Screaming into the void?
Are you a capacity manager who seems to just be screaming into the void?

As a Capacity Manager you will often start out telling people that something is going to happen in 6 months.

And you’ll tell them every month for 6 months

Until there’s a loud “twang” and something has to be done immediately to solve the problem you knew was going to happen 6 months ago.

At that point it’s likely the question will be raised “Why didn’t anybody tell us?” That’s when you need to be ready to strike. Ready with all the warnings, and a request to sort out the process to give you some teeth. Whatever senior manager just had their fingers burnt, just became your sponsor.

Of course ideally they will take notice of you from the start and avoid the issue. But....experience says otherwise.
Topics

- Just a Tickbox position
  - Capacity Management does what?
- Process, People, Products, Partners
- Proving value
- Workshops
- Services
- Replication
Ok so someone has been reading the internet and their ITIL documentation. They know someone is supposed to be doing Capacity Management, and until someone is doing it, they have an awkward question to answer in their yearly review. This is a problem. To solve this they get themselves someone to pin the role on. This lucky soul is now the Capacity Manager. Problem solved. The managers yearly review is now one item easier. But the Capacity Manager can find they have no real process to properly interact with, and nobody with any teeth to implement a process.

This can be the downside of frameworks like ITIL. People implement things, to follow the framework, with less consideration around “why” something should be done than is wise. They want to tick the box, call it a success, and move on.

The way we get the Capacity Manager integrated, is education. You may very well have to educate everybody from Sys-Admins to the CIO, and fingers crossed even the CEO, about what Capacity Management is actually trying to do. So what is it Capacity Management is trying to do?
So this is my personal example of a tick box position. I applied for and took the position of Capacity Manager at a large utility company.

They had recently given up trying to be an outsourcer for other companies and were “merging” IT positions back into the main company. In the process they annoyed the Capacity Manager and he was leaving.

In my interview it was clear that tooling wasn’t in place, “just a few spreadsheets” was what they had. But part of the role would be to choose proper tooling and do the job properly.

In reality I found a second PC under a desk, running thousands of lines of VB code, and multiple cascading MS Access databases...with a few spreadsheets on top. You could keep this running by seeing what failed, and editing the code and so on. But really it was a warning sign that Capacity wasn’t taken seriously. After starting the process to replace this in house system, it became clear the process would be too expensive (not the tool...the process which was 2/3 the total cost), and the plug was pulled. Just keep the existing system running.

Then I started to notice that if I made recommendations, that could save hundreds

<table>
<thead>
<tr>
<th>Personal Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Utility Company</td>
</tr>
<tr>
<td>• Capacity Manager position suddenly vacant</td>
</tr>
<tr>
<td>• Home grown MS Access solution on a PC under desk</td>
</tr>
<tr>
<td>• Promise to implement proper tooling</td>
</tr>
<tr>
<td>• Budget withdrawn</td>
</tr>
<tr>
<td>• Capacity recommendations ignored</td>
</tr>
<tr>
<td>• Continue to over provision</td>
</tr>
</tbody>
</table>
of thousands of pounds, they were ignored and over ruled by senior managers. “Because we did it that way last time”.

This is a tick box position. It didn’t matter if I did my job or not. So long as I turned up, sat in the seat, and could vaguely describe my supposed purpose. The problem that management saw, was solved. They had a Capacity Manager. They didn’t know why the position was important to them, and they were not going to be told what to do by some new guy in his 20s.
Ok so here it’s time for some audience participation.

(Obviously if you’re reading this as a PDF, you should feel free to set a 30 second timer on your phone and give your own elevator pitch. Oh go on, it’ll be good practice I promise).

Because volunteers are usually hard to come by, lets see a show of hands for anybody who’s job involves capacity management.
Ok so who’s been doing this the longest?

You’ve got 30 seconds to describe your role.

Your Elevator Pitch

- Hands up if your job is “Capacity Manager” or similar
- Who’s been doing this the longest?
- You’ve got 30 seconds to tell us your job
- Do you ever feel like you are screaming into the void?
We’ve probably all heard the story about the cleaner/janitor at NASA, which go’s as follows.

*During a visit to the NASA space center in 1962, President John F. Kennedy noticed a janitor carrying a broom. He interrupted his tour, walked over to the man and said, “Hi, I’m Jack Kennedy. What are you doing?”*

“Well, Mr. President,” the janitor responded, “I’m helping put a man on the moon.”

The simple mechanics of his role are not so important. The outcome of that person’s work as part of a bigger enterprise, will be a man on the moon.

So what do I think we are there to do?
Ultimately the Capacity Managers role is about balancing money vs capability. We don’t want to spend money, and we want to make the most of what’s already been spent, while making sure the customer get’s good service.

How do we do it?
We improve the RoI on infrastructure, while reducing software and staff costs. By doing this we actually improve customer service, increase profitability, and avoid
negative publicity. Part of the way we do that is by being a link between the Business and IT.

If you can educate the right people into understanding what the Capacity Manager does, then you can work on the mechanics of the role, and correct failures within the organisation that may be blocking the positive impact the role can have.
Assuming you do feel like you’re “Screaming into the void”, you’ve probably got a relatively immature capacity management situation happening. Most things will be happening within individual Silos and at a component level. Then there will be monitoring and alerting for each of the silos, with people running around putting fires out, followed by some back slapping and crowing about a job well done.

Services will be checking that they are running in some manner. Simulated transactions, that kind of thing.

Capacity is probably something that’s worried about when a project is run to implement something, then once implemented and thrown over the fence, it’s supports problem.

There will of course always be incidents that have nothing to do with capacity, and reactive processes need to be in place to catch those and fix them. But we really should be aiming to have no capacity related incidents. i.e. things we should have predicted 6-12 months ago, or before a project happened.
How do we end up in that situation? Well, Capacity isn’t something that bites you on the ass on day one. Not being able to handle incidents or change, well that’s going to hit you right away. But Capacity...it can wait. Of course by then other processes are sort of undertaking the role well enough that it appears to be a low level problem, and implementing anything is seen as more difficult and costly than it requires.

Partly this is because all the benefits are unclear. “We can just buy more when we need it”. We’ve all heard that one. With unclear benefit’s it’s difficult to get resource applied to the task. To top it off, the technologies required for capacity management are not understood. “We already have monitoring and it show’s me what’s used”.

Basically it’s a bit of chaos stirred in with assumptions and misunderstanding.
Capacity Management is based on information. At each level we need to get the correct information to provide the correct answers. Where capacity management is immature, this flow of information tends not to be happening.

From a business level we’re looking for information about the long term strategy, and plans. If the business is intending to double in size over the next 3 years, then it’s important to start planning the IT capacity to support that now. It’s important that the business understands that you will need that information to support them also. Much like the technical teams, they tend to like to hold onto their domain and treat any outsider asking for information as trouble. Educate these people so they realise you’re trying to support their plans, but you can only do it if you know what the plans are.

For the service you need to understand how it fits together. What bits of the service will grow with increased workloads, how’s it currently performing (how should it be performing). While some technical elements might be involved in this information it’s still mostly about concepts. E.g. there’s a batch run that happens every month and it takes 8 hours running on the x y and z systems, it has a window of 9 hours. The service owner needs to realise that you are looking to help them, look after their service. There should be regular meetings (possibly short), between Capacity and the Service owner, to explain what the capacity risks are that the service is currently exposed to. The capacity team should be looking to adapt their information and reporting to the service owner to support that service.

Finally the component monitoring is the technical part. The collection and storage of the underlying performance and capacity data, combined with knowledge of how the platforms work from the platform experts. This is the easy bit to get across to other people. Again technical Silos tend to resent someone looking into their realm. But with education around what you are looking to achieve, and that what you’re doing is looking to make their life easier, this should be able to be overcome.
Ok so lets say you have a process and all the information is at your fingertips. What are you going to do with it. Organisations that have not taken Capacity Management past the “oh the support team do that” stage are likely to have a bunch of monitoring tools, none of which talk nicely to each other and definitely don’t present the data in a single location or format. What they do present may not even be trending forwards into the future. If it is it’s probably simplistic trending, rather than modelling the enviornment.

If you want to implement anything you’ll have to fight on a number of fronts. No agents: Nobody want’s more software running on servers. The world is complex enough, and don’t we have monitoring agents already? Nobody want’s to take on the costs of change and maintenance of a new set of agents. Costs: Tools are generally not free so you’ll have a licence cost Connectivity: If you want to bring all that data into a single place, you have to deal with a lot of people all presenting their own obstacles. But you need a central point to bring that data together to do this.
We’ve probably hit enough of the negatives. What can we do about it. Well we can focus on the 4 Ps. They are in this order for a specific reason.

Get the process sorted, which will justify the people and give them something to do. Then they might need products to do it and Partners can fill gaps left if needed.

Process, People, Products, Partners.
The deliverable from the Process is the Capacity Plan. So everything else we are doing is focussed on that.

Process Governance
The process needs to offer value.

Interface and Information
The Capacity Process will be interfacing with a number of other processes. Change, Event Management, Financial Management, the Service Desk etc. These interfaces need to be defined up front, and the other teams buying into the process to ensure it works. Information needs to flow both ways in most cases. Capacity need to be aware of changes coming down the line, as well as feeding into change requests going into the change process.

Metrics, report and thresholds
Without data the process is useless. Platform metrics are easily gathered, but this also covers KPIs from the service and business forecasts feeding into the future plans and risks. As well as the metrics, thresholds need to be defined and just as importantly tuned over time. The quality of alerting relies on thresholds being correct. Too many false alerts can destroy trust in the process.

Deliverables
We already mentioned the capacity plan. It’s not necessarily a single document. There are likely to be specific plans for each service, provided to each service owner, and a plan for the business presented to the business.
Once we have a process, we need the people to make it happen.

Not every organisation is going to be lucky enough to have a full time Capacity Manager. It’s likely to be a part of someone’s position. However that’s organised that time has to be spent by someone, on running the process. Ideally yes it’d be a full time position, but sometimes it’s useful for the Capacity Manager to have connections with other teams.

The Capacity Manager needs to have the right set of skills. It’s not enough to understand the intricacies of how HP-UX virtualisation works. The capacity manager and/or team need a broader set of skills. One of our most successful clients, who started with no real process or people, imported a business analyst into the capacity team (of two). By selecting the right candidate, they have merged into the team someone who knows the service owners, knows many of the services, and how they work. Combining that, with the technical knowledge of the original team member has created a formidable team.

The team need to have the authority to implement the process, and to get time with the service owners. Otherwise it’s back to being a tick box position, and no matter how great the process and the people, if they are being ignored, bad things will
eventually happen.

At Metron we like a bit of Greek Mythology. You don’t want to be Cassandra. Cassandra had the gift of prophecy, but was cursed so that nobody would believe her. Being a Capacity Manager with no authority, is very similar!
In any organisation there are going to be plenty of products that contain useful data.

- BI tools can tell you about the business volumes going through the services
- APM tools can tell you how long transactions are taking to be processed, and where they are being processed
- Monitoring tools (SCOM/OEM) can provide you with raw performance data from the hardware
- Platform Management tools can give you insight and raw data for specific hardware platforms (Storage tools)
- The CMDB should be the definitive list of what hardware exists out there, to support which service, so you can know what to monitor.
- Then there’s all sorts of home grown scripts producing data files and databases with useful information in them.

That’s great, but what are you going to use to collate all that data into your capacity plan(s)? Of course the business would like you to use Excel as you already have it. Which is fine, if you’re just doing a very basic Capacity Plan on a little data and have all the time in the world to do it.

But a dedicated capacity tool is probably wise, particularly if there is a requirement
for the capacity information to be accessible via self-service dashboards and for the capacity alerts to be automatically integrated with the service desk. You may even be feeding the results back into the BI tool(s), and eventually you end up with something that looks like this. You’ll note that Excel didn’t vanish. If I go a day without using Excel I’d be surprised. It’s just so useful for manually playing around with data.
Well the final P was Partners.

You don’t have to have Partners, but sometimes they are needed. Possibly the team is shrinking, and you might want them to play a larger role in the Capacity process. They might be your software supplier, or someone else. With a properly laid out process, and toolset, it should be possible to bring partners in to cover gaps you might have. Be that a lack of skills, a lack of time, or a specific exercise you need to get completed.
Near the start of this I showed a slide with an “immature” capacity management process. It had quite a lot of blank space on it.

If the 4 Ps have been put in place, then all these items should exist.

Unified reporting across Component, Service and Business.

Planning based on the requirements of the business, and services, not just looking at hardware usage and predicting forwards hoping everything will stay the same.

Integration will lots of other processes and teams.

A shift from Reactive monitoring, to Proactive management and planning.
To get to that mature capacity process, you’ll need to educate other people about what you do. One of the better ways we’ve found to do this is create a short “workshop”. Something that’s say 15 to 20 minutes, and that lays out the purpose of the Capacity process, and how the people invited to that workshop will benefit from it.

You’ll probably want to start with the sponsor (or people you are trying to recruit as sponsors). Extend you elevator pitch (that you’ve all been thinking about since someone got picked on!), to present at a high level how Capacity Management is going to save money, and make happy customers. Use relevant examples from within the organisation. Point out the pain the organisation has gone through recently because of capacity problems, and use that as the leaver to introduce better Capacity processes.

Once you have the Sponsor on board, you can start to meet with Service Owners, take them through the process and talk about how it’ll improve their life. They’ll have KPIs they need to hit, use the workshop to find out what they are for each service, and talk about how you’re going to help the service achieve them.

With one or more services on board, you can start to interact with the support

---

Workshops

- Extend your elevator pitch
- Sponsor
  - Savings
  - Pain avoidance
- Service Owners
  - Customer Service
  - KPIs
- Support teams
  - Pro-Active
  - Less out of hours work
teams, lay out the plan and who’s behind it, but expect an interactive session. Ultimately you want to make their lives easier, and let them spend more time at home with the kids, rather than answering the “on call” phone. You’re not looking to change their tools, but you probably want to call on their data and expertise with their platform.

In all of these, workshops you don’t need every Service Owner, or every platform team. Pick the ones who recently experienced pain, do a good job for them, and keep doing that. What we find is that after a while, service owners will approach the Capacity Team asking for the same reports as X, Y and Z get.
So let’s say we have a service on board. We’re collecting the relevant platform and service data. We’re sending alerts to the service desk, giving them plenty of time to avoid capacity related incidents for this service. So everything is ticking along nicely. With one exception. We’re spending money on a process with nothing to show for it.

You need to identify your own KPIs and start monitoring them.

If you do a good job, what is going to happen?
• Capacity Related Incidents Reduce
• Emergency Purchases Reduce
• Service Quality Improves

I’ve spoken to Capacity Managers who collect the stats from the service desk showing the number of capacity related incidents. They are then reporting on that.

The same with keeping their eye open for emergency purchases of hardware due to a lack of capacity. Worryingly this still seems to be disk space in the main. One of the easier things to monitor and predict when it’ll run out.
Service Quality is a tricky one. But you might have a correlation between the reducing capacity incidents and improving service.

Where possible turn these into cash. What does a capacity related incident typically cost the business. How many did you avert. Do the simple addition. Yes it’s a wishy washy estimate. But estimates are MUCH better than shrugged shoulders.

So now you have more stats, showing what a great job you are doing, but you have to get them in front of people. The workshops got the process started, but regular review meetings will keep it going. They don’t need to be long. Just a quick chat through the capacity of a service with the service owner, risks for the future, and any successes you’ve seen over time.
This is an example of a slide used to review services by one of our clients.

You can pretty much expect the business to look at nothing except the first 3 cells for Overall, Trend and Outlook. Is it green? Will it still be green? Yes. Move on.

The center part will be interesting to the Service Owner. These are all the different applications and technologies that go into supporting the service, and you can see where there are some risks that need attention.

Finally at the bottom is a plan of actions to resolve those risks. It shouldn’t take long to review this slide and any supporting evidence with a Service Owner. But it will keep them informed and make them aware of up coming issues or costs that they will need to deal with.
We've worked over several years with a rapidly growing financial company. They didn’t have a centralised capacity role. But they did have a very expensive upgrade to an Oracle Cluster that they needed to size. They came to us to help them capacity plan that upgrade. I doing so, they saved a significant amount of money, by selecting different hardware to their original plan. This work, on a single platform, was used internally to push for a centralised Capacity role, that would ensure other platforms were also not wasting money.

Chris here was given the role. With a largely UNIX background he’d had UNIX capacity as part of his role and done a good job hand crafting some reporting. But it was limited. So they purchased a tool, to free up Chris’s time, to be clever with the data, not worry about getting the data. This allowed Chris to start interfacing with other technical teams and silos. Finding out what mattered to them, and what they needed to know. But he was struggling to engage the Service Owners.

At this point they imported Wayne. Now Wayne had little IT knowledge, but was a business analyst. He knows data, he knows the business services, and importantly, he knows the people. Now the door is wedged open and reporting can be tailored to the Service Owners requirements. The introduction of APM tooling, further strengthened that relationship by being able to give a direct correlation between IT

<table>
<thead>
<tr>
<th>Lets talk about that client</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Growing Financial Company</td>
</tr>
<tr>
<td>• No formal capacity process</td>
</tr>
<tr>
<td>• Large DB Server upgrade</td>
</tr>
<tr>
<td>- (Reduced costs via capacity planning)</td>
</tr>
<tr>
<td>• UNIX admin given the Capacity Manager role</td>
</tr>
<tr>
<td>• Cross Platform tooling purchased</td>
</tr>
<tr>
<td>• Interfacing with technical teams/silos</td>
</tr>
<tr>
<td>• Business Analyst joins team</td>
</tr>
<tr>
<td>• Interface with Service Owners</td>
</tr>
<tr>
<td>• APM Tooling</td>
</tr>
</tbody>
</table>
capacity, and business volumes.
So here’s a simple chart. Taken completely in isolation (Infrastructure Capacity Management), it’s not immediately obvious what the capacity of this system is. We’ve got two possible peaks. Experience will tell us that the peak in the early hours is probably a bit of offline work. Backup or a batch run. While the online day peak is probably transaction related.

But do we have transactions 24 hours a day? We have another 30% CPU we can probably use before performance starts to degrade. But we’ve no idea what that means to the business. Is that 20% more business, 40%? We can’t really say.
If we move into Service Capacity management and start putting the transaction volumes against utilisation. We can quickly see that there is a 20% CPU overhead for the system doing “nothing”. When transactions peak, there were 340, vs 50% CPU. Given the 20% overhead, that’s 113.3 transactions for every 10% CPU. Or in simpler terms, we can double the number of transactions before we hit 80% CPU.
Of course we could trend existing CPU usage in isolation and make predictions based on history. E.g. this system has much more CPU than it needs, it’s not going to have a capacity problem in the next few months. But the business doesn’t stay static, and they need to know what’s possible in numbers that they do understand. They also can’t be looking at 1000s of charts to make that call.
If we’re going to report service capacity projected into the future, we need a more condensed view of the IT estate. Something to give a single glance, and get some idea of how concerned we should be.

Perhaps rather than trend charts, we’re looking at a heat map, with values to indicate how long until a threshold is reached.
With the Service Capacity reporting we also want to be indicating the costs of the capacity used (and perhaps also the costs of the unused capacity, a future capacity KPI perhaps?). Ok so this again might be an estimation. But if you can find out what was paid for the hardware, how long before it’s expected to be replaced, and it’s expected energy consumption. Then you can make a stab at it. If you can get to an agreed value, and you can report on it, it’s really useful information for the business.
This is why that cost information can be really useful.

Let’s say we are running in a cloud environment. So our costs are very well understood and can be taken into the reporting. But with systems spinning up and down as required, our costs are not static. If we can put costs against the transactions from the service capacity reporting, we can start to see how much each transaction is costing us to process. Now we might have a spend limit/budget that we’re not allowed to exceed. This will then tell us how many transactions we can handle before we are at risk of over spending.
Just to review the difference between Infrastructure and Service Capacity.

With just the infrastructure we have nothing to relate the used capacity to. Was it wasted effort? All we can do is monitor the past and trend forwards, in the hope the current workload grows as it has previously. The platform teams might be on top of that, but experience says…probably not.

By stepping up to Service Capacity Management, now we’re starting to talk the companies language. They can see what will be affected if a capacity problem exists, and if we can include costs we can start to provide a “unit IT cost” for each transaction.
Chances are that you are going to get buy in from a few people related to one service to start with. Probably whoever just experienced some pain. But you need to be prepared for the next service, and the next, and so on. So when you’re making your process and reporting. Consider how you are going to replicate this to other services. Make sure the 4 Ps are not only appropriate to one service.

Once you have your first service, demonstrate what you are doing, to other services. Make sure though that meetings are two way affairs. You want to understand their service, as much as you can, so the reporting makes sense to them.

Options/Customisation. There’s a reason car manufacturers offer options. People like to have input into the item they receive, they’ll buy into it more if they can make it theirs. So make the deliverable theirs. If you have to call Transactions “Trades”, or servers become “hosts”. Do that for them. If they only want a few high level reports, or if they want more supporting charts. Do that for them. They may even have some good ideas you can use on other services.
Some of you may recognise this face. Malcom has been around the industry for a couple of years.

So he knows what needs doing, and he’d moved into a new role at a large outsourcer.

As with most outsourcers Capacity reporting was included in the contract to some extent or other, and some clients cared more than others about what that actually means. So there was a mix of solutions for different clients and it was unfortunately largely a manual process. Skilled people used to copy and paste data from one application to Excel. Possibly highlight a row and calculate a total or an average. That kind of thing.

What it meant was that the reports being produced were limited, and the value to the consumer of the report was low. Malcolm is where I stole my “You spend 6 months telling them, then when it breaks, you tell them you told them, that’s when you can get something done” idea. He’s seen it over and over again. When capacity related pain was experienced by the business, he could jump in and say “I can stop this happening again”.

The other client I wanted to mention

- Knows Capacity Management
- New Role at large growing outsourcer
- Capacity Reporting in contracts
- Heavily Manual Process
  - Copy, Paste, Copy, Paste, Copy, (please just kill me)
- Tooling recently implemented
- Defined 3 Levels of Capacity Management Offer
- Automated Process as far as practical
Not surprisingly where I’m involved is that they had purchased a Capacity Tool to get rid of the manual process as much as possible. But not just that. Malcolm wanted to refine the process, and get all the clients into a standardised capacity service offering clearly defined levels of output. Gold, Silver and Bronze if you will. Where Bronze might be the basic automated reporting, Silver would include more data, and a small amount of time from the team to assess things, and Gold being a full on capacity plan with models and so forth. Creating the levels allows each level of service to be automated as far as possible. It also allows the capacity service to “advertise” exactly what you can get at each level.
Roundup

- Know your value (Elevator pitch)
- Wrangle the 4 Ps into place
- Run Workshops
- Meet service owners (Coffee)
- Track your value
- Standardise
- Replicate