Customizing Servers for Emerging Scale-out Workloads Using CloudSuite

Michael Ferdman, Almutaz Adileh, Onur Kočberber, Stavros Volos, Mohammad Alisafaei, Djordje Jevdjic, Cansu Kaynak, Adrian Daniel Popescu, Boris Grot
Anastasia Ailamaki, Babak Falsafi

écocloud
Cloud Server Efficiency

• Cloud computing is pervasive
  – User base growing exponentially
  – New services appearing daily

• Constant need for more servers
  – Many costs: HW, space, power
  – Want max server efficiency

• Modern servers not designed for scale-out apps
  – Cloud apps are scale-out, but server built for scale-up

How efficient are scale-up servers for scale-out apps?

© 2012 EPFL PARSA
Which Benchmarks to Use?

- Examined traditional benchmark suites
- Benchmarks designed for scale-up

Don’t represent scale-out applications
Key Scale-Out Characteristics

- Serve independent requests/tasks
- Operate on huge dataset split into shards
- Communicate infrequently
Introducing CloudSuite 1.0

Data Serving
Cassandra NoSQL

MapReduce
Machine learning on Hadoop

Media Streaming
Apple Quicktime Server

SAT Solver
Symbolic VM constraint solver

Web Frontend
Nginx, PHP server

Web Search
Apache Nutch
Micro-arch. Properties of CloudSuite

- Instruction working set exceeds L1 & L2 capacity
- Large data working set doesn’t fit on chip
- Low degree of data sharing
- Low ILP and MLP
- No floating-point operations

CloudSuite is different from traditional benchmarks
Conclusions

• Cloud services require efficiency
  – Architecture should match needs of scale-out workloads

• Scale-out apps are different than traditional apps
  – Both at software and micro-arch. level

• CloudSuite: A set of emerging scale-out apps
  – Supported and updated to reflect the state-of-the-art

Step toward efficient scale-out servers
Download CloudSuite 1.0
parsa.epfl.ch/cloudsuite

Coming
June 2012

CloudSuite Tutorial & Disk Images @ISCA 2012
parsa.epfl.ch/cloudsuite/isca12-tutorial