IT SERVICE OPTIMIZATION
AN INTRODUCTION
SUMMARY

IT Service Optimization (ITSO) is aimed at optimizing the planning and delivery of IT services. The goal is to consistently meet IT service levels while minimizing infrastructure costs and mitigating risks. This increased efficiency makes it easier to manage IT resources, adapt to business changes, and align with business objectives.

This paper introduces ITSO and discusses tools and processes that can improve the planning and delivery of IT services.
What is ITSO?

IT Service Optimization (ITSO): Consistently meet IT service levels while minimizing infrastructure costs and mitigating risks

ITSO outlines a five-step, iterative process that leads to optimized IT service performance, improving both the operational efficiency of IT infrastructure and the proactive planning for IT services that align with business objectives. Without performing these steps, IT will not realize its full value to the business.

ITSO Five-step Iterative Process

The first four steps of ITSO deal with the planning and provisioning of IT services. These are important, yet sometimes overlooked, steps to identify business priorities so that IT can align with them and understand which services are most important. The more important the service, the less acceptable is the risk, which is a consideration when mapping IT resources to services.

Pre-deployment scalability testing and provisioning predicts service performance under varying demands, identifies optimal hardware configurations to support the service, and also controls costs through accurate provisioning.

Step Five deals with managing the ongoing performance of IT services once they are launched. Performance management tools and techniques ensure that service levels are met and reported on an ongoing basis. Services are monitored and problems are detected and resolved before they impact users.

Lastly, a circle back to the beginning of the ITSO process addresses the need for ongoing performance and capacity management for IT services, which is critical for mitigating service delivery risks. It is important to re-visit business priority, service level objectives and demand fluctuations over time, so IT resources can be allocated appropriately and service levels can be maintained. Adjustments for changes in business goals and service demand must be made to support each service with an appropriate level of risk.

This five-step, iterative process helps IT organizations increase efficiencies, minimize waste and mitigate risks, to maximize the business value of IT.

The ITSO Process

1. Understand Business Objectives
2. Prioritize Services & Assess Risk Levels
3. Establish Services Levels
4. Plan & Provision Services
5. Manage Service Performance

Step One: Understand Business Objectives

The first step in creating a service is to fully understand the business requirements. What is that the business is trying to accomplish? To answer that, IT professionals require a business perspective and knowledge of the strategic goals of the organization. Sometimes those goals have yet to be documented, making the job more difficult, but no less necessary.

Understanding the priority of business goals enables IT to align with those priorities and determine where time and resources are best focused.

Taking into account the business goals of the organization, it is important to review processes to be sure that they accomplish meaningful work in a sensible and efficient manner. This analysis should be conducted prior to automating a process. In other words, it does no good to automate an inefficient process. For example, a major computer vendor implemented a Sales Force Automation (SFA) system. After spending $8 million on implementation, the company saw a decline in sales. It turns out they had automated a bad process and were making mistakes more efficiently!

Step Two: Prioritize Services and Assess Risk Levels

It is important to understand the relative importance of different IT services and prioritize them so that resources are allocated appropriately. Determine which services are aligned with critical business objectives. Such business-critical services should receive more attention and more planning to minimize risks and ensure consistent delivery of those services.

Prioritizing IT services helps IT focus attention and resources where it is most needed in order to generate business value.

If you like what you have read, click here for a FREE SUBSCRIPTION to the TeamQuest Business Intelligence Series.
**Step Three: Establish Service Levels**

Service level requirements should be set to ensure that the business goals underlying IT services are met. What level of availability is needed in order to ensure an efficient operation? What are response time or throughput requirements?

Service level requirements are generally documented in a service catalog, which contains definitions for every service provided for the business by the IT organization. Sometimes a Service Level Agreement (SLA) is also negotiated and agreed upon by both the IT organization and the business unit, but isn’t necessary. Required service levels should be achievable and not so detailed as to introduce unnecessary overhead and complexity to the management of the IT service. A good service definition aids communication with business unit clients and ensures that expectations are in line with reality. Ideally, required service levels are expressed in business terms meaningful to clients.

**Step Four: Plan and Provision Services**

Applications must do more than simply fulfill functional requirements. Performance and scaling requirements are just as important. Such requirements should be taken into account as early as possible during the product evaluation or development process. A variety of development techniques, tools, and profilers can be used to tune applications during the development process, mitigating the risk that new applications will fall short when a production-level workload is applied.

The process for preparing critical applications for production should include steps for determining the optimal configuration for systems that will host the new applications, taking data center architectural policies into account. Especially when rolling out new applications, a load-testing tool to simulate transactions coming from end users or other systems can be used to benchmark applications on test systems to be certain they can support required service levels.

For most systems, time and cost considerations make it prohibitive to conduct such tests on the actual hardware using production-level workloads. Full-sized machines and software are too costly to dedicate for testing purposes, and finding the optimal configuration through empirical testing can take a long time. Instead, smaller, but representative loads can be applied to a scaled-down set of test servers and software while performance analysis software takes a baseline reading of performance. Analytical modeling (see “Capacity Planning Methods,” below) can then be used to rapidly predict how various configurations will perform under a scaled-up production-level workload, all without the need to purchase the actual configurations under consideration. Analytical modeling is a cost-effective, fast and accurate alternative to brute-force testing.

**Step Five: Manage Service Performance**

As a follow-up to establishing service level requirements, it is necessary to monitor and report how well IT is meeting those requirements. Monitoring allows IT to react when problems are threatening to impact the business, and also for client business units to know whether they are getting their money’s worth from IT services. Ideally, proactive performance management practices such as capacity planning are used to ensure problems are addressed before service level requirements are jeopardized.

Performance issues can never be avoided completely, because circumstances are constantly changing. Business plans, forecasts and technologies change, and unpredicted events occur. That is why IT Service Optimization, like Total Quality Management, Six Sigma, or the Service Level Management sub-process of ITIL, calls for continuous analysis and adjustment. Problems are continually detected and remedied by making adjustments or changes to correct those problems.

**Step Six: There is NO step six!**

However, it is very important to revisit steps one through five routinely for ongoing performance and capacity management; there are many variables that affect services that are in constant flux and can affect IT resource allocation.

As business forecasts, technology and other factors change over time, those changes must be incorporated into revised capacity plans. Otherwise, fire-fighting and inefficiencies will prevail.

It is also important to adjust capacity requirements to be sure that risk levels align with business priorities. If a critical service shows substantial growth in demand, capacity plans should be adjusted to reduce the risk of falling short of service level requirements, allow for unexpected spikes, and ensure uninterrupted quality of service. Services of lesser importance do not require such close attention.

Performing ongoing capacity planning keeps the data center running at high efficiency and minimizes waste. This also allows the data center to adapt more quickly to changes in demand or technology.
**ITSO and ITIL**

Many organizations today are implementing IT Infrastructure Library (ITIL) best practices. ITIL is a framework that provides best practice guidelines for managing and delivering IT services. It offers structured, scalable, common sense processes that organizations should adopt and adapt to fit their own environments.

ITIL is compromised of seven core areas, of which one is Service Delivery. Five processes are identified under Service Delivery: Service Level Management, IT Financial Management, IT Service Continuity Management, Availability Management, and Capacity Management.

ITSO supports ITIL best practices, focusing primarily in the Service Level Management and Capacity Management areas of Service Delivery.

Service Level Management (SLM) is an important ITSO-related process that calls for the use of service definitions to ensure improved communication with client organizations and as a means for determining whether IT services have been acceptably delivered. SLM is a cyclical process of improvement where service definitions are established, IT services are measured, performance is reported, and then adjustments, improvements and corrections are made.

The ITIL Capacity Management process is crucial to IT Service Optimization and includes the monitoring, performance analysis, and capacity planning processes used to ensure that service definitions are fulfilled using a minimum of computing resources. Capacity Management ensures that the overall organization’s goals and priorities are fulfilled by IT without wasting capacity.

**ITSO and BSM**

Business Service Management (BSM) is another term for the complete set of ITIL processes (see “ITIL and ITSO,” above). The often-stated goal of BSM is to allow an organization to manage IT from both the business and IT perspectives.
**ITSO Tools**

Tools that support ITSO processes can be classified into four core areas:

1. Event Monitoring
2. Performance Management
3. Performance Reporting
4. Capacity Planning

---

**Event Monitoring**

The most basic form of Event Monitoring is illustrated when an end user calls to complain about response time, or worse yet, when sales inexplicably plateau as online customers are “encouraged” by bad performance to look elsewhere for the products they seek. Such is the fate of a completely reactive IT organization that has failed to implement the principles of IT Service Optimization.

In reality, even IT organizations with a low level of process maturity have some processes in place to track service levels. At a basic level, availability is monitored and outages are reported by an event monitoring system. This is still a reactive form of management, but at least IT has a chance to react before the damage is extensive.

More advanced event monitoring spots adverse trends and anomalies in performance or uses educated “rules of thumb” as a means to identify impending problems and notify management before service levels suffer. This proactive event monitoring can prevent end users and the business from being adversely impacted.

**Performance Management**

Performance management tools provide a means of collecting performance data for problem detection and analysis. Often, companies use a variety of free, purchased and homegrown performance tools. This makes it difficult to manage the performance of services across multiple tiers or plan capacity requirements for business-critical services. Standardizing on a single, best-of-breed solution is recommended.

Performance Management is comprised of two principal activities: reactive performance analysis and proactive performance analysis. Reactive analysis occurs after an adverse event. Proactive analysis is concerned with preventing such events from affecting the smooth delivery of IT services.

Reactive analysis of an adverse event includes activities such as:

- Correlating performance data to locate the IT component that is the root cause
- Drilling down from an anomalous point on a performance graph to determine who or what is responsible
- Visually comparing current performance graphs and charts with past performance in an effort to understand the cause of an event

Proactive analysis includes activities such as:

- Watching trends and cycles in system performance and projecting future performance levels as compared to service levels specified in the service definition
- Analyzing trends to determine how best to ensure that service levels will be met in the future
- Identifying underutilized capacity for potential redeployment

Proactive analysis is preferable and much more efficient than reactive analysis, since it prevents disruptive events, ensures a smooth-running data center, and eliminates time spent on fighting performance fires. Organizations that spend inordinate amounts of time doing reactive analysis (a.k.a. fire fighting) should make changes to reduce it. If no additional time is available for the proactive analysis needed to prevent problems, then time should be made. One way of making time is by limiting access or changes to the applications and infrastructure with lower business priority or that generate the most problems. Fragile systems should be isolated and protected to give the IT organization time to shore up its performance management efforts.

Any performance management tool will be involved with collecting performance data, and if any historical analysis capabilities are provided, there will necessarily be a performance database. When selecting a performance management tool, it makes sense to understand how the tool implements performance data collection and storage policies, as this can impact the scalability and manageability of the tool. The efficiency of the data collection agents and performance database are also important factors.

**Performance Reporting**

Performance reporting is a crucial function that fits tightly with several ITSO processes. It is important to provide timely reports with the appropriate level of detail to various levels in the organization:

- Exceptions are reported during Event Monitoring
- Analysis reports, tables and graphs are generated during both reactive and proactive performance analysis
- Planning reports and analyses are generated during Capacity Planning

There is a need for both interactive analysis reports and also periodic reports provided to management for tracking performance against service levels. Furthermore, financial reports can demonstrate IT’s contribution to organizational goals, resources consumed by IT, and consumption of IT resources by various departments.
**Capacity Planning**

This is the area where IT organizations can reap the most benefits. Capacity Planning is about predicting performance, allowing services to be provisioned with just the right hardware configuration to accommodate forecasted changes or spikes in business workloads. The judicious use of capacity planning techniques can help prevent performance fires, freeing up human resources that would otherwise be spent troubleshooting.

Capacity planning is truly the key to a well-oiled, smooth-running data center, helping savvy IT organizations to:

- Accurately and efficiently provision new applications
- Reduce, delay and sometimes completely avoid costs
- Justify upgrades with objective analyses
- Optimally configure systems to accommodate forecasted changes in business
- Migrate systems to take advantage of new technology
- Consolidate servers to simplify management and gain economies of scale
- Mitigate risk, by playing out potential scenarios in advance, and then preparing for them.

**Methods for capacity planning include:**

**Trending**

Trending uses simple extrapolation of resource utilization over time. The advantage of this technique is that it does not require sophisticated tools. A spreadsheet will do, but it can be difficult to take into account non-linear behavior buried deep within a multi-tiered system.

**Simulation Modeling**

Simulation, a much more sophisticated technique than trending, actually simulates the queuing events that occur during execution. The downside of simulation modeling is that it can be very time-consuming to build and run the models.

**Analytical Modeling**

Analytical Modeling uses mathematics to calculate how a queuing network will perform. To use analytical modeling, a baseline set of measurements is taken on a system and a model is then built based on a description of the system. Results from the model are compared with the baseline, and when they match, the model is considered calibrated. From there, hypothetical changes can be made to system configuration or business workloads, and the model will predict how the changes will affect performance. Done properly, analytical modeling is both fast and accurate.
ITSO and TeamQuest

As specialists in performance management and capacity planning, TeamQuest has defined IT Service Optimization as a process for meeting IT service levels while minimizing infrastructure costs and mitigating risks. Supporting ITSO, TeamQuest software returns value to the company quickly, increasing IT performance while decreasing costs.

TeamQuest software gathers performance data from all of the IT components comprising a service, providing a comprehensive view of end-to-end performance as a service moves through the tiers. It also supports a variety of Unix, Linux and Windows platforms. This integration across the entire enterprise provides a single point of reference with a unified look and feel, making it easier to measure performance against service levels.

TeamQuest Performance Software is a suite of four integrated products that helps organizations optimize IT services. Used individually or in combination, TeamQuest software scales to thousands of servers in complex heterogeneous and virtualized environments.

TeamQuest Model

TeamQuest Model is a capacity planning package that accurately projects the amount of resources required to support consistent service delivery at appropriate risk levels. TeamQuest Model is used to:

• Accurately provision for services, taking business priority, demand fluctuations, and architectural policies into account
• Determine optimal configuration for services to minimize waste
• Predict if service levels will be maintained as workload increases
• Analyze components of response time

TeamQuest View

TeamQuest View offers powerful performance management as well as reporting capabilities. TeamQuest View can be used to ensure quality service delivery by assisting with early problem detection, in-depth investigation and analysis, and trending capabilities. TeamQuest View is used to:

• Analyze service performance
• Investigate performance issues that affect service levels
• Monitor and report service levels
• Identify potential problems before service levels are affected
• Measure service response time

TeamQuest Alert

TeamQuest Alert monitors activity throughout the data center and watches for problem conditions that could impact services. TeamQuest Alert is used to:

• Monitor services across the entire enterprise and report on events
• Spot and report impending problems before they impact service levels

TeamQuest On the Web

TeamQuest On the Web displays performance reports using a Web browser, making it easy to report on service levels. TeamQuest On the Web is used to:

• Report service levels to management
• Automatically produce and distribute performance reports to business units and IT management

Conclusion

IT Service Optimization is a five-step process for consistently meeting IT service levels while minimizing infrastructure costs and mitigating risks. ITSO calls for the use of a variety of tools and techniques to ensure that IT efficiently achieves business goals, improving both operational efficiency of IT infrastructure and proactive planning for IT services. Performing these steps will help IT realize its full value to the business.

ITSO supports ITIL best practices, focusing primarily in the Service Level Management and Capacity Management areas of Service Delivery. ITSO also supports BSM goals, focusing on areas where optimization is required to achieve intended results.

There is no reason to delay implementation of ITSO. In fact most organizations are already practicing some aspects. Rather than starting from scratch, these organizations can benefit from reviewing their current processes with respect to ITSO and filling in the gaps.

Bibliography
