Center for Experimental Research in Computer Systems (CERCS)

A CISE-funded Center

Georgia Institute of Technology, Karsten Schwan, Director, 404.894.2589, schwan@cc.gatech.edu
Georgia Institute of Technology, Calton Pu, 404.385.1106, calton@cc.gatech.edu
Georgia Institute of Technology, Doug Blough, 404.385.1271, doug.blough@ece.gatech.edu
Georgia Institute of Technology, Sudhakar Yalamanchili, 404.894.2940, sudha@ece.gatech.edu
Ohio State University, Jay Ramanathan, 614.565.4187, jaram@cse.ohio-state.edu
Ohio State University, Rajiv Ramnath, ramnath@cse.ohio-state.edu

Center website: http://www.cercs.gatech.edu/

Scalable Management for Cloud Computing Environments: Monalytics

Cloud computing offers tremendous benefits to organizations by providing on-demand access to configurable computing resources, while lowering costs and enabling entities to increase computing assets with minimal effort. Data centers are the backbone of cloud infrastructures. They must be able to efficiently and rapidly adapt to support the increasing demands of cloud clients. As a result, automation in facility management is a key challenge, particularly when given the large-scale nature of data center and cloud systems.

Cloud providers have difficulty dealing with variability, both in the demands of client applications and in the resources available for the computation. By monitoring all aspects of hardware and software performance, service providers can detect and address performance problems. Unfortunately, for very large scale cloud environments, this cannot be done, since the amounts of monitoring data generated would be staggering. Therefore, cloud vendors are hampered in extending and growing their facilities to meet future demand.

Researchers at CERCS have developed Monalytics, a manageability software program that facilitates the development of next-generation scalable data center management products for very large cloud computing environments. It operates like the many ‘”big data” applications used by companies to mine data about customer preferences, but in contrast to those systems, its purpose is to flexibly perform management functions in a manner that scales to the size of the task. It enables cloud service providers to more effectively operate the ever-increasing numbers of hardware and software components in the data center, to make vast amounts of computing power available on demand and to enable cost savings and reduced energy demands. With this scalability, Monalytics will make it possible for future cloud data centers to provide levels of computing power equivalent to today’s largest supercomputers, thereby realizing the full potential of cloud computing. It does so by processing monitoring data online – to rapidly extract data of interest, in-place – where data is generated so that data center networks are not overwhelmed.

Monalytics has attracted attention of multiple companies, such as HP and VMWare. These entities are particularly interested in the ability of the technology to enable significant performance improvements in large scale cloud environments while reducing system management costs. At HP, groups have used the approach to monitor utility data centers. At VMWare, online management methods like those enabled by Monalytics are routinely used to consolidate data center systems.

> **Economic Impact:** The manageability software market for clouds, in particular, is estimated by International Data Corporation to be a $2.5 billion market by 2015. Monalytics will directly contribute to this market and will accelerate the growth and scope of cloud computing by allowing cloud environments to grow ever larger. It will allow cloud service providers to better meet the challenges of system management for the millions of managed objects in future large cloud environments. Experimental evaluations have shown the Monalytics approach yields up to 92% reduction in time to insight and 86% lower cost compared with traditional approaches to performance management. This advance will result in lower-cost, more efficient data centers run by public and private cloud providers. It will also improve end-user’s online experiences, ensuring consistent and reliable performance in their cloud applications, leading to greater adoption of clouds and their use. This can lead to new business opportunities in clouds, such as improved efficiencies offered by online methods for managing data center power consumption, or better service offerings like “premium” services that provide improved service quality to end users.

For more information, contact Karsten Schwan, 404.894.2589, schwan@cc.gatech.edu.

Attaining Predictably Fast Responses: The "Travelport Flight Shopping Engine"

Researchers at the Center for Experimental Research in Computer Systems (CERCS) have been working with industry to improve the performance of and consider new services and functionality for Travelport, a shopping engine for making travel arrangements (flight shopping). Several prototypes have been developed to address topics that include early problem detection, traffic distribution, multi-core processing, and new services.

One result of this work was a 35% improvement in the average response time of end user requests for travel options (e.g., possible flights), and a 10% reduction in failures to meet response time requirements for such requests. The economic impact of the breakthrough translated into 20% less hardware purchases by Travelport (the shopping engine provider) and a more competitive position for Travelport's shopping engine in the market place. These technology breakthroughs, therefore, improved an existing product for airline shopping, i.e., Travelport's flight shopping engine. These breakthroughs are relevant to all products that require extensive calculations such as flight shopping, ticketing, hotel reservations, and similar. More importantly, they are relevant to any application for which the amount of processing performed for each request can vary significantly depending on the nature of the requests.

Recent work with Travelport is exploring new services and service opportunities, the idea being to find new ways to monetize the rich shopping and booking information available to Travelport. Prototype software built by CERCS students has demonstrated that such monetization based on online data analytics can even be done online, by observing and then mining customer traffic.

> **Economic Impact:** The flight shopping/booking industry is undergoing a rapid evolution, in part because of new players like Google that recently acquired one of the three international flight shopping and booking services. Well-positioned to capitalize on increasing travel activities worldwide, Travelport operates both domestically and internationally. By helping local companies develop new services, we help protect and strengthen its financial position, creating jobs in the Atlanta area.

For more information, contact Karsten Schwan, 404.894.2589, schwan@cc.gatech.edu.

Power-Efficient Data Centers

In the U.S., data center facilities consume approximately 2% of all electricity, with an estimated growth rate of 12% per year. The majority of this data center power is either consumed by the computers themselves or by the HVAC units required to keep the computers from overheating.

The use of virtual machines has allowed data center managers to conserve power by consolidating data processing onto fewer servers in times low demand. However, that consolidation had not previously been performed while taking into account its impact upon the cooling demands of the data center.

The Georgia Tech CERCS project "CoolIT", developed in collaboration with the Mechanical Engineering department, allows the synergistic and cooperative management of IT and cooling system resources, adjusting air velocity and the location of active computing simultaneously to minimize power consumption. Initial results attained in CoolIT highlight the interesting trade-offs faced by a coordinated management solution. For instance, at lower cooling air velocities, for a homogeneous set of server systems, an awareness of hot spots in the data center permits the IT management system to operate at close to 100% maximum performance load, whereas without such awareness, there are situations in which only 40% load is achieved. These results illustrate the significant benefits of coordination between computer and facilities power management. Similar insights have been gained for the electrical distribution systems associated with machines, where unequal loads can cause inefficiencies in power delivery, for instance.

> **Economic Impact:** CoolIT researchers have interacted with the major IT vendors in the US. IBM's data center technologies group, for instance, has interacted with our research team to create new technology for data center thermal monitoring (a CoolIT student worked with the IBM team). There have also been interactions with smaller companies, one being OSISoft, a leading vendor in large-scale monitoring software, which has used its exposure to CoolIT to market its software in the (for the company) new domain of data center monitoring and management. Most recently, topics on IT system management have affected VMWare's product offerings, through joint work of CoolIT researchers with VMWare's cloud computing and data center management teams.

For more information, contact Karsten Schwan, 404.894.2589, schwan@cc.gatech.edu.