WattDB – A Cluster of Wimpy Processing Nodes to Approximate Energy Proportionality

Theo Härder, Daniel Schall
AB DBIS, University of Kaiserslautern
{haerder, schall}@cs.uni-kl.de

Extended Abstract

Energy efficiency gained a lot of attention lately. In data management research, much work focused around the use and optimization of SSDs as middle-tier buffers or replacements of HDDs. Initial research goals aimed at performance improvement and energy saving when processing large volumes of data stored on external devices. Using SSDs, data-intensive applications achieved impressive performance gains, whereas energy saving was limited to the amount needed by the persistent storage, typically in the range of 10 – 20%.

Energy consumption is far from being proportional to system utilization for a *single computing node*, because main-memory power usage is more or less independent of system utilization and linearly grows with memory size, i. e., the number of RAM modules. Increasing main-memory size by adding more RAM modules would rapidly shift the relative power use close to 100%, even in the idle case.[1] Due to this fact, the scope for optimizing the relative energy efficiency by software means is limited and would almost disappear when a computing node is equipped with a large memory. Only if hardware optimizations would be invented, e. g., by a substantial reduction of RAM's energy consumption, more software-related opportunities would be present.

Using a single computing node, we would never come close to the ideal characteristics of *energy proportionality*. Depending on the workload, we cannot just switch off RAM chips, especially in the course of DBMS processing, because they have to keep large portions of DB data close to the processor. Preserving this reference locality is the key objective of each DBMS buffer.

Google servers mostly reach an average CPU utilization of ∼30%, but frequently even less than ∼20% [BH09]. Peak load situations often occur only in less than ∼5% of the time a server is running [SHK11]. With our current flash-based optimizations, we do not often obtain any noticeable effect on overall energy saving – except for continuous peak load situations. Given normal load patterns and arrival times, an average request is processed more efficiently, i. e., system resources are allocated for shorter intervals, thereby reducing system utilization even further – but not the *overall energy consumption*.

[1] Memory of enterprise server quality currently consumes ∼10 Watt per 4GB DIMMS [THS10].

Our energy-related DBMS experiments reported in [HHOS11] are in accord with the general observations by Tsirogiannis et al. [THS10] that "within a single node intended for use in scale-out (shared-nothing) architectures, the most energy-efficient configuration is typically the highest performing one".

What kind of system architecture enables a close approximation of ideal energy proportionality? Current research to *data management on new hardware* almost exclusively focus on high performance for continuous peak loads in specific application areas and – to achieve this goal – primarily rely on extremely large main memories. But from a "green perspective", it is unreasonable to build systems, e. g., main-memory DBMSs for OLAP applications, which have a much lower average utilization.[2] For this reason, we have started the WattDB project, where a cluster of wimpy, *shared-nothing* computing nodes replaces the powerful DB server machine. The cluster core consists of a *single node* with remote storage[3] and can attach further nodes without interrupting DB processing. In this way, the cluster can scale up to n nodes and is able to smoothly grow and shrink dynamically – depending on the current workload needs.

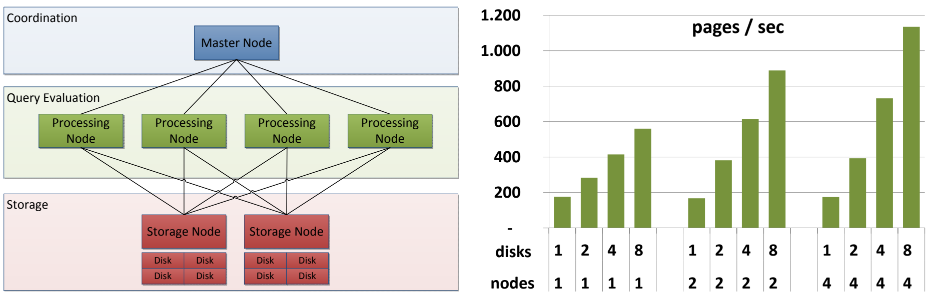
To minimize the energy footprint of each node, *Intel Atom D510* light-weight CPUs are used. In combination with the installed 2 GB of main memory and the mainboard, each node consumes less than 30 W in idle mode. Storage disks are designed for mobile use, so each disk consumes only 3 W. Our hardware is *Amdahl-balanced*, hence, processing power and data throughput are matched. Still, a single node is not energy proportional; when idle, about 70% of the node's peak power is consumed. Although the single nodes are not energy proportional, the desired energy proportionality is approximated by load-dependent deactivation/reactivation of cluster nodes. Apparently, due to this dynamic node attaching/detaching, WattDB as a cluster will stepwise approximate the ideal course of power usage, i. e., its behavior is becoming energy proportional. Note, the cluster dynamics, i. e., the time span where low-utilized nodes are disconnected from the cluster and deactivated or where overload situations are resolved by reactivating switched-off nodes, is a key question to be answered by the project.

Each of the individual computing nodes must be able to access the entire database. As a consequence, we need to build an I/O architecture, where – at each point in time and each cluster configuration – all external storage devices (SSDs or HDDs) can be dynamically shared by all attached processing nodes, i. e., the *shared-nothing* processing architecture of the cluster has to be supported by a *shared-disk* I/O architecture (see Figure 1(a)).

As a consequence of dynamic node fluctuation, DB cluster coordination becomes a frequent task to optimally support DB processing and maintenance as well as concurrency control and logging/recovery, etc. Static task assignment to specific computing nodes does not allow flexible reconfiguration and may quickly lead to unbalanced system behavior. Therefore, static allocation of processing nodes and storage structures is impractical. Hence, new partitioning schemes and procedures allowing dynamic reconfiguration have to be developed. Instead of allocating physical partitions, flexible physiological DB partitioning is needed – a new outstanding challenge to make WattDB work.

[2] Such a DB server would steadily consume 2.5 kW for each TB of main memory installed, no matter whether it is idle or working. Furthermore, substantial energy is needed for cooling and uninterruptible power supply.

[3] In this case, all coordination, query processing, and storage-related tasks have to be performed by this node.



(a) Logical cluster architecture

(b) Storage benchmark results

Figure 1: WattDB overview and some preliminary performance results

Currently, our WattDB implementation only allows to specify configurations using static data partitioning and a fixed number of nodes to run database query workloads. But this is sufficient to deliver a *proof of existence* that a cluster of wimpy computing nodes can approximate energy proportionality while processing DBMS tasks.

Initial measurement results confirmed our expectations concerning performance behavior and energy consumption. Figure 1(b) shows some isolated performance results of different cluster sizes. Using an identical benchmark, we ran a massive-parallel OLTP simulation on a predefined number of processing and storage nodes. Obviously, the results measured show a strong correlation between the number of nodes and the system performance. Energy use scales with the number of nodes as well, therefore, the system has the ability to approximate energy proportionality.

This observation is only a starting point for further research towards flexibility of cluster configurations. Flexible allocation of computing nodes and data partitions is needed to achieve dynamic cluster adaptations to the actual workload demands. Following this way, our goal is to seriously observe energy saving, especially for low-load situations and even idle times. Furthermore, we want to specialize WattDB towards differing directions to provide tailor-made support for application classes such as OLTP, OLAP, and MapReduce.

References

[BH09] Luiz Andre Barroso and Urs Hölzle. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Synthesis Lectures on Computer Architecture*. Morgan & Claypool Publishers, 2009.

[HHOS11] Theo Härder, Volker Hudlet, Yi Ou, and Daniel Schall. Energy Efficiency is not Enough, Energy Proportionality is Needed! In *DASFAA Workshops, 1st Int. Workshop on FlashDB, LNCS 6637*, pages 226–239, 2011.

[SHK11] Daniel Schall, Volker Hudlet, and Manuael Kern. Towards an Enhanced Benchmark Advocating Energy-Efficient Systems. In *TPCTC*, 2011.

[THS10] Dimitris Tsirogiannis, Stavros Harizopoulos, and Mehul A. Shah. Analyzing the Energy Efficiency of a Database Server. In *SIGMOD*, pages 231–242, 2010.