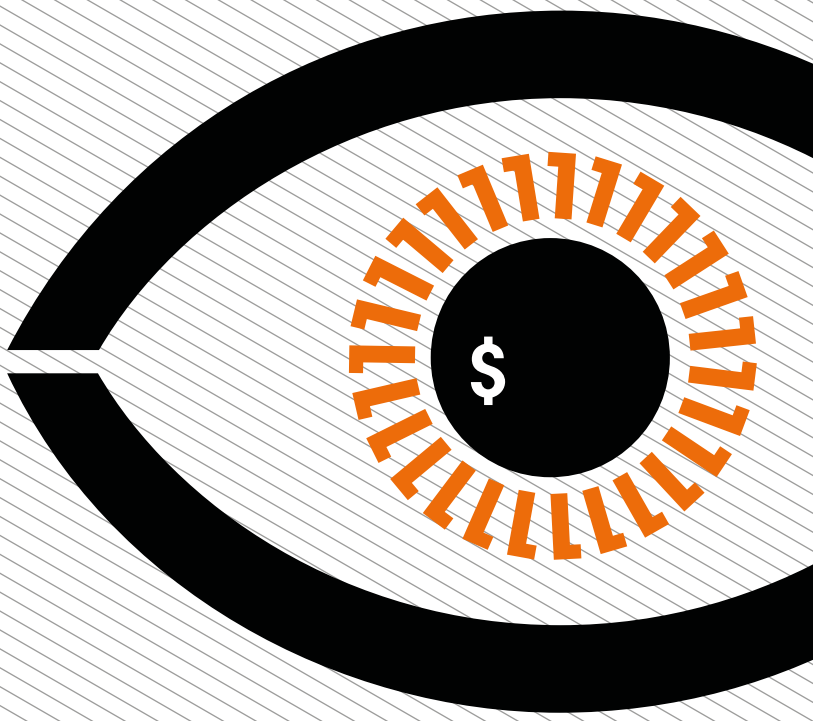

The 3rd age of capacity planning

The return of capacity planning as an essential part of IT management

sumerian®
Forward Thinking

White Paper

Why you need capacity planning more than ever: The 3rd age of capacity planning has arrived.

The 1st Age—In the beginning

Management and efficiency

I am reminded of a scene similar to that of 2001 A Space Odyssey when primates first discovered tools as they sat in front of a large black monolith. Looking back, for all intents and purposes that monolith may as well have represented the mainframe and the rock the tools available at the time.

Within the single silo'd environment of the mainframe, supporting tools were rudimentary but fit for purpose as the relative complexity of the estate had well defined boundaries and provided a hard line constraint. Compute resources were treated as finite as the cost of expansion was prohibitive and it was considered time well spent to hone and chisel code to the nth degree. Human effort was cheaper than machine time.

Life was a lot simpler then.

The 2nd Age—Rise of the machines

Orchestration and performance

With the consumerization of technology and the birth and rapid acceleration of the Internet the challenge of the capacity planner shifted into the world of distributed computing.

In this age capacity management focused more on load balancing and avoiding bottlenecks within the myriad of loosely coupled parts that made up a company's website and application stack.

Throughput and performance replaced efficiency as a capacity planner's focus of activity. This wasn't an insignificant task as the objective of consistent performance was often at odds with the unpredictable demand profile that the Internet could generate.

As more and more services were digitized this complexity grew, and the role became more about the plumbing. Horizontal scaling offered new system design options but first you had to manage the pipe, and tool up to identify the bottlenecks. Capacity management became more about the day to day monitoring and management of threshold alerts and less about management of scarce resources.

Scheduling was also becoming more of an issue. The traditional transactional daytime period and night time batch were long gone, fueled by consumer expectation of IT services available all day every day. The challenge all the more difficult as mobile technology added yet another layer of complexity as that expectation became all day, every day, everywhere.

Capacity planning was starting to resemble a giant game of Tetris, identifying workloads that could co-exist on the same platform to squeeze as much value as possible from an organizations IT investment.

Capacity planning is dead…Long live capacity planning.

In the highly automated virtualized world of today it is questionable what the role the capacity planner now needs to play—or if indeed there is a role to play.

The 3rd Age—Virtualization

Containment and control

With companies expanding and increasing their global reach the concept of an "on-line day" was no more. High availability was now the norm and thanks to Moore's and Kryder's laws the improved hardware was there to be exploited. Virtualization technologies and "the cloud" meant that, in theory at least, computational power was limitless.

As Agile and then DevOps came to be recognized as mature and alternative approaches to the waterfall method of software development, the pace and speed of change increased significantly.

Code is being pushed into production faster than ever without any real appreciation for resource management. With the ability to use software to mimic hardware components, operations teams could be bypassed in the name of speed. And speed is essential in the digital market place as it provides first mover advantage in a "land & expand" strategy to seize market share.

We have come full circle. Compute is now relatively cheap and people costs represent a significant part of a business operation.

With the software defined datacenter turning everything into a virtual world, capable of supporting DevOps continual deployment methodology, ITIL based processes are viewed as standing in the way of progress. The new industry norm of a multiple service provider model, and the overhead of the formal request processes they bring with them viewed as yet another deadweight for development teams to carry as they race towards implementation. Far quicker to hard code an IP address than ask the outsourced network team to provide one.

And it's going to get worse. Massive investment from Goldman Sachs, Northern Trust and Sequoia Capital amongst others, in containerization technologies such as Docker have accelerated partnerships with firms like Red Hat, Microsoft and Amazon Web Services. Maturing this technology faster than anything else that has gone before it.

As developers, seeking greater flexibility and less reliance on operations teams to provide the environments they need, move from virtual machines to the more granular units that containerization technologies offer, it is the operations teams that are left trying to make sense of the complexity while carrying the associated cost.

Strong controls are needed to control Dev's ability to spin up additional compute on a whim, and ensure that they don't forget to tidy up after themselves. Instances need to be tagged as part of build to establish ownership, usage and effective financial management through attribution or allocation of cost.

The alternative is virtual sprawl, the inability to effectively manage the sheer number of virtual machines and the additional support burden, security, license and compliance issues it brings with it.

It used to be said the best developers carry pagers. In the world of elastic compute capacity I'd proffer that the best developers carry cheque books too.

And yet…why worry?

Cloud based technology comes with the promise of taking care of all your operational needs.

It is after all highly available(?[1]), resilient, scalable and by and large cost effective. Does capacity management have a part to play in this environment? You could be forgiven for thinking that the role of capacity planner has run its course and has no place in the highly automated virtual world.

It's not that simple, consider Elastics Hosts recently unveiled SPRING.IO for containers offering. Billed as true "on demand" compute, pay only for what you use, it seems we have finally reached the promised land of IT as a commodity, no different to gas or electricity. But read the small print.

While they talk about billing based on actual use rather than on the amount of provisioned capacity they also point out that each instance will never drop below a quarter of the maximum set by the user. This is to guarantee capacity and protect the service from spikes. Comforting, but they will charge you for the privilege. And heed the warning of Congo[2]. When you get it wrong or when code runs amok, auto-scaling may leave you with a hefty bill that needs to be paid.

Time to tool up

There is still a very definite need for capacity management, however capacity management needs to evolve with the technology it needs to manage. From being about management and efficiency, through orchestration and performance it is now about containment and control. And the tools available to the capacity planner need to be able to support this.

In a world of artificial intelligence, nano sized computers and virtual everything the tools of the past won't cut it. To stand half a chance in this environment you need to harness the power of the very thing that needs to be controlled.

The size and scale of the IT estate means that automation is essential for data collection, validation and cleansing. Structuring it so that it can be mined to provide meaningful insight. Visualization is a necessity to turn the mass of data generated into readily accessible information. Predictive analytics a must if you are to identify meaningful trends and patterns that allow preventative action to be taken in a controlled and planned way. All of these elements are required to manage the modern IT environment effectively.

[1] http://uk.pcmag.com/internet-products/67162/news/aws-outage-takes-down-netflix-pinterest

[2] http://dcsuk.info/news_full.php?id=38768&title=Forecast:-Cloudy,-with-growing-demand-for-insight

Fully equipped the capacity planner is key to helping shape the decisions that businesses can derive true benefit from.

Not just the nickel and dime tactical…

- Are you running on the correct instance type?
- Do you have sufficient reserve compute?
- Are there redundant workloads to be removed?
- Are there required workloads that can be optimized?
- How do you prevent the sprawl of virtual machines?

…but more importantly, the 6 million dollar strategic.

- When should you convert costly on-demand to lower cost reserved compute?
- When is it time to move to the next family of hardware?
- Where and when is it right to "cloudburst"?
- Would the operation be more cost effective if it was provisioned in-house?
- What model best meets business demand, in-house, hybrid or cloud?

The benefits of this approach for an organization are not confined to capacity planning…

- Incident management, reducing the number of incidents.
- Problem management improves as detailed performance information becomes readily available.
- Change Management can easily and readily assess the impact on existing workloads confirming whether or not adequate headroom is available.
- Release Management benefits by knowing which areas are capable of accepting a planned distribution
- When augmented with cost information Financial Management gains useful insight into the true cost of provision and can replace rudimentary attribution models with greater accuracy for charge back mechanisms.
- Disaster recovery plans can be easily re-checked based on current consumption levels
- Development activity can be better measured and accommodated reducing the risk of a failed and costly implementation.

All of which leads to greater levels of availability, thereby increasing Service Management Teams' ability to not only meet, but exceed agreed Service Levels.

In conclusion

By maturing these processes and moving service from a reactive activity to a proactive one, business areas start to recognize the value IT can add to the organizations strategic decision making process and actively seek their input. IT departments are then viewed as less of a cost base and burden that needs to be reduced and more of an asset to an organization. One they are willing to invest in.

The route to proactive service management requires 3 steps:

1. Establish a Baseline. The first step in planning future capacity is gaining a clear understanding of what your infrastructure looks like today, a holistic single view of your current estate's utilization and headroom for growth.

2. Deploy predictive analytics to identify trends and provide a forward view of anticipated service threats. This allows issues to be avoided in a suitable timeframe with less impact on people resource.

3. Using your Baseline as a starting point model 'what if' scenarios to see the impact of your actions. With the speed of delivery and change present this needs to support rapid decision making and be intuitive, as simple as dragging and dropping workloads and resources around your estate.

More information

To find out more about Sumerian Forward Thinking® predictive capacity planning and using Sumerian Capacity Planner CPaaS, just give us a call on 0131 226 9300, drop an email to sarah@sumerian.com or visit our website at www.sumerian.com

sumerian®
Forward Thinking