Treating Energy Consumption as a First Class Performance Goal in Data Processing Systems

Jignesh M. Patel
University of Wisconsin—Madison
jignesh@cs.wisc.edu
http://www.cs.wisc.edu/~jignesh

Data processing systems, which includes “traditional” database management systems and more modern MapReduce-like systems, have traditionally been designed, implemented, and operated based on optimizing traditional performance metrics such as query response time or throughput. These large data processing systems consume tremendous amounts of energy, and both environmental and economical reasons now demand that we play close attention to the energy consumption characteristics of these systems. This new interest in energy efficiency along with the move towards consolidation of computing resources – namely the cloud computing movement – provides unique opportunities to rethink data processing systems in fundamental ways.

More specifically, we now need to make energy consumption a first class “performance” goal in data processing systems, which may co-exist with other traditional performance goals such as query response time. Viewed from this perspective, nearly every aspect of traditional data processing needs to be rethought, and in some cases fundamentally redesigned to become more energy efficient. For example,

a) Traditional database management systems may now optimize for the lowest energy consuming query plan while guaranteeing that the query plan meets certain response time performance constraints as may be specified in some Service Level Agreements (SLAs);

b) MapReduce clusters, which are well known to operate well below peak utilization most of the time, may want to consider turning off portions of their cluster, at non-peak hours;

c) Mobile devices may need to consider producing the best answer to a query while staying below a certain power budget;

d) Similarly, an entire cluster may want to shed some of it current work to quickly get below a new power cap;

e) A database management system may dynamically modify hardware-level power v/s performance characteristics (e.g. memory parking, or core parking) to reduce the energy consumption of a query without impacting its response time.

While the notion of energy efficiency has an inherent natural appeal to many, to really get the industry excited and engaged we must also consider how energy efficiency translates into actual economical gains. Thus energy efficiency must be considered within a broader framework that aims to minimize the total cost of owning and operating data centers/clusters. In other words, one must take a holistic systems approach to this problem or energy efficient data processing, and design, develop, and implement techniques that produce actual energy gains when measured end-to-end on actual systems/clusters.