DETERMINING THE TEST FOCUS THROUGH RISK ASSESSMENT

Ross Collard, Collard & Company
Presented to WOPR2, April 15, 2004

Rationale for Risk Assessment

We cannot test everything, so we need to make conscious decisions about where to focus the depth and intensity of testing. Typically the most fruitful 10% to 15% of the test scenarios uncover 75% to 90% of the significant problems (Petschenik).

Risk assessment provides a mechanism with which to prioritize the test effort. The questions which risk assessment help answer are (1) where to direct the most intense and deep test efforts, and (2) where to deliberately test lightly, in order to conserve resources for the areas which need intense testing. Risk-based testing can find more of the significant problems quickly and early, by testing only the most risky aspects of a system first.

The Case Study

I recently participated in a performance and robustness testing project, which provided an opportunity to apply and refine my risk assessment methodology in order to determine the test focus. The business and technical situation is documented in a series of appendices (see later), which I use in teaching classes on performance testing.

The situation described in the appendices is fictional. I do not have permission to reveal the firm's identity. However, the situation and issues described in the appendices though disguised are close to the real situation.

Major Risk Factors

In my experience, most system performance and robustness problems occur in these areas (which are not necessarily listed in order of importance):

- Resource-intensive features.
- Timing-critical or timing-sensitive uses.
- Likely bottlenecks (based on the internal architecture and implementation).
- Customer or user impact, including visibility.
- Prior defect history (observations of other similar systems in live operation).

- What's new and modified from the prior system version.
- Heavy demand: heavily used features.
- Complex features.
- Exceptions.
- Troublesome (poorly built or maintained) portions of the system.

Risk Identification Questions

I have developed this generic list, which may need to be modified to fit the particular context:

1) What circumstances are likely to cause heavy demand on the system from external users (i.e., remote visitors to a public web site, who are not internal employees)?

2) Under what circumstances is heavy internal demand likely (i.e., by the internal employees of a web site)?

3) What uses of the system are likely to consume a high level of system resources per event, regardless of how frequently the event occurs? The resource consumption should be significant for each event, not high in aggregate simply because the event happens frequently and thus the total number of events is high.

4) What system uses are timing-critical or timing-sensitive?

5) What uses are most popular, i.e., they frequently happen?

6) What uses are most conspicuous, i.e., have high visibility?

7) Based on your understanding of the system architecture and support infrastructure, where are the likely bottlenecks?

8) What specifically is new or changed in the coming version of the system or its support infrastructure? Areas which are new or modified are more likely to have performance issues than areas which have already been running satisfactorily and have not been touched. However, if most or all of the system is new, everything is at risk and answering this question will not help.

9) What has been your prior experience with other similar situations? Which features, design styles, subsystems, components or systems aspects typically have encountered performance problems? If you have no experience with other similar

systems, skip this question.

10) Are there any notably complex functions in the system, for example, in the area of exception handling?

11) Are there any areas in which new and immature technologies have been used, or unknown and untried methodologies?

12) Are there any other background applications which share the same infrastructure and are expected to interfere or compete significantly for system resources (e.g., shared servers)?

13) What is the architects' and developers' level of confidence in the system's adequacy? Do we know where in the system these people feel comfortable that performance will not be an issue, and in which areas are they nervous? I am assuming that the testers have access to the architects and designers – if not, this question and the next one may be unanswerable and thus irrelevant.

14) What are the architects' and developers' reputations for delivering systems which fail to meet the performance and robustness goals, and their credibility in spotting potential problems? Since these people usually understand the system internals better than anyone else, their suggestions could be invaluable – but only if they know what they are talking about.

15) What can we learn from the behavior of the existing systems that are being replaced, such as their work loads and performance characteristics? How can we apply this information in testing the new system?

16) What areas of the system operation, if they have inadequate performance, most impact the bottom line (revenues and profits)?

17) What *combinations* of the factors, which you identified by answering the previous questions, deserve a high test priority? What activities are (a) likely to happen concurrently, and (b) cause heavy load and stress on the system?

18) What areas of the system can be minimally tested for performance without imprudently increasing risk, in order to conserve the test resources for the areas which need heavy testing?

19) In summary, considering the total picture, what areas should the performance test focus on? Consolidate your answers to the prior questions in this exercise to form an answer for this question, by completing the following table.

Instructions for the Risk Prioritization Work Sheet

First list the areas to be tested, which are the results of the risk identification exercise, as

a series of row titles in a table (see later for an example). You can categorize the areas in various ways, depending of what feels comfortable for you. Generally these areas can be lifted directly from your answers to the previous questions. An "area" can be a system feature or set of related features (such as an on-line search), an event (such as a sales promotion), a part of the system architecture (such as a firewall), and so on. Expect to list between about 10 and 25 areas in your table.

Second, use a five-point scale to assess the importance of each entry in the table, ranging from 5 – high, 4 – moderate-to-low, 3 – moderate, 2 – moderate-to-low, to 1 – low. Use a question mark ("?") in place of a score when you are uncertain about an entry. Since uncertainty is a sign of risk, the question mark should be considered equivalent to a score of at least 3 or 4 (moderate or moderate-to-high). Exposure is the combination of probability and cost – do not add or multiply their numbers together in order to compute the exposure, just assign what you think is the right weight to their combined importance. In a similar way, the priority is the combination of your assessments of the exposure and the relative ease of testing.

Third, jot down your assumptions and rationale behind each score you assign, if you do not think these assumptions will be obvious.

I have filled in a few rows of the table to give you a sense of what I am asking you to do, but you may choose to disagree with my assessments (the sample entries in the table).

Do not bother to include areas which are obviously low priority, unless you need to think the situation through and fill in the entries in order to recognize or confirm the low priority. The asterisk (*) in this following table example indicates a row that you probably would not bother to include, but has been included here to help make the example well rounded. In practice, most areas of low priority are filtered out and not listed, to reduce clutter and render the table manageable.

Risk Prioritization Work Sheet Example

Area to be Tested	Likelihood or Probability of Performance Problems in this Area	Likely Cost or Consequences of the Performance Problems	Exposure (Combined Importance of Likelihood and Consequences)	Relative Ease of Testing in this Area	Test Priority for this Area
Database maintenance (updates, re-indexing, and back-ups)	High (5), because this is a highly data-dependent system with a data-centric system architecture.	High (5), because the database performance affects all parts of the system operation and is highly visible.	High (5), as this is based on the combined effect of the entries in the two columns to the left.	Moderate to high (4), as automated test cases already are available, and a tool is being acquired to run them.	High (5), as this is based on the combined effect of the entries in the two columns to the left.
Denial of service (DOS) attack	Low (1), because an attack is assumed to be unlikely.	High (5), because without adequate DOS controls an attack will shut the system down.	Moderate (3)	Moderate to low (2), because a large test load must be generated and delivered.	Moderate (3)
Incoming telephone calls to the call center, after a promotion	Moderate to high (4), but expected to decline over time as more people switch to directly ordering via the Internet.	Moderate to high (4), because people are sensitive to phone delays. Members will be irritated and business lost. However, non-telephone work is not affected.	Moderate to high (4)	Low (1), because a small army of testers are needed to manually place phone calls, or specialized, expensive call generators.	Moderate to high (4)
Transaction which changes the quantity on an order (*)	Low (1), because this is a low-volume transaction which uses few resources.	Low (1)	Low (1)	High (5)	Moderate to low (2)

Answers to the Risk Identification Questions

The answers we developed for the book club system are:

1) What circumstances are likely to cause heavy demand on the system from external users (non-employees)?

A promotion, particularly the surge of orders stimulated by a promotion.
A newly available e-catalogue, when a large number of people download during a small interval of time.
A denial of service attack.

2) What circumstances are likely to cause heavy internal demand (by book club employees)?

Printing the paper version of the catalogue to mail to customers.
Management ad hoc reporting.

3) What uses of the system are likely to consume a high level of system resources per event, regardless of how frequently the event occurs?

A book search.
Data mining (not part of the order entry system, but runs on the same shared servers).
Transmission of graphics and video clips of authors and their books.
Database maintenance, such as updates, re-organizations and backups (the nightly "crawl").
On-line order processing.

4) What system uses are timing sensitive?

A call center phone call.
Database access in support of a phone call.
Senior management and VIP queries (urgent or high priority ones).

5) What uses are most popular, i.e., frequently happen?

Home page hits and home page downloads.

6) What uses are most conspicuous, i.e., have high visibility?

Home page hits and home page downloads.

7) Based on your understanding of the internal system architecture, where are the likely bottlenecks?

The centralized database, shared across multiple applications.
Database maintenance, such as the nightly database back-up.
Shared servers across multiple applications.
Printing large jobs such as the catalogs: the printers are not hard wired, but on the common local area network (LAN).
Voice telephone calls: the voice servers (using computer-telephone integration or CTI technology), also are on the common LAN.
The firewalls: there is no separate front-end proxy servers; the firewalls reside on the web servers.
The T1 lines to the outside world. The Web access and the voice call center share the same T1 lines.

8) What specifically is new or changed in the coming version of the system or its support architecture?

It is a new system -- nothing can be singled out as being added or modified, because it is all new.

9) What has been your prior experience with other similar situations? Which features or systems aspects typically have encountered performance problems?

Promotions.
Video clips and graphics.
Management ad hoc reporting.

10) Are there any notably complex functions in the system?

Book search.
Database backup.
Management ad hoc reporting.
Return of an order (includes re-stocking and processing a credit to reverse a credit).

11) Are there any areas in which new and immature technologies have been used, or unknown and untried methodologies?

Wireless is new to the system architects and the book club.

12) Are the any other background applications which share the same infrastructure and are expected to interfere or compete significantly for system resources?

Billing system.
Company e-mail.

13) What is the architects' and developers' level of confidence in the system's adequacy? Where in the system do these people feel comfortable that performance will not be an issue, and in which areas are they nervous?

No information on this topic is available from the architects and developers.

14) Also, what is the architects' and developers' reputation for delivering systems which fail to meet performance goals, and their credibility in spotting potential problems?

No information is available.

15) What can we learn from the behavior of the existing systems that are being replaced, such as their work loads and performance characteristics? How can we apply this information in testing the new system?

The architecture and work load patterns of the new systems are not similar enough to draw meaningful parallels.

16) What areas of the system operation, if they have inadequate performance, most impact the bottom line (revenues and profits)?

Unavailability of the web site.
Inability to place an order on-line.

17) What combinations of the factors, which you identified by answering the previous questions, deserve a high test priority? What activities are (a) likely to happen concurrently, and (b) cause heavy load and stress on the systems?

To be debated. We'll leave this answer to the group debate.

18) What areas of the system can be minimally tested for performance without imprudently increasing risk, in order to conserve the test resources for the areas which need heavy testing?

Any area or aspect of the system that was not mentioned in the answers to the previous questions falls into this category.

19) In summary, what areas should the performance test focus on? Consolidate your answers to the prior questions in this exercise, to form an answer for this question.

High Priority Test Scenarios

The most fruitful scenarios, based on the risk identification, are:

1. Heavy external demand
   1.1. Promotion and new e-catalogue
   1.2. Denial of service attack

2. Heavy internal demand

2.1. Printing the paper catalogue
2.2. Management ad hoc reporting

3. High resource consumption per event
   3.1. Book search
   3.2. Data mining
   3.3. Graphics and video clips
   3.4. Database maintenance
   3.5. On-line order processing.

4. Timing-sensitive
   4.1. Call center phone calls
   4.2. Database access in support of phone calls
   4.3. Senior management and VIP queries

5. Most popular and visible
   5.1. Home page hits and downloads.

6. Likely bottlenecks
   6.1. Centralized database
   6.2. Firewall(s)

7. Complex functions
   7.1. Return of an order

8. Unknown or immature technologies
   8.1. Wireless.

9. Applications sharing the same infrastructure
   9.1. Billing system.
   9.2. Company e-mail

10. High impact on the bottom line
    10.1. Unavailability of the web site
    10.2. Inability to place an order on-line

11. Combinations of factors
    11.1. Web access and the voice call center (share the T1 lines)

Findings and Results

So how well did this test focusing strategy work? The story (which I will present verbally at WOPR2) may amuse you or vex you.

Table of Contents

I deliberately hid the table of contents back here so you could get immediately into the paper on the first page.

APPENDICES

Introduction to the Case Study

This case study describes a situation in which a performance test is needed. Your assignment is to identify and analyze the issues and develop a test strategy for the situation. This strategy describes how you will use testing to predict the system's actual live performance, in order to assess whether its performance is likely to be acceptable when the system is placed in operation.

Rationale for this Case Study

What's the problem?? Today's personal computers run a thousand times faster than IBM's first personal computer. Network speeds have increased even more, and database storage costs have dropped by a factor of more than ten thousand. System performance engineering and software optimization have advanced from a few incomplete rules of thumb to respectable professional fields. So why bother to test system performance and robustness?

I'd like to describe the problem by telling a story. Imagine that you have made a heroic effort to test a system under unreasonable deadlines and with limited staff and equipment. You scrupulously test to ensure the features work as expected, and then release the system. You monitored the performance as a byproduct of feature testing, and everything seemed fine in the test lab. A few days later, you receive a call from a senior executive in the client community. You are expecting words of appreciation, but he can only moan and rant about what you have done to him.

Our story ends on a sad note. The features work, *but*:

- The system response time is slow.
- The throughput (the volume of concurrent demands that the system can handle), is low.
- The system cannot handle peak loads or sudden surges in demand.
- It is fragile and readily crashes or hangs, and cannot recover from errors.
- It does not work on all the users' configurations.
- The system worked fine in the test lab but does not scale up – there appear to be bottlenecks.
- The system's resource use is prohibitively expensive. It is a "resource hog".

Ouch. If you identify with this story -- you have "been there, done that" -- and would prefer not to re-live the experience, then this case study is for you. Or if you have never been there and want to keep it that way, this case study is for you too.

Appendices: Understanding the Situation

Organization of this Case Study

The case study contains six parts:

Part 1: Understanding the Situation: In this part (which starts below), you will address the test objectives and focus, justification, scope, technical and business issues, risk factors and overall test approach.

Part 2: Understanding the Situation -- Suggested Answers to Part 1: You will critique suggested answers to the questions in Part 1, and either concur or suggest revisions and improvements to these answers.

Part 3: Reviewing the Proposed Test Strategy: You will review and issue an opinion on a proposed test strategy. This strategy is presented in the form of an executive summary and a comprehensive set of appendices with the supporting details.

Part 4: Reviewing the Consultants' Report: Subject matter experts (the consultants) have rendered an opinion on the adequacy of the proposed test strategy in Part 3. Your job is to determine which of the consultants' findings and conclusions are valid and worth acting on, and what corrective actions to take.

Part 5: Reviewing a "Practical" Test Strategy: In this part, an alternative approach is presented, which was developed based on the consultants' feedback. You will review and compare this alternative to the original proposed approach, and determine which of the two strategies is the most suitable, or what mix of the two you recommend.

Part 6: Developing the Robustness Test Strategy: You will review and critique the proposed strategy for testing the systems' reliability and recoverability. Part 6 is not included in this set of documents.

Learning Objectives

The lessons to be learned from this case study and series of exercises include:

- How to develop a performance test strategy in a typical mixed-technology, mixed-vendor environment with multiple interdependent application systems.
- How to set performance goals and testing objectives.
- How to perform a risk assessment and use it to focus and prioritize the test efforts.

Appendices: Understanding the Situation

- How to design the test lab, and allow for the differences between the lab and the real-world infrastructure which will be utilized in live operation.

- How to design realistic test work loads.

- How to use automated load testing tools effectively, and minimize opportunities for tool mis-utilization.

- How to decide what to measure and what data to collect.

- How to interpret the harvest of performance data and form meaningful, trustworthy conclusions.

- Why and how system performance testing is fundamentally different from performance engineering, system optimization and feature testing.

Using the Case Study

I have designed this case study with the intent that you will work through the series of exercises, not just skip them and read the suggested answers. The case study can be worked through individually or with a small team.

If you are working through the case study by yourself, simply skip the team exercises (in Part 1, these are Exercises 1.2 and 1.8).

I provide an estimate of the time to allow for each exercise. These estimates are guidelines: some people complete the exercises in a third of the estimated time while others take up to triple the time. This is not a race. People who zoom through the exercises may fail to see the nuances, and deliver unsophisticated answers which are not well thought through. Give the exercises enough time to absorb their lessons.

List of Exercises in Part 1

Exercise 1.1: Determining the Testing Objectives (60 to 90 minutes)
Exercise 1.2: Team Discussion of the Testing Objectives (45 to 60 minutes)
Exercise 1.3: Determining the Test Focus and Coverage (75 to 90 minutes)
Exercise 1.4: Developing the Test Scenarios (45 to 60 minutes)
Exercise 1.5: Calculating the Work Load (20 to 30 minutes)
Exercise 1.6: Outlining the Test Environment (30 to 45 minutes)
Exercise 1.7: Outstanding Issues (20 to 30 minutes)
Exercise 1.8: Assessing the Effectiveness of a Test Scenario (30 to 45 minutes)

Appendices: Understanding the Situation

Exercise 1.9: Team Discussion of the Test Environment and Issues (30 to 45 minutes)

Total time required for all Part 1 exercises: approximately 6 to 9 hours.

Abbreviated Exercise for Part 1

Use this abbreviated exercise if your time is limited, instead of the full series of seven exercises listed below. Perform the following activities. (Allow 30 to 45 minutes for the abbreviated exercise.)

1. Review the questions listed in Exercise 1.1, but do not actually work your way through the questions in this exercise and develop answers.

2. Then read the Description of the Situation, Part 1A only. Browse through the related appendices.

3. In Part 2 of the case study, the Suggested Answers, read and critique the answers to Exercise 1.1.

4. Do you agree with these answers? If you do not, what issues do you have with the answers and what improvements do you recommend?

5. As a separate follow-on activity, review Exercise 1.3 but do not work your way through the questions, and critique the suggested answers to Exercise 1.3 in Part 2. (Allow another 20 to 30 minutes for this follow-on.)

Disclaimers

The organization described in this case study is fictional, and is not intended to represent any real-world organization.

Testing tools from various vendors are mentioned in this case study and its accompanying series of exercises. None of these mentions should be construed as an endorsement of a particular vendor or a recommendation of a particular tool.

A Friendly Warning

The questions in the exercises may seem easy at first reading, but they can be fairly difficult to answer. The questions are hard because they require us to use judgment and think the situation through. They are also hard because the information you are provided in these exercises (as you will see in the next few pages) is not complete and perfect -- just like in the real world.

Appendices: Understanding the Situation

There is some degree of uncertainty, unresolved issues and information which is open to interpretation in this case study -- again, like the real world. Sometimes testers get into "analysis paralysis" -- they feel that they cannot do anything at all unless they first know everything about the situation. However, we cannot procrastinate and wait until we have perfect information; the test strategy will never be formed.

Ironically, despite the temptation we sometimes have to want to know everything about a situation, attempting to master all the information can be an overwhelming task. A great deal of information is presented in the next few pages, as background for the exercises, and an important part of the test planners' job is to simply to get a handle on this situation. The test planners have to review the information in the next section, determine what's relevant, and sift out what's not. Just like in a real test project.

Because the questions you have been asked to address are not that easy, there is a temptation to disregard them on a real-life project and "forge ahead" regardless. Though we do not want to become paralyzed in indecision, it is not a good idea to skip these questions. If we do not think them through at the beginning of a testing project, sooner or later we will be forced to return and reconsider them, and perhaps will have to change the direction of the testing project.

Exercise 1.1: Determining the Testing Objectives
(Allow 60 to 90 minutes for this exercise.)

Instructions

This exercise is the first in a series where you, working individually or in small team, will develop a performance testing strategy. Your purpose in this first exercise is to understand a typical business situation, analyze the issues and consider how to test its system's performance and robustness.

Read the background to the case study, in the attached Description of the Situation, Part 1A only. (You do not need to review Part 1B of the description for this first exercise.) Expect that this background reading will take you about 20 minutes, including a browse through the appendices to Part 1A. You may say: "Why should I have to read this? The only thing I ever read is TV Guide." Actually, this reading is important. Testing is context-specific, and we can talk endlessly about test strategy, but there is no substitute for actually getting in there and doing it ourselves. This background reading provides the context – it describes a typical challenging situation you are likely to encounter on the job.

Based on the background (i.e., the description of the situation), answer the questions listed below. Each answer need be no more than a few lines long. In the time available for this exercise

you may not be able to get through all the questions. The intention here is not to rush so that all questions are covered, but to think about the issues in developing a performance test strategy. These questions are tough. We do not expect perfect answers, but we would like your best thinking. If you become bogged down on a question, however, it is not worth agonizing over, so move on after a few minutes to the next question.

Questions to Address

(1.1.1) What do we want to accomplish with this performance testing project? In the process of determining what we want to accomplish, it is helpful to consider a few other questions:

- *Why* are we doing this test?
- *Who* do we need to satisfy? Who are the vested interests: who have needs and expectations of the test results? These individuals or groups sometimes are called stakeholders or constituencies.
- *What* do they want to know?
- *What* SHOULD they want to know?
- *How* will they use the results of the performance test?

(1.1.2) A set of business objectives are included in the description of the situation. Which of these business objectives can reasonably be addressed in a performance testing project? Which cannot?

(1.1.3) Performance goals are stated in the description of the situation (Part 1A). Overall, are these performance goals for the system (i) relevant and significant, (ii) realistic (i.e., probably feasible to attain), and (iii) testable or measurable? Has any major goal been omitted?

(1.1.4) Performance testing objectives also are stated in the description of the situation. Which of these testing objectives can be directly linked back to one or more specific business objectives? Which cannot? (A performance goal differs from a testing objective: an example of a testing objective is to evaluate whether response time is adequate, whereas the goal is a target for comparison with the measured response time. The goal is the level of response time that the system is expected to meet. It is difficult to evaluate if performance is satisfactory without having reasonably specific goals.)

(1.1.5) Are all these testing objectives appropriate, valid and within the scope of the performance

testing project?

(1.1.6) Are these testing objectives specific enough so that we can evaluate if they have been accomplished? (The specificity of the testing objectives depends on the specificity of the performance goals. If you felt that the stated performance goals are not specific enough, assume for the moment that these goals have been revised and now are realistic and measurable. In other words, answer this question as if the revised goals can be used to evaluate the system's performance.)

(1.1.7) In which ways do you think the corporate culture is likely to encourage effective testing practices? In which ways is it likely to discourage them?

Exercise 1.2: Team Discussion of the Testing Objectives
(Allow 45 to 60 minutes for this exercise.)

Instructions

Form a team of three to four people for this exercise. Find a comfortable place to gather around and work together. Together with your teammates, compare your answers to the previous questions in Exercise 1.1. The intention in comparing answers is not necessarily to reach consensus, though that's fine, but to obtain a deeper appreciation of the issues by seeing others' perspectives. Note that we are not looking for polished and detailed answers at this time, just an in initial sketch of your thoughts, ready for discussion with the whole class.

Allow about 10 minutes for each question, to discuss them in your team. (Some of these questions could take 3 weeks each to discuss, but we only have limited time in the classroom.) It is OK if you do not get through all the questions in the time allotted, but move on to the next question if you feel that you are becoming bogged down on any one question. At the end of the exercise, be prepared to discuss and justify your team's answers with the class.

Appendices: Understanding the Situation

Description of the Situation, Part 1A

A.1. Overview

The following sections provide background information needed for the exercises. These sections are labeled:

The Context
- A.2. The Business Background
- A.3. Your Responsibilities
- A.4. Basic Functions of the System
- A.5. The Physical Environment
- A.6. The Database
- A.7. Interfacing Systems
- A.8. Business Operations and Processes

The Objectives
- A.9. Business Objectives for the System
- A.10. Performance Goals
- A.11. System Requirements which Influence Performance
- A.12. Performance Testing Objectives
- A.13. Perceived Vulnerabilities

Appendices
- Appendix 1A.1: System Work Flows
- Appendix 1A.2: System Architecture

Later, more information about the situation is provided in another section, the Description of the Situation, Part 1B, but this information is not needed for the first three exercises.

A.2. The Business Background

Testing Books (TB) is a book club which specializes in selling testing and quality assurance books. The official company slogan is: "Test Geeks 'R Us" and the web site is testingbooks.com. The senior managers of the book club describe the business as thriving, growing and profitable. They believe the book club has a core of loyal fans that prefer its services to generalists like Barnes & Noble and Amazon, and they want to build on this success by providing even better service and more competitive prices.

To support this business goal, TB is in the process of building a comprehensive new information system to support its core business operation, which is the ordering and shipping of books to its

members. This new system will essentially replace the existing automated and manual systems which the book club uses for ordering and shipping books, and also replace an existing web site which is not considered to be very effective.

A.3. Your Responsibilities

Your job will be to test the performance of this new system, because it is critical to the success of the book club. The scope of this testing project includes measuring and evaluating the system response time, throughput, error rates, resource utilization, scalability and ability to handle peak loads. Congratulations (or condolences). You will report directly to the vice president of information systems for the purposes of this project.

Your immediate assignment is to draft a high-level strategy which describes how you will proceed and the approach you recommend for this testing project. You will be presenting and discussing this performance test strategy with the senior managers of Testing Books *next week*. In this presentation, they will want to know how you will test the performance of the system -- not the details as yet, but your overall approach. The managers expect an insightful, cogent analysis, and they are confident that what you say next week will be well thought through, organized and pertinent. (No pressure here at all!)

A.4. Basic Functions and Work Flow of the System

The new system processes book orders: it provides order entry and fulfillment capabilities. With this system, customers can order books directly from the web site and book club employees also can enter orders through a client/server network. The system also manages the fulfillment of these orders. In all, it will support the following business activities:

- Ordering of books from the book club.
- Picking, packing and shipping of books to members in fulfillment of orders.
- Answering queries on the status of memberships, orders and shipments.
- Publication of electronic and printed catalogs which show what books are available to order.
- Broadcasting of special offers and promotions.
- Reporting of the management information needed by the book club executives to run the business.

This management information, which is automatically generated by the system, includes order volumes, trends (which books are selling swiftly or slowly), order backlogs, the turnaround times to fulfill orders, book inventory levels, etc., and it is used by the senior managers who are running the business. While the volume of transactions in this category is likely to be relatively low, it is important to be able to provide timely answers – executives do not like to wait.

Please note that this is not a complete list of the business activities and the system functions which support them, just the major ones. The complete list of functions which the system is expected to provide is listed in the system requirements document, and a related set of use cases describe the services provided and how the functions should work. The system requirements and the use cases are not attached to this exercise, but this overview will provide enough information for you to do the exercises.

To realistically measure performance, testers need to know how the system works. At this stage, you may have many questions about the functionality and work flow, such as: "How does one become a book club member?", "Can non-members access all parts of the web site?", "Is this particular function even on the web?", and "How does this system tie into other systems?" Appendix 1A.1: System Work Flows, briefly explains the system-user interactions. (Later, Appendix 1B.1 lists the features and outlines how frequently each feature will be utilized, but this Appendix 1B.1 is not needed to answer the questions in the first three exercises.)

A.5. The Physical Environment

The new system will run on a multi-tier server architecture which will be shared by other applications needed for the book club's business operations, such as the billing system. The system environment includes the database, the web site, wireless capabilities, voice call center facilities and high-volume printers. Appendix 1A.1: System Architecture describes the technical infrastructure of the operating environment, in terms of the network topology, the existing equipment which can be re-used and incorporated into the new infrastructure, the various types of servers and the support software. You do not need expertise as a system architect, system administrator or network engineer to read this appendix.

A.6. The Database

The system has a centralized database which contains data about books in inventory, members, orders, shipments and bills. A new database is currently being converted from other existing databases of book club members, orders and books. The senior managers have identified the database conversion as a complex, high-risk activity, and they are concerned that several cycles of database tuning may be needed before the system performance goals are met.

When it has been converted from existing files, the database will contain approximately 100,000

active members. (Active members are those who have notified the club of their desire to receive a monthly catalog of books, or who have ordered a book within the last year.) The database will also contain another 200,000 inactive people who have ordered at least one book in the past, but not recently, or who have expressed an interest in the book club. The data on these other, inactive people is used in marketing campaigns. People are deleted from the database after three years with no activity.

A.7. Interfacing Systems

In addition to the book ordering system, the plan is to migrate other existing application systems to the same technical environment (the client/server network), and run them in this environment also. The main systems which share the same resources are:

- Member Billing: generates invoices for books ordered from the book club by members, and tracks shipping and payment status of these orders.

- Publisher Ordering: generates orders for books ordered by the book club from publishers, and tracks their delivery status.

- Communications: handles internal and external e-mail messages.

A.8. Business Operations and Processes

The book club business is organized into five main groups: (a) senior management, (b) the customer service group, (c) the catalog publishing group, (d) the warehouse distribution group, and (e) the information systems group. Each group has an assigned set of responsibilities, and there is little or no cross-over of tasks among the five groups.

The *customer service group* works directly with the book club members. Members can access the club's Web site or telephone (speaking to a book club employee), in order to place orders, query order status, make complaints, and change information about their memberships. The Web and phone orders and member data changes are processed while the member is on-line or is on the phone. The customer service group is intended to have approximately 100 personal computers (PCs) for its employees.

The *catalog publishing group* chooses the books for the club to offer, collects book reviews and prepares a monthly catalog of available books, which is either printed and mailed or is distributed via the Internet to the members. (These monthly catalogs do not contain the full list of books available for sale, just the most topical ones.) The catalog publishing group is intended to have 25 PCs to support their work. The entire staff of the catalog publishing group is located at the remote satellite office. (All the other business groups are located in the headquarters

building.)

The *warehouse group* distributes books to the club members, either from a member order or as the default monthly selection. A copy of the current monthly book selection is automatically sent to each active member, unless he or she explicitly informs the club that he does not want this book. Stocks of high-demand titles are maintained at the warehouse for filling customer orders. When the supply of any title becomes depleted the warehouse issues a request to replenish the stock to the publisher. A book shipment to a member can originate only from the warehouse -- other system users cannot authorize shipments. Personnel at the warehouse pack and ship approximately 2,000 books to members in a typical day, or about 45,000 books a month (working 22.5 days per month on average). The warehouse will have 50 PCs.

The *senior managers* make ad-hoc queries and receive on-line status reports and graphs which help them to manage the business. They also expect to receive regularly printed-out monitoring reports from the new system. There are 25 PCs planned for this group's use, including support staff such as the administrative assistants to the managers.

The *information systems group* supports and maintains the computer systems and will be expected to support the new system. This group will have 25 PCs including a test lab.

A.9. The Business Objectives for the System

The overall business goals of the senior managers are to (a) grow revenues, (b) improve profitability, and (c) increase member satisfaction. They have agreed to fund the new system, on the understanding that it will facilitate meeting these goals.

The specific business objectives for the new system are as follows:

- Support the operations of the book club, by providing the order entry, fulfillment and related features. (These features are described in more detail in the system requirements documents.)

- Double the volume of orders entered directly via the web within a year, and increase their percentage to 80% of all orders within three years (i.e., reduce the number of mail and telephone orders to 20% of the total).

- Improve productivity in the customer service and warehouse distribution groups by 25%.

- Distribute 95% of all books within 15 working days from the date of an order and at the least available shipping cost.

- Distribute books with 98% accuracy, i.e., only 2% of the books or less should be returned because the wrong book was shipped to a member or because the member's address was incorrect.

- Answer 95% of customer telephone calls (e.g., queries about order status), within 3 minutes.

- Support the projected growth in membership for the next 5 years. Today's system configuration does not have to support the load predicted for 5 years in the future, but the system must be upgradeable to meet the projected demand. The projected growth is described later in this document (in the Description of the Situation, Part 1B).

A.10. Performance Goals

In order to fulfill the business objectives, the system must meet these performance goals:

(a) Response times must be satisfactory when the system is operating under a typical realistic load.

(b) The system must operate correctly when accessed simultaneously by multiple users.

(c) The system must be able to handle heavy loads.

(d) Performance and reliability levels must be maintained over an extended period of use (24x365 operation).

(e) The system availability (uptime) must be adequate in live operation.

(f) The entire system must be tuned optimally in order to efficiently utilize the computing resources.

(g) The system must degrade gracefully, not fail catastrophically, when it is pushed up to, and beyond, it's planned maximum capacity.

(h) The system must be scalable, so it can be upgraded in the future to accommodate the projected growth.

A.11 System Requirements which Influence Performance

System requirements in areas like usability, security and maintainability usually are not performance requirements per se but they often are performance-related. Testing that these other requirements have been met is outside the scope of the performance testing project. Nevertheless, these system requirements may significantly influence performance. If the performance goals are met, but without these requirements being satisfied, then the executives will consider the performance test results invalid. The performance-related system requirements are:

Usability

The system must be user-friendly, so that visitors to the web site will be encouraged to browse and order books.

The system features must have a consistent format and methods of navigation.

It should use high quality, high definition graphs and video and audio clips.

The web site must be viewable on all popular browsers and platforms.

Users should be able to set their own preferences.

Data Availability and Integrity

Members must be able to retrieve their full history, if desired, via the web site.

The system must be able to provide ad hoc management reports on demand.

The database must be continuously updated in order to provide the latest information.

The call center must be able to replay any recorded voice message on demand, for up to three days.

Security

The system must allow new members to establish secure accounts' from which to order books and authorize payments, etc.

Members' financial information and payment transactions must be encrypted using an approved standard encryption algorithm.

All of a member's data must be password protected, and no other member's data can be accessed with the first member's ID and password.

The system must be secure from hackers.

Maintainability

The system must be easy to maintain.

A.12. Performance Testing Objectives

The purpose of this performance testing project, as outlined by the senior managers, is to assess the adequacy of the system's performance in live operation. Each of the following testing objectives focuses on whether a particular performance goal is likely to be met:

(a) Are response times satisfactory when the system is operating under a realistic load?
- Response time goals have already been specified in service level agreements (SLAs) which have been negotiated between the IS group and the business units. These service level agreements are described in the Description of the Situation, Part 1B. As mentioned earlier, you do not need to review Part 1B in order to answer the questions in Exercises 1, 2 and 3.
- The guideline for response time consistency has been set as plus or minus 25% of the average. That is, if the average response time for a transaction is expected to be 2 seconds, then 90% or more of the measured response times should fall within the range of 1.5 and 2.5 seconds.

(b) Does the system operate correctly when accessed simultaneously by multiple users? Problems such as features interfering with each other, database optimization, resource contention and transaction priorities may need to be considered here.

(c) How well does the system handle heavy loads? We cannot assume that the performance degradation from adding additional users or performing extra work is linear: a significant increase in response time may occur when only a few users are added or the workload increases by an increment.

(d) Will performance and reliability levels be maintained over an extended period of use? Insidious problems such as memory leaks will not reveal themselves in short run tests or by testing with a small number of users. Such problems usually lead to performance degradation and eventually to system failures.

(e) Will the system availability (uptime) be adequate in live operation? Apart from the planned downtime for maintenance, is there reasonable confidence that the unplanned downtime will be held to acceptable levels?

(f) Is the entire system tuned optimally? Successful tuning requires the empirical testing of server and client configurations using carefully controlled workloads.
- Are upgrades to dual-processor or quad-processor servers required to achieve the performance goals?
- On the other hand, an optimally tuned system may allow the book club to

increase the number of users without costly hardware upgrades. (Experience has shown us that very few systems are initially tuned for optimum performance.)

(g) Does the system degrade gracefully, or fail catastrophically, when it is pushed up to, and beyond, it's planned maximum capacity?

(h) Is the system scalable: can it be upgraded in the future to accommodate the projected growth over the next 3 to 5 years, without major software re-writes or a major re-structuring or conversion of the database?

A.13 Performance Vulnerabilities

The managers of the book club business groups have compiled this list of the perceived threats to the system performing adequately.

The likely sources of spikes in demand on the system from external users (non-employees) are responses to promotions in the first hours after they are broadcast, requests to download the e-catalogue in the first hours after it becomes available, and denial of service attacks. Heavy internal demand (caused by book club employees) is likely to occur when printing the paper version of the catalogue to mail to customers.

Uses which are resource-intensive, e.g., likely to consume a high level of system resources per event include book searches, data mining (this not part of the order entry system, but it runs on the same shared servers), downloading graphics and video clips of authors and books, the ad hoc management reporting and the database backup (the nightly "crawl").

System uses which are particularly timing sensitive include the call center phone calls, the database access for phone call support, and senior management and VIP queries (especially high priority ones).

Other background applications share the same infrastructure and may interfere or compete significantly for system resources. The biggest sources of contention are likely to be the billing system and spikes in e-mail traffic. Since these business managers have no knowledge of the system internals, vulnerabilities caused by inadvertently designed-in bottlenecks or poor implementation (e.g., inefficiently written software code) are not addressed in this section. (See the architecture evaluation performed by a review team of experienced architects, in Appendix 1A.2.)

Appendices: Understanding the Situation

Appendices to Part 1A

These appendices provide supplementary details about the book club's situation. While the details are worth reviewing, it is not necessary to analyze them in depth in order to answer the questions. Appendix 1A.1 outlines the main interactions among the system and its users. Appendix 1A.2 describes the technical environment in which the system will operate.

Appendix 1A.1: System Work Flows

People can order books either through the web site, or by phone or mail request. Using the web site, interested parties can search for books by topic, author, etc., query the availability and price of a book, and query the status of an order in progress. If they choose, people can enroll as book club members via the web site, phone or mail. A person does not have to become a member to order a book, but members get special privileges such as early notification of sales and occasional discounts. There are no membership fees. Book club members receive monthly catalogs and promotions, either electronically or in printed form.

An order can contain books from various publishers, with different quantities of each book ordered. Orders can be modified or cancelled at any time up until the day of shipment. If a book is not in stock, the system informs the person (or the internal book club employee) and asks if he or she wants to place a backorder.

A database record is created for every new order, and this record is updated to track the progression of the order through to fulfillment. The ordering system maintains records on the book inventory as well as on orders. When an order is entered, the system generates directives to the warehouse staff to pick, pack and ship the order, as well as printing the paperwork needed to ship the order. The system provides the capability to query and update the inventory of books on hand, and automatically decrements the inventory as books are shipped. Ordering books from publishers and book distributors is not part of this system.

The ordering system itself does not generate bills or process payments, but triggers these actions by the separate billing system. People usually pay by credit card, and members with acceptable credit history can be billed for payment within 30 days. To collect payment, the ordering system sends a transaction to the separate billing system. In the event of returned books, the billing system issues refunds but these do not flow through the ordering system. The system also provides the capability for the catalog publishing group to update book information such as descriptions and reviews, and to compose and publish the catalogs.

Appendix 1A.2: System Architecture

a. Architecture Overview

The system architecture has been derived from a business model, which is in the form of a set of use cases describing the services needed. (You will not need to review the use cases to do the exercises.)

The architecture is based on these principles: (a) the business drives the services, and the services drive the technology; (b) agility to respond to change is a fundamental business requirement, and the architecture is expected to be in flux on an on-going basis; (c) the system is as platform-independent as possible; (d) the system is loosely coupled, in order to minimize bug propagation (so that a code change on one server, for example, does not necessitate code changes on other servers); (e) interfaces will be mainstream and standards-based (i.e., only reasonably mature, non-proprietary interfaces will be utilized); (f) and the system is designed for robustness (e.g., fault-tolerant with fail-over capability, load balancing, redundancy and reserve capacity).

The service-oriented architecture uses Internet protocols to interface among platform-independent, semi-autonomous applications. XML, messaging and the Web have been utilized to re-architect and build new, integrated applications.

The architecture utilizes several relatively small, dedicated-function servers. These include database, web, wireless, application, print, voice telephone, network and remote location servers. The design provides flexibility and agility; improves reliability through high server redundancy, and facilitates troubleshooting by simplifying problem isolation and helping to pinpoint the location of a problem to the nearest server.

b. Architecture Evaluation

An experienced group of architects has reviewed the proposed system architecture and confirmed that it should be able to support the system's functional requirements. The architects also have identified some possible bottlenecks, based on their review of the system design on paper. None of these possibilities may actually become bottlenecks. Eliminating any one possibility will incur additional costs for hardware, networks, data storage, software and support, which the executives are not convinced are needed. The possible bottlenecks are:

- The centralized database, which is shared across multiple applications in a data-rich, highly data-dependent environment.

- The nightly database back-up, when other data-related transactions may slow to a crawl.

- Shared servers across multiple applications. For example, during a bill processing cycle the billing system mat pre-empt the application servers.

- The printers, which are not hard-wired directly to the print servers, but are on the local area network (LAN). The reviewers had a split opinion about the printers. Some believe that the monthly catalog printing will be a bottleneck. Others disagree, pointing out that the electronic print file has to be downloaded to the printers only once, regardless of how many thousands of copies will be printed.

- The voice or CTI servers also are on the LAN, leading to the possibility of heavy telephone traffic interfering with other work and vice versa.

- The web site access and the voice call center share the same T1 telecommunications lines.

- There are no separate front-end proxy servers; the firewalls reside on the web servers.

Every system contains bottlenecks. This number of possibilities (seven) is not unusual, and is not an indicator of an incompetent design. No opinion or information on this topic of possible bottlenecks is available from the system architects and developers, other than they are confident their design will work.

The remainder of this appendix describes the planned initial system configuration.

c. The Internal Network Topology

While book club members and visitors can access the system only through the web site, the internal part of system used by the book club employees will operate on a client/server network. This network will include approximately 250 users, with 225 in fixed locations with one workstation per user, and 25 mobile wireless users with hand-held devices.

Of the 225 fixed-location workstations, 200 will be physically located at the book club headquarters, which is an office building in Los Angeles, and will be on a local area network (LAN). The other 25 fixed-location workstations will be located at a satellite office, which is several miles away and which will be connected to the headquarters building by a wide area network (WAN). The 25 wireless devices will be used in the warehouse, which is attached to the Los Angeles headquarters building.

The client workstations will be supported by approximately 16 servers. Following the system architecture principles expressed earlier, each server will be dedicated to a specialized use and coupled to another redundant, load-sharing server for the same specialized use. It is anticipated that the server configuration will include two database servers, two application servers, two Web servers, two voice telephone CTI (computer telephony) servers, two held-held device severs, two print servers, two network servers, and two servers remotely located at the satellite office. This server configuration may possibly change as the system design is refined and finalized, or as necessary to balance

and tune the system, or as and when the system load characteristics change.

The servers in the cluster are intended to fail-over, that is, to back each other up as necessary without a loss of functionality. In the event that one server is disabled, the overall system should remain operational, even if it provides a degraded, slower level of performance to the users until full capability can be restored.

Initially, each server is intended to have a single processor. However, the flexibility has been left in the budget to upgrade selected servers to dual-processor or even quad-processor machines, if necessary, depending on the results of the performance testing and the performance tuning and debugging.

d. The Existing Infrastructure

The parts of existing infrastructure which can be re-used, namely the equipment and facilities which are already in place and support the existing application systems, will be integrated into this new environment, as described above. The equipment which will be carried over to the new operating environment has already been included and counted in the inventory of equipment for the new system, which is described in this section.

The parts of the existing infrastructure which cannot be re-used, such as older clients and servers, will be discarded after the new environment is operational.

There are no hand-held devices or wireless systems currently being used by the book club.

e. The Database Servers

Each of the two database servers will contain a complete copy of the database, and each will mirror the database on the other database server. In other words, all updates will be made concurrently to both copies of the databases, and these copies will be periodically and automatically monitored for consistency by the database management system (DBMS). The database servers will use RAID (redundant disk) technology, so that each database server in itself has redundant copies of the data.

A complete back-up copy of the database will be taken once every night. During this back-up process, which is expected to be completed within a period of 30 minutes, user response times may be slow. Few users are expected at the time, probably 3.00 am, so that slow responses are a minor issue. No guarantees have been made to the user community as to the worst-case performance of the system in this situation. Nevertheless, since the managers want to know if the user response times are likely to be extremely slow during the database back-up, measuring the user response times during the back-up is within the scope of this project.

A full set of the application systems will reside on each of the two application servers, so that both servers will have the capability to support all the system features and process all

transactions.

Load balancing software will be employed to automatically share (balance) the load among the servers: it will route transactions between the two application servers, to route query transactions between the two database severs, and also to route print requests between the two print servers, in order to optimize performance. This load balancing capability is expected to be physically located prior to the application and database servers, and since the system design includes no front-end proxy servers it thus will reside on the web servers... It is anticipated that no extra add-on load balancing tool will be utilized; the balancing will be done by the built-in capabilities of the server operating system.

g. The Print Servers

The print servers are connected to eight high speed laser printers, which will be shared by several users, and which are located in the work areas close to their main intended users. These printers are used to print in high volume the paper copies of the book club catalog. Although the catalog is available on-line, printed copies are available on request. Approximately 5,000 copies of the 300-page catalog are printed and distributed per month. The printers are also used for a variety of internal business reports.

In addition, dedicated local printers will be directly connected to individual personal computers and normally will not be shared across the network. Measuring the print times on these local printers has been declared to be outside the scope of this test project. Though the performance testers are not interested in the level of service provided, they will need to consider the impact, if any, of the localized printing on overall network performance. If network bandwidth is tight enough for local printing to be a problem, even with large print runs, then the system probably will have other bandwidth-related bottlenecks too. Testing the local print response is relatively simple and could provide early feedback on the existence of bandwidth problem areas.

h. The Web Servers

The Web servers will be connected to routers, which connect to the Internet through multiple T1 lines. Each T1 line has a rated throughput of 1.544 megabits per second (Mbps), or sufficient capacity to accommodate up to 25 simultaneous sessions using 56 kbps modems, assuming no data compression technology is used. The book club expects that the Web servers will be accessed from a variety of browsers on a variety of remote platforms. There will be firewalls on the Web servers which may slow their performance. Since not all the security software tools and controls have as yet been determined, the impact of the security controls on the system performance is not as yet fully known.

The Web servers will have their own databases, physically located on these servers. The Web databases will contain redundant copies of the same data and will be used to store

frequently downloaded data, such as FAQs, and to help isolate the Web site visitors from the internal systems of the book club.

The Web server databases will not contain all the data needed to answer queries or process orders The Web servers will need to access the database servers for some data which is not available on the Web servers. (Although the Web servers are physically connected to the database servers, some of the confidential company data on the internal databases is intended to be essentially invisible to Web visitors for security reasons -- in other words, this data cannot be accessed by the Web servers.)

i. The Voice Telephone Servers

The voice telephone servers will be used to coordinate incoming calls with customers' profiles, so that the person answering a call has ready access to the caller's account status. For example, a customer's data will be retrieved automatically, based on the telephone number of an incoming call, and displayed on the user's screen as he or she answers the call.

The voice telephone service will be supported by call center software which resides on the voice response servers. The voice service is handled by the same set of T1 lines which handle the Web service.

j. The Wireless Servers

The wireless servers will be used to communicate with the hand-held devices in the warehouse. These devices will help direct the order fulfillment, specifically the picking, packing and shipping of books.

In the future, the hand-held devices will be shared with other application systems in addition to the order processing system. The main other application is expected to be inventory management, where the hand-held devices will be used to take and report inventory counts and report out-of-stock conditions directly from the warehouse floor.

The inventory management system is not expected to be implemented until at least 9 months after the order processing system goes live. When the inventory management system is fully operational, it is expected to account for approximately 25% of the total wireless message traffic. (The order processing system will account for the other 75%.)

The system designers have not worked with wireless technology before.

k. The Remote Location Servers

Two servers will be situated at the satellite office. These servers will support a local area network for the client workstations in the satellite office, and will connect the satellite office to headquarters by a wide area network.

The business group which is located at the satellite office, and thus is the main group supported by these servers, is the catalog publishing group. The role of this group is described in a later section of this document.

l. Network Technologies

Internally within the headquarters building, the fixed-location clients and servers will be connected by a fast Ethernet local area network (LAN), using the 100Base-T standard and rated at 100 Mbps. If the performance measurements indicate that the LAN is a major bottleneck, the speed of the LAN could possibly be increased to a gigabit per second (using gigabit Ethernet). However, an allowance for this contingency has not been included in the project budget.

The Web and CTI (telephone) servers will be connected to a set of T1 lines, which can be shared by both the Web and voice telephone traffic. These lines have a rated capacity of 1.544 megabits per second for each line. The performance of the protocol conversions from external networks, such as SONET, to internal Ethernet LAN, is assumed not to be a significant issue and will not be measured as part of this project.

Internally within the satellite office building, the clients and servers will be connected by a traditional Ethernet LAN, rated at 100 Mbps.

The wide area network (WAN) connecting the headquarters and the satellite office will utilize one or more dedicated T1 lines.

m. Software

The decision has been made to use Windows XP on both the servers and on the clients.

The database management software used by the book club will be Oracle 9i.

The Web servers will run Apache software to service Internet traffic.

n. Physical Installation

A contractor is in the process of installing and wiring the computers and network equipment needed for the new system.

This contractor will test and confirm that this equipment has been installed and works, and also will confirm that the support software, such as Windows 2000 and device drivers, has been installed correctly on the equipment.

Description of the Situation, Part 1B

This section addresses these topics:

B.1. System Usage Demographics
B.2. Growth Projections
B.3. Service Level Agreements
B.4. System Development and Feature Testing Methodology
B.5. Automated Test Facilities
B.6. Conditions and Constraints
Appendix 1B.1: Feature List and Operational Profile

B.1. System Usage Demographics

The transactions processed by the order processing system and other related systems are listed below, together with their expected frequencies of utilization. The transaction volumes are based on actual counts of transactions within the existing system, which have been adjusted for the expected differences between the existing system and the new one. (There is not always a one-to-one relationship between the transactions in the existing system and the new one.) These counts reflect today's use and do not include the project growth in the future, and do not reflect the anticipated shift from primarily internal ordering (by book club employees) to primarily external (by book club members and visitors through the web site). The peak hours of demand for the business groups may not all happen at the same time. The following table summarizes the demand:

User Group / Feature or Transaction	**Frequency of Utilization** (Transactions per hour)	
	Typical hour	Peak hour of month
Customer service group		
Web site hits by external users	950	7,750
Demands generated internally	250	1,500
Catalog publishing		
Monthly catalog of books		
Printed catalog	-	5,000
Distributed via the Internet	-	15,000
Monthly promotion	-	50,000
Warehouse group		
Shipping and bill of lading papers	150	1,000
Picking instructions	250	1,250
Information systems group		
(not including periodic database back-ups)	250	6,000
Senior management group		
Queries and reports	Under 25	1,750

See Appendix 1B.1 for the supporting details behind this table.

Occurrences of Peaks

Information about when the peak demands occur during a week or month (i.e., at which specific hours), for each feature and type of system use, is not yet available.

B.2. Growth Projections

The system must be scalable or upgradeable to support book club member growth rates of 15% per year, and order volume growth rates of 25% per year, for the next four years.

The growth is expected to be achieved primarily by providing an enhanced Web site for book club members and potential members. For this reason, the mix of books sold over the Internet versus through the traditional channels is expected to change, as follows:

Time Frame	Percentage of Books Sold	
	By Traditional Channels	By Internet
This year	80%	20%
2 years ahead	50%	50%
4 years ahead	20%	80%

Please note that this table does not include default orders. By default, members receive the monthly book selection without placing an explicit order. A book club member must state explicitly that he or she does not `want the book of the month; otherwise the book is shipped to him.

The mix of ordering demands is expected to evolve to [20% traditional channels – 80% web], with a corresponding expected shift of work from the telephones to the web. This means that the internally generated workload in response to telephone calls is expected to fall over the next 5 years, not rise, even after allowing for the overall growth in business volumes. So the available or spare CTI (call center) capacity will effectively increase.

B.3. Service Level Agreements (SLAs)

A separate service level agreement has been developed for each of the three operational groups, the customer service group, the catalog publishing group, the warehouse distribution group. Each service level agreement has been signed by the manager of the pertinent business group and by the manager of the information systems group. The adherence to the service level agreements will be monitored and reported to senior management monthly, and the information systems group will be evaluated in part on their ability to satisfy the service level agreements.

The following statements summarize the key points in each of these service level

agreements. The agreements state that, for priority 1 features under normal operating conditions (e.g., with an average workload), the internal users' response times should be less than 0.5 seconds, 90% of the time or better, on the client machines. (Priority 1 is the highest priority.) Under peak load, the internal users' response time should be 1 second or less, 90% of the time. This response time goal must be met for each business unit individually, such as the customer service group, the catalog publishing group, and the warehouse distribution group. In other words, it is not sufficient if the average response time across all the business units meets this goal, while the average for any one business does not meet the goal.

For priority 2 features, the internal users' response times should be less than 2 seconds, 90% of the time or better. For features with priority levels 3 or more, no response time guidelines have been established as part of the SLAs.

For the external users who visit the Web site, the requirement is that the book club's Web site is "noticeably faster" than competitors' Web sites. It has been determined that this means that the book club's home page on the Web site can be downloaded in no more than 4 seconds on average, under an average work load, and no more than 10 seconds under peak load. These times are measured from when a user clicks to initiate an action until when the user starts to see a response, e.g., when a page starts to be rendered, not when the page is fully visible.

These response time targets are the averages for all Web site visitors. (A person using a 9.6 Kbps modem will have a slower response than someone with a dedicated T1 line.)

The book club also anticipates that these Web-based response time targets will need to be revised at least once every 6 months, and tightened, as competitors' Web services become progressively faster.

On-going processes will need to be put in place, to monitor and report the level of compliance with the SLAs during the on-going live operation of the system. Developing these mechanisms to monitor and report on SLA compliance is not part of the scope of the performance test project.

The service level agreements also address other areas besides response time and throughput, such as acceptable levels of errors in processing, system availability, etc., but these areas are outside the scope of the performance testing.

B.4. System Development and Feature Testing Methodologies

The new book club order processing system is currently in the process of being developed. Where possible, the developers are re-using existing software components, which have either been purchased from external sources or derived from other internal systems.

Rapid application development (RAD), which is also called the iterative spiral approach,

is being used by the developers. In this approach, the feature testing starts early and overlaps development, and components and sub-assemblies of components are tested as they become available, and then re-tested as the components are debugged or enhanced.

There is a feature testing team, which has been organized to test the features of the new order processing system. This team is not responsible for performance, load or stress testing, but its members have been asked to cooperate and provide assistance to the performance test team, as appropriate and without interfering with the feature test project.

The developers and testers are planning to build component-level test drivers, in order to start unit-testing each major software component as soon as it becomes available.

B.5. Automated Test Facilities

This feature testing team is in the process of building a library of automated feature test cases for each of the features listed above. Depending on the criticality, risk and complexity of each feature, the library is expected to contain anywhere from one to ten test cases for each feature.

These test cases are being built and will be executed and maintained by using WinRunner, a tool from Mercury Interactive of Sunnyvale, CA. The feature testing team has purchased five copies of WinRunner, which could be available for performance testing when the feature testers are not using them.

Note that the mention of any particular tool in this case study, such as WinRunner, should not be construed to be a recommendation or endorsement of that tool.

B.6. Conditions and Constraints

At this time, the senior managers have not set a deadline or budget limit for this performance testing project. This does not mean that they will be willing to accept an indefinitely long time frame or an indefinitely large budget for this project, but they are waiting until they hear what you propose.

Appendix 1B.1: Feature List and Operational Profile

User Group / Feature or Transaction	**Frequency of Utilization** (Transactions per hour)		**Priority**
	Normal use	Peak use (**)	
		Week *Month*	
Customer Service Group			
(a) Web site hits by external users (direct external demands)			
Main or home page only	450	1,800 3,600	1(*)
Book search	250	1,000 2,000	2
Book query (availability, price)	250	1,000 2,000	1
Book order	50	250 500	2
Credit card authorization	50	250 500	2
Query status of existing order	25	50 75	3
Add new member (sign up)	5	15 50	1
Change membership information	5	15 50	3
Link to another site	125	500 1,000	2
These transactions are all initiated from external users via the Internet.			
(b) Demands generated internally in response to telephone calls (the average duration of these telephone calls is 12 minutes each)			
Book query	100	300 600	1
Book order	50	200 400	2
Credit card authorization	35	100 350	2
Status of existing order	10	25 100	2
Add new member	1	5 10	1
Change membership information	1	5 10	3
Delete existing member	0.5	2 5	3
Complaint	0.5	2 25	1
These transactions are all initiated from the personal computers in the customer service group.			

(*) Priority of each type of transaction, where 1 is the highest and 4 is the lowest priority.
(**) These peaks represent the expected load during the worst-case hour in a typical week, and in a typical month, in the two columns respectively. The peaks for all these types of transactions are not expected to occur within the same one hour of the week or the month.

User Group / Feature or Transaction	**Frequency of Utilization** (Transactions per hour) Normal use	Peak use (**) *Week*	*Month*	**Priority**
Catalog Publishing Group				
Book list				
Add book to list of available books	2	50	100	3
Change book information	1	50	100	3
Delete book	1	10	20	3
Add book review	5	25	50	3
Change book review	1	5	10	3
Monthly catalog of available books				
Printed catalog (***)	0	0	5,000	4
Distributed via the Internet (***)	0	0	15,000	4
Monthly promotion (***)	0	0	50,000	3

These transactions are all initiated from the personal computers in the catalog publishing group.

(***) These loadings on the system happen only in occasional bursts; normally the transaction traffic is zero.

Warehouse Group				
Shipping and bill of lading (Shipping papers and invoice)	150	500	1,000	3
Inventory management (****)				
Inventory query (Internal levels of inventory)	10	50	100	1
Publisher query (e.g., review of new book titles)	10	50	100	2

These transactions are all initiated from the personal computers in the warehouse group.

(****) These transactions are not processed by the ordering system, but by other systems which share the same resources.

User Group / Feature or Transaction	**Frequency of Utilization** (Transactions per hour) Normal use	Peak use (**) *Week*	*Month*	**Priority**
Warehouse Group (continued)				
Inventory management (continued)				
Publisher order	5	25	50	3
Inventory update (Receipt of new book data from publishers)	10	500	500	3
Order fulfillment				
Picking instructions and confirmations	250	750	1,250	2

These transactions are all initiated from the hand-held wireless devices in the warehouse group.

Information Systems Group *(****)*

System development and maintenance	50	250	500	2
Internal e-mail traffic -- among all departments	150	1,000	5,000	2
System administration				
On-going administration	50	250	500	1
Periodic (e.g., database back-up)	0	50,000	50,000	4

These transactions, except the e-mail, are all initiated from the personal computers in the IS group. The e-mail traffic is distributed equally across all the clients in the network.

Senior Management Group

Ad-hoc query	0	250	500	1
Daily on-line status report	10	100	250	2
Weekly status summary	5	500	500	4
Monthly financial and operating data (****)	0	0	500	4

These transactions are all initiated from the personal computers in the senior management group.

Other Systems on the Shared Servers

The ordering system interacts and shares resources with other systems, such as billing, publisher ordering, market data analysis (data mining) and communications (e-mail). The extra work loads include overhead and background transactions like e-mail between the book club employees, and features which are not part of the ordering system but

which are hosted on the same application servers. (Other systems which are not listed here have their own separate application servers, and these are not included in the equipment inventory in Appendix 1A.2.)

The work loads for the features of the other systems which run on the shared servers are as follows:

User Group / Feature or Transaction	**Frequency of Utilization** (Transactions per hour)			**Priority** (*)
	Normal use	Peak use (**)		
		Week	*Month*	
Billing group				
Generate e-bill	50	250	1,500	1
Print bill	50	250	500	1
Generate reminder notice	10	25	250	1
Process payment	50	250	500	2
Adjust bill or payment	10	25	150	2
Publisher ordering group				
Order from publisher	10	25	150	1
Receive and store	10	25	150	1
Pay publisher	10	25	150	2
Adjust order or receipt	1	10	50	2
Marketing group				
Data mining query	10	25	250	5

Microsoft and Me

Microsoft has published a similar performance testing case study, for a fictional firm called the Duwamish Book Store. The Web-based business functions supported by the system include point-of-sale, order entry, shipping and receiving, and a book catalog. My case study in this book was published in a substantially complete version well before Microsoft published theirs. The Duwamish Book Store case study is worth reviewing as a counterpoint to this one. Microsoft's case study is simpler and less rich in issues, so it may be worth looking at first, before you get into the complexities and nuances of this case study.