Performance Engineering Cookbook
Ingredients for Performance and Capacity Success

Metrics: Are you measuring up?

Peter van Eijk

This is a series of brief articles explaining the basic concepts of systems performance and capacity planning. Motivated by the Computer Measurement Group, these concepts are applicable to IT systems and beyond.

Metrics

Metrics is about quantifying our systems and their outcomes. In our series we have already encountered metrics for capacity, availability and response time. In this issue we will focus on the 'demand' side of those metrics. Who wants to know, and why? These stakeholders often require so-called KPIs, Key Performance Indicators. Let's begin by dissecting that expression.

An indicator is anything that ascribes some value to something, whether observed or inferred. For example, we can see that a certain server has the color brown or yellow. We can infer that a server has a CPU utilization of 75%. Neither of these has much business significance in itself. For example, a high server utilization is probably bad in an interactive environment where it increases the response time for users, but could actually be considered low in a batch environment. In fact, when the batch job is CPU intensive, a low CPU utilization is bad.

So we use the word 'performance indicator' generally if there is a direction that we judge to be better. Low interactive response times are good, low transaction throughput rates are bad. Often there is a 'good enough' limit, beyond which improvement is no longer relevant. E.g. humans can hardly observe response times smaller than 50 milliseconds.

Finally, what are 'key performance indicators'? Imagine a small sized data center with a couple of hundred servers. Each server and its network interfaces can have dozens of performance indicators. None of these thousands of indicators would be considered very relevant by an external stakeholder.

What is a good indicator? Key performance indicators have significance to stakeholders: they are simple and few, and express the concerns of the stakeholder in such a way that the indicator either reassures, or leads to clear action. A good example would be "time before we run out of capacity", if you are in a growth scenario.

A common pitfall is to use performance indicators that are easy to measure and report, rather than performance indicators that are meaningful to the stakeholders. Utilization is, once again, a good example of that, because even though utilization may be easier to measure, response time is more meaningful for the users.

Link farm

Wikipedia: http://en.wikipedia.org/wiki/Performance_metric (much wider than IT, with some valid points)
http://en.wikipedia.org/wiki/Metrics_(networking) (very much in need of a clean-up, so feel free to volunteer in working on that)

Note to readers: are there any concepts here that need further elaboration? We want volunteers to find more link worthy pages in sources such as the CMG archives, Wikipedia, and for linking back from Wikipedia to these pages. Please write to the author: Dr. Peter HJ van Eijk at pveijk@nlcmg.nl.