

Scale-out processors: Bridging the efficiency gap between servers and emerging cloud workloads

March 21 2012

Cloud computing has emerged as a dominant computing platform providing billions of users world-wide with online services. The software applications powering these services, commonly referred to as scale-out workloads and which include web search, social networking and business analytics, tend to be characterized by massive working sets, high degrees of parallelism, and real-time constraints – features that set them apart from desktop, parallel and traditional commercial server applications.

To support the growing popularity and continued expansion of cloud services, providers must overcome the physical space and power constraints that limit the growth of data centers. Problematically, the predominant processor micro-architecture is inherently inefficient for running these demanding scale-out workloads, which results in low compute density and poor trade-offs between performance and energy. Continuing the current trends for data production and analysis will further exacerbate these inefficiencies.

Improving the cloud's computational resources whilst operating within physical constraints requires server efficiency to be optimized in order to ensure that server hardware meets the needs of scale-out workloads.

To this end, the team of HiPEAC member Babak Falsafi, a Professor in the School of Computer and Communication Sciences at EPFL, the director of the EcoCloud research center at EPFL (founded to innovate

future energy-efficient and environmentally friendly cloud technologies), presented [Clearing the Clouds: A Study of Emerging Workloads on Modern Hardware](#), which received the best paper award as ASPLOS 2012. ASPLOS is a flagship international computer systems venue with a high citation index.

“While we have been studying and tuning conventional server workloads (such as transaction processing and decision support) on hardware for over a decade, we really wanted to see how emerging scale-out workloads in modern datacenters behave.” says Falsafi. “To our surprise, we found that much of a modern server processor’s hardware resources including the cores, caches and off-chip connectivity are overprovisioned when running scale-out workloads leading to huge inefficiencies.”

Mike Ferdman, a senior PhD student team member explains: “efficiently executing scale-out workloads requires optimizing the instruction-fetch path for up to a few megabytes of program instructions, reducing the core complexity while increasing core counts, and shrinking the capacity of on-die caches to reduce area and power overheads.”

Scale-out Processors

“The insights from the evaluation are now driving us to develop server processors tuned to the demands of scale-out [workloads](#)”, says Boris Grot a postdoctoral team member. “In a paper that will appear in the flagship computer architecture conference, ISCA, this year, our team proposes the Scale-Out Processor, a processor organization that unlike current industrial chip design trends does away with power-hungry cores and much of on-die cache capacity and network fabric to free area and power for a large number of simple cores built around a streamlined memory hierarchy.” Not only do these improvements lead to greater performance and efficiency at the level of each [processor](#) chip, they also

enable a net reduction in the total cost of ownership in datacenters.

Provided by EuroCloud

Citation: Scale-out processors: Bridging the efficiency gap between servers and emerging cloud workloads (2012, March 21) retrieved 27 April 2024 from <https://phys.org/news/2012-03-scale-out-processors-bridging-efficiency-gap.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.