Rough Set Based Green Cloud Computing in Emerging Markets

P.S. Shivalkar
School of Computing Science and Engineering, VIT University, India

B.K. Tripathy
School of Computing Science and Engineering, VIT University, India

INTRODUCTION

Cloud computing is all about the usage of resources – everything from Software-as-a-Service (SaaS) to Platform-as-a-Service (PaaS) to Infrastructure-as-a-Service (IaaS) – over the Internet on a pay-per-use basis. It is either a kind of revolutionary new paradigm for Information Technology (IT) service delivery of the products or a new name for a service delivering products as old as IT itself. When we deal with cloud computing, the very first thing comes to our mind is data sharing. A lot of data gets stored on the data servers from various regions of the world, which is stored over cloud running various applications. These applications consume small amount of data and generate large data sets.

The basic cloud computing services provided by organisations are often subdivided based on whether they provide SaaS, PaaS, and IaaS, but their needs for standards appear similar. The cloud is being used over a wide range of locations, for various purposes in different areas which include healthcare, education, government, communities, e-commerce, etc. The cloud computation involves a large network of servers, dumb terminals, data storages which all together work in real-time environment.

BACKGROUND

National Institute of Standards and Technology (NIST) defines cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability (Mell & Grance, 2011). Recently, e-commerce or electronic commerce experiences a rapid development in the world. Half of large scale enterprises and more than one fourth of small scale ones are involved in e-commerce (Hongbo et al., 2012). A recent World Economic Forum report emphasized how cloud computing, coupled with sound business and software strategies, can speed up the creation of innovative products or services and help companies provide better customer service (Murugesan, 2011). The data used over the cloud is tremendously huge and is known as BigData.

To make these data to be brought in use efficiently, researchers are working in various aspects so that it must not make any harm to the nature in any way. Many companies have started using cloud computing increasing their profit and also started saving more power, which is causing the nature to get affected in a good way i.e. it makes more power utilization. There are techniques to minimize the overhead of the structure, extract only the relevant information reducing the power consumption by lowering the energy to handle data. This article explores the overhead load reduction of data transmission over Cloud by making the use of analyzing the mechanism that support the goal of energy saving. Rough Set is one such technique which can be used to reduce the data content. Also the extensions of the Rough Set theory (rough mereology) have developed new concepts like data mining in distributed systems.

DOI: 10.4018/978-1-4666-5888-2.ch103

E-MARKETING AND GREEN CLOUD COMPUTING

With the growth of high speed networks, there is an alarming rise in their usage. This ever-increasing demand is handled through large-scale datacenters, which consolidate hundreds and thousands of servers with other infrastructure such as cooling, storage and network systems. The commercialization of these developments is defined currently as Cloud computing (Buyya et al., 2008). High growing demand of Cloud infrastructure by Google, Amazon, etc has drastically increased the energy consumption of data centers (Kumar & Buyya, 2012).

Cloud computing, through the use of large shared virtualized datacenters can offer large energy savings. However, Cloud services further increase the Internet traffic, hence growing information database which in contrast increase energy consumption. Energy-efficient solutions are required to minimize the impact of Cloud computing on the environment.

According to a study by GfK (Gesellschaft für Konsumforschung) Custom Research (Murugesan, 2011), cloud computing is being used on large scale in these emerging markets. Multinational companies are taking measures on regional level to penetrate the market for good service provision and are also taking care of localization (Murugesan, 2011).

The issue of energy consumption in IT equipment is receiving more attention in recent years and there is growing recognition of the need to manage energy consumption across the entire information and communications technology (ICT) sector. In cloud software services, power consumption in transport is negligibly small at very low screen refresh rates. As a result, cloud services are more efficient than modern midrange PCs for simple office tasks. At moderate and high screen refresh rates, power consumption in transport becomes significant and energy savings over midrange PCs are reduced. The number of users per server is the most significant determinant of the energy efficiency of a cloud software service.

The power consumption during data transfer can be reduced with the help of techniques like Rough set which is very useful in getting the reduced dataset in form of reducts and cores. The data worked over cloud is huge and the drawback with the Rough set is that it cannot be applied over a huge data set. So for the same purpose the new methodology of MapReduce is used to make small clusters of the BigData being transmitted over the network. Next, we shall define Rough set, how it can be used to reduce BigData and hence saving power consumption during data transfer.

MAPREDUCE

The enormous amount of data accumulated by organisations creates new challenges in dealing with BigData in order to acquire useful information. For the purpose of processing BigData, Google developed a software framework called MapReduce to support large distributed data sets on clusters of computers (Dean et al., 2008), which is effective to analyse large amount of data. As one of the most important cloud computing techniques, MapReduce has been a popular computing model for cloud computing platforms. Followed by Google's work, many implementations of MapReduce emerged and lots of traditional methods combined with MapReduce have been presented until now.

Apache Hadoop is a software framework that helps constructing the reliable, scalable, distributed systems (Hadoop, 2010). Phoenix is a shared-memory implementation of Google's MapReduce model for data-intensive processing tasks (Ranger et al., 2007). Mars is a MapReduce framework on graphic processors (He et al., 2008). Twister is a lightweight and iterative MapReduce runtime system (Ekanayake et al., 2010).

Some other developments in this direction are obtained by combining traditional methods with MapReduce. Apache Mahout can help to produce implementations of scalable machine-learning algorithms on Hadoop platform (Mahout, n.d.). Menon et al gave a rapid parallel genome indexing with MapReduce (Menon et al., 2011). Blanas et al proposed crucial implementation details of a number of well-known join strategies for log processing in MapReduce (Blanas et al., 2010). Ene et al developed fast clustering algorithms using MapReduce with constant factor approximation guarantees (Ene et al., 2011). Lin et al presented three design patterns for efficient graph algorithms in MapReduce (Lin et al., 2010).

MapReduce is a programming model, which is described as follows:

The computation takes a set of input key/value pairs and produces a set of output key/value pairs. The user