

Performance Engineering Cookbook

Ingredients for Performance and Capacity Success

Peter HJ van Eijk, PhD.

5. The impact of downtime: where does it hurt?

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.

Impact of downtime

Downtime is the opposite of availability, which we discussed earlier. In this installment we look at the cost of downtime for the business and its users.

Downtime annoys users and interrupts business processes, which leads to lost revenue, lost savings opportunity, or enhanced risk. A good monetary valuation of the impact of downtime can be used to support a business case for investing in more capacity or redundancy of an IT system. But contrary to popular believe, there is no objective measure of impact cost.

When developing business cases, I have learned that the most important aspect of them is that all monetary costs are believable and acceptable to the stakeholder who is asked to make the investment. This is nothing specific to IT. Approaches can also differ around the valuation of physical stock (historical, current market, etc).

For example, what is the cost of an ecommerce site not being available for 1% of the time? One approximation is that it is 1% of profit (or would revenue be better?). However, if prospective customers faced with a broken site switch to a competitor site and won't return ever, the losses could be substantially higher. Take whatever argument your stakeholders accept.

Websites are often used to enable self-service for customers. Downtime can then imply that the customer will call or email instead, which might incur a substantially higher cost per transaction. This could be termed 'substitution cost'.

A way to quantify customer 'nuisance' that goes well with some marketing and sales people is to ask which gift or 'goodie' would compensate the customer for her suffering. Each unavailable transaction would then incur this cost. This can be called the 'cost of recovery' approach.

Downtime costs often are not linear in time. A brief outage might be easily survivable, but as the outage becomes longer, more and more slack in the environment is used up, deadlines start to be missed, and compensating mechanisms for resilience break down. It is like having no air. You'll survive for a few seconds without any subsequent permanent damage, or maybe even a minute or two, but not much longer.

For a more extensive discussion see my paper 'The true cost of downtime' available in the CMG 2011 conference proceedings.

Link farm

Wikipedia: http://en.wikipedia.org/wiki/Business_continuity and http://en.wikipedia.org/wiki/Disaster_recovery give some more substance to avoiding and handling downtime and its cost. See http://en.wikipedia.org/wiki/Business_case for more extensive methodology in developing business cases.

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.