Capacity Planning Use Case: Mobile SMS
How one mobile operator uses BMC Capacity Management to avoid problems with a major revenue stream
## TABLE OF CONTENTS

**INTRODUCTION** ................................................................. 1

**THE SMS SERVICE ARCHITECTURE.** .................................... 2

**THE PROCESS** .................................................................... 2
  » Step 1: Infrastructure and business data collection: sources and metrics ........................................ 2
    • Infrastructure data ........................................................ 2
    • Business data .............................................................. 2
  » Step 2: Data analysis / model building .................................. 3
    • Performance and load analysis ........................................ 3
    • Correlating business and performance – the key to building capacity models ......................... 4
  » Step 3: "What-if" analysis .................................................... 6
  » Step 4: Capacity alert setup ................................................ 7

**CONCLUSION** ................................................................. 8
INTRODUCTION

A large mobile operator uses BMC Capacity Management to track the performance and capacity of its SMS service – a major contributor to its subscriber revenue. BMC Capacity Management allows the operator to locate bottlenecks and find the most appropriate solution through "what-if" analyses. Thanks to BMC Capacity Management, the operator is also able to identify the resources required to send a single SMS, and the average load of a single subscriber. As a result, the effects of marketing campaigns designed to increase the subscriber base or increase the number of SMS messages sent can be calculated simply and accurately.

This document describes the capacity management process using BMC Capacity Management for this use case.
THE SMS SERVICE ARCHITECTURE
The mobile operator’s SMS architecture is made of a dedicated application server cluster (totaling eight CPUs) running a UNIX® operating system. It is connected to a shared database cluster running Oracle® on UNIX. There are also other ancillary systems that, for the sake of simplicity, we will consider out of scope.

THE PROCESS
The capacity planning process supported by BMC Capacity Management is made of the following logical steps:

1. Infrastructure and business data collection
2. Data analysis / model building
3. “What-if” scenario simulation
4. Capacity alert setup

STEP 1: INFRASTRUCTURE AND BUSINESS DATA COLLECTION: SOURCES AND METRICS

Infrastructure data
BMC Capacity Management can import performance data from the mobile operator’s multiple network, server, and database management applications. For example, according to the operator’s internal customer policies, each layer of the SMS service is monitored and managed using a specific tool: HP OpenView, Oracle’s Enterprise Manager, and EMC’s Storage ECC. BMC Capacity Management allows data from all these management tools to be consolidated into a single repository, aka Capacity Database.

Relevant performance metrics include:

» CPU and memory utilization for each server
» File System status on SAN and local discs
» Database utilization and storage occupation

Business data
Business metrics are imported from OSS/BSS systems using custom connectors that can easily created by configuring general-purpose SQL connectors provided by BMC Capacity Management. Relevant metrics of the SMS application are:

» Number of subscribers (hourly)
» Number of SMS processed (hourly), detailed by type of service

All data gathered is summarized by BMC Capacity Management to provide homogenous time granularity (hourly). Some sources, such as the server management systems, provide information more frequently. Data that is older than a configurable threshold (typically nine months to one year) is further summarized into daily granularity, with an option to store the hourly average for each day, if desired.
STEP 2: DATA ANALYSIS / MODEL BUILDING

Performance and load analysis

The first analysis provided consists of performance and load graphs for the time periods of interest. BMC Capacity Management builds most of these graphs automatically and guides the mobile operator’s users through the process of finding correlations through intuitive wizards.

Figure 1. Breakdown of total load on the SMS System based on four types of service offered during previous six months.

Figure 2. Hourly drill-down of total SMS sent during the last 30-day period.

Figure 1 shows the breakdown of the total load on the SMS system based on four types of service offered (S1, S2, S3, S4) during the previous six months. The graph displays two important aspects: a) there is a clear weekday/weekend pattern; b) the vast majority of load is determined by two services (the red and green ones).
Figure 2 shows an hourly drill-down of total SMS sent during the last 30-day period. It also shows that SMS are sent more frequently at some hours of the day than at others.

The moving average line, however, shows no obvious trend, indicating that while some hours see much higher volumes than others, there is no particular trend in total SMS volume, something that we can also see from the longer term graph in Figure 1.

![SMS Server 1 CPU, Mem, Disk (hourly)](image)

Figure 3. Three critical server metrics for one server used for the SMS application over a month

Figure 3 looks at performance metrics rather than load, showing a combination graph of the three critical server metrics for one server used for the SMS application over a month. The red line on the graph represents the 70 percent maximum utilization threshold according to the policy set by system administrators for this class of servers. As can be seen, disk utilization is minimal, while memory utilization runs at an average of around 50 percent. Neither disk nor memory utilization seems to be in danger of becoming capacity constrained. CPU utilization, on the other hand, is peaking at over 90 percent for certain periods of the day — not just once, but for many days — implying that the system is approaching saturation. Thus it becomes clear that, for this server, the CPU is the critical metric, and hence, CPU performance should be increased.

**Correlating business and performance – the key to building capacity models**

Performance vs. load analysis indicates a relationship between average CPU utilization (per hour) and the number of SMS sent in the same period. Thus, to identify the capacity of the SMS system, it is necessary to determine the average CPU impact of a single SMS. BMC Capacity Management is able to display the two metrics and calculate the best correlation between them. The identified correlation is then automatically fed into the modeling engine.
Figure 4 shows the plot of CPU utilization vs. SMS sent. BMC Capacity Management has been able to identify a linear correlation between the two metrics. The correlation is not perfect — there are some hours when average CPU utilization varies from the trend line — but it is good enough to use as a basis for extrapolation. The extrapolation is then shown in Figure 5, where we identify the number of SMS sent per hour, which leads to CPU utilization hitting 70 percent. It is also notable that the system has a base load of around 50 percent CPU utilization even when no SMS are sent. This is due to a second application running on the same system (outside the scope of this study).

This graph is of cluster performance, rather than the performance of an individual server as was shown previously in the load section.
STEP 3: “WHAT-IF” ANALYSIS

CPU performance is already unacceptably high for the current workload of the mobile operator’s SMS application, causing frequent violation of company performance policies. Furthermore, the goal set by the operator’s marketing department was the ability to handle one million SMS during peak periods. By using the modeling capabilities of BMC Capacity Management, it is possible to perform a “what-if” analysis to see how different upgrades affect the CPU bottleneck just discovered. The SMS system uses a cluster of eight servers, and a possible improvement would be to add one or more CPUs to the cluster.

Figure 6 shows the effect on performance of adding four CPUs to the cluster. As the graph shows, with four additional CPUs, the SMS capacity would now be close to one million SMS per hour, rather than the limit of 400,000 or so for the existing eight-CPU cluster.

Using BMC, the mobile operator’s capacity management team was also able to examine the possibility of adding one, two, or three additional CPUs. In the end, the decision was made to add four because the operator was expecting additional SMS volume growth.
STEP 4: CAPACITY ALERT SETUP

Clearly, discovering the CPU limit for the SMS application was critical. However, analysis was not only aimed at existing performance issues. The BMC Capacity Management solution’s ability to set performance alerts through modeling and trend forecasting also allows future performance issues to be reliably predicted.

Figure 7. Historical growth in storage space needed for the database of the SMS application over a five-month period

Figure 7 illustrates the historical growth in storage space needed for the database of the SMS application over a five-month period. Although there is a drop in December, caused by the purging of some older data, the size of the database is gradually growing. The three lines show the maximum size, the allocated size, and the used size.

Figure 8. Forecasted tablespace utilization of the database plotted against the maximum allocated filesystem space reserved to the SMS application
Figure 8 shows the forecasted tablespace utilization of the database plotted against the maximum allocated filesystem space reserved to the SMS application. The red area on the right of the graph shows the critical date for the system. Planning an update of the storage/database configuration at a suitable time before then avoids what could otherwise be a severe business outage.

CONCLUSION

Based on reports such as this, BMC Capacity Management is able to automatically generate alerts warning the relevant IT personnel that a limit will be hit within a configurable time interval — in this case, six months. As a result, IT management can schedule purchases, maintenance windows, and upgrades in a timely fashion, thus reducing the likelihood of either unplanned outages or wasteful capital expenditure due to advance over-provisioning.

To learn more about BMC Capacity Management, please visit www.bmc.com/capacitymanagement.