Performance Evaluation
9/23/2014

- CPU 2006 Working Set Size - Gove Paper
  - Working set size of SPEC 2006 programs
  - Darryl Gove from Oracle
    - Experimentalist, detailed craftsman
    - Author of Solaris Application Programming (book)
      - Performance counters and system tools
  - Metrics of memory usage
    - Spec people did not want the whole program to fit into caches
    - Virtual Size (VSZ)
      - Virtual memory allocated by the operating system
      - Address range
    - Resident Set Size (RSS)
      - Measure of how much physical memory is actually being used
      - RSS is less than or equal to VSZ
    - Working Set Size (WSS)
      - How much you are working on in a short interval or the recent past
    - Example
      - Application with footprint of 100M (RSS)
      - Application calls mmap to allocate space for a 1 GB file
      - VSZ is now 1.1 GB
      - Application only reads 100 M of the file
      - VSZ stays at 1.1 GB
      - RSS is now 200M
      - WSS is 100M (max of 200M) assuming application keeps iterating over only a tenth of the program
    - Working Set
      - Concept introduced in the 1960s (Denning 1968)
      - WSS is an estimate of the size of memory in use during application execution. It is usually estimated over a small interval.
      - Small physical memory is ok if the working set size can fit.
      - “The Working Set Model for Program Behavior” - Peter Denning
    - WSS estimated in Gove
      - Estimated at 64 byte block level
      - Array records if each block was touched or not during an interval
        - 1 billion memory operations chosen at the interval
          - Smaller interval leads to small WSS
          - If there is no locality, WSS means nothing

                - Requires temporal locality to give the model meaning
                - Large interval demonstrates if there is temporal locality or not
                - This should correspond to roughly 4 billion instructions executed
                - Must be large relative to the cache
            - Granularity of reference size (cache block vs. page size)
                - Depends on what you are trying to study
                - Gove's was more applicable to caches
            - Interval should depend on what footprint you are studying
                - Paging/Pages vs. Caches
        - Core Working Set Size (CWSS)
            - Number of blocks in a sample used in the previous sample
            - Can be considered a working set within a working size
            - Relevant when the cache can not hold the working set
- SPEC memory usage goals
    - In 2000 it was 200 MB
    - In 2006 it was 900 MB
    - RSS and VSZ very similiar for SPEC integer workloads
        - Nearly identical for SPEC INT 2006
    - Make sure your platform can handle the memory requirements for the benchmarks
    - RSS and VSZ for SPEC FP workloads exhibit some difference
    - Streaming data in leads to an RSS smaller than the VSZ
    - The WSS for integer workloads do not exhibit much change from SPEC INT 2000 and SPC INT 2006
    - Blocked algorithms such as for matrix multiplications enable better reuse of data to enable high performance
    - Blocking is used to reduce the WSS or increase locality
    - WSS can be used to measure the impact of algorithmic changes
    - SPEC FP 2006 WSS shows more separation than SPC INT, but still lots of similarity for WSS between 2000 and 2006
- Gove looked at WSS over time
    - deal2 had a very spikey WSS over time
    - Some programs had a near constant RSS over time
    - named had lots of fluctuations but the spikes were not as large
- Gove was done on a Sparc system using Spot and SHADE
- Always give references about the tools you are using to profile so the experiment can be replicated
- Standard deviation for WSS is helpful because it allows us to see how much the WSS changes
- Raw data is a good reference, even if graphs are easier to read

        - Can be used as a sanity check even if things like architecture or compiler are different
    - Lots of errors can be introduced when running large benchmarks
        - ALWAYS DO SANITY CHECKS