

Business Objective Based Resource Management

S. Aiber D. Gilat A. Landau N. Razinkov A. Sela S. Wasserkrug

IBM Haifa Research Lab
Haifa University, Mount Carmel, Haifa 31905, Israel
972-4-8296211

{ sarel, dagang, ariel, natali, sela, segevw }@il.ibm.com

ABSTRACT

Enterprises today wish to manage their IT resources so as to optimize business objectives, such as income, rather than IT metrics, such as response times. Therefore, we introduce a new paradigm, which focuses on such business objective oriented resource management. Additionally, we define a general simulation-based autonomous process enabling such optimizations, and describe a case study, demonstrating the usefulness of such a process.

Categories & Subject Descriptors: C.4 [Performance of Systems]: Modeling techniques and Performance attributes. I.6 [Simulation and Modeling]: I.6.3 Applications and I.6.4 Model Validation and Analysis.

General Terms: Management, Measurement, Performance, Economics.

Keywords: IT Policy, Optimization, Simulation, Business Objective, Economic Considerations, Modeling Techniques.

1. INTRODUCTION

The major challenges in the runtime resource management in any enterprise are both the initial optimization of these resources, and keeping these resources optimized when changes occur. Traditional IT optimizations typically focus on IT measures (e.g., increase the site's availability by 1 %). Nowadays, however, many enterprises are keen on achieving business objectives, optimizing the total revenue generated by the infrastructure. Therefore, resource management should focus on these business goals, rather than the more conventional IT metrics.

2. RELATED WORKS

Current IT related runtime resource management efforts focus on optimizing IT level metrics, such as round-trip response times [1]. However, these IT metrics are not directly related to business level goals, such as revenue or return on investment (ROI). Some products attempt to manage IT so as to better align it with business level objectives. For an example, the Peakstone eAssurance product [2] attempts to manage a web site so that a set of service level agreements (SLAs) is not violated. However, we are unaware of any works in which a generalized approach, an architecture and technologies are introduced for runtime resource management according to high level business objectives.

3. OPTIMIZED RESOURCE MANAGEMENT

We define a general automatic optimization process implementing business objective based resource management. Figure 1 depicts this process. First of all, business goals are defined. Then, based on these definitions, both business rules detailing how IT level metrics affect these business goals, and SLAs (or other contracts), are generated. Then the following continuous loop is executed: 1. business objective based optimization is carried out (see the following sub-section). 2. The IT policy defined by the optimization process is enforced to the runtime environment (by effectors) that manages the resources of the IT infrastructure according to this policy. 3. Available monitoring tools monitor the IT resources, the conformance to the contracts, and the business objectives achieved by the infrastructure. 4. The Significant Change Detection (SCD) component uses the results of the monitoring process to recognize a significant enough change between the model and the business environment. 5. If a significant change is determined, the model used for the optimization is updated, and the loop begins again starting with the optimization process.

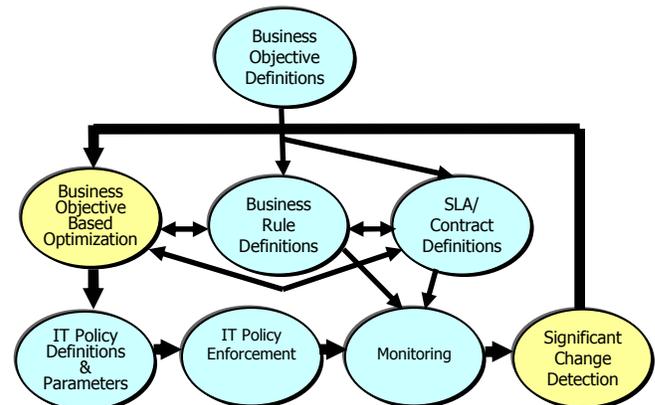


Figure 1. Business Objective Based Resource Optimization

3.1 Models for Business Objective Based Optimization

In order to carry out business objective based optimization, a model of the system is required. In our implementation, this model is composed of three main sub-models: a business level model, an IT model and an IT to business level impact analysis model.

The business level model supports the calculation of the business metrics, e.g., gains from commissions, explicit penalties paid to customers whenever SLAs are violated, etc. In order to enable

Copyright is held by the author/owner(s).
WWW 2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

optimization, this economic model should calculate an “end result”—a single quantity that can be used to quantify the alignment of the IT with the business objectives. We term this quantity the *overall business metric* (OBM). An example of such an OBM is the total income generated by the IT infrastructure.

The IT model is composed of the system model and the system user behavior model. The system model takes into account items such as the hardware configuration of the IT, and the software installed on this hardware. The system user behavior model takes into account the manner in which the users of the IT infrastructure use the systems supported by this infrastructure. In our current implementation, the system user behavior model and the system model are modeled using hybrid simulation models (i.e., simulation models combining both conventional detailed simulation modeling techniques, and other techniques).

The IT to business level impact analysis model defines how events at the IT level impact the business objectives defined by the business level model. Examples of this are how individual response times measured for a customer impact the penalty paid to a customer and the measurement of a customer’s dissatisfaction with poor response times.

In order to optimize the IT according to the business objectives, an optimizer is coupled with the model. The optimization process itself is a search over the space of possible actions/policies, which attempts to find the setting in this space that optimizes the business objectives. Figure 2 depicts the architecture of the optimization mechanism.

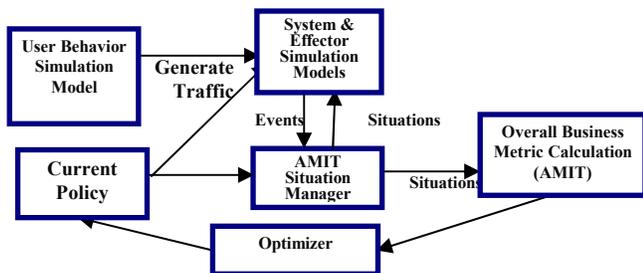


Figure 2. Resource Optimization Architecture

4. CASE STUDY

In order to validate the above approach, an e-Trading web site (i.e., a web site allowing the buying and selling of stocks on the internet) was optimized according to business objectives. This e-Trading web site was based on IBM’s Trade3 benchmark application [3], i.e., a three-tier J2EE e-Trading web application. The architecture of this site appears in Figure 3.

The business objectives in this case study were the following:

- For each buying/selling of a stock, a commission of maximum of 4% or \$25 is collected.
- Two types of SLAs exist: Platinum and Gold. A customer pays a daily flat fee of \$50 and of \$20 for Platinum and Gold SLA respectively.
- A customer is promised an average response time of 1 second for platinum and of 2 seconds for gold agreement respectively.
- For each 1% deviation from this response time, the site pays a penalty of \$5.5 or of \$3.5 for platinum and gold agreement respectively.

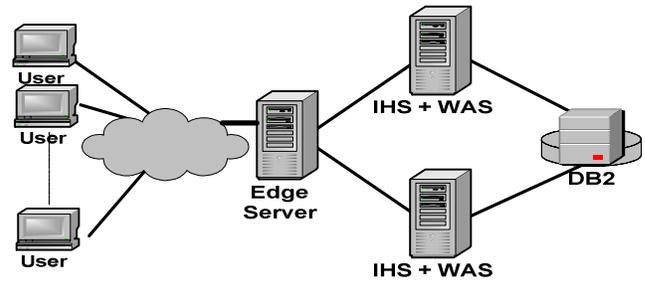


Figure 3. Use Case Configuration

The effector used in this case study was the bandwidth allocation component of IBM’s Edge Server’s TQoS component [4]. This component allows allocating the network bandwidth of the user’s HTTP requests, thereby allowing prioritizing the requests of one customer over another.

The edge server policy was optimized using the architecture described in Section 3.1. With no bandwidth limitation (i.e., all customers receive equal priorities, and, therefore, equal resource allocation), the total income was -\$475,376. The optimization process, after 7 iterations, found a policy improving the total income and brought it to \$179,088, i.e., changed an initial negative income to a positive value. Figure 4 shows these results.

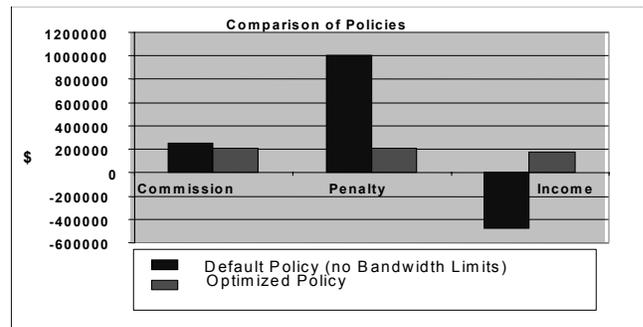


Figure 4. Comparison of Default & Optimized Policies

5. SUMMARY

This paper presents a process and algorithm for managing the IT resources of an IT business infrastructure according to high-level business objectives. This approach has the advantage of constantly keeping the infrastructure aligned with business objectives, and clearly connects IT related policy decisions to business level metrics such as profit or ROI. In addition, a case study demonstrating the usefulness of this approach was described.

6. REFERENCES

- [1] D. Menasce, D. Barbara, and R. Dodge, “Preserving QoS of E-commerce Sites Through Self-Tuning: A Performance Model Approach”, Proceedings of 2001 ACM Conference on E-commerce, Tampa, 2001, pp. 224–234.
- [2] <http://www.peakstone.com>
- [3] <http://www-3.ibm.com/software/webservers/appserv/benchmark3.html>
- [4] <http://www.ibm.com/software/webservers/edgeserver>