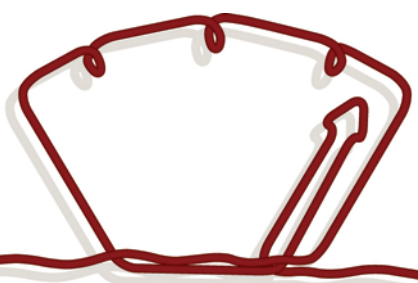
Capacity planning

Achieving business-aligned capacity planning with analytics

Implemented well, capacity planning can significantly improve IT-business aligment, reduce risk – and, in today's tough budgetary climates, keeps costs in check by "sweating the assets."



Introduction

Although capacity management should be viewed as one of the most important IT planning processes to get right, many organisations find it difficult to implement a strategy that focuses adequately on the needs of the business. Implemented well, capacity planning can significantly improve IT-business alignment through an improved understanding of business processes and behaviour, reduce risk by establishing a planning horizon in which to provision adequately - and, in today's tough budgetary climates, keep costs in check by "sweating the assets" and re-deploying under-used infrastructure elsewhere. So why do so many IT organisations find it difficult to implement?

The need for smarter capacity planning

When it comes to capacity planning today's IT teams are faced with a challenging balancing act: that of delivering IT services to meet current and future demand, but not falling into the trap of over-provisioning and wasting scarce budget. However, traditional capacity planning techniques can often be one of the limiting factors; typically driven by a bottom-up approach, they have a tendency of being too siloed and crude. For example, different teams across the IT organisation might manage capacity discretely within their own particular domain or application field, basing their forecasts purely on historical performance and projecting out when they believe it will hit a bottleneck. Although this approach might satisfy short-term capacity needs, its disconnection from the business, as well as other parts of the infrastructure, introduces inefficiency and risk. Instead, what's needed is a top-down approach that first considers the way the business is consuming IT (both currently and factoring in its plans for the future), with the IT organisation then translating these needs into its services and infrastructure provisioning.

Managing capacity for the business

Much has been written about the disjointed planning processes between IT and the business. In the remit of capacity planning, failure to bridge this void is particularly problematic, fuelling inefficiency and risk on both sides of the fence. To address this disconnect, IT organisations first need to improve their visibility into the business functions and processes they support, by building a complete top-down picture that correlates business processes, transactions and demand with IT service utilisation and performance. From this view, quantified baselines of current headroom can be reliably determined, enabling IT organisations to improve their decision making and proactively manage capacity provisioning with precise forecasts.

Key terms and definitions

- **Capacity** – the amount of work that your IT services can do at a given level of performance
- **Load/Peak Load** – how much business demand is placed on your services and what the highest demand is
- **Performance** – is how well the service is fulfilling the load. For example, the number of business transactions completed in a certain time frame
- **Utilisation** – is how much capacity is being used at any given time
- **Headroom** – how much space is left available for growth

This improved visibility not only reduces risk and inefficiency, but enables IT teams to significantly improve the dialogue they have with their business counterparts – by presenting fact-based evidence on how IT is fulfilling demand for key business processes - and, importantly, how further investment can drive even greater value. Such evidence is now possible with the advent of sophisticated data mining and analytical techniques. And it is with these advances in mind that analytics is providing the necessary degree of accuracy to support business-aligned capacity planning.

How analytics helps

Analytics is helping a growing number of enterprise IT organisations achieve a new level of capacity planning capability. Unlike traditional approaches that only provide simplified infrastructure utilisation forecasts, analytics brings a powerful combination of analytical techniques and expertise to bear for truly business-aligned capacity planning.

By capturing and combining data from across the IT estate and mapping it to key business processes and transactions, analytics provides a "joined up" model of the current working environment (see Fig.1) exposing hidden correlations that indicate where changes on both the IT and business side of the equation, will bring about the most returns from a performance, capacity and cost perspective.

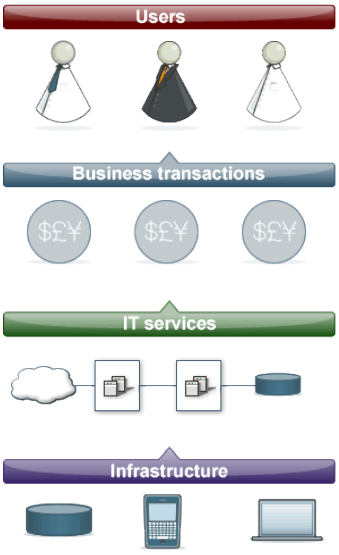


Fig. 1 – IT Analytics' top down approach – 4 levels of correlationand visbility

Copyright © Sumerian

Creating a baseline model

The first step in the analytics approach is to create a baseline of the IT estate, establishing the current end-to-end performance and capacity of each service within an organisation's IT portfolio. This is achieved by taking a service-by-service modelling approach, capturing samples of performance and utilisation data from a service's underlying system components, such as server CPU, physical and virtual memory, I/O, network latency and disk storage used. This baseline is then tracked over time to ascertain the service's utilisation by hour, day, week, and month – capturing average as well as peak usage or busy hour highest demand), along with any seasonality trends (see Fig. 2). The peak hour load is a particularly crucial measurement as it represents the highest identified utilisation of a service – a key metric for planning capacity headroom and forecasts.

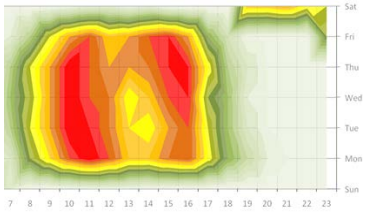


Fig. 2 – Sumerian heat map visualisation illustrating CPU utilisation

The second part of the baseline process profiles the business load placed on the service, and captures business process metrics such as transactions types, volumes, durations for users – again, across regular time intervals, and accounting for trends and seasonality. Finer granularity in the business metrics can also reveal some interesting findings. For example, some business transactions will have minimal processing requirements and make relatively small demands on load, whereas some might have several queuing stages to their processing and require far greater bandwidth and capacity.

Understanding behaviour

Using these baseline models, the effects of business load on IT service utilisation can then be data mined and analysed. In particular, this stage examines the correlation between theprocessing of a business transaction and the way system components interact with it and one another. For example, by applying correlation and linear regression analysis, analytics can extrapolate how memory and CPU utilisation are impacted by business volumes, or how database transactions are consuming disk storage. If a strong correlation is found (see Fig. 3), it produces a quantified relationship that can be used to determine and predict the component or metric's behaviour under different variances – the vital key to accurate scenario modelling and forecasting.

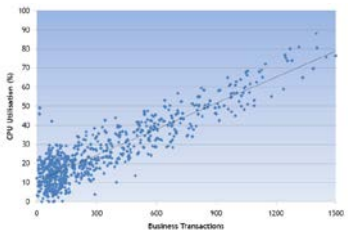


Fig. 3 – finding the correlation between business transactions and CPU

Case study – Sumerian helps investment arm of UK bank to align capacity to trading types and "sweat the assets"

The challenge

The investment arm of large UK bank with complex trading applications across multiple asset classes was undergoing a large-scale merger and needed to understand the impact of increasing business volumes on its IT infrastructure.

What we did

We captured utilisation data for key platforms: application servers, databases, batch and storage, along with business volume data from trading databases. We then correlated the incoming business volumes with utilisation data to create a footprint model for each type of transaction on all infrastructure components.

The value delivered

We established the capacity limitations of the service across all components, enabling the bank to provision resources accurately and "sweat the assets". These improvements meant that a 100% increase in trade volumes could be supported, backed up by quantifying the minimum investment required.

Calculating current headroom

The outputs from the two previous stages provide the necessary elements in which to calculate the current available capacity headroom and, in conjunction, ascertain if capacity has, or is at risk of, breaching saturation limits. To enable visibility and clear understanding of headroom, Sumerian uses Capacity Flowpipe visualisations (see Fig. 4) to help both IT teams and the business units understand their services' current capacity utilisation and headroom. If any areas are breaching system component limits, the direct correlations formed between infrastructure components will determine what impact this is currently having on business demand and the backlog it is generating.

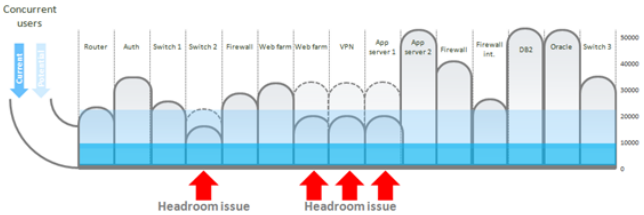


Fig. 4 – Sumerian visualisation for service capacity and headroom

Scenario modelling and forecasting

From understanding current capacity and headroom, sophisticated scenario modelling techniques can then be applied to gain an even deeper level of strategic intelligence for forecasting growth requirements or making predictions around the impact of architectural changes and upgrades. The advantage of scenario modelling is that it can safely answer any number of "what if?"

Copyright © Sumerian

questions around capacity – essentially providing a “pre-sandbox” environment in which to test out ideas before they’re green-lighted in reality.

For example, if delays are being experienced in a batch job, scenario modelling can predict the impact on performance, and capacity if the job is parallelised, or processing power is increased. Or in the case of predicting the effects of a 20% increase in users for an application that has a particularly complex processing queue, scenario modelling can predict the impact of growth on utilisation over the next few months so that IT can plan adequate capacity for it, as well as open up a dialogue between IT and the business to make improvements to the queue’s processing chain. For architectural changes, scenario modelling can, for example, answer whether a service could scale horizontally and quantify whether more servers can be added to the architecture and what the likely improvements will be - or, if the application can’t scale horizontally, model how increasing the capacity of the hardware that’s already there could be achieved.

Keeping page with change

The continual monitoring of service utilisation and business demand is vitally important if capacity is to be successfully managed on an ongoing basis. Business behaviour and demand will fluctuate over time, and even with the most precise forecasts to hand, there will be unforeseen changes that can’t be predicted. Therefore, keeping a running view of each service’s performance and capacity against demand is critical.

Analytics’ ongoing data capture of these key metrics enables the capacity planning position to be constantly tracked, highlighting trends and indicating where changes will bring about the most business benefit. If capacity planning is to be genuinely successful, it should be based on real metrics correlated between both IT and the business. Using the advanced capacity planning techniques offered by analytics gives both the IT organisation and the business powerful intelligence that enables a truly optimised set of services that maximise the value of IT now and for the future.

Case study – Sumerian helps European investment bank identify optimal investment strategy and de-risk change

The challenge

A European investment bank with highly complex trading environment (2000 nodes/3 grids) was seeking to increase its market share by doubling the volume of trade portfolios processed.

What we did

We captured utilisation data from existing platforms and determined the aggregated compute demand. We then created a queuing model for portfolio flows to determine the relationship between flow rate, task duration and resource usage. By applying change analytics scenario modelling, we could then quantify the impact of change on these factors.

The value delivered

We identified improvements to increase processing capacity by 40% with no additional hardware investment. In addition, we recommended the optimal investment strategy to further increase capacity to 80% above current levels.

More information

For further information on Sumerian or to arrange a demonstration of our services, contact us on 0141 229 7580, e-mail us at **info@sumerian.com** or visit our Web site at **www.sumerian.com**

Copyright © Sumerian