



OCP
SUMMIT

March 20-21
2018
San Jose, CA

OPEN. FOR BUSINESS.



Hybrid SMR Disk Drives: Who, What, Why, When, and Where?

Theodore Ts'o
Technical Lead
Google

OPEN. FOR BUSINESS.





Who am I?

- Linux Kernel Maintainer
 - Ext4 Maintainer
 - Served program committees for LSF/MM, Linux Plumbers Conference, Linux Kernel Summit, Usenix FAST and Usenix ATC
- Technical Lead at Google
 - Kernel Storage Team
 - Hybrid SMR Project
 - Co-author of the “Disks for Data Centers” white paper

OPEN. FOR BUSINESS.





What are SMR disk drives?

- Disk drives where the tracks partially overlap
 - Relies on the write head being wider than the read head
 - Like roof shingles
- Increases the capacity of disk drives by up to 25%
 - <https://sata-io.org/developers/sata-ecosystem/shingled-magnetic-recording-boosting-capacity-and-lowering-costs>
- Disk broken up in 256 MiB zones
 - Writes must be append-only within a zone
 - To re-record a zone, it must be erased
 - Much like an erase block on Flash storage



What are Hybrid SMR drives?

- Allows portion of the media to be converted from Conventional Magnetic Recording (CMR) to SMR
- Backwards compatible with Conventional HDD's
- Conversion can be done while the disk is serving
 - Non-destructive for unaffected regions of the disk
- Uses separate LBA ranges for CMR and SMR
 - Conversion brings CMR sectors off-line and SMR sectors on-line (or vice versa)

OPEN. FOR BUSINESS.





Why Hybrid SMR drives?

- Managing hot and cold data in a cluster file system
 - PDSW Keynote in 2017: “Colossus: Cluster-Level Storage @ Google”
 - <http://www.pdsw.org/keynote.shtml>
 - To use all of the bytes and IOPS of each disk
 - Spread hot data across all of the disks in the cluster file system
 - Use the rest for cold data

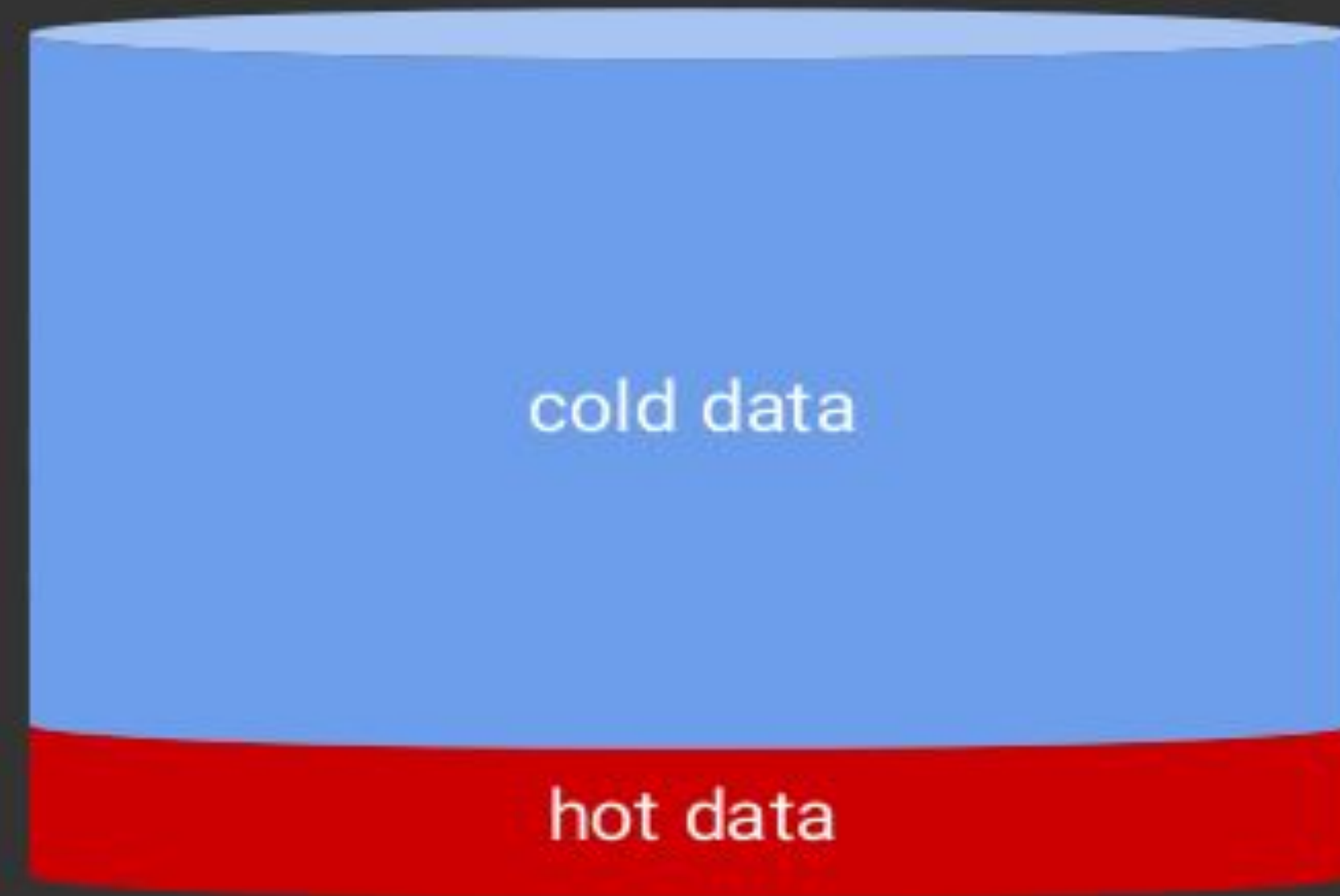
OPEN. FOR BUSINESS.



What we want



small disk



big disk

Equal amounts of hot data (spindle is busy)
Rest of disk filled with cold data (disks are full)

OPEN. FOR BUSINESS.



Why Hybrid SMR drives?

- Managing hot and cold data in a cluster file system
 - PDSW Keynote in 2017: “Colossus: Cluster-Level Storage @ Google”
 - <http://www.pdsw.org/keynote.shtml>
 - To use all of the bytes and IOPS of each disk
 - Spread hot data across all of the disks in the cluster file system
 - Use the rest for cold data
- From “Disks for Data Center” paper
 - <https://research.google.com/pubs/pub44830.html>
 - SMR friendly data tend to be cold
 - Disk IOPS is a perishable resource -- “use it or lose it”
 - To avoid stranding IOPS, store SMR and CMR data on a Hybrid SMR disk



When?

- As soon as possible. :-)
- Both Seagate and WDC have announced that they are working on Hybrid SMR
 - <https://blog.seagate.com/intelligent/new-flex-dynamic-recording-method-redefines-data-center-hard-drive>
 - <https://blog.westerndigital.com/dynamic-hybrid-smr>
- T13 held a Study Group Meeting on Hybrid SMR on March 2nd with another scheduled on March 26th

OPEN. FOR BUSINESS.





Where?

- T10/T13 standards committee
 - Device Interface standardization only
- OCP's Cloud-HDD Storage community
 - Cloud Product Requirements specification
 - Similar division of labor as in the Fast-Fail Read proposal
 - Work on Hybrid SMR drives slated to begin 1-2 months from now.
 - New Joint Development Agreement coming soon
 - To be drafted by the OCP Leadership Team
 - Please consider signing it and joining this activity!

OPEN. FOR BUSINESS.





What Else?

- At the 1st Study Group Meeting it was clear there was interest in converting between CMR, SMR, and HSMR modes of operation
 - From HDD vendors: SKU reduction
 - From at least some customers: being able to use a single pool of drives that can be used for traditional 100% CMR and 99.9% Host-Managed SMR drives
- This will be on the agenda for next week's study group meeting in Milpitas

OPEN. FOR BUSINESS.





Conclusion

- Hybrid SMR represents an new opportunity to use SMR technology
 - Takes advantage of the unique characteristics of hyperscale cluster file systems
 - Will help lower the cost of cloud storage for customers
- OCP will accelerate the unification of the product requirements of Hybrid SMR for Cloud.

OPEN. FOR BUSINESS.





OCP SUMMIT