

Capacity Management for IT Production

...

Theory and Practice

Dennis Adams Associates Ltd.

November 2004

What is Capacity Management?

- Capacity Management is the discipline that ensures IT infrastructure is provided
 - at the right time
 - in the right volume
 - at the right price
 - and ensuring that IT is used in the most efficient manner.
- This involves input from many areas of the business to identify
 - what services are (or will be) required
 - what IT infrastructure is required to support these services
 - what level of Contingency will be needed, and
 - what the cost of this infrastructure will be
- (Source: ITIL & ITSM World.)

The Objective of Capacity Management

- To ensure that cost justifiable capacity always exists for the following resource types:
 - Hardware (Operating Systems, Databases, but also power, heating, datacentre space...)
 - Networking Equipment (LANs, WANs, bridges, routers)
 - Peripherals (storage, printers)
 - Software (OS, network SW, purchase/in-house)
 - HR (where it impacts IT capacity)
- To ensure that this capacity is matched to the current and future business requirements
- (ITIL definition)

Why have Capacity Management?

- Business wants to focus on their Business Planning
- IT should enable the Business Strategy

- The Rate of change in the Business world is very high
- This results in a increasing rate of change demanded from the IT component, and increasing flexibility.

- Business are increasingly expecting “On Demand” IT.
- IT needs to deliver agreed, measurable services, for an agreed, measurable price, within an agreed, measurable timescale.

- IT Production, in particular, needs to become Pro-Active.
- This is a key element in the new Service Orientated Architecture (SOA) initiatives.
- IT PRODUCTION TEAMS MUST MANAGE CAPACITY!

Additional Benefits of Capacity Management

- Deferred expenditure
 - Defer the cost of new equipment to a later date
 - The money can be spent in other ways.
 - It may be possible to defer the expenditure permanently ?!
 - With the pace of technical improvement, the later a purchase is done, the more Capacity is obtained for the money
- Economic provision of services
 - Capacity is matched to business need.
 - Cost Savings as a result of not maintaining (and paying for) Unnecessary Capacity
- Planned buying
 - is always cheaper than panic buying

Capacity Management in context: ITIL, ISO & SOX

- The ITIL Library covers a number of IT management planning categories, of which Capacity Management is one.
- Although they do not have specific references to Capacity Management, the ISO 9000 and CMM (Capability Maturity Model) highlight the importance of processes which are:
 - Consistent
 - Repeatable
 - Auditable
 - Verifiable.
- The recent US Sarbanes-Oxley Legislation also emphasises this
 - pressure on European companies to comply with SOX
- Dennis Adams Associates are aware of these best-practices as a basis for designing IT processes for clients.
- At the same time, we strive to keep in touch with the practical, pragmatic issues involved with managing IT Production.

Capacity Management: Inputs and Outputs

- The inputs into the Capacity Management processes:
 - Performance monitoring
 - Workload monitoring
 - Application sizing
 - Resource forecasting
 - Demand forecasting
 - Modeling
- From these processes come the results:
 - the Capacity Plan itself
 - Forecasts
 - Tuning data, and
 - Service Level Management guidelines.
- (Source: ITIL & ITSM World.)

Capacity Management Inputs

- Performance monitoring
 - Capture metrics from running systems
- Workload monitoring
 - Breakdown total performance metrics by workload to analyse them.
- Application sizing
 - What are the requirements of the new applications coming on stream?
- Resource forecasting
 - Predict which resources will be needed for future growth.
- Demand forecasting
 - Allow for Organic (“planned”) growth and unplanned demand growth.
- Modeling
 - Trend Analysis = the easiest form of modeling
 - Can create a “pilot” system, measure the response & extrapolate results.

Phase 1: Performance & Workload Monitoring

- Capture the following basic performance statistics.
 - CPU Utilisation
 - Memory Utilisation
 - Disk Storage Utilisation
- On NT Servers, this can be done using “perfmon”
 - (See slide 17 re “Tools”)
- Breakdown of these figures by Application Workload
 - E.g. where two different types of activity are running on the same server, this should be split off.
 - Demand will change during the day, and night, so need to have a record of this. In practice, just extract “night” & “day” figures.
 - It is the peak demand which is important
- Focus on back-end servers first.

Possible Structure of Monitoring Stats

Date	Period	Application	ServerType	ServerName	CPU Equivs	Memory MB	Disk GB
10-Nov-04	Day	Resource_Planning	Application	LONS0010	5130	502	8.4
10-Nov-04	Night	Resource_Planning	Application	LONS0010	4389	512	9.6
15-Nov-04	Day	Resource_Planning	Database	LONS0020	89189	987	14.9
10-Nov-04	Day	Resource_Planning	Desktop	LOND2341	6102	62	2.5
10-Nov-04	Night	Maintenance_Schedules	Database	LONS0190	78910	751	10.5
10-Nov-04	Day	Main_Intranet	Web	LONS0234	7891	127	2.6

Notes:

CPU figures captured as, e.g. “90%”, should be presented as CPU-equivalents, based on power of CPU.

This structure enables analysis by Application, where an application may consist of multiple machines.

May need to make this more sophisticated by capturing both Moving averages AND Peaks of utilisation.

Results from Phase 1

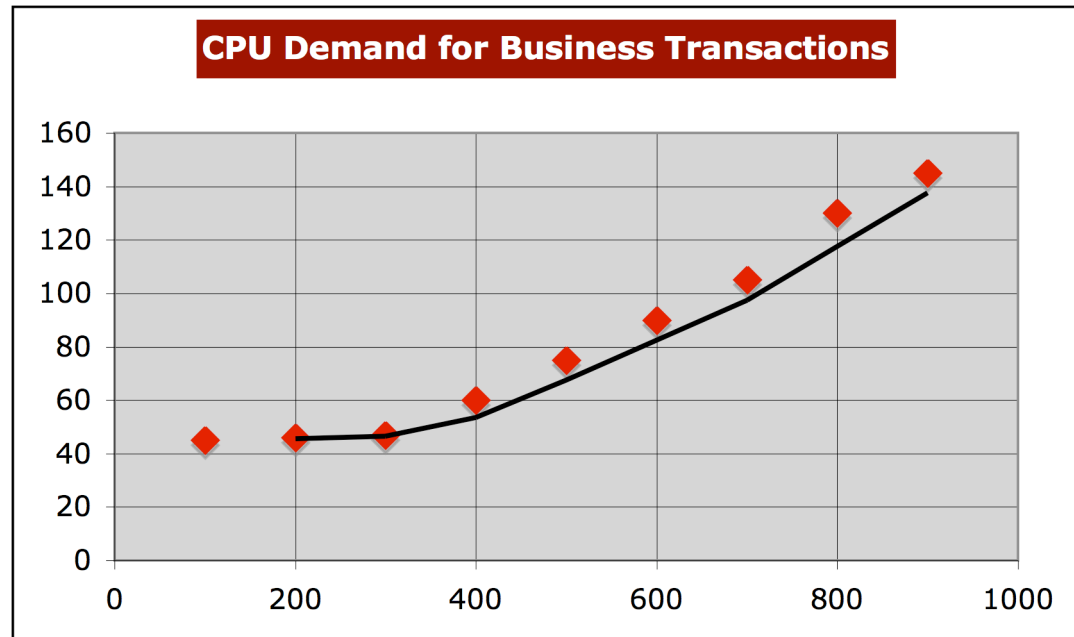
- Identify the amount of resources needed for each main system.
- By comparing this with the existing system configurations, we can identify where systems are over-specified, and which systems appear to be experiencing bottlenecks.
- We can identify systems where there are obvious bottlenecks.
- By capturing this data over a period, we can do basic predictive trend analysis, and identify where systems may need upgrading.

Phase 2: Business Metrics, Resource Forecasting and Modeling

- Need to capture basic Business Transaction metrics, e.g.
 - Number of business transactions (e.g. maintenance records edited)
 - Number of end-users (note that end users may be of different types, so need to differentiate these).
- Correlate these with the Performance Statistics.
- Identify correlations between Business Transactions and Performance Statistics.

Results from Phase 2

- Correlating Business Transactions with Performance Statistics means that we can predict the amount of (e.g.) CPU, Memory and Disk space required for Business Transactions.
- If the number of Business Transactions increases by $x\%$, we can show how much the CPU demand should increase by.
- This is not always
 - linear!



Phase 3: Application Sizing and Demand Forecasting

- Once we have figures from Business Transactions, and a way of correlating them with the Computing Resources, we are able to do more predictive work, based on up-coming projects.
- For example, a new project may be proposed which is expected to be technically similar to the existing Resource_Planning application, and consisting of 25 users.
 - By looking at the figures for the Resource_Planning application, we can forecast the likely CPU, Memory and Disk space required for the new application.
 - This will assist purchasing when the application is prepared for deployment.

Final Objective: Capacity Management Outputs

- Capacity Plan
- Forecasts
- Tuning data
- Service Level Management Guidelines

Ongoing Capacity Management: Processes

- Capacity Management is not a “once-off” process.
- It is essential to create Processes and Procedures which integrate Capacity Management in an ongoing way, e.g.
 - New Project Initiation should include Application Sizing as part of the design process.
 - Project Deployments (and handover to live) should include Demand Forecasting.
 - New Servers should automatically be included in the Capacity Planning model when they are commissioned.

Ongoing Capacity Management: Tools

- Software tools need to be in place to “close the loop”
 - Capture actual activity from the systems into a central repository.
 - HP Open-View, Tivoli, InfoVista & PAWZ are all able to extract this data.
- If there are no tools available, using NT “perfmon” registry stats run from NT schedules
- Data on performance of systems should be collected in a central repository.
 - In the beginning, it would be possible to use Excel, but the quantity of data would require an Access / SQL Server database subsequently.
- The central repository would be used to generate application sizing information, and feeding into Modeling tool(s).
 - Also to provide a data source which other users can access directly, .e.g. Intranet of performance statistics.

Appendix

about

Dennis Adams Associates Limited

Our Key Consultancy Skills

Work with IT Production Managers to

Audit the current environment and help them
Define and Deliver an IT Production Strategy

based on the key components:-

Metrics

Operational Tools

Processes & Procedures

Standards

Metrics and Operational Tools

- **Metrics**

- What methods are in place for collecting and publishing Activity key performance indicators, including man-hours ?
- What method are in place for collecting Technical metrics (such as CPU, disk utilisation etc.) ?

- **Operational Tools**

- What tools are in place for collecting the metrics, above ?
- What tools are in place for delivering the Support function (e.g. Unicenter) ?
- To what extent are the tools integrated (Help Desk fed from Asset Management, into Time Tracking etc.) ?
- Do they have historical analysis (e.g. help-desk should include problem resolution), so that trends can be detected and lessons learnt.?

Processes and Standards

- **Processes and Procedures**

- How do the existing processes and procedures facilitate the day-to-day running of IT Production ?
- How do they facilitate the relationship with the Business Sponsors and IT Development ?
- What processes are in place to support the changes to Production Standards (hardware & O/S upgrades etc)?

- **Standards**

- Are there Technical standards within IT Production against which developers should develop solutions?
- Are there Baseline Configurations & Configuration control ?
- How are these Standards updated? What processes are in place for engaging with other technical teams to discuss emerging technologies?

Creating an IT Production Strategy

- **Assist IT managers to Define and Implement a tailored IT Production Strategy**
- Based on IT Production Audit to identify issues and help IT managers identify a better way of working
 - MOPS matrix and recommendations
- Analysis of MOPS matrix
- Determine the gaps, and prioritise
- Costings and timescales
 - Contacts with “best of breed” suppliers, where necessary.
- Recommendations:
 - Metrics / Tools / Processes / Standards
- Assist with Presentations of Strategy to Stakeholders

Implementing an IT Production Strategy

- **Assist IT Managers in Implementing Strategic Steps**
- Can act as the external advisor in dealing with Tools suppliers.
- It is essential that anyone implementing an IT Production Strategy understands the culture of IT Production
 - IT Production is a risk-averse culture
- Any Changes need to be implemented Incrementally.
- No disruption to the day-to-day running and support of existing Infrastructures
- Essential to have clearly documented “light” processes.
- The primary focus of the technical teams must remain on the technical support of Infrastructures.

IT Production Experience



Capacity Management for IT Production

...

Theory and Practice

Dennis Adams Associates Ltd.

<http://www.dennisadams.net>

November 2004