Chapter 5
Quality of Service Monitoring Strategies in Service Oriented Architecture Environments using Processor Hardware Performance Metrics

Ernest Sithole
University of Ulster at Coleraine, UK

Sally McClean
University of Ulster at Coleraine, UK

Bryan Scotney
University of Ulster at Coleraine, UK

Gerard Parr
University of Ulster at Coleraine, UK

Adrian Moore
University of Ulster at Coleraine, UK

Dave Bustard
University of Ulster at Coleraine, UK

Stephen Dawson
SAP Research Belfast, UK

ABSTRACT

The sharp growth in data-intensive applications such as social, professional networking and online commerce services, multimedia applications, as well as the convergence of mobile, wireless, and internet technologies, is greatly influencing the shape and makeup of on-demand enterprise computing environments.

In response to the global needs for on-demand computing services, a number of trends have emerged, one of which is the growth of computing infrastructures in terms of the number of computing node entities and the widening in geophysical distributions of deployed node elements. Another development has been the increased complexity in the technical composition of the business computing space due to the diversity of technologies that are employed in IT implementations. Given the huge scales in infrastructure sizes

DOI: 10.4018/978-1-61350-432-1.ch005

and data handling requirements, as well as the dispersion of compute nodes and technology disparities that are associated with emerging computing infrastructures, the task of quantifying performance for capacity planning, Service Level Agreement (SLA) enforcements, and Quality of Service (QoS) guarantees becomes very challenging to fulfil. In order to come up with a viable strategy for evaluating operational performance on computing nodes, we propose the use of on-chip registers called Performance Monitoring Counters (PMCs), which form part of the processor hardware. The use of PMC measurements is largely non-intrusive and highlights performance issues associated with runtime execution on the CPU hardware architecture. Our proposed strategy of employing PMC data thus overcomes major shortcomings of existing benchmarking approaches such as overheads in the software functionality and the inability to offer detailed insight into the various stages of CPU and memory hardware operation.

MAJOR DEVELOPMENTS AND CHALLENGES IN ON-DEMAND COMPUTING

Given the current developments in eEnterprise implementations, the infrastructure planning tasks of accurately determining performance that can be delivered by on-demand computing resources and in turn, obtaining accurate estimates of the appropriate infrastructure hardware performance capabilities required for business computing solutions are becoming an increasingly challenging exercise to undertake. One major cause that has led to the difficulty in quantifying performance in business computing systems has been the huge amounts of user-generated data as a result of the exponential adoption of on-demand hosted computing services and applications such as social networking, e-commerce, and multi-media content sharing services. Yet another development that has caused a huge increase in consumer-generated data is the convergence of mobile, wireless and web technologies into a ubiquitous computing platform. In response to the challenges arising from these major trends, there has been a phenomenal growth in the size of deployed computing infrastructures, with the magnitude of the infrastructure expansion being characterised by three main dimensions of growth: (a) the increases in the number of computing machines brought together to form server domains, (b) the wide geographic locations, over which participant server nodes are physically deployed and (c) the different types of technologies that are used to produce computing solutions.

Challenges for Performance Evaluation

Arising from the physical distribution of compute nodes due to the dispersion of resources in the infrastructures, are the performance-related challenges pertaining to the need to quantify network delays. The delays emanate from the communications of status and coordination messages as well as the actual data transfers between host machines. The calculation of overall performance metrics in distributed systems, which are dependent on network delays, is not straightforward to perform given that application routines running inside server nodes generate data in quantities that can vary dynamically. As a result of the changing loads, traffic levels introduced on network links usually follow irregular patterns leading to congestion and bandwidth delays that cannot be easily established. It is important to emphasize that while network delays do not feature in the proposed performance evaluation strategy considered in the latter sections of this chapter, they nevertheless make a key aspect which is captured in the Service Taxonomy presented in Figure 4.

Related Content

Cost Estimation Methods in Quality Management and Continuous Improvement: A Reflection of the Relationship
Brian J. Galli (2021). *International Journal of Service Science, Management, Engineering, and Technology (pp. 38-61).*
www.irma-international.org/article/cost-estimation-methods-in-quality-management-and-continuous-improvement/267179

Always On: Managing Online Communities in a 24/7 Society
Sarah Williams (2017). *Promotional Strategies and New Service Opportunities in Emerging Economies (pp. 314-328).*
www.irma-international.org/chapter/always-on/175560

Perceived Value and Brand Loyalty in Fine Dining Service
Jyoti Kainthand Harsh V. Verma (2013). *International Journal of Service Science, Management, Engineering, and Technology (pp. 1-12).*
www.irma-international.org/article/perceived-value-brand-loyalty-fine/77359

A Study on Relationships Among Software Engineering Capabilities in Japan Using Panel Analysis
Yasuo Kadono (2013). *Best Practices and New Perspectives in Service Science and Management (pp. 180-192).*
www.irma-international.org/chapter/study-relationships-among-software-engineering/74992

The Influence of UGC on Consumers’ Information Process on Service Failure
Te-Lin Chung (2012). *International Journal of E-Services and Mobile Applications (pp. 32-47).*
www.irma-international.org/article/influence-ugc-consumers-information-process/73715