

Performance Engineering Cookbook

Ingredients for Performance and Capacity Success

9. Reporting: 'check please'

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. Brought to you by Dr. Peter HJ van Eijk.

Reporting and stakeholders

A lot of performance engineering is about numbers. Though we have not discussed monitoring techniques and reporting styles, we have covered the basics of the numbers that matter. The question is: to whom do they matter? In this note we will highlight the various audiences that have a stake in the systems under consideration and their very differing needs. We will look at this by moving up timescales. It is important to tailor and minimize the information that is presented to a specific stakeholder. This speeds up their actions, and avoids 'data deluge' and 'analysis paralysis'.

Incident response. Most urgent in time is to know when a system is not performing well enough, and the provider is in danger of violating his SLA (Service Level Agreement). The essential data should point incident response staff to predefined actions. Typically you would alert on deviations of customer service levels, such as web server response time, as well as on component malfunction and capacity thresholds of high risk infrastructure components, such as disk space. Each alert should have a predefined response or escalation with a pointer to further information. Nothing more. A set of red/amber/green lights should suffice.

Problem management needs the capability to drill down in the chain of transactions (see previous note on transaction structures). More detailed metrics are necessary to investigate the root cause of a problem, though the time scope is likely to be limited to a few hours at most.

SLA management needs reports that describe the customer facing performance on a longer time scale, up to several months typically. These people need information that enables them to discuss with clients sources of SLA violations or other customer dissatisfaction. Additional information needed would be service volumes and the performance of any subcontractors under the control of the customer (i.e. the customer's internet provider).

Systems architects and other engineers also need longer term reports for evaluating the system architecture. Is the system performing optimally? How can sources of SLA violations be engineered away? Are capacity allocations and cost structure optimal? These people need the SLA reports, but also the capability to drill down into the performance and capacity of components as well as zoom in and out in the time scale. They would also need exports of data to feed into their models.

As you develop reporting systems, don't get carried away in trying to make it perfect. Focus on the most imminent information for the specific audience. As you do this you should bear in mind that the usefulness of these reports is highest in the early phases of the lifetime of the systems they report upon. Any half report is better than no report or the wrong report.

Link farm

Wikipedia: http://en.wikipedia.org/wiki/Service-level_agreement

Anything missing in this series, or something you don't agree with? Please write to the author.