Building Energy-Efficient Clusters for I/O-Intensive Workloads: A Fast Array of Wimpy Nodes (FAWN)

David Andersen
(Carnegie Mellon U.)

Abstract

Power and cooling have rapidly become a barrier to scaling computer systems, from individual chips to entire data-centers. In this talk, I will present our work on designing a new cluster architecture that uses up to an order of magnitude less power to serve I/O-bound workloads. Particularly for seek-intensive loads, today's servers suffer an enormous gap between the capabilities of the processor and what the storage subsystem can deliver to that processor, but still spend tens to hundreds of watts powering the CPU and chipset. Our new architecture, termed a FAWN--a Fast Array of Wimpy Nodes--operates by coupling low-power embedded-system-class processors with fast flash memory, effectively re-balancing the I/O and computational capabilities of the cluster. Programming such a cluster can be challenging; we present our initial work that shows how a FAWN can be organized to easily support high-volume key-value lookup workloads, such as those experienced by Facebook or Amazon, as well as throughput intensive workloads derived from machine learning projects that need to "grep the web."