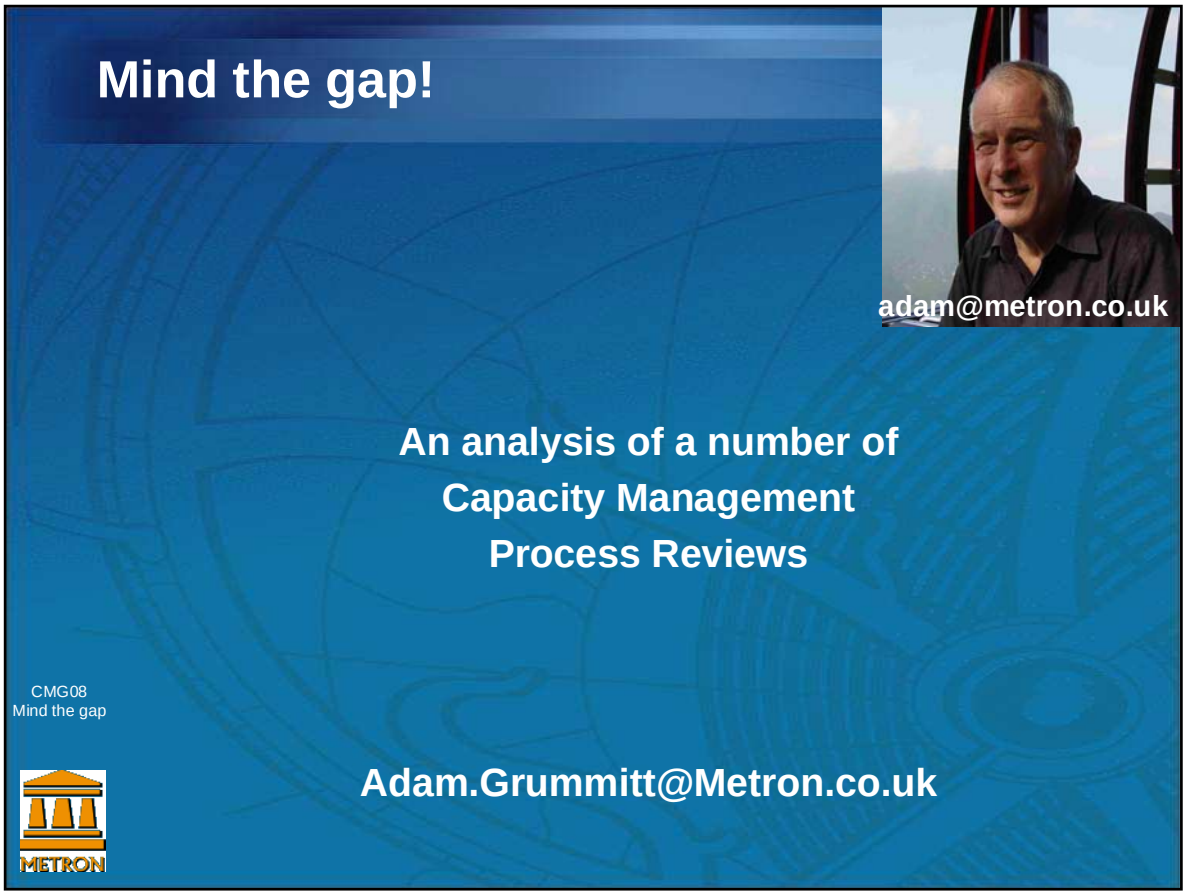


There is an increased awareness of the need for governance of ITSM processes. This is promulgated by the management consultancies offering extended audits to enterprises to check that their processes are reasonable in comparison with their peers and with good practice within their industry. Such audits are usually more of a major attitude survey based on extensive interviews with an arbitrary scoring method than a detailed technical audit. People are given a mini-lecture on CMMI and then asked to rate themselves and their peers with an arbitrary score of 1 to 5. These scores are faithfully recorded and aggregated and analysed and wondrous graphs produced.

Such surveys add to their own bad name by producing Kiviat diagrams showing scores of different aspects to unrealistic accuracy and with a false assumption that the academic attainment of all that is described in ITIL is necessarily a Good Thing in all cases.

Each activity varies for each service during its own life cycle, as well as in the light of changing company circumstances.

Each activity may or may not be appropriate for each service, server or whatever, depending on its importance to the business, its cost and its reliability.

Abstract

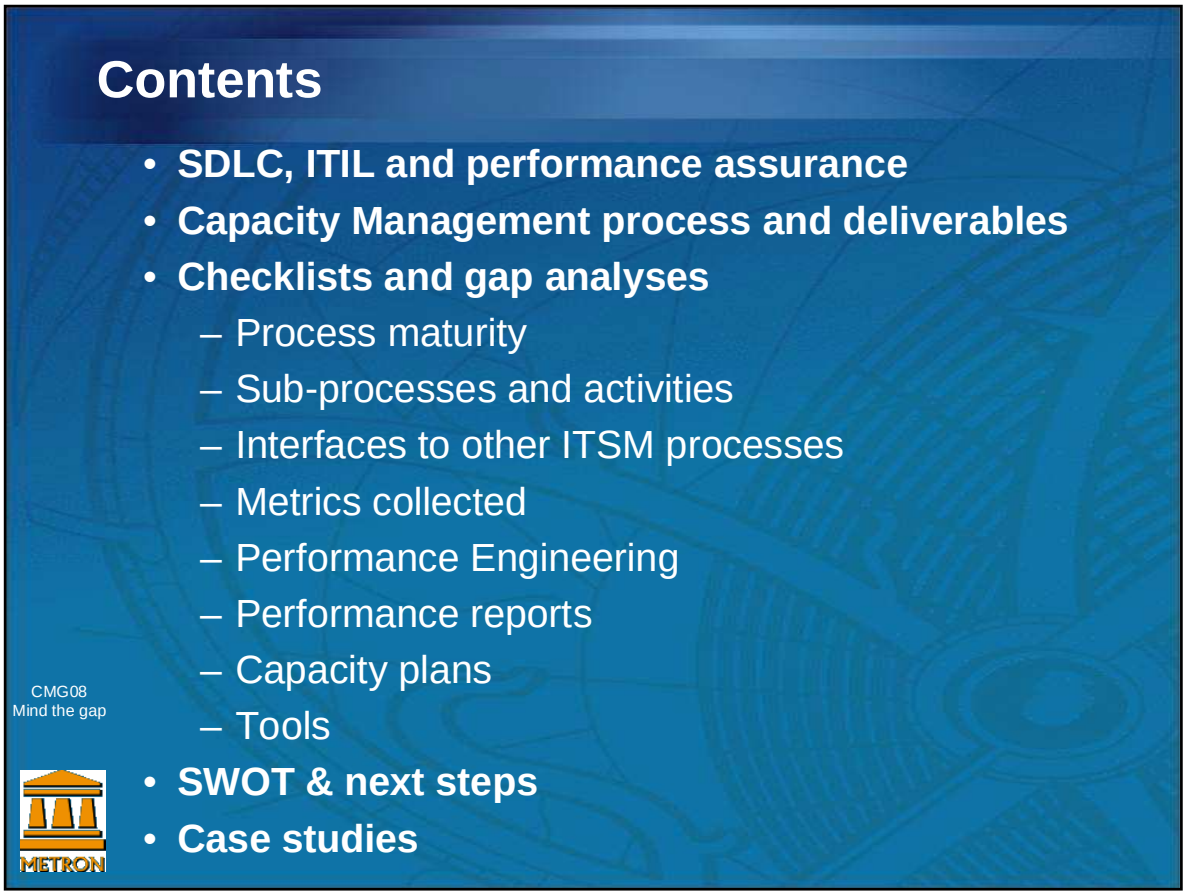
- **Although management consultancies offer audit services based on ranking sites' IT processes against their peers, many companies prefer to identify any gaps in their adoption of good practice for capacity management. This is more objective, reveals "unknown unknowns" and relies less on attitude surveys.**
- **This paper reviews a number of such gap analyses. Irrespective of the enterprise culture (performance "just adequate", or "mission critical") the same approach is applied. Good capacity management measures workloads and performance to ensure that business needs are met now and in the future.**

CMG08
Mind the gap

What most sites already well immersed in ITSM find is a more valuable service is a consultancy review of their actual processes, why they have made whatever decisions they have, and an indication of where there are risks, issues or dangers. So rather than say a process is mature with a snapshot score of 3.571, it is more useful to say that although performance reports are published to the web, there are few hits on that site from outside the performance team members who publish it. A manager is more inclined to say although the process is well in place, there is a need to discuss some of the known unknowns and also to reveal any unknown unknowns.

This paper is based on a number of such gap analyses. Although arbitrary scoring is eschewed, there is still a need to start with an outline template of what is generally considered to be Good Practice and that is usually based on ITIL as a starter with pragmatic experience adding a lot more practical detail. But it is vital that the company culture is recognised and that the objectives of the service are appreciated before stating that any of the usual practices are vital for that site.

It seems remarkable at first that every site is so different in its detailed implementation of an IT development and infrastructure environment, but then each one is a living organism and has huge opportunities for variation.



This paper begins with a brief review of the SDLC and ITIL and their relationship. Each side tends to view the universe from their own point of view without due recognition of the other and the need for coordination. It then gives a summary view of the capacity management process and deliverables, largely as outlined in ITIL. It then drills down into more detail and considers a range of spreadsheets, each reviewing one aspect of the process Eight are shown. There are sixteen in the full list used. Each site tends to make its own selection as to which areas are of most import.

The examples used in this paper are essentially from a combination of similar sites – it is not easy to follow the logic if the examples are scattered across the spectrum of responses. So the ones used have been chosen as they are typical of the hard pressed retail sector where the economic downturn is having an impact on everything. Thus, rather than capacity management defining what equipment is needed, it is more likely to be told it will have 10% less money to provide the same services to double the number of users (due to mergers and acquisitions) with the same hardware and 10% less staff. Be frugal, do things just in time, but make sure that the mission critical services continue to be supported with high availability and good performance.

Most of these analyses will clearly finish with a SWOT and outline of next steps. This paper will give a summary of some case studies and finish with a set of typical reports.

A number of acronyms are used. The base ones are:
IT = Information Technology (though ICT is the preferred ITIL acronym)
ICT = Information and Communications Technology
ITIL = Information (and communication) Technology Infrastructure Library
ITSM = IT Service Management
itSMFI = ITSM Forum International
itSMFUK = ITSM Forum United Kingdom
SWOT = Strengths, Weaknesses, Opportunities, Threats

#4

SDLC – Development Life Cycle

Systems	Software	Deliverables
Feasibility	Requirements	TOR, Scope, PID, SOR
Analysis	Architecture	Functional Spec
Design	Design	SPE, System spec
Implementation	Implementation	Program spec, modules
Testing	Testing	system, α, β, load, trials
Maintenance	Deployment	pilot, production, retire

Steps: Review, Feedback, Prototype, Cyclic
Routes: Waterfall, Iterative, Scrum.
Priorities: Project vs. Process, Panic vs. Procedure

CMG08
Mind the gap

METRON

The SDLC is mentioned in ITIL V3, as is almost every aspect of alternative approaches to IT and Service Management.

The S in SDLC is described as system, software or even service by different authorities. The description of the precise steps in any project varies in detail, as do the many approaches to development outlined over the years. Early "waterfall" development was soon improved by increased prototyping and more iterative approaches, with focus enhanced by use of scrum and sprint teamwork.

SDLC introduces its own set of acronyms as on this slide.

TOR	Terms Of Reference
PID	Project Initiation Document (or sometimes PIC for Charter)
SPE	Software Performance Engineering
α and β test	Internal systems testing and external pilot site testing

For sites who work this way, any project or more than so many days has to be defined in a project management – project initiation style. The deliverables used to establish a project are often called a PID (Project Initiation Document) or a PIC (Project Initiation Charter) or some other document containing terms of reference.

However, there is a need to map this project view into a matrix of management since equally important to the service is the infrastructure view of applications which is just as interested in growth of an existing application as it is in the arrival of new one.

#5

ITIL® is

- **IT Infrastructure Library - books & descriptions**
  - *V2 exams on Service Support & Service Delivery*
  - Business Perspective, Infrastructure, Development, Security & Service Management
- **V3 Service Life Cycle: Strategy, Design, Transition, Operation and Improvement (and introduction)**
- **Good practice for ITSM, basis of ISO/IEC 20000**
- **V1 replaces GITIMM by CCTA in 80's, V2 OGC in 90's, V3 now**
- **Metron key contributor to initial Demonstrator**
- **itSMF, itSMFI and itSMFUK**
  - The IT Service Management Forum for users
  - Promotes exchange of info & experience
  - UK, NL, B, AUS, ZA, F, USA (over 40 chapters).
- **OGC owns it; TSO prints; itSMFI keeps; itSMFUK actions; APMG QAs; APMG, ISEB, EXIN, Loyalist & Dansk IT examine; HP, Pink, Purple, FGI, Fox, Metron & many others teach it**
- **Maybe some fragmentation with ITSM Library & ISO exams**

CMG08
Mind the gap

METRON

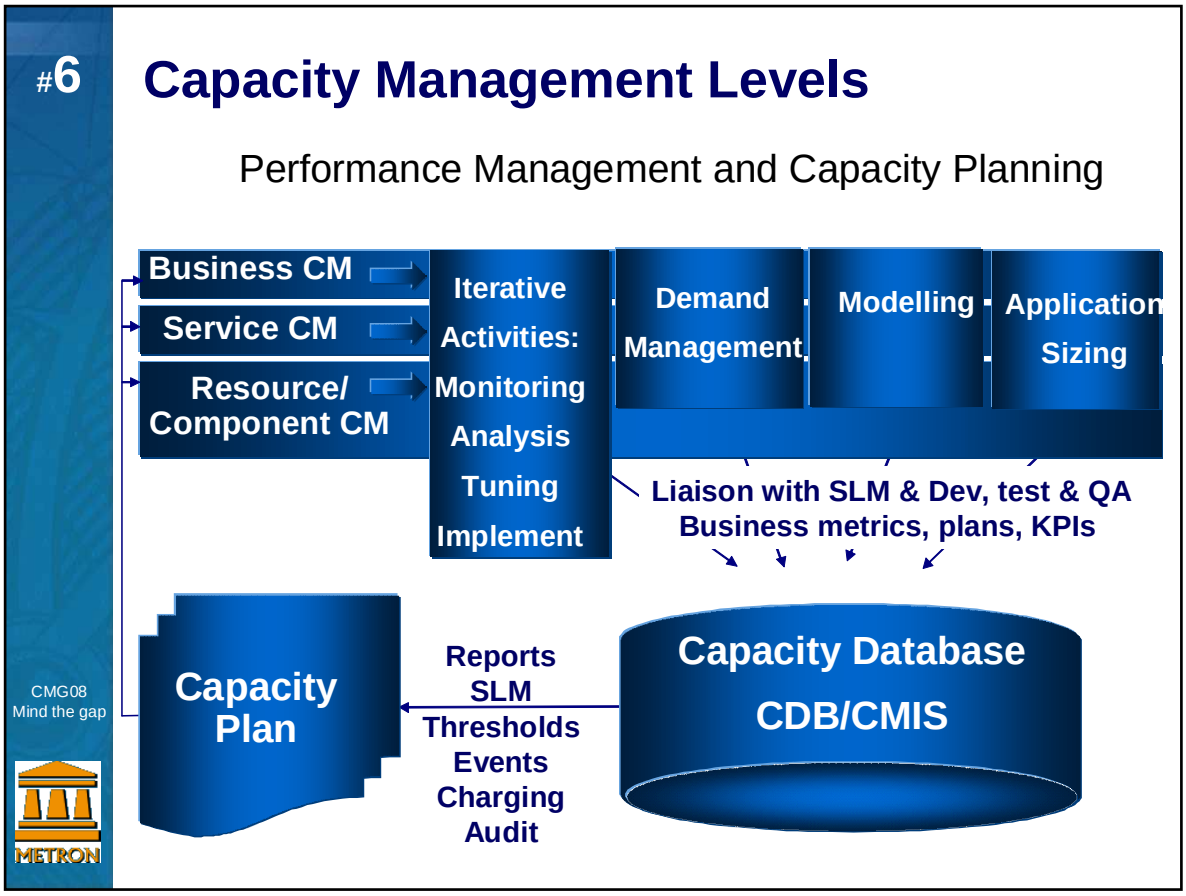
This slide refers to a lot of organizations:

CCTA	Was the Central Communications and Technology Agency (UK), now part of
OGC	The Office of Government Commerce (UK)
TSO	The Stationery Office formerly HMSO (Her Majesty's Stationery Office)
APMG	The APM Group
ISEB	Information Systems Examination Board allied to the NCC and BCS
NCC	National Computing Centre (UK)
BCS	British Computing Society
EXIN	The EXamination INstitution for information science
Loyalist	Loyalist College of Applied Arts and Technology
Dansk ITexamine	The Danish IT examination authority

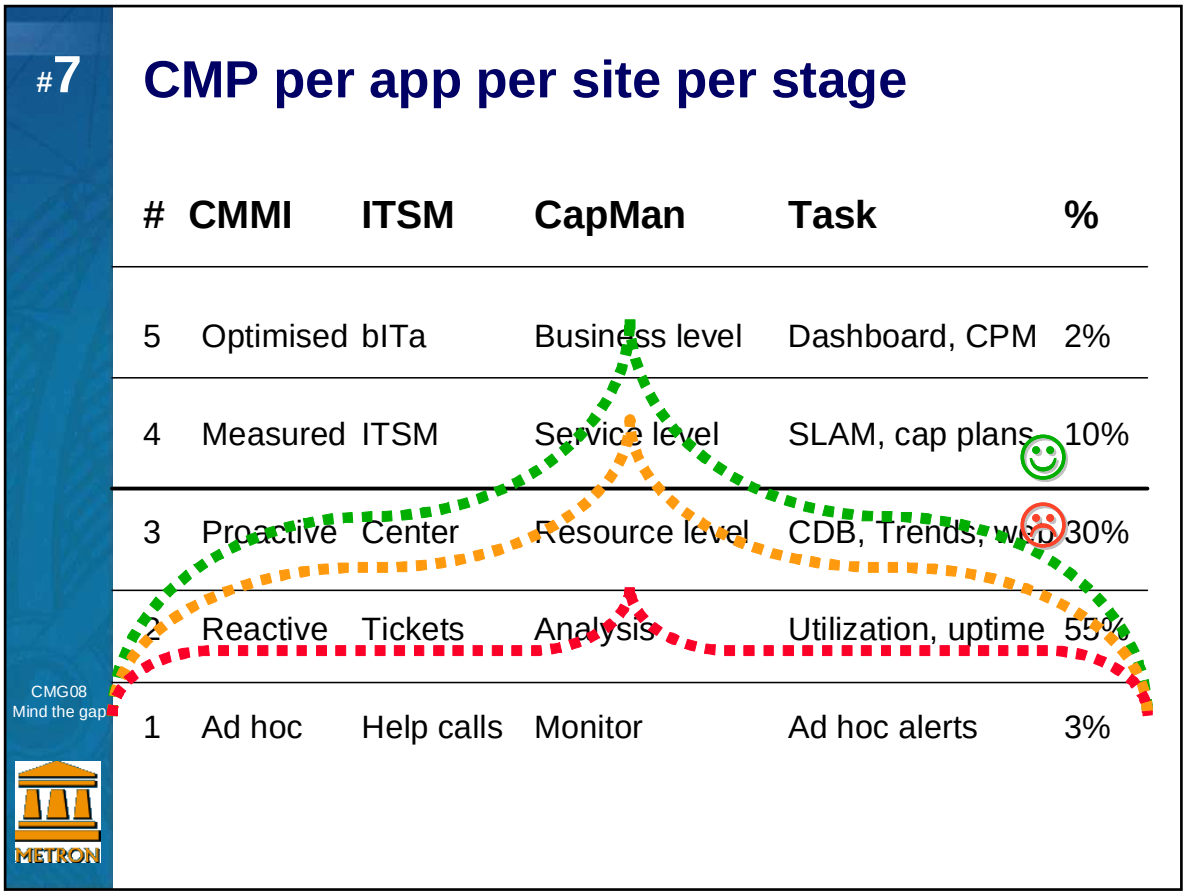
The infrastructure view has gained more impact with the momentum of ITIL. The library is now well known, albeit at a superficial level. The history has been discussed in many places with different levels of authority and memory accuracy. The key things to remember are why it was introduced and to what purpose. The "centre of IT expertise" for the UK government was aware in the 1980's of the increasing skills shortage in the public sector and the fact that for each new site, they paid significant money to external consultants to provide what was effectively an operations manual for ITSM. So they gathered together a general description of good practice from a number of sources, with a view to publishing it at no profit. I think that if the same team had tackled it now, it would be a free download off the web.

It was meant to be just a general description, independent of hardware ,operating system, software, database, network or any other variables. As such it was a question of "take it or leave it, adopt and adapt at will" without the implied "correct" answers for which of many processes would tackle which activity within the detailed dataflow definitions for any one site.

It does now carry such a large revenue from foundation training and certification that a whole army of false prophets have raised it to a new gospel-like level in order to drive the material into new areas and new markets. Maybe fragmentation of interests will cause fragmentation in the deliverables…



This slide is as detailed as ITIL gets in describing the capacity management process, sub-processes and activities. It presents a useful summary of the main activities and presents them at three levels being the resource or component level, the service and business levels. It is a neat overview and shows the essential nature of the capacity database and capacity management information system at the heart of the process. It is a key part of the 50 pages or so that describe the CMP in ITIL. The same topic is also discussed in ISO/IEC 20000 but summarised to just a few pages.



This slide shows the spectrum of implementation of ITSM and capacity management within it across sites. There are levels of process maturity suggested in this slide, which correspond roughly to typical levels of ITIL implementation. However, for any one application, it may move from one level to another during its life cycle – which on average is only around 18 months. Thus, at the start, it may be subjected to intense performance engineering and QA trials to derive resource requirements. But once it has settled down it may well be little monitored, or perhaps simple utilization trends maintained.

Also, for any site, given limited resources, it is likely that decisions will be made to go further up the maturity grid for servers that are expensive (mainframes and super-servers) and for services that are mission critical. Thus there is a “curly bracket” indicating that most sites will choose to monitor all servers and services, but only trend significant services and only model expensive servers.

My experience is that the higher up the grid that curly bracket is implemented, the higher is the regard for IT within the company and its perception externally. These are important considerations for a CIO whose life cycle on average is also around 18 months… so he or she is likely to be very interested in a gap analysis of what is actually going on in their capacity management domain.

#8

Summary of gap analysis procedure

- **Agree scope of study and checklist of targets**
  - Enterprise Infrastructure, Service/App Portfolio…
  - Mainframe, warehouse, UNIX, Linux, Windows…
  - LAN, WAN, SAN, PAN, RDBMS, ERP…
- **Assess vs. relevant checklists for Good Practice**
- **Review the Capacity Management Process (CMP)**
- **Talk to the Capacity Management Team(s) (CMT)**
- **Readily available reports submitted for analysis**
- **Interviews within a week to assess activities**
- **Review available evidence/ deliverables**
- **Reveal gaps (known and unknown)**
- **Identify SWOT and next steps.**
- **Report within a week for total focus on material**

CMG08
Mind the gap

The procedure for each gap analysis consultancy will vary somewhat in the nature of things. A lot of effort will be put into a Statement of Work or similar document ahead of the assignment. Much of that may well be changed in the light of the emerging realities. Nonetheless, the approach is usually much as indicated in this slide.

The agreed range of platforms has to be defined so that the scope of the review is clear from the start. This will in turn help to identify the key areas and hence the key players to be interviewed.

However, each interview, rather than trying to "score" activities on some arbitrary scale, is focused on current activities and deliverables. Each interviewee is asked to submit sample performance and capacity reports, ideally some that they are proud of and some that they had pain in creating – and any that they felt they should be able to generate but can't.

All the discussions are held in the light of the checklists for Good Practice in the areas selected by the CIO.

All of the interviews and sample reports are then analysed and a review produced, revealing gaps and identifying SWOT (Strengths, Weaknesses, Opportunities and Threats) and next steps.

Depending on the site and the number of target systems this exercise is typically two to four weeks. It is essentially a short snapshot where the consultant is totally immersed in the material to produce a quick report to management. It is not a vehicle for an extended on-site army of management consultants.

#9

Some case studies – CMP KPI

- **Ecommerce site**
  - Too busy installing new machines every day
- **Telecoms provider**
  - Doing it well but poor external perception
- **Public Sector**
  - Documented it well but doing it poorly
- **Finance House**
  - Implemented but low coverage outside center
- **Retailer**
  - Too busy reacting ad hoc to project demands.

CMG08
Mind the gap

Considering five major studies at particular sites, they can be reviewed at a high level to demonstrate the variety of approaches and expertise involved.

A short study at a successful ecommerce site demonstrated the difficulty of establishing business processes when the business growth was such that capacity planning amounted to deciding how many new machines should be added to each pool of the multi-tier solution every day.

A long study at a telecoms provider showed the situation where a CMT was doing CMP well and ensuring good use of existing capacity and planning well the future, but without reporting it widely or well.

A short review of a public sector site showed that it had understood the requirement for ITSM processes and documented proposed ITIL processes in some detail but had little resource to actually do it.

A short study at a finance house showed that there was the experience and expertise within the data center for effective and efficient capacity management, but less coverage outside their own domain.

The ad hoc retailer was too busy reacting to ad hoc project demands to establish any processes.

#10

Some case studies – all need numbers

- **Ecommerce site**
  - Non-optimal upgrades OK whilst money is spare…
- **Telecoms provider**
  - Enough for normal traffic, but not Christmas peak…
- **Public Sector**
  - Slow changes in applications or workloads…
- **Finance House**
  - Duplex everything and allow for 50% spare capacity on top of 50% for hot standby…
- **Retailer**
  - Poor results, economic downturn, be frugal and cover more machines with less staff…

CMG08
Mind the gap

Each of these five sites has different levels of CMP. Reporting on their own CMP activities varied from nil to extensive. But for different reasons, all five required the capacity management process to provide the measurement numbers they needed for all the services provided by IT.

The ecommerce site felt that non-optimal upgrades were fine so long as the growth continued. A short review of the capacity would remove the panics for performance.

The telecoms provider had the process in place but had no demonstrable deliverables or external measures of performance.

The public sector site had limited resources that were absorbed into daily performance issues and project work rather than establishing the framework to avoid such issues.

The finance house had previously worked to a protocol of triplexing everything and making sure that there were always three levels of support for every key component (such as power supply by grid, generators and solar panels with triplexing of each). However, as times change, so the move was towards duplexing and 50% spare capacity based on peak of peak predictions.

For the retailer, life was dominated by the economic downturn and so everything was ad hoc with less staff and more servers.

In all cases, much the same approach was used for the gap analysis study.

The rest of this paper shows the results for two of these studies in more detail. First the Finance House which was doing effective capacity management but wanted to improve the governance. Then the retailer which was somewhat less mature in its capacity management process.

#11

Triplex Finance House - TFH

- Service is vital, avoid single points of failure
- Triplex everything from datacenters to discs
- Power supply – 3 mains suppliers on 3 cables
- Standby generators
- Solar panel standby-standby
- All discs triplex mirrors
- Mainframes & non-stop, UNIX & Windows
- Already doing capacity plans for all domains
- Establishing an enterprise capacity plan
- And a dashboard & risk management regime for it
- Bleeding edge ITIL/ITSM site

CMG08
Mind the gap

The Finance House, Triplex Finance House, TFH, was known to have a highly qualified and large team of IT professionals (nearly 2000 with about half involved in development, with up to another 1000 sub-contractors called in on demand for major projects).

The service provided is so critical that downtime has to be minimized and performance optimized.

Although there is a mix of domains, as applications become more and more multi-tier, so it became felt that the capacity plan needed to be enterprise wide. However, the degree of metrics and planning in each is somewhat variable. Also, the aggregation of a number of separate plans from different authors into a single document takes a lot of time and editorial passes before it is acceptable to all. Such a large document tends to develop a life of its own.

#12

TFH - findings

- Capacity Management Team does both PE & CP
- PE growing on success of mainframe sizing project
- CP mature on mainframe & non-stop, less on distributed
- Services focus on ~90 category 1 & then ~90 category 2
- Process weaknesses have been identified internally recently and are being addressed but it will take time
- Need for:
  - clearer measures of drivers & KPIs
  - automated reporting to track, predict and police SLA threshold violations
  - integrated CDB
  - enhanced liaison with development, testing, SLAs, standard builds, changes, CMDB, monitoring …

CMG08
Mind the gap

METRON

Although the capacity management processes were in place, the coverage was not complete and a lot of reports were out-of-date. The CMDB changes were not advised to the CMT, so there were a significant number of essentially defunct machines still being reported on by various hand-crafted reporting regimes over the years.

The main areas needing enhancement lay in those of communication with other teams such as development and testing.

SLAs had some performance criteria, essentially on throughput and often related to what were effectively batch jobs updating the data warehouse.

#13

TFH – initial gap analysis

- **RCM well established for cat.1 & 2 services**
- **SCM being introduced for cat. 1 services**
- **BCM identified as next stage**
- **CMP dashboard initially "all green" – so**
- **No apparent shortcomings in objectives**
- **And no approval for any more expenditure**
- **Thus now need to show:**
  - current penetration of agreed target services
  - current gaps in detailed deliverables
  - underpinning checklists for deliverables etc
  - underpinning dataflows between functions

CMG08
Mind the gap

The initial gap analysis found that services had been effectively categorized but the service catalogue was still emerging. The resource/component capacity management processes were well established but service capacity management was just being introduced for some category 1 services. Business capacity management is identified as the next stage and will require more work on business drivers, KPIs and QoS.

An initial dashboard for management on the capacity management process itself was well structured but was completed to show everything as "all green". This had the unsurprising but unanticipated effect that the next request for expenditure was rejected as there were no current problems. The reporting process was then designed to incorporate an accurate reflection of some key metrics like coverage of services, servers, deliverables and underpinning metrics.

#14

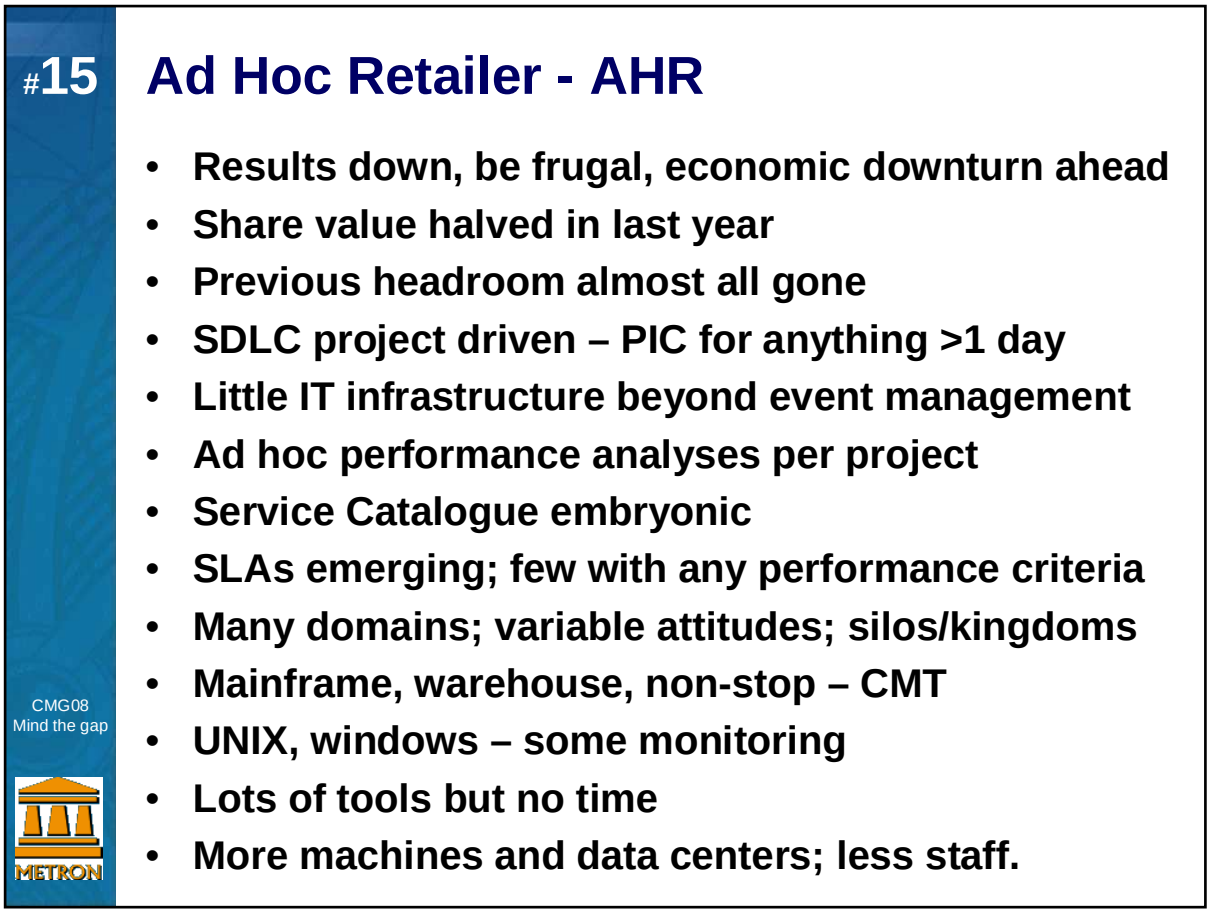
TFH – need to enhance:

- **Quality of info and dataflows between CMT &:**
  - Service Management
  - Change Management
  - Configuration Management
  - Production Monitoring
- **Measures of business drivers & KPIs:**
- **Reporting to track and predict SLA violations**
- **Service-server mapping and configuration changes**
- **Ability to clone reports with automatic trending etc**
- **Ability to analyse reports and identify bottlenecks**
- **Workload characterisation, workload forecasting**
- **Modelling scenario handling**

CMG08
Mind the gap

METRON

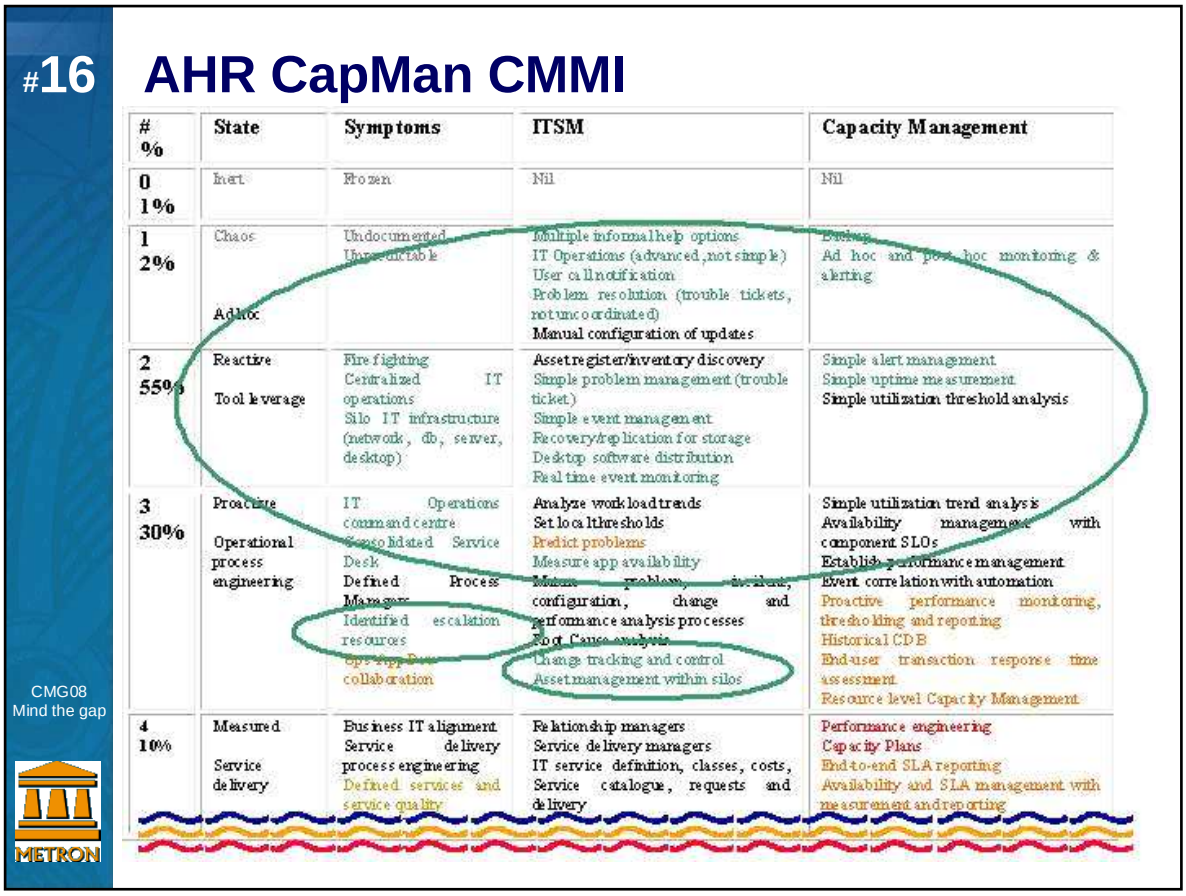
The main conclusions from the study were as shown in this slide. Essentially it revealed a need for more coverage, better information flow and more automation.



The Ad Hoc Retailer was a similarly large enterprise with major data centers and a large IT staff, including maybe 1000 developers. However, it had a chequered history in recent years with mergers and nw investors but an overall drop in market share, revenues and profit. The net result was that its share value had halved in the previous year and staff morale was low. Every decision was taken in the light of potential job losses and the need to be frugal. Headroom that had been established in the past was now effectively taken up so that capacity was becoming a major issue.

The enterprise had an SDLC culture with PICs (Project Initiation Charters) for any task requiring more than one days individual effort. However, there was little infrastructure leeway left, and there was little infrastructure management left in place. Previous processes had been dropped as staff had gone and more project work was being done by fewer people with more services on more servers.

There were ideas about a Service Catalogue and SLAs but they were emergent. There were many domains and also a history from different data centers so that there was a large number of what felt like separate kingdoms with barriers and lack of communication. Some of the project meetings were the first time that capacity management people in different teams had met.



This is the first of a number of sample reports for a typical (but for the purposes of anonymity, mythical) retail IT datacenter. I appreciate that it is difficult to read the detail on the screen, so it is copied below. However, it is only the principle of its contents that matter here.

The first few columns show the CMMI maturity levels and an indication of the population of sites that are in each; the state it reflects (active/reactive) and some ITSM symptoms that indicate the level attributes. Then there is a column for the corresponding ITSM activities and the capacity management activities typical of each level.

The cells for each intersection show relevant features, with traffic-light colors used to indicate where there are risks and dangers associated with the current level of maturity. The traffic-light is extended throughout the reports to allow for five colors. Grey or black is used to indicate comments that are not significant at this site. Red is clearly used for major issues. Orange is used for issues that are of concern and an off-yellow is used for areas where things are “on the edge” of becoming an issue.

The circles indicate where this particular site mostly lies.

#17

AHR CMP sub-processes

	Resource/Component	Service	Business
Monitoring			
	Some Metrics from some platforms, databases and a few applications collected and used for pagers but not held in a CDB Thresholds defined for metrics to allow exception reporting	Categorize resources by the services they provide as in the Service Catalogue If response time or throughput is part of the SLA ensure relevant data is collected	Categorize and weight services to reflect business significance in the light of business strategy and plans Access and collect Business metrics, Business Measures of Interest and KPIs
Analysis			
	Some CDB data is analyzed regularly at an agreed frequency covering an agreed period	Analyze response time to identify major contributor	Compare Business drivers with achieved service levels and resource utilization levels
	Some average and maximum values are recorded for each period, with bottlenecks, patterns and trends analyzed	Check SLAs measurable and achievable and report exceptions to SLAs – done for some key availability metrics and batch throughput	Measure and report on Business Metrics of Interest (BMIs), Quality of Service (QoS) and KPIs
Tuning			
	Areas where resource utilization can be improved are identified but proposed tuning activities are not modeled to assess potential benefit	Assess and model level of benefit of proposed tuning activities to the Service	Assess relevance and accuracy of Business drivers etc and feedback to the appropriate business unit
Implementation			
	Resource utilization figures are published to the agreed recipients	Publish achieved Service Levels to agreed recipients	Publish business volumes, KPIs and BMIs to agreed recipients
Demand Managem't			
	Identify resource utilization of individual services to enable Variable Charging	Identify and agree Services that provide Business Critical Applications	Agree weighting of criticality within the enterprise
	Identify utilization patterns of individual services to enable off-peak scheduling	Identify mechanisms for Variable charging on a per service basis	Agree charging mechanisms and options for workload smoothing
Modeling			
	Use trend analysis to assess likely resource utilizations per device – sometimes Model systems behavior under varying workloads and provide tuning recommendations – not done	Use trend analysis to assess changes in service workloads – sometimes Model systems behavior under varying business forecasts for workload changes – not done	Use trend analysis to assess changes to meet new business demands – some new projects Model systems behavior under varying changes to meet business needs - no

CMG08
Mind the gap

METRON

The second typical report is that relating capacity management to other main ITSM processes.

Across the top are the three main sub-processes, resource/component, service and business capacity management.

Down the side is a list of all the related activities.

In each cell there is a standard description of typical relationships and their manifestations.

Again, traffic-light colors are used to indicate where there are issues and risks with appropriate local edits to each description.

In this case there are some significant areas in red – though some risks have been left in grey as the site has make conscious decisions as to their lack of local relevance.

#18

AHR CMP interfaces to ITSM

To Capacity Management	From Capacity Management
From Availability Management	**To Availability Management**
Business processing and resilience requirements are passed to the CMT – maybe for some new projects with the new PIC Availability technologies ,both used and planned, are made available to the CMT	Make Availability Management aware of potential non-availability of resources due to capacity issues – maybe if liaison good with particular app dev team Feed regular performance, monitoring and alerting data to Availability Management
Component failure Impact Analyses are passed to the CMT – are incorporated in informal processes	Inform Availability Management of requirements for additional infrastructure needed for required level of resilience- on a project basis
From Change Management	**To Change Management** **Email changes and weekly review of risk level**
Change requests are forwarded to the CMT to review capacity and performance implications. – only for known major projects Details of changes to existing workloads are forwarded to the CMT - no	Provide predictions on the performance and capacity aspects of planned changes – only for some major projects Capacity Management is represented on the CAB as required for major matters
From Configuration Management	**To Configuration Management**
The CMDB is accessible to the CMT for read and update with CI changes communicated - no	Input the capacity effects of changes to the CMDB - no
Provide details of IT components and workload deployment across them - no	Input the capacity details of newly available resources to the CMDB - no
From Continuity Management	**To Continuity Management** **Ad-Hoc-Stores services run disaster recovery trials on full mainframe workload**
Business Continuity Plan and other service continuity considerations are passed to the CMT - no Vital Business Functions and their minimum processing requirements are defined for the CMT -no	Provide updates on minimum configuration requirements as live processing levels change Incorporate all recovery options into Capacity Plan
Raise RFCs as live processing changes - no	Assess the impact of RFCs on recovery options

CMG08
Mind the gap

METRON

A third report shows the dataflows across ITSM processes with the effectiveness of interfaces between capacity management and other processes.

The first column describes what is expected from other processes to capacity management. The second column shows what is expected from capacity management to other processes.

This is then addressed for availability management, then change management, configuration management, continuity, financial, incident, problem and so on.

Again, traffic-light colors are used to highlight local edits where there are issues of concern.

	Resource/Component	Service	Business
Monitoring			
	Some Metrics from some platforms, databases and a few applications collected and used for pagers but not held in a CDB Thresholds defined for metrics to allow exception reporting	Categorize resources by the services they provide as in the Service Catalogue If response time or throughput is part of the SLA ensure relevant data is collected	Categorize and weight services to reflect business significance in the light of business strategy and plans Access and collect Business metrics, Business Measures of Interest and KPIs
Analysis			
	Some CDB data is analyzed regularly at an agreed frequency covering an agreed period	Analyze response time to identify major contributor	Compare Business drivers with achieved service levels and resource utilization levels
	Some average and maximum values are recorded for each period, with bottlenecks, patterns and trends analyzed	Check SLAs measurable and achievable and report exceptions to SLAs – done for some key availability metrics and batch throughput	Measure and report on Business Metrics of Interest (BMIs), Quality of Service (QoS) and KPIs
Tuning			
	Areas where resource utilization can be improved are identified but proposed tuning activities are not modeled to assess potential benefit	Assess and model level of benefit of proposed tuning activities to the Service	Assess relevance and accuracy of Business drivers etc and feedback to the appropriate business unit
Implementation			
	Resource utilization figures are published to the agreed recipients	Publish achieved Service Levels to agreed recipients	Publish business volumes, KPIs and BMIs to agreed recipients
Demand Managem't			
	Identify resource utilization of individual services to enable Variable Charging	Identify and agree Services that provide Business Critical Applications	Agree weighting of criticality within the enterprise
	Identify utilization patterns of individual services to enable off-peak scheduling	Identify mechanisms for Variable charging on a per service basis	Agree charging mechanisms and options for workload smoothing
Modeling			
	Use trend analysis to assess likely resource utilizations per device - sometimes Model systems behavior under varying workloads and provide tuning recommendations – not done	Use trend analysis to assess changes in service workloads – sometimes Model systems behavior under varying business forecasts for workload changes – not done	Use trend analysis to assess changes to meet new business demands – some new projects Model systems behavior under varying changes to meet business needs - no

#19

AHR - Initial findings re CMP

- Monitoring and event management in place
- Performance and capacity management less so
- Largely living off previous “fat” in headroom
- Essentially now reactive based on firefighting
- Ad hoc reactions to meet specific project needs
- SDLC development driven rather than ITIL
- No IT Infrastructure or IT Service delivery view
- Project driven rather than process (need both)
- Application oriented rather than service
- Silo oriented, kingdoms, frontiers, lack of liaison
- Little performance or capacity detail in PIC
- PIC tends to be sized only for major new projects
- Little workload review of growth or business need
- SLA tracking and reporting mostly on availability
- Total service and server views missing.

CMG08
Mind the gap

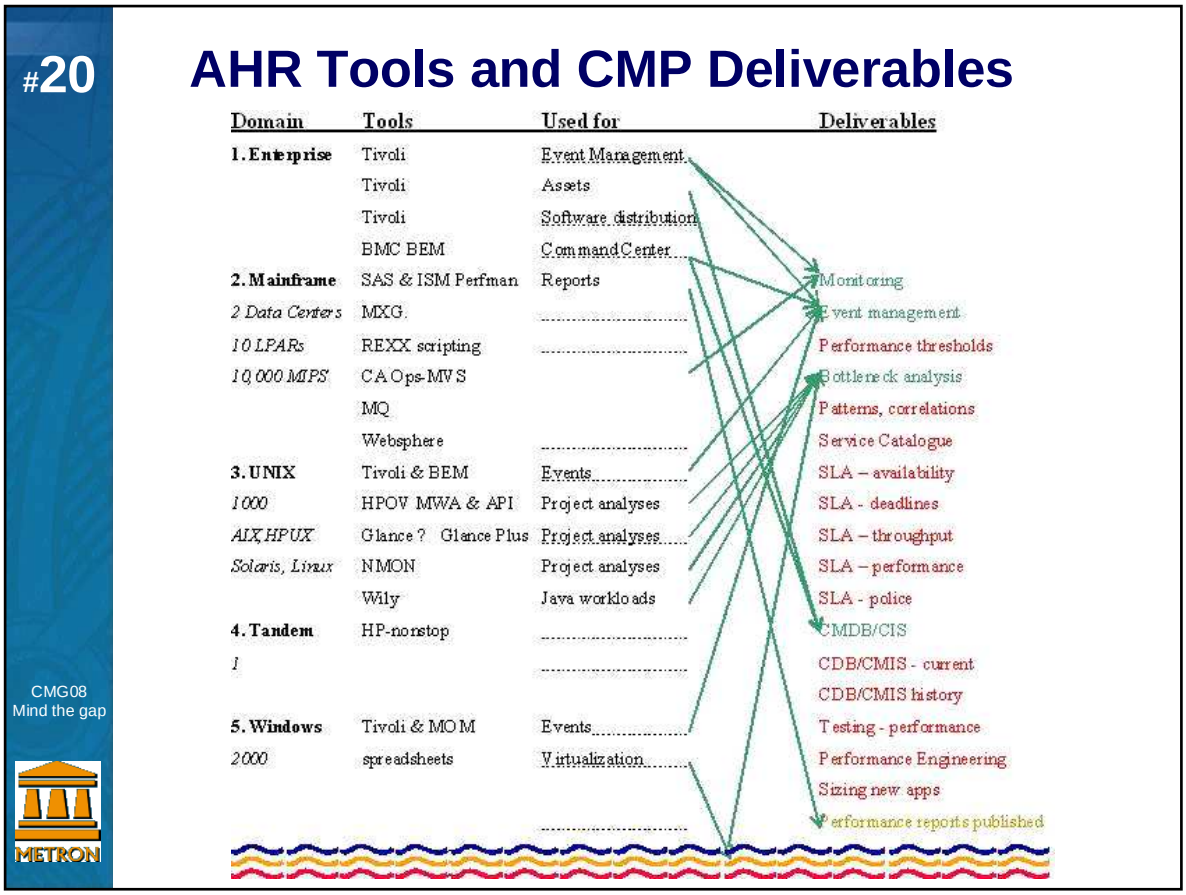
METRON

These initial reports allow a preliminary view of the capacity management process to be discussed early in the project. This initial deliverable is a useful test of the water to make sure that there are no surprises later.

At this sample site, an amalgam of real sites, development rules the roost and projects are well coordinated. The IT infrastructure is viewed largely as an event management facility to ensure availability. Performance targets are few and are essentially deadlines for batch completion overnight.

The main concerns revealed in the interview process was the lack of communication or awareness of other plaforms or processes. A silo oriented set of fiefdoms ruled each domain.

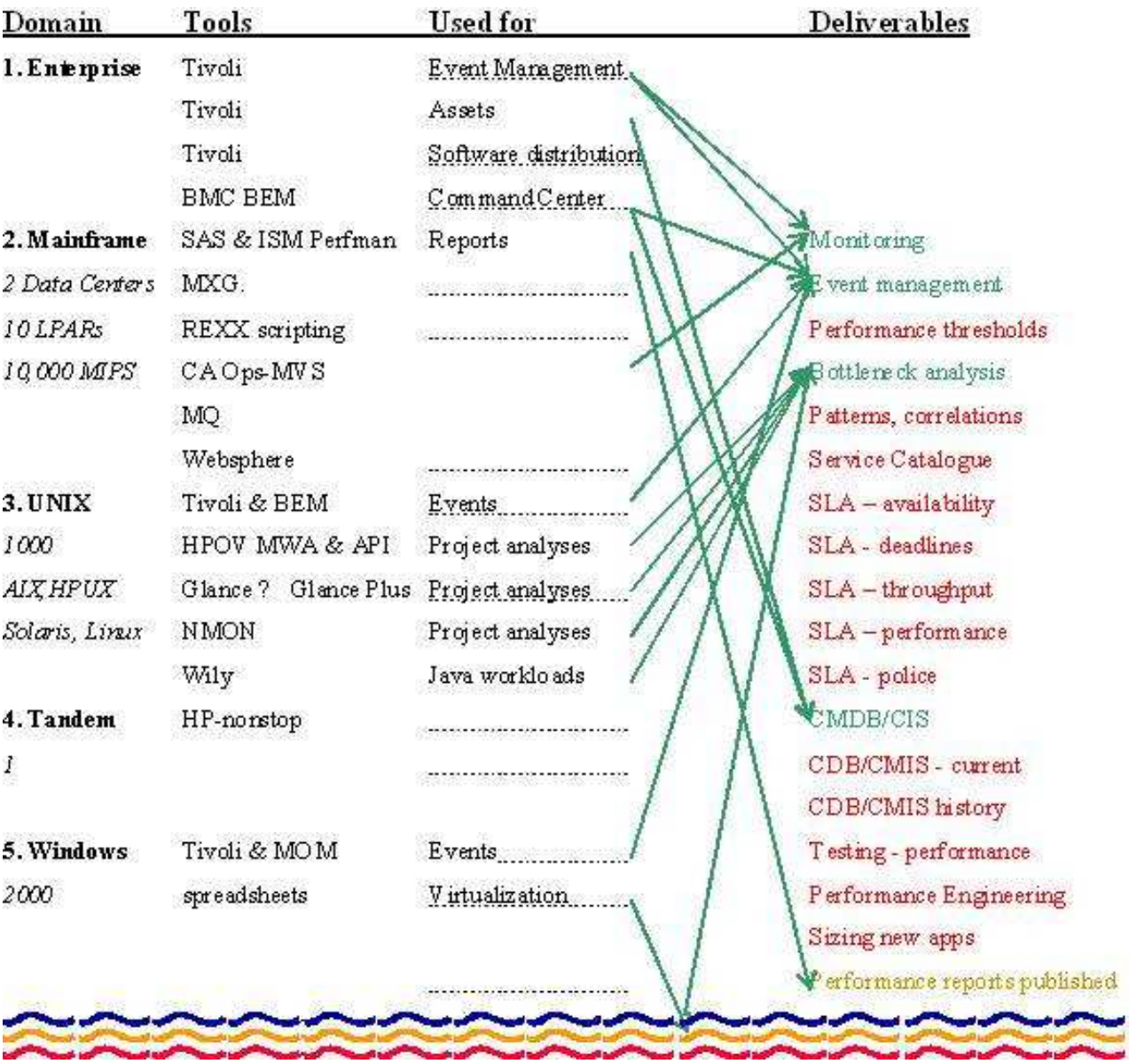
Sometimes such observations are already well known to those who asked for the project, but not always, so presentation of this initial tester often yields an interesting meeting.



This report relates the deliverables from the capacity management process to the tools available per target domain and how well they are used to meet the need.

The first column in this case shows the various target domains, the second the tools available. The last column shows the deliverables. The arrows indicate which parts of the tools are exploited to meet the required deliverables.

In this site there is a wide mix of tools from different datacenters that have been merged slowly over the years. So as well as supporting a number of different variants, there is also a central objective of "standardising" which typically adds an extra dimension in terms of support requirements which is an overhead until fully established and the previous standards retired.


Domain
Tools
Used for
Deliverables
1. Enterprise
Tivoli
Event Management
Tivoli
Assets
Tivoli
Software distribution
BMC BEM
CommandCenter
2. Mainframe
SAS & ISM Perfman
Reports
2 Data Centers
MXG.
10 LPARs
REXX scripting
10,000 MIPS
CA Ops-MVS
MQ
Websphere
3. UNIX
Tivoli & BEM
Events
1000
HPOV MWA & API
Project analyses
AIX, HPUX
Glance ? Glance Plus
Project analyses
Solaris, Linux
NMON
Project analyses
Wily
Java workloads
4. Tandem
HP-nonstop
1
5. Windows
Tivoli & MOM
Events
2000
spreadsheets
Virtualization
Monitoring
Event management
Performance thresholds
Bottleneck analysis
Patterns, correlations
Service Catalogue
SLA – availability
SLA - deadlines
SLA – throughput
SLA – performance
SLA - police
CMDB/CIS
CDB/CMIS - current
CDB/CMIS history
Testing - performance
Performance Engineering
Sizing new apps
Performance reports published

#21

AHR Metrics collected

Group	Metrics
CPU:	utilization (total; system; User; I/O wait; idle)
	queue length
Memory	available; utilization
	page in rate; page out rate, swap in rate; swap out rate
	% reads cached; % writes cached
Disk	utilization (total; by logical drive; by physical drive; by reads/writes
Workload	Number of logons; concurrent users
	CPU total by user/ command/ process
	Logical I/Os by user/ command/ process
Database	Number of users; sessions
	Cache effectiveness
	I/O profile (logical/physical) by user/ session
	Database size
	Long table scans
Filesystem	% space free/ used and GB space free/used
Application	Number of users; number of transactions
	CPU usage by user/ transaction type
	Response times and throughputs by transaction type

CMG08
Mind the gap

METRON

The next reports reflecting more detailed analysis starts with this review of the typical metrics available and those that are collected at this site circled in green. There is a potential issue with this slide in that each platform tends to be somewhat different in its level of detail. That is, most mainframe servers collect and store a wide array for SMF/RMF record types, whereas most UNIX servers will store a limited selection of statistics and Windows servers likewise. However, there is often a reasonably consistent attitude to storage and exploitation of metrics across the domains at this level of detail.

#22

AHR Information required for new apps

System Summary:	
Outline Description	A brief description of purpose of the systems
Environment	Batch, development or OLTP, mission critical service/contingency
	Software versions (e.g. OS, TP, Database, Development)
Timescales	Start Development/System Trials/Implementation, Complete Implementation or phases thereof
Development Workload	Number of Programmers terminals and Testing Requirements
Service Levels:	
Availability	How many days/week and which days
	How many hours/day and which hours
System Serviceability	% uptime target; Maximum recovery time
	Stand by arrangements: hot stand by, cold stand by
Terminals/Network	Mean Time between failures; Acceptable repair time
Data Base	Standby needed? Alternative network access routes
	Transaction logging; Duplexing of database
	Currency - e.g. must the database be 100% up to date, Frequency of security copying
Performance Requirements	Does any single user or group of users have priority?
-	Does any class of work have priority
For on-line transactions:	Mean response time; 90% response time (where relevant and measurable)
	Number of simultaneously active users
For Batch workloads:	Key deadlines for starting and/or completing jobs
-	Any requirement to run concurrent streams of work
For dev't workloads:	Mean & 95%ile response times for a simple interaction
	No of simultaneously active terminals
	Session throughput (sessions/hour)
Workload Definition:	
Session Characteristics	For each type of definable session:
	Type of user
	Number of sessions/day; Typical session length
	[illegible]

CMG08
Mind the gap

METRON

This report reflects the level of data required for effective performance engineering and the need for good dataflows with development and testing. Again circles are used to indicate what happens at this mythical site.

#23

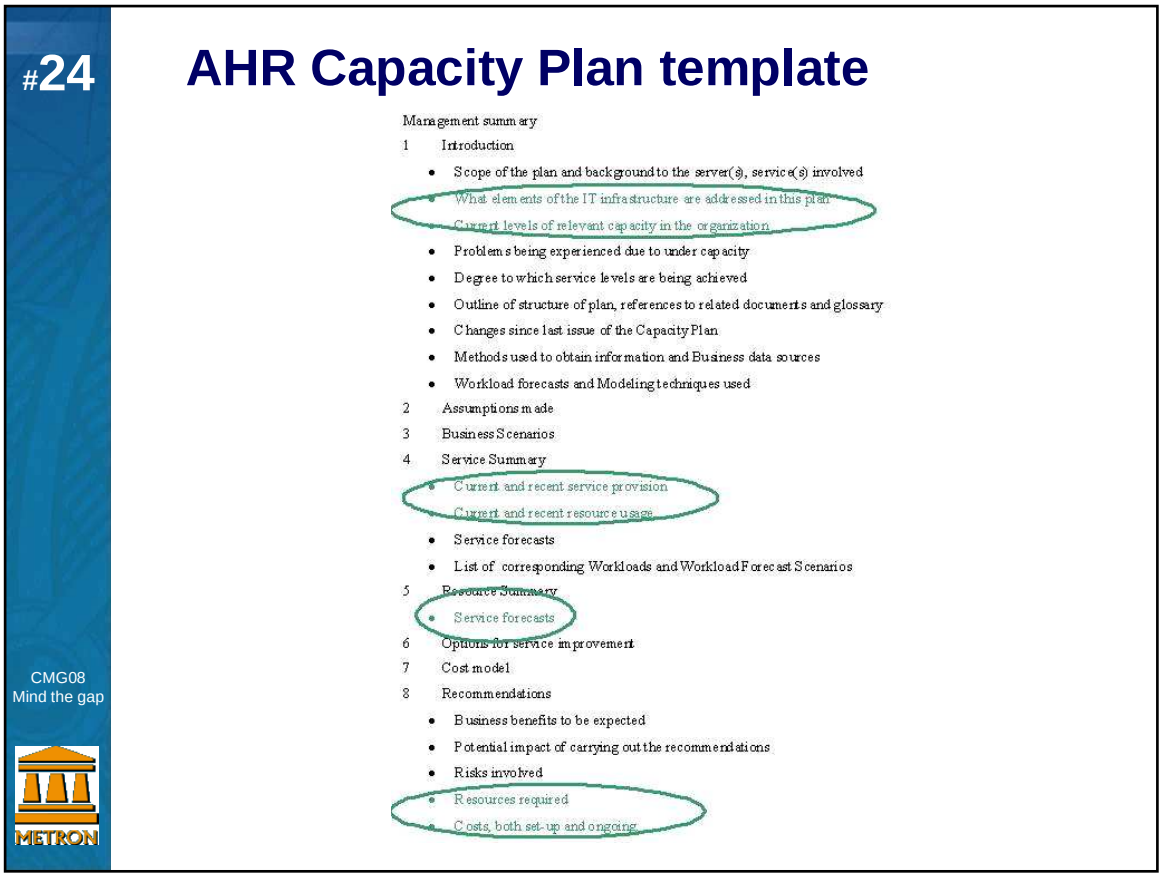
AHR Reports required

Analysis by System per node	
CPU by processor or total per interval:	% utilization (/system, /user, /IO Wait, /Idle).
Block device by device or total per interval:	% utilization, Read/Write requests/sec, etc
IO per interval:	Physical/Logical reads & writes per sec, % cached, etc
Free Memory per interval:	free memory pages, pages freed/scanned/sec, etc
Paging/Swapping Summary per interval by CPU or total:	Page/swap faults per sec, physical paging IO /sec, etc
Process per interval by CPU or total:	Process switches/sec, processes running, etc
System Calls per interval:	System calls/sec, System reads & writes/sec, etc
Analysis by user, user/command, command, process	
Activity per interval:	Username or command/process, elapsed time, No commands, etc
Analysis per file system per interval:	
	Total/% inodes/blocks/KB available/used/free
RDBMS Analysis	
Overview (per interval):	CPU usage, current logons, reads/writes etc
File IO per file id/name per interval or over period:	physical disk reads/writes, response time, Buffer cache hit ratio Logical I/O summary, per session, per user etc
Session Analysis per session	Metrics, Profile, Table scans, Redo log, Latches

CMG08
Mind the gap

METRON

This slide shows a typical list of reports required, with the related key metrics in the second column. Once again green circles are used to highlight the areas addressed within this site.



A standard capacity plan template is used to highlight those areas that are incorporated in the sites internal capacity plans. In this case it is clear that there are few formal capacity plans but within the capacity management team a number of key practical areas are assessed. The lack of a formal report leads to a poor perception of the abilities of those involved which could readily be rectified by a more open and transparent reporting system.

Management summary

1 Introduction

- Scope of the plan and background to the server(s), service(s) involved
- What elements of the IT infrastructure are addressed in this plan
- Current levels of relevant capacity in the organization
- Problems being experienced due to under capacity
- Degree to which service levels are being achieved
- Outline of structure of plan, references to related documents and glossary
- Changes since last issue of the Capacity Plan
- Methods used to obtain information and Business data sources
- Workload forecasts and Modeling techniques used

2 Assumptions made

3 Business Scenarios

4 Service Summary

- Current and recent service provision
- Current and recent resource usage
- Service forecasts
- List of corresponding Workloads and Workload Forecast Scenarios

5 Resource Summary

- Service forecasts

6 Options for service improvement

7 Cost model

8 Recommendations

- Business benefits to be expected
- Potential impact of carrying out the recommendations
- Risks involved
- Resources required
- Costs, both set-up and ongoing

#25

AHR SWOT with respect to the CMP

S	W
Monitoring	Performance Reporting
Event Management	Capacity Management
Project Management	CDB (capacity database)
CMT people	CMDB (asset register)
Previous experience	Service Processes
O	**T**
Few perf metrics in PID	Too few staff to action
Few perf targets in SLA	Too little money to drive
Current tools	Too little liaison
Scalable tools	Aggregated traffic changes
Automated tools	Business drivers & peaks

CMG08
Mind the gap

The Strengths-Weaknesses – Opportunities – Threats (SWOT) for each site is entirely local. In this case, there are the skills, expertise and tools available, but resources are so stretched across so many machines that matters are essentially ad hoc or reactive at best. This is compounded by the project culture which does not encourage good infrastructure-wide processes.

#26

AHR Highlights

Most potential impact:

- **SDLC project driver & ad hoc demands OK but add**
- **Matrix control with IT Infrastructure processes**
  - Process workgroups to coordinate activities

Most impact already in progress:

- **PIC needs more performance input**
  - Previous questionnaire “too long and complex”
  - Current drafts “too bland to instrument”
  - Needs more inter-team liaison to complete
- **SLA needs more performance criteria**
  - Business input of real requirements
  - And real traffic related to business drivers.

CMG08
Mind the gap

The highlights of a study again are site specific and need to be aimed at the management to try to establish a path forward.

#27

AHR Actionable Items

Strategic	Tactical
Set up CDB (database)	Review metrics per domain Review use of existing & potential tools Establish common core
Set up CMIS (info system)	Review reports per domain Review use of existing & potential tools Establish common core
Add to SLA	Review top SLAs Add performance and capacity criteria
Establish CMT	Core team per domain + inter-CMT liaison
Outline CMP	Inter ITSM team liaison: monitoring, tracing, tuning, testing, performance engineering development, business change management & SLM.

CMG08
Mind the gap

This summary slide tries to identify the key items to address in the roadmap. For each strategic target, tactical activities are identified.

#28

AHR Next steps = Review and improve:

- **Current Infrastructure processes and interactions**
- **All 16 CMP Good Practice checklists in detail**
- **Real benefits of VMware program**
- **Pilot prototypes on top 2 services per domain:**
  - CDB
  - Performance reports
  - Performance patterns and thresholds
  - Performance analyses and trends
  - Capacity plans
- **Pilot prototypes on top 2 servers per domain**
  - Workload characterization
  - Workload forecast scenarios
  - Resultant resource demand scenarios

CMG08
Mind the gap

This slide suggests some immediate next steps to help focus the management on practical solutions to address the worst of the gaps in the current process.

#29

AHR CMMI "now" and "then"

#	Objective	CMMI Now ⊙⊙-⊙	CMMI Then ⊙⊙-⊙	Caveats, notes
1.	**ITSM Capacity Management is in place**			Only CMT is in mainframe area
i	CM is mandatory	1	?	Essentially ad hoc or reactive only
ii	CM determines agreed services for action	2	3	Services mostly well known, but not logged.
iii	Capacity Plans are produced	1	?	Plans actioned by teams to ensure headroom
iv	A CDB is maintained.	1	?	No formal history or CDB across platforms
v	Future capacity demand is forecasted.	0	?	% increase for major new projects
2.	**Capacity Management is active.**			
i	CM activities are in place.	2	3	Reactive for most significant services
ii	CM sub-processes are in place.	1	?	Only resource level and only for some servers
Iii	CM team has documented processes	0	1	No specific team as such in most domains
3	**CM inputs are in place**			
i	Tools are in place for agreed metrics	1	?	Tools in place but not fully exploited
ii	Agreed data is provided by business units	0	1	Starting to ask for more info in PIC

CMG08
Mind the gap

METRON

This slide relates back to the original maturity review and suggests current levels of implementation of the main ITIL capacity management objectives. It then adds a suggested target level after an agreed project to address them, with some notes and caveats.

#	Objective	CMMI Now ⓪❶-❺	CMMI Then ⓪❷-❸	Caveats, notes
1.	**ITSM Capacity Management is in place**			Only CMT is in mainframe area
i	CM is mandatory	❶	②	Essentially ad hoc or reactive only
ii	CM determines agreed services for action	❷	③	Services mostly well known, but not logged.
iii	Capacity Plans are produced	❶	②	Plans actioned by teams to ensure headroom
iv	A CDB is maintained.	❶	②	No formal history or CDB across platforms
v	Future capacity demand is forecasted.	⓪	②	% increase for major new projects
2.	**Capacity Management is active.**			
i	CM activities are in place.	❷	③	Reactive for most significant services
ii	CM sub-processes are in place.	❶	②	Only resource level and only for some servers
Iii	CM team has documented processes	⓪	①	No specific team as such in most domains
3	**CM inputs are in place**			
i	Tools are in place for agreed metrics	❶	②	Tools in place but not fully exploited
ii	Agreed data is provided by business units	⓪	①	Starting to ask for more info in PIC

I have only indicated the findings for five sites and discussed two in more detail, showing a sample selection of reports for one of them. The same approach and extended set of checklists can be used for every site, but the factors that make each site different will apply to a gap analysis project too, so every study has its own character. I hope I have given you an appreciation of a couple of extreme examples.

Thank you for your time.

Any questions or comments?