# 1 System design process

## OVERVIEW

This book describes an approach to system architecture design specifically intended to support successful geographic information system (GIS) operations. With it in hand, you have a process—not a prescription—and a practical tool (see the Capacity Planning Tool on the DVD) to help you plan the infrastructure for a GIS specific to your needs. Knowing what you need is all-important to this process, and it is the assumption of this book that you have already identified what your organization requires from a GIS. This is because system architecture design defines the relationship between peak user-workflow-processing needs and the hardware and network infrastructure required to meet them.

The purpose of this book is to provide managers with not only the tools but also the information to help them build a successful infrastructure for GIS. A fundamental understanding of the supporting information technology, GIS performance and scalability, and the models and methods in this book fosters the type of management framework that leads to success. These fundamentals of system architecture design could help in building a system for any technology. But, as consultants at Esri, we have applied them to guide our customers to successful GIS deployments.

We share our understanding of the technology in a purposeful order similar to the way you should approach building your GIS. Chapter 1 provides an overview of the system design process and introduces the Capacity Planning Tool (CPT). The CPT includes a variety of tools that describe your business needs, help you select the right software and hardware technology, and model the performance and scalability of your GIS operations. It is an open-source application designed to help GIS and IT managers understand their business needs and promote more productive GIS operations. Chapters 2–4 focus on GIS software technology, software performance, and data administration and illustrate software patterns and key performance parameters that can be used to help quantify your GIS business needs. Chapters 5–7 show how to select and configure an IT infrastructure to meet your system performance and scalability needs (the needs that were developed as you worked through chapters 2–4). Chapter 8 demonstrates the need for building GIS operations that address your security needs. Chapters 2–8 identify the building blocks that contribute to successful GIS deployment—getting all of these pieces of the design puzzle put together right from the start can save money and reduce implementation risk.

Chapter 9 shares our understanding of performance and scalability fundamentals. These are the terms and relationships that contribute to user productivity throughout your system. Understanding these fundamentals will help bring all of the pieces of the system design puzzle together. These are the terms and relationships used by the CPT to complete the system architecture design analysis. Chapter 10 provides an overview of the CPT user functions and information resources.

Finally, chapters 11–12 explain the process for completing a system design in which user requirements lead to hardware specifications. If you are a system architect, you may be tempted to jump to chapter 11 where the system design process is spelled out step by step. Don't do it; rather, look before you leap. Even system architects are not ready to begin until they understand the technology and the CPT, explained progressively within the first 10 chapters, because technology changes dramatically and continuously. (I would not send our system design consultants to a customer site until they understand what's inside the first 10 chapters of this book.)

At one time you could simply define user requirements, select a software solution of choice, then identify hardware and network specifications to fulfill system capacity needs. Nowadays, it's more complex. Selecting the right technology requires a more specific understanding of

GIS user needs and system infrastructure limitations. Enterprise GIS operations today are supported by a variety of software and hardware technologies that must work together to meet user performance needs when operating at peak system capacity. Technology selection must take into consideration performance and scalability requirements and infrastructure limitations. An integrated business needs assessment—evaluating GIS user needs along with projected system performance and scalability throughout the requirements definition process—is the recommended way to ensure deployment success.

For this reason, system architecture design is not at heart a step-by-step process, but rather is best conducted as an integrated business needs assessment. It is more like working on a puzzle. There are procedures within the process, such as those that must be followed in establishing business requirements, but first, as in a puzzle, a system must be understood as a whole before it can be properly put together. Like the pieces of a puzzle, many interacting components compose the whole, each related to another in a very special way. Each special relationship emerges during the system architecture design process to reveal to you its significance to your design. The components of this puzzle are what we examine in *Building a GIS*, and our discussion is divided into three parts.

## Part I: Understanding the GIS software technology

**Chapter 1.** System architecture design—defines this process at the highest level and identifies its purpose along with the pieces of the system design puzzle (figure 1-1). Chapters 2–4, which complete the first section, describe the contributions of software technology, answering the following questions:

**Chapter 2.** Software technology—what are the software technology options?

**Chapter 3.** Software performance—an overview of software performance considerations: how will your software technology decisions affect system performance and scalability?

**Chapter 4.** GIS data administration—what are your options for maintaining and providing the required GIS data resources?

## Part II: Understanding the IT infrastructure

Once you have identified what you want out of the system, you will need to identify the network and platform architecture required to deliver what you want. You also need to build an infrastructure that will protect and secure your business resources. The IT infrastructure is the focus of the second part of the book, chapters 5–8. You can obtain information about cost from the software and hardware vendors once you identify the right solutions for your organization.

**Chapter 5.** Network communications—what effect will a given technology solution exert on the existing infrastructure?

**Chapter 6.** GIS product architecture—what configuration options match up with your availability and scalability needs?

**Chapter 7.** Platform performance—what do the vendors have to offer, and how do you select the hardware that will provide the required processing power?

**Chapter 8.** Enterprise security—how do you adjust your architecture solution to accommodate your security needs?

## Part III: Putting it all together

Putting all the pieces together completes the puzzle. The performance fundamentals, the CPT, and the actual system design process itself are presented in chapters 9–11. The book concludes with an overview of system implementation best practices—how to deliver within established performance budgets.

**Chapter 9.** Performance fundamentals—an overview of the performance relationships between the pieces of the system design puzzle.

**Chapter 10.** Capacity planning—a fully illustrated and detailed description of how to use the CPT (a composite of several integrated information management tools that identify user-workflow requirements and select the right hardware and network solution), including how to customize it for your own situation.

**Chapter 11.** Completing the system design—a walk through the system design process, using a case study to show how to bring the pieces of the puzzle together in the system architecture.

**Chapter 12.** System implementation—guidelines for implementing the selected design solution, which take into account the maintenance and tuning that follow.

## The CPT on the DVD

The performance components that contribute to success are clearly defined and modeled in the CPT, which is provided on the accompanying DVD and thoroughly described throughout the book. Templates are provided in the CPT for collecting user requirements, as are standard workflow models that translate peak user loads to selected platform processing environments. The CPT translates the peak user workflows to specific platform and network specifications, applying the performance models shared in chapter 9 so that you can use it to evaluate which mixes of technology will best meet your needs. In doing so, you may find the selected software technology will not support users over the available network infrastructure. You may find your favorite vendor does not provide the best hardware for your preferred solution. You may find you need to upgrade your network infrastructure or change your business processes to conform to a more distributed architecture. As you change each piece of the puzzle, its relationship with the other pieces of the puzzle will change and the overall solution will be reevaluated. The CPT is intended to make this process of reevaluation easier for you. For each solution, the CPT takes each component into consideration and evaluates performance of the integrated system: Does the selected solution meet your organization's needs? If not, what other alternatives will better support your needs?

Also on the DVD, you'll find teaching and learning aids, including a list of multiple-choice questions and capacity planning exercises associated with each chapter. These might be useful for teachers and students or simply for readers interested in reviewing the main points of discussion. A full set of CPT demonstration videos are included to help you understand how to use the CPT tools. Also included is a full-day slide and audio presentation of the *System Architecture Design for GIS* preconference seminar presented at the 2010 Esri International User Conference.

A system is only as strong as its weakest link. Enterprise GIS operations are supported by a large number of infrastructure components, wherein the weakest component will limit system capacity and user productivity. Selecting the right software technology, building proper applications, establishing an effective database design, and procuring the right hardware all play critical roles in fulfilling system performance and scalability expectations. Understanding your infrastructure needs before you buy can make the difference between success and failure and significantly reduce the overall cost of system deployment. The key to a successful system design is understanding what you need, establishing measurable performance milestones, and managing incremental progress toward performance and scalability goals throughout the development and deployment process.

Computers are part of our life because they save people—and organizations—time and money. Tasks that used to take hours or weeks can now be done in minutes, even seconds. People use GIS for the same reason they use computers: because it makes them more productive, more efficient at work. Anything that can improve efficiency and productivity is welcome. System performance—how fast and reliably the components of a computer system can process an application and display the results for you—directly contributes to user productivity. You need to configure all the elements in a system (software, hardware, network communications) so that they work well and efficiently together in doing what you need to get done.

Technology is changing and, unless people take the time to understand it, they can get in over their heads and their systems can suffer performance problems. Have you ever experienced a problem with a system at point A, only to discover that the actual cause of the problem was located way over at point E, a seemingly unrelated area? Usually these realizations come after spending quite a while exploring a lot of dead-ends trying to get to the root of the matter, often settling for a temporary work-around. No one needs to tell you, these times spent chasing symptoms interrupt the workflow and, if they happen over and over again, can significantly reduce productivity. They also could be a sign that there may be something wrong with the system design.

Far better, of course, is to get the system right the first time around. Then you can spend your tinkering time doing serious planning and maintenance, to keep your system at the ready to adjust, as advancements in technology and organizational goals present new opportunities for growth through change. Getting the system right means achieving *performance*: a proper system works the way you want and as fast as you need. A system able to take advantage of opportunities and grow is a *scalable* one.

The key to achieving both performance and scalability—the prerequisite for getting the system right and for keeping it right as the demand on it grows—is the same task: understanding the nature of each component of a system, the interrelationships between components, and how changes in one affect the other. Each component technology affects the overall system performance. The reason you might not have known right away the symptomatic connection between point A and point E in the example above is no doubt due to the intricacies of these interrelationships. They seem complex, but after we've sorted them out, you will see how simple system design really is.

## System architecture design for GIS

Computer systems were first used to automate cartographic map production in the late 1960s. Before that, geographic analysis, a method for displaying the relationship of many layers of geospatial information, relied on more time-consuming methods. Traditional methods of creating Mylar representations of each dataset and overlaying the layers of spatial data could take weeks to complete a simple information product. Early computers were able to automate the analysis process and reduce map production timelines from weeks to hours. Current computer technology can complete this same type of analysis and generate dynamic map displays in less than a second.

The term GIS was introduced by Dr. Roger Tomlinson in the 1960s, when he initiated, planned, and directed the development of the Canada Geographic Information System, the first computerized GIS in the world. Tomlinson's effort was later promoted by professors at Harvard University in the 1970s, inspiring several geographic consulting companies to develop and expand GIS technology. One of those companies was Esri, beginning in 1969. Consider that Esri software runs on more than one million desktops in more than 300,000 organizations, and you can see that the evolution of GIS has followed the advance of computer technology.

Local government and business started deploying large GIS operational systems in the early 1990s, and soon it became clear that the success of such distributed GIS operations was strongly connected with an understanding of GIS performance and scalability in a distributed computer environment. Distributed GIS operations were characterized by high performance computer systems coupled with a high demand on network communications. GIS deployment required several years' investment in data migration and custom application development. That, along with the required system infrastructure investments, was expensive. An understanding of system design requirements was critical in supporting successful GIS deployments. This is why a software company became so interested in hardware system performance early on.

My initial responsibility when I joined Esri was to develop a team that would support successful implementation of turnkey (out-of-the-box) GIS sales. Part of that was the acquisition and installation of hardware and software technology adequate to the task of supporting an organization's GIS workflows—or, as we call them, GIS peak operational system level requirements. We established a Systems Integration Department with four specific teams (system architecture design, project management, system installation, and project control). We helped complete several hundred successful GIS turnkey project implementations between 1992 and 1998, and learned a few things along the way. The best practices developed to facilitate implementation of turnkey GIS sales were also used to address failed system implementations. We learned from our mistakes and those of our customers.

Technology has changed tremendously, but the three fundamental building blocks required for each project implementation have not changed in all this time:

1. A clear definition of the peak user workflow requirements (user requirements analysis)
2. A clear understanding of infrastructure requirements to support the peak user workflows (system architecture design)
3. An implementation strategy that provides proper systems integration management from initial contract authorization to final system acceptance (project management)

A variety of best practices were soon established to facilitate a proper user requirements analysis. It had been evident during earlier implementations in the 1970s and 1980s, and then became abundantly clear, that identifying your user requirements—knowing exactly what you need to get out of a GIS (being able to describe the information products)—was essential for a GIS implementation to be successful. Implementations without a clear goal and purpose would fail.

Distributed processing systems introduced in the 1990s required a proper platform and network infrastructure to ensure the GIS benefits identified in the user requirements analysis could be achieved in a distributed environment. Many systems were deployed without a clear understanding of the system infrastructure requirements, and many of these early systems failed to meet user performance expectations.

In the early 1990s, I began to develop a system architecture design process for Esri that would identify the proper hardware and network infrastructure required to support successful GIS implementations. Like a system itself, this process has been developed, maintained, and fine-tuned over the years, through working with colleagues and clients who have helped make it as productive as it is today. GIS is usually integrated into an existing system, so this design process takes into consideration an organization's infrastructure limitations. The process can be used to make specific recommendations for hardware and network solutions that are actually based on existing and projected user needs. (You'd be surprised how many projects just guess.) Based in reality, such system designs reduce implementation risk and are more likely to be approved. In any case, a system design is often a prerequisite for gaining the go-ahead for your GIS project from upper management.

This system design methodology recognizes that people, application technology, and selected data sources are equally important in determining the optimum hardware solution, as shown in figure 1-1.

What people need, what it takes to create what people need, and how all the components of a system work together—every day and reliably—are the elements that comprise a balanced system design. System capacity requirements were initially represented by peak concurrent user workflows, where each concurrent user would generate system-level server processing and network communication loads. A user workflow would represent the system loads generated by a single client use case. When web services were later introduced, system capacity requirements were represented in terms of peak service transaction rates (peak pages or displays per hour). Display transaction rates proved to be a

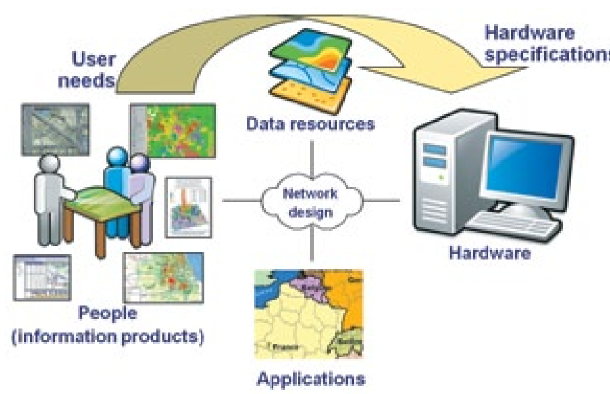



**Figure 1-1**

What is system architecture design?

- People: Understanding user needs—information products and the procedures for making user products—establishes a rationale by which to estimate peak system workflow loads.
- Applications: Software technology determines the processing requirements (system loads) that must be handled by the hardware solution.
- Data: The type of data source and what's required to access it (data access requirements) show you how the processing load is distributed across the system architecture.

more accurate sizing model than basing throughput on concurrent user headcounts.

In an enterprise design, both user and service workflow throughput rates must be combined to identify shared system loads—we need to identify a common throughput unit of work. A new term, user productivity, was introduced to convert peak concurrent user workflow loads to display transactions. User workflow loads can be translated to average system displays per minute (DPM), also known as transactions per minute (TPM), by multiplying peak concurrent users by user productivity represented in terms of displays per minute per client (DPM/client). For example, 20 concurrent users with a user productivity of 10 DPM/client would generate a system throughput of 200 DPM. We found early on that real user productivity is best understood as DPM rather than display transactions per hour (TPH). For example, a power user might generate 10 displays per minute, while a casual user might generate 1-5 displays per minute—user productivity of the 10 DPM/client for a specific workflow would be twice the system load as the 5 DPM/client. One can envision a GIS user requesting 10 map displays per minute (1 map display every 6 seconds) as a power user. The final system processing loads (whether generated by user workflows or by web services) now have a common unit of work (display) that can then be used to combine system loads for platform and infrastructure planning.

The Esri system architecture design process shares specific deployment strategies and associated hardware specifications based on identified operational user and service workflow requirements.

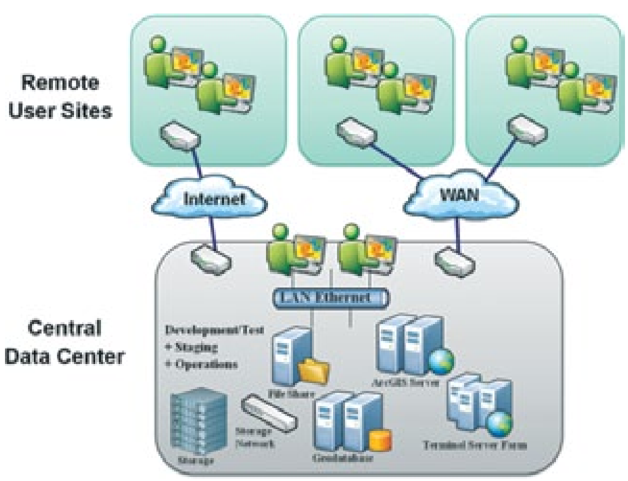



**Figure 1-2**

Identifying business requirements.

## Why we do planning

The primary reasons for planning include identifying business needs, defining project requirements, and reducing implementation risk. In practical terms, we need a plan if we hope to get something done. The plan provides a foundation for successful implementation. Figure 1-2 represents typical distributed GIS operations. GIS operations include both central and remote user locations connected over a distributed network communication infrastructure providing access to centrally managed business information systems. What will it cost to build such a system?

Distributed computing environments support users located at multiple remote sites connected over wide area network (WAN) and Internet network connections with applications and data sources supported in a centralized data center.

Business requirements may include a variety of user workflows and web services to improve business operations. These workflows and service definitions must be identified early during the planning process because they establish a foundation for the rest of the planning process.

The peak user workflow and web services processing requirements must be identified and translated to appropriate system design loads—server processing times and network traffic. These processing loads provide a foundation for generating hardware specifications (performance and capacity requirements) and network infrastructure needs.

A project must be defined before you can procure and implement the required technology. The project will include a budget and a schedule along with a variety of project performance milestones, as shown in figure 1-3.

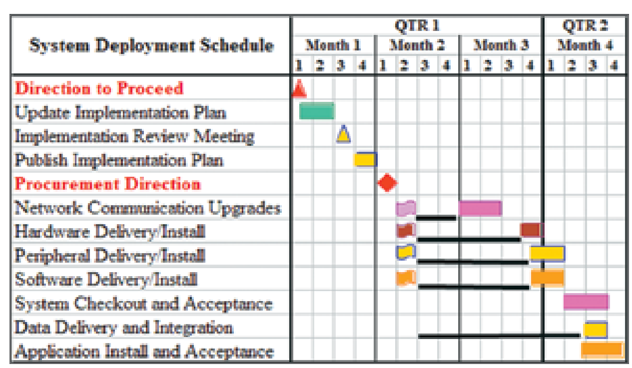



**Figure 1-3**

Defining project requirements. Project schedules provide a framework for managing system delivery within established time and cost budgets. Performance milestones should be included in the project schedule to validate compliance throughout the design and delivery process.

Project approval is required before any major design and/or implementation efforts begin. Project approval is based on identified or perceived benefits that will accrue to the business resulting from the proposed level of time (schedule) and financial (budget) investment.

Delivering a system that will satisfy identified business requirements is fundamental to project success. Understanding key system performance parameters, identifying incremental system performance targets, and establishing a system performance validation plan can chart a solid framework for managing implementation risk. Success requires doing your homework. Planning is an incremental process driven by rapidly changing technology. The most effective plans are designed to adapt to changing business needs and evolving technology opportunities.

## Why system architecture design is important

Vendor computer hardware and network infrastructure expenses represent a significant percentage of the overall cost of deploying and maintaining distributed GIS operations. For many implementations, costs for hardware procurement and information technology (IT) administration and support exceed the cost of the GIS software. These costs must be identified during the project approval process and managed throughout system deployment to ensure that resources are available for implementation of the technology.

In order to define the overall project requirements, you must understand the relationship between GIS user needs, GIS software processing requirements, and computer hardware and network infrastructure performance and scalability. Understanding the technology (both GIS software and vendor hardware) is fundamental. To meet peak user performance requirements, a distributed computer environment must be designed properly.

Standard IT best practices confirm the importance of deploying a balanced system environment. The weakest link in the design will limit system-level performance. User productivity can be improved by monitoring system performance, identifying system performance bottlenecks, and allocating system resources to upgrade and expand capacity of these "weak links" in the system design. Understanding the distribution of software processing loads and the associated network traffic generated by GIS workflows during peak operations is an essential prerequisite to selecting the right hardware and network bandwidth to handle and transport these workflow loads. That's why such understanding is the foundation for the system architecture design process.

Since 1992, Esri has given my team the opportunity to develop and continue to improve a system architecture design process that can share specifications for a balanced IT solution. Investment in hardware and network components based on a balanced system load model provides the highest possible system performance at the lowest overall cost.

The chain (figure 1-4) represents the several factors that are linked in a system and therefore affect system performance. User workflows must be designed to optimize interactive client productivity; work request queues should be established to manage heavy batch processing loads and enhance user productivity. Some information products are slower in the making than others, so user display requirements should be carefully evaluated: do you really need such a high-quality map when a simple one would do? Simple information displays promote quicker display performance and therefore a more productive user interface. The same holds true for the geodatabase design and database selection: how can they be optimized to make the best of user performance and productivity? If complex data models are needed for data maintenance, then possibly a simple distribution database replica could balance that out by providing high-performance view and query operations. The system platform components you select (servers, client workstations, storage systems) must perform adequately and provide the capacity to fulfill peak user workflow requirements. In addressing performance needs and bandwidth constraints over distributed communication networks, the system design strategy should strike a balance between power (or quality) and economy (efficiency). The technology and

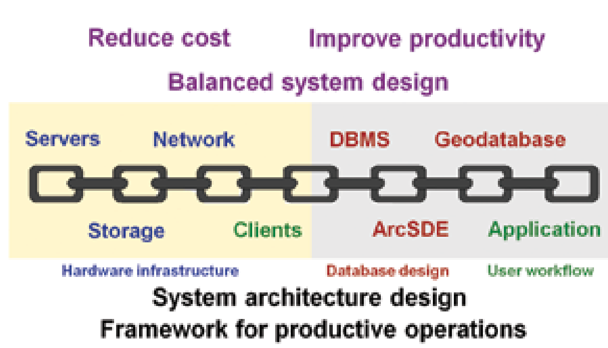



**Figure 1-4**

Balanced system design. A balanced system load leads to the highest system performance at the lowest overall cost.

configuration must be selected to conserve these shared infrastructure resources.

The weakest link (performance bottleneck) in this chain will determine the final system performance and capacity. That's why, in building a foundation for a productive operational environment, the system architecture design process must take into account every component of the system.

The primary infrastructure components contributing to system performance are identified in figure 1-5. These components include the user, application, and data source and how they are connected. What comes of their connections and relationships in the service of a system is the stuff of which performance models are made.

The application design and peak user loads impact network traffic and drive communication bandwidth requirements. The application design can contribute to communication chatter, which combined with network latency can increase user display response times and reduce user productivity. As traffic volumes approach available network bandwidth capacity, additional delays are caused by traffic contention (queuing). These key factors impact display response time and productivity for the remote user community.

Hardware selection (processor core performance and physical memory) impacts server performance and capacity. The software (software component installation) and platform architecture (single- or multi-tier) impacts system capacity, system load profile, and display service times.

The data center network configuration also can impact system performance. Server network interface cards must have adequate bandwidth capacity to handle peak traffic loads and avoid traffic contention in the data center—this is where all user workflows come together to access shared applications and data resources.

Hardware selection, physical memory, and the selected storage configuration can impact access to available data resources. The type of storage format (data source) and how data resources are distributed in storage can contribute to performance bottlenecks.

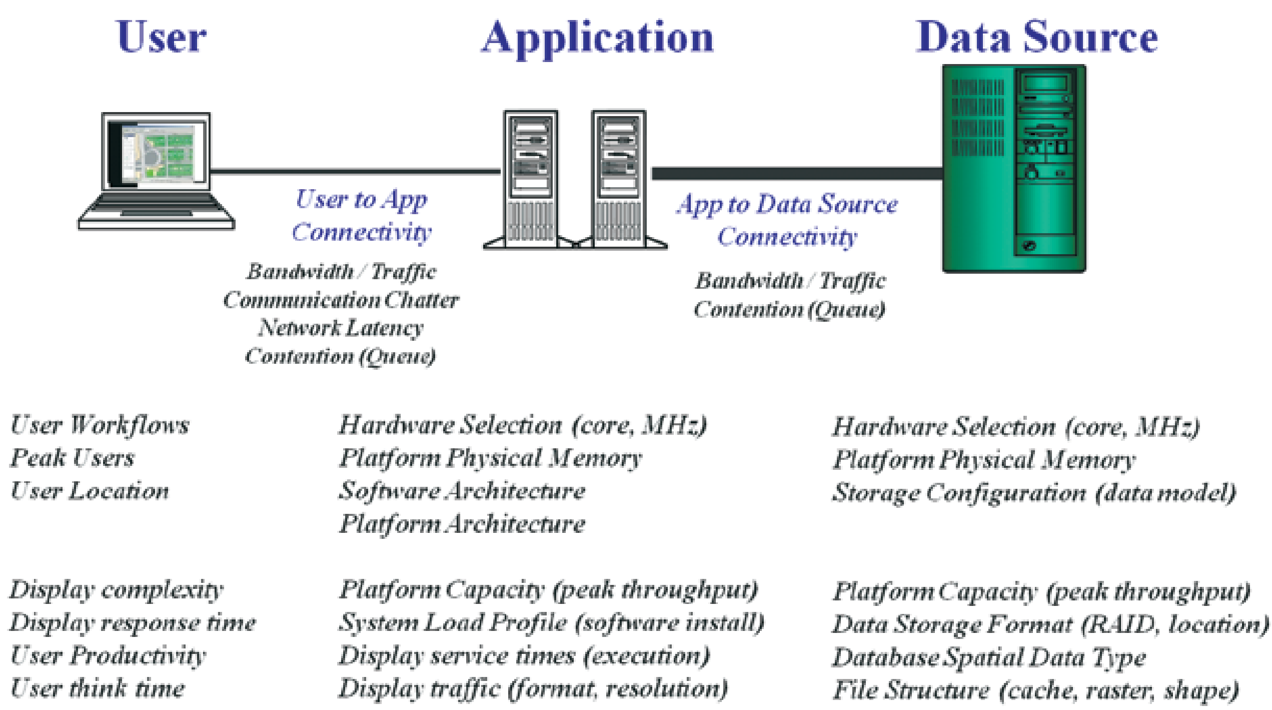



**Figure 1-5**

Performance terms and measures. Many factors contribute to system performance and scalability. Some are related to established business requirements (user workflows, peak users, user productivity, user think time, user locations), and others depend on how we design the software applications (display complexity, display response time).

Putting these pieces together is what we call system architecture design.

Almost every interrelationship can be factored in for use in your thinking about and modeling of a system. The peak number of concurrent users and their locations represent the user access requirements. As far as display complexity, how simple can we make the display or efficient can we make the application workflow? The requirements for display response time and user think time translate to user productivity needs. The application workflow establishes what you need in the way of platform capacity to satisfy the display service-time requirements. Selecting the right hardware, the appropriate platform architecture, and understanding how these decisions impact system processing loads will define the cost and performance of the system. The data source determines the database server capacity and storage requirements. The application and data architecture, combined with peak user workflow requirements, establish the network communication requirements (bandwidth, traffic, latency). These are the pieces that make up the system architecture design. We will be examining each of the pieces throughout this book as we identify the primary terms and relationships that contribute to system performance and scalability and a proper system architecture design. We will also describe how these critical design components perform under load, and define the relationship between component utilization and system delays (queue time) as we explore user performance and productivity.

## What is the system design process?

The traditional GIS planning process includes a GIS needs assessment and a system architecture design. The GIS needs assessment identifies what you want out of the system, and describes specific information products, data requirements, and application needs that will improve business decisions and workflow productivity. The system architecture design uses system design guidelines and platform-sizing models to establish vendor hardware specifications and identify peak network traffic requirements. Platform performance specifications are calculated based on user workflow requirements pinpointed in the GIS needs assessment.

Timing is important in the decision-making process, so the most effective system design approach is one that considers user needs and system architecture constraints throughout the design process—an integrated business needs assessment. It's a kind of holistic approach, if you want to look at it that way, but methodical nonetheless, and we don't plan to take into account everything that affects user productivity (only what's important to promote implementation success). Figure 1-6 provides an overview of the system design process as an integrated business needs assessment.

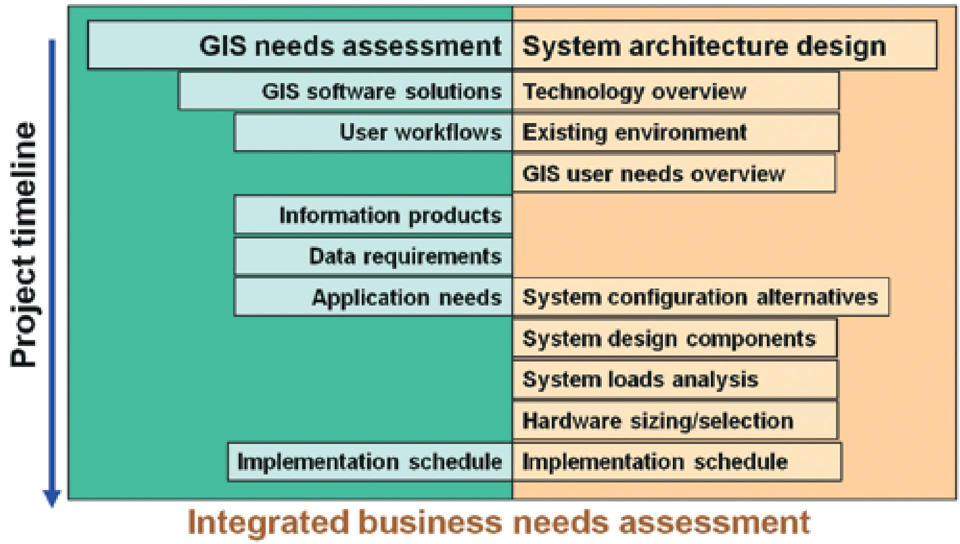



**Figure 1-6**

Integrated business needs assessment. System architecture design should be conducted throughout the user needs assessment.

### GIS needs assessment

The GIS needs assessment identifies how GIS technology can be leveraged to improve organizational or business operations. A fundamental objective is to clearly understand and identify what information products you want out of the GIS. Once you understand what information you want from the GIS, you are prepared to identify GIS application and data requirements and develop an implementation strategy for meeting these identified needs. The user requirements analysis is a process that must be accomplished by the user organization (improving user workflows, identifying more efficient business practices, and setting organizational goals and objectives). This is about reviewing how your business works and identifying how GIS can be leveraged to make your business more efficient and successful. An experienced GIS consultant who understands GIS technology and system design best practices can help facilitate this planning effort.

### System architecture design

The system architecture design is based on the peak user requirements identified by your GIS needs assessment. You must have a clear understanding of your GIS functional and data requirements before you are ready to develop system design specifications. System implementation strategies should schedule hardware purchase requisitions *just in time* to support user deployment needs.

The system design normally begins with some level of what's called a *technology exchange*. In other words, people working together to understand and share what they need to know (the information in the first eight chapters of this book). User participation is a key ingredient in the design process. Once the technology exchange is complete, the process includes reviewing the existing computer environment, GIS user requirements, and current GIS design alternatives. The system design capacity planning tools translate peak user workflow requirements to specific platform specifications. An integrated implementation strategy is developed to achieve GIS deployment milestones.

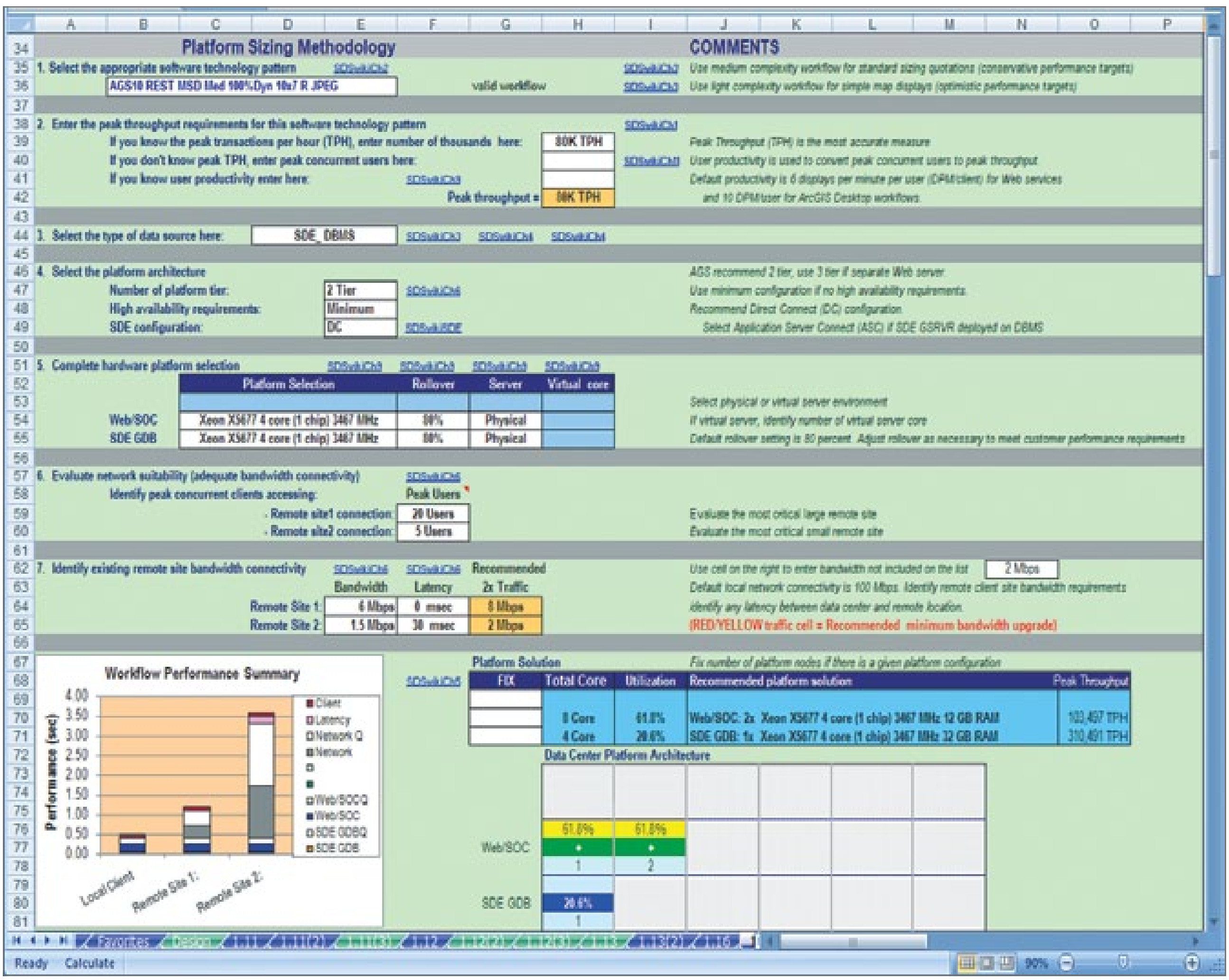
Traditionally, the user needs assessment and the system architecture design were two separate efforts. But there are some key advantages in completing these efforts together. GIS software solutions should include a discussion of architecture options and system deployment strategies. The existing hardware environment and information on peak user workflows and user locations can be identified during the user needs workflow interviews. Technology selection should consider configuration options, required platforms, peak system loads for each software technology option, performance and scalability, and overall system design costs. And finally the system implementation schedule must consider hardware delivery milestones. A primary goal for developing the CPT presented with this document is to automate the system architecture design analysis so that GIS system architects can use the capacity planning tools to complete an integrated business needs assessment, considering system architecture design implications throughout the user needs assessment process.

An integrated business needs assessment (user needs with an integrated system architecture design assessment) is by far the best way to complete your planning process. You can easily integrate the system architecture design process with your user needs assessment by understanding the content presented in this book and using the CPT. You can use this tool—and the several modules within the tool—to document your workflow requirements and support your technology decisions during the user needs assessment process: this will be your integrated business needs assessment. In this step-by-step process for completing the integrated business needs assessment (chapter 11), we will focus on the system architecture design and touch lightly on the user requirements analysis. (Roger Tomlinson's book, *Thinking About GIS: Geographic Information System Planning for Managers*, provides a more complete step-by-step process for the user requirements analysis.) The same City of Rome case study provided in this book is presented from a user needs perspective in chapter 9 of Tomlinson's book.

You can use the CPT to provide an overview of your business needs, infrastructure specifications, and to establish your hardware vendor platform requirements. The CPT shows you the expected system performance for a variety of technology options. You can identify your network and platform environment and see if the technical solution you are considering can work for you. You can demonstrate many of the system performance problems you may be experiencing in your current environment.

Technology decisions can be made based on a full understanding of user workflow requirements and properly established system performance expectations. Once the system is operational in your environment, you can continue to base technology decisions on a complete understanding of your system performance and scalability. Establishing performance target milestones, based on credible information about the technology, will reduce implementation risk and build a framework to manage your implementation success.

Figure 1-7 shows a view of the CPT Platform Sizing Worksheet. Inputs required to complete this selection will be discussed in the first 7 chapters of this book. This is a simple tool designed to walk you through the workflow platform-sizing methodology. The process involves selecting your software technology patterns (chapters 2–3), entering your expected peak throughput loads (peak transactions per hour or peak concurrent users), selecting your data source (chapter 4), choosing your desired platform architecture (chapter 6), making your platform selection (chapter 7), completing your network suitability analysis (chapter 5), and viewing your final platform solution (platform solution list, platform architecture drawing, and Workflow Performance Summary).




## Figure 1-7

CPT Sizing tab.

The Platform Sizing Worksheet (CPT Sizing tab) walks you through the design process. Workflows are selected from the CPT Workflow tab, and the system design analysis is completed using the CPT Calculator functions:

1. Select the user workflow or web service.
2. Enter the peak throughput requirements (TPH or concurrent users).
3. Select the type of data source.
4. Select the platform architecture.
5. Complete the hardware platform selection.
6. Identify key remote site peak concurrent user loads.
7. Confirm adequate bandwidth connectivity (network suitability analysis).
8. Display analysis results (Platform solution, data center multi-tier hardware graphic, and Workflow Performance Summary).

**Primary and secondary factors to consider**

Distributed GIS solutions bring together a variety of vendor products. Each vendor technology is a part of the total enterprise that must be integrated into the existing system environment. Integration of any multivendor environment is made possible through voluntary compliance with generally accepted industry interface standards. When each new component is integrated into the system, the entire functionality, performance, and security of the system can be affected, and the following primary factors must be considered:

- Functionality: Does the integrated system meet your functional workflow requirements?
- Performance: Will the integrated system satisfy your performance needs during peak workflow operations?
- Security: Does the final system environment support your enterprise security requirements?

The primary design factors of functionality, performance, and security often dictate the initial system architecture design strategy. In many cases, policy issues, driven by established compliance factors or other technical or nontechnical issues, limit platform technology options and restrict deployment on older software versions. At first glance, the system architecture design can appear to be preordained, requiring only a simple implementation decision.

However, several secondary design factors should be reviewed since they often dictate implementation success or failure. These factors include the following:

- Cost considerations
- Scalability (ease of handling more users or higher volumes)
- Reliability (eliminating single points of failure)
- Mobility (support field editing or viewing)
- Availability (dependence on internal and external services)
- Quality of service or data
- Software stability
- Maintainability (centralized versus distributed architecture)
- Flexibility (adaptability to change)

Reviewing and understanding these secondary design factors are necessary to ensure success in the context of your specific implementation strategy. A complete design process will include a review of both the primary and secondary design factors to ensure proper technology selection and appropriate deployment strategy.

**What system architects do**

System architects establish the target architecture during the system design process, purchase and install the hardware needed to support the design, and resolve any performance issues during final implementation. Once the design is approved and the project is funded, the system architect is responsible for a final review and update of the hardware specifications right before procurement. That person is also responsible for scheduling the vendor installations and for participating in the monitoring and testing of system functionality and performance at each of the implementation milestones (see chapter 12).

Enterprise GIS environments include a broad spectrum of technology integrations; in other words, the system architect must ensure all these integrations play well together. Most environments today are composed of a variety of hardware vendor technologies, including database servers, storage area networks, Windows Terminal Servers, web servers, map servers, and desktop clients—all connected by a broad range of local area networks (LANs), WANs, and Internet communications. All these technologies must function together properly to provide a balanced computing environment. So if the business did not begin with GIS, the system architect will be integrating new systems with existing business operations when you "go live" or launch a GIS. A host of software vendor technologies, including database management systems, ArcGIS Desktop, and ArcGIS Server software, web services, and hardware operating systems all must operate seamlessly with existing legacy applications (existing business and user applications are added to the integrated infrastructure environment to enable the final implementation). This large mix of technologies must work together properly and efficiently to fulfill user workflow requirements.

Final purchase decisions are influenced by both operational requirements and budget limitations, which introduce unique challenges for the system designer. Good leadership, qualified staff, and proven standard practices lead to successful deployments, so a wise system architect creates a team for the GIS project to build from there. (Tomlinson's book, *Thinking About GIS*, describes the ideal GIS project team composition and methodology from the beginning stages of planning for a GIS. The book you have in hand takes off from there.)

## Success with GIS

Selecting the right technology is the first step toward deployment success. Software and hardware vendors develop their technology to support what they believe is generally accepted community interface standards. Ideally, all these vendor solutions will work together as a seamless system environment. In reality, both technology and standards are constantly changing. Technology that is used most together works best together. New technology introduces risk, and in most cases must be integrated with other system components over some period of time for interface anomalies to be identified and resolved. It can seem as if, just when you figure out one, another element comes along, and you must figure it out all over again. Each component introduced, whether it's a first launch or an upgrade, must be reevaluated in the context of the whole. So, how often are you going to get lucky if you do this all by happenstance? For this reason, it is paramount to plan carefully, implement in stages, and ideally stay one step ahead of the changes. Several critical deployment stages lead to a successful implementation. Understanding the importance of each and the key objectives within them result in more effective enterprise implementations. Figure 1-8 shows the different stages of GIS deployment and the advantages of taking an iterative approach to managing implementation risk.

### Requirements stage

Understanding technology alternatives, quantifying user requirements, and establishing an appropriate system architecture design deployment strategy are critical


**Figure 1-8**

Building a GIS process.

during the requirements stage. Capacity planning during this phase can establish preliminary hardware performance specifications and validate that budget and performance expectations can be achieved.

Figure 1-9 provides an overview of the CPT Calculator identifying the component sections used during the software technology selection process. Workflows are defined by selecting the appropriate software technology and software performance parameters (E1:I6). A complete workflow performance analysis is conducted by completing the remaining selections (user inputs are identified by the white cells) in the CPT Calculator. The references to E1:I6 and others like this throughout the book are Microsoft Excel shorthand for identifying a particular spreadsheet location. Cells are identified by the intersection of a column (E) and a row (1). Arrays are identified by identifying the top left and bottom right cells. In this case, for example, E1:I6 describes the box in the Excel spreadsheet identified by cell E1 and I6.

The software technology and selected performance parameters generate GIS workflow performance targets that can be selected for the system architecture design. The CPT Calculator generates workflow performance targets from technology baselines and performance adjustment factors derived from Esri performance validation test results. Workflow performance targets are updated with each software release to incorporate the latest performance metrics.

The CPT Calculator also provides system architecture design and performance information for a single workflow selection. User requirements (peak users and productivity or peak service transaction rates) generate system processing loads, and platform architecture (tier configuration, high availability, spatial database engine (SDE) connection, data source) and platform technology selection complete the system design inputs. The Calculator can also select Standard and Project workflows from the CPT Workflow tab for simple platform-sizing review.

Client network bandwidth connectivity inputs incorporate a preliminary communications assessment and provide local and remote user workflow display performance estimates. Platform configuration options include both physical and virtual server environments. System design solution is displayed in the platform-sizing section.

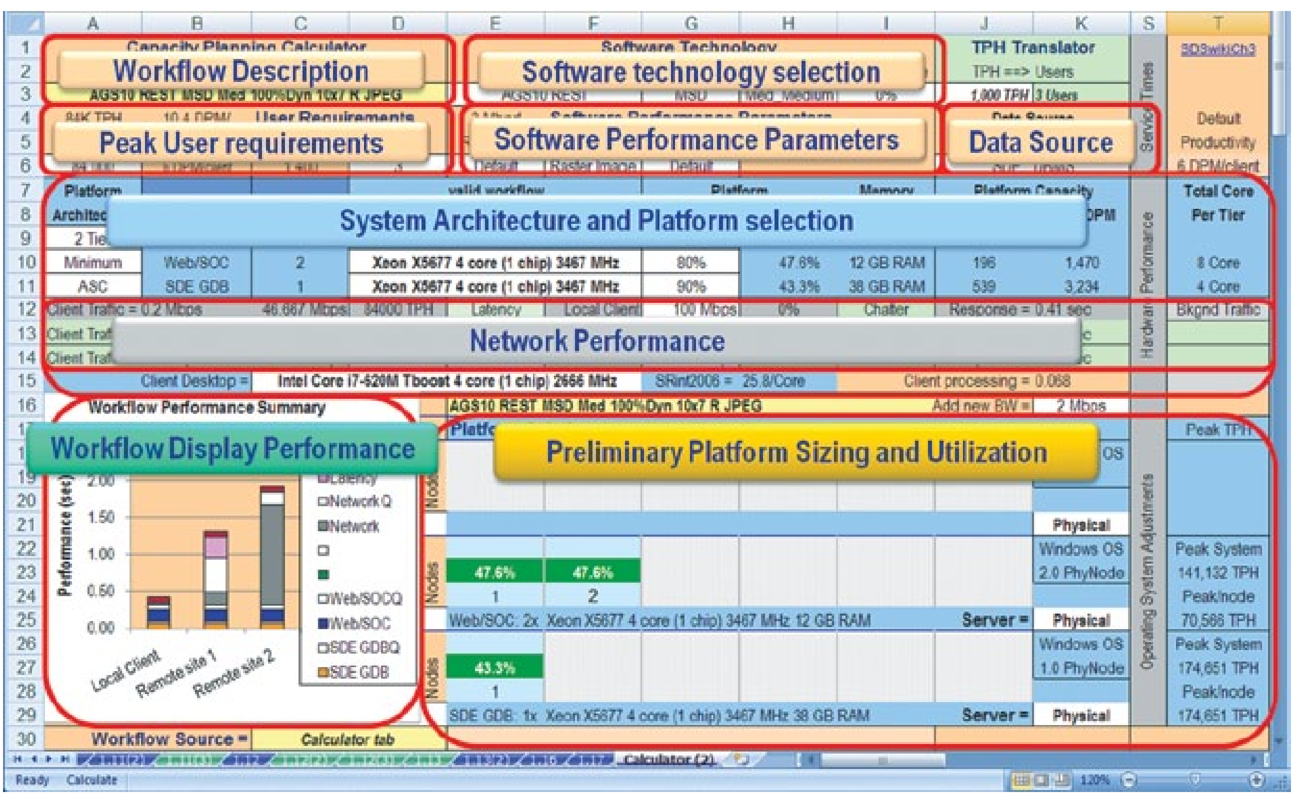



**Figure 1-9**

CPT Calculator tab. The CPT Calculator was developed to address user workflow analysis and software technology selection. Understanding user needs is the first step in building a GIS. Identifying the information product specifications and selecting the right software technology is an important part of the user needs analysis. The CPT Calculator includes drop-down inputs for software technology selection (software, map document, complexity, and percent data cache) and key software performance parameters (display resolution, density, and output format). Chapters 2–3 will describe the available GIS technology patterns and software performance in more detail.

The CPT Calculator provides a complete system design assessment for an informed software technology selection based on your estimated user requirements and your software technology selection. Functionality includes a single workflow platform-sizing assessment, which addresses most customer platform-sizing questions confronted by sales, support, and technical marketing staff.

A review of available software technology patterns and software performance parameters can provide foundation for proper technology selection. We will review software technology patterns and discuss software performance in the next two chapters. User workflow analysis, aided by the CPT Calculator, can help you select reasonable workflow performance targets for use in the system architecture design.

Figure 1-10 highlights the role of the CPT Workflow tab in selecting appropriate user workflow performance targets in preparing for your system design. Software technology deployment patterns are explored in chapter 2. Software performance for each of these technology patterns is discussed in chapter 3 (along with the results of Esri benchmark testing efforts to baseline standard software performance profiles) and identifies key software performance parameters for establishing workflow performance targets.

A workflow nomenclature provided by the CPT Calculator identifies the recipe used to generate the selected software performance targets. This recipe identifies the software technology deployment pattern and the selected software performance parameters used in generating the workflow service times. Several Standard Esri Workflows are provided to help estimate reasonable user and service workflow load profiles. Understanding the business workflows, processing complexity, and information product requirements are key in making a proper technology selection.

The Standard Esri Workflows share performance goals for establishing reasonable performance budgets that can be delivered to satisfy your peak business

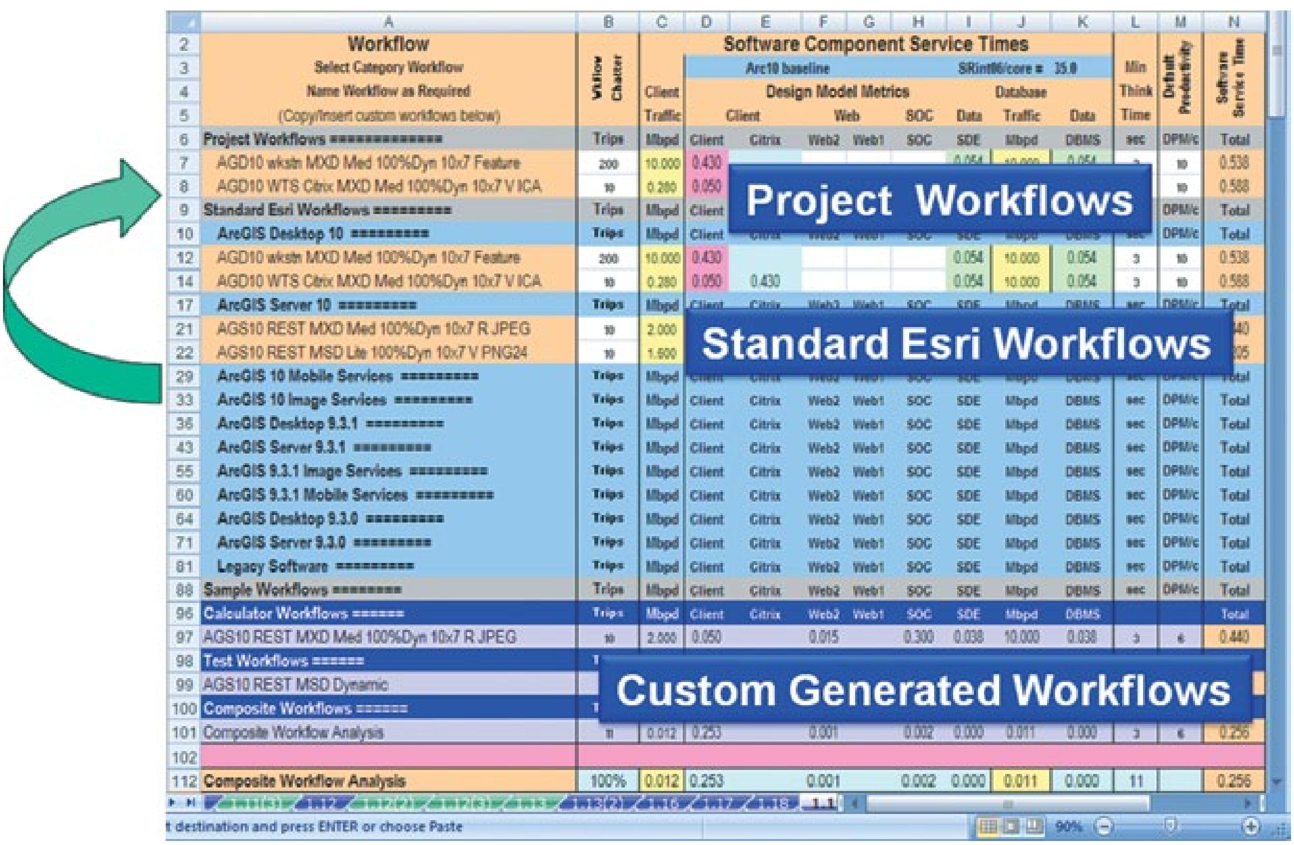



**Figure 1-10**

CPT Workflow tab. The CPT Workflow tab provides a lookup table for workflow performance targets (component software service times) used in the CPT Design tab. Standard Esri Workflows are maintained in the Workflow tab for easy selection in establishing project workflow performance targets. The Standard Esri Workflows were generated from the CPT Calculator and the workflow name identifies the software technology and performance parameter selections. Workflows not included in the Standard Esri Workflow list can be generated from the CPT Calculator tab, with the Calculator workflow service times provided for use on the CPT Workflow tab.

requirements. High complexity workflows generate heavier system loads that result in hardware and network infrastructure specifications that can increase your project budget. Light complexity workflows require less hardware and infrastructure bandwidth, which can reduce your project budget; tight budgets require careful project management during application and database design to deliver with the lower performance targets (this increases project risk).

A composite workflow analysis section is included on the CPT Workflow tab used to generate representative workflow service times for a combination of two or more workflow services. The CPT Workflow tab provides the primary source for the selected workflows. Once the composite workflow analysis is complete, the composite workflow service times are available (line 101) to include in your design.

The CPT Calculator can generate custom project workflows by completing the software technology selection and the associated software performance parameters (E1:I6 in figure 1-9). The CPT Calculator generates workflow service times based on the selected performance parameters, and the result is displayed on the CPT Workflow tab (row 97 in figure 1-10) for use in your design. CPT Calculator selections can be used to provide a broad variety of focused workflow load profiles.

The CPT Test tools can also be used to better quantify workflow loads with throughput and platform utilization values collected from prototype test results. Results generated from the CPT Calculator and Test tabs can be used to establish custom workflow performance targets for most any user workflow or service. Prototype testing during the early planning stage can be expensive and reduce the budget available for project implementation. Thus, this testing normally is appropriate only when a cost savings can be realized by a more refined validation of workflow performance targets. Remember, these are performance targets, and there will be many opportunities for test and validation during the software application design and system implementation phases.

The peak number of current users and estimated web service rates are identified during the requirements phase and developed during your user needs assessment. Peak concurrent users and workflow productivity establish system throughput loads that drive the system architecture design. System loads generated by a variety of software technology solutions from local and remote user workflows are common characteristics of a distributed GIS system design. Understanding user locations and infrastructure limitations will often impact software technology selection. An initial workflow performance evaluation can be completed using the CPT Calculator (figure 1-9). Figure 1-11 highlights the role of the CPT Design in completing the user requirements analysis.

The requirements analysis module includes a row for each user workflow configured on each user network location. Network connections are represented by *gray*- and *green*-colored rows. *Gray* rows represent the data center network connections, including LAN, WAN, and the Internet. Remote site network connections are represented by *green* rows located within each data center network. User workflows are represented by

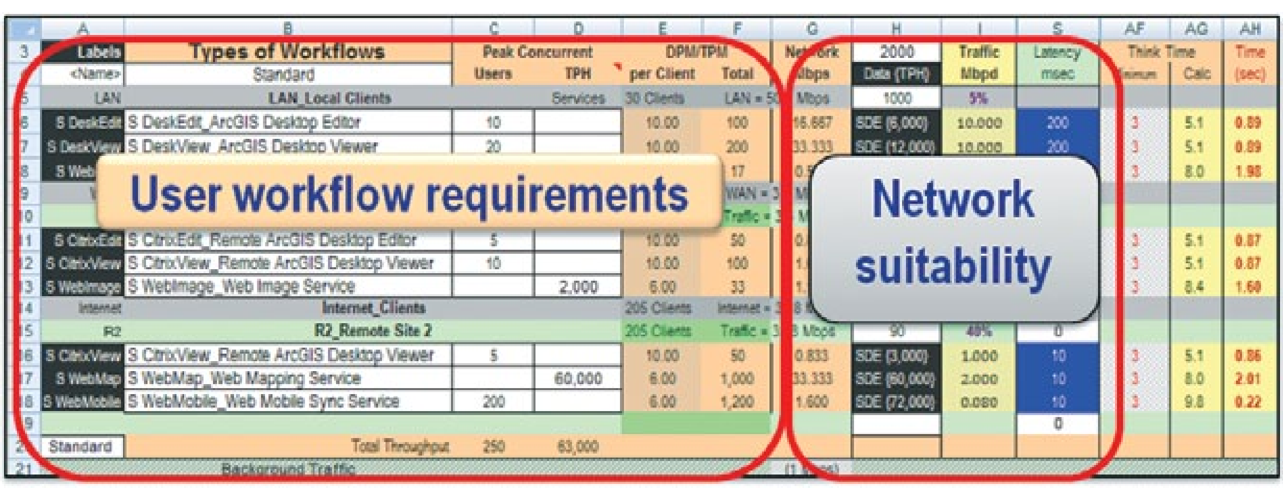



**Figure 1-11**

CPT Design Requirements Analysis module. The CPT Design tab includes a Requirements Analysis module for identifying business workflow needs. User workflow requirements are identified based on user locations and peak workflow loads (concurrent users or service transaction rates) for each system implementation milestone. The CPT generates the network traffic loads to complete an enterprise-wide network suitability analysis. Software configuration (figure 1-13) and platform selection (figure 1-14) complete inputs for the system architecture design.

the *white* rows located within the appropriate data center network. You can add network rows and user rows as required to represent your enterprise configuration. Network bandwidth is selected in column H.

The *white* cells located on the user workflow rows are where you identify your user requirements. Workflow names are selected in column B from a drop-down list of workflows available on the Workflow tab. Peak users and services are entered in columns C and D.

Once you have entered your user requirements, the CPT Design tab will complete a network suitability analysis. We will discuss configuring the user requirements in more detail in chapter 5 on network communications.

The CPT Design provides a framework for collecting and modeling your business requirements and the network infrastructure available to support peak system loads. User requirements should be identified by your business operations staff. Your IT network administrators identify available network infrastructure capacity. The CPT Design completes a preliminary network suitability analysis once you identify user locations, peak system loads, and the associated network bandwidth connections. Chapter 5 will discuss how to address network communication constraints in your design and how infrastructure limitations can impact user display performance and workflow productivity.

**Design stage**

System development and prototype testing validate functional interfaces and performance target compliance. Functions must work and performance targets must be met to enable follow-on deployment. This is where time and money are invested to build and test the selected environment. Initial prototype testing demonstrates functionality and reduces system integration risk. Preliminary software performance metrics can validate initial capacity planning assumptions and confirm that performance targets can be satisfied.

The system design includes vendor hardware selection. Figure 1-12 highlights the role of the CPT in sharing vendor hardware performance and capacity benchmarks. The CPT Calculator and Design tabs use the CPT Hardware tab as a resource to complete the hardware-sizing analysis. Vendor hardware performance

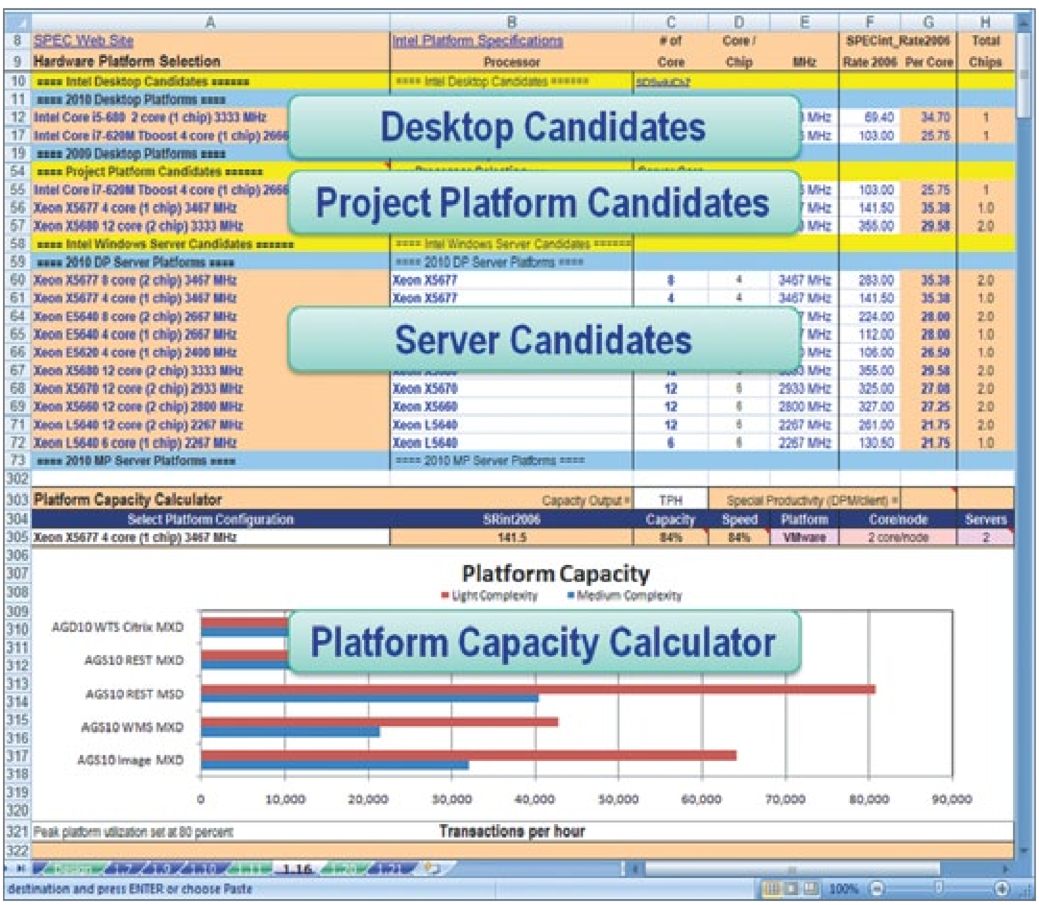



**Figure 1-12**

CPT Hardware tab. Hardware platform lookup and selection tools.

benchmarks are used to translate peak business workflow loads to the selected hardware platforms (the system architecture design analysis is completed by Excel as you complete your data center configuration). The CPT Hardware tab also includes a simple Platform Capacity Calculator, generating peak capacity ranges for five sample GIS workflows based on your platform selection. History on platform performance and the use of hardware vendor benchmarks will be discussed in detail in chapter 7.

The selected platform architecture and software installation will determine availability and scalability of the final hardware design. Figure 1-13 shows the CPT Design software configuration module. Software components for each workflow identified in your user requirements analysis are highlighted in this section and can be installed on a selected platform tier located in the CPT Design platform selection module. Chapter 6 will discuss available software architecture patterns and how they can be represented in the CPT Design. The CPT software configuration module provides flexibility to define software component platform installation and data source selection for a variety of data center platform configurations.

Figure 1-14 provides a full view of the CPT Design tab. Hardware selection for each platform tier is made from a drop-down list provided in column B in the platform selection section (A21:R63). Platform performance can significantly impact user productivity and peak system capacity. The CPT Design tab provides an adaptive data center architecture for configuring up to 10 separate platform tiers, with server nodes on each tier able to expand (scale out) to represent peak system capacity design requirements. Platform performance and scalability will be discussed in chapter 7. Platform capacity is a key design component contributing both to hardware platform and software licensing system deployment costs.

The CPT Design tab provides the platform solution (number of nodes and platform utilization) based on an integrated analysis of the user workflow requirements and platform selection. A platform tier can be configured as a Physical or Virtual server environment. A Workflow Performance Summary provides an average display response time for each user workflow. CPT Design includes the ability to adjust user workflow productivity to represent performance within selected network and platform constraints.

Each CPT Design tab models system performance for a comprehensive project deployment milestone. Each design sheet identifies individual user workflows, peak workflow loads (concurrent users or service transaction rates), network bandwidth and projected network utilization, selected platform architecture and software and data source configuration, platform selection and projected platform utilization, final hardware selection requirements, and expected workflow average display response times and productivity for each user site location.

Figure 1-14 highlights an overview of the CPT Design graphic Platform sizing and utilization display (AE23:AP39).

User productivity is directly related to workflow processing complexity. Workflow display performance depends on user location, network connectivity, software install, platform selection, and peak system loads. Figure 1-14 shows calculated user workflow display performance in the CPT Design Workflow Performance Summary (AI3:AP20). This diagram provides average display performance for each CPT Design user workflow based on the selected platform solution, network connectivity, and a comprehensive overall system-loads analysis. This valuable representation of design performance expectations is based on user location and selected workflow performance targets. User workflow display performance at locations throughout the enterprise environment is an important consideration in making a final software technology selection. The selected design solution should be configured to satisfy user performance needs.

#### Construction stage

The CPT Test tab provides several tools to translate measured performance metrics to software technology workflows used for capacity planning. The goal is to develop and deploy software technology solutions that perform within the planned performance targets. This is crucial during the construction phase, because if design software solutions exceed initial workflow performance targets, system design adjustments must be made to meet peak business performance needs. A change control review process should be in place to evaluate these design changes and ensure adjustments as needed.

The CPT Test tab provides four different ways to translate live software technology performance measurements to CPT workflow service times. Measured throughput transaction rate (transactions per hour) is the most accurate. Throughput (TPH) entry will override the other three Test tab input options. Peak users is the second throughput input option. In this case, peak users is multiplied by user productivity (displays per minute/user) to estimate the throughput transaction rate. The third and fourth options use measured map display render time to generate workflow service times. Cell B12 configures the tool for measured ArcGIS Desktop or ArcGIS Server render times (an ArcGIS Desktop author can use the MXDPerfstat tool or ArcGIS Desktop Map Publishing tool to measure ArcGIS Desktop MXD (map document) render time. The PerfHeatMap tool can be used to generate

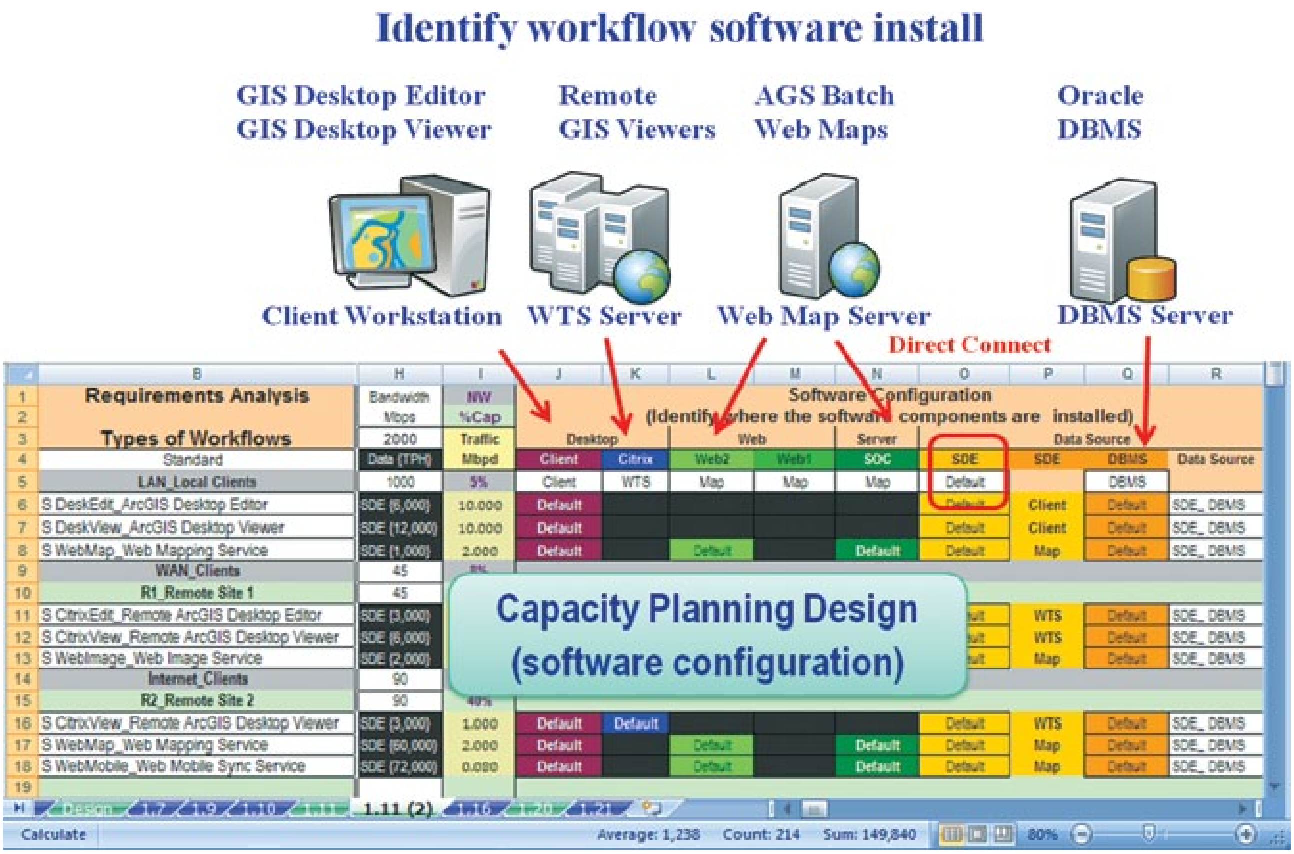



**Figure 1-13**

CPT Design Software Configuration module.

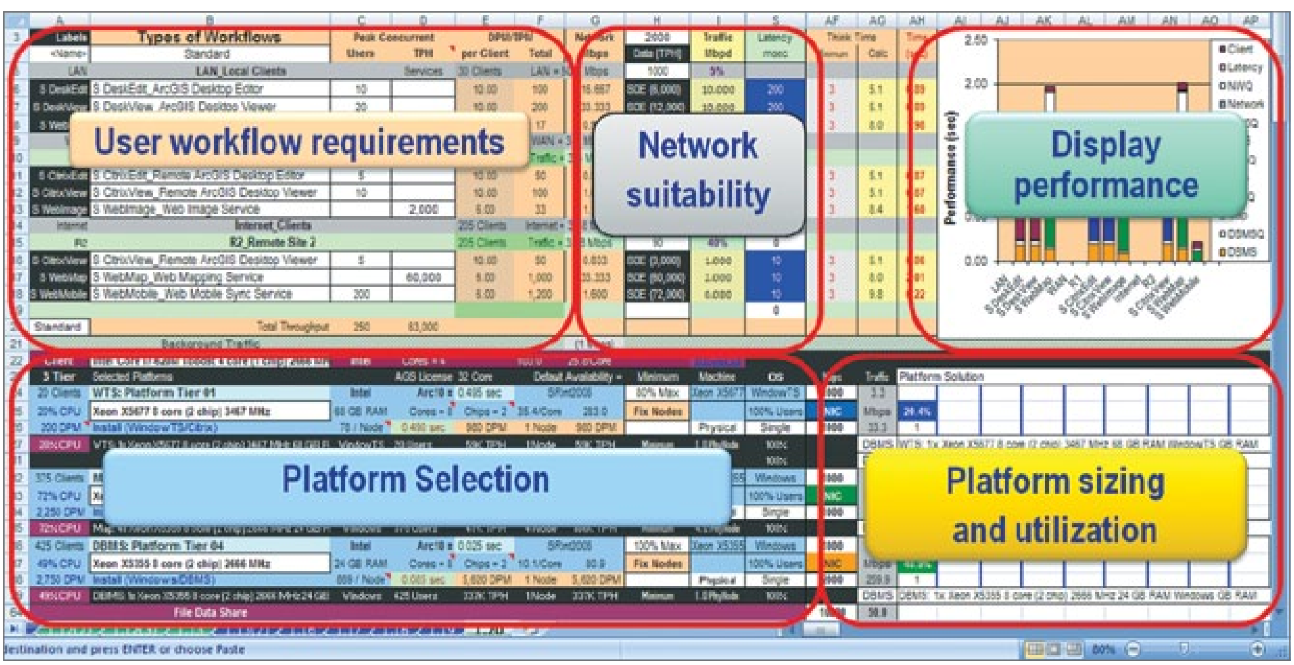



**Figure 1-14**

CPT Design tab. The CPT Design tab was developed as a framework to configure and execute an enterprise level system architecture design analysis. User workflow needs are identified in the Requirements Analysis module (figure 1-11). Network traffic loads are generated to complete an enterprise remote site network suitability analysis. Software configuration was addressed in figure 1-13.

ArcGIS Server map service rendering times throughout a defined service area. These are samples of tools that can be used to measure map display complexity (render time) during initial map service design and deployment. The Test tool links map display complexity to the selected web-mapping software technology pattern to generate baseline workflow service times.

The CPT Test tab can be used by GIS authors who create MXDs for web publishing. Map display rendering time can be measured during map publishing optimization or after initial server deployment. This is the first milestone in validating workflow performance. During initial deployment, measured throughput values (map display transactions per hour or peak concurrent users) can be used to validate workflow performance. During live operations, system performance measurements can be used to validate design compliance.

The first tool shown in figure 1-15 is used for translating ArcGIS Desktop-measured display rendering times to map service workflow service times. ArcGIS Desktop users can use this tool when they create and publish map documents as a web service. The ArcGIS Desktop map document rendering time can be used as a measure of map display complexity. You can use the Measured Performance Desktop tool to estimate web service workflow service times during the authoring process (based on map display complexity). Professional GIS cartographers responsible for map design should have a performance target identified for their specific workstations to measure display complexity. This is an opportunity get the map display complexity right before the display becomes part of the map service.

Once the web service is published, the ArcGIS Server version of this same tool can be used to estimate published workflow service times. A PerfHeatMap test script is available on the Esri ArcScripts website that can be used to measure and map web service render times for a specified map extent. The render-time measurements along with the ArcGIS Server Measured Performance Translation tool can be used as a quantitative crosscheck that the published web service is performing within the established performance targets. The CPT translated workflow service times can be used to validate your established workflow performance targets. You can quickly see if these services perform within your budget and make appropriate display complexity adjustments to meet your goals. Too many performance problems slip through during this phase and end up in production with embarrassing and sometimes painful operational shortfalls, especially if

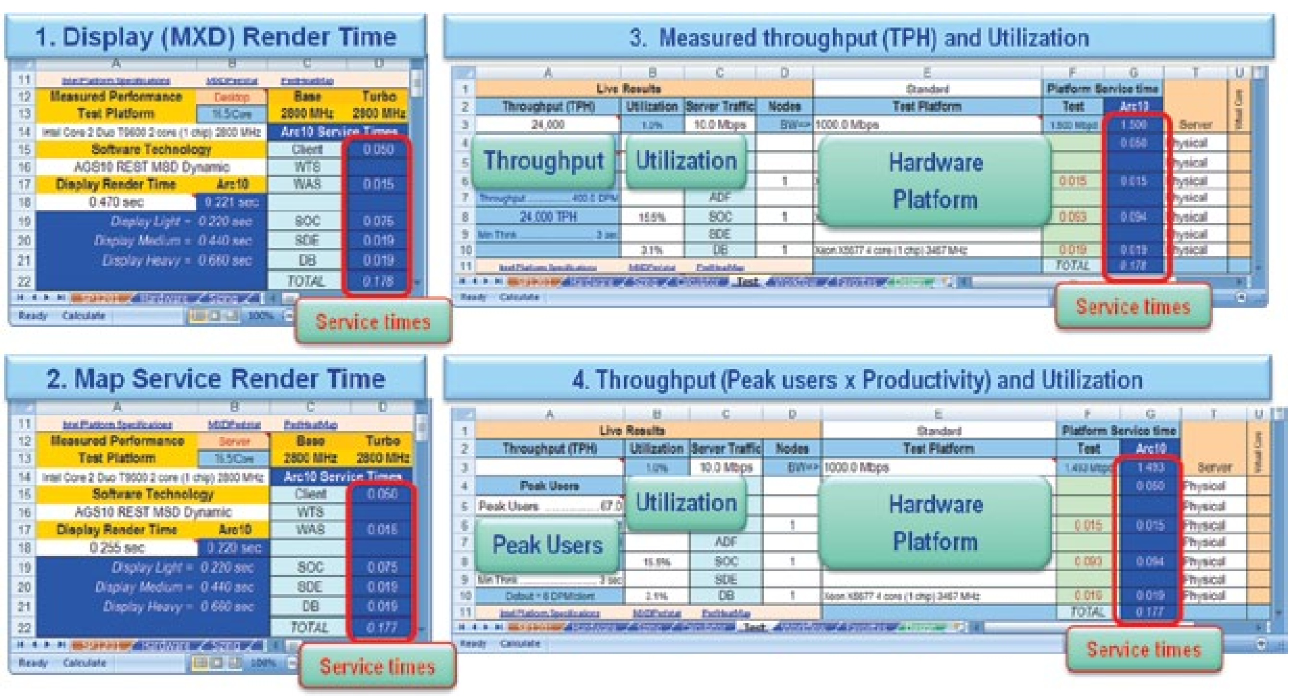



**Figure 1-15**

CPT Test tab. The CPT Test tab was developed to validate workflow performance compliance. The CPT Test tab does just the opposite of the Design tab, it takes live system-level throughput and utilization performance measurements as input values to calculate workflow software component service times (workflow performance targets that would generate these results on the Design tab). Planned workflow performance targets can be compared with design and deployed service measurements to validate workflow performance compliance (if peak business requirements can be supported by the planned design solution). The CPT translates results measured on existing hardware to performance expectations on the selected platform design solution.

the estimated peak user requirements were critical for operational success.

The third and fourth tools shown in figure 1-15 translate system throughput and utilization measurements to CPT workflow service times. You can input web service throughput measurements to calculate workflow service times. Peak concurrent power users can also be used as an input if measured transaction throughput measurements are not available. These tools can help translate live system throughput (peak user workflows loads) and platform utilization metrics (loads on the servers) to workflow service times to determine if your system is within performance budgets. These tools can also generate workflow service times from test workflow loads. You can decide whether to spend your money on lab testing or use live system performance measurements. System performance fundamentals are discussed in chapter 9 along with a complete review of the CPT Test tab and how you can translate system measurements to workflow service times for your design.

#### Implementation stage

Final system procurement and deployment demonstrate operational success. Capacity planning metrics can be used to monitor and maintain system performance goals and guide system performance tuning. Good planning, development, and testing will result in a smooth deployment, productive operations, and satisfied users.

Establishing workflow performance target milestones and managing software performance to achieve performance goals throughout deployment is a positive recipe for success. Building systems without regard to performance and scalability can lead to disappointing results and costly recovery. The cost of having to make a technology change increases exponentially as the project implementation proceeds through deployment. A change or correction made toward the end of the process can cost a thousand times more than if you'd identified the problem and made the change at the beginning. This demonstrates the importance of getting it right during the planning phase and making small adjustments along the way to optimize the final system deployment.

In the 1990s, much of the system cost was in data migration followed by the relatively high cost of technology (software and hardware). Systems would be deployed over several years and spread costs over its lifecycle. The environment today has changed. Data is much more readily available (much faster acquisition time), and still leads as the highest overall GIS investment. Hardware and software costs have dropped by orders of magnitude and are approaching commodity prices.

Technology is changing much faster these days. Because implementation timelines are short, managing implementation risk continues to be a top priority. Getting it right from the start is best done by taking the time to understand the technology, quantify user requirements, select the right software technology, and deploy the right hardware. Not getting it right from the start will cost money to make it right later—in system, labor, and costs. Sometimes, if repeated failures erode upper management's confidence, the cost is that you don't get to have a GIS at all, and the potential benefits are never realized.

## Why use capacity planning tools?

Capacity planning tools were developed to simplify the system design process and reduce implementation risk. GIS workflows generate network and server processing demands across the enterprise. GIS user productivity depends on the available capacity of these shared resources. There are many software technology patterns and performance parameters that impact workflow performance. It is not easy to put all these pieces together without modeling your system. The combination of many interrelated factors influences system performance. Capacity planning tools allow users to focus on technology selection and information product design. Workflow performance and the system architecture design solution are generated as a function of technology selection.

The CPT provides a model that connects user workflow requirements with system performance and scalability. Building a GIS based on a system design model that can both define and validate user workflow productivity requirements establishes a foundation for successful GIS deployment. Identifying proper system development and deployment performance milestones and validating workflow design compliance within established performance design goals will reduce implementation risk and ensure system deployment success. The CPT is designed to help GIS and IT managers understand their business needs and promote more productive GIS operations.

## Planning for success

Most enterprise GIS deployments evolve over years of commitment, planning, and hard work. It is essential in today's world to plan for technology change and update these plans on an annual basis. GIS project planning should be scheduled to support the annual business cycle. Enterprise GIS is an evolving program that changes each year to support business objectives and keep pace with technology.

The best plan is to understand what you're doing, every step of the way. Then, the technology won't get away from you and you won't run amok with it either. What do system architects do? They plan for success. They understand that computer systems are only as strong as their weakest link, and sometimes that weakest link is us. To understand a system as a whole, you have to know all its parts. To be unfamiliar with or misunderstand any one of them is equivalent to missing more than one opportunity to save money and time. For the sake of simply making it work, you owe it to your GIS project to move forward with an intelligent plan. Planning ahead before investing in software and hardware saves money and reduces risk. GIS projects most often fail because system performance requirements are not satisfied. Setting appropriate performance targets, following development best practices, monitoring performance throughout deployment to ensure performance targets are met, and using the models and tools provided in this book increase your likelihood of success.