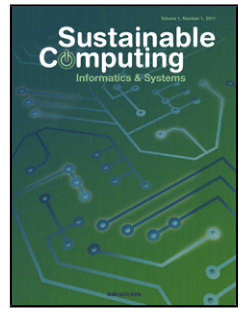
Accepted Manuscript



Title: Energy Efficient Computing, Clusters, Grids and Clouds: A Taxonomy and Survey

Author: Muhammad Zakarya Lee Gillam

PII: S2210-5379(17)30070-7
DOI: http://dx.doi.org/doi:10.1016/j.suscom.2017.03.002
Reference: SUSCOM 164

To appear in:

Received date: 12-11-2014
Revised date: 29-10-2016
Accepted date: 7-3-2017

Energy Efficient Computing, Clusters, Grids and Clouds: A Taxonomy and Survey

Muhammad Zakarya[1,2] and Lee Gillam[1]

[1]Department of Computer Science, University of Surrey, UK
[2]Abdul Wali Khan University, Mardan, Pakistan
{mohd.zakarya,l.gillam}@surrey.ac.uk

Abstract. Cloud computing continues to play a major role in transforming the IT industry by facilitating elastic on-demand provisioning of computational resources including processors, storage and networks. This is necessarily accompanied by the creation, and refreshes, of large-scale systems including cluster, grids and datacenters from which such resources are provided. These systems consume substantial amounts of energy, with associated costs, leading to significant CO_2 emissions. In 2014, these systems consumed 70 billion kWh of energy in US; this is 1.8% of the US total energy consumption, and future consumption is expected to continue around this level with approximately 73 billion kWh by 2020. The energy bills for major cloud service providers are typically the second largest item in their budgets due to the increased number of computational resources. Energy efficiency in these systems serves the providers interests in saving money to enable reinvestment, reduce supply costs and also reduces CO_2 emissions. In this paper, we discuss energy consumption in large scale computing systems, such as scientific high performance computing systems, clusters, grids and clouds, and whether it is possible to decrease energy consumption without detrimental impact on service quality and performance. We discuss a number of approaches, reported in the literature, that claim to improve the energy efficiency of such large scale computing systems, and identify a number of open challenges. Key findings include: (i) in clusters and grids, use of system level efficiency techniques might increase their energy consumption; (ii) in (virtualized) clouds, efficient scheduling and resource allocation can lead to substantially greater economies than consolidation through migration; and (iii) in clusters, switching off idle resources is more energy efficient, however in (production) clouds, performance is affected due to demand fluctuation.

1 Introduction

Large scale computing systems as observed in the top500 [1] supercomputers, clusters, grids [2], and clouds [3] consist of a large number of Information & Communication Technology (ICT) resources that are connected through a network. Supercomputers and clusters are non-distributed systems which are used to solve large problems quickly where large mathematical calculations are involved like weather forecasting, defence & control systems etc. Distributed systems (grids and clouds) are preferred over non-distributed systems for three main reasons including reliability, distributed nature of applications and concurrent execution of applications [2]. These systems provide their services to users on either best or commercially reasonable effort policy.

Cluster, grid, cloud and datacenter service providers maintain a large pool of computational resources, that needs more energy to (i) operate properly and (ii) to cool down the heat generated. A recent report [4] show that in 2015, across the world almost 416.2 terawatt hours of energy was consumed by datacenters which is higher that the UK's total consumption. The amount of energy consumption will continue to increase with increasing capacity unless energy efficient management techniques are established and applied [5]. The resource allocation and management algorithms along with the physical infrastructure of the datacenter are needed to reduce the environmental impact (CO_2 emission) and make them more energy and cost efficient [6]. In this paper, we discuss energy consumption of clusters, grids and clouds, and whether it is possible to minimize energy consumption without detrimental impact on service quality and performance.