Perspectives

Optimizing IT Cost
The CIO's Handbook

TATA CONSULTANCY SERVICES
Experience certainty.

Greetings

As our worldwide clients have turned to us for help in this challenging period, it heightened the need for continuous new thinking. In the blink of an eye, financial pain became acute so our customers rapidly did everything that was obvious. The challenge in the past 12 months was moving beyond the obvious and finding new and unexpected ways of making IT cost less, and yet do more to help businesses. This issue of *Perspectives* is dedicated to explaining some of the novel ideas and practices we have discovered.

N. Chandrasekaran
CEO & Managing Director

The theme of this issue of *Perspectives* is practical creativity. This past year, we have learned more than usual as the pressures to provide increased value have naturally led to specific innovations.

This is not a time of massive reconstruction but one of incremental innovation. Each of the techniques discussed in this issue revisits or expands on trends and activities that are already a part of your IT landscape. But many of the ideas such as SOA are not working as expected, or technologies such as cloud computing and virtualization are not delivering as much value as they should. Throughout these articles, you will find pragmatic suggestions to gain greater returns from these activities.

We also look at some of the enduring problems of IT, such as multi-vendor program management or organizational change management, and suggest ways of addressing problem areas that have become more acute in the current environment of scarcity.

In addition, you will find some "aha" moments. I am particularly keen about the idea of using XBRL, usually something that is approached begrudgingly, as a new way to shine light on risk in a company. This is the kind of creative thinking that helps our clients navigate current challenges while gearing up for sustainable growth.

J. Rajagopal
EVP & Head, Global Consulting Practice

CONTENTS

Optimizing IT Cost: The CIO's Handbook

As the business cycle ebbs and flows, the rationale for IT investment tends to extremes. In boom times, IT is asked to support growth and accelerate performance at all costs. But in times of a bust, like the current financial downturn, IT becomes an engine of cost savings and a focus of cost cutting.

At first, IT cost cutting is an urgent matter. Executives ask where spending can be cut right away. When IT is applied to business operations to reduce spending and increase efficiency, the returns must be achieved as quickly as possible.

But after the first wave of slashing is over, a more measured approach takes hold, which is the focus of this issue of *Perspectives*. More than in any other previous downturn, new techniques and approaches offer the creative practitioner a whole host of options for improving IT and business operations.

Some of the options explored in this issue *(fig 1.1)*, like cloud computing, have been overhyped and underanalyzed. The reader will find new analysis of how to put cloud computing to work and how and when to make specific tactical moves, such as moving data to cloud storage. Other areas such as new ways to envision change management practices and the application of XBRL (eXtensible Business Reporting Language) in analytics are innovations that sprang from the vast experience of TCS consultants around the globe. Professor Howard Rubin's ideas about Technology Economies represent a new paradigm for IT management.

Fig.1.1: **Refocusing IT cost optimization**

Social networking, Enterprise 2.0, New age analytics (like XBRL), Remote infrastructure management, SaaS, Service catalog, Cloud storage, Multi-sourcing, Private data center fabric, On-premise storage, ITIL, Cloud development, Organizational change management, SOA, IT costing, Federated identity, IT spend and portfolio, Virtualization, CMMI, ERP / Supply chain optimization, Enterprise architecture, Enterprise scorecarding /analytics

Benefit Realization: Untapped, In pursuit, Saturated

Discipline / Technology Emergence: Nascent, Breaking in, Commoditized

Diameters of the circles reflect the quantum of impact on cost-benefits

The companion illustration shows the number of issues we had to choose from in making our analysis. Our focus was determined by the general applicability of the techniques we analyzed, and their potential for outsized returns in both the short and long term. That said, a quick look at the diagram reveals many worthwhile issues to be addressed at a later date.

As in every issue of *Perspectives*, the findings come from the field. The compelling new technologies and disciplines described, the mechanisms employed, and the pathways to immediate benefits are the products of the creative energy of thousands of consultants seeking to make IT a positive force in these challenging times.
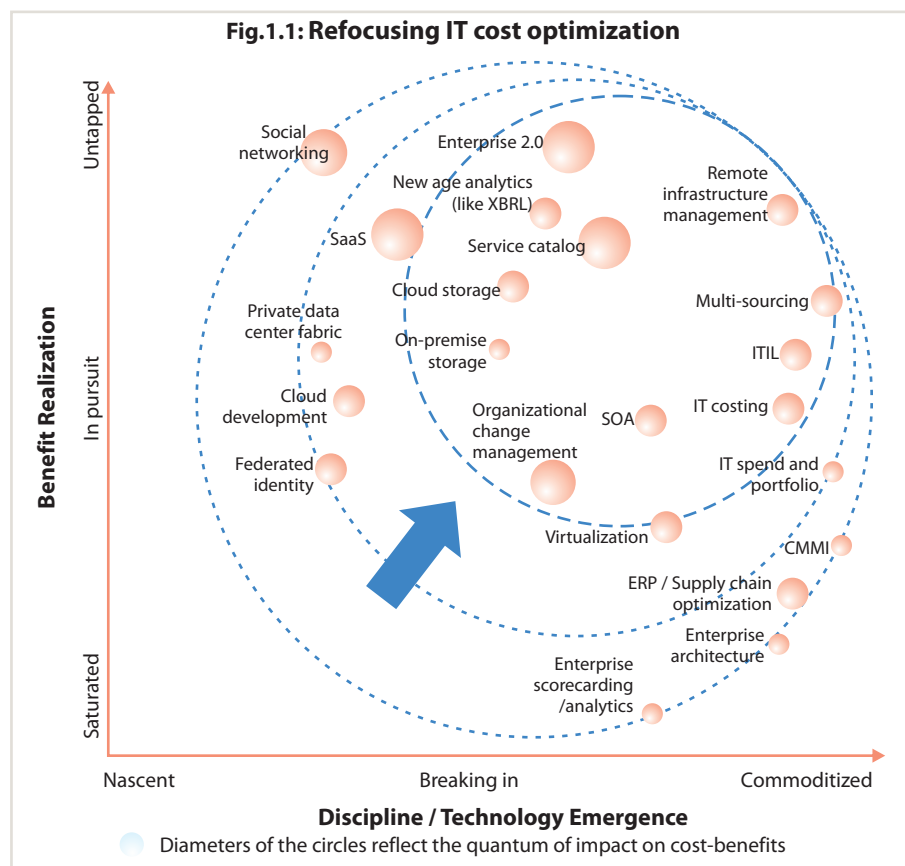
Aligning IT and Its Costs

When the focus is on supporting growth, systems are implemented to hit a moving target. They are designed to support the business that will be, not the one that is. As a result, sometimes such applications of technology may be underutilized or at worst become shelfware. Even when applications are in regular use, it is not clear what data and functionality are the most important. As pressure to allocate IT costs to specific business units intensifies, a deeper understanding of business and IT alignment is required.

The transformations required by new strategies can be far reaching. In the Telecom industry, for example, companies are seeking advertising revenue through application services - a departure from traditional wire-line business. The IT used to provision these services becomes part of the cost of goods sold and must then be accounted for as a variable cost of the products and services.

In the first section of this issue of *Perspectives*, the use of service catalogs and service-based costing are explored as a way to develop a more accurate view of how IT is providing business value to specific types of activities.

Service Management – Aligning IT using the Service Catalog Lens

This article explores how the service catalog approach that was introduced originally by ITIL can provide a unified view of how IT services support business services. By creating and maintaining an extensive service catalog, important aspects of IT can be confidently identified. Areas for investment in improvement or expansion of IT services become much clearer, as do the set of services that are providing the least value. With such an understanding in place, it sets a foundation for charging IT on the business value it provides, which is a major step forward in sustaining alignment and establishing clear priorities for IT portfolio strategy.

Takeaways

- The service catalog is emerging as a way to manage and plan the IT portfolio. It is a unified and self-contained methodology to manage business-IT alignment

- The contents of the service catalog should be determined by analyzing business needs

- The price for services should include the service cost and the service levels agreed on with business units. The service level for a service may vary from one business unit to another, based on individual service level needs

IT Financial Management– Making Chargeback Workable

While mapping the IT services used in a business is a great first step, once the most valuable services are defined, the challenge becomes quantifying their worth. Traditionally, chargebacks have always been performed in terms of IT costs, with the actual prices determined by spreading the direct costs and associated overheads for each service over the departments that utilized the service. But such an approach ignores many important factors and fails to assess the value and the demand of the services provided in business terms. If the end game described in the last article is value-based pricing for services, the goal of this article is to find a way to establish a foundation for such a structure.

This article makes a strong argument for making a deeper analysis of the costs and value provided by services using categories developed from the field. The goal is to find a common understanding of IT cost between business units and IT. Obviously, cost accounting must still be performed, and establishing a notional price based on business value will not be without its complications, but to achieve this vision even in part transforms the relationship between IT and business into one that is far more fact-based, collaborative, and intimate.

Takeaways

- The technology for utility chargeback will take a few years to mature

- The purpose of chargeback is making business units accountable for IT consumption, giving them a price that they can trust

- Costing techniques less rigorous than utility costing can still achieve this purpose as long as they are practical and trusted by business

- Service-based costing helps in achieving plausible estimates for chargeback by following some rules of thumb

Getting More from Operations

The rush to transform both business and the supporting IT infrastructure to accommodate the new economic reality has led to lurches toward programs that promise quick benefits. In many cases, plans for data center virtualization or for off-shoring various kinds of business processes or support functions were hastily accelerated. But after these initial spurts of activity, in most companies it became clear that many sorts of transformations were needed that would not just be quick projects but would represent long-term changes that are gradually achieved.

The articles in this section focus on issues related to the long-term changes needed to improve operations. Some of these efforts, such as SOA, have been underway for a long time and now clearly need to be refocused to succeed. Skills for multi-vendor program management and for change management must also be improved to achieve transformation over the long haul. In three articles, *Perspectives* offers some insights for making SOA practical and improving the basic skills required for long-term change.

Service Oriented Architecture - Quick SOA

SOA has been haunting IT for so long that some analysts have declared it dead. And not dead on arrival but dead after a long stay. But even the analysts making such pronouncements do not dispute the fundamental value of SOA, which shows why it has been pursued so far.

SOA's promise to provide business agility by making processes dictate how applications would be used has always been attractive. But this vision has been held back because legacy applications have been difficult to change, and some architects have attempted to enforce rigid hierarchies.

This article argues that it is time for some compromises to be made in the pursuit of SOA in order to accelerate progress. The limitations of legacy applications should be accepted, and services that can be built should be created from the bottom up rather than waiting to figure out how to create the services that should be built based on a more top-down design. In addition, REST (Representational State Transfer) -based services, which are often easier to create, should be employed, not just those that rely on the more robust WS* standards. The main thrust of the article is that it is time we got started with services however they may be created, rather than accepting delays in pursuit of perfection.

Takeaways

- The traditional services hierarchy (i.e. business services supported by applications, supported in turn by infrastructure services) is not a requisite in designing for SOA, and should not be a constraint

- SOA should maximize reuse of legacy assets

- The focus on reuse forces us to look at a new categorization of services based on the older standards that legacy applications would support

- Modern SOA standards like ones by OASIS (WS*) are important only in core processes and in critical cross-enterprise interfaces like supply chain

- Supporting services may retain legacy interfaces

- Explore REST interfaces to support SOA

Program Management – A United Front: Coordinated Multi-vendor Programs

With globalization creating an ecosystem of diverse specializations, it is common for many different consulting and outsourcing organizations to have to work together on large inter-related projects, often referred to as programs. While the fact that each firm brings expert knowledge in specific areas helps reduce risk, there is growing systemic risk that all of the partners will not effectively work together.

This article asks the question: Why have IT vendors been so slow to find a way to work together in large programs? In the manufacturing industry, cooperation among many firms is the rule, and, has been the engine of new forms of supply chains and distributed design and manufacturing processes. But in the world of IT, close collaboration seems to be the exception. Too often partners operate in silos and are not concerned with the overall success of the client. This article argues for several tactical approaches to information sharing and service levels that bring multi-vendor programs into better alignment.

Takeaways

- Vendors are more willing today to work collaboratively and show program benefits

- Program management has to bring in mechanisms to foster this trend

- Unifying the interfaces with multiple vendors is the first step towards a coordinated multi-vendor program

- Program management should strive towards having a common knowledge management environment. There are ways to achieve this while preserving vendor interests and Intellectual Property (IP)

- Program-level analytics should move from contractual analysis of SLAs to metrics that reflect coordinated benefits and cross-vendor performance

- Adding some risk-reward mechanisms to contracts can help in incentivizing a concerted effort by vendors

Organizational Change Management – Harnessing the Desire for Change

In large IT transformation programs, the staff involved feel that the solutions being adopted are inflicted on them by senior management. In such cases, the people who are most important to the change, the people who will work differently, feel alienated because they have little input. It is no surprise that many such programs fail.

The cynical view is that people just hate change, but a more accurate interpretation is that people hate being pushed into changes that they do not understand and for which few clear benefits are explained.

This article argues that most people are actually open and eager to change if they have a stake in framing the change and can see they will enjoy benefits. In other words, large transformations must be pulled as well as pushed. This chapter explores various tactics to increase the pull for change.

Takeaways

- Most change management initiatives start on the wrong foot by assuming that people resist change. On the contrary, change is welcomed if people are able to relate to it

- In the process of driving change, one should avoid changing the identities and natural behavior of people

- Organizational culture, and how one identifies with it, is an important aspect that should be preserved

- When such principles are followed, change is embraced, not resisted

New Frontiers

Compared to 20 years ago, the men and women playing the CIO role today are suffering from an embarrassment of riches. The CIO is left with too many unexploited options, like enterprise-wide virtualization, cloud, new forms of Business Intelligence (BI) , and so on. Are IT managers up to the task of leveraging these powerful capabilities? Is a new paradigm for IT management required?

By and large, techniques such as virtualization have been applied only in obvious ways to decrease data center costs through server consolidation. Service management techniques have allowed standardized support to be delivered from remotely located centers helping to exploit labor arbitrage, but this is just one form of commodititzation. While these basic options are soon exhausted, pressures to cut costs remain. Many more ways of using virtualization and exploiting commodity prices and cost cutting arbitrage are waiting to be exploited.

It is time we asked the questions: Is this all we can get from virtualization? Does the true potential of virtualization go beyond captive datacenters to shared cloud-based services where virtualization can open the door to different money-saving technology economics?

Professor Howard Rubin, Gartner Fellow interviewed in this edition, suggests that many of the enduring challenges of balancing IT cost cutting and investment can be addressed through sharing IT with peers, and cloud seems to be the promising way to do it.

Is IT also missing opportunities that would soon appear as burdens? For instance, the newly energized appetite for information-based management asks for more investments in sophisticated infrastructure and analytics tools. Can analytics use new standards that make it simpler? XBRL, which has mostly been adopted at gunpoint, could be seen as a blessing, a technique that could bring efficiency to information processing.

Enterprise Cloud Computing – Taking off for the Cloud

This article asks IT to reconsider the meaning of cloud computing. How can the use of the cloud be expanded, not only through Software as a Service, but also in ways that make the best use of the economics of elasticity- a property that emulates infinite capacity? The areas of testing, analytics, and disaster recovery are all analyzed in ways that are sure to suggest many other possibilities.

Takeaway

- Cloud computing provides unprecedented capacity availability due the property of elasticity. Today, developments in cloud computing are boosted by parallel computing technologies

- Enterprises can start off with cloud computing for IT applications for which both economics and flexibility make sense

- Application testing, analytics, and disaster recovery are three areas where the cloud has immediate merits

Enterprise Cloud Computing – Pros and Cons of Cloud Storage

"The Pros and Cons of Cloud Storage" takes a deeper dive into the potential uses of cloud storage. An economic analysis is being used to discover where cloud storage makes the most sense. This article categorizes business use of storage, and then suggests which of the three tiers is most amenable to cost-saving migration to cloud storage. The benefits, risks, and responsibilities are analyzed.

Takeaways

- Storage economics on the cloud is dependent on multiple factors

- It provides added advantages in terms of data integrity and availability

- A rule of thumb is that one should use cloud for tier-2 storage: enterprise data that is active but less frequently used

- Cloud storage also raises some concerns in terms of compliance, as security perceptions evolve

Enterprise Risk Management – XBRL and Real-Time Analytics

"XBRL and Real-Time Analytics" attempts to turn compliance lemons into risk-management lemonade. Regulators have insisted that financial information be reported using the XBRL XML standard. The motivation for XBRL adoption is that once the information is brought into the regulator's data center it will be much easier to consolidate and analyze.

This article argues that the same benefits apply to expanding the use of XBRL within companies to analyze data for risk management. In effect, XBRL allows the ETL (Extract, Transform, Load) function to be bypassed by proper coding of information at the source. When used for risk management data, it means that you can get access to the big picture much faster than through traditional methods, which is exactly what risk management is all about.

Takeaways

- Enterprise Risk Management is striving to be a self-correcting function in business by being more intelligent on emerging risk parameters and sources

- At the same time, a myriad of new compliance regulations are calling for a unified approach to risk information system

- Traditional approaches to analytics consolidate data from disconnected and heterogeneous applications into a datamart, which is expensive

- XBRL can help work around these challenges by gathering information from legacy applications in a different way

- In the process, it reduces application operating costs and makes analytics more timely and intelligent

Tête-à-Tête

Technology Economies and "Change the Business" Investment

The closing chapter of *Perspectives* is an interview with Professor Howard Rubin of the City University of New York, a pioneer in IT management philosophy. Professor Rubin suggests that companies envision their IT infrastructure as a Technology Economy. This avoids the lurching behavior from investment for growth to draconian cost cutting, and replaces it with a continuous process of optimization to maintain balance. Rubin predicts that the fundamental economics of IT will lead to a world of more services that are collaboratively developed and shared across businesses. Open source is one model of this phenomenon but the Cloud will usher in many more. Rubin suggests CIOs prepare to take on the role of stewards of the Technology Economy in their organizations.

Takeaways

- In down economies Change the Business (CTB) spending may need to increase to support beneficial transformations. However, one should be careful that this is not the absolute driving metric

- Cost cutting should focus on demand management by lowering excessive IT and look for commodity pricing

- Virtualization is underexploited. Companies should seek expert help

- The cloud is an early form of "tech-commons", a paradigm where IT will be shared and collaboratively developed by peers in competition

- Value of open source lies in harnessing collaboration, not in the cost of software acquisition.

Summing Up

While the going has certainly gotten tough, this issue of *Perspectives* provides some novel and specific advice about how to move beyond simple cost cutting toward a more agile posture. By introducing new ideas and revisiting some established concepts, we beleive that this edition illuminates a path that goes beyond budget slashing to a new vision of IT.

In the near future, we believe that our clients will find a way to transform IT from a sometimes inert mechanism into a constantly adapting Technology Economy, an organism that maintains its health and balance with proper care. The specific steps recommended in this document provide a foundation that will help this organism come to life.

As always, we look forward to hearing your thoughts about the insights presented in these articles. Please email us at *global.consulting@tcs.com*

Aligning IT and Its Costs

IT Service Management

Aligning IT Using the Service Catalog Lens

Robert Mahler
Consultant, IT Process and Service Management

The service catalog is at the core of aligning IT to business goals, but there are trade-offs when implementing it

Abstract

Looking at a service catalog as a mere repository of IT services is misleading. A service catalog that takes into account the needs of internal customers goes a long way in helping IT align itself with business goals by defining the services that the business would consume. Today, this has emerged as a core discipline for determining the role of IT in business.

Yet, the approach to developing an effective catalog may vary depending on the sophistication of a company's IT department. This article discusses two approaches: top-down (starting with customer requirements) and bottom-up (emphasizing reuse of existing assets). Balancing these approaches leads to a trade-off that can start companies on a journey to a cost-effective service catalog.

The "build it and they will come" approach does not work

Currently, most organizations create their service catalog based on what they think their internal customers need and are willing to pay for (this article focuses on internal customers, but many of the ideas can be applied to services created for external customers as well). IT organizations then go about designing services using what they feel is the most cost-effective method for themselves as well as for their customers, without actually asking their customers what they want. This is the "build it and they will come" view of service design. However, what is frequently missed are questions such as "Are these the right services?" or "Do my customers really need or want these services?" or "How will I determine if these services are meeting the needs of my customers?" The answers to all of these questions are critical to determine the optimum content of the service catalog.

The best approach is to first obtain a thorough understanding of the organization's current or potential consumer base and then design the services that are needed. To that end, a company that is implementing services would first determine what customers they are serving, the needs of those customers, the key business drivers to use, and how to measure performance. For example, the customers might be internal software users who need minor enhancements to existing applications and functionality. The key business drivers would be to reduce support costs and preserve existing levels of quality. The IT services organization would then decide how to measure success, how to charge for the service, and how to continue to improve the service catalog.

To effectively design the contents of a service catalog, great thought must be given to measuring the quality of services. Not only should the efficiency of services be measured, but their effectiveness from a customer perspective should be measured as well. Traditionally, the only measurement of a service is whether it met the Service Level Agreement (SLA), but this only tells part of the story and can be particularly misleading if the SLA itself is flawed. For e.g. in the airlines industry, passenger analytics is mission-critical to the reservations department and they would seek 99.9% availability. The marketing department, on the other hand, could do with less since they would work with periodic data.

 In addition to using the SLA as a benchmark, there is an emerging trend towards measuring the customer's quality of experience, which gives the organization more information about how well it is meeting the customer's requirements, and, more importantly, their needs. The main tool for measuring the quality of experience is often a customer survey. However, such surveys rarely provide valid and actionable information. Rather, a best practice in ensuring quality of experience is having a beta phase, as a new or changed service is rolled out, where users' interaction with the service is monitored in terms of business benefit and experience. This helps establish two important factors for chargeback: the service level required by each business unit and the perceived value of the service. The combination of these two factors determines whether a business unit will be assigned a bronze, silver, gold, or platinum level of service.

Chargeback a price, not the cost you incur

Chargebacks frequently fall into two main categories: the cost of the service plus an overhead markup, and the cost of meeting the service levels agreed upon with the business unit.

The first method is generally effective in covering the provider's costs but rarely effective in pricing the service based on the true value to the customer.

The second method is more effective at establishing a price based on customer value, as long as the SLA is defined by the customer based on business needs, and, not just defined by the service provider's ability to provide the required service. For instance, one business unit may see a service as mission critical and want 99% availability while another business unit may not require this level of availability. There should be higher pricing for the more demanding business unit.

To implement the second method, SLAs should be dynamic, with the same services having different service levels depending on the customer *(Fig 2.1)*. The service levels can be obtained from the quality of experience measured during the beta phase.

A service catalog should be flexible to allow a customer to select the services that they want as well as the service level they need. The service level definition should incorporate factors such as application availability and response time.

Fig. 2.1: Supporting service levels different for each customer
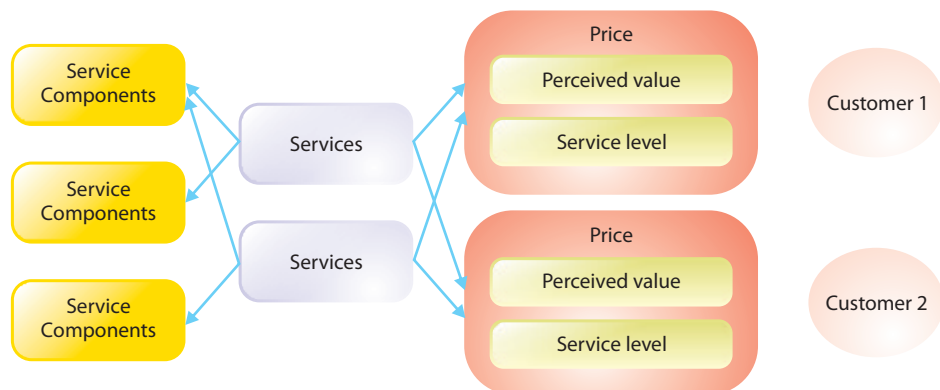
Source: TCS Global Consulting Practice – Research Desk

This "pick and choose" model would require services to be modularized to allow reuse of components. Effective sharing of lessons learned and improvements across the services can help in building such a modularized service catalog quickly. What is achieved through this process is a catalog that could have different prices for different customers, based on what they want and the level of service they need.

In traditional thinking, business requirements are traced to system requirements, which then result in design, code, and testing; the new paradigm will lead to business requirements traced to service requirements that will then become the source of software requirements. This shift will mean that the true product of the organization will become the contents of the service catalog, which will be enabled by the software - the opposite of the current paradigm *(Fig. 2.2)*.

Fig. 2.2: THE PARADIGM SHIFT

Traditional services approach

Service catalog approach

Costing

Direct cost

Overhead

Business interface

SLA definition

SLA compliance

IT services

Applications

Infrastructure assets and services

Chargeback

SLA compliance

Quality of experience

Overheads

Business services

Services menu

Service pricing structure

Modularized IT services

Application components

Infrastructure components

IT cost elements

An ideal approach to the service catalog abandons the "build it and they will come" philosophy. It moves closer to a user-centric model that is fully aligned with customer needs.

Balancing Scalable Modules with Best-of-Breed

Many legacy IT assets will go into a new service catalog. However, these assets may not deliver the desired level of service or even meet the emerging needs of the customer. Ideally, these applications should be reengineered to support the service catalog. Yet major reengineering is usually deferred because of budget constraints. One of the goals of the service catalog is to provide the best variety of services and service levels from the existing IT infrastructure to help customers find the best fit. This is a trade-off between the top-down approach, where the service catalog is built from scratch starting with business needs and the bottom-up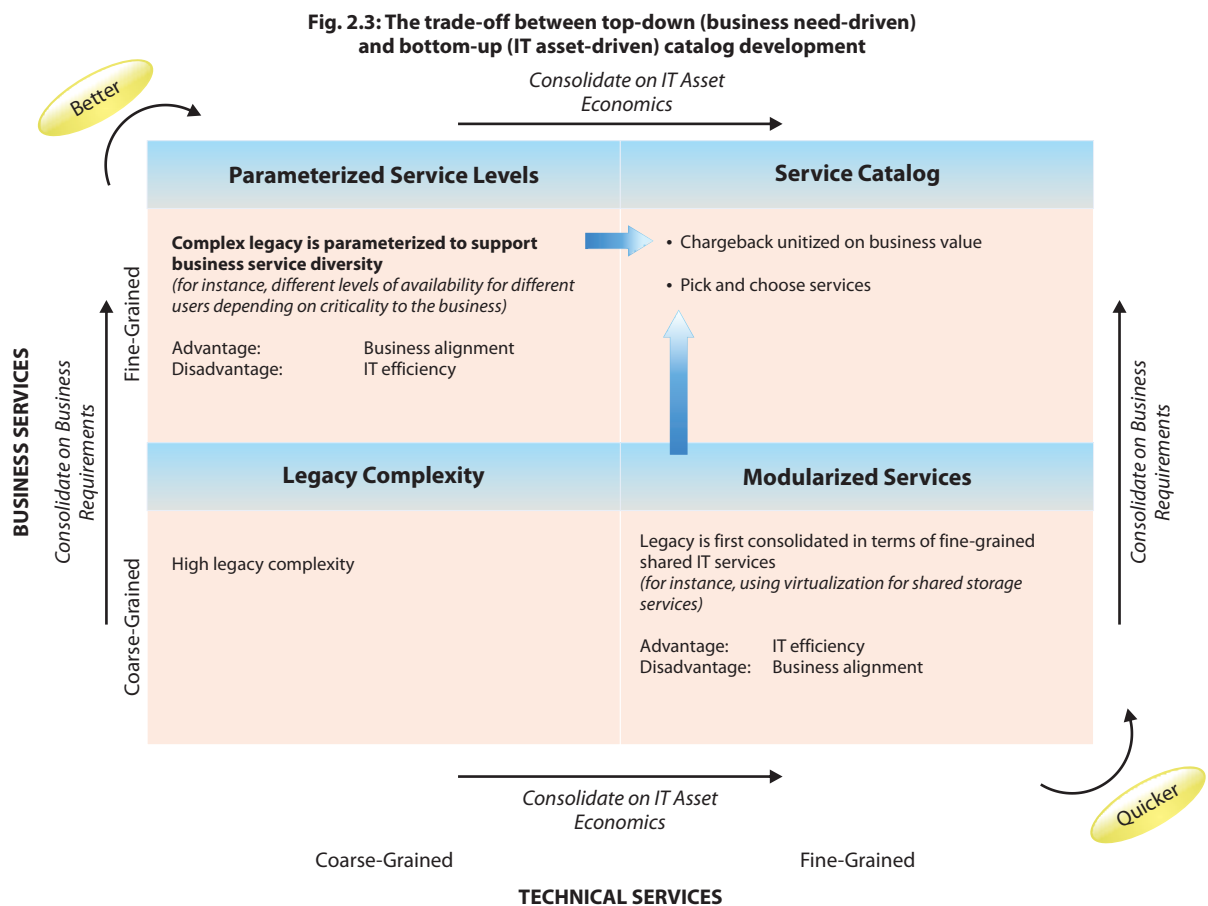 approach, which reuses existing IT assets as much as possible. The best approach is to modularize existing IT inventory to the extent that it supports flexible and changing services. This should be done in tandem with defining services from a customer point of view *(Fig. 2.3)* to find the services that are common to both approaches. While the former helps create a "pick and choose" catalog of services, the latter creates a gradual transformation of chargebacks to the business units by leveraging existing assets.

Fig. 2.3: The trade-off between top-down (business need-driven) and bottom-up (IT asset-driven) catalog development

Consolidate on IT Asset Economics

	Parameterized Service Levels	Service Catalog
Fine-Grained	**Complex legacy is parameterized to support business service diversity** *(for instance, different levels of availability for different users depending on criticality to the business)* Advantage: Business alignment Disadvantage: IT efficiency	• Chargeback unitized on business value • Pick and choose services
	Legacy Complexity	**Modularized Services**
Coarse-Grained	High legacy complexity	Legacy is first consolidated in terms of fine-grained shared IT services *(for instance, using virtualization for shared storage services)* Advantage: IT efficiency Disadvantage: Business alignment

BUSINESS SERVICES — Consolidate on Business Requirements

Consolidate on Business Requirements

Consolidate on IT Asset Economics

Better

Quicker

Coarse-Grained   Fine-Grained

TECHNICAL SERVICES

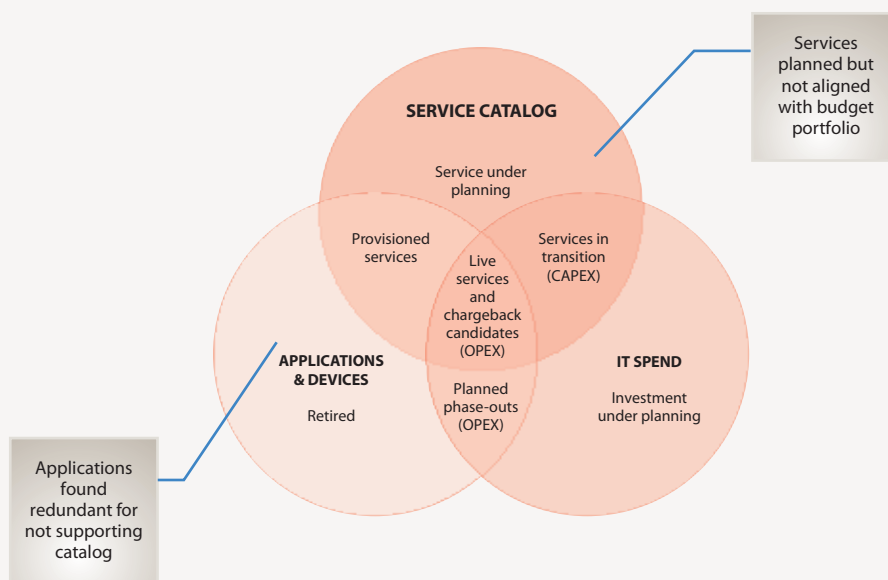Source: TCS Global Consulting Practice - Research Desk

Service Catalog for Application Portfolio Rationalization

The service catalog is said to be defined from the top down when it starts with capturing business requirements and moves to reengineering applications to develop the needed IT services. This approach seems idealistic, especially when the business has sizeable legacy applications that may not be suitable for the service catalog. Nonetheless, this approach can play a valuable role in aligning IT with business goals.

For readers who are aware of enterprise architecture, a top-down service catalog plays a similar role in aligning IT. In both cases, IT identifies redundant assets as well as those in need of reengineering.

In other words, a top-down service catalog helps weed out costly assets that make little contribution to the business.

This diagram identifies enterprise applications that fall outside the service catalog and thus are candidates for retirement or reengineering.

Service catalog is a lens to the IT portfolio helping continous rationalization. This makes many of the other disciplines and approaches to rationalization of IT portfolio secondary.

Source: TCS Global Consulting Practice – Research Desk

Self-correcting IT

While the discipline of service catalog creation is evolving rapidly, some best practices have emerged: First, services have to be defined in terms of what the consumer needs. Second, quality of experience should be factored in the chargeback. The best way to do this is by capturing the user experience in the beta phase. Third, charging back a price rather than the cost makes IT more aligned with business goals. The service may have different service levels for different business units based on the demands of each unit. The difference in service levels would determine the price. Lastly, IT services should be defined by consolidating IT assets into granular services that can be shared across multiple units with different service levels.

To change the current view, a change in mindset must occur. No longer will we be able to define ourselves based on the software we create. Rather, software will be defined in terms of the services the customer would need. The product would cease to be mere software or software tools, but contents of the service catalog. Defining requirements and building solutions then becomes self-correcting – with solutions being validated using the service catalog lens. Effectiveness and efficiencies of the catalog would emerge as the driving metric for business-IT alignment.

IT Financial Management

Making Chargeback Workable

Santhanam Mukund
Head, Business Analytics, Global Consulting Practice

Service-based costing can help achieve many chargeback goals even with rough estimates

Abstract

IT financial management often grapples with the problem of handling chargebacks using methods that business units either don't understand or won't accept. Most costing methods fail to go beyond simple asset accounting because of complexities in IT resources and their usage.

Utility computing, in which IT resources are charged back similar to the way utilities charge customers for energy, will take a few years to come to fruition while the technology to support it matures.

However, IT can still help business units find value for their money by involving them when developing costing methods. Service-based costing being one of those.

A joint effort between business units and IT can help achieve chargeback goals while still relying on relatively rough estimates. Agreement about the methodology used for chargeback is key to success.

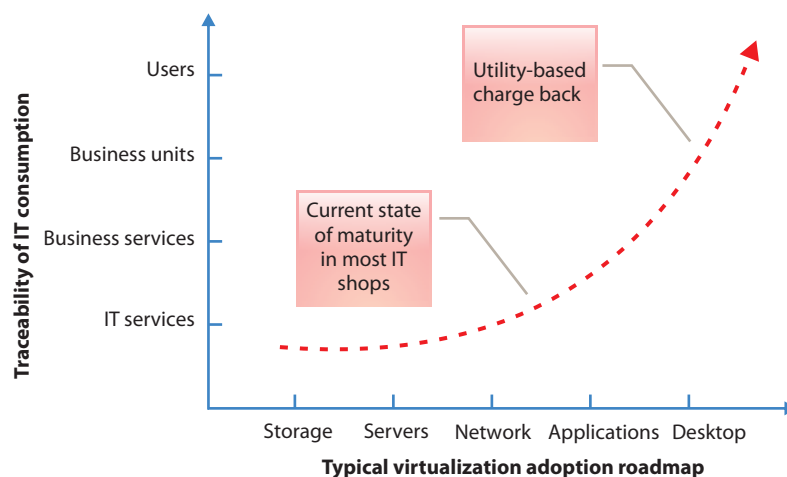# Are business and IT speaking the same costing language?

In the past few years, there has been intense talk about utility costing being brought into enterprise IT. Utility costing means charging back for actual usage, much in the same way that electricity and water are billed to consumers. Today, even best-in-class organizations are finding it hard to implement an accurate chargeback model. Legacy applications, ad hoc IT investments, and shared datacenters are among the complexities that make such chargebacks difficult.

But how important is utility chargeback, really? What matters most is charging back IT costs using trustworthy estimates. Business units often distrust IT chargeback because they think they're being charged for more than they're getting. The lack of trust stems from a costing model that excludes the business units from discussions on the costing methodology. Most methods suggest assigning IT cost elements like datacenter overhead on flat averages. This places an unfair burden on business units that consume fewer resources.

Fig. 3.1: Utility-based chargeback not ready for deployment

Source: TCS Global Consulting Practice, Research Desk

This article proposes that businesses should not pursue utility chargeback for the time being because technologies such as end-to-end virtualization and service orientation that supports metering IT usage still need a few years to mature *(Fig. 3.1)*. Rather, practical costing methods should be employed so that business users can understand and accept the trade-offs in estimates. A costing culture needs to be shared between business and IT so that they understand the costing model and interact with it using a common costing terminology.

Service-based costing can help IT achieve this. To understand service-based costing, we need to look at the typical maturity levels in IT financial management.

IT financial maturity depends on IT service management maturity

The following illustration shows how maturity levels in IT financial management relate to IT service management maturity. For instance, a good configuration management capability supported by a well-defined Configuration Management Database (CMDB, an IT Infrastructure Library best practice) would help link IT assets with services to their consumers and make assigning costs easier.

Fig. 3.2: Maturity levels in IT financial management correlates with service management maturity

ITFM COST CULTURE MATURITY		IT Asset Management		IT Configuration Management		Analytics
Optimized						Influencing business processes; self-sustaining IT by creating measurable business value
Integrated					Demand- driven chargeback & budgeting (like utility-based chargeback)	
Focused				Service-based costing / cost transparency		
Structured			Traditional IT costing (high-level drivers)			
Basic		Standalone accounting				
		Asset	Lifecycle	Inter-relationships	Analysis	Process analytics

IT SERVICE MANAGEMENT MATURITY

The fig. 3.2 shows that a business would reach its most mature level in IT financial management when IT becomes self-sustaining, has an optimal portfolio, and is constantly monitoring the business value of IT investments.

Most IT organizations are grappling with costing methods at the second level (see structured level in the illustration), where IT assets are only accounted for as either capital expenditures (like hardware purchases), or operating expenses like (IT staff Full Time Equivalents). A broad average is then used on the booked cost to charge back to the business units.

The immediate goal, hence, is to reach the third level, where there is a culture of service-based costing. Here, IT is able to draw relationships between IT assets and services and move into a more unitized chargeback that is trusted by business units.

Service-Based Costing (SBC)

In SBC, IT resources are traced to IT applications, and eventually to services rendered by the application. For example, the cost of servers running an Enterprise Resource Planning (ERP) system would be assigned to the ERP application and then to the services within it, like *create journal entry* or *close books* services supported by ERP. The services are costed to the consumer of the services based on agreed-upon rules.

Resources	Services	Cost objects
• Hardware • Software • Network • People • Facilities • Outsourcing	• Applications services • Enterprise service • Personal technology service	• Lines of business • Shared services and functions • Direct to customers
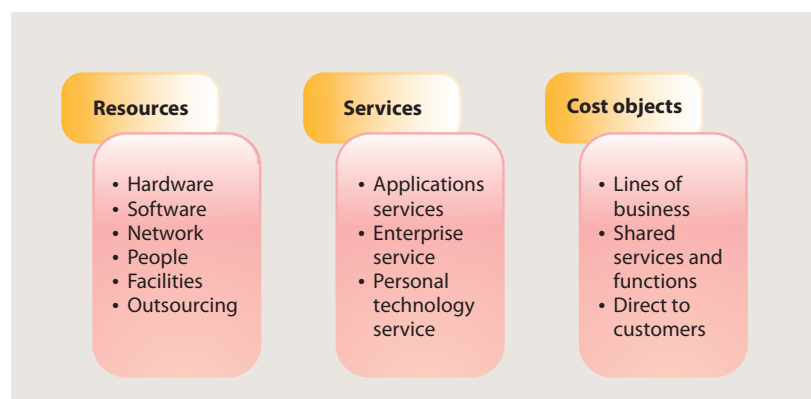
However, tracing IT resources to applications, and then to services, is easier said than done. First, defining the services involves unraveling the intricate relationships between business processes and applications. Second, resources consumed by an application are often shared by multiple applications. Getting around these complexities is onerous. Is there a simpler way?

Focus on services with high costs

Services can be categorized into three groups: enterprise services, personal technology services and application services.

Enterprise services are shared services like the local area network (LAN), datacenter overhead, power usage, and so on. Such services are typically apportioned using device technologies. For instance, allocation of the LAN to business units might be proportional to the bandwidth allocated through devices like network routers. In addition, enterprise services can be translated into a catalog of services with prices and

Apparent similarity to Activity-Based Costing

Activity-Based Costing (ABC) is another costing method based on usage. However, it has not made significant inroads due to complexities in legacy systems.

It can be argued that service-based costing is ABC as applied to services, by translating IT inventory to services. However, there is a tradeoff in that service-based costing in practice measures usage via tangible parameters that may not always reflect true consumption. One example is charging back datacenter power based on floor space rather than on the number of servers supporting an application for the simple reason that the latter is not always traceable.

By definition, ABC in its ideal form, is a utility-like chargeback methodology.
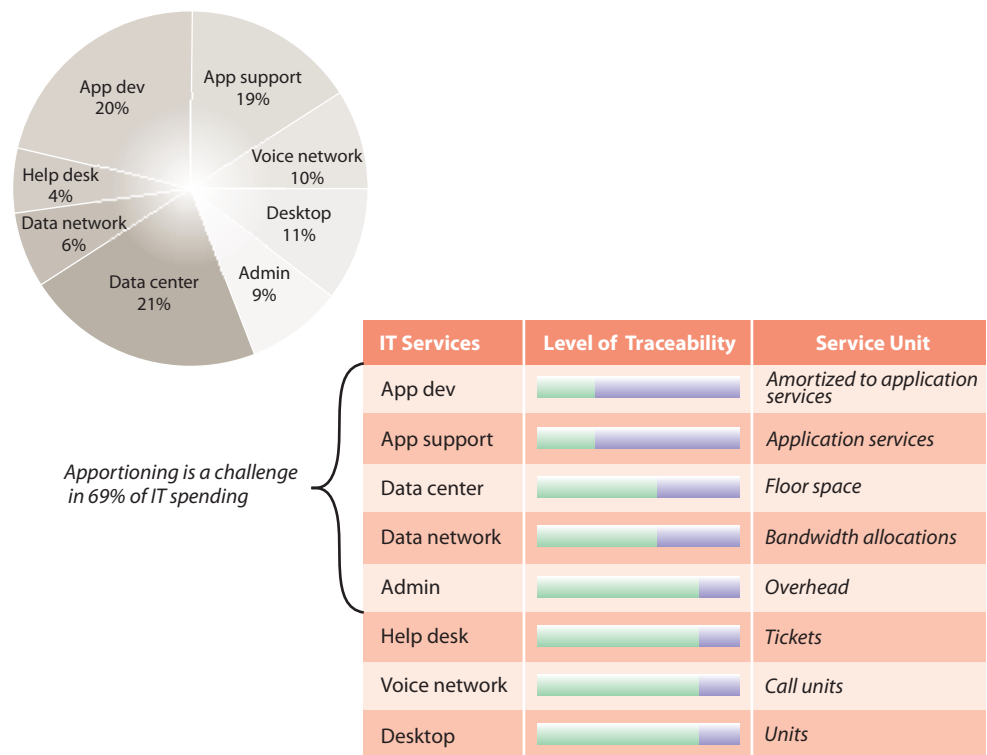
practical units attached to them. For example, datacenters may be charged back based on the relative floor space consumed by the business unit, in case the floor space is more clearly demarcated than are the servers.

Personal technology services are assets dedicated to users, like desktops, client software, and the like. By definition, they support direct costing to business units.

Application services use one or many applications to enable a business transaction. For instance, creating a goods movement would involve both ERP and supply chain applications. Most costing complexities lie in application services. In general, we find that application services are costly, and, at the same time, difficult to apportion in any tangible way. Mapping the underlying resources (servers, network, licenses, maintenance, and so on) to an application is a challenge. For instance, an ERP system might use groups of synchronized servers (known as clusters) that are in turn shared with other applications.

The best place to start building a costing model is with key applications that have significant costs for the business. Typically, the top few systems represent about 80 percent of total costs. The question is, which IT assets make up the 80 percent? *(Fig. 3.3)*

Fig. 3.3: Breakdown of average IT spending

Apportioning is a challenge in 69% of IT spending

IT Services	Level of Traceability	Service Unit
App dev		*Amortized to application services*
App support		*Application services*
Data center		*Floor space*
Data network		*Bandwidth allocations*
Admin		*Overhead*
Help desk		*Tickets*
Voice network		*Call units*
Desktop		*Units*

Source: TCS Global Consulting Practice - Research Desk, spending data sourced from Gartner 2009

One way to identify such applications is by looking at the capacity management history to discover the frequency of capacity upgrades. Applications that must upgrade their capacity often are typically transaction-intensive, consuming substantial resources. Define application services by looking into the transactions within the application. For example, a business might have workflow applications, usually classified as middleware,

that stitch multiple applications together into common processes. The services provided by the middleware could include:

- Notifications, which inform users about actions due

- Transactions, a series of actions connected to a business activity such as a payment to a supplier

- Work definition, a service that enables business units to define their processes

While well-defined service management processes classify important applications by the services they perform, apportioning the necessary resources requires documenting the interrelationships between these services and their resources. Such information can be maintained in a CMDB. CMDB is the database that keeps an inventory of all IT services and assets, along with information on how they are related. In reality, though, very few IT shops have adequately updated CMDBs. Most CMDBs succeed in keeping an inventory of IT resources alone, without application relationships. While modern CMDBs have the ability to autodiscover and update IT resources such as CPU and application usage, tracing the interrelationships is largely a time-consuming manual activity. This makes apportioning a challenge.

Hence, it makes sense to choose an abstraction of IT resources that is practical for making estimates.
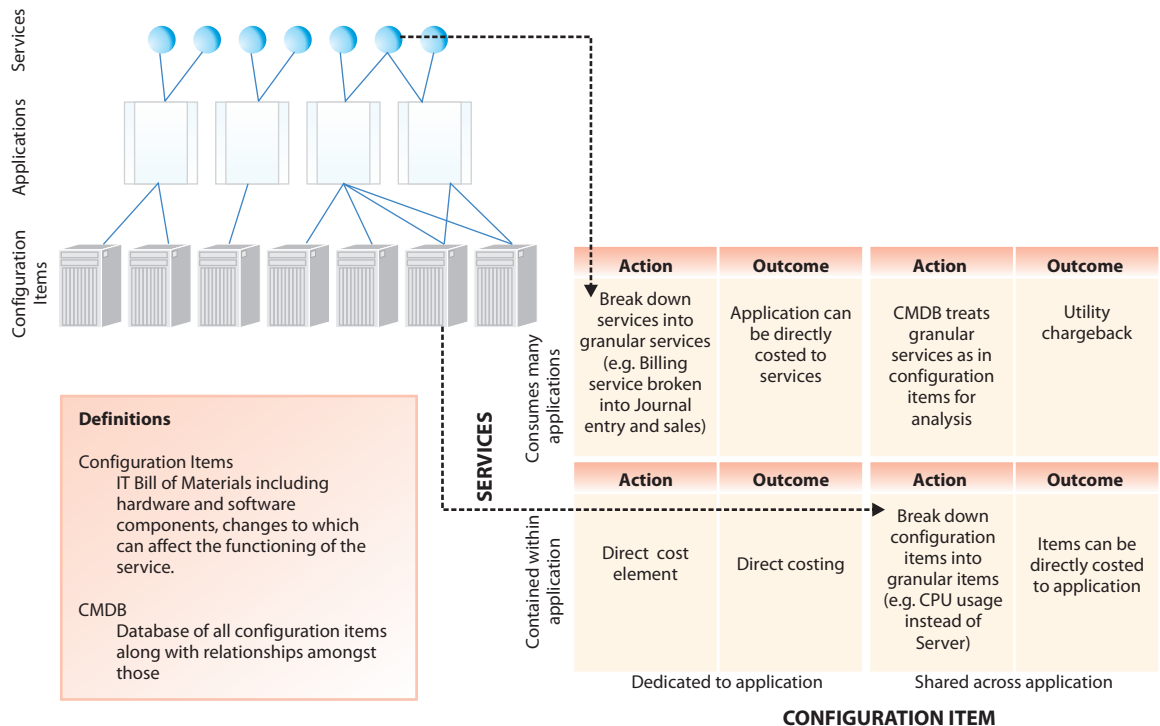
Addressing the cost assignment challenge using granularity

Let's go back to the workflow application example. Such an application includes multiple tools, database licenses, and hardware. Many of these resources are shared across services provided by multiple applications making assignments complex because the inter-relationships between services and the resources they consume are many-to-many.

For dedicated resources, the costing model is as simple as direct assignment. In the case of shared resources, the resources may be broken down in granular terms to support direct assignment. For example, a server may be broken down into CPUs and RAM *(Fig. 3.4 on next page)*. Similarly, if a service is jointly rendered by two applications, the services can be broken down into smaller services in order to assign the application cost to the services. For example, a sale of goods transaction service would involve both supply chain and ERP applications; it could be broken into goods issued (in supply chain) and sales invoice (in ERP) services.

Even though this type of assignment is becoming easier with the adoption of technologies like virtualization, choosing a granularity that is practical is necessary when defining services in order to make plausible estimates. However, a weighted score can be used for apportioning resources when finding granular information is difficult.

Fig. 3.4: Choosing the granularity for assigning cost element

	Action	Outcome	Action	Outcome
Consumes many applications	Break down services into granular services (e.g. Billing service broken into Journal entry and sales)	Application can be directly costed to services	CMDB treats granular services as in configuration items for analysis	Utility chargeback
	Action	**Outcome**	**Action**	**Outcome**
Contained within application	Direct cost element	Direct costing	Break down configuration items into granular items (e.g. CPU usage instead of Server)	Items can be directly costed to application
	Dedicated to application		Shared across application	

CONFIGURATION ITEM

Definitions

Configuration Items
IT Bill of Materials including hardware and software components, changes to which can affect the functioning of the service.

CMDB
Database of all configuration items along with relationships amongst those

Source: TCS Global Consulting Practice - Research Desk

So far, we have looked at IT inventory and service management together to address a service-based costing model. Yet, despite developing and implementing the model, the business may fall short of achieving an environment where IT chargeback is trusted by business units. Many of the costing methods discussed could still be far from beneficial unless business users understand the services and the costing mechanisms. A gap exists between business and IT when it comes to sharing the complexities of technology because business users see IT as being the domain of geeks. In many instances, better communication is needed more than accurate costing techniques.

Development of any costing model should include participation from business units, especially while defining services to include in chargebacks. This helps business units respect the diligence that goes into the estimates.

Softer methods matter more

In our quest for better costing methods, the lack of tangible measurements for resources is often overcome by softer methods like better collaboration with consumers.

The primary objective of a costing model is not accuracy, as one would believe, but finding mechanisms that are trusted by business consumers. A costing model, even one that relies on broad estimates, can be a business driver if business units adopt it. By making business units part of the costing methodology development process, costing decisions meet many of the goals of an accurate chargeback.

The success of service-based costing is in sharing a common mindset about IT usage, developing a culture of understanding regarding the limitations of IT costing, and working out practical ways to address the inherent challenges. This is a milestone in IT maturity. True utility costing may ultimately be of help, but it will take a few more years before the technology to support it is available.

What to charge back? The cost versus price debate

There is a growing debate about whether the cost charged back on services should factor in the contribution to the business, not merely the money spent. This means putting a price tag on services where cost is added to an agreed-upon value in terms of revenue or profitability.

While this may sound philosophical, the debate is understandable. Costing often relies on broad estimates that are subject to as much speculation as perceived value would be, at least in some areas of IT. This subject is better dealt with from the perspective of the service catalog, a topic discussed in an earlier article.

Getting More from Operations

Service-Oriented Architecture
Quick SOA

Dhakshinamoorthy Renganathan
Senior Consultant , IT Architecture, Global Consulting Practice

A workable approach to SOA for tough economic times

Abstract

Justifying ambitious investments in Service-Oriented Architecture (SOA) is tough.

As it would be with any transformation, SOA can turn into a bunch of messy projects in its pursuit of agility for the enterprise. Now, with the global economic malaise hitting IT spending, such investment with a long-term payback will be scrutinized.

This calls for a different approach to SOA in the near term.

The building blocks of SOA are services. This article proposes that a piecemeal approach can be taken to implement SOA by creating additional categories of services that reuse legacy standards and systems, enabling smaller SOA investments to turn into quick wins.

Agility is elusive

Since its inception, there has been a strong case for adopting SOA for business agility. Services are bits and pieces of multiple applications that bring flexibility to processes. They provide data where it is needed and allow processes to be adapted to the needs of the business. At its core, SOA is process-centric, which means SOA is about making processes choose the services during process execution. The services chosen could be different at each instance of the execution, making the process flexible.

> **Verbatim**
> "I have also seen a number of companies that are starting to realize small cost savings and increased agility, but it's taken them 6+ years to get there, and they have not yet recouped their initial $15 million investment. It will probably take them another 3-4 years to break even."
>
> **Anne Thomas Manes,**
> **Research Director, Burton Group**

Today, a significant proportion of IT investments are in SOA projects. The risks are high since projects tend to be too transformative for the processes and applications involved. It is common to find these projects mired midstream with the original rationale for introducing SOA long forgotten.

Many of the hard lessons learned stem from starting with too grand a vision. In a SOA transformation, processes are redesigned for flexibility, and then applications are altered to support the services required in the process model. However, processes that looked good on a blueprint did not go as smoothly when tied to legacy applications and work cultures (the automated and the human parts of processes). Changing legacy systems is expensive. As a result, it makes sense to set an achievable target, one that is modest but good enough to get SOA started with quick returns for the business.

This requires that agility be brought into areas where it matters most and that the relevant processes be made flexible using SOA.

Agility simplified

A business generally looks at agility in terms of its current and emerging business models. For instance, a telecom provider today has competition from players outside the telecom industry. Nokia is seeing competition from Google who off late introduced the Android platform for hand-held telecom devices. Hence, lines are getting blurred, and there is no standard way to establish the scope of a company's operations. A company has to outline sections of its operations that need to quickly respond to this dynamic environment. Typically, there are three common perspectives:

1) *New market readiness* - This is the ability to expand to new markets, including factors such as handling new currency, distribution model, or compliance requirements. Predicting potential markets and charting out the prospects helps in creating the process foundation.

2) ***Market responsiveness with new products*** - This means responding to market dynamics with new products or services. Processes need to be provisioned for different options in the product lines as well as new support services for customers.

3) ***Cross-enterprise collaboration*** - This is the ability to add new suppliers and partners. Businesses should identify potential partners and move quickly to operate seamlessly with them.

To understand the implication of these dynamics on processes, consider the case of introducing a new product, which is very common in agile businesses. A new product development may look far easier on its way through R&D until that product reaches production. On production, the assembly would require collaboration with suppliers using a supply chain model to support the economics. Much of the process would require changes in existing applications like ERP and supply chain management systems. SOA helps processes change without changing applications. Had these applications been wrapped using well-defined services, the new process could choose relevant services dynamically. The process would then pick sections of the processes that would suite the new product. This way, the business is able to introduce new products more quickly.

Yet many processes redesigned with SOA encounter unforeseen complexities. First, our reliance on SOA standards, such as Simple Object Access Protocol (SOAP) and Web Services Standards (WS*), demands better infrastructure to overcome performance challenges. Second, parts of the processes reside in legacy applications that are rigid, making reengineering expensive.

Start off by picking the few processes that are critical to business agility by considering these three factors:

• Processes that are critical to cross-enterprise efficiency with important partners

• Processes that are customer-centric (example, customer acquisition and service management)

• Degree of straight-through processing (how far a given process can be automated without human involvement)

Table 4A suggests the relevant criteria to determine the processes most amenable to immediate application of SOA.

Table 4A: Typical factors and criteria in selecting processes for SOA

		Business Analysis		Technical Analysis		Information Analysis	
		Which processes need to change?		*What standards can help?*		*What business intelligence is needed?*	
		Efficiency	Flexibility	Global standards	Local standards	Data domains	Data quality
Agility Factors	Cross-enterprise collaboration	• Automated cross-enterprise transaction	• Drivers for dynamic supply chains (interoperability across suppliers)	• WS* • Global data synchronization network	• Message oriented middleware • EAI compatibility	• Order to cash • Procurement to pay • Master data	• Synchronization with partners and suppliers (such as for identity federation)
	Customer-centricity	• Cross-sell/up sell • Service desk operations	• Product portfolio management • Service and product configuration management	• Quality of service • Service management standards (such as ITIL)	• CRM maturity	• Customer data information • Product information management • Configuration items	• De-duplication ratio
	Straight-through processing	• Length of automated workflows • Discretionary human involvement	• Quality of decision support in human interventions	• SOAP/REST orchestration bus	• Legacy middleware integrated with users (like Lotus Notes workflows and EAI)	• Users in organization hierarchy and structure • Role-responsibility mapping	• Workflow digressions and redundancies • Number of corrective delegations
Criteria to determine core processes		• New market readiness • New product readiness • Compliance readiness		• Feasible within short timespan • Maximum reuse of legacy • Low cost of implementation		• Domain consolidation definitively leads to higher data quality • Timely flow (real-time or disjointed) • Information security and compliance	

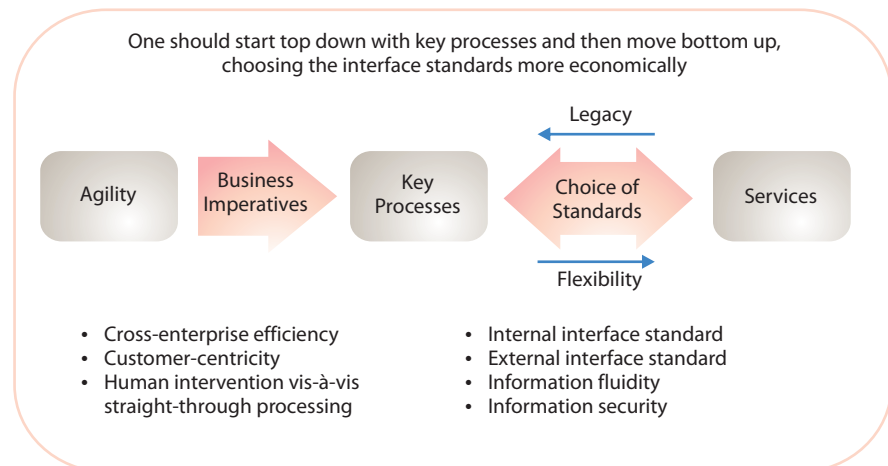Source: TCS Global Consulting Practice – Research Desk

During the first stage of SOA implementation, processes are identified and designed for flexibility. After that, services to support the processes need to be defined. In an incremental approach to SOA, we would require new categories of services in addition to what is normally used. The traditional definition of services in SOA has three layers: 1) the services consumed by business users, 2) the underlying application functionality exposed as services, and 3) technical infrastructure services such as networks and servers. This approach, while desirable, forces us to consider making sweeping changes in order to implement SOA, often overlooking the cost of reengineering legacy applications.

Service definition driven by choice of standards

The goal at this point is to limit the services to a critical few processes, reusing legacy systems as much as possible. The key parameters are:

• Reusing legacy functionality
• Reusing legacy architectures
• Increasing revenue and profits
• Reducing operating costs

While factoring in these parameters, one should still define services from the perspective of the consumer, that is, the business users who run the processes. More often than not, services are implemented to make processes flexible for a different purpose from what the consumer expects. For instance, the user originating a loan at a lending company usually contacts an external agent to validate the customer's credentials; the information exchanged with the agent could use a simple email workflow automation instead of integrating with the agent's software using WS* standards.

One should start top down with key processes and then move bottom up, choosing the interface standards more economically

| Agility | Business Imperatives → | Key Processes | Legacy ← Choice of Standards → Flexibility → | Services |

- Cross-enterprise efficiency
- Customer-centricity
- Human intervention vis-à-vis straight-through processing

- Internal interface standard
- External interface standard
- Information fluidity
- Information security

Source: TCS Global Consulting Practice - Research Desk

The choice of standards for services plays an important role in SOA economics. To take this into account, along with the SOA services hierarchy explained earlier (business services, application services, and infrastructure services), we should consider three new categories of services explained with an example on the following page:

1) **Core services:** Services for cross-enterprise and customer-centric processes important for agility. These would typically use SOAP services. Here, WS* specifications, an accepted industry-wide standards for better collaboration and service orchestration, supports contextual changes in processes.

2) **Inner services:** *Services* needing flexibility and can follow legacy standards like the native interfaces provided by the applications or traditional HTTP interfaces.

3) **Peripheral services:** These could remain standalone in their legacy form due to their small role in flexibility and their high reengineering cost. Standards do not matter much in these services. These services could remain disconnected from the processes with batch interchanges.

Example

Consider a lending institution. Agility is defined by how quickly the lending institution can respond to different market situations with lending products. Its lending products offer a mix of fixed and variable interest rates.

It identifies three business services important to this process: *Create Loan, Accrue Income,* and *Close Account.* The Accrue Income service is seen as core because it determines how the business reacts to varying market rates. It needs to interface with an external agency to gather current market rates. On the other hand, the *Create Loan* and *Close Account* services are less important for flexibility.

Further, analyzing the applications for compatibility reveals that three application services are involved: *Create Loan, Close Account,* and two variants on the *Accrue Income* service: *Accrue Income at Fixed Rate* and *Accrue Income at Variable Rate.* These services are a part of legacy applications, which are quite old. Modifying these services would be expensive.

However, flexibility is needed in *Accrue Income* at Variable Rate for strategic reasons. It is reengineered to use SOAP messaging and picks up market rates from external agencies and updates interest rates dynamically.

The *Create Loan* service is already an online application. It needs customer-centricity so that the customer interface can be improved frequently for a better user experience. It is reengineered using an existing standard (traditional HTTP) and marked as an inner service.

The other two application services are kept as is, making them peripheral services.

The classification of services could change based on more information.
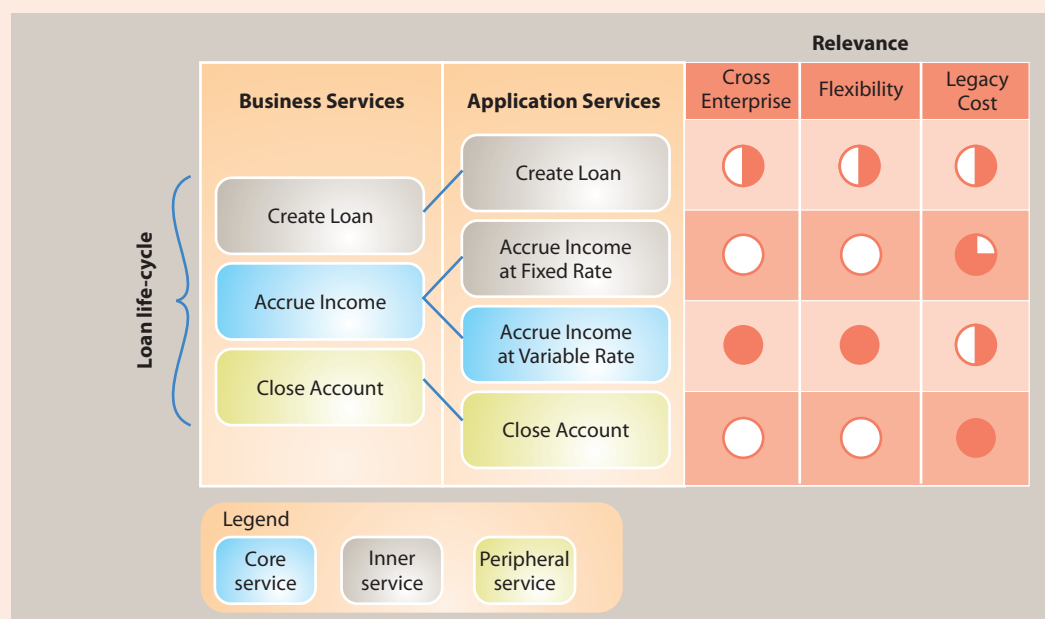
New categories of services are needed for an incremental approach

When creating a transformation roadmap, look for quick wins. Limit the meaning of the word "agility" to help identify core services that offer the best returns in order to find the top candidates for SOA. Other services can remain unchanged or implemented with fewer integration standards.

Afterword: The SOAP versus REST debate

REST (Representational State Transfer) is an alternative standard that is trying to make SOA simple by avoiding the myriad of SOAP/WS* standards that continue to proliferate.

Readers who are current with trends in SOA and who are interested in advanced information on SOA interfaces will note increasing use of REST architecture in place of SOAP. It seems unclear which one would eventually drive SOA; yet, each has different merits.

The debate is relevant in the context of this article, especially in the choice of standards that determine *core services versus inner services*.

REST is a model that takes its architectural principles from the traditional Internet, which proves the success of REST. Roy Fielding put forward the principles in his doctoral dissertation in 2000. Fielding is one of the minds behind Hypertext Transfer Protocol (HTTP), the familiar Web protocol. Fielding's thesis has caused many SOA practitioners to rethink the SOA approach suggested by the OASIS consortium on WS* specifications, sparking a debate. Today, REST is used as much as SOAP. For instance, people using Amazon Web Services, a platform for application development on the cloud, have used REST for 80% of the services. However, in this case, one could assume that the applications are built from scratch, not inherited from legacy systems. The choice is more difficult when one has to reuse legacy applications for SOA.

REST proposes that applications be designed with more "nouns" (like sales transaction, invoice, goods movement), with each noun exposed as an identifiable URL. SOAP, on the other hand, relies on context-sensitive interfaces with functions ("verbs") encapsulated in XML. For example, REST would fetch a record as "a sales order numbered <some number>" while SOAP might use "find sales order <some number>". Note that the latter would use the same URL for all records while former would have distinct URLs for each sales order.

REST is resource-friendly but limiting in design while SOAP requires more computing power (due to heavy XML to carry context data) that makes data interchange richer.

Also, REST is easy to implement when an application is designed from scratch since in that case the design can support REST principles. SOAP is suitable for legacy systems when the system is heavy on functions and contextual interfaces (or, in developer terms, "stateful").

Today, there is no golden rule for deciding the right interface for the services, other than the architectural intuition one would have regarding interaction needed between applications. If the data interchange carries a lot of contextual information that needs to be persistent across a process over time, SOAP is the most suitable choice. One example is user authentication, which may traverse multiple applications in a process; WS* security standards are well-suited for this. On the other hand, applications that require short-lived transactions across applications may use REST if the legacy application supports it. For example, fetching or updating documents in a document management system may use REST, when each document has a distinct URL in the legacy system (as in Microsoft SharePoint).

Many of the inner services as defined in this article could use REST if the legacy application has a REST-like design.

Source: TCS Global Consulting Practice – Research Desk

Program Management

A United Front: Coordinated Multi-Vendor Programs

Shouvik Banerjee
Head, Service Management Center of Excellence, Strategic Sourcing

Abhijit Mazumder
Head, Strategic Sourcing

Management of multi-vendor programs is ripe for change

Abstract

Large IT outsourcing deals typically operate with multiple vendors. Vendor diversity helps keep prices competitive and mitigates the risk of a vendor failing.

Yet there's a price to pay for this model because projects are often unwittingly designed in a way that vendors work in silos, overlooking the program goals as the vendors compete.
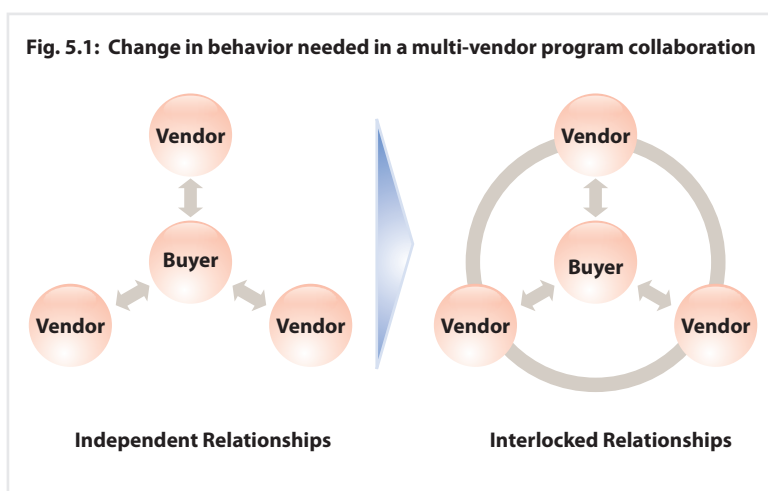
It's time we found a new way. IT outsourcing providers need to become strategic partners rather than pure suppliers. Today, vendors are developing the depth and breadth of their services within the IT outsourcing ecosystem. Buyers, on the other hand, are consolidating on a core set of vendors. Shouldn't this lead to a whole new style of running large multi-vendor programs? There are only a few practical things being done today that can bring about this much needed change.

This article provides an overview of the changing face of program management in IT outsourcing in the context of vendor consolidation.

Multi-vendor programs fall short of expectations

Companies have been striving for efficiency and agility in large outsourcing programs ever since multi-vendor outsourcing became common practice. To date, the achievements are far from the desired outcomes. Without adequate oversight, service providers end up working in silos and don't act as members of a cohesive, collaborative team. Multi-vendor programs are framed to leverage the complementary capabilities of the various providers. When successful, they achieve impressive economies of scale while maintaining a consistent quality of service. In practice, however, successful results are far from the norm.

The biggest barrier is a lack of proactive cross-vendor collaboration. Frequently, multiple providers compete for their slice of the pie and are unable to transcend their competition in order to collaborate. Vendors become reluctant to share data with peers or to perform their assigned project tasks transparently. In the process, stakeholders have little visibility into the program as a whole. The buyer eventually loses control of the program and faces a growing total cost of ownership (TCO). A recent study suggested that the cost of managing multi-vendor programs consumes up to 30 percent of the total contract value, or TCV (Aquaterra 2009, "How to Optimize Complex Multi-Provider Outsourcing Contracts").

Fig. 5.1: Change in behavior needed in a multi-vendor program collaboration

Independent Relationships

Interlocked Relationships

Multiple vendors give buyers negotiating power, but at the expense of vendor cooperation. This defeats the purpose since having multiple vendors was never the goal; the goal was to have many vendors act as one to achieve the business objectives behind the program *(fig. 5.1)*.

Evolution of IT Outsourcing

The evolution of IT outsourcing has yet to see the phases of maturity seen in other industries. For instance, in manufacturing, the journey towards efficiency started in the 70s, and made a significant leap in the 80s when the Japanese quality philosophies swept many production facilities. Next, the accelerating globalization triggered the development of new supply chain models. Multi-vendor IT outsourcing should take a page from the Japanese supplier network models to pursue concurrent service and value delivery.

The Toyota supply network's workaround to a catastrophic fire at an Aisin Seiki factory in 1997 showed the value of the Japanese practice of making vendors work in closely knit families (referred to as *keiretsu*). The fire had paralyzed Toyota motor production because Aisin was the

main supplier of a part called the p-valve. The production was restored in a matter of days when other suppliers rushed to the rescue to establish a makeshift production arrangement. This behavior was in contrast with many of the global auto manufacturing companies, which made vendors compete on cost and price. Yoshio Yunokawa, General Manager of Toyoda Machine Works Ltd., stated that "Toyota's quick recovery is attributable to the power of the group, which handled it without thinking about money or business contracts" (Valerie Reitman, "Toyota Motor Shows Its Mettle After Fire Destroys Parts Plant," Wall Street Journal, May 8, 1997).

In contrast, when a large multinational capital goods manufacturer used separate infrastructure and implementation vendors for a multi-country SAP rollout, the vendors didn't try to compensate for each other's shortcomings. For instance, while the program grappled with frequent outages, the implementation vendor would pass the blame to the infrastructure provider instead of improving its own deployment model for better fault tolerance. While the client thought it had a best-of-breed solution, the individual vendors were each focused on meeting their own set of deliverables. This multi-vendor program faced various hurdles of coordination and strategy integration. Ultimately the need for seeing it as a single program became clear. This led to reconstituting the Program Management Office (PMO) in a way that there was more focus on compliance with business requirements than on contracts with individual vendors.
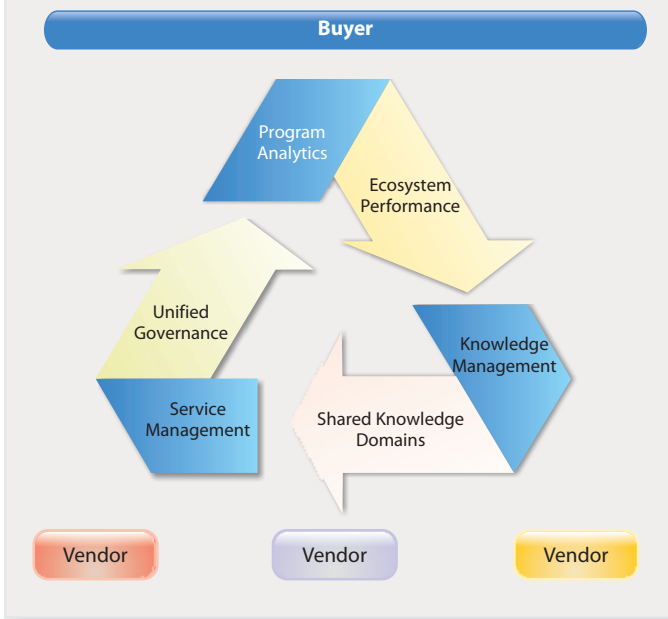
Three factors can improve multi-vendor outsourcing programs

The primary challenge in making multiple vendors operate in unison lies in each vendor having its own service-level agreement (SLA) with the buyer. While each vendor has its own set of processes for meeting the SLA, multi-vendor synergy can be developed using three common elements *(Fig. 5.2 on next page)*:

* Unified service management
* Shared knowledge management
* Centralized program analytics

Unified service management

Service levels should be designed to promote frequent interaction between vendors in a manner that remains anchored to program goals. This contradicts traditional SLAs, which are usually set up as an agreement between the buyer and each individual vendor. In such agreements, clauses encouraging cross-vendor collaboration are often simply words on a page with no effective practical implications. This gap between words and action can be filled through unified service governance, often driven by the primary vendor, with adequate representation from the buyer. A critical responsibility for such a governance body is to create the underlying Operating-Level Agreements (OLAs) that foster cross-vendor collaboration and automated workflows across all domains to support an integrated view of the entire program.

Fig. 5.2: Bringing about multi-vendor synergy

Buyer

Program Analytics

Ecosystem Performance

Unified Governance

Knowledge Management

Service Management

Shared Knowledge Domains

Vendor

Vendor

Vendor

It's also important to have representation from people who are actually involved in the business processes because they look at program performance, not from the point of view of contractual compliance, but in terms of the value they need from the vendors as a group. Their participation helps ensure that vendors share knowledge in the context of business goals. Hence, a common knowledge management system is useful in facilitating this knowledge sharing.

Shared knowledge management

Traditionally, each vendor keeps its own knowledge repository. In large programs, it is even common to find multiple repositories kept by the same vendor which may be servicing more than one project. A multi-vendor shared model works best when the participating vendors proactively contribute knowledge and assets to a shared repository.

Nonetheless, it is natural for competing vendors to safeguard their intellectual assets. Two factors help vendors become more forthcoming in a common knowledge management environment. First, a vendor would like to test the maturity of its own assets. For example, its software development lifecycle (SDLC) processes in a collaborative development. Second, most frontline IT providers are starting to realize that open communities and co-development are more effective than proprietary methods. Knowledge-sharing among vendors helps the provider enrich assets in a shorter period of time.

In a departure from tradition, new factors are encouraging vendors to share knowledge. Multi-vendor programs should take advantage of this tendency by promoting certain knowledge channels:

• Media - Focusing on availability of structured content (data and information)

• Collaboration - Focusing on traditional and non-traditional communication channels (Web 2.0 forums, instant messaging)

• Community - Focusing on networks and communities as alternative sources of knowledge

One would find that these principles apply differently in various knowledge domains in a large program. Knowledge sharing gets more participation in those domains where such sharing helps vendors with their projects. For instance, two vendors engaged in a common development project need to leverage a common specification repository. This helps in concurrent design. Also, some domains need a community-driven approach to encourage sharing of best practices across vendor boundaries, like in technology planning. A broad categorization of knowledge domains and suggested scopes of collaboration are shown in *Fig. 5.3*.

Fig. 5.3: Shared knowledge domains across various multi-vendor interfaces
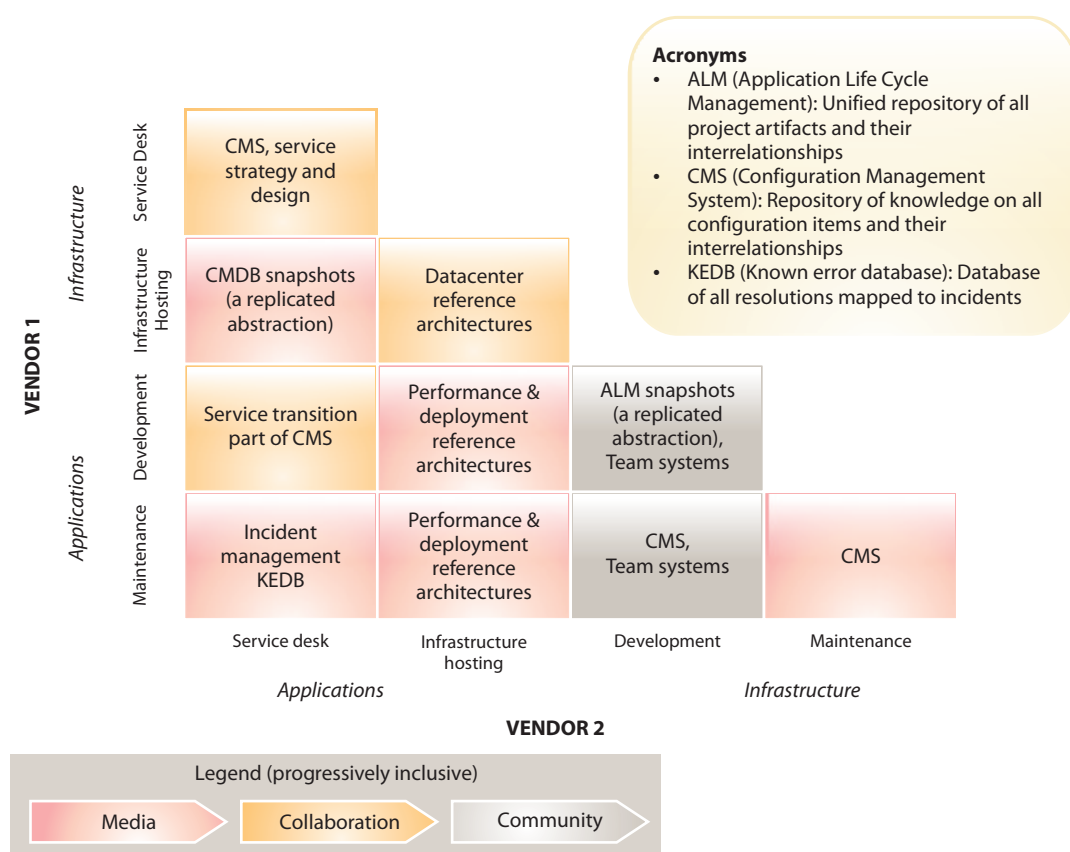
Source: TCS Global Consulting Practice – Research Desk

As we go deeper into the illustration *(fig. 4.3)*, we observe that multi-vendor programs often require sharing a configuration management system (CMS). While it makes sense to have a common CMS for the program as a whole that all vendors can use, in real life vendors usually have their own systems to support their homegrown processes. Often, a provider may end up using outdated inputs on their part of the project which were supplied by another vendor. A possible workaround toward common configuration management might be to use meaningful data replications across vendor systems.
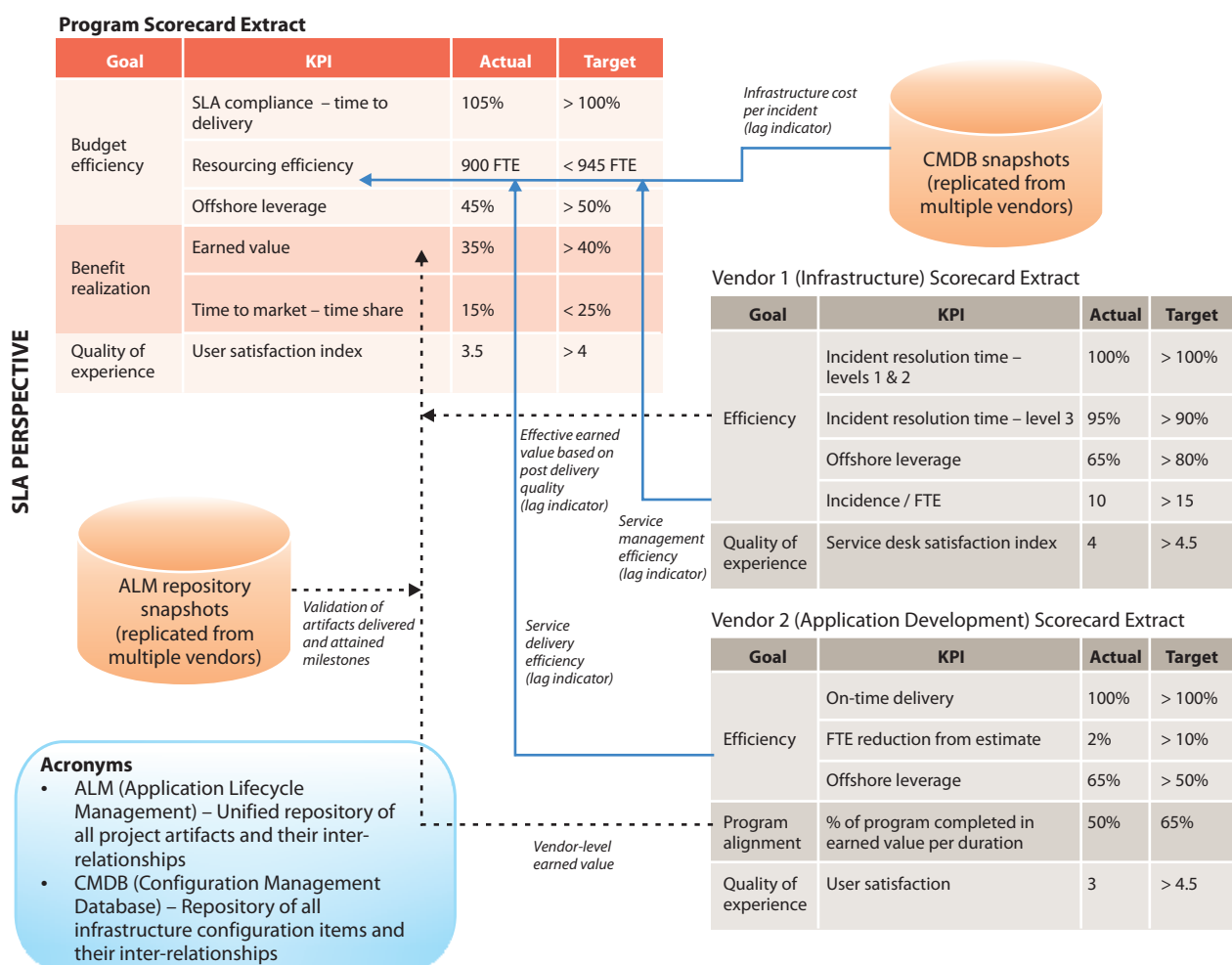
Shared knowledge management calls for better mechanisms to process program data and derive meaningful information about its impact on goals.

Centralized program analytics

It's common for traditional IT governance to use analytics based on balanced scorecards to determine whether or not a project is in line with operational goals. Yet, program analytics distinguishes itself by synthesizing many data sources and deriving contextual meaning. It opens up interesting possibilities when we consider using centralized analytics in a multi-vendor environment, even though more shared knowledge management practices are needed to support this vision. The measurement model must track the interaction between vendors and not just look at a project or vendor in isolation.

Effective multi-vendor program dashboards are created using traditional scorecarding methods such as setting a program vision, cascading metrics, and developing contextual dashboards. But, the effectiveness largely relies on involving vendors at the very inception in order to gain their buy-in. The important metrics are the ones that trace the effectiveness of the project ecosystem within the program. Such scorecards mean that measures dictated in multiple SLAs be connected and rolled up to determine program performance. The *fig. 5.4* is an abstraction of such analytics.

Fig. 5.4: Nature of scorecard analytics in a multi-vendor program

Program Scorecard Extract

Goal	KPI	Actual	Target
Budget efficiency	SLA compliance – time to delivery	105%	> 100%
	Resourcing efficiency	900 FTE	< 945 FTE
	Offshore leverage	45%	> 50%
Benefit realization	Earned value	35%	> 40%
	Time to market – time share	15%	< 25%
Quality of experience	User satisfaction index	3.5	> 4

SLA PERSPECTIVE

Infrastructure cost per incident (lag indicator)

CMDB snapshots (replicated from multiple vendors)

ALM repository snapshots (replicated from multiple vendors)

Validation of artifacts delivered and attained milestones

Effective earned value based on post delivery quality (lag indicator)

Service management efficiency (lag indicator)

Service delivery efficiency (lag indicator)

Vendor-level earned value

Acronyms
- ALM (Application Lifecycle Management) – Unified repository of all project artifacts and their inter-relationships
- CMDB (Configuration Management Database) – Repository of all infrastructure configuration items and their inter-relationships

Vendor 1 (Infrastructure) Scorecard Extract

Goal	KPI	Actual	Target
Efficiency	Incident resolution time – levels 1 & 2	100%	> 100%
	Incident resolution time – level 3	95%	> 90%
	Offshore leverage	65%	> 80%
	Incidence / FTE	10	> 15
Quality of experience	Service desk satisfaction index	4	> 4.5

Vendor 2 (Application Development) Scorecard Extract

Goal	KPI	Actual	Target
Efficiency	On-time delivery	100%	> 100%
	FTE reduction from estimate	2%	> 10%
	Offshore leverage	65%	> 50%
Program alignment	% of program completed in earned value per duration	50%	65%
Quality of experience	User satisfaction	3	> 4.5

Source: TCS Global Consulting Practice – Research Desk

Contractual perspective: Incentivizing collaborative behavior

There is a growing emphasis among vendors to collaborate and achieve program objectives. However, fostering this behavior requires program management to incentivize project contracts. Until now, contracts tended to digress from the program mission while scrutinizing individual vendors. Clauses should be introduced for risk-reward sharing, ones that would spur vendors to complement others. The benefits achieved should produce incentives for this type of behavior.

The following chart offers an analysis of emerging best practices for typical contract types that help promote a risk-reward culture.

Table 5A

Contract type	Project best suited	Buyer		Seller		Who has the upper hand?	Risk-reward best practices	
		Risk	Advantage	Risk	Advantage		Service level riders	Example
Fixed price	Outsourced application development and maintenance (ADM)	Demand management	Predictable costs	Strained relationship with buyer stemming from frequent change requests (CRs)	Resourcing flexibility	Seller	Both parties report on superfluous requirements and resources; agree to engage the two at lower rates	*A project is dogged with CRs. The supplier has redundancy in resources, which are allocated to pending CRs. CRs are resolved at a concessional rate*
Time and material	In-house ADM	Low efficiency	Resourcing flexibility	Lack of long-term commitment	Revenue predictability	Seller	Buyer intends to extend contracts to other projects based on performance	*An in-house developer is tracked on defect per work unit (say function point). The resourcing is extended to other projects based on better results*
Open book/ Cost plus	Infrastructure setup	Lack of ownership of baseline cost by supplier	Costing variability without renegotiation; more visibility on cost	Burden of added disclosure impeding up-sell initiatives	Fixed margins	Both evenly placed	Buyer does independent estimate of baseline cost; supplier reduces baseline cost and takes a share of it	*Such pricing models are suited to projects with complex cost structures and large procurements – like datacenter setup. A rigorous estimate is done by customer on baseline cost; helps supplier benchmark the savings that they could drive*
Unit-based/ Transaction based	Service management	Buyer may lack maturity to unitize service consumption	Scalability with added consumption by business	Break even point critical to the pricing	Higher profitability through shared service components	Buyer	Fixed number of transactions, beyond which unit-based pricing is applied	*Typical, when buyer engages seller for service desk. A minimum number of tickets need to be paid for, beyond which seller is paid on the number of tickets*
Incentive-based	Infrastructure management	Benchmarking seller performance to justify incentive	Payout is more benefit driven	Impractical benchmarks	Fixed cash flow	Buyer	Agreed performance bands tied to the incentives	*The project may be required to provide 97% availability; incentive on higher level achieved*
Gain sharing	Incubation (e.g. cloud computing setup)	Reliant on vendor maturity	Seller shares investment and stakes	Reliant on buyer maturity	Better returns on successful completion	Buyer	Naturally risk-reward in nature; clauses in line with co-owned KPIs	*Suits high-risk technology project, such as private cloud computing. The KPI is usually product specific, like % of infrastructure virtualized in a private cloud*

Definitions

Open book: Refers to contracts where vendors disclose all material cost and charge a markup on that cost. It is normally used when the material cost is expected to vary significantly.

Gain sharing: A pricing structure that emulates joint ventures without equity participation.

Source: TCS Global Consulting Practice – Research Desk

The next generation of program management is about incentivizing behavior

Large programs rely on multiple vendors for a couple of reasons. The first is risk management; the customer would like to hedge the failure of any one vendor by having more than one. Second, by having multiple vendors, customers gain more negotiating power. Even so, most programs still face the pain and cost of having multiple vendors working for different purposes.

The multi-vendor practices suggested here, are a natural evolution in outsourcing where program goals are met by making vendors leverage complementary competencies. However, the practical solution to achieve multi-vendor synergy rarely lies in making vendors subscribe to a common methodology. Rather, it lies in creating incentives that foster a more collaborative environment (*Table A* on previous page).

Shared knowledge management and centralized analytics would have seemed far-fetched a few years back. Yet, with the consolidation of competencies in the IT industry, mature vendors are embracing multi-vendor models, knowing that it is a highly sought-after competency today. Modern program management needs to encourage cooperation and collaboration in multi-vendor projects.

Organizational Change Management

Harnessing the Desire for Change

David Michelson
Head, Business Process & Change Management , Global Consulting Practice

The assumption that people resist change starts change management programs on the wrong foot

Abstract

IT optimization should stop pushing new processes to people unless it finds as much pull from them.

Whether it is about adopting new methodologies, like Agile development, or relocating an IT unit, as in Remote Infrastructure Management, the changes affect people who find the cultures and processes new to them.

To handle this, such rollouts are often pushed with the assumption that people resist change. On the contrary, people seek change constantly - provided one doesn't try to change people's behavior.

How can change management programs exploit the human desire to seek change, striking the right balance between push and pull?

This article takes a fresh look at change management, showing how the principle applies in IT cost cutting.

An apparent paradox

"People don't resist change, people resist being changed" – *author unknown*. Peter Senge cites this quote in his acclaimed book *The Fifth Discipline*[1]. It clarifies a common misperception about change management. The question is how to avoid changing people while effectively changing processes. This is both a paradox and a key to successful change management.

Presuming that change will be resisted has caused many change management strategies to fail. Typically, change management approach tells us that we must expect resistance, even plan for it, and use sophisticated techniques to get around it. This creates a self-fulfilling prophecy. People resist someone else's preconceived idea of what is best for them. Hence, most change management strategies start off on the wrong foot, creating resistance by encroaching on people's identities. So, how can we avoid this?

Preserving identity

There are three aspects of individual identity.

The first is personality. A relevant understanding of this can be drawn from the works of William Sheldon[2], where he identifies three classes of people – easy going, aggressive, and intellectually artistic. Another aspect of personality can be drawn from Florence Littauer's work Personality Plus[3], where she suggests four categories: phlegmatic (calm and docile), melancholy (privately emotional and artistic), sanguine (cheerful and happy-go-lucky), and choleric (quick-tempered and arrogant). Personalities are distinct, and we should be careful to factor in the personalities of the people impacted by any process change.

The second aspect of identity is how people perceive the roles they play. It is common for roles to change in business transformation, but people often expect that their responsibilities will remain the same or even be increased. However, a change in process could mean having to accept a reduced level of responsibility. A sudden change causes resistance. To reduce resistance, one should look at how the affected person sees his/her role and places self within the process. A good way to measure this is using the RACI chart in the CoBIT[4] (Control Objectives for Information and related Technology) framework, where a person is seen as connected to an activity in one of four ways: Responsible, Accountable, Consulted and Informed . A person who has been responsible for a process should at least be consulted about it after the process is changed. If he is merely informed, he will most likely be resistant.

The last aspect of identity involves how a person sees the organization's culture. While every organization has a distinct culture, each employee connects with it in a different way. For instance, an organization known for higher family values will find it easy to retain employees who value socialization more than compensation. A change that

departs from this tradition won't be taken well by those who have identified with it for a long time.

Change management shouldn't attempt to change people's personalities. With well-defined boundaries, change can be welcomed, striking a balance between push and pull.

Striking a balance between push and pull

More often than not, people will choose change if you satisfy four basic criteria that greatly influence their ability to choose *(Table 6A)*:

First, people must understand what the change is and why it is important. This enables them to contemplate the change and how it affects them. Without this basic information, people cannot possibly embrace the change. In essence, this information frames the decision.

Second, people must believe that the change is practical so that they have the confidence to change. Change always involves doing things in new ways, which in turn requires new skills, information, or tools. People have to believe that they can bridge the gap from where they are today to where they will be in the future.

Third, people must believe that change is the right thing to do. While understanding the what and the why that frames the choice, believing that it is the right thing allows them to choose. Embracing change requires a personal belief that the change is the best course of action.

Finally, people must be willing to step out of their comfort zone. This involves letting go of existing commitments and relationships. No matter how compelling the change may be, getting out of one's comfort zone takes courage. Without courage, people will not choose to act.

Table 6A: Criteria for balancing between push and pull

Characteristic	Definition	Implication
Rationale	Understand what the change is and why it's important	Fundamentally frames the decision to change
Practicality	Find the change workable	Creates the confidence to change
Belief	Embrace the change	Inspires people to change
Commitment	Provide willingness to get out of the comfort zone	Creates courage to act

The first two criteria rely on push mechanisms and the latter on pull behaviors. Traditional change management approaches are quite accomplished in the first two areas: communicating the rationale and practicality of the solution through different types of "push" strategies including town halls, intranet web sites, and training sessions. All these strategies are designed to communicate what has already occurred, often answering

frequently asked questions and rallying support for the predetermined solution. What is missing is instilling the desire to change and engendering the courage to act on that desire.

In order for this to happen, people need to feel a sense of control. They need to feel that they've been given an adequate forum where their needs and concerns have been understood and incorporated. They should believe that there has been a fair opportunity to be heard. In short, they need to be part of the decision-making process.

Allowing people to choose change

In the context of IT, in large automation projects one often finds a gap between the automated process and those expected by users. Eventually new process is imposed on people who are sceptical on the efficacy. The adoption of the system faces a natural resistance.

Rather than creating a solution and pushing it to stakeholder groups, project teams should make the stakeholders pull the solution by choosing to change. This push/pull balance is formed by creating four pressure points as stakeholders move through the various levels of commitment to a program.

1) **Build a foundation on change methodology -** Having a planned approach for change, involving external facilitation to guide project teams and stakeholders through the change process, provides a foundation upon which change programs succeed.

2) **Engage stakeholders in decisions and the vision behind them -** The project team identifies stakeholders to participate, defines the decisions to be made, builds the solution, and trains stakeholders. This push frames the decision for stakeholders and provides them with the means to bridge the gap between current and future states.

3) **Help leadership break down barriers -** Mobilize and align the stakeholder groups' top management around the program's priorities, vision, and solution. This creates an environment that fosters widespread involvement and participation in the program; it provides the leadership necessary to break down barriers and guide the program to success.

4) **Identify agents of change and make them champions within their groups -** When stakeholders participate in the creation of the solution and solicit widespread commitment from their peers, they become agents of change who pull the solution from the project team and make it their own. This is an important distinction. As senior leaders take on more responsibility for the success of the program, the project team has to learn how to work with much broader participation from stakeholders.

Working out the push-pull dynamics while optimizing IT

The primary metric driving IT optimization is reducing the ongoing maintenance portion of IT spending in order to divert funds for new IT investments. This metric continually helps IT meet emerging business needs while at the same time helps cap IT spending. Here, we analyze typical push-pull dynamics in some common changes that many businesses seek while optimizing IT.

Scenario 1: Transition to Agile development model from the traditional Waterfall model

Agile development is about adapting to changing business requirements, a departure from the Waterfall model where requirements and design are frozen before implementation. Agile development keeps stakeholders in sync using less documentation through frequent interactions with their peers and business units. Many stakeholders may find the Agile model to be loose and risky since it may be based on requirements that don't seem clearly planned or well-documented. On the contrary, Agile works around such risks by choosing iterative development where alignment with stakeholders is verified at each iteration.

Parameter	Grounds of resistance	Push characteristics		Pull characteristics	
		Rationale	**Practicality**	**Belief**	**Commitment**
Iterative design instead of frozen design	Risk of rework	Rapid development to meet changing business requirements; fewer surprises	Initially pick a module to pilot the process model and find benefits	Seek a document agnostic approach	Ask for frequent meetings to clarify requirements *(more MoMs)*
More tacit knowledge; less documentation	Lack of documented reference may create conflicts later	Saving time in ritualistic documentation	Build team system-enablers *(collaborative development tools like Collabnet or Microsoft Team System)*	Subscribe to the collaborative development tools usage *(more user registrations seen)*	Active use of collaborative tools *(more sign-ins and updates observed)*
Quality assurance	Resistance to fewer metrics and base artifacts	Focus on critical few metrics	Show that the metrics are still traceable but with less documentation *(e.g. module wise defect density)*	QA team focuses more on best practices than metrics *(such as introducing Scrum Master)*	QA standardizes best practices for cross-project use *(Rebuild Process Assets Library for Agile)*

Contd...

Scenario 2: Infrastructure cost optimization - Transition to Remote Infrastructure Management (RIM)

Remote service desk is an emerging trend (often referred to as Remote Infrastructure Management), where the service desk is remotely operated either as dedicated centers or by specialized providers. This entails interaction of business users and the IT organization with a remote team. The users submit requests and complaints as tickets, which are resolved from a remote site using service desk practices (like ITIL) and automated tools. This usually leads to restructuring part of the IT organization and often meets with initial resistance from business users and IT staff.

Parameter	Grounds of resistance	Push characteristics		Pull characteristics	
		Rationale	Practicality	Belief	Commitment
Business case	Questionable net business benefit	Noncore processes offshored to divert funds to core competencies	Cost savings from labor arbitrage, an industry-wide trend	Become a spokesperson	Seek a role in the new scheme
Efficiency	Net savings may not be same as labor arbitrage	Provider is a specialist in infrastructure services and therefore has a better economy of scale	Design a pilot to quantify benefits (e.g. start with remotely managing level 1 services)	Analyze net benefit (e.g. reduced ticket resolution time by remote desk)	Propose additional services for remote management (e.g. add level 2 services)
Cross-cultural synergy	Cultural disparity between remote location and client staff	Globalization has to deal with diversity	Cross-cultural sensitization (joint workshops and exchange sessions with pilot onsite and off-site teams)	See improvements from the sensitization programs (e.g. reduction in tickets reopened because of miscommunication)	Engage in developing best practices (e.g. categorize service requests in a way that overcomes communication gaps)

Source: TCS Global Consulting Practice – Research Desk

Capitalize on the innate desire for change

Despite our focus on deploying new processes and technologies, the ongoing business challenge is to get people to work in new ways. Traditional change management approaches often fail by attempting to get people to accept someone else's preconceived idea of what's best for them. We tend to overlook that it is human nature to want to change and to evolve ourselves. People rarely resist change when given the opportunity to address a problem by helping create a solution.

While most change management handbooks focus primarily on using communication to sell change, they tend to overlook the importance of seeing the signs of people wanting change. Ironically, traditional change managers will need to change their approach and leave behind their presupposition that people always resist change. By capitalizing on the natural desire for change and enabling people to pull the change, change management efforts can better succeed in meeting their business objectives.

References

- [1] Senge, Peter M. (1990), The Fifth Discipline, Doubleday/Currency,

- [2] William H Sheldon(1940) - The varieties of human physique: An introduction to constitutional psychology, New York: Harper & Brothers

- [3] Florence Littauer - Personality Plus: How to Understand Others by Understanding Yourself

- [4] COBIT® – http://www.itgi.org/cobit

New Frontiers

Enterprise Cloud Computing

Taking Off for the Cloud

Dr. Gautam Shroff
Vice President and Head, TCS Innovation Labs Delhi

How can enterprises leverage cloud computing in the near term?

Abstract

The term "cloud computing" has been used recently in a variety of contexts, including Software as a Service (SaaS), Infrastructure as a Service (IaaS), grid computing, parallel computing, and so on. The SaaS model has been popular for a number of years, especially with small and medium-sized businesses. Additionally, remotely hosted data centers have been a part of corporate IT for a long time. So exactly what about cloud computing is new, and what does it mean for enterprise IT?

This article demystifies cloud computing and shows how enterprises can start using publicly available cloud services for short-term benefits, factoring in possible risks.

What is cloud computing?

What differentiates the cloud from traditional infrastructure and systems is its scale. Google, Yahoo, Amazon, and eBay have applications running on thousands of servers. These applications are all highly parallel and accessed by millions of concurrent users. This scale is unprecedented in enterprise IT. At this scale, the usual assumptions no longer hold true - servers fail but applications cannot afford downtime, and using hot standby and disk mirroring simply does not work with a hundred thousand servers. As a result, the infrastructure has to be fault-tolerant and self-healing (incidentally, for all the talk of autonomic computing from the traditional computer industry, only Internet services have achieved true self-healing). Together, this kind of scale, this level of fault tolerance, and massive parallelism have resulted in a number of fundamental innovations, which are now becoming the focus of the computing community at large.

First, scale and fault tolerance, especially when dealing with hardware failures, requires development of completely automatic provisioning systems, which would sense and respond without any human intervention whatsoever, something that enterprise IT has never really seen. This capability of spontaneous provisioning and extension of resources is referred to as elasticity.

Second, provisioning has become a true Internet service. Infrastructure can be ordered, provisioned, managed, and used on demand, either by humans or as a self-managed system. This property was required by large Internet services like the Amazon shopping portal to drastically reduce deployment and system management costs, as well as time by automating traditionally human-managed tasks. In doing so, Amazon achieved economies of scale and now are extending this capability to enterprises.

Next, to exploit this scale of parallel computing required the evolution of new parallel programming paradigms and data organizations, like BigTable and MapReduce from Google. The importance of parallel computing is likely to grow rapidly in the future, driven by cloud computing on the one hand and multi-core processors on the other. Dedicating processor cores to virtual resources would increase the efficiency of virtualization. From an enterprise perspective, these technologies have the potential to drive radically new approaches for a wide range of analytical applications.

To summarize, cloud computing technology includes large-scale fault-tolerance, infrastructure on-demand, and a higher level abstraction for parallel computing known as *cloud programming*.

From the perspective of enterprise IT, it has been said that cloud computing is an opportunity to leverage some of the services available in Internet clouds. Yet, one may also attempt to optimize applications and IT operations being managed inside the organization by leveraging similar technology. This is called *private cloud*. The jury is still out on whether private or public clouds will eventually have more impact on enterprise IT, but we believe that there are significant opportunities to leverage public clouds in the near term. While building private clouds is feasible using technology available today, only very large enterprises with large-scale virtualization would be able to amortize workloads,

find economies of scale, or use low-cost power the way that public clouds do.

How can enterprises leverage public clouds in the near term?

As the technology around cloud computing has come of age, an important factor driving adoption is the "pay per use" model whereby cloud services are available with usage-based pricing such as CPU hours or storage days consumed. This helps businesses by eliminating capital expenses, while still retaining the ability to scale up with demand, bringing in a new economic model for IT investments.

Now, with virtualization and SOA both gaining adoption, we will soon have both dynamic resources and dynamic processes. Let's examine how we can build upon this combination of technologies.

Taxonomy of emergent services in cloud computing

Software as a Service (SaaS) - Software like Salesforce.com provides access to complete applications in the cloud.

Platform as a Service (PaaS) – Offerings such as Google's App Engine and Microsoft's Azure provide a development environment plus a built in multi-tier deployment architecture and automatic scaling and fault tolerance. The price for these features is that applications need to be migrated to the specific development environments provided by these vendors, which differ in many ways from the standard Java or Microsoft software stacks.

Infrastructure as a Service (IaaS) – Services such as Amazon's EC2 offer direct access to virtual servers in the cloud. While provisioning is immediate in IaaS, fault tolerance, deployment architecture and scalability are in control of, as well as the responsibility of, the users.

Finally, planning and system integration on cloud systems requires continuous monitoring, analytics and real-time response, which opens up a space for Cloud Consulting Services.
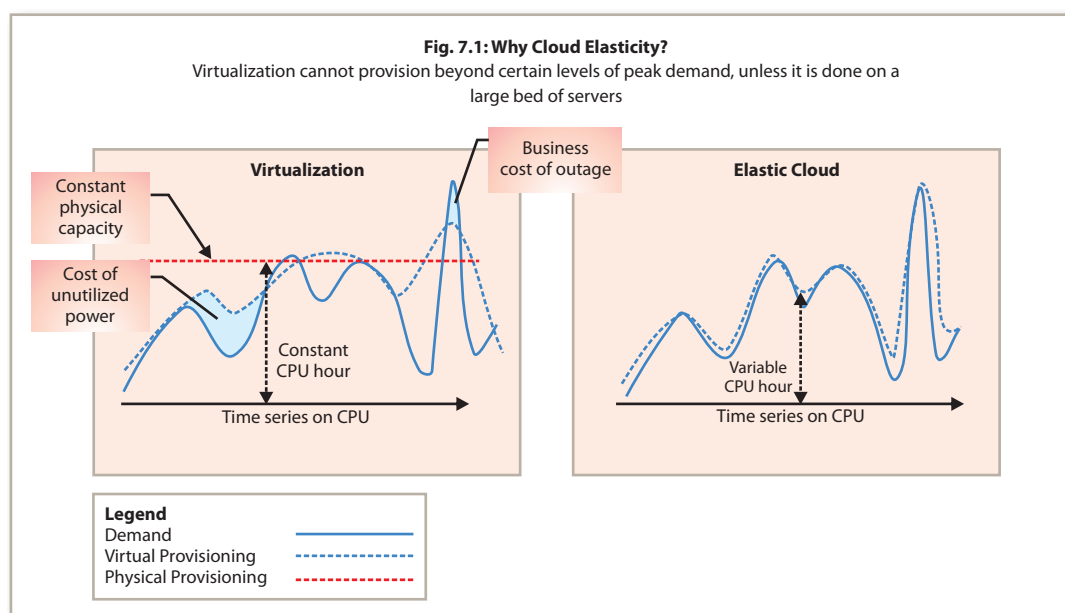
Cloud Service	Description	Examples
Cloud Services	Consulting and system integration on cloud	Advisory, security assurance, system integration, enterprise data management
SaaS	Publicly hosted application rendered on browser	Salesforce.com, WebEx
PaaS	Development platform, IDE (Integrated Development Environment)	Google App Engine, Microsoft Azure, TCS InstantApps
IaaS	Virtual servers, storage	Amazon EC2, Eucalyptus (Open source private cloud)

Cloud computing economics

Statistically significant data about the economics of the cloud has yet to emerge to clearly establish its value in terms of Total Cost of Ownership (TCO) for enterprise IT. Yet, on some important parameters, cloud computing indicates a real economic advantage.

Many of these cost advantages stem from cloud providers leveraging economies of scale. First, the provider's cost of purchasing network capacity and storage is 3 to 7 times cheaper when they have tens of thousands of servers versus the number at a medium-sized enterprise data center. Second, providers have been able to use high levels of automation in order to amortize the cost of administration over a larger number of servers, an estimated gain by a factor of 7. Also, the cloud providers are all leveraging significantly lower power costs (by a factor of 3) by locating their data centers in power-producing regions, such as Idaho and Washington in the United States. Finally, cloud providers are able to enjoy far higher degrees of server utilization, in the range of 60-80% with large scale virtual provisioning amortized across many customers. Not the least, it is important to understand that leading cloud providers, such as Google and Amazon, developed these capabilities for other businesses (search and retail, respectively), and so there was marginal investment involved in extending infrastructure to cloud services, hence opening up a new business model.

From the enterprise user's perspective, cloud computing provides the illusion of infinite capacity *(Fig. 7.1)*- the ability to rapidly leverage additional capacity when needed, thereby avoiding upfront investments for peak load and instead paying only for what one consumes. In contrast, the need to provision for peak capacity in private data centers constrains the utilization levels that can be achieved. Further, the often similar operational profiles for the most critical and resource-intensive applications limit the improvements possible by sharing capacity between applications using virtualization. As a result, most data centers operate at average utilizations of 20% to 40% - very low utilization rates indeed and certainly not cost-effective.

Fig. 7.1: Why Cloud Elasticity?
Virtualization cannot provision beyond certain levels of peak demand, unless it is done on a large bed of servers

Source: TCS Global Consulting Practice - Research Desk

Consider this: running an 8-core server costing US $2000 for two years works out to roughly 128 CPU hours per dollar. Equivalent usage on the Amazon cloud for two "small" dual CPU servers would cost $2.56 for the same CPU hours. At first, it seems that the cloud is more expensive. However, if we factor in low utilization of most servers in the data center (say 40%), whereas cloud computing power can be easily scaled up or down with demand, therefore enjoying a far higher utilization (say 80%) - we see that the cloud cost becomes competitive again in terms of net payout. Adding the cost of power and cooling, which normally doubles the cost on average, brings out a clear advantage in the cloud. The difference is more striking in storage when one considers that cloud storage is automatically replicated at least 3 times, thereby eliminating backup and archival costs. Finally, network bandwidth from cloud providers is far cheaper than what is available to most enterprises.

Even if one manages to improve utilizations in the data center through virtualization, critical applications will still need to be provisioned for peak load. Typical transaction processing loads can vary by factors of 5 or 10 in the course of a day, with even higher factors needed for peak times of the year. Nonetheless, the demand profile for many business applications within an enterprise is likely to be similar, limiting the amount of capacity sharing that can be achieved by virtualization. Thus, even with virtualization one would still need to maintain excess physical capacity to handle peak load, which would remain idle when demand is normal. On the other hand, using a cloud provider, computing resources can be scaled up and down in minutes, thereby significantly lowering infrastructure costs without compromising performance.

First steps toward the cloud

If we agree in principle that the cloud can be cheaper and of course faster due to automation of provisioning, how can an enterprise begin leveraging the cloud?

Consider that there are still valid concerns with cloud computing from an enterprise perspective, including data confidentiality, compliance, lock-in, auditability, and software licensing. Enterprises are still wary of placing production or sensitive data in the cloud, since current cloud offerings are essentially public networks and hence exposed to more attacks. While there are no fundamental obstacles to making a cloud environment as secure as an in-house data center, this requires careful planning using encrypted storage, virtual LANs, and network middleware. Further, it is likely that at least some cloud providers will begin to offer these levels of security in the future. Regarding software licensing, most independent software vendors (ISVs) have yet to come to terms with usage-based pricing. So even if one leverages cloud infrastructure, the need to pay for licenses up front can obviate some of the cost advantages of the cloud.

An often cited obstacle is large-scale data transfer. For example, transferring even a few hundred gigabytes of data over 20 Mbps can take a couple of days; for larger data sets, it is impractical. Our view is that this is not a problem since data in disks can be physically shipped, which is faster, cheaper, and more reliable. Amazon has recently announced support for physical data shipment for both import and export functions.

Even in the light of current limitations, we see the following three immediate opportunities for leveraging public clouds *(Table 7A on next page)*.

1) ***Development and test environments in the cloud -*** The infrastructure needs for development and testing enterprise applications differ from those of a production environment. Data security requirements are lower for such environments. At the same time, the cloud allows for a high degree of variability (the ability to configure diverse environments for development and testing) and volatility (the ability to retire and restore infrastructure). Servers can be set up instantly as virtual machines for each new development project environment, many of which can be released once the application is put into production. Further, the time required for provisioning and configuring a development environment can often incur significant overhead in large organizations due to procurement and service desk procedures. Leveraging cloud services for development and test servers is therefore cost-effective, low risk, and agile. It also potentially improves business agility by improving the response time of IT to new business needs.

Now consider performance testing. Stress testing an application on production hardware is difficult, especially early in the development cycle, simply because such an environment is often not available. Using the cloud, one can create a production-class infrastructure that can be provisioned on demand and disbanded quickly on completion of the testing. Further, such an environment can be configured to simulate diverse production scenarios, which is far more difficult to do easily on-premise.

2) ***Disaster recovery in the cloud -*** Maintaining a disaster recovery site that can be rapidly brought into production for business continuity requires replicating hardware infrastructure at least partially, which in normal circumstances might not be utilized. It is possible to store a virtual image of the production environment in the cloud so that actual servers can be provisioned only when required. Similarly, production data backups can be physically shipped to a location near the cloud provider on a regular basis and loaded into the cloud only when needed. Alternatively, updates can be replicated regularly over the network but exported to disk remotely rather than locally. Such cloud-based disaster recovery can be significantly cheaper than replicating infrastructure while offering similar levels of protection.

3) **_Analytics in the cloud -_** New programming paradigms, such as MapReduce and BigTable, were developed for the cloud to enable massively parallel computations while automatically compensating for hardware and software failures. These models are especially suited for analytical tasks such as data warehousing, customer segmentation, manufacturing optimization, and so on. Normal enterprise analytics can also benefit greatly from elasticity. Enterprises need to run analytics on customers, supply chains, manufacturing operations, and the like on a daily basis. Such jobs may run for a few hours on dedicated hardware, and occasionally require even larger capacity, which is often unpredictable, necessitating over-provisioning. Using cloud computing, the required infrastructure can be provisioned for a theoretically infinite capacity and disbanded quickly thereafter.

Note that analytics often involves large volumes of data. Here too, physically shipping data and transferring only small volumes over the network is a way to get around this issue.

Table 7A: Cost-Benefit Analysis for Three Use Cases

Use Case / Decision Parameters		Commissionability	Variability	Volatility	Elasticity
Definition		*How quickly and easily can one set up the environment*	*How quickly and easily can one alter configuration*	*How easily and cost-effectively can one retire and restore environment*	*To what extent can the environment scale up spontaneously to demand*
Development & Testing Environment	Cost avoided	Fewer machines required to emulate diversity of production environment	Quick change in configuration supports diverse test bed (e.g., performance versus memory)	Quick restoration of environment (image) in future maintenance and enhancement	Better productivity during load testing overcoming unpredictable capacity shortage
	Cost added	Network consumption and bottleneck in staging environment on cloud		Cost of image storage	
Disaster Recovery	Cost avoided	Excess provisioning	Scaling up when contingency is anticipated	Contraction of capacity during idle time	Spontaneous scaling up prevents excess provisioning
	Cost added	Compliance costs (auditing and implementation)			Spontaneous elasticity would need security assurance
Analysis	Cost avoided	Full-time licensing of expensive tools	Tuning to data volume and analysis intensity	Archival of datamart and algorithm image for future reuse	Capacity for analysis is unpredictable (theoretically infinite)
	Cost added	Staging costs (moving data)	On-premise datamarts require appliances to connect with business applications	Recurring storage cost	Security assurance costs

Source: TCS Global Consulting Practice - Research Desk

A glimpse at adoption: *Cloud quick win at a leading pharmaceutical company*

A global player in the pharmaceutical industry is using public cloud computing in two use cases. First, its developers in drug research are extremely pleased with the ability to rapidly provision new servers whenever required since this has improved their effectiveness and agility. Second, bioinformatic computations require heavy computational power, and in a research environment, it is difficult to predict and plan for the maximum capacity needed, which in theory is infinite. Our client regularly uses the public cloud platform for this purpose, paying only for what the organization actually uses while still being able to do extremely large simulations whenever required.

Cloud consulting: An emerging discipline

New cloud platforms, infrastructure services, and applications are emerging every day. The industry is poised to see disruptive IT adoption models with a new set of economics. Currently, the cloud services landscape sees some players pitching in with niche competencies. Amazon can be seen as an attractive option for elastic infrastructure. Google, on the other hand, distinguishes itself with an interesting development platform supported with APIs that connect its SaaS services (like Google Maps and Docs). Microsoft provides a wide solution landscape in the form of a hybrid of on-premise and hosted platforms. Enterprises are exploring setting up private clouds using open source tools to achieve large-scale virtualization. In the future, we will see a mix of private and public services blended together. At the same time, we also see new aspects of security and compliance surfacing in the cloud.

With most players offering different competencies, enterprises need to take a best-of-breed approach in adopting cloud computing. Businesses need to explore, incubate, and plan a roadmap for cloud adoption in terms of their business models and the emerging cloud economics. This makes cloud consulting services for system integration, service management, and technology advice especially important to connect to the evolving ecosystem and bring it to life for an enterprise.

Pros and Cons of Cloud Storage

Prateep Misra
Research Manager, TCS Innovation Labs for Open Storage

Is the cloud ready for corporate data storage?

Abstract

Cloud storage is becoming mainstream. Storage costs consume a significant proportion of IT infrastructure spending. Interestingly, the bulk of the cost goes towards maintaining storage infrastructure, which includes power, cooling, and other data center costs. The advent of cloud storage, with its "pay-per-use" proposition, gives us a compelling storage alternative.

However, the current state of cloud storage technology has yet to overcome some shortcomings. A storage strategy should factor in both advantages and disadvantages to determine what types of data should be moved to the cloud. This article provides guidelines for making that decision.

Storage strategy: More than mere economics

There are still some questions about the practicality of wide-scale enterprise storage in the cloud. Two factors must be considered: economics and technical capability to support enterprise IT. The cloud is an emerging facet of IT with economic benefits that cannot be ignored. Storage cost is mushrooming and business has little control over it. Storage optimization can be more difficult than consolidating applications since there is less visibility into data consumption than there is into applications which are in use. The cloud is economically attractive, but how will it eventually overcome the perceived technical challenges? Cloud storage helps businesses adopt the "pay per use" model, which reduces, if not eliminates, capital investment in storage. The question is, What types of storage make sense in the cloud?

Our study suggests that both the economics and feasibility of storage may vary across classes of data determined by usage factors. The benefits of flexibility must be considered in addition to the value proposition in economic terms.

Why cloud storage?

Cloud computing refers to applications, platforms, and infrastructure delivered as services by providers who run these services in their own data centers. A public cloud storage service like Amazon S3 offers storage on a "pay-per-use" model. On the other hand, private cloud storage refers to a storage environment that emulates the cloud within captive data centers. Technology for private clouds has grown along with virtualization technology. Note that this article discusses only public cloud storage services.

The cost of storage in the cloud is comparable to internal storage, sometimes even less expensive. The ballpark acquisition cost of storage for non-mission critical data is $1/GB per month, considering power, cooling, and space costs over a storage lifespan of 36 months. Moreover, for such data, storage is replicated twice for data durability. Considering these factors, total cost of ownership (TCO) for internal storage over 36 months is $8 per GB, which works out to roughly $0.22/GB per month. Cloud storage is priced at $0.17/GB per month, provided the initial data, if voluminous, is uploaded by physically shipping the data. There is usually an added cost for incremental uploads and downloads (usually about $0.20/GB). The breakdown of these calculations is provided in detail later in the article. However, it is apparent that the economics of cloud storage depend on the usage of data, so let us explore the technical and usage factors that make the cloud a viable and profitable option *(Table 8A on next page)*.

Table 8A: When is Cloud storage useful?

Parameter	Conventional/ Virtualized	Cloud	When is cloud useful
Capacity scalability	Physical or virtually provisioned	Capacity spontaneously fetched on demand	When storage requirements are unpredictable (e.g. analytics)
CAPEX	CAPEX 30% of TCO OPEX 70% of TCO	Variable OPEX - "Pay-per-use"	For data sets with long life span
Access mechanism	Tightly coupled (not fault-tolerant unless on RAID)	Loosely coupled in SOAP* or REST** services (can be message-queued)	For applications that can connect to SOAP/REST storage (e.g. modern knowledge management systems such as Microsoft SharePoint)
Identifier	Physical/Network/Virtual Path	Globally unique namespace	When storage is accessed by distributed applications (as is the case with XBRL data)
Tenancy	Single tenant – allocated to consumer else read only Multi-tenancy can be achieved with certain virtualization technologies	Multi-tenant - secured logically through partitioning and replication	For personal storage
Recovery	Requires internal recovery tools	Automated recovery (self-healing)	For permanent data storage, or mission critical data
Service management	Depends on maturity of internal configuration management processes	Cloud storage vendor provides online service desk; configuration usually automated.	When internal storage service cost is high (can be analyzed from current service desk incidents)
Usage	Data lookup and updates have negligible marginal cost	Data lookup and updates have costs in form of upload and download cost per GB	When data stored is greater than data fetched; large database with contextual lookups (e.g. knowledge management)
Availability	If availability is limited by capacity, provisioning can reduce downtime to the extent free physical capacity is available	Availability is normally not constrained by resources (disk space) since capacity can scale up elastically	When storage capacity is the primary constraint to availability

* Simple Object Access Protocol   ** Representational State Transfer

Source: TCS Global Consulting Practice - Research Desk

The practical side of the cloud

While the cloud offers some advantages for enterprise storage, a set of constraints emerges that could be critical and costly if not included in the storage roadmap.

Cost and performance of bandwidth is often a limiting factor in cloud storage. Adding to this, security concerns and compliance requirements make businesses wary, despite the credentials of cloud providers such as Amazon, Microsoft, and Google.

Five technical factors should be considered when developing a cloud storage strategy: data lifespan, cost of initial data loading, cost of usage, application architecture and security.

1) **Data lifespan -** The longer the data is stored, the better the economics of cloud storage. The lifespan of data falls into three broad stages, referred to as tiers *(Fig. 8.1)*:

Tier 1 –The data is frequently used

Tier 2 – The data is considered active for reference purposes, but used occasionally

Tier 3 – The data is considered inactive and is archived for possible future references

Fig. 8.1: Typical distribution of enterprise storage across tiers when optimized

In the course of this discussion, we would find that Tier 2 data are the best candidates from cloud, with some exceptions.

As data is exchanged across computing nodes and storage nodes, network bandwidth and latency (the speed of transmission) determine performance. In storage clouds, location of the data is not known beforehand. As a result, calculate the worst-case impact of latency and bandwidth to ensure that performance requirements are met. Accessing storage over a wide area network using interfaces like SOAP and REST is much slower than conventional access over a storage area network that uses ethernet or fibre channel.

Source: TCS Global Consulting Practice – Research Desk

2) **Cost of initial data loading -** Opt for physical transit whenever possible. There can be significant bandwidth cost and time consumed in transferring legacy data. For instance, 1 TB of data would take a week to upload with 100 MBPS bandwidth and 80% utilization. Some service providers allow users to send disks or tapes physically to them for bulk data transfer, which is a worthwhile option.

3) **Cost of usage -** Choose storage for which information provisioning (real-time backup) is higher than usage.

Cloud storage providers charge for download and upload in addition to the rental on storage. This is charged per GB of upload or download. Hence, data that is frequently updated or downloaded may be costlier in the cloud. Therefore, delineating data with high storage costs but less frequent usage is an important factor to consider in finding the sweet spot for cloud storage.

Consider a scenario in which a business has an active document management system with uploads and downloads every minute. After a span of time, such databases have huge stockpiles of Tier 2 data in the form of old but active documents. The probability of a document that is 6 months old being downloaded in a month could be less than 10%. In such a case, it makes business sense to transfer the Tier 2 data experiencing low
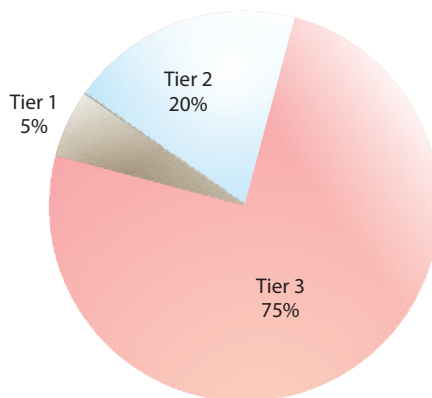
download frequency to the cloud. The system should have a qualification policy that transfers such data automatically (in fact, most document management systems provide such usage analytics). This reduces total cost of storage ownership because the internal infrastructure need not be configured for high redundancy.

The business case for storing data in the cloud is driven by usage economics *(Table 8B)*:

Table 8B: When should Tier 2 data move to the cloud?

	Unfavorable business case	Favorable business case
Cloud cost breakdown	Cost of storage for 12 months = 0.17 x 12 = $2.04 Cost of 12 downloads (data needed 12 times per year) = 0.17 x 12 = $2.16 Cost of 12 uploads (data updated 12 times per year) = 0.10 x 12 = $1.20	Cost of storage for 12 months = 0.17 x 12 = $2.04 Cost of 2 downloads (data read twice per year) = 0.17 x 2 x12 = $ 0.34 Cost of 1 upload (data updated only once) = 0.10 x 1= $ 0.10
Cloud TCO	$ 5.4 per GB per year	$ 2.48 per GB per year
On-premise cost	Storage costs for 12 months = $0.33 per GB (assuming a lifespan of 3 years and hardware acquisition costs of $1 per GB) Operating costs (power, data center overhead, usually twice acquisition costs for a 3-year lifespan) = $0.66 per GB TCO per replication = $0.33 + $0.66= $0.99 per GB per year Cost of provisioning (one original and two replications) = 2x $0.99= $.1.98 per GB per year + $0.99	
On-premise TCO	$ 2.97per GB	

Source: TCS Global Consulting Practice – Research Desk. Data sources: TCS Innovation Lab, UC Berkeley, and Amazon Web Services

4) ***Application architecture -*** Avoid cloud storage when an application's architecture is not conducive. Many applications may not directly support the data access mechanisms of cloud providers (SOAP or REST interfaces). Depending on the architecture of the legacy application, there could be significant reengineering and performance-testing costs associated with connecting these applications to cloud storage. Some providers offer on-premise "converters" (NAS gateways supplied as an on-premise appliance) to get around this issue. These converters can also cache data in transit so that bandwidth and latency do not impede performance.

For new application development, the application architecture and programming model should be carefully chosen to ensure good performance using cloud storage. New applications that are data-intensive but small should move computation of data to get around bandwidth issues (for example, the MapReduce programming model by Google in which computation is handled by the database platform in the cloud).

Moreover, while cloud storage provides high fault-tolerance and elasticity (spontaneous increase in capacity as demand picks up), it might require a different architecture for session-intensive applications to use these elasticity features. Elasticity features in some cloud services provide limited state management capabilities (the ability to hold the session data while the connection is being transitioned from one server to another). Session data is data which the user sessions need to hold while navigating from one application page to another; for instance, in online shopping, the items added to the shopping cart are held from page to page until the user checks out and completes a purchase. A mission-critical application accessing data stored in the cloud should keep session data either in the cloud database or on the client (the browser in a thick client application). In general, developers of database applications that access data in the cloud should explore moving the entire application to the cloud.

5) *Security* - One important difference between cloud storage and conventional storage is that the precise data center is not always known in the cloud (this is how cloud providers support high elasticity—provisioning of capacity may span multiple data centers). It's difficult, if not impossible, to guess where data actually resides and how and when it moves. Data can move across political boundaries and perhaps violate regulatory requirements. Although data centers for any one service provider are usually located within the same geographic borders, this factor should be considered in light of compliance requirements.

In addition, most clouds do not provide encrypted storage. It is typically the customer's responsibility to ensure that cloud data is encrypted.

Careful due diligence is required before storing mission-critical or personally identifiable data in public clouds. The customer must carefully review the storage service for governance issues such as backup, recovery, offsite storage, physical location of data, and removal of data after contract termination as well as security and privacy of data on shared servers. Companies that use public clouds must be aware that they, not the cloud storage provider, are considered the custodians of the data. In a legal context, the customer has the obligation to produce data stored in the cloud, not the cloud provider.

For instance, data stored in the cloud may be used as evidence in legal proceedings in which the customer is involved. The customer is liable to preserve and produce records stored in the cloud during court proceedings. Cloud storage services must therefore include the ability to search, retrieve, and validate the forensic integrity of data so that it is admissible as legal evidence.

Our research suggests that cloud storage could affect a significant proportion of COBIT (Control Objectives for Information and Related Technology) control scenarios. It is a methodology to evaluate and deploy security and risk controls across IT infrastructure. As a result, it is important that businesses carry out an assessment before moving storage to the cloud.

From a business continuity perspective, it is not advisable for data requiring rapid recovery to be stored in the cloud since the actual data center may be either unknown or remotely located. During contingencies, one may need to physically restore data from an alternative location (such as a mirrored data center).

While using cloud storage has distinct merits, there is confusion about the right cloud storage strategy. A few rules of thumb can be helpful here.

General guidelines for implementing cloud storage

We find that Tier 2 data are best candidates for cloud storage. For example, document management systems contain a section of documents that are rarely used; it can be bulky with the digital media content and large PDF files. At the same time, these files need to be kept active to support search engines. Such system can have a policy to identify Tier 2 content of this nature and transfer it to the cloud.

Tape drive storage is normally used for Tier 3 data of historical importance that could be recalled in instances like an audit. Accounting records older than five years would normally reside in tape drives. Tape drives are much less expensive than the cloud (TCO is $0.01/GB per month over a 5-year span) and offer significantly faster data transmission. However, businesses have started considering storing Tier 3 data in the cloud to make it disaster-proof. In fact, meeting disaster recovery requirements is a good way to start using cloud storage since there is less complexity in migrating Tier 3 legacy data.

To summarize, Tier 2 data qualifies for cloud storage since it lowers TCO by avoiding the cost of power and data center operations. Tier 3 data, on the other hand, could be considered for the cloud to meet disaster recovery requirements.

Table 8C on next page provides some guidance for making decisions about cloud storage in terms of technical and economic factors; it considers three storage-intensive use cases: business intelligence (enterprise reporting and analytics), desktop storage, and email.

Table 8C: Dependency of applications on storage parameters
(The moon scale indicates degree of dependency)

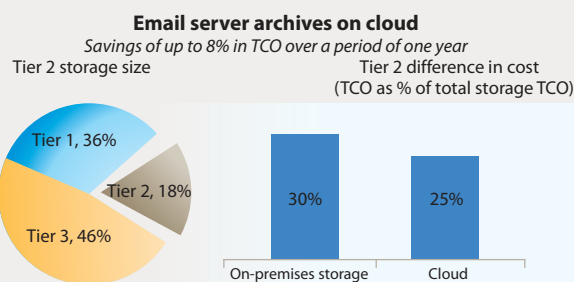| Storage	Performance Factors	Data Lifecycle		Loosely Coupled		Frequency of Use		Size of Transfer		Recommendations
		Impact	Reason	Impact	Reason	Impact	Reason	Impact	Reason	
Business Intelligence	Analytics	Medium	Easy flushing of old data marts	Low	ETL is usually loosely coupled (uses batch processes)	Medium	Datamarts periodic; snapshots frequent	Medium	Initial staging large	Data marts on cloud Queries can be serialized and queued in SOAP
Business Intelligence	Standard reporting	High	Usually periodic; historical importance	Medium	Tied to applications unless in SOA	High	Frequent updates	Low	Transactional updates	Not advisable on cloud unless the architecture requires large amounts of intermediate and static data
Desktop	Business artifacts	High	Files references are ad-hoc (not systematic)	High	Files by nature are loosely coupled	High	Day-to-day use	Medium	File wise transfer (incremental transfer)	Advisable on local SAN or NAS
Desktop	Incidental files	Low	Obsolete files pile up (often overlooked)	High	Files by nature are loosely coupled	Low	Usually one-time use	Medium	File wise transfer (incremental transfer)	Should be filtered and archived on temporary cloud storage (such data consumes a lot of space) Integrating with desktop/enterprise search engines is a good way to filter
Email	Current online	High	Archived frequently	High	SMTP }(email protocol) is loosely coupled	High	Every moment	Low	Transactional updates	Not advisable on cloud unless email is on SaaS
Email	Server archive	High	For compliance	High	Can be queued	Medium	For analytics and historical importance (compliance)	High	Archived email consumes significant storage	Cloud should be considered since this data is bulky and carries high operating cost Frequent recall and tracing makes tape drive alternative inconvenient

Legend: Impact ● High ◑ Medium ○ Low

Source: TCS Global Consulting Practice - Research Desk

Estimate of savings from storage of Tier 2 data in the cloud
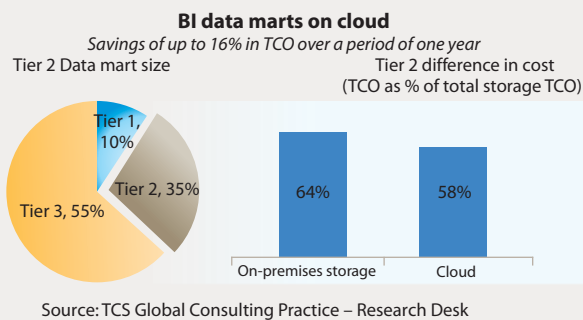
Archiving email in the cloud

Tier 2 email data typically includes archives for the last year kept on a backup storage server as well as in active storage, usually accessible via fibre channel so that data is available to applications that are integrated with email. The mail server fetches the data on a transaction-by-transaction basis with a set of email records being fetched at one time. The advantage of archiving is improved performance on the primary mail server while making older messages still available on demand. The cloud is an option for this type of active archive storage. However, initial transfer of the data (which could be several TBs) should be physically shipped to the service provider, and the email server should be configured to connect to cloud storage using Simple Object Access Protocol (SOAP) or a Network-attached storage (NAS) gateway.

Email server archives on cloud
Savings of up to 8% in TCO over a period of one year

Tier 2 storage size

Tier 2 difference in cost
(TCO as % of total storage TCO)

Tier 1, 36%
Tier 2, 18%
Tier 3, 46%

On-premises storage: 30%
Cloud: 25%

Source: TCS Global Consulting Practice – Research Desk

BI datamarts in the cloud

Storing BI datamarts in the cloud is a good option when snapshots of datamarts are taken to enable quick queries. This type of data is usually bulky with high redundancy. Datamarts are obsolete within a short timeframe because fresh datamarts are generated and thus should be considered Tier 2 data. Storing and retiring datamarts in the cloud is economical since the cost of such storage on-premise is high. Moreover, to overcome performance issues on datamarts in the cloud, ad hoc queries can be queued.

BI data marts on cloud
Savings of up to 16% in TCO over a period of one year

Tier 2 Data mart size

Tier 2 difference in cost
(TCO as % of total storage TCO)

Source: TCS Global Consulting Practice – Research Desk

Cloud storage is inevitable

Cloud storage has matured adequately in terms of TCO and technology for businesses to seriously consider placing some types of data in the cloud. Two primary concerns remain: performance and security of intellectual property.

Regarding performance, cloud storage technology has more to offer in the near future. For instance, Ethernet technology is moving to the Internet, a move that may ameliorate many performance and technical issues. Also, bandwidth is increasing at a higher rate today. While many of these developments are not yet mainstream, the current state presents enough of a business case to get started with cloud storage.

As far as security is concerned, businesses should also consider traditional security problems that may be better addressed in the cloud. For instance, internal data pilferage, one of the prevalent forms of security breach today, can be better managed in the cloud because of interfaces that provide more traceability. Regarding concerns about data crossing political boundaries, cloud providers need to address this question, explaining how they would offer value-added managed services for compliance.

As we store more data in the cloud, the cloud will have more in store for the enterprise, both in terms of economics and flexibility.

Enterprise Risk Management

XBRL and Real-Time Risk Analytics

Prabha Bob
Principal Consultant, Financial Services, Global Consulting Practice

Sachin Shinde
Consultant, Financial Services, Global Consulting Practice

XBRL is set to emerge as the much awaited answer to real-time risk analytics within the enterprise

Abstract

The financial downturn has exposed the vulnerability of businesses to diverse risks. Enterprise Risk Management (ERM) has become a prime business activity, and regulatory compliance is only a small part of it.

To make ERM truly "risk-aware," large investments are being made in a new breed of analytics and tools, and the cost is growing.

Meanwhile, the eXtensible Business Reporting Language (XBRL) standard for financial reports seems to be an imposition on businesses. On the contrary, it is set to emerge as the much-awaited answer to real-time internal analytics for risks. In other words, XBRL can be used beyond financial reporting, making it an internal information medium.

This article will show how to use XBRL to find new, cost-effective ways to analyze risk.

ERM needs adaptive analytics

Enterprise Risk Management (ERM) is a framework designed to continually sense and mitigate business risk. It deploys a governance model supported by controls in processes to avoid crises caused by internal or external factors. Quality of ERM is based on the coordination between the governance model and the underlying risk management information system.

Recently, there has been renewed focus on ERM, especially after the economic downturn exposed flaws in the ability of many enterprises to sustain themselves. Therefore, regulatory oversight has been more rigorous and new compliance measures have been introduced. For instance, the Securities and Exchange Commission (SEC) and Standard & Poor's both require that businesses be explicit about their ERM processes.

A key concern about ERM is the gap between the governance model and the control mechanisms in business processes. This gap can best be seen by looking at the underlying analytics *(Fig. 9.1)*.

Fig. 9.1: How regulations appear on the ERM analytics radar
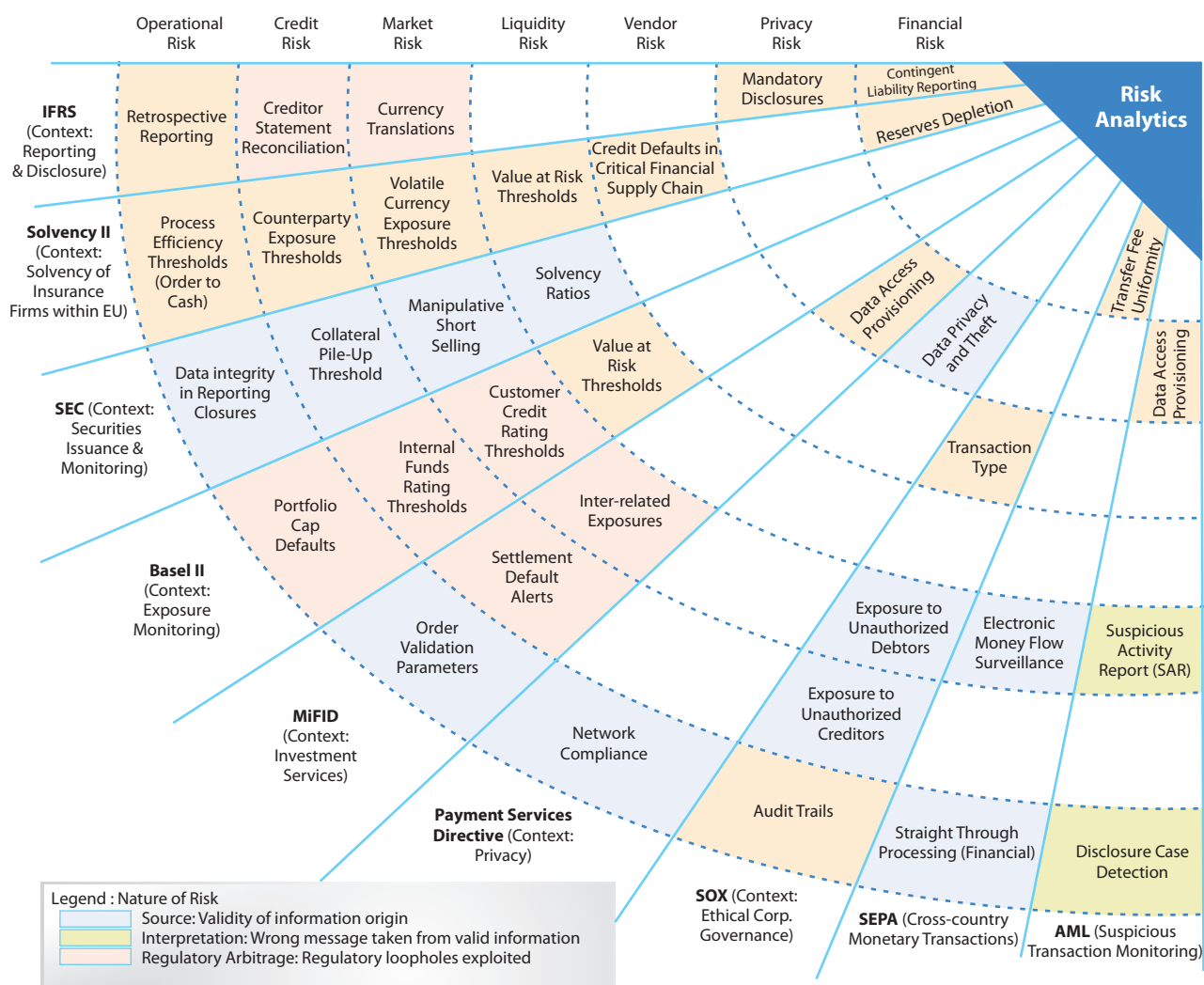
Source: TCS Global Consulting Practice - Research Desk

In the light of ERM, information systems need to be re-evaluated, which would ask for new investments in analytics. However, the task does not end there. The speed at which regulations are changing with new forms of risk, making IT adapt is a continuous process. Information systems must be built in a way that can adapt to new parameters as they emerge.

Risk-aware analytics that cost less is what's needed right now.

Barriers to risk-aware analytics

Risk analytics is usually part of Business Intelligence (BI), thus subject to the traditional problems associated with it.

Clusters of disconnected applications are very common in even the most mature businesses due to frequent process changes, ad hoc IT upgrades, and the merger/acquisition history within the company. Many times, old applications are maintained so that they can be ready to produce legacy data when required in compliance with laws such as the Sarbanes-Oxley Act (SOX).

With myriad applications having heterogeneous architectures, information flow depends on a patchwork of batch processes to support analytics. This arduous activity prevents companies from proactively adding risk parameters and changing algorithms unless compliance issues intervene. Therefore, internalizing many of the principles of risk management beyond meeting compliance requirements sounds like a philosophical question to both business and IT.

For example, if an investment banking firm needs to monitor large customers for failure to pay (known as *default risk*), it would probably have to delve into individual applications for each product line and connect customer lists across them to consolidate direct and indirect exposure.

Adaptive analytics for spontaneous risk triggers

In a dynamic environment (where new business models and marketplaces emerge), new risk parameters get introduced and old ones are revisited. This requires analytics to be adaptive. This aspect is often ignored in centralized analytics, where data is pooled from multiple parts of the business with predefined algorithms. It tends to miss the risk parameters emerging in less scrutinized sources (sections of the environment, which we call spatial blind spots). As an example, in the mortgage crisis, there was lack of visibility on Special Purpose Vehicles (SPVs), the subsidiary entities created to securitize defaulted assets. Such a syndrome calls for federated analytics, where each business unit and division has its own local intelligence yet they are coordinated.

On the other hand, despite analytics being federated, it would still rely on human intervention to periodically mine data and run reports. Sometimes the periodicity could be costly if it misses interim triggers related to imminent risks. At least, a part of analytics would need to be real-time to raise spontaneous alerts. For instance, a rise in fuel pricing in a region could trigger a rise in global oil prices, which may disrupt the supply chain plan of the business.

Maturity levels in ERM analytics

Source: TCS Global Consulting Practice - Research Desk

Such analytics are often very different from reports provided by ERP systems since accurate updates of risk positions are not discretely captured using standard accounting procedures.

However, today the outlook has changed. Global recession lessons are biting at everyone's heels. How to support real-time risk analytics while avoiding a wide-scale IT overhaul is an important question.

Meanwhile, regulators have come up with a new standard for financial reporting – the eXtensible Business Reporting Language, or XBRL. This apparently entails change in the reporting systems and their supporting applications, but the implications of this standard go beyond mere regulatory reporting.

The XBRL "imposition"

XBRL is a form of the XML standard specific to financial reporting and it is now being mandated by the SEC. The primary goal is to have a common reporting format for financial reports so that they can be read by computerized systems. For instance, the regulatory body can easily evaluate the net non-performing assets (NPA) across an industry by consolidating reports from multiple businesses.

It may seem that XBRL is only here to support regulatory bodies while adding to the compliance cost incurred by businesses. This perception will change once the true potential of XBRL is realized and its importance to the internal analytics of the enterprise is recognized.

Information federation with XBRL

Until now, we have seen that analytics for ERM needs to merge with the mainstream information flow of the business, making it adaptive to new parameters. However, applications and data sources are not joined in a manner that easily supports this paradigm.

Instead of standardizing data exchange channels across applications (which may require integration efforts or SOA services), we should explore how disparate data sources can be standardized in a cost-effective way *(Fig. 9.2 on next page)*.

One of the tenets of XBRL is data federation. That is, data and reports may be distributed across multiple locations, but the data remains traceable. For instance, XBRL could facilitate drill-downs from customer balances to transactions with the customer from

> **Why regulations introduced XBRL**
>
> XBRL provides an opportunity to store reports in a data structure rather than a document structure. This facilitates the ease and speed of attribute-based data analysis within and across organizations, which in turn provides substantial intelligence to regulators and the markets. The standardized structure is called XBRL taxonomy. It defines tags for different elements in the financial reports. The most important one is a tag to represent the code of the entity (customer, supplier, or bank) involved in the financial figure. This code is standardized for each entity across the industry. With the standardization, XBRL helps the regulatory bodies do industry-wide consolidation and find risk patterns that cut across companies so that it can be traced back to specific business entities.

one application to another. This could help gauge customers' solvency. Such drilling down into details, even when the data resides elsewhere, is a mechanism essential in analytics and a boon when applications are disconnected. The question is, How can XBRL be implemented to achieve this goal cost-effectively?

Fig. 9.2: A functional abstraction of XBRL taxonomy for federated risk analytics
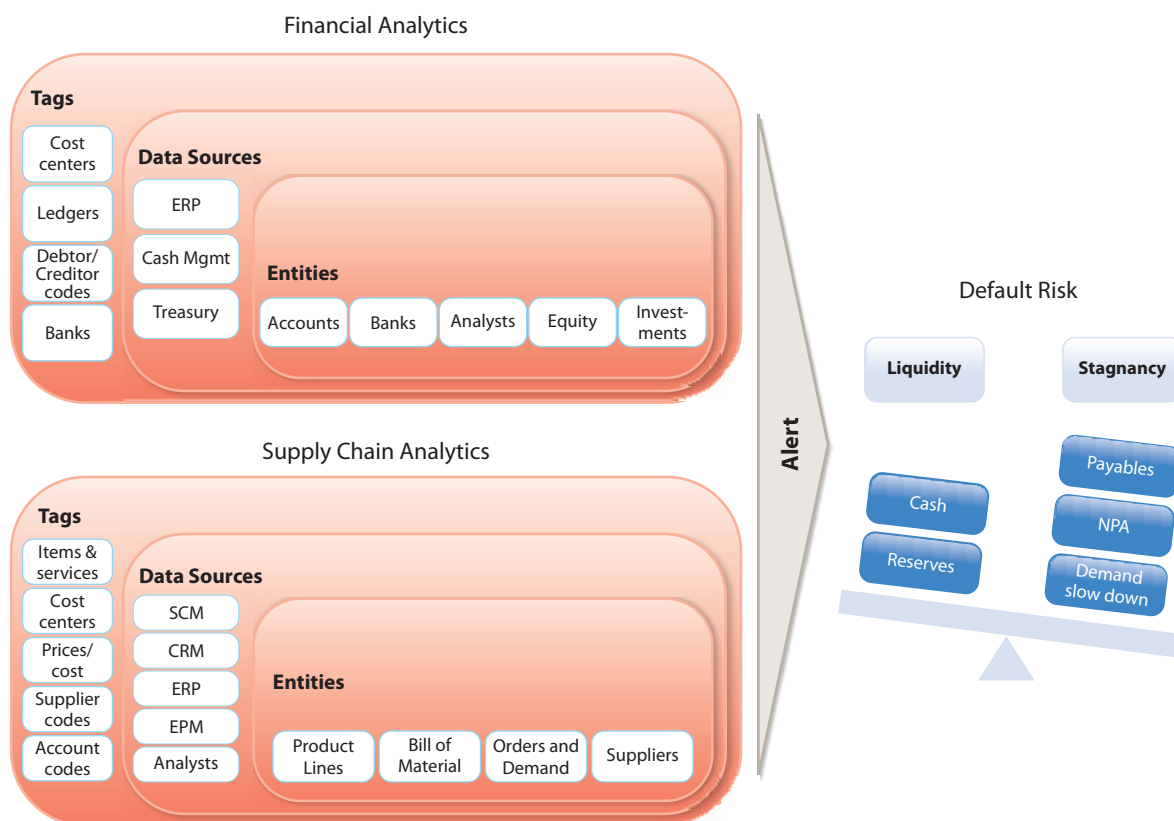
Source: TCS Global Consulting Practice - Research Desk

Replicating data sources to avoid reengineering applications

BI middleware pulls data from applications and supports analytical reports by interacting with diverse data sources. The most common approach is to have a data interchange hub (that performs ETL or Extract, Transform, Load) that deals with each application separately to pull data and build a centralized repository, called a *datamart*. As a result, reports don't deal with the applications directly, but with the datamarts. Such projects often have an ongoing cost.

The bigger problem with this approach is that analytical reports do not get real-time feeds from the source applications since they have to wait for datamarts to be updated. This detracts from the goal of spontaneous alerts for risk analytics.

Rather, such BI tools should pull relevant data on demand by intelligently choosing the right source. If all applications are XBRL-compliant, BI tools don't need to pool data but can contextually drill down into it. This makes analytics more real-time and federated. But how can legacy applications support such a paradigm?

Transforming legacy databases to XBRL is impractical and it poses risks to the functioning of applications. Yet most databases have good replication capability using very simple tools. It makes sense for these databases to create regular snapshots in an XBRL-compliant form. These snapshots in turn act as data sources in the information hierarchy, without any need to reengineer the application *(Fig. 9.3)*.

Even in today's age of sophisticated databases, a good amount of enterprise data resides in Excel or Access. This data is an important part of the information hierarchy. Microsoft offers a plug-in, certified by Edgar Online, to support XBRL transformation on Excel. Similar tools or plug-ins are available for more sophisticated databases and applications. It's a way to continue using legacy systems while supporting a real-time information system.

The question now becomes, Has the role of BI changed? Traditionally, BI pools data from various sources to deliver reports. In the proposed paradigm, there is less pooling of data. Rather, BI tools talk to real-time XBRL sources and create intelligent risk alerts and reports, which in turn are XBRL-compliant. In this way, new-age BI is already here. Most BI tools have started supporting XBRL. It has been an easy transition since most BI tools are already XML-enabled; XML is the generic form of XBRL.

Fig. 9.3: How XBRL driven BI would be different from traditional BI to support real-time analytics
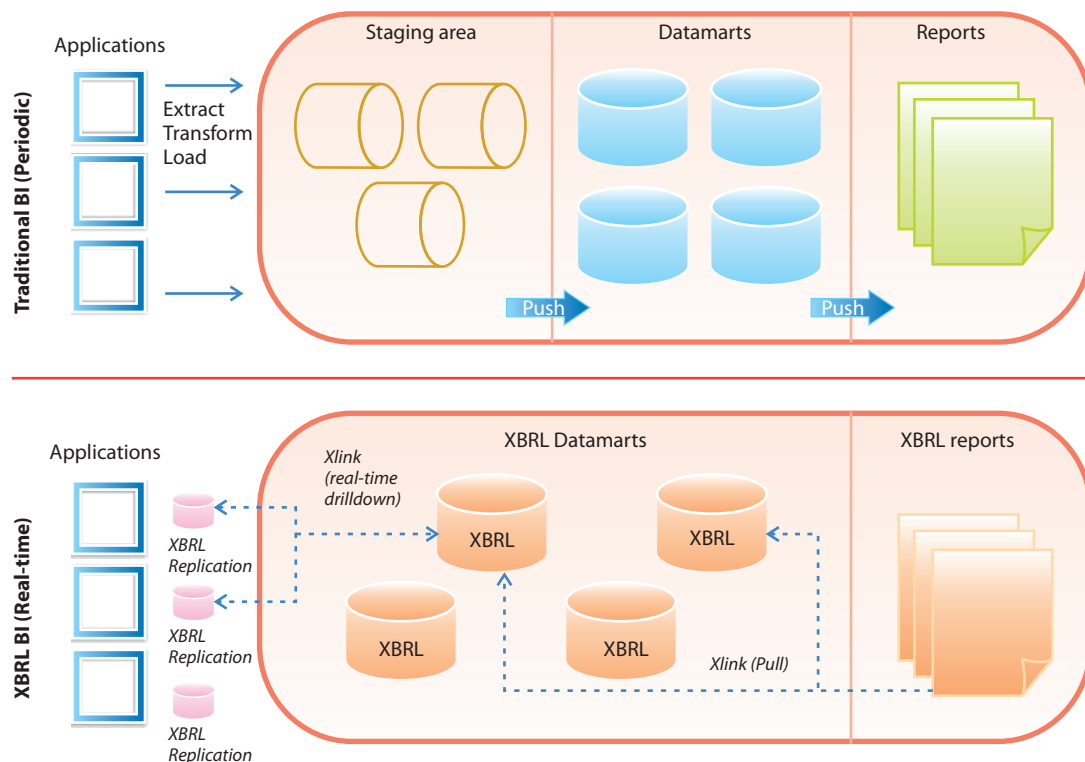
Source: TCS Global Consulting Practice - Research Desk

Cut the cost of running legacy applications with XBRL

While we see that XBRL is less expensive than perceived, it can also cut IT costs in another way.

Many legacy applications retain historical data for either compliance reasons or possible historical analysis but are not currently in active use. Such applications are common and difficult to get rid of. The cost of running such applications is high in terms of licenses, power consumption, and administrative staff. By replicating the data into XBRL snapshots, the applications can be retired but the data still has a place in the information hierarchy.

At the same time, when using data federation, enterprises can rely less on one single high-powered BI tool. Rather, they can use bits and pieces of BI at different layers of the information hierarchy.

XBRL makes ERM cost-effective

The economic downturn has taught us that risk governance is not about procedures but about mechanisms to make the enterprise self-correcting. Enterprises are increasingly accountable to shareholders regarding ERM which is seldom gauged on governance procedures but on risk intelligence systems. It reflects the changing face of compliance and business outlook toward internalizing risk management. Businesses have to show their information competency to increase shareholder value. ERM is as important as any other performance metric that drove shareholder value in the past.

A federated BI seems to be the answer to risk analytics. ERM becomes practical and low cost when we adopt XBRL. This is because XBRL supports structured documentation of information with a distributed hierarchy. Leveraging historical data from reports, along with current data for analysis will provide leading-edge intelligence and adaptive analysis. With simple implementation and conversion tools, it plays well with legacy systems.

This refutes the common perception that ERM is costly. In fact, it is costly only when ERM is not incorporated into mainstream information systems.

Tête-à-Tête

"During a down economy, firms should be growing CTB (Change the Business)"

Dr. Howard Rubin

Interview with Dr. Howard Rubin,
Professor Emeritus of Computer Science at Hunter College CUNY USA

Tête-à-Tête

This time, *Perspectives* met with Howard Rubin, Professor Emeritus Hunter College of City University of New York. Rubin is a veteran analyst in technology economics. A senior advisor to Gartner and a research associate with MIT Center for Information Systems Research, Rubin has advised global companies in IT strategy, and consulted for the US federal government and United Nations in strategic initiatives.

He runs the analyst firm Rubin Worldwide, which specializes in technology consumption economics. Rubin is now propagating the concept of "tech-commons," a philosophy that he believes would drive technology adoption in large enterprises in the future. This philosophy says, IT agility would be shown by how companies share applications and infrastructure with peers, and find best-of-breed solutions. His thoughts stem from the patterns he saw in open source and cloud computing, and examples of these in enterprises he worked with.

Publications and analysis by Rubin can be viewed at rubinworldwide.com and from MIT-CISR releases. Rubin can be reached at *howard.rubin@rubinworldwide.com*.

Here, Rubin shares his thoughts about optimizing IT cost both in the near term, and over time when technologies would present more opportunities to business.

Ray Strecker, North America Head of TCS Global Consulting Practice, interviews him.

Thanks, Howard, for being with us in this edition of Perspectives.

Optimizing IT spend has never been more important with revenues and IT budgets still down in absolute terms. At the same time, many of the best ideas for revenue growth and competitive advantages depend on emerging technology, from new mobile devices to databases for advanced analytics. How does business balance between cutting cost and nurturing innovation?

I believe that "balance" is the key word in your question.

If an enterprise is managing its "Technology Economy" properly it will also be in the process of economic optimization and should have the necessary economic agility – upward scalability, downward compressibility, and fixed versus variable expense structure – so that cost cutting is never an issue.

And the other side of the balanced model is a rigorous investment program of high transparency that recognizes that technology is a lever than can be used to grow revenue, protect revenue, reduce operational cost in the short term and avoid costs in the long term, and also to manage risk.

If the investment model is in continuous use and coupled with optimization, then "the balance between cutting cost and nurturing innovation" is an inherent part of the whole process.

The kind of model I am suggesting makes all this visible in business terms and supports attaining such a balance.

IT costs are often divided into run the business (RTB) and change the business (CTB) with RTB costs for ongoing operations and maintenance, typically accounting for 75 percent of total spend. To affect change, CIOs have more tools today, including workforce globalization, a broader array of outsourcing options from both traditional and newer players, and technologies like virtualization, green computing, and cloud. Is the metric really moving over time? If not, why not?

Actually the metric is both moving over time and failing us all at once.

The simplistic RTB-CTB model was a great leap forward when we introduced it 10 years ago, but since then we have learned that it masks the dynamics it was intended to illuminate.

The RTB view is simply too broad to lead to judicious action and interpretation is hard because people usually put all of infrastructure/production into that bucket. And today that is an area of major investment.

So yes, the metric is moving over time although it is hard to "read".

From the portfolio viewpoint, CTB climbs when the economy is good and shrinks in tight times when discretionary projects are cut. We've learned that this is the inverse of successful competitive behaviors. If companies use the continuous optimization model we should see CTB continue to grow. In fact, during a down economy firms should be growing CTB, as IT is their only expense that can be leveraged to get operational leverage and drive overall operating expense down.

For the average financial services firm, IT expense is about 9% of Non Interest Expense (NIE), and some are as high as 16-18% which means that somewhere between 82% and 91% of NIE is non-IT. Investment in IT can bring the whole thing lower!

What are the top two or three components of IT consumption that offer the best potential for cost reduction today?

A big component that most people miss is "demand". Simple things like having too many desktops per employee or mobile devices make a big difference. Demand also takes other forms – like processing capacity in servers and storage and MIPS. Even extreme service levels with little business value take their toll on expense. Additionally, "bloated" application portfolios offer a great opportunity for demand reduction – both in maintenance labor and the downstream impact on infrastructure capacity.

At the same time, firms should look to the marketplace for commodity pricing in well-known areas of sourcing like mainframe and the help desk and email. Plus the global labor market still abounds with opportunities for labor rate arbitrage.

Virtualization has been a major theme in reducing IT infrastructure spend over the last few years, because reducing an organization's server requirements not only lowers hardware cost but also software license cost, power consumption, labor, and space costs. Have the easy gains been made at this point, and if so, what are the next options to contain infrastructure costs?

This may be hard to believe but our data shows that virtualization is relatively untapped and has low penetration. The "easy gains" that have been made, have been in testing and non-critical areas of low economic value. Firms need to drop the idea of doing it themselves and get external expertise and move virtualization into production environments, the mobile work environment, and other applicable areas.

Is cloud a practical alternative today? How do you see this area evolving?

"Cloud" is in the early adoption stage. I find most firms don't understand what it is; don't know the real applicability; aren't thinking about public and private clouds; and don't understand cloud offerings and their likely evolution.

To me "cloud" is the first manifestation of the potential of what I call "technology commons." If you look at the recent Nobel Prize in Economics, it was awarded for work somewhat related to "business sharing". The cloud is a perfect vehicle where applicable for low cost, sharable, mass market computing resources.

The first stages of use are likely in testing and development for in-house IT, and then I can see its large scale use for critical elements of analytics, business intelligence, and searching.

You have been advocating the concept of "tech-commons," arguing that enterprises should share IT functions through utility-like services. We have seen open source tapping community innovation. We also see concurrent development happening on cloud platforms. Do you see these technologies emerging as the building blocks for tech-commons? What other building blocks are needed?

Most definitely, yes. Open source is pivotal.

More importantly, new technologies would make IT finely divided in core and shared services, where the latter are the candidates for tech-commons, and the former the competitive differentiators.

And for the commons, I see services such as standard processing (probably best on cloud), email, network services. For services that are competitive differentiators, specific business product focused systems and customer interface systems, etc.

More specifically on open source, a number of organizations are making a serious commitment to increase the use of open source. However, resulting total cost of ownership for open source has remained a black box for business critical applications. Do you find this becoming more tangible?

Most definitely it is becoming more tangible. With new levels of IT transparency and a better understanding of the true IT cost of goods, the open source dynamics and value will become clearer, and that will drive adoption.

I would say that open source is a manifest of another form of the "commons". The value is not in low cost but in collaborations.

Social networking and other Web 2.0 technologies are emerging as business tools, but this seems to be one of many areas where corporate computing is chasing innovation in personal computing. One could also note the iPhone, web search, and gaming as areas where our computing environment at home is often ahead of what we see at work. What should the business world be doing to keep up?

I don't think it is as much "keeping up" as it is a change of attitude. The distinction between IT and "technology" is becoming more and more blurred. Hence companies need to view technology holistically and not operationally. Once you take the total view into account, innovation and adoption can happen anywhere.

For example, look at media companies like those in cable and broadcasting, the line between technologies is truly blurred - where does IT end and "transmission" begin? Similarly in banking where core banking driven by technology has changed the age-old concepts in banking.

With total technology governance at the enterprise level, the full potential of technology innovation can be realized. I think technology governance, yet again, is about the governance of the commons.

What does all this mean for the role of the CIO? How does the CIO combine the role of general business executive, tech visionary, cost cutter, and reliable service provider into a manageable job?

The new CIO is the keeper of the enterprise's technology economy. The CIO in my view is charged with making its dynamics visible and making it effective. As my colleagues at MIT CISR put it, this is all about creating IT savvy enterprises in which the value and opportunities presented by technology are understood, leveraged, and managed communally.

Yet as the economy improves, long-term optimization often gets put on the shelf, and in the next downturn, the company is faced with the same set of short-term choices. How does a company avoid being like a fad dieter who never loses weight on a sustained basis?

This is where my idea of Technology Economics comes in. As I mentioned in response to one of your earlier questions, companies need to take charge of their own technology economy – or perhaps it is a microclimate. They need to put in place balanced programs of continuous optimization with a strong investment model. Doing this is a far better and sustainable program, which should stop/prevent fad diets.

About TCS' Global Consulting Practice

TCS' Global Consulting Practice (GCP) is a key component in how TCS delivers additional value to clients. Using our collective industry insight, technology expertise, and consulting know-how, we partner with enterprises worldwide to deliver integrated end-to-end IT enabled business transformation services.

By tapping our worldwide pool of resources - onsite, offshore and nearshore, our high caliber consultants leverage solution accelerators and practice capabilities, balanced with our knowledge of local market demands, to enable enterprises to effectively meet their business goals.

GCP spearheads TCS' consulting capacity with consultants located in North America, UK, Europe, Asia Pacific, India, Ibero-America and Australia.

About Tata Consultancy Services (TCS)

Tata Consultancy Services is an IT services, business solutions and outsourcing organization that delivers real results to global businesses, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled services delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development.

A part of the Tata Group, India's largest industrial conglomerate, TCS has over 143,000 of the world's best trained IT consultants in 42 countries. The company generated consolidated revenues of US $6 billion for fiscal year ended 31 March 2009 and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

Contact

For more information about TCS' consulting services, email us at global.consulting@tcs.com or visit www.tcs.com/consulting.