CoolEmAll - Models and Tools for Optimization of Data Center Energy-efficiency

Micha vor dem Berge, Wolfgang Christmann
Christmann Informationstechnik + Medien
Ilsede, Germany
Email: [Micha.vordemBerge,Wolfgang]@christmann.info

Eugen Volk, Stefan Wesner
High Performance Computing Center Stuttgart
Stuttgart, Germany
Email: [volk,wesner]@hlrs.de

Ariel Oleksiak, Tomasz Piontek
Poznan Supercomputing and Networking Center
Poznan, Poland
Email: [ariel,piontek]@man.poznan.pl

Georges Da Costa, Jean-Marc Pierson
IRIT
University of Toulouse
Toulouse, France
Email: [dacosta,pierson]@irit.fr

***Abstract*—In this paper we present objectives and expected results of the CoolEmAll project that aims at improving energy-efficiency of data centers. To achieve this goal, CoolEmAll works on data center simulation and visualization tools (SVD Toolkit) and models of data center efficiency building blocks (DEBBs) that will be used by the SVD Toolkit to facilitate simulations and ensure their high quality. Using these models and tools data center designers and operators will be able to analyze and optimize both existing and planned data centers. By taking into consideration several important aspect that affect energy efficiency, the project is going to enable analysis of data centers under various conditions including low levels of loads, diverse management methods, specific cooling systems and IT equipment configurations.**

***Index Terms*—data centers; energy efficiency; simulations**

I. Introduction

The FP7 CoolEmAll project addresses the problem of energy efficiency of data centers. So far their efficiency has been mainly driven by minimization of Power Usage Effectiveness (PUE) [PUE] metric and adjustment of data center configurations based on experience and basic measurements. This approach faces many limits as it does not allow predictions of energy efficiency and heat transfers for data centers that do not yet exist. In order to cope with this problem some approaches apply Computational Fluid Dynamics (CFD) simulations to analyze data centers' efficiency but they are often found to be difficult to use. Additionally, existing tools and metrics usually concentrate on peak or average loads without deep analysis of data center efficiency depending on a level of load, application type and management policies.

CoolEmAll tries to address these missing issues and facilitate a process of the detailed analysis at the same time. Therefore, the main goal of the project is to enable data center designers and operators to reduce its energy impact combining optimization of IT, cooling and workload management.

CoolEmAll will develop simulation, visualization and analysis tools to enable data center designers and operators to plan and run energy and resource-efficient facilities.

These tools and models should help to minimize the energy consumption, and consequently the CO_2 emissions of data centers. To accomplish these goals CoolEmAll will deliver two main results. First, it will design diverse types of data center building blocks including hardware specification, geometrical model, and energy efficiency metrics. Second, it will develop simulation, visualization and decision support toolkit (SVD Toolkit) to enable analysis and optimization of data centers built of these building blocks.

Both building blocks and the toolkit will take into account aspects that have major impact on actual energy consumption: hardware characteristics, cooling solutions, properties of applications, workload variability, and resource management policies. To achieve it, the energy efficiency of data center building blocks will be precisely defined by a set of metrics and parameters expressing relations between essential factors listed above and the energy consumption or heat dissipation. In this way CoolEmAll will enable optimization of data center energy-efficiency also for low and variable loads rather than just for peak loads as it is usually done today.

The project also aims at selecting or, if needed, proposing new metrics that define energy and thermal efficiency of building blocks under various conditions. The models and simulation tools will be created and verified using a testing environment based on the servers prototypes, delivered by one of the project's partners, allowing fine grained monitoring and control of resources.

The structure of the paper is as follows. Section II contains a brief description of the CoolEmAll ComputeBox concept along with RECS hardware prototype used in the project as an environment for verification of models and simulation tools. In Section III we describe the concept of data center building blocks which can be applied in modeling

978-3-901882-46-3

and planning data centers. In Section IV the simulation and visualization toolkit is presented. Section V concludes the paper.

II. CoolEmAll ComputeBox

One of the results of CoolEmAll will be blueprints of ComputeBoxes - bundles of pre-configured and ready-to-use components for data centers. As such they are similar to the Open Compute Project [OpenCompute] but more focused on modules for building data centers. These blueprints will range from single racks up to larger modules such as containers or server rooms.

The first blueprint, called ComputeBox1, has been already defined and it allows to deliver a complete data center in one rack. The ComputeBox contains high-density and energy-efficient servers, which will be the RECS Multi-Node Computer [Christmann2009] in the CoolEmAll testbed case, as described in following paragraphs. The RECS Computer system has been chosen due to its high energy-efficiency, density, rich reconfiguration capabilities, and integrated efficient monitoring and controlling solution which enables us to monitor the complete rack at a very fine granularity with a negligible overhead for the computing and network resources. The ComputeBox1 also includes a special 19" rack with an integrated cooling solution as well as an efficient monitoring architecture that provides temperatures and power consumption of all components. A central and very fast storage will also be integrated and connected via a fast interconnect which is coincidently the backbone for the computing nodes.

In order to build and verify simulation models a partial prototype of ComputeBox1 is being developed by the project as a project testbed. The cluster server RECS consists of 18 single CPU modules, each of them can be treated as an individual PC. The mainboards are COM Express based CPU modules, each mounted on a standardized baseboard which makes it possible to use every available COM Express mainboard that has the basic size. In CoolEmAll we will evaluate which CPU module will be the best for each particular use-case. Each baseboard is connected to a central backplane. This backplane has two functions, first it forwards each Gigabit Ethernet Network of the CPU modules to the front panel of the server, and second it connects the baseboards' microcontrollers to the central master-microcontroller. The first versions of the RECS prototype has been created and tested within other FP7 project - GAMES [Kipp2011]. An example of one RECS is presented in Figure 1. It contains three types of CPUs and can be inserted into a rack in a single 1U chassis. The most powerful settings in this example include Intel i7 Quad-Core CPUs with 16 GB of RAM each.

The novel monitoring approach of the RECS Cluster System is to reduce network load, avoid the dependency of polling every single compute node at operation system layer and build up a basis on which new monitoring- and controlling-concepts can be developed. Therefore the status of each compute node of the RECS Cluster Server is connected to an additional independent microcontroller in order to manage the measured data. Each node is equipped with a thermal and current sensor. All sensor data are read out by one microcontroller per node which acts as a slave and thus waits to be pulled by the master microcontroller. In this way potential overheads caused by measuring and transferring data are reduced; in particular in a large-scale environment this approach can play a significant role. This microcontroller-based monitoring architecture is accessible to the user by a dedicated network port and has to be read out only once to get all information about the installed computing nodes. If a user monitors e.g. 10 metrics on all 18 nodes, he would have to perform 180 pulls which can now be reduced to only one.

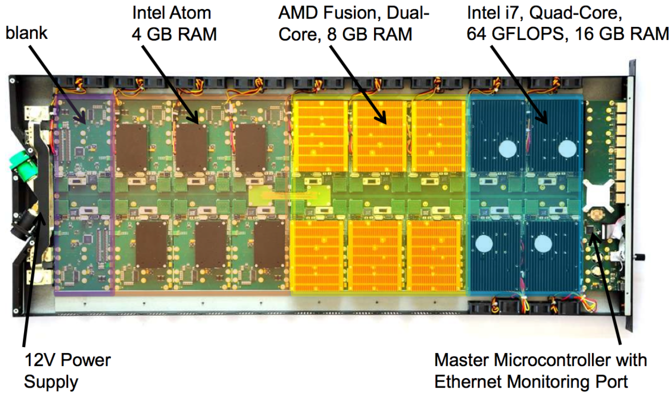

Fig. 1. RECS example

The ComputeBox is a blueprint that can be used by data center designers as a source of good practices and a module for building or extending data centers. Nevertheless, a ComputeBox consists of smaller elements and users should be able to use these elements as predefined input to simulations. To this end, CoolEmAll introduced a concept of Data center Building Blocks (DEBBs) presented in the next section.

III. Data Center Efficiency Building Block (DEBB)

The main goal of DEBB is to facilitate a process of data center modeling, simulation, and visualization. DEBBs are defined in such a way that users can insert them into simulation tools rather than build models from scratch. The CoolEmAll project takes a holistic approach to data center modeling and simulation by allowing end user to analyze data center from a number of perspectives. Therefore DEBB definitions must consist of various types of information needed for workload and resource simulations, CFD simulations, calculation of metrics and visualization. Hence a DEBB, as presented in Figure 2, is an abstract description of a piece of hardware and other components defined by:

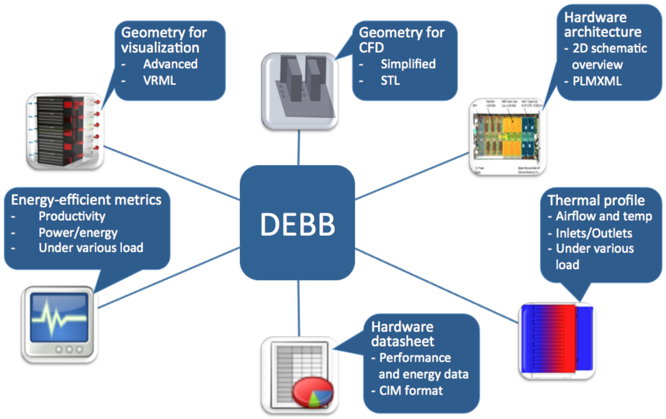

Fig. 2. DEBB structure

- 3D model of DEBB components needed for visualization; these 3D models are described using well known VRML format,
- 3D model of DEBB components needed for CFD simulations; the model can be simplified compared to the visualization model in order to speed up CFD simulations; these 3D models are described using common STL format,
- Specification of components and sub-building blocks; they optionally include arrangements of components and sub-building blocks within particular DEBB (white-box description); the PLMXML standard [PLMXML] is considered to describe the hierarchy of building blocks,
- Thermal profile describing air-flow (including direction and intensity) for air-based cooling and temperature on inlets and outlets for different load-levels,
- Hardware datasheet containing information about performance and power consumption of components for different load levels; information about the power consumption should include mainly CPU and memory, and optionally IO and storage; for this purpose the CIM format [CIM] is considered,
- Metrics describing energy efficiency of a DEBB.

DEBBs are designed to model data-center building blocks on diverse granularity levels, from a single node up to a complete data center. In this way they can support users in modeling and simulating a data center. Within CoolEmAll, the following DEBBs will be supported:

1) Node Unit is the smallest element of a data center to be modeled. This unit reflects a single computing node, e.g. a single blade CPU module or a RECS CPU module.
2) Node Group reflects an assembled unit of building blocks of level 1, e.g. a complete blade center or a complete RECS unit (currently consisting of 18 node-units).
3) ComputeBox1 reflects a typical rack within an IT service center, including building blocks of level 2 (Node Groups), power supply units and integrated cooling devices.
4) ComputeBox2 building blocks are assembled of units of level 3, e.g. reflecting a container filled with racks or even complete compute rooms.

As the focus of CoolEmAll is to simulate thermal behavior of a DEBB it will be modeled as the smallest unit in the thermodynamic modeling process. The thermodynamic processes within a Node Group should provide models to calculate boundary conditions for simulations of larger modules, e.g. full racks or server rooms. To perform these simulations the arrangement of the Node Groups with the temperature and air throuhput at the Node Group outlets over time will be provided as an inbound boundary condition. Results of these simulations will contain the room temperature over time at the outlet of the Node Group.

In this way DEBBs should improve and facilitate the process of modeling, simulation, and visualization of data centers by delivering predefined models with comprehensive information concerning performance, power consumption, thermal behavior, and shape of data center components. To achieve it DEBBs will be used by the CoolEmAll simulation, visualization, and decision support toolkit presented in the next section.

IV. SIMULATION, VISUALIZATION AND DECISION SUPPORT TOOLKIT

The main result of the CoolEmAll project will be a simulation, visualization, and decision support toolkit (SVD Toolkit) which is a simulation tool to allow data center planners and operators to model the energy-efficiency implications of physical placement of IT within the facility. It will also examine different approaches to cooling as well as the role played by applications and workload. The platform will enable the optimization of both the data center design and operation in the process of planning new infrastructures as well as improving the existing ones.

The simulation platform will integrate models of applications, workload scheduling policies, hardware characteristics and cooling. The results of this platform will include estimations of energy consumption and energy-efficiency metrics as well as air and thermal flows obtained from computational fluid dynamics (CFD) simulations. The SVD Toolkit will be accompanied by advanced visualization tools that will allow users to easily analyze the simulation result for many configuration options. They will be also able to study impact of various loads, thermal- and energy-aware workload scheduling and resource management on the energy efficiency. A modular approach to these simulations (according to DEBBs defined in Section III) provides many extension possibilities and a high level of customization.

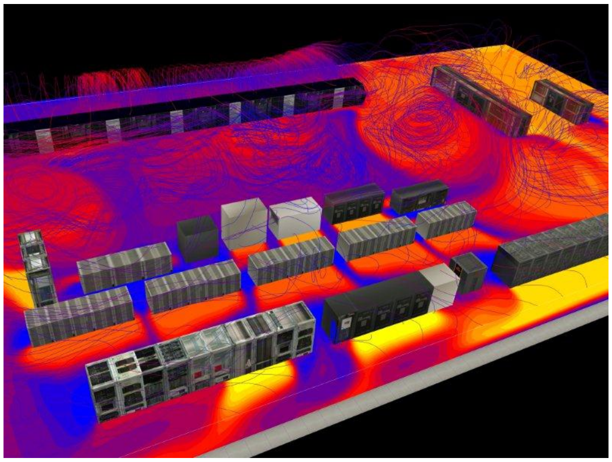

Fig. 3. Visualization of temperatures and air flow in a computer room

The main CoolEmAll outcomes will be based on existing technologies, which are extended appropriately to meet the objectives. The following subsections present the architecture of the toolkit (Section IV-A) and planned simulations performed using the SVD Toolkit, namely heat transfer using CFD (Section IV-B), workload and resources (Section IV-C), and applications (Section IV-D).

A. Architecture

The main goal of the SVD Toolkit is to provide an open solution for assessing the energy efficiency of a data center by simulations. Simulations integrated into the toolkit include CFD simulations for assessing a cooling system of the rack or computing room. Workload simulations will provide estimated heat generation information for the CFD simulations as well as estimation of resource utilization and perfomance. Concepts of Compute Boxes and DEBBs defined in Section III will allow modeling various hierarchies, ranging from single nodes and racks to containers or even whole data center rooms. The full CoolEmAll workflow for modeling, simulation and visualization of data centers is presented in [Berge2012]. A user of the SVD Toolkit will be able to use its tools separately and pass data between them if needed or to execute the whole workflow in an automated way. To enable this the Simulation and Visualization platform COVISE [Wierse1993] will be applied as an orchestrator for the SVD Toolkit. It can be used to steer the simulations included into the SVD Toolkit and to provide advanced visualization of results. An example of temperature distribution and air flow visualization is illustrated in Figure 3.

B. Heat transfer simulations

For assessing heat issues in a data center, the heat dissipation and distribution of the nodes and its components have to be known. Heat generated by the compute nodes as well as the air throughput at the server outlets will be determined using the workload simulations (Section IV-C) based on earlier tests and application models (Section IV-D). Output of workload simulations will be passed as an input to simulations of heat transfer and cooling. Several cases are considered: passing heat dissipation and air throughput from a selected time point interesting for a user, calculating average values or ranges for a given period, or passing values changing in time for non-stationary simulations. As the cooling devices in operation nowadays are usually air or liquid based, simulations of heat transfer and fluid dynamics can be used to assess the effectiveness of the cooling and to predict the room temperature distribution. In CoolEmAll an OpenFOAM [Weller1998] based solver will be the main choice for a CFD tool integrated into the SVD Toolkit as it is released as open source under the GNU General Public License. In this way CoolEmAll will enable heat transfer and airflow simulations leading to the assessment of the cooling for the computing nodes without additional costs for end users.

C. Workload simulations

The CoolEmAll goal is to enable analysis of data center energy-efficiency for various levels of load rather than for peak values as it is commonly done to date. To this end, the SVD toolkit will enable analysis of data centers with respect to important aspects that may affect their energy efficiency including a workload to be executed in a data center. Although impact of workloads is often neglected they may strongly affect amounts of energy consumed by a data center and values of energy-efficiency metrics. The extent of this influence depends on workload variability, size, and type. For instance, PUE values can differ significantly between periods of peak and low loads. Workload characteristics also determine energy saving methods that can be applied. For example, virtualized workloads can be migrated and consolidated at runtime.

In order to allow estimating data center energy efficiency for diverse loads CoolEmAll will take into consideration workloads both from HPC and Cloud domain. These workloads will come from data centers operated by institutions belonging to the project consortium. To simulate workloads and workload scheduling policies the CoolEmAll Data Center Workload and Resource Management Simulator will be developed. Its functionality will include simulations of different kinds of workloads and analyzing impact of scheduling and resource management policies on resource utilization and power usage. The output of the simulator will include load, heat generation, and servers air throughput information for CFD simulations and visualization within SVD Toolkit. The simulator will be developed as an enhancement of the GSSIM simulator [GSSIM]. This enhancement will be facilitated by basic energy efficiency modeling features that GSSIM already supports and available mechanisms to insert custom workloads and scheduling policies [Krystek2010]. The simulation environment will also include means to insert application-specific performance and energy consumption models. More details concerning application modeling is presented in the next section.

D. Application modeling and profiling

The analysis of data center energy-effiiency should be performed with applications ultimately running in this data center in mind. Usually representative benchmarks are used to reflect diverse sets of applications. So far, their main focus was solely on performance (e.g. FLOPS for HPC and time to result or requests/second for services in industry). Hence there is a need to complete metrics with power consumption and thermal impact and provide appropriate benchmarks. To achieve it there are several challenges to overcome, as defined in [DaCosta2012] and [Berge2012].

In particular, the thermal impact of applications needs to be measured and modeled. The thermal impact is based on a combination of power and time, as most applications have a non constant behavior over time. Therefore, a handful of unique (but yet representative) applications that can serve as a basis for benchmarks should be classify in the multi-dimentional impact space (performance, power, thermal). To achieve such a classification, the ability to recognize phases of application [DaCosta2011] during runtime is necessary in order to acquire the knowledge leading to the classification. CoolEmAll will address these questions in order to build application models for simulations.

V. Conclusion

The CoolEmAll project presented in this paper aims at delivering a set of innovative models as well as simulation and visualization tools that will enable comprehensive analysis of data center energy efficiency in an easy way. CoolEmAll will allow studying scalability of data center power usage with load. In particular, the project will provide means to analyze impact on energy consumption factors such as IT equipment arrangement, hardware configuration, cooling systems types and settings, workload and resource management policies, types and parameters of applications. CoolEmAll models will be verified on the RECS system enabling very fine grained efficient monitoring and rich re-configuration options. The CoolEmAll SVD Toolkit along with data center building blocks will be an open framework enabling integration of various tools for workload simulation, CFD simulation, and visualization. These tools will support data centers designers and operators in planning and optimization of both new and existing data centers.

Acknowledgment

The results presented in this paper are partially funded by the European Commission under contract 288701 through the project CoolEmAll.

References

[GSSIM] Bak, S., Krystek, M. Kurowski, K., Oleksiak, A., Piatek, W., Weglarz, J. (2011). GSSIM - a Tool for Distributed Computing Experiments, Scientific Programming, vol. 19(4), pp. 231-251, 2011

[Berge2012] M. vor dem Berge, G. Da Costa, A. Kopecki, A. Oleksiak, J-M. Pierson, T. Piontek, E. Volk, and S. Wesner (2012). Modeling and Simulation of Data Center Energy-efficiency in CoolEmAll, to appear in Lecture Notes in Computer Science, vol. 7396

[Christmann2009] Christmann (2009). Description for Resource Efficient Computing System (RECS). http://shared.christmann.info/download/project-recs.pdf

[DaCosta2011] Da Costa G., and Pierson J-M. (2011) Characterizing applications from power consumption: A case study for HPC benchmarks. In *International Conference on ICT as Key Technology for the Fight against Global Warming*, Toulouse, Springer.

[DaCosta2012] Da Costa Georges, Hlavacs Helmut, Hummel Karin and Pierson Jean-Marc. (2012) Modeling the Energy Consumption of Distributed Applications. In *Handbook of Energy-Aware and Green Computing*, Chapman & Hall, CRC Press.

[Kipp2011] A. Kipp, L. Schubert, J. Liu, T. Jiang, W. Christmann, M. vor dem Berge (2011). Energy Consumption Optimisation in HPC Service Centres, Proceedings of the Second International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering, B.H.V. Topping and P. IvaÌĄnyi, (Editors), Civil-Comp Press, Stirlingshire, Scotland

[Kipp2012] Kipp, A., Jiang, T., Fugini, M. et al. (2012) Layered Green Performance Indicators, 478-489. In Future Generation Computer Systems 28 (2).

[Krystek2010] Krystek M., Kurowski K., Oleksiak A., Piatek W. (2010). Energy-aware simulations with GSSIM, 1st Year Proceedings of the COST Action IC0804

[Pierson2011] Pierson Jean-Marc (2011) Energy: A New Criteria for Performances in Large Scale Distributed Systems in "Performance Evaluation of Computer and Communication Systems. Milestones and Future Challenges". Lecture Notes in Computer Science, 2011, Volume 6821/2011, 38-48

[Wierse1993] Wierse A., Lang U., and Rühle R. (1993) A system architecture for data-oriented visualization. In *Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization*, pages 148–159, London, UK, Springer-Verlag.

[Weller1998] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. (1998) A tensorial approach to computational continuum mechanics using object-oriented techniques. In *Computers in Physics* Volume 12, Issue 6, pp. 620–631

[Volk2012] Volk E., Buchholz J., Wesner S., Koudela D., Schmidt M., et al. (2012) Towards Intelligent Management of Very Large Computing Systems. In Proceedings of the *Competence in High Performance Computing 2010.*

[CIM] The Common Information Model (CIM) website, http://dmtf.org/standards/cim

[OpenCompute] The Open Compute Project website, http://opencompute.org

[PLMXML] The Open Product Lifecycle Data Sharing (PLMXML) format, http://www.plm.automation.siemens.com/en_us/products/open/plmxml/

[PUE] The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE, http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE