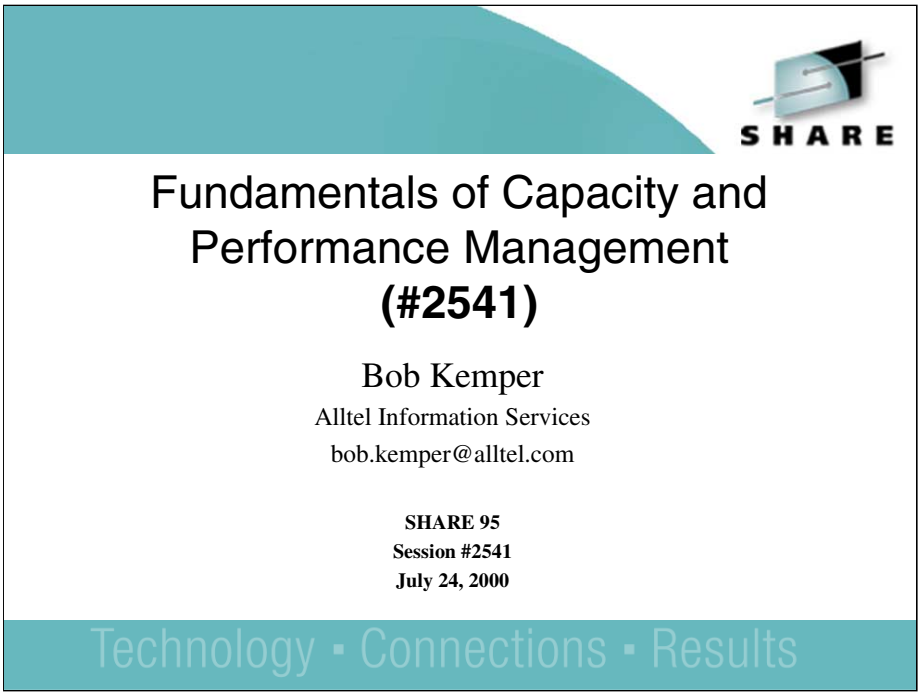
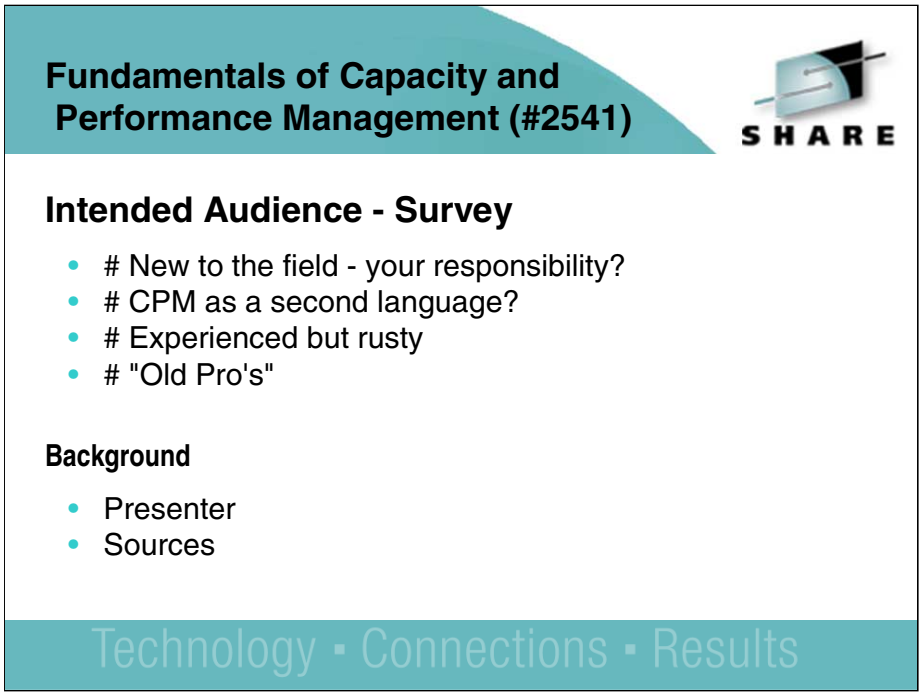


About the presenter:



Survey the audience

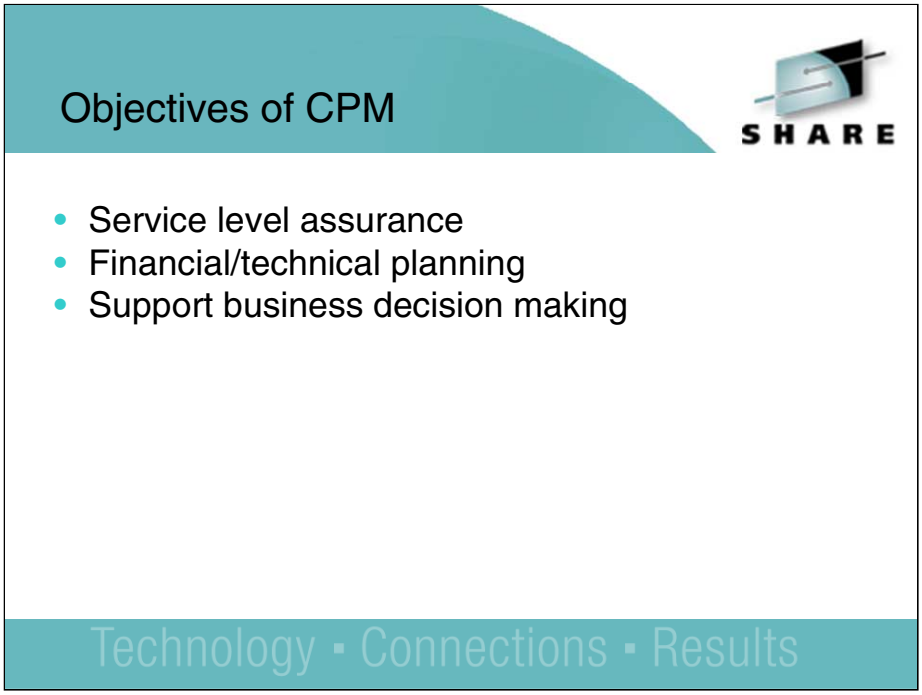
- New to the field - Your responsibility (You're in the right place)
- CPM as a second language (You're in the right place)
- Experienced, but "rusty" (OK for refresher, organization of thoughts)
- "Old Pro's" (You're welcome for my benefit, but you might get more return on investment elsewhere)

Background

- Presenter
  - 15 yrs in CPM
  - 15 yrs MVS/mainframe
  - 1 yr appl dev mgr - Unix/Oracle
- Sources
  - Published works of: J. Buzen, H. P. Artis, C. Watson, S. Samson, many others
  - Training/consultation with experts (both theorists & practitioners)
  - Personal experiences & observations

Outline

- Overview
  - 3 Objectives
  - Definition & Scope
  - The Org Chart
  - Staffing Tips
- The (5-Step) Process
  - Basic Systems theory
  - The 5 Sub-processes
  - Process Maturity
  - The Tools of CPM
- Challenges



Objectives of CPM

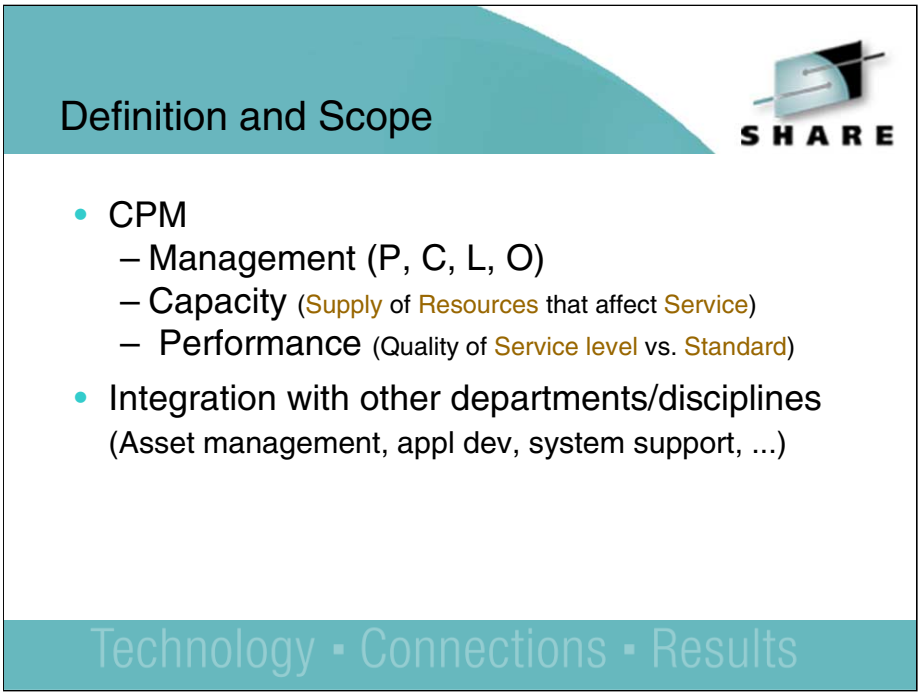
The scope of interest for Capacity and Performance Management people can vary a great deal, but in general, the CPM department's charter consists of three main objectives. These are…

1. Service level assurance
2. Financial/technical planning
3. Support of business decision making

Service level assurance means ensuring that interactive response time and batch turnaround are consistent and acceptable. This means, of course, that everyone must know and agree upon what is acceptable for each service.

It's not hard to maintain acceptable service with unlimited budget. A charter compliment of service level assurance is **Financial/technical planning**. This involves providing acceptable service at the lowest possible cost.

But, our technology environment is not static. Things change. Business volumes change. Markets change. Technology continually advances. Prices change. Consequently, to sustain optimum cost/performance equilibrium over the long haul, it is important to develop and evolve a technology strategy. As the third leg of the CPM Charter, this is stated as **Support of Business Decision Making**. There is another aspect also. Effective technology planning must be done in the context of the specific needs of the business it serves. Support of business decision making, then, involves supporting making business decisions as well as making technology decisions. (e.g. can I offer to my customers 24x7 service availability?, offer premium service level options?, offer "dynamic capacity growth"?, …)



Capacity and Performance Management - Definition and Scope

"CPM" consists of three key words:

Management in the HR sense (text book) involves 4 functions: Planning, Leading, Organizing and Controlling. This also describes fairly well the meaning of Management as it relates to the CPM discipline.

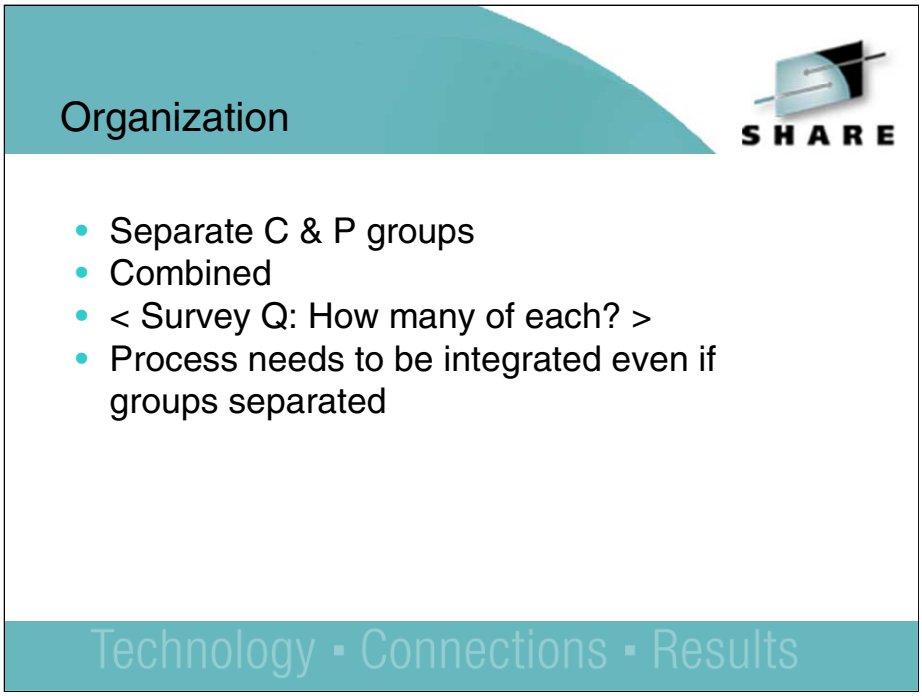
Capacity, as it relates to CPM, can be defined as the supply of a resource that effects service.

Performance is the "quality" of service provided as compared with a predefined standard or goal.

All together then, CPM is… the act of planning for and controlling the supply of resources that effect service, such that service levels meet or exceed a predefined standard or goal through leadership and organization.

Related disciplines

There are a number of disciplines which are closely allied with CPM. So much so, that the boundaries vary. Related disciplines include Asset Management, Application Development, System Support, Problem/Change Management, Strategic Planning, Product Planning.



Organization of the CPM Function

As we will see in a few minutes, CPM is a single functional discipline. However, that doesn't mean there's only one way to organize the people who perform the process.

Some organizations have separate teams which focus on capacity vs. performance. Frequently, the team of performance specialists is called the "Performance Group" or "Response Time Improvement (RTI)" team. They focus mainly on tactical, day-to-day tuning of system and applications.

The other team then, focusing on long term planning, trending, etc. are referred to as "Capacity Planning" or "Strategic Planning".

Survey Q: How many have separate groups? Combined? None?

Regardless of how people/groups are organized, the process must remain integrated.

Remember: Capacity and Performance inherently affect each other.

If you tune a workload which was using a lot of capacity to use less, it will change your capacity requirements (i.e. delay the next upgrade).

Staffing

Many organizations use the Myers/Briggs or similar tests to match people with specific jobs. It correlates your answers to “preference” and personality questions with those of “happy” or “satisfied” members of specified jobs/disciplines.

Now popular in HR circles is a focus on “Competencies”. Related tools allow organizations to define the mix of competencies required for each job and match them with the profiles of those in the available labor pool.

CPM Competencies / Personality Traits

The function of CPM is highly numerical/analytical. Consequently, those averse to statistics or math are likely to be unhappy and unsuccessful.

CPM’s role is mainly consultative. Therefore, effective communication skills are crucial.

The CPM scope cuts across many technical disciplines. Ability to speak the language of practitioners in various areas is a plus (e.g. Systems, Application development, Network, Telephony)

Since much capacity is normally consumed by business applications, CPM must have one or more people who are intimately familiar with the company’s or division’s specific applications (and preferably have connections in the Application Development organization(s)). <June CIO Mag. article - Bus-specific knowledge will be the most crucial competency for success as a technologist in the coming economy. >

CPM is a nexus of business and technology issues and decisions. A business background and a working knowledge of the principles of accounting are a big plus, as are any knowledge of commercial purchasing, leasing or business law.

The CPM charter calls for reaching conclusions and making recommendations based upon incomplete information. Consequently, no perfectionists.

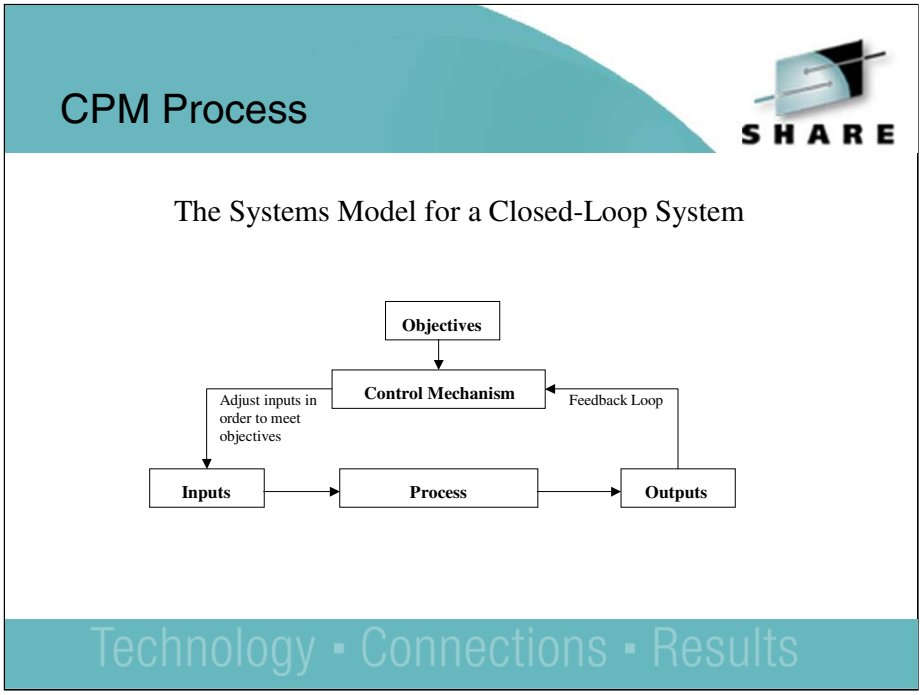
Having a map vs. knowing the territory

Like any profession, experience is valuable. There’s an old cowboy saying that “it’s nice to have a map, but it’s better to know the territory.”

Checkpoint - Part 1 Summary

- Objectives
  - Service Level Assurance
  - Financial/Technical Planning
  - Support business decision making
- Definition / Scope
  - Planning for & Controlling the Supply of Resources that affect Service,
    such that Service levels meet or exceed predefined standard/goals
    through Leadership and Organization
- Organization
  - Process must be integrated, even if functions are separated

Technology ▪ Connections ▪ Results



The CPM Process

Terminology from basic "Systems theory"

System: "An integration of elements, all working together toward an objective" from Introduction to Computer Based Information Systems, Ch1

2 Types:

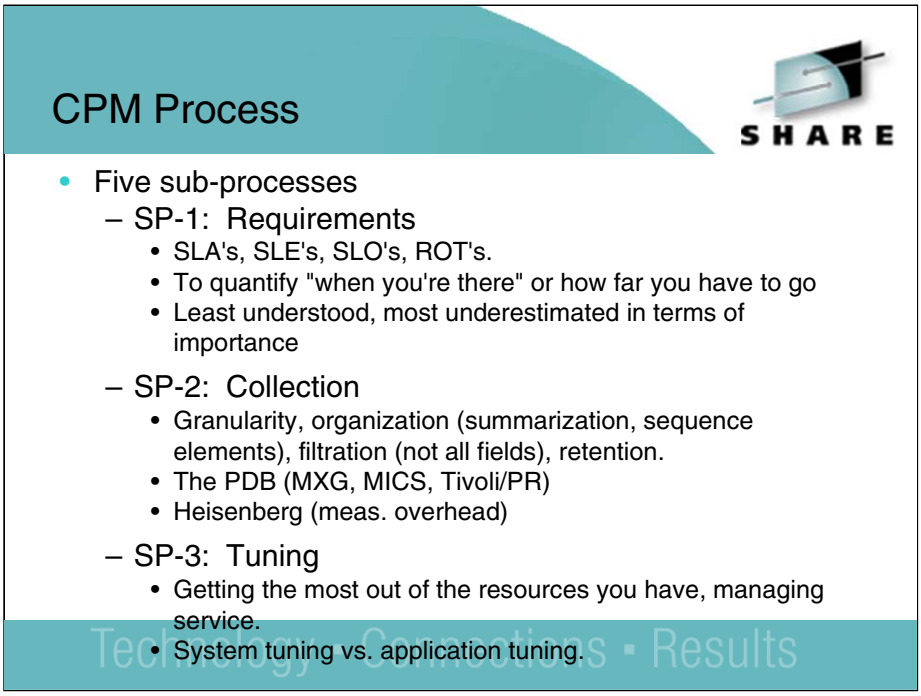
Open (or Open Loop) system

- Start it, and it goes.
- Ex: Fireplace

Closed Loop system

- Difference is "Feedback Loop"
- Adjusts inputs to meet objectives
- Ex: HVAC with Thermostat

CPM is the 2nd type - Closed Loop



The CPM Process

The singularity in the wording “The CPM Process” is deliberate. In this model, management of resource capacity in the context of defined minimum service levels is accomplished using a single five-step integrated process. The steps are referred to as “sub-processes” and are numbered 1-5.

SP-1 - Requirements

The first sub-process is the identification and quantification of requirements. Included are…

• Service level requirements such as system availability, online response time and batch window.

• Business plan growth requirements including new accounts, new applications, customer or revenue growth projections

• Reporting requirements including customer, inter-organization, management and executive reporting, feeds to budget or financial systems, audit or tracking reports.

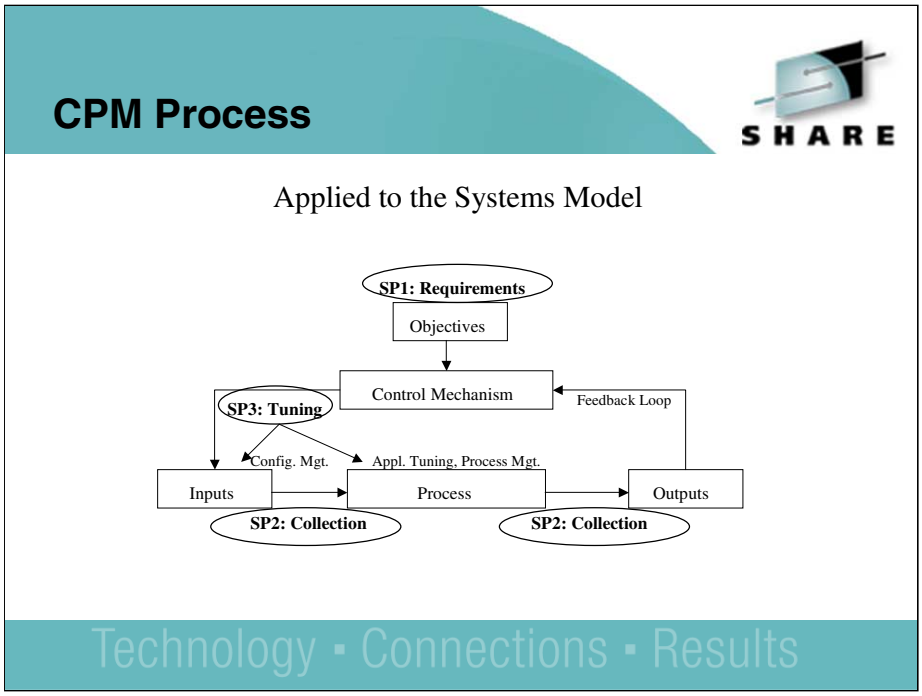
This sub-process is generally executed once at the initiation of service or contract, then revisited when changes to the contract or service definition occur, or when substantial changes are made to the business plan. Often, requirements are quantified in formal service level agreements (SLA’s) with customers, but even without SLA’s, customers’ expectations should drive the service level objectives (SLO’s) for each distinct service.

SP-2 - Collection

This sub-process involves the collection and organization of data about resource consumption and business and technology events which supports delivery of the requirements from SP-1. Examples include: CPU consumption, Service delivery data, business or technology events and historical trend data. This data is often consolidated and concentrated into a Performance Database or “PDB.”

SP-3 - Tuning

The essence of this sub-process is extracting maximum value from the resources devoted to IT service delivery. It involves the evaluation of data collected in SP-2 against the requirements identified in SP-1 and initiating actions to close any existing gaps or take advantage of improvement opportunities while minimizing expense and risk. Examples



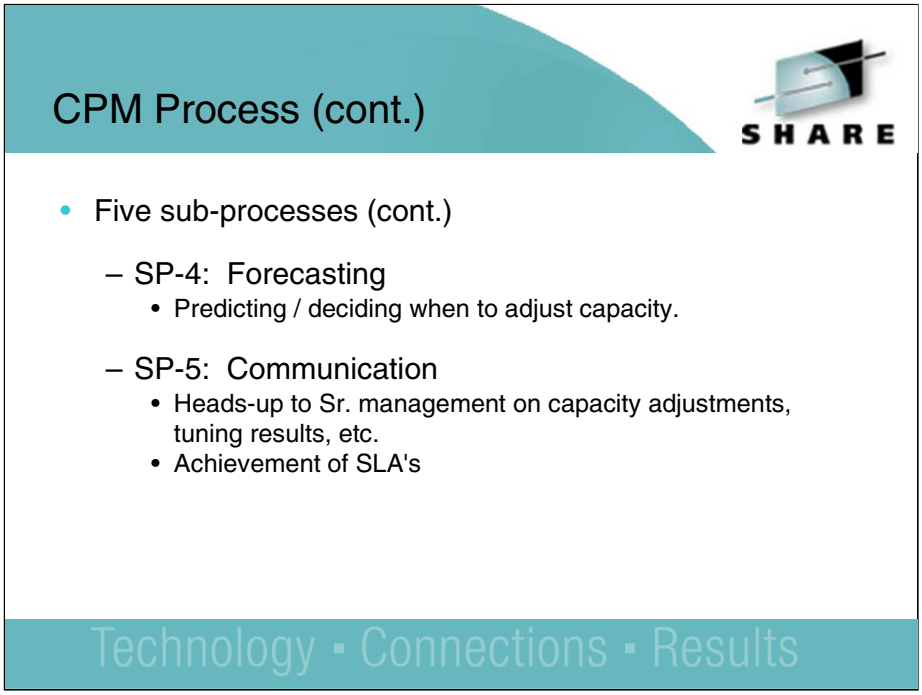
SP1-SP3: Applied to Systems Model

SP1: Relates to defining objectives

SP2: Relates to measuring inputs and outputs

SP3: Relates to making adjustments (e.g. system resources, process design)

Together, these steps constitute the “Performance Management” aspect of CPM



The CPM Process (cont.)

SP-4 - Forecasting

SP's 1-3 define the "performance management" component of the CPM process. SP-4 involves combining knowledge of business plans and requirements with resource consumption data to provide the information necessary for asset and budget planning. This defines the "capacity management" component of the unified CPM process. Examples include: Modeling, Trending, Forecasting, Estimating and Sizing.

Examples of SP-4 output include…

Next capacity upgrade is recommended mm/yyyy. Estimated impact of upgrade delay. Analysis of alternatives.

Total billing units estimate for next year, for rate management

Value of new technology to service levels or expenses

Impact of projected business growth on response time or batch window

Impact of configuration management decisions on response time or batch window

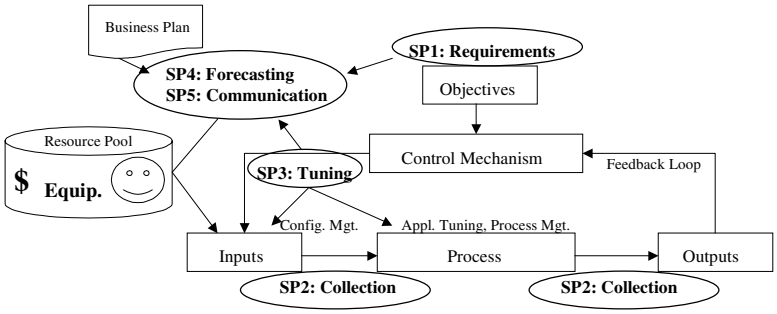
SP-5 - Communication

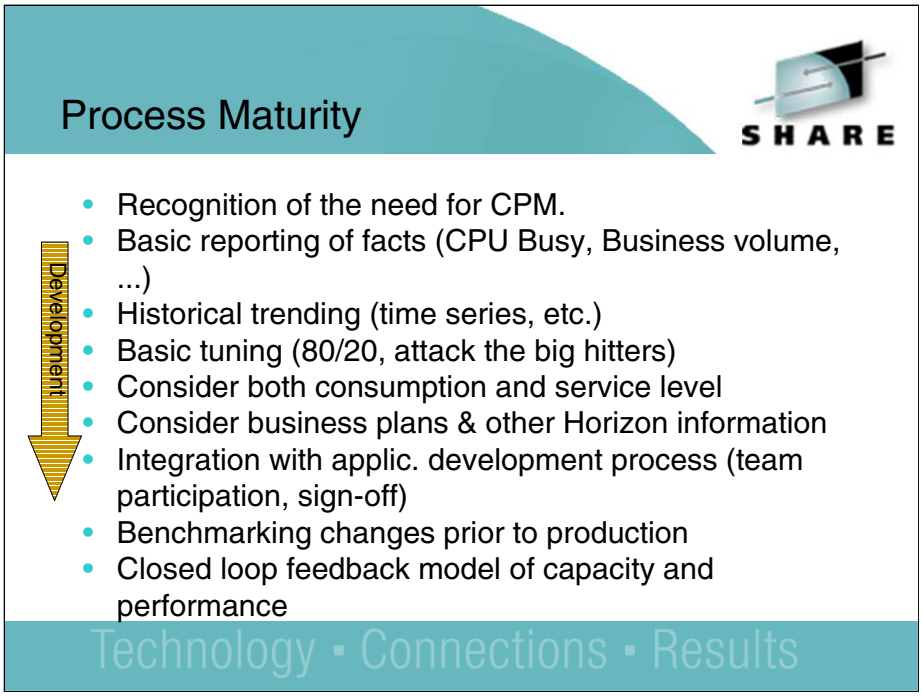
SP-5 involves providing information about IT resources to all stakeholders which is timely, accurate and tailored to their requirements. It also involves obtaining feedback from customers regarding changes to business plans or requirements.

Examples of SP 5 output include: Resource consumption (for CPU, disk, tape), Service level attainment reports, Period exception reports, Changes to hardware/environment, Monthly capacity plan updates, Recommendations, Change impact assessment, Ad-hoc reports.

CPM Process

Applied to the Systems Model





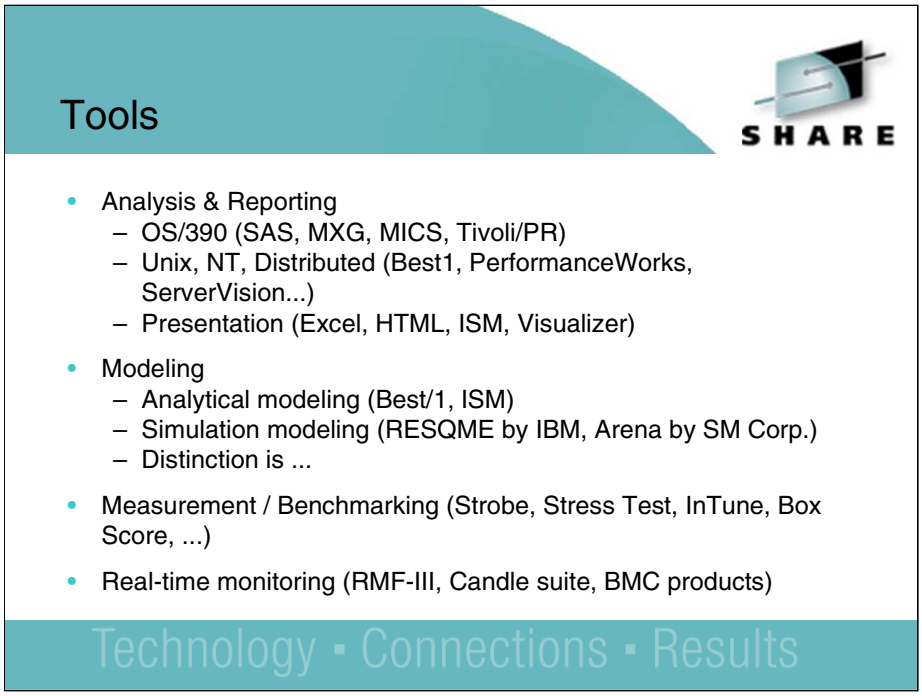
Process Maturity

A Maturity Model measures an organization's progress toward complete and consistent deployment and execution of a specific process. (The concept was formalized by Watts Humphrey in his book "Managing the Software Process," where he proposed the "Capability Maturity Model" or CMM.)

Rather than a formal "maturity model" these phrases are intended to suggest attributes along a continuum from "introduction" to "best practices" for the 5-step CPM process model.

A common misconception is that all organizations should achieve the highest level of maturity. Depending upon the importance of IT to a organization's success and the total investment in IT, a lower level of maturity may be most cost effective.

Example: Benchmarking, Feedback model are unnecessary if you only run back-office work on your machines, workload is highly stable and investment in IT resources is low.



For analysis and reporting

- OS/390: MXG and MICS are SAS-based. Tivoli/PR is DB2/SQL-based
- Unix, NT, Distributed: Best1 from BMC, PerformanceWorks from Landmark,
  ServerVision from Beta Systems.

For Modeling

- Analytical: Best1 from BMC, ISM from ISM, Inc.
- Simulation: RESQME from IBM, Arena from Simulation Modeling Corp.
- Distinction:
  Analytical modeling is a point-in-time snapshot. Assumes steady state.
  Simulation modeling is time-sensitive. Can model impact of “divergent” periods.

Meas. / Benchmarking

- Strobe from Compuware, Stress Test from Demand Technology Software, InTune from BMC, Box Score from Watson & Walker.

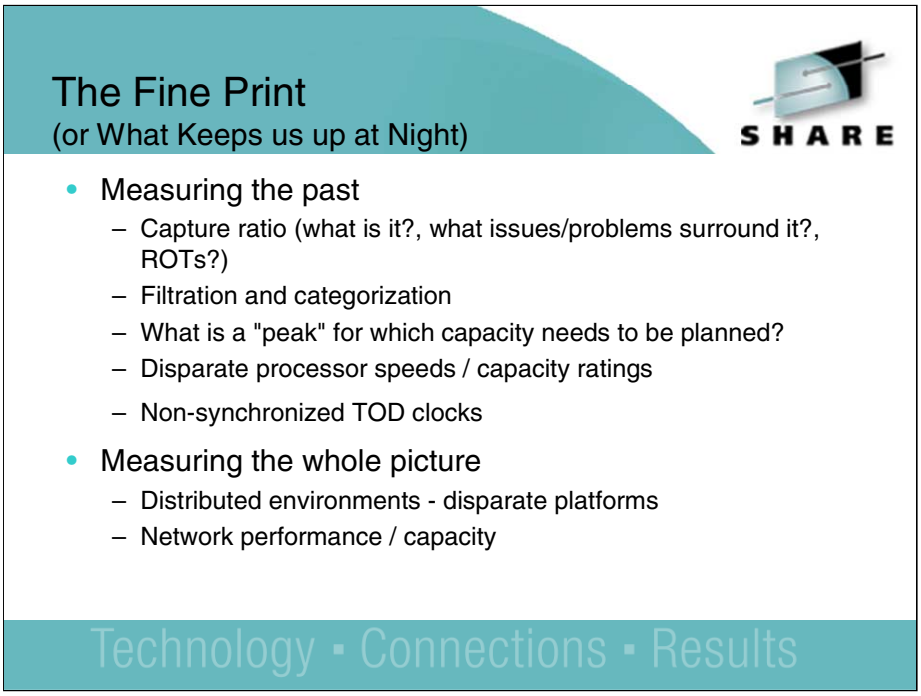
Real-time Monitoring

- RMF-III from IBM, Omega*, Control*

Checkpoint - Part 2 Summary

- 5 SP's
  - SP1: Requirements
  - SP2: Collection
  - SP3: Tuning
  - SP4: Forecasting
  - SP5: Communication
- Process Maturity
  - Continuum from Recognition to Best Practices
  - Appropriate levels, practices differ
- Tools

Technology ▪ Connections ▪ Results



The Fine Print (or What Keeps us up at Night)

Measuring the past

Capture Ratio is the proportion of all consumed resource which is attributed to specific resource consumers. For example, CPU capture ratio is the sum of the individual workload CPU consumptions divided by the total measured CPU busy time for an interval. "Normal" CR can vary widely based upon operating system, workload mix and measurement tools. It should be tracked as a trend indicator rather than against a specific ROT number. A declining CR may indicate increasing non-productive or "overhead" time, and should be investigated.

Filtration and Categorization

CPM is chartered with tracking and forecasting resources for a wide variety of purposes and people. It is necessary to be able to categorize resources and performance specifically tailored to each. For example, the System Support people may need numbers by "region" or "instance", while developers need information by "user group." Service Management may need similar information by "client" or "product line" and the Executive Team needs resource allocation by Business Unit. Performance and capacity data must be filtered, sorted and categorized strategically such that all of these needs can be met with a minimum of CPU and Storage.

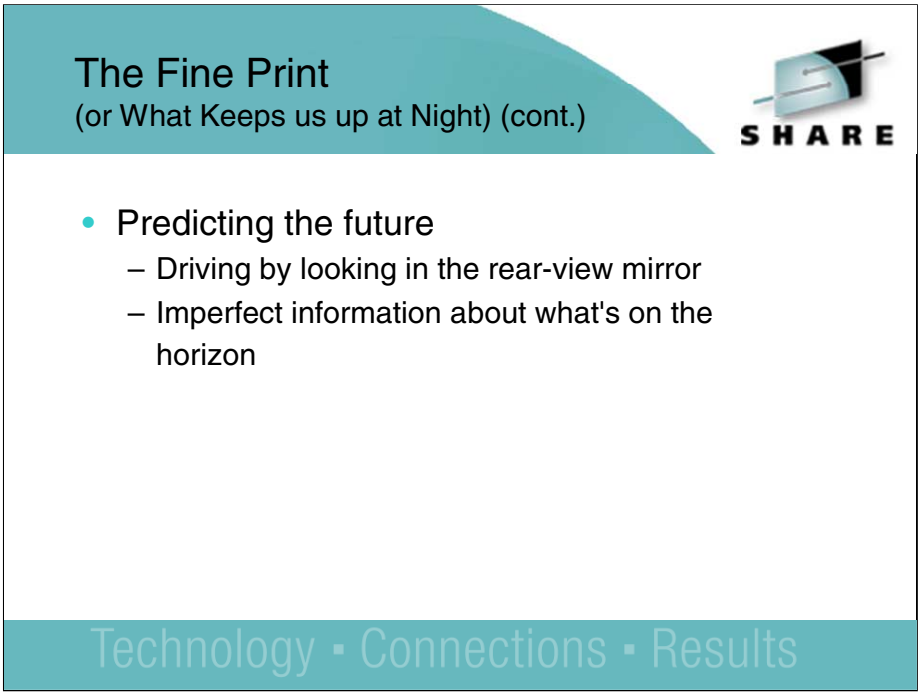
Definition of "Peak"

This concept is best described using the metaphor of a sports stadium. Most days it is not full. However, it is crucial that there be enough seats to "meet service level" (each ticket holder have a seat) on a game day. Service level agreements must be specific about not only response time or turnaround, but also over what interval the performance data are to be aggregated. I'll need to have a lot more capacity if I can't afford to miss service level even during a 10 minute spike than if I'm held to only an hourly average.

Disparate Processor Speeds

For purposes of measuring work completed, using "CPU Seconds" can be dangerous. This is because if you double a processor's speed, the same unit of work will then use half the CPU seconds. This has implications for both billing and capacity planning. Cheryl Watson has written eloquently on this topic in her TUNING Letter.

It has always been true, but particularly with newer CMOS class processors, relative changes in processor throughput are workload dependent. Cheryl Watson: "differences can exceed 30%".



Predicting the future

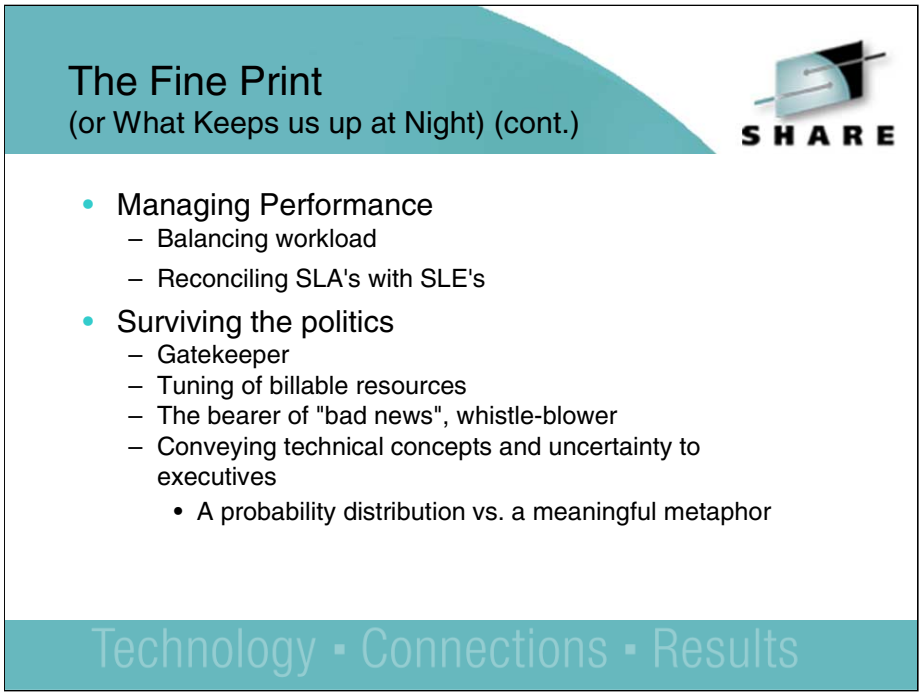
CPM Sub-process 4, Forecasting, involves predicting not only what will happen in the future, but also what the impact will be and how many changes will interact with each other.

Trending

A simple and traditional approach to forecasting is to use trending. That is, measure the historical growth of workload(s) and draw a (straight or curved) line into the future from it. The biggest danger to over-dependence on this approach can be understood by likening it to driving your car by looking in the rear view mirror. If the road ahead is similar to that behind, it works for a while. However, at the first unexpected curve or object in the road, somebody's likely to get hurt.

Event-sensitive forecasting

The solution, of course, is to look through the windshield. However, few of us are blessed with a transparent window or well-aimed headlights. The "trick" lies in identifying the factors which most directly impact capacity for each resource, and then finding reliable sources for forecasting each factor. Examples of common factors include Account Volume, Customer volume, Business Transaction rate. Indirect influences frequently include economic factors such as interest rates or news or other media events. Sources for future values of these latter factors are often, at best, speculative. However, business plans often exist which provide assumptions about future business volumes. Examples include sales forecasts, product release schedules and advertising campaign schedules. The benefit of using these sources is that if the business volumes which would drive capacity do not materialize, or hit early, then both revenue and expenses are affected in synchrony. The challenge is then to pay for additional resources only as needed, but have them soon enough to protect service levels.



The Fine Print (or what keeps us up at night) (cont.)

Managing Performance

When workload is spread across platforms, there is a certain amount of “fragmentation.” Recent developments are helping (e.g. OS/390 Parallel Sysplex, CICSPLEX, WLM Goal mode, Scheduling Env’s.). However, workload balancing remains a significant challenge.

Reconciling SLA’s with SLE’s

Even when SLA’s exist, actual users who are accustomed to service far exceeding the agreements can become very disappointed with service near the SLA threshold.

Surviving the Politics

Gatekeeper - Loved Ones.

Tuning of billable resources - Saving CPU via tuning when CPU is a billable resource.

Bearer of bad news - Need to spend $’s for upgrade.

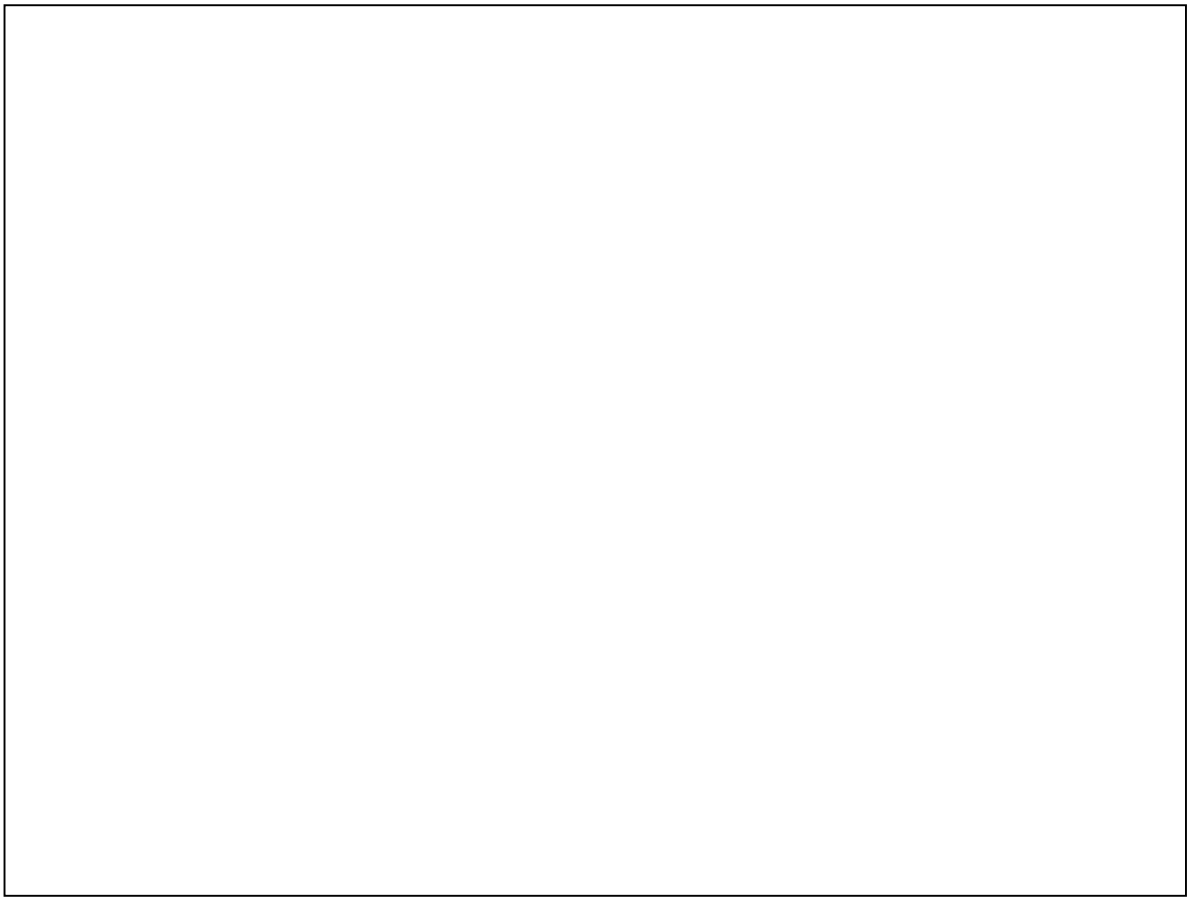
Whistle-blower - Identifying resource/cost intensive applications.

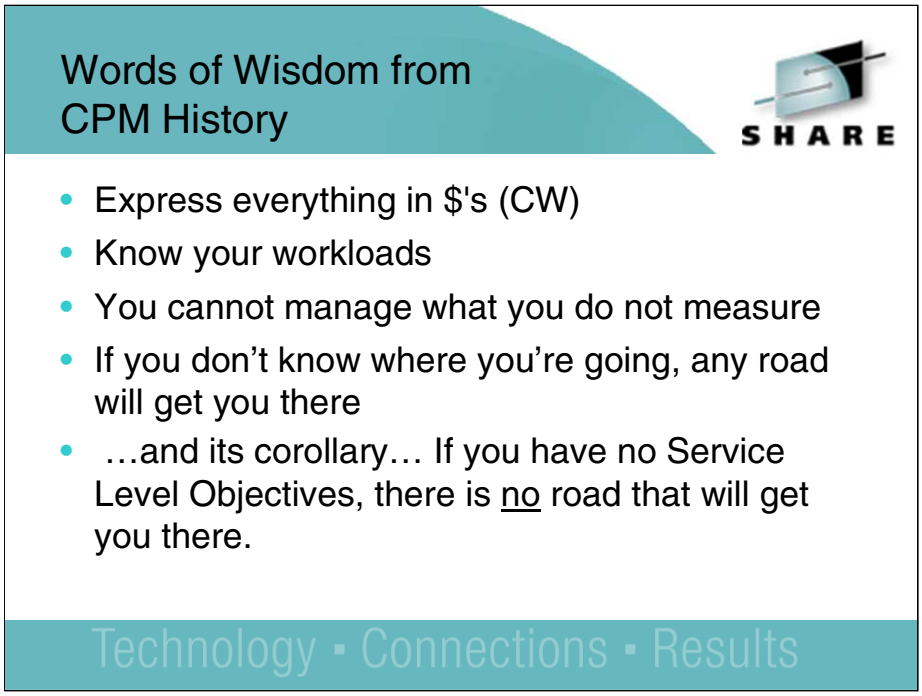
Conveying technology & uncertainty to Executives -

- “Guarantee no problems.” Lawn service & no dandelions
- Apollo-13. 20 Amps? “About enough to run that coffee pot for 9 hours.”

Summary

- Ensure acceptable service at the lowest possible cost and enable improved decision making for executive management.





Gems of Wisdom from CPM folklore

- **Express Everything in $'s**: (Cheryl Watson) Availability, Response time, Completion times,
  always convert to $'s. Means you need to build & maintain a conversion table.

- **Know your workloads**: (CPM folklore) Unique application characteristics, critical dependencies, reasons things were coded as they were (& are they still valid?). Bus Unit
  ownership, relative priority, who owns each…

- **You cannot manage what you do not measure**: (TQM folklore) Popularized during the
  TQM craze of the 1980's. If you don't look and keep track, then you have no context or direction. A testimonial for SLA's.

- **If you don't know where you're going…** : (Cultural folklore; Roy Rogers?)
- **Corollary…** : Effective management requires Gap Analysis (another TQM term) ability,
  which compares service delivered (measured) against service committed (SLA). If there is
  no goal, you don't know when you're "there."

Other EWCP Sessions this Week

- Session grid available at the door
- Other Introductory/Tutorial Sessions

Technology ▪ Connections ▪ Results

Other sources of Information for CPM Professionals

- Publications
  - Cheryl Watson TUNING Letter
  - IBM Redbooks series (Available on the Web)
  - Steve Samson's book on OS/390 CPM
- Web sites
  - ICCM
  - Cheryl Watson's (plus lots of links to others)
  - CMG

  … attachment at back of handout with addl. URL's
- IBMMAIN Newsgroup

Technology ▪ Connections ▪ Results

Questions

- (please state Name, Installation)

Technology ▪ Connections ▪ Results

Thank You

Bob Kemper
Alltel Information Services (CPI)
Jacksonville, FL
bob.kemper@alltel.com
904-854-3157

Technology ▪ Connections ▪ Results