Storage

G54CCS – Lecture 6

Richard Mortier

http://www.cs.nott.ac.uk/~rmm/teaching/2011-g54ccs/

Overview

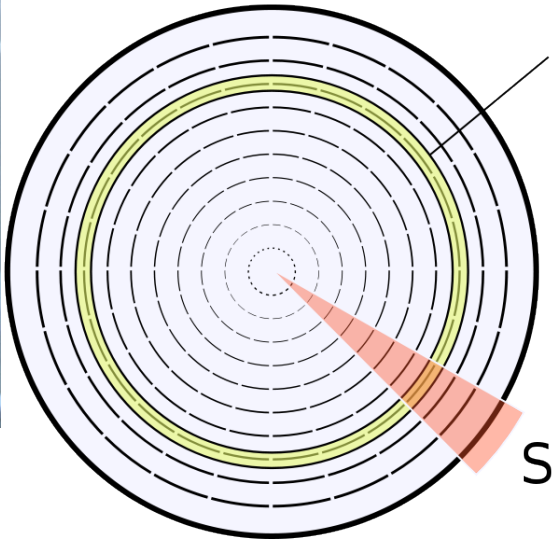
- What is Storage?
- Managing Many Disks
- Tradeoffs

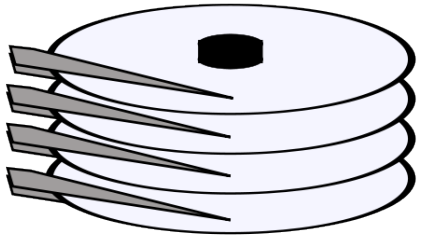
Overview

- What is Storage?
  - Disks
  - Capacity growth
  - Demand growth
- Managing Many Disks
- Tradeoffs

What is *Storage*?

- How you maintain data after loss of power
  - Used to involve a wide range of media
  - Now usually focus on (magnetic) hard disks
  - Surge of interest in solid-state devices (SSDs)
- Disk metaphor led to consideration of *blocks*
  - Derive from disk geometry
  - Fixed-size chunks of addressable storage





Heads

8 Heads,
4 Platters

What is *Storage*?

- Used to maintain data after loss of power
  - Once involved a wide range of media
  - Now usually focus on (magnetic) hard disks
  - Recent surge of interest in solid-state devices (SSDs)
- Disk metaphor led to consideration of *blocks*
  - Derive from disk geometry
  - Fixed-size chunks of addressable storage
- The most basic abstraction offered
  - E.g., Amazon Elastic Block Store (EBS)

Many Disks

- Disk capacity has grown dramatically
  - My first PC had 40MB HDD
  - Have 1TB USB HDD on my desk as cheap backup


WD
Western Digital

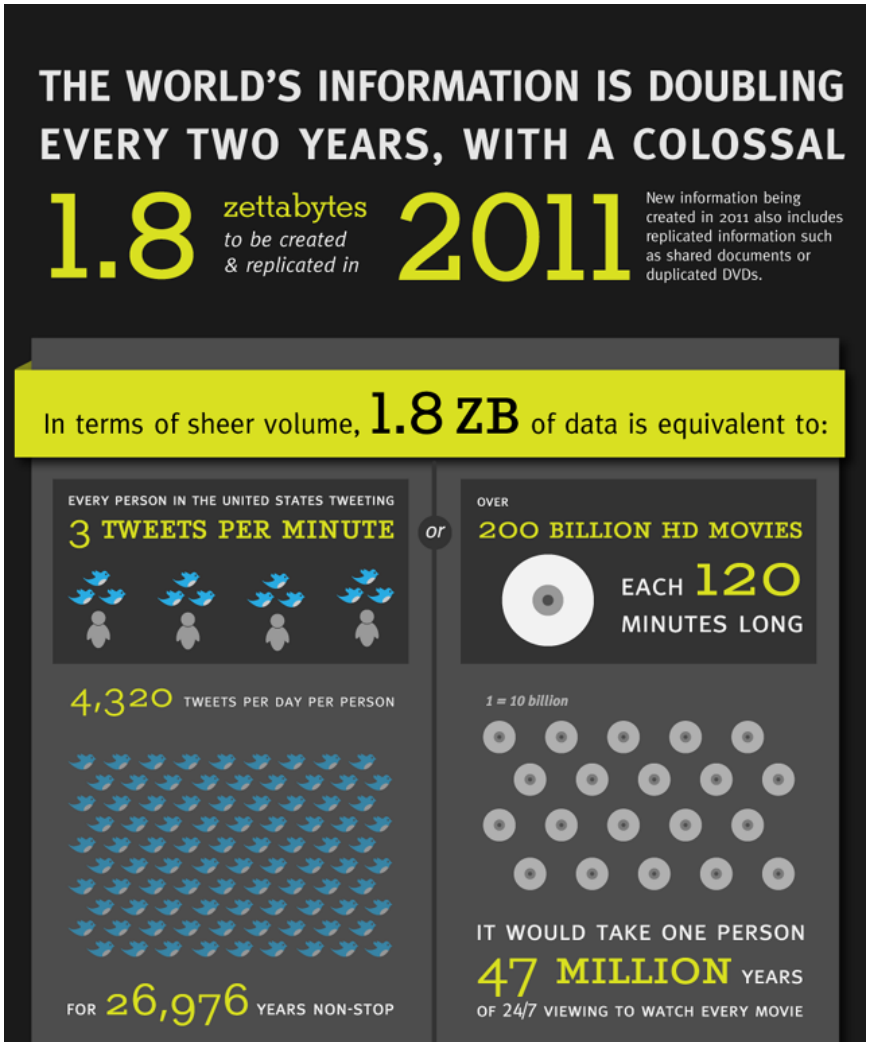
Many Disks

- Disk capacity has grown dramatically
  - My first PC had 40MB HDD
  - Have 1TB USB HDD on my desk as cheap backup
- Not just capacity
  - HDD capacity: 3.75MB to > 1TB (> 270,000:1)
  - Size: 87.9 cub. ft. to 0.002 cub. ft. (40,000:1)
  - Price: $15k/MB to < $0.0001/MB (> 1.5M:1)
  - Latency: > 0.1s to < 0.004s (> 40:1)

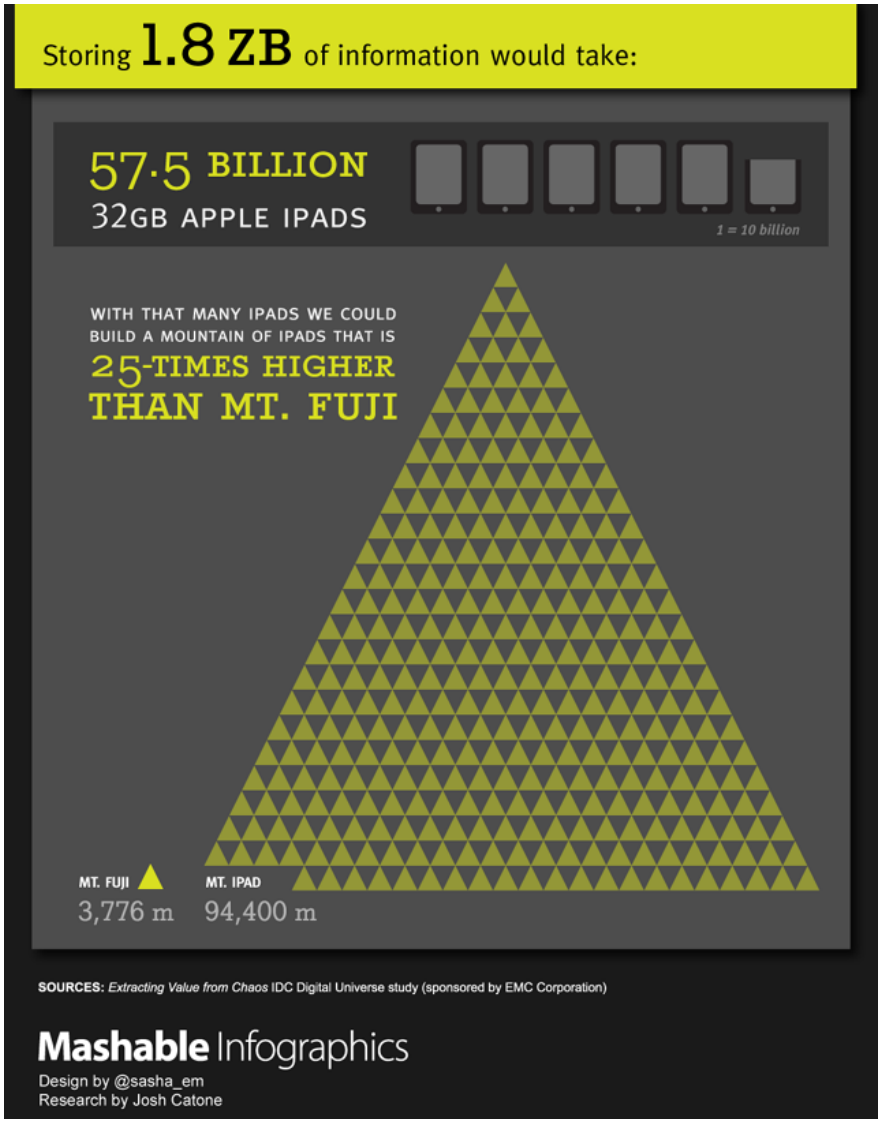
Many Disks

- Disk capacity has grown dramatically
  - My first PC had 40MB HDD
  - Have 1TB USB HDD on my desk as cheap backup
- Not just capacity
  - HDD capacity: 3.75MB to > 1TB (> 270,000:1)
  - Size: 87.9 cub. ft. to 0.002 cub. ft. (40,000:1)
  - Price: $15k/MB to < $0.0001/MB (> 1.5M:1)
  - Latency: > 0.1s to < 0.004s (> 40:1)
- So has data

1 ZB = 1000 EB = 1 Billion TB



1 ZB = 1000 EB = 1 Billion TB



Overview

- What is Storage?
- Managing Many Disks
  - Higher-layer abstractions
  - SQL vs. NoSQL
- Tradeoffs

Many Disks

- Disk capacity has grown dramatically
  - My first PC had 40MB HDD
  - Have 1TB USB HDD on my desk as cheap backup
  - So has data
- Can be organized in different ways
  - Single Large Expensive Disk (SLED)
  - Redundant Array of Inexpensive Disks (RAID)
  - Just a Bunch of Disks (JBOD)
- JBOD is the *de facto* standard in the cloud
  - Let the software do the hard work at scale

Higher Layer Abstractions

- Blobs
  - Sequence of bytes, no exposed structure
  - E.g., Amazon S3, Google Cloud Storage, Azure BLOB Service
- Object-Relational Mappings
  - Often really a key-value store
  - E.g., Google App Engine, Amazon Simple DB, Azure Table Service
- Relational Tables
  - Expose relations within data (rows, columns)
  - E.g., Amazon Relational DB Service (RDS), SQL Azure

SQL *vs*. NoSQL

- Not about SQL *per se*, but the structure of the data store
- Relational store
  - More structure, availability of indexes, good for SQL
  - More design required by developers, schema migration issues
  - Column store *vs*. Row store – Map-Reduce
- Key-value store
  - Commonly used, easy to *shard*, easy to think about, often SQL-like
  - Lack of rich indexing can cause performance cliffs
- Document store
  - Little structure, no schema issues, very flexible
  - Cf. Key-value store
- Performance at scale claims are *highly* contested
  - E.g., MongoDB war stories http://bit.ly/ugrtmx

Overview

- What is Storage?
- Managing Many Disks
- Tradeoffs
  - Price
  - Problems of scale
  - Example of scale: Facebook Messages

Tradeoffs

- Managed *vs*. Unmanaged
  - Price for managed is much higher
- Local *vs*. remote
  - Care about latency and bandwidth
  - In the cloud, have network overheads and costs
- Read *vs*. write performance
  - Need to know your workload
  - In the cloud, multi-tenancy gives high variability

Service	Resource	Price
EC2	Local disk	Free, but no persistence across reboots
EBS	Block store	$0.11 / GB-month
	IO Requests	$0.11 / 1M
S3	Data	$0.14 / GB-month
	Reduced Redundancy	$0.093 / GB-month
	GET Requests (DELETEs free)	$0.01 / 10,000
	PUT,POST,COPY,LIST Reqs	$0.01 / 1,000
Simple DB	Machine hour (normalized)	$0.154 / hour
	Storage (+ 45B / item)	$0.275 / GB-month
RDS	Instance, Storage	$0.12 – $2.96 / hour, $0.11 / GB-month
	Multi Availability Zone	$0.24 – $5.92 / hour, $0.22 / GB-month
	IO Requests	$0.11 / 1M
	Data transfer OUT (IN is free)	$0.00 – $0.05 / GB OUT up to 500 TB

Scale Creates Problems

- Concurrency
  - Perfect consistency is *hard*
  - Eventual consistency is in vogue
  - http://bit.ly/qSIpxP
- Management
  - Large numbers and MTTF suggests disks fail *all the time*
  - …but we build software for the *common case*
  - …so can be (counter intuitively) *more reliable*
- Complexity
  - Demonstrates distributed systems problems
  - E.g., Etsy http://bit.ly/rhdpGJ

http://bit.ly/rt2csc

Facebook Messages

- Small messages (HDFS/Hbase)
  vs. Photo store (Haystack)
  - 6B+ messages/day
  - 75B+ R+W ops/day (1.5M/sec at peak)
  - 55% R vs. 45% W – each W inserts avg. 16 records
  - 2PB+ online data in HBase (6PB+ with replication)
    - Compressed: message data, metadata, search index
    - 250 TB / month