A NEW JOB MIGRATION ALGORITHM TO IMPROVE DATA CENTER EFFICIENCY

The increase of the computational power leads necessarily to more complex approaches in the resources management. One of the main problems nowadays for a computing center is given by the under exploitation of the available resources. It may happens that in a heterogeneous batch queue system, available for both serial single core processes and parallel multi core jobs, one or more computational nodes of the cluster are serving a number of jobs lower than their capability. A typical case is represented by more single core jobs running each one over a multi core server, while more parallel jobs - requiring all the available cores of a host - are queued. The idea is to pile up the maximum number of jobs over the minimum number of hosts, compatibly with the available CPU and memory. This way the running jobs do not suffer any performance loss, and at the same time the farm may gain contiguous job slot able to host new parallel multi-core jobs otherwise stuck in queue. A prototype of job mover has been developed with the aim of freeing the best part of otherwise unavailable resources in a computing cluster. We started implementing a batch system and queue simulator, in order to test the efficiency of several job rearrangement algorithms. Defining an exploitation parameter, strictly connected to the cluster load, we implemented two algorithms able to increase the computational resources load - limited for each server by the number of available cores. The problem of the permutations of a set of jobs, each one requiring a variable number of core, over a set of server is described by an NP-complete complexity class. Due to the difficulty in finding the best solution, we focused on searching a solution able to improve the current load status of the cluster. In order to ensure a full hosts exploitation, and prevent at the same time the overload of one or more nodes in the cluster, the job migration takes place only under certain conditions. We also paid a special attention to avoid a too frequent job displacement, damaging the global performance. A secondary effect, probably not less appealing by the point of view of the "green computing", is represented by the power efficiency improvement through a dynamic job rearrangement, with an energy saving up to 90% in some particular cases - more frequent than expected. The system, developed at Scuola Normale Superiore, in collaboration with the Computer Science Engineering Department at the University of Pisa [Italy], is able to provide an increase in the number of running jobs. The increase is from 15% in case of heterogeneous single and multi core running processes, to 90% in case of long term single core jobs running over several different hosts, and a large number of parallel multi core processes - requiring the entire number of processors of a single computational node - stuck in queue.

Primary authors : Dr. CALZOLARI, Federico (Scuola Normale Superiore - INFN Pisa) ; Dr. VOLPE, Silvia (Computer Science Engineering Department, University of Pisa [Italy])