Computer Science 146
Computer Architecture

Fall 2019

Harvard University

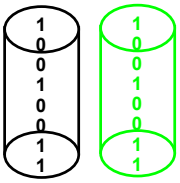
Instructor: Prof. David Brooks
dbrooks@eecs.harvard.edu

Lecture 23: Clusters and Wrapup

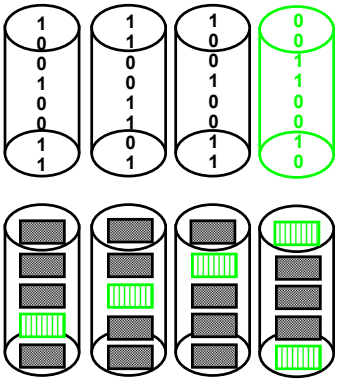
Summary: RAID Techniques

Goal: Performance, popularity due to reliability of storage

- ***Disk Mirroring, Shadowing (RAID 1)***
  - **Each disk is fully duplicated onto its "shadow"**
  - **Logical write = two physical writes**
  - **100% capacity overhead**



- ***Parity Data Bandwidth Array (RAID 3)***
  - **Parity computed horizontally**
  - **Logically a single high data bw disk**
- ***High I/O Rate Parity Array (RAID 5)***
  - **Interleaved parity blocks**
  - **Independent reads and writes**
  - **Logical write = 2 reads + 2 writes**



I/O System Example

- Given
  - 500 MIPS CPU
  - 16B wide, 100 ns memory system
  - 10000 instrs per I/O
  - 16KB per I/O
  - 200 MB/s I/O bus, with room for 20 SCSI-2 controllers
  - SCSI-2 strings (buses) – 20MB/s with 15 disks per bus
  - SCSI-2 1ms overhead per I/O
  - 7200 RPM (120 RPS), 8ms avg seek, 6MB/s transfer disks
  - 200 GB total storage
- Q: Choose 2GB or 8GB disks for maximum IOPS?
  - How to arrange disks and controllers?

I/O System Example (cont'd)

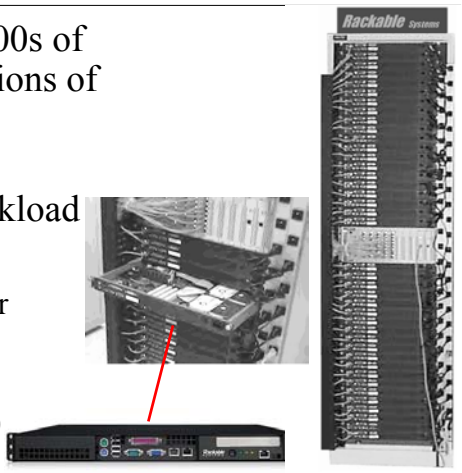
- Step 1: Calculate CPU, memory, I/O bus peak IOPS
  - CPU: 500 MIPS/(10000 instructions/IO) = *50000 IOPS*
  - Memory: (16-bytes / 100ns) / 16KB = *10000 IOPS*
  - I/O bus: (200 MB/s) 16KB = *12500 IOPS*
  - Memory bus (10000 IOPS) is the bottleneck
- Step 2: Calculate Disk IOPS
  - T_{disk} = 8ms + 0.5/120 RPS + 16KB/(6MB/s) = 15 ms
  - Disk: 1 / 15ms = *67 IOPS*
  - 8GB Disks => need 25 => 25 * 67 IOPS = *1675 IOPS*
  - 2GB Disks => need 100 => 100 * 67 IOPS = *6700 IOPS*
  - 100 2GB disks (6700 IOPS) are new bottleneck
- Answer: 100 2GB disks!

I/O System Example (cont'd)

- Step 3: Calculate SCSI-2 controller peak IOPS
  - T_{SCSI-2} = 1ms + 16KB/(20MB/s) = 1.8ms
  - SCSI-2: 1/1.8ms = *556 IOPS*
- Step 4: How many disks per controller?
  - 556 IOPS/ 67 IOPS = *8 disks per controller*
- Step 5: How many controllers?
  - 100 disks/ 8 disks/controller = 13 controllers
- Answer: 13 controllers, 8-disks each

Google Cluster Architecture

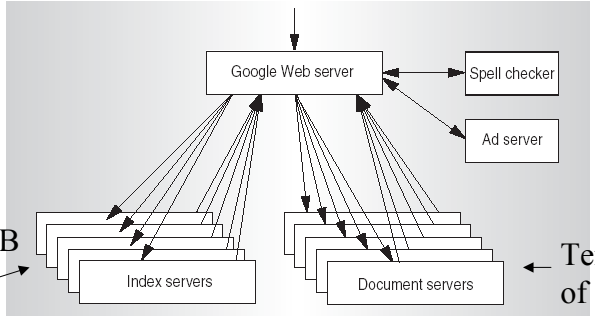
- 1 Google Query reads 100s of MBs of data, 10s of Billions of CPU cycles
- Massive parallelism, throughput oriented workload
- Key Metrics
  - Energy Efficiency (power and cooling issues)
  - Price-Performance Ratio (10,000-100,000+ CPUs)



Price-Performance Ratio

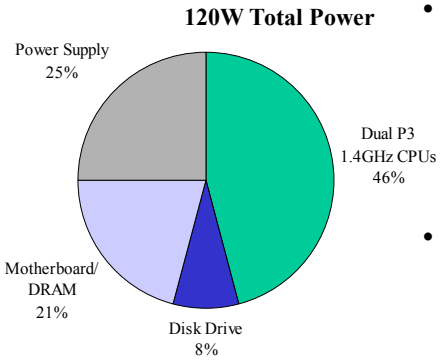
- Reliable computing via unreliable commodity PCs
  - Reliability provided with software redundancy
  - Clusters distributed geographically (catastrophic fault tolerance)
  - Replicated services and data across many machines
  - Google stores dozens of copies of the Web across its clusters

Distributed Query Processing



- Index and Document Servers provide
  - Redundancy, ability to maintain, update database HW/SW
  - Lots of easy to exploit parallelism
- Index "shards" allow queries to be parallelized into machine pools

Energy Efficiency



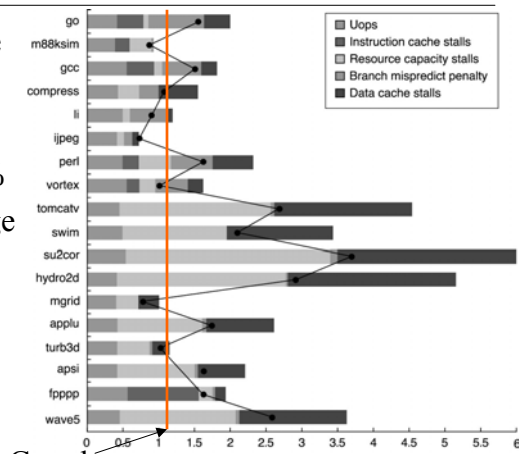
- 120W per rack, 88 racks per cabinet
  - ~10KW per 25ft^2 cabinet
  - 400 W/ft^2 power density
    - Up to 700 W/ft^2 with newer CPUs!
  - Typical data centers only support 70-150 W/ft^2
- Energy Bill?
  - 10 MW-h per month (including cooling, about 25-30%)
  - @15c/KWH, $1,500 per month

How many machines does Google have?

- Not publicly disclosed… (paper says 15,000+)
- 4/30/04 SEC filing says 250 Million spent on hardware
  - From the paper, 278K per 88 dual-CPU rack
  - ~900 racks, 80,000 machines

Application Characteristics

- Measurements for the index server
  - For a P3 CPI of 1.1, Branch mispredict 5%
  - Actually about average
- Still SMT/CMP do make sense (30% SMT performance benefit)

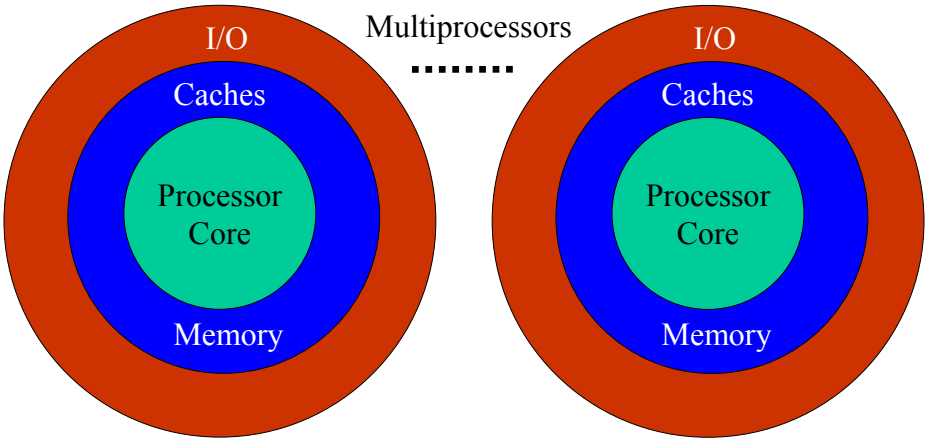


Google

What will the Final cover?

- The final will be:
  - Comprehensive (full course)
    - With emphasis on post-midterm work
    - Pre-midterm work that was not covered on the midterm
  - Similar in flavor to the midterm
    - A bit longer
    - But you will have much more time
    - Provide as much explanation as you can for partial credits

Big Picture of Computer Architecture



Core CPU Architecture

- Pipelining
- Superscalar
- Dynamic Execution/OOO
- Register Renaming
- Branch Prediction
- How to maintain precise interrupts?
  - Reorder buffers

Memory Hierarchy

- Caches
  - Cache organizations
  - Cache performance optimizations
- Physical and Virtual Memory
  - Address translation, TLBs
- Main Memory

Multiprocessors and I/O

- Cache Coherency Protocols
- Other difficulties with multi's
- Simultaneous Multithreading

- Basics of Disks and I/O
- RAID

Questions?

- Final is Tuesday, May 25th at 2:15
- Schedule review for Friday May 21st?