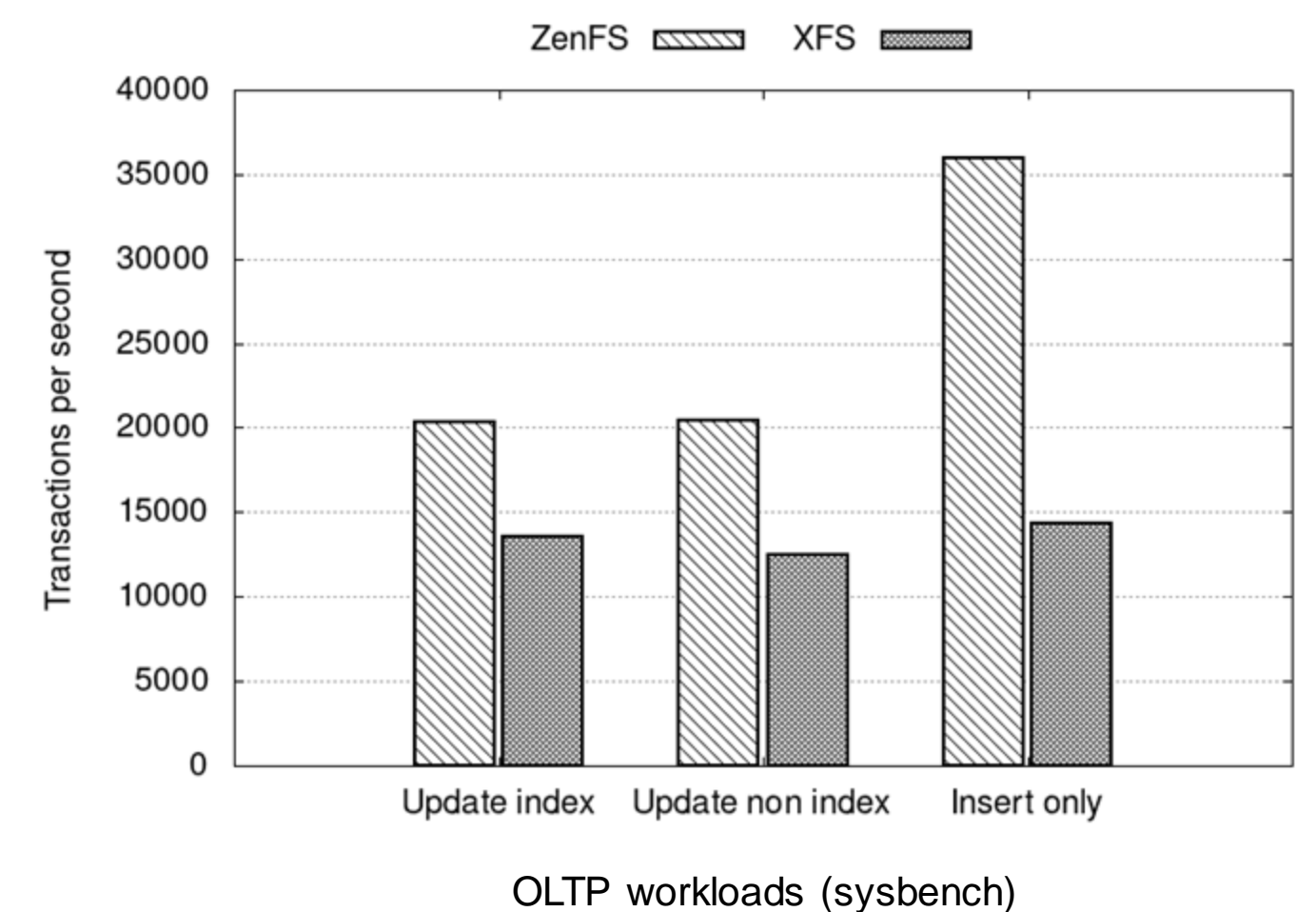
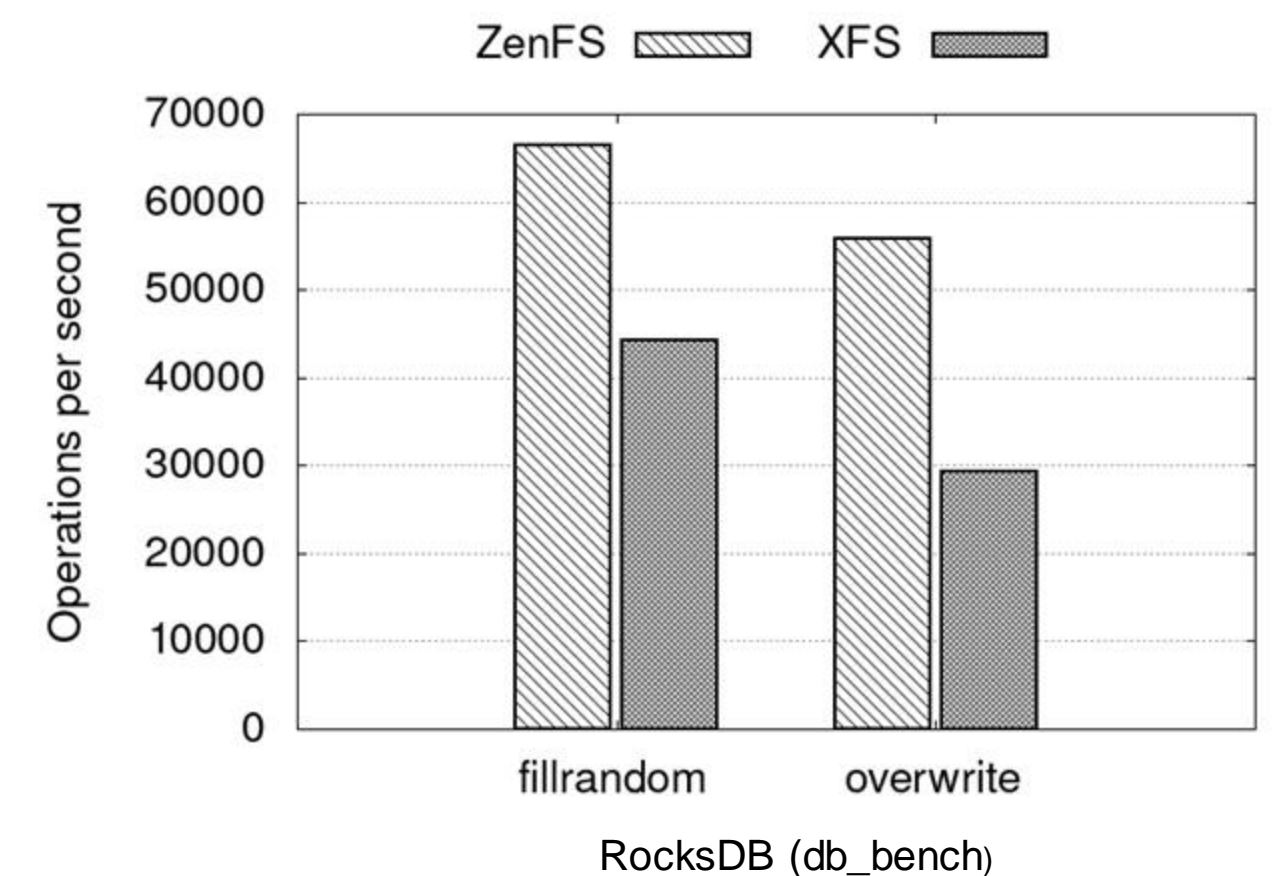
ZenFS performance

- ❑ Outperforms XFS for write and mixed workloads
- ❑ Improves both throughput and QoS
- ❑ Space amplification on par with XFS



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022





Lessons learned

- **file size = ~ zone capacity**
 - No GC needed, great performance, OK space amp
- **file size << zone capacity**
 - Tighter app integration required (done for TerarkDB)
 - Good performance, small amount of GC needed



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022



A decorative graphic of a green pipe network with various fittings, valves, and elbows, framing the central text.

ZenFS limitations

- ❑ Limited to RocksDB/TerarkDB
- ❑ Not a general purpose file system
- ❑ Can we bring some of the lessons learned into
 - ❑ f2fs?
 - ❑ btrfs?
 - ❑ bcachefs?



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

A decorative graphic of a green pipe network with various fittings, valves, and elbows, framing the central text.

What can be improved?

- ❑ **Can we improve the FS block allocators to play nicely with ZNS friendly workloads to reduce GC?**
- ❑ Yes we can!
 - ❑ Separate files into different zones when writing
 - ❑ Use write life time hints to co-locate files



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

Suggested way forward

- ❑ **Start experimenting with block allocators**
- ❑ Data placement improvements translates to less GC
 - ❑ Better performance (throughput, QoS)
 - ❑ Small improvements can lead to big wins
- ❑ Not trivial to implement (but fun!)
 - ❑ Automated benchmarks available, easy to test



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

A decorative graphic of green pipes with valves and elbows, running vertically on the left side of the slide and curving at the top and bottom.

Links & further reading

ZenFS (benchmarks available in tests/)

<https://github.com/westerndigitalcorporation/zenfs>

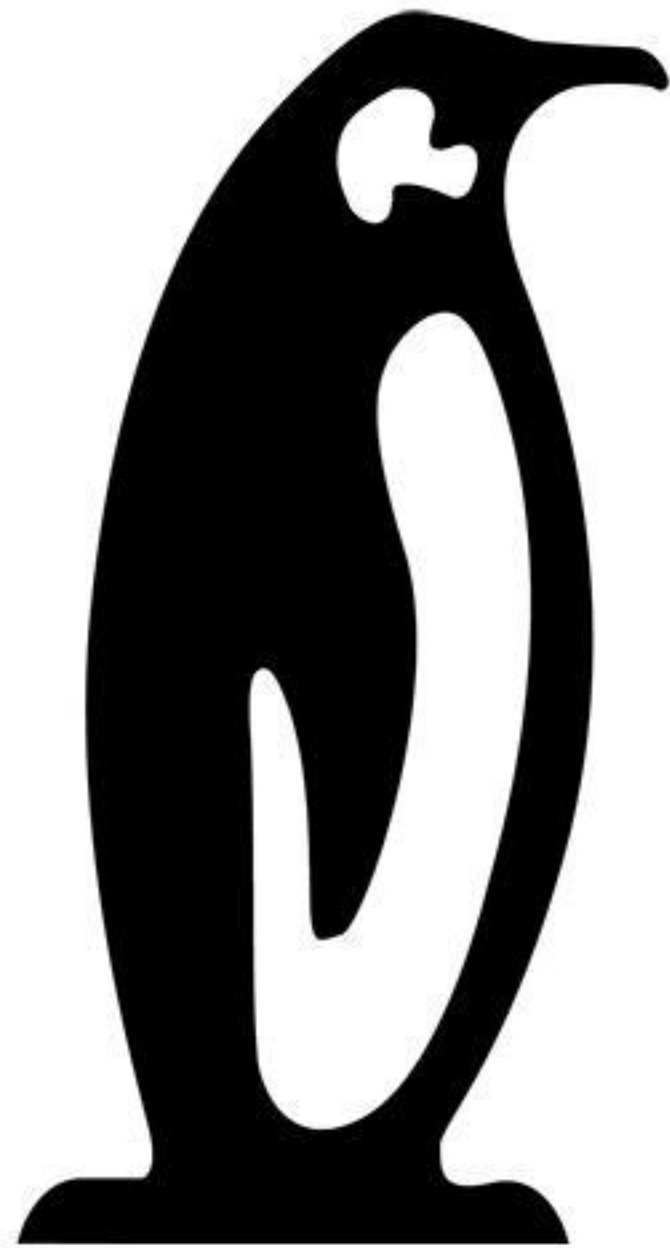
ZNS: Avoiding the Block Interface Tax for Flash-based SSDs

<https://www.usenix.org/system/files/atc21-bjorling.pdf>



Linux

Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022



Linux Plumbers Conference

Dublin, Ireland **September 12-14, 2022**