OPEN
Compute Project

OCP U.S. SUMMIT 2017

Santa Clara, CA

Cloud-Optimized HDD Standardization Process

Lawrence Ying, Google Inc.

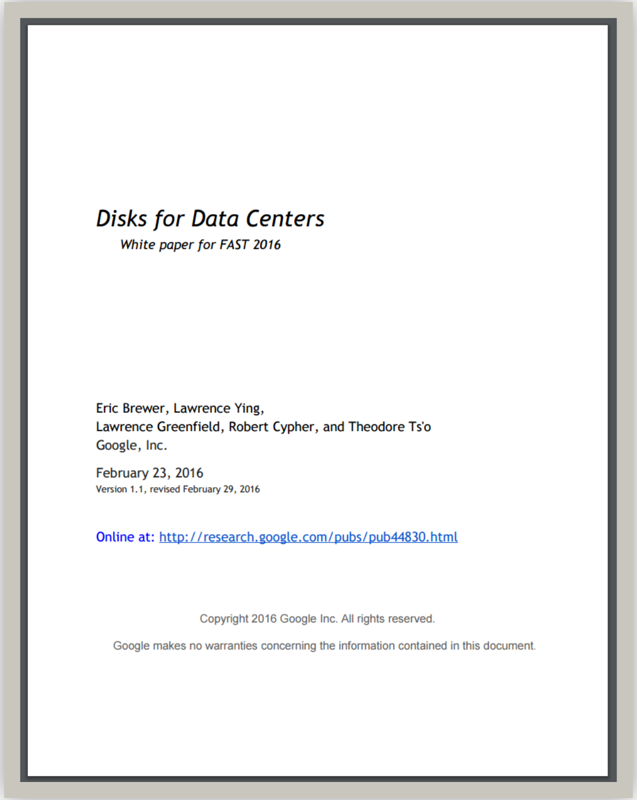
Michael McGrath, Microsoft Corp.

Jason Adrian, Facebook Inc.

A brief history…

Disks for Data Center white paper <*research.google.com/pubs/pub44830.html*>

- Presented by Google in both 2016 FAST and 2016 OCP Summit



Disks for Data Centers

White paper for FAST 2016

Eric Brewer, Lawrence Ying,
Lawrence Greenfield, Robert Cypher, and Theodore Ts'o
Google, Inc.

February 23, 2016
Version 1.1, revised February 29, 2016

Online at: http://research.google.com/pubs/pub44830.html

Copyright 2016 Google Inc. All rights reserved.

Google makes no warranties concerning the information contained in this document.

Jason

A brief history…

OCP Storage Call follow-up

- Microsoft and Facebook acknowledged that they also have many similar ideas to those published by Google. Examples include:
  - Flexible disk capacities and error rate trade offs
  - Host managed (or aware) advanced queueing and caching
  - Alternative form factors and parallel (multi) accesses
- With support from the OCP Storage Lead, Google, Microsoft and Facebook have worked together through the OCP collaboration principles to set the foundation and process that can accelerate the implementation and adoption of these ideas for everyone.

Michael

Proposal: A new process

- **Goal:** A new OCP standardization process to facilitate consensus for Cloud Storage around a set of use cases and associated interfaces, in order to accelerate technology development and augment existing standards bodies (T10, T13, SATA-IO, etc)
- **Scope:** Scale-out storage deployments with >10,000 HDDs
- **Status:** Targeting submission to the OCP Incubation Committee for approval within the next 1-2 months

Example:

NCQ “Prio” – What does the Prio bit mean in scenario X, Y, and Z?

Michael

Process in a nutshell

Propose within OCP

- problem statement
- scope and usage
- initial spec draft
- plan and schedule

Iterate within OCP

- socialize proposal
- iterate the details
- gain consensus
- OCP standardization (interfaces and test cases)

Enable & Standardize

- OCP Accepted devices become available
- T10/T13 standardization work begins (if applicable)

For more info, please see: *http://goo.gl/O08iJl*

Fast-fail read: Process trial-run

If the process proposal is approved within OCP

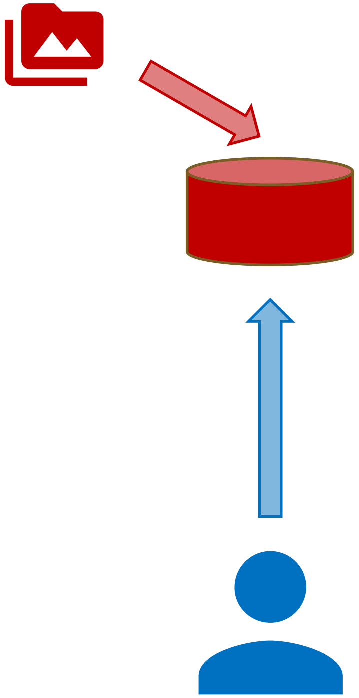
... would like to test it with the "**fast-fail read**" proposal.

Lawrence

Fast-fail read: Process trial-run

Problem Statement :

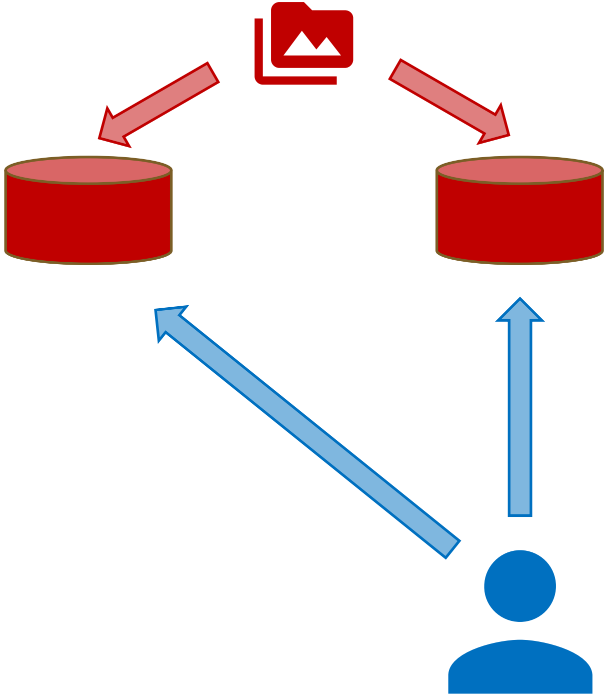
- HDD can ***sometimes*** be slow to read
  (Ex/ 500ms read latency at 99.9%tile)



Fast-fail read: Process trial-run

Problem Statement :

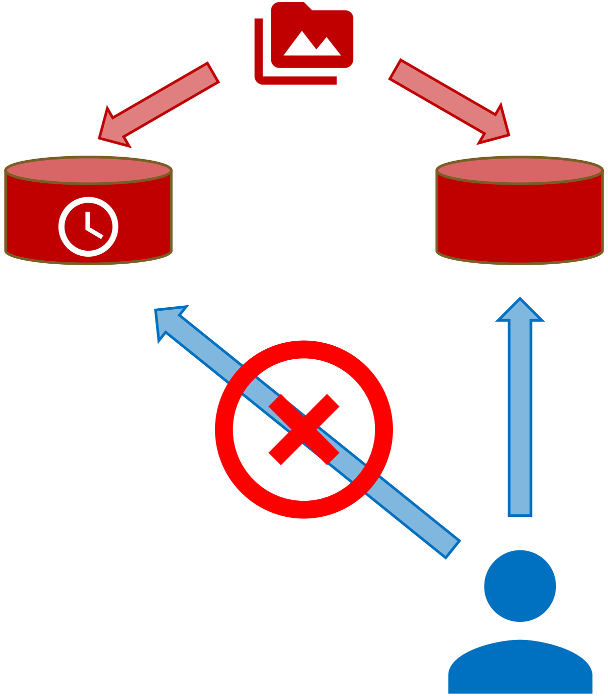
- HDD can ***sometimes*** be slow to read (Ex/ 500ms read latency at 99.9%tile)
- Data is stored on >1 HDD in data center



Fast-fail read: Process trial-run

Problem Statement :

- HDD can *sometimes* be slow to read (Ex/ 500ms read latency at 99.9%tile)
- Data is stored on >1 HDD in data center
- When one HDD is slow to read, we can just read from another HDD instead

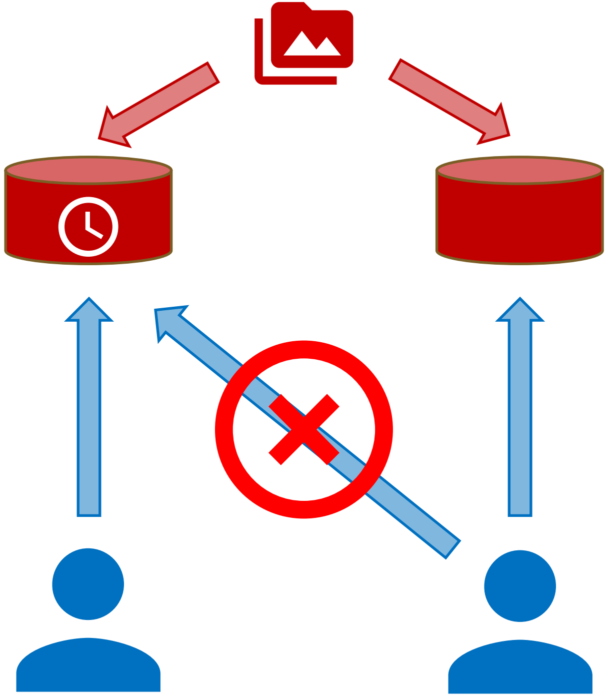


Lawrence

Fast-fail read: Process trial-run

Problem Statement :

- HDD can *sometimes* be slow to read (Ex/ 500ms read latency at 99.9%tile)
- Data is stored on >1 HDD in data center
- When one HDD is slow to read, we can just read from another HDD instead
- When this happens, would prefer the first HDD to abandon the read request (so it's "freed up" to do something else)



Fast-fail read: More details

Proposed interface needs:

- Two policies for reads: (1) fast-fail read, and (2) regular read

Out of scope: (some examples)

- Advanced queueing and caching management
- Advanced host management of disk background activities
- Advanced logging or health monitoring

For more info, please see: http://goo.gl/ZaeMiy

Q & A

OPEN
Compute Project