

Energy-Aware for Database Descriptions and Manipulation

MAPi PhD Proposal

January de 2014

- Thematic Area: Green Computing
- Supervisor: João Saraiva
- Co-Supervisor: Orlando Belo
- University: Minho
- Research Institute: HASLab
- External Researcher: Joost Visser, SIG & Radboud University Nijmegen, The Netherlands.

For the last couple of years, we have begun to express heavy concern over the impact done to the environment with activities like oil-drilling, CO2 emissions and energy-consumptions.

Now, with recent studies also accrediting the last two to the IT sector, people/ businesses have begun to opt for a more Green-IT approach, resorting to IT-external (cooling systems/lights/shutdowns/etc) and hardware changes (where almost all the research in Green-IT has been done). But the rising demand is exceeding efficiency improvement, since energy-efficiency on the hardware level is cancelled out by inefficient software, the most responsible for energy-consumption, where studies show that "Up to 90% of energy used by ICT hardware can be attributed to software"!

This PhD project plans to focus on data-centers, the heavy energy-consumption side of IT, where the goal is to develop theories and techniques for reducing energy-consumption, specifically in databases/data warehouses on the schema design and query level, and afterwards conducting a validation.

1 Objectives

As already mentioned, data-centers consume immense amounts of energy, spending close to 100-200 times the amount of energy as a same size workplace building. Consequently, they are a great candidate to apply "Greenness" to, reducing

both electrical consumption (the environmentalism side) and electrical expenses (the financial side), as we can easily understand just by looking at what areas Green computing has focused on.

But unlike previous Green-IT work where hardware and physical components were the principal target, this PhD project aims to research, analyze, and understand how the other side of the Green spectrum (software) can be used to make Green data-centers, specifically focusing on database/data warehouse designs, and queries. Most data-centers store many databases and data warehouses, where both have their niche and way of being designed. But their design can have many huge impacts, not only on response times, but also consumption costs. Certain database/data warehouse architectures demand more resources to calculate, which leads to high consumption costs. If consumption measurements were taken into consideration during architecture design, energy costs would reduce heavily.

Along with the design of databases/data warehouses, the query languages also have a huge impact on calculation costs. Knowing the execution order of the relational algebra functions tree, we can analyze and may be able to translate the language to a Greener calculation. In both of these cases, we also need to take into consideration the performance costs of making something Greener, balancing between Greenness v. Performance is fundamental. To do so, we can take advantage of SIGs energy-efficiency metrics on software and apply the same principals in this work, answering questions such as "Which queries/schemas consume most energy?", "How much energy is consumed per request?", or Which schema/query optimizations improve energy-consumption?.

The final goal of this project will be to validate the developed theories/techniques for Greenness in a large data-center.