Capacity Planning

Capacity is defined as the maximum amount or number that can be received or contained. For example, the amount of data that a computer hard disk can store is the disk's capacity. The maximum possible data rate received over a communication channel under ideal conditions is its capacity. Capacity can also refer to non-technical things such as the maximum amount of work that an organization is capable of completing in a given period of time or the maximum number of people that can physically occupy a room.

Discrepancies in capacity and demand results in inefficiencies associated with either under-utilized resources or unmet user demand. The goal of capacity planning is to minimize this discrepancy and to provide satisfactory service levels in a cost-efficient manner. The Information Technology Infrastructure Library defines capacity management as supporting the optimum, and cost effective, provisioning of services by helping organizations match their resources to the business demands.

Capacity planning doesn't always mean planning for periods of peak demand. Capacity requirements can vary greatly from times of peak demand to times of limited demand. As a result there may be drastic differences in the resources required to maintain normal operations during periods of peak demand. When planning for what levels of capacity will be supported it's important to understand and weigh the costs, benefits, and risks of delivering, or not delivering, certain level of capacity. Because of this, capacity planning is more of an art than a science. At a high-level such planning includes three steps:

1. Determine Capacity Requirements

Define required system performance to support associated business and process workloads. Capacity planning defined in business terms rather than utilization of CPU, memory, hard drive space, etc. simplifies planning for impacts on people, processes, growth, etc. Once defined, capacity requirements are then later used as inputs into identifying the planning of the physical infrastructure (technology, building, staff, etc) necessary to support the business requirements.

Contact prospective capacity providers as early in the project's life cycle as possible. It's important to make individuals and departments aware of the prospective impact a project may have on their organization. The earlier this can be done the more easily it can be mitigated.

Agree upon a unit of measure to assess conformance to defined system performance requirements. Note that the goal of capacity planning is to provide satisfactory services in a cost efficient manner. As a result, this unit of measure should be defined in terms of business units as well as the technical requirements necessary to support them. Associate it with a measurable quantity of work as opposed to the amount of system resources required to accomplish that work.

Test the system against defined business requirements and make adjustments where needed. This relates more to IT performance, as opposed to business functions, necessary to meet business needs. This may be accomplished by adjusting IT resources such as CPU utilization, memory usage, disk space, etc.

Agree upon minimum levels of service (capacity) the system must provide. These requirements are often defined by clients/users and should outline measurable system performance in terms that make sense to clients/users. This typically includes items such as response time, processing time for each request, minimum number of requests that can be processed in a given period of time, etc. Defined in these terms it's easier to illustrate success and make adjustments to the technical aspects of the system as necessary to meet business objectives.

2. Analyze Current Capacity

Understand the organizations current capabilities.

Before assuming that additional capacity is necessary a detailed understanding of current available capacity is required. A capacity study may need to be performed to evaluate the organizations current capacity.

Check usage of system resources necessary to support business requirements (CPU, memory, hard drive, etc). This relates more to IT performance, as opposed to business functions, necessary to meet business needs. If no current system exists it becomes even more important to understand and evaluate all available information regarding expected demands upon that system. This may include information such as historical project archives, industry standards, information obtained from vendors/customers, etc

*Record and track utilization of system resources to determine where capacity adjustments need to be made to support business processes as defined by the clien*t

3. Plan for Future Capacity

Forecast expected workloads for a particular period of time. Understand how changes in workloads affect business processes and the system that was built to support them. Translate that into technical requirements necessary to maintain the system at a level that satisfies user demands.

Plan for future system usage and plan adjustment to accommodate varying demands upon the system. Create a capacity plan that outlines the current configuration, required future configuration, and the steps necessary to accomplish any necessary system changes. Identify key capacity areas/items needing to be monitored and expected growth rate of these items. Include a defined response strategy for growth. Thresholds should be defined that represent utilization levels requiring action necessary to increase capacity. Such as utilization at 75%, 1,000 users, 10,000 visitors per day, etc. The goal is to mitigate the risk of a capacity bottlenecks before it becomes an issue. Specific actions should be taken early in response to increase in defined. Contingency plans then define what actions may be taken, in response to identified capacity triggers, in hopes of reducing the impact of capacity issues not mitigated through responses to growth thresholds.

Depending on the organization capacity planning may be applied using one of three strategies:

a. Lead strategy adds capacity in anticipation of an increase in demand
b. Lag strategy adds capacity after demand has increased beyond existing capacity
c. Match strategy adds capacity incrementally in response to changes in demand

For more information and tools related to capacity planning, the CDC Unified Process, or the Project Management Community of Practice please visit the CDC UP website at http://www.cdc.gov/cdcup/. ▍

Project Management Community of Practice

- *August 26, 2011*
  Leadership
- *September 23, 2011*
  Understanding Section 508
- *October 28, 2011*
  Information Security 101 for Project Managers
- *December 09, 2011*
  Enterprise Architecture

For more information on the Project Management Community of Practice visit the PMCoP website at http://www2.cdc.gov/cdcup/library/pmcop/ ▍

CDC Unified Process Presentations

The CDC UP offers a short overview presentation to any CDC FTE or Non-FTE group. Presentations are often performed at your facility, on a day of the week convenient for your group, and typically take place over lunch structured as one hour lunch-and-learn style meeting.

Contact the CDC Unified Process at cdcup@cdc.gov or visit http://www.cdc.gov/cdcup to arrange a short overview presentation for your group. ▍

Contact the CDC Unified Process

The CDC Unified Process Project Management Newsletter is authored by Daniel Vitek, MBA, PMP and published by the Office of Surveillance, Epidemiology, and Laboratory Services.

For questions about the CDC UP, comments regarding this newsletter, suggestions for future newsletter topics, or to **subscribe to the CDC UP Project Management Newsletter** please contact the CDC UP at cdcup@cdc.gov

http://www.cdc.gov/cdcup/

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention