ELE 585 Scribe Notes

02 October 2017

Last time: Performance of Database Workloads on Shared-Memory Systems with Out-of-Order Processors
Benchmarks used are standardized. Very database-centric.
TPC = Transaction Processing Performance Council
TPC-B: Memory-intensive form of TPC-C?
TPC-C: Banking-style application
TPC-D: Very different application, much less used. Data-mining.

Trace-based simulation: wrong.

Alpha processors: no longer exist.

Results:
OLTP and DSS can improve by adding optimizations
No discussion of energy; No one cared about energy during the time of publication
Very little discussion of area; not a very constrained study

Stream Buffer:
Prefetches next block in instruction memory. If there is a miss in L1-I, first check stream buffer.
Energy analysis:
Performs extra fetches from memory, potentially increasing overall energy
If well utilized, the performance improvement can save overall energy

Always use plots that work well with greyscale printers!

Clearing the Clouds
Authors
Michael Ferdman -- Professor at Stoney Brook, possibly a senior student at CMU when Babak Falsafi transferred to EPFL
Babak Falsafi -- Professor at EPFL, previously CMU
EPFL: Swiss Federal Technical school of Lausanne. Part of the same institution as ETH Zurich.

What is this paper about?
Scale-out: In contrast to scale-up. To improve performance, it's more beneficial to add more machines rather than adding more resources to the same machine.

Today, both scale-up and scale-out are popular as shown in personal desktops and cloud service
Benchmarks released as CloudSuite. Initial reaction was negative due to difficulty in use. Later releases added scripting to make it more accessible.

SQL: Interface with databases providing queries to database
NoSQL: Big Table, Cassandra, widely used in industry
In-memory databases which you do simple queries against

MapReduce
Map: Apply a function to data in a dataset. Essentially tagging
Reduce: Take every element that starts with same key and coalesce them
Enables coarse-grained parallelism
Programming model
Often shows up in offline processing in datacenters.

Media Streaming
Definitely parallel on servers; some parallelism on clients also.
Media server which streams videos

SAT Solver
What boolean inputs satisfy a particular boolean expression.
Not exactly a "cloud" benchmark. Not likely very widespread in the cloud.

Web Frontend
Course-grained parallelism
Nginx -- no threads; state machine with non-blocking I/O. Return data on a per-flow basis
PHP Backend with Advanced PHP Cache

Web Search
Use Nutch and Lucene benchmarks
Good first-order approximation for PageRank
Serving side of web-search. Assuming index is already created.

Parallelism Comparison
Most applications are very parallel
MapReduce is probably slightly less parallel because of communications or imbalanced workload
SAT Solver probably less parallelism

PARSEC/Olden -- Parallel benchmark developed at Princeton

What are the contributions?
The benchmarks themselves
The in-depth analysis of the performance of these benchmarks
Results Section
Well-structured and well-written. Often authors also split out a discussion section from the data/results section.
This paper and OLTP/DSS paper disagree
This paper argues for smaller cores, less out-of-order features
This paper argues to remove LLC

Only run on one machine and configuration. More design space exploration would give stronger evidence to support the claims in the results section.
Arguments about having too much off-chip bandwidth and ineffectiveness of L2 cache were later found incorrect in their future papers.