

Introducing a Capacity Management Maturity Model

Business units are demanding more services and greater reliability from IT, while also trying to constrain, or even reduce, budgets. In those rare cases where budget is not a constraint, power, cooling, floor space or administration soon become a limiting factor to installation of new equipment. As greater demands are placed on IT for efficiency and productivity, it is becoming more important for IT organizations to adopt more sophisticated methods; in other words, IT management processes must become more “mature.”



About the Author

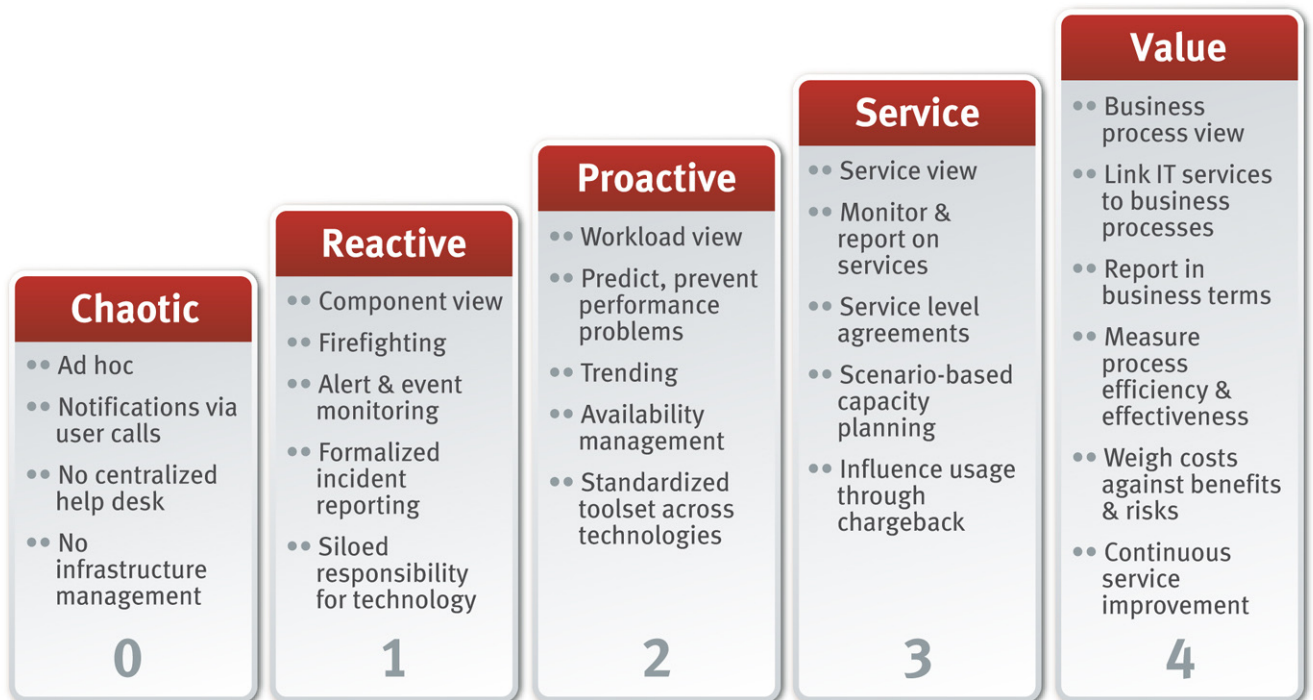
Per Bauer is the Technical Account Manager at TeamQuest Corporation EMEA. Per supports key customers across Europe, Middle East and Africa with technical assessments and managing large implementation projects. Prior to joining TeamQuest, Per worked for a global ERP-vendor for 10 years, holding a position as manager for worldwide QA.

The recent adoption of virtualization technologies lowers the amount of hardware required and reduces power consumption to some extent. But it also raises management complexity since there is no longer a permanent and exclusive relationship between physical resources and the software that runs on it. If you then choose to implement a Service Oriented Architecture (SOA) on top of that, the increase in complexity multiplies.

Bridging the gap between service demands and IT capabilities requires a higher level of performance management for IT resources. Moving to mature Capacity Management tools and processes lowers costs, improves service quality and increases IT productivity since staff can focus on the most important duties rather than fighting fires.

Implementing Capacity Management, however, is an evolutionary process. To help guide businesses along the path, TeamQuest has developed a five-level Capacity Management Maturity Model. The scale is based on an IT Management Process Maturity Model described by Gartner. Interested readers are encouraged to check out other predecessors as well, such as Carnegie Mellon’s Software Engineering Institute’s Capability Maturity Model (CMM) and Phil Cosby’s Quality Management Maturity Grid.

In this paper we describe each of five levels of Capacity Management maturity, how organizations can determine what level they are at and what is needed to move to the next higher level.



Level 0 — Chaotic

If you're at this level you probably...

- Have no centralized service desk;
- Don't predict any kind of incidents (they all come as surprises); and
- React to events in an ad hoc manner.

The most primitive level – Chaotic – is characterized by an overall lack of operations management discipline. Rather than trying to operate a step ahead and acting before users are affected, the IT organization only learns of performance problems when users call in to complain. At the time of such events, performance tools and techniques are assembled ad hoc. Typically, the people involved in the troubleshooting only have access to snapshots of the most recent activities and lack information about the period that led up to the incident.

To make matters worse, the organization often lacks a centralized service desk for reporting of incidents and end users don't necessarily know where to turn. Without a service desk, there's no focal point for coordinating the problem-solving process with various technical teams.

Companies operating at this level of maturity do not have a clearly organized approach to solving performance- and capacity-related problems. They do not have any standard or uniformity as to what toolsets and processes to use; it's all run on a "best effort" basis. If they get something right, it is more by luck than by analysis.

Level 1 — Reactive

If you're at this level, you probably...

- Have rather detailed information on how components are performing.

However, you also probably...

- Still lack an overview;
- Have a hard time planning since you're busy reacting to events outside your control; and
- Misdirect your efforts sometimes, focusing too much on less important incidents.

Companies operating at this level are more mature than those at the Chaotic level, but tend to have a fragmented view of what is taking place within their environment. Typically, they have an assembly of different tools to monitor the activities of different pieces of hardware, applications or services. These tools are not very well integrated and will only give an isolated view of a particular component. These silos of information do not correlate to one another and do not provide a comprehensive view of the full spread of a service or application offered to customers.

When a situation is discovered, whether through an alert from a monitoring tool or through a customer complaint, procedures are in place as to who is responsible for resolving the issue. The tools sometimes allow the IT department to respond to problems before getting customer complaints, but the response is based on information at the component level. They do not identify what services are being impacted or how it affects the business. IT, therefore, can misdirect its efforts by addressing problems that are not necessarily important to the business.

This reactive approach worked better some years ago, when there was less complexity in the IT environment: only a few tiers; simple applications and services. Since there were only a few easily defined technical silos, it was easier to localize the problem. Now, with SOA and virtualization adding new levels of abstraction to the technology, it is misguided to think of technology in terms of distinct silos or even tiers. Most components are thoroughly interconnected and interdependent. That increase in complexity calls for a more mature approach to capacity management.

Level 2 — Proactive

If you're at this level, you probably...

- Focus on the workload analysis rather than the performance of technical components; and
- Rely on trending to give you early warnings regarding future incidents.

However, you also probably...

- Have a hard time predicting the outcome of complex scenarios with enough confidence.

Organizations in lower maturity levels are triggered by events, without putting much effort into understanding the cyclicity or seasonality of applications. At this level you are taking a more forward-looking view to try to predict and avoid those problems. Tools are in place for quicker diagnosis of performance problems and trending is used to predict performance bottlenecks or incidents in the near future. When detected, work is commissioned to mitigate the risk. This enables you to predict and avoid downtime in many cases.

In order to do trending you need historical information, not just a view of current performance or during the last few days. The tools used must be able to continuously collect and aggregate data and put it into a historical database. Another key attribute of this level is the focus on workloads instead of just the compound activity of a technical component. Rather than looking at infrastructure components as entities, the activity of a component is broken down by application or service. These streams of activity are then the focus for analysis and reporting.

Combining the trending capabilities with the notion of different workloads makes for a new level of proactivity. This is the first stage where an organization truly automates the prevention and resolution of problems that are now known to recur. This is where the IT department starts caring about the availability of applications, not just equipment.

To move to this stage, you need to see the full spectrum of an application, not just one layer or one tier of a multi-tiered application. This requires the use of a standardized toolset that covers the complete technology stack of a server, as well as the diversity of different platforms when looking across numerous servers.

In order to keep it all together, this stage requires a stricter approach to monitoring. If you are stuck with a continuous flood of unrelated alerts and events from a multitude of sources, it's hard to prioritize and optimize the effort. You need to identify the key factors beforehand and monitor, store and analyze that subset of information. Of course, one still might need to put out any fires that occur. But if you already know where the fires are likely to occur and which fires will hurt the most, you can prioritize actions to minimize damage.

Level 3 – Service

If you're at this level, you probably...

- Compile workload information across tiers to analyze and report on services;
- Have established a dialogue with the business, trying to anticipate changes; and
- Use analytical modeling to predict the outcome of those scenarios.

However, you also probably...

- Are not accustomed to reporting IT results in business productivity or financial terms.

The Service level is a continuation of the Proactive level, but with a stronger focus on workloads representing services. The view is extended to cover the full spectrum of workloads that make up a particular service, from the end user to the backend systems. By automatically gathering the information needed from each of the components that comprise the service, one can detect, prioritize and execute those actions that will improve end user experience.

To achieve this, simple trending is not enough to predict future needs. When looking at the complex environment needed to deliver a service, it is necessary to have more advanced tools that predict the effects of future business changes.

Analytical modeling offers a simple yet powerful way to address different scenarios. Based on the empirical data collected during the monitoring of systems, models of the systems are created. These models can then be used to accurately predict the impact of growth, changes to the infrastructure, migration of workloads, etc.

Reaching this level takes more than a focus on technology. You also need to establish processes for exchanging information. To identify and verify prediction scenarios, there must be a two-way communication between the business and IT. If the business wants IT to provide a certain level of service without excessive over-provisioning, it needs to supply IT with growth projections and business plans. On the other hand, to demonstrate the value of the information received, IT needs to share cost-benefit analyses of this planning with the business. Building this mutual trust is integral to success.

At this level, since IT knows the characteristics and importance of different services, they can start to optimize the usage of them by implementing chargeback mechanisms where different parties pay for the actual use of resources. This will yield not only a fair distribution of cost, but also a way for IT to influence the usage patterns. Offering a discount during low activity hours might smooth out the peaks and troughs often seen in a data center.

Level 4 – Value

If you're at this level, you probably...

- Measure and report on IT in business terms;
- Have a set of tools and processes enabling you to align IT with the business; and
- Are one of the very few organizations that have reached this level of maturity.

This is the final endpoint of Capacity Management maturity. At this point, there are no additional technical steps to take. Rather it is a matter of taking advantage of the earlier work to fully integrate IT into the business. Everything is looked at from a business point of view, rather than a technical standpoint. The focus is on actions that will benefit the company as a whole.

At this level, the organization realizes the full value, not just of implementing Capacity Management, but of IT as a whole. Since IT's actions are fully and seamlessly aligned with those of the business, they are directed toward helping the business achieve its goals.

IT and business performance are tied together. In addition to IT services, business processes are measured and audited for efficiency and effectiveness. It is possible for IT and business units to accurately determine the cost of IT services and weigh those costs against business benefits and risks in order to make the best decisions regarding the use of IT.

A central process in organizations on this level is Continual Service Improvement. Even though the highest level of maturity has been reached, there is a continuous need to realign IT processes to the changing business needs.

Conclusion

The ambition of this paper has been to describe a generic model for Capacity Management maturity. Most businesses today are in either the Reactive or Proactive stages. To take the next step and reach the Service stage, homogeneous monitoring across heterogeneous platforms and technologies in combination with predictive modeling becomes crucial. This also prepares you for the ultimate level, the Value stage.

It should be kept in mind however, that gaining maturity is a continuous process and not just about the tools. It's neither a matter of black or white. For some less critical services, it might make sense to leave them at the Reactive level and fix problems as they arrive. Giving them the full proactive treatment is probably a waste of resources. And, no matter how proactive you are, there will always be unexpected problems that call for an immediate reaction. But the key business processes should all be operating at the highest possible level, so that the organization is getting the maximum competitive advantage out of its IT investments.

TeamQuest Corporation

www.teamquest.com

Follow the TeamQuest Community at:

Americas

One TeamQuest Way
Clear Lake, IA 50428
USA
+1 641.357.2700
+1 800.551.8326
info@teamquest.com

Europe, Middle East and Africa

Box 1125
405 23 Gothenburg
Sweden
+46 (0)31 80 95 00
United Kingdom
+44 (0)1865 338031
Germany
+49 (0)69 6 77 33 466
emea@teamquest.com

Asia Pacific

Units 1001-4 10/F
China Merchants Bldg
152-155 Connaught Rd Central
Hong Kong, SAR
+852 3571-9950
asiapacific@teamquest.com

**Copyright ©2010 TeamQuest Corporation
All Rights Reserved**

TeamQuest and the TeamQuest logo are registered trademarks in the US, EU, and elsewhere. All other trademarks and service marks are the property of their respective owners. No use of a third-party mark is to be construed to mean such mark's owner endorses TeamQuest products or services.

The names, places and/or events used in this publication are purely fictitious and are not intended to correspond to any real individual, group, company or event. Any similarity or likeness to any real individual, company or event is purely coincidental and unintentional.

NO WARRANTIES OF ANY NATURE ARE EXTENDED BY THE DOCUMENT. Any product and related material disclosed herein are only furnished pursuant and subject to the terms and conditions of a license agreement. The only warranties made, remedies given, and liability accepted by TeamQuest, if any, with respect to the products described in this document are set forth in such license agreement. TeamQuest cannot accept any financial or other responsibility that may be the result of your use of the information in this document or software material, including direct, indirect, special, or consequential damages.

You should be very careful to ensure that the use of this information and/or software material complies with the laws, rules, and regulations of the jurisdictions with respect to which it is used.

The information contained herein is subject to change without notice. Revisions may be issued to advise of such changes and/or additions. U.S. Government Rights. All documents, product and related material provided to the U.S. Government are provided and delivered subject to the commercial license rights and restrictions described in the governing license agreement. All rights not expressly granted therein are reserved.