Lecture 3 Notes – Topic: *Benchmarks*

- What do you want in a benchmark?
  - benchmarks must be representative of actual workloads
  - first few computers were benchmarked based on how fast they could add/multiply
    - Main use of computers originally
- In the 70's, first time people though of actual programs
  - matrix multiply early first program benchmark
  - Sorting
  - Stone Age synthetic benchmarks
    - Whetstone (fp)
    - Dhrystone (int)
    - 25 years ago papers begging coming out about how bad they were
      - not representative of actual workloads
    - Synthetic - artificial constructs, Fortran, Ada, Pascal
    - Dhrystone - has a lot of C. character data type
  - Linpack
    - Linear Algebra Package
    - not synthetic
    - actual package scientist use to solve linear algebra problems
  - Lawrence Livermore Loops (LLLs)
    - LL Labs.
    - Pulled out 14 different loops from their programs
    - Then added 8 more
- Paper - An Overview of Common Benchmarks (IEEE Computer magazine, Broader readership, not journal)
  - Called old benchmarks "Stone Age Benchmarks"
  - Wet and Dry because rubbing knife by rubbing against stone
    - Program checked for quality
  - Sieve of Ratosthenes
    - Prime Number algorithm
    - Benchmark used for many years
    - Great as a power benchmark
      - the only natural program that takes the most power was prime number generation
  - EDN Benchmarks
    - Electronic Design Use (magazine) 80's
    - Magazine came up with their own benchmark, from the point of view of journalism
    - One of the first
  - 1989, SPEC released first suit
    - This paper shows what was done before SPEC

  - Table A. Statement Distribution in percentages
    - Dhrystone just tries to have a certain percentage of each operation
  - Table B. operator distribution in percentages
    - Linpack is almost always 1 add/1 multi
    - C features in Dhrystone only
  - Table C. Operand data-type distribution in percentages.
    - Distribution of structures that are used by benchmarks
  - Table 1. Procedure profile for Whetstone.
    - optimized math library were and still good way to increase performance.
  - Table 2. Procedure profile for Linpack
    - SAXPY - (single precision) A * X + Y
    - DAXPY - (double precision) “”
    - one loop y[i] = y[i] + a*x[i]
    - 77.1 percent of operations
    - Real program
- Stone Age Benchmarks
  - People use to use global variables a lot
    - Principles now use local variables
  - No I-Cache because of such little code
  - People run around rules by using compiler that would eliminate code.
    - Tried to add rules and print statements so people didn’t
    - some compilers didn’t compile if operand not used
  - 1988 - People complaining about competitors cheating
    - So SPEC and TPC founded in summer of 1988
    - How to stop cheating
  - SPEC - desktops use this for benchmarking
  - TPC - transaction processing council, for large data centers/servers
- Benchmarks
  - CPU benchmarks
    - Centered on pc’s microprocessor
    - (SPEC
    - SPLASH - from Stanford, parallel intensive
    - STREAMS - popular for streaming workloads, 90’s. Worked at UT TACK John McCalpin
  - Transactions Processing Benchmarks
    - Useful for big company (bank)
    - Transaction like bank transactions, ordering, etc.
    - Computer could handle more or less “transactions” = more money
  - Embedded and media Benchmarks
    - For small devices with computers, cell phones
    - Majority of processing these days, consumer market
  - HPC (Supercomputing) Benchmarks
    - Scientific computing
  - Cloud/Big Data Benchmarks

        - New
        - Can fall under TPB
        - "Big Data"
        - Need to analyze large data sets
        - State of benchmarking not as mature, just coming out
    - Web Server Benchmarks
        - SPEC web benchmark
            - But recently retired
        - These benchmarks popular because big thing you do
    - Browser Benchmark
        - client side web benchmark
        - javascript
        - load times
        - processing
    - PC Benchmarks
        - make for drawing lines, small operation
        - Office related tasks
        - Photoshop
        - Common things
    - Parallel Processing Benchmarks (NPB)
        - SPLASH
        - Nas Parallel Benchmark (NPB)
            - has a lot of communication between threads
        - very parallel benchmarks
        - Developed by NASA
        - NP (embarrassing parallel)
            - purely vectorized program, no dependency, fully parallelizable
        - Character: Vectorizable
    - Java Benchmarks (SPECjvm98, SPECJjvm2008)
        - newer, popular
        - huge part of computing world
        - phones, cloud
    - GPU Benchmarks
        - Application vs. processing
- HINT Benchmark
    - Will be reading soon
    - Scientific computing domain
    - very different then above benchmarks
- SPEC CPU Benchmarks
    - Most papers in last 25 years about SPEc
    - SPEC CPU 1989 (10programs, 4c INT, 6Fortran FP)
        - equintot (equation to truth table) - workstations first coming out
            - for first automated tools which came out

        - place & route
        - design automation
    - SPEC CPU 1992 (20, 6C int, 14 fp (2c, 12f))
        - most architecture papers talk about integer side
        - integer program are more random with jumps and loops
        - floating point more predictable program flow usually because of application
        - math more interesting
    - 1995 (18, 8C int, 10 F FP)
    - 2000 (26, 12 int (11c, 1C++), 14 fP (10F, 4C))
    - All code in spec is real program, submitted outside of SPEC and analyzed
        - code must compile on most compilers
    - 2006 (29, 12int, 17 fp (c, c++, F)
    - next iteration: 14/15
- TPC Benchmarks
    - Database and Transactions Benchmarks
    - 8 different companies came together to form TPC
        - sun, mips, hp, etc.
    - for transactional workloads
    - cloud workloads, for big players
        - not as wide as SPEC
    - Dr. Jim Gray - guru of database systems
- Jim Gray
    - 1988 at Tandem Computer - promoted by paper in '85
        - what happens when you are buying in credit cards
        - made benchmark for these loads
        - Wrote TPC-A
    - TPC is synthetic (cannot have real world input)
    - Business side, tried to understand stores, prices, stocking shelves, timing, etc.
    - Started Microsoft Research group
    - went sailing after mothers death, never found
- Transaction Categories
    - OLTP (OnLine Transaction Processing) - looking on amazon, checking bank
        - lots of transactions from multiple sources
        - each transaction small
    - DSS (decision support system) - looking at a lot of data, analysis to make strategic decisionss
        - data from OLTP side periodically filtered to DSS side
        - runs for a long time trying to figure out one thing
        - "data-mining"
    - OLAP (On Line Analytics Processing) - combination of both OLTP and DSS
        - book suggestions
        - google search personalized

- TPC Benchmarks
  - TPC-C (OLTP)
  - TPC-H (DSS)
  - TPC-E (OLTP)
  - TPC-DS (DSS)
  - TPC-W (OLAP) - release 2003, retired 2004, inadequacies
  - TPC-VMS - run rules for how to run other benchmarks on a virtual machine
  - TPC-Big Data
    - in the works