# 1 Executive Summary

Energy has always been a concern in large scale distributed platforms. Nowadays, it becomes even more critical due to the transition to the next generation extreme scale High Performance Computing platforms (HPC) and its convergence with the concepts of Cloud, Big Data and the Internet of Things. An increase of the computing performance of HPC platforms by a factor of 100 is targeted while staying at the current order of magnitude in terms of Power consumption. This huge gap clearly shows that reaching the exascale needs a revolution in the way of handling resource management problems. Reducing the energy consumption to obtain radically more Flop/s per watt than today's systems is the major challenge. In order to address big societal applications such as health, security or climate, HPC platforms evolve toward heterogeneous many-cores (hundreds of cores in the same chip) composed of general purpose components along with specialized cores. The number of computing units will drastically increase but the I/O and interconnection networks are evolving much slowly while the memory hierarchy will be even deeper than today. In addition, more processing capabilities will obviously lead to more data produced, stressing even more the interconnects. Data movements will continue to grow (more data produced and more complex traffic), both within nodes and between nodes. Hence, the energy challenge is further complicated because of data movement and storage that are expected to consume more than 70% of the total power [7]. Several complementary steps are needed to allow the system management software to adapt to this ever-increasing scale and complexity [29]. The ability to build a physical exascale system is not a guarantee for running exascale applications. **Efficient tools to utilize such platforms at a sustained rate** must also be provided. A key element for application design and management will be to better use memory hierarchies and optimize data movements [24].

The ENERGUMEN project will study efficient and practical tools for managing the allocation of jobs to the various components of an extreme scale HPC platform. There exist various mechanisms for reducing the energy in large scale HPC platforms. First, it is possible to decrease the clock frequency of the computing units (known as *Dynamic Voltage and Frequency Scaling* - DVFS). Another way is to switch-off some nodes (*shutdown*) or to put them into a sleep mode for some time (*power-down*). Both mechanisms can be used simultaneously. There exist many studies in this direction, mainly theoretical assuming idealized and restricted models. Alternatively, saving energy can be obtained as a consequence of reducing data movements by adequate communication-aware allocations of the jobs (resulting to internal communications with high locality, close to I/O nodes, etc.). Current approaches use this leverage, but are limited by assuming that they do not have any influence on the applications themselves and by considering them as black boxes. Increasing the scale will even more increase the amount of data to move. Recent studies [7] emphasize that today's platforms spend most of their time moving data instead of computing. To the best of our knowledge, no work has been devoted so far to study explicit methods of saving energy as a consequence of enhanced allocations and reduced data movements using the knowledge extracted from applications. In this project, we propose two new complementary mechanisms for addressing the energy/performance trade-off in extreme scale HPC platforms. First, we will revisit the classical speed-scaling and power down mechanisms by using a *malleable model*, which allows to shrink or stretch dynamically the execution time of the jobs according to the current energy profile. We will also study optimized policies for energy-aware data allocations at the software level. These mechanisms aim to introduce more flexibility into the management of the heterogeneous resources of extreme scale HPC platforms. Both mechanisms involve the design of sophisticated models and methodologies for the efficient exploitation of the idle periods at running time. There are hard scientific and technological challenges to determine the best trade-off between both mechanisms. The success of designing adequate models and efficient optimization methods depends heavily on the collection and the analysis of the huge amount of data produced in HPC platforms. We also propose to develop and test several software products in actual datacenters.

**In the quest of extreme scale HPC platforms where a trade-off between performance and energy consumption is necessary, the originality of ENERGUMEN is to revisit the principles of existing resource managers and to investigate new functionalities by harnessing applications' malleability.**