

Synopsis – *Performance Assurance for IT Systems* by Brian King

ISBN: 0849327784 (Published by Auerbach August 2004)

Introduction

Performance Assurance for IT Systems is a comprehensive guide to the approaches and techniques for addressing IT performance and related technology issues over the system lifecycle. Literature on this subject is extremely scarce; it is frequently limited to advice on tuning the system after it has gone into production. It is difficult to find any information or guidance on addressing the subject during that part of the lifecycle before a system goes into production, which is invariably the time when key decisions have to be made with respect to the technical architecture and application design. *Performance Assurance for IT Systems* has been written by someone with many years of industry experience in addressing performance and technology issues throughout the system lifecycle, on a wide variety of projects and technologies. It contains much practical advice, easily applicable technique and useful reference material.

Summary

IT Performance is often treated as a standalone topic whose lifecycle starts around the time that a system goes into production. It is therefore hardly surprising that the performance and scalability of new applications is frequently found to be unsatisfactory. In severe cases it can even endanger an entire business. Although problems can occasionally be remedied through a modest amount of tuning, or via a moderate hardware upgrade, this is only likely to be the case on smaller systems. Medium to large systems tend to suffer from more fundamental problems, ranging from: woefully inadequate volumetric information on which to base the original solution, an under-sized technical architecture, through to poor design / product selection and implementation. Rectifying such problems can frequently be expensive and time-consuming.

Ultimately, performance is simply a by-product of: the business requirements and associated volumetrics, the technical architecture, any employed software products, application design and implementation. In theory, the majority of post-production problems can be pre-empted by understanding the issues and tackling them before the system goes into production. The process should begin during the feasibility or procurement stages. Unfortunately, there is frequently a scarcity of concrete information during the early stages, which induces apprehension and a lack of willingness to address the subject. Hardware sizing is an obvious example. It is often regarded as a 'black art', as fear of the unknown often prevents the subject from being tackled. The ever more aggressive time scales that projects are faced with can lead to snap decisions that result in the configuration of inappropriate hardware, thus feeding the fear and forming a vicious circle.

Books that cover performance tend to be vendor or product specific, and they often assume that the system already exists, i.e. it is either already in production or about to go into production. More general publications focus on highly technical topics, such

as queuing theory. Material that covers the real-life issues that are encountered during the early stages, when the solution is being formulated and important decisions have to be made, is extremely scarce.

Sizing and performance can be tackled pre-emptively in a focused and cost-effective manner. *Performance Assurance for IT Systems* is a critical book that discusses the issues, approaches and techniques that can be adopted throughout the system lifecycle, from reviewing hardware and software products at the feasibility stage, estimating capacity requirements with minimal information, designing and building for performance and scalability, through to stress testing and post-production aspects. It includes discussion on effective procurements from the perspective of both the purchaser and the bidder. The importance of understanding the performance capabilities and potential constraints of hardware and software technologies is emphasised by the inclusion of a variety of “technology tasters” that briefly summarise a given topic and highlight the potential performance issues. The watchword throughout the book is “fit for purpose”, as the emphasis is on pragmatism, a necessary prerequisite in this day and age when there is increasing pressure to deliver in ever shorter timescales.

Contents

The book is split into two parts; the first part addresses individual areas of interest, roughly following the system lifecycle. The topics comprise:

Preparing For The Challenge argues that performance and technology issues should be addressed proactively throughout the system lifecycle: from initial investigations, through the bid or feasibility study, through development, to post-production. It introduces the skills that are required during the various pre-production stages, and it summarises the concept of Performance Assurance, an approach to tackling performance proactively throughout the development lifecycle.

Caveat Emptor (Let the buyer beware) looks at the issues that surround the review and selection of hardware and software products, including the discussion of typical product lifecycles and vendor marketing approaches. It provides guidance on a focused approach to such reviews, particularly of software products.

Lies, Damned Lies and Benchmarks discusses the problems that surround the use of benchmarks to aid the sizing of a system, in particular the methods that can be employed, knowingly or unknowingly, to produce highly optimistic, unfeasible results. Typical “tricks of the trade” that are often used by vendors are outlined. It discusses how to approach in-house benchmarks, or benchmarks that are mandated from potential vendors, to ensure that they provide useful results.

Non-Functional Requirements and Solutions looks at the effects of poorly written requirements on the tendering process. It is of particular interest to the authors of ITTs, discussing the benefits of clear, unambiguous, and realistic requirements. In addition, it discusses general approaches to hardware sizing and deployment, plus the implications of high availability and business continuity requirements on the complexity and cost of a solution.

Hardware Sizing: The Crystal Ball Gazing Act focuses on the problems and potential approaches to this difficult subject when there is limited information to work with early in the system lifecycle. It is an area that many people seek to avoid because of the scarcity of information. A worked example is included, along with guidelines and useful rules of thumb.

Bids can be stressful, especially for people, such as the Hardware Sizer, who are at the end of the solution chain and who consequently have extremely limited time to complete their work. **Bid Logistics** contains advice for tackling bids in a coherent and pre-emptive manner, with the emphasis on catering for the shorter procurement cycles that are becoming more prevalent.

Awareness Of Queues discusses the need to understand where they form, and their effects on performance, particularly the knock-on effects when there are networks or hierarchies of inter-related queues. It demonstrates that software queues and locks can easily prevent the full exploitation of the available hardware. A brief introduction to performance modelling tools is included, highlighting their advantages and disadvantages.

Designing for Performance, Scalability and Resilience commences with a review of some of the basic principles, including: the need to avoid the extravagant use of hardware resources; partitioning techniques; the need for balanced configurations; and catering for resilience. A range of design factors that can affect performance are discussed, along with suggestions for possible approaches to effective design.

Non-Functional Acceptance focuses primarily on stress testing, covering the perceived need for such testing, the tools that can help, and general guidance on the preparation, set up and running of the tests. Soak Testing and Operational-Readiness Testing are also briefly covered.

Post-Implementation provides an overview of the classic Capacity Management processes that have been used for the last 20+ years. It goes on to discuss a minimalist approach that will be appropriate for those organisations that have limited resources to cover the subject in any detail.

Softer Issues deals with non-technical items. Personal skills are frequently as important as technical skills in producing effective solutions within ever more aggressive time scales. Many areas are touched on, including: communication; the danger of over-engineering solutions; playing devil's advocate; awareness of cost and time constraints; team balance; mentoring; and client responsibilities.

So What looks at the approach where performance and scalability are not addressed, typically on cost grounds, covering the risks that are inherent in such an approach. It affirms the author's view that it is possible to tackle these subjects in a focused, cost-effective manner that will allow management to make informed decisions.

The second part of the book contains reference material that has been called **Technology Tasters**. The foundation for sizing and performance work is a solid understanding of hardware and software technologies. All technologies have 4-5 key technical factors that heavily influence performance and scalability. These sound bites attempt to distil the key points of a given technology. They are in no way meant to

replace a comprehensive understanding that can only be obtained by reading specific books, papers, and vendor material on a given subject matter. For example, at the time of writing the author has in excess of 100 detailed papers on CPUs and server technology, in addition to several books that include sections that cover some of the same ground. The objectives of the tasters are simply to provide a useful, initial understanding of a technology and its related performance issues that the reader may not be fully familiar with. The tasters are:

- Operating Systems
- CPU Basics
- Memory
- Hard Disk Basics
- Server Infrastructure
- Multiprocessor Systems (Shared Memory)
- Hard Disk: Fibre Channel, SAN and NAS
- File Access Basics
- Relational Database Basics
- Back-end Server Clusters
- Network Basics
- Firewalls
- Server Load Balancing
- Web servers and Cache servers
- LDAP servers
- Modern Development Technologies (primarily .NET and J2EE)

Target Audience

The book is aimed at a general IT audience. To that end, conscious efforts have been made to minimise the use of jargon.

Technical Architects and Performance Analysts who have limited experience of addressing hardware sizing, performance and scalability issues before a system goes into production will find much useful information. There are so few people who are prepared to tackle hardware sizing early in the lifecycle that this topic on its own should be valuable to most people. Discussions on procurement issues will be helpful, particularly for those who are inexperienced in either writing or responding to an Invitation To Tender (ITT), while the technology tasters will provide a slightly different perspective on those technologies that the reader may already be familiar with, plus a useful introduction to unfamiliar topics.

Application developers will gain a sound and extremely useful appreciation of a variety of non-functional issues, helping them to appreciate the potential effects of decisions that they make during development, particularly in the areas of software product selection and application design.

Part 1 provides awareness of some of the non-functional issues that **IT management** is likely to encounter. Project Managers, Bid Managers, IT Middle Management, and Sales & Marketing can all find useful information to aid an understanding of the issues that their technical colleagues are confronted with. In addition, the technology

tasters can provide an introduction to individual topics that they may be encountering for the first time, helping them to make informed decisions on technical risk.

The Author

Brian King has worked in the IT industry for 37 years. Up to 1985, he had spent 11 years in Application Development (of commercial systems) and 7 years in Technical Services, including the development and modification of low-level systems software. He joined a large UK systems house in 1985, specialising in hardware sizing and performance activities for new applications, from the initial bid stage, through development, to production launch. Other work included: Pre and Post Sales Support for an analytic performance modelling tool; helping clients to introduce Capacity Management systems; and performing many individual Capacity Planning studies for production systems. He acted as the champion for a sizing and performance initiative within the company, where the objective was to raise the general skill level of the technical consultants. He has lectured and given courses on various aspects of Capacity Management, particularly sizing and performance. He became a freelance consultant in 1996, specialising in technical infrastructure solutions during the bid and design stages, including: vendor and product evaluations; hardware sizing; and designing for performance, scalability, and resilience. Since 1985, he has undertaken circa 450 technical assignments. His hobbies include theatre, art, reading and gardening.

Contact

Brian King
9 Vernon Drive, Ascot, Berks, SL5 8TN, UK
Telephone +44 (0)1344 291872, Mobile +44 (0)7770 898873
Email: b.king@dsl.pipex.com