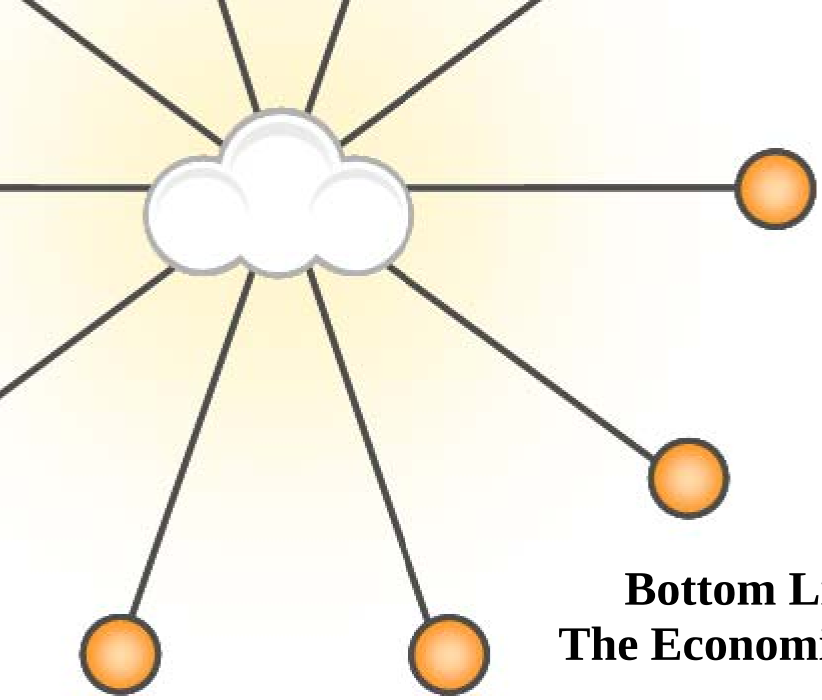

Bottom Line SOA:
The Economics of Agility

Marc Rix
marc@marcrix.com
http://www.marcrix.com

April 2007

Graphics by Brian Teifer

Service-oriented architecture (SOA) has emerged as a promising harbinger of agility, but certainly is not without its critics. Although SOA is not a new concept, it is not a mainstream approach and has not yet demonstrated widespread, repeatable success. It is no surprise then that it is often met with passionate resistance. Given the lack of SOA case studies and that it usually demands a dramatic departure from the norm, SOA is often sidelined to make way for the more traditional, generally accepted point-to-point (P2P) integration approach.

In the absence of objective SOA performance data, the industry needs a simple model that compares the distinguishing characteristics of SOA and the P2P approach on a fundamental level so that we can make more objective judgments about the fit of SOA in our enterprises. This article presents such a model and demystifies the short and long term financial effects of each, illustrating that although each is aimed squarely at the goal of enterprise integration, their effects on the bottom line are vastly different – in fact exactly opposite.

This model demonstrates that a service-based IT infrastructure is significantly more financially sound than a point-based IT infrastructure when integrating a heterogeneous IT environment. Applying traditional enterprise architecture practices to the challenge of building agility is not only a recipe for disaster, but companies that currently rely on point-to-point solutions are likely to already be caught in financial quicksand.

The Anatomy of Agility

In order to effectively compare P2P and SOA they must first be put on a level playing field. This model assumes that the primary goal of IT is to leverage its technology and resources in a way that maximizes its value back to the business. Since today's businesses demand quick action in the face of accelerating change, speed of delivery is an integral component of the IT value proposition. Therefore, each of the two approaches will be evaluated on its ability to provide and sustain IT agility while maximizing return on investment (ROI). Before we dive into the math we need to define what agile IT looks like.

Fundamentally, companies need their IT organizations to turn strategic business ideas into valuable technology solutions as quickly as possible. However, what is considered quick for one company may be lethargic for another, which makes it difficult to establish a universal benchmark of agility. But let’s assume that the market will select for the quickest companies, who are likely to have achieved as close to zero latency as practicable for their industries. For our purposes, agility is characterized by IT’s ability to turn around solutions almost instantly.

In order for IT as an organization to operate in real time, its infrastructure must also operate in real time. This means that the technology that goes into producing solutions must be ready to be integrated and delivered at a moment’s notice. Building such a product invariably involves blending the capabilities of many software functions from multiple sources into a unified package. Therefore, each component function must be ready to be called into the game at any time.

Furthermore, functions are chained together in different sequences to serve different business needs. And with potentially hundreds, or even thousands, of useful functions teeming in the typical enterprise, the number of possible combinations is enormous. To be agile, IT must assume that any of these integration “value chains” may need to be shrink-wrapped and delivered to the business at any time. The only way to meet this extreme, real-time integration challenge is to preemptively link every component in the enterprise to every other component. Otherwise, the integration work must occur after the order comes down from the business, thereby increasing time to delivery and diminishing agility. Through total connectedness IT can assemble applications by simply activating pre-integrated chains of components in a sequence that suits the business. This takes the integration step out of the delivery cycle and moves IT toward real-time execution. But what does a fully connected, agile IT infrastructure cost?

Measuring ROI

The primary unit of investment in the agile IT environment is the dedicated connection – the physical bit of infrastructure that links one component directly to another. Physical connections provide a transport mechanism through which information can flow between otherwise isolated functions. Component integration, and thereby total enterprise integration, cannot happen without them.

Each connection, of course, comes with a price tag, which represents the sum of all short- and long-term expenses associated with keeping the connection in working order. Short term costs encompass the work required to build the connection, such as planning, designing, implementing, and testing. Long term costs stem from work required to maintain the connection over time, such as bug fixing, upgrading, and enhancing. Since these costs are directly related to work performed by humans, we will consider them all labor expenses and base them on an average, blended labor rate. The investment portion of the ROI equation, therefore, will be the total ownership cost (i.e., the aggregated lifetime labor cost of owning each dedicated connection) of operating a fully connected, change-ready IT infrastructure.

Each time a component is linked to another, value-added information emerges from the IT environment. The connectedness of that environment allows IT to combine information in new ways that are of value to the business. The more integrated nodes there are in the network the more ways information can be creatively assembled, hence the more value can be extracted. Of course, not all relationships between connected nodes will be active all of the time, which suggests a distinction between *active* links and *passive* links.

Active links are integration relationships that are “online.” They are in use and serve a concrete business need. Passive links are relationships between components that exist but are inactive because the value provided by the relationships is not currently in demand.

When a business need calls for them, though, they are quickly switched on and become active. Whereas active links provide *actual* value to the business and can be used to measure the *present* value of the network, passive links provide *potential* value and can be used to measure the *future* value of the network. Since agility is a measure of both the present and future performance of IT, we are interested in both active and passive links. Therefore, the total return delivered by an agile IT infrastructure is the combined value of all active and passive connections. The ROI of the network is simply an expression of the difference between total value and total cost.

Now we are ready to begin the analysis. The remainder of the article compares two approaches to achieving total enterprise integration and evaluates them based on the ROI the resulting integration network offers back to the business. The first is the “classic” approach to enterprise architecture, characterized by simple point-to-point connections between isolated components and often favored for its simplicity and historical success at meeting business needs. The second is the SOA approach, characterized by loosely coupled, reusable links between components and gaining popularity for its overall flexibility.

The ROI of P2P

IT traditionally tackles integration challenges reactively, responding to the arrival of concrete business requirements with the commencement of a software development lifecycle. Since the process unfolds within the scope of a funded project with defined schedule and resource constraints, the traditional model rewards IT for quick solutions that do not inject complexity beyond that which is required to satisfy the immediate business need. Any deviation outside the defined boundaries of the project is generally construed as unnecessary risk. Consequently, components are integrated using methods that have minimal impact on the project plan. The method that fits this profile best is the P2P approach, in which all isolated components are joined together through direct, physical connections.

The benefits of this approach are its simplicity, low up-front costs, and long-standing reputation for getting the job done. Development teams are not burdened by technical requirements that do not relate directly to the business solution and managers are not accountable for expenses that do not directly support their project objectives. The consequence, though, is that each integration solution provides little or no value to any business case other than the one for which it was specifically designed. Because reuse is not a consideration, each IT component must have a dedicated, physical connection to all others in order to achieve total connectedness. Let’s examine the economic implications of this approach.

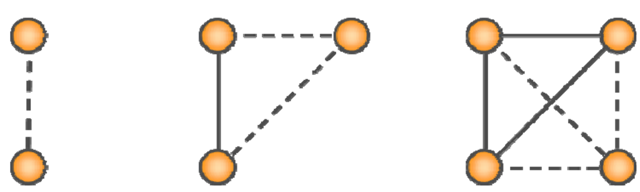

Figure 1: Evolution of a P2P Network

Suppose we have an IT infrastructure, or segment thereof, consisting of 10 components that each perform a business function (e.g., "create application," "calculate credit score," "authorize user") that has value outside its native system. We wish to link them all together into an agile, change-ready network. As illustrated in Figure 1, we begin by linking two nodes together, yielding one connection. Adding a third node, we must link it to each of the first two. Adding this node has caused the total number of connections to increase to three. Adding the fourth node requires linking it to each of the existing three, producing a grand total of six dedicated connections.

Interestingly, each time we add a just single node to the network, we must build more than one connection to link it in. In fact, the number of new links required is proportional to the size of the network at the time the component is added. Specifically, linking every component to every other component will require $N(N-1)/2$ physical connections, where N is the total number of components, or nodes, to be included in the network. When we are finished networking all ten of our example nodes we will be left with $10(10-1)/2 = 45$ physical connections, as depicted in Figure 2.

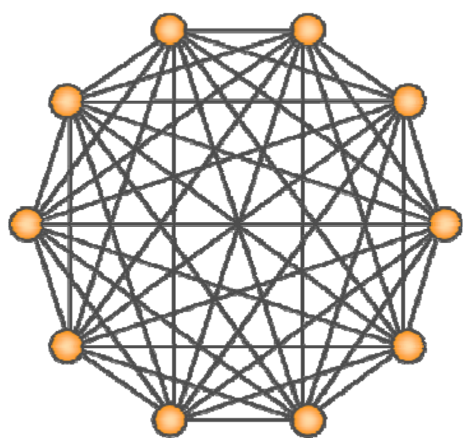

Figure 2: ***N(N-1)/2*** **Dedicated P2P Connections**

Let's now assume that each connection has a fixed lifetime cost. (In reality, of course, individual build-out and ongoing maintenance costs will vary, but every IT organization should be able to calculate a reasonably accurate approximation based on average labor costs and useful life.) For example, let's say it takes, on average, 40 hours of labor at a blended rate of $50 per hour, or $2000, to build each connection. This puts our total up-front integration cost at $2000 * 45 = $90,000.

As for maintenance, let's say that each component will need to be materially changed – such that the nodes it is connected to must be correspondingly changed – once during its useful life. Each change by itself requires an additional 10% of the initial build-out cost – in this case, $200 (.10 * $2000). At face value this may seem perfectly manageable, but remember that since each node is connected to every other node, *every change will affect every node* regardless of where the change originated. A change to one node will cause all nine of its liked nodes to change. At $200 apiece, the bill comes to $2000. Multiply this by the ten independent changes that we anticipate and the total maintenance cost of the network comes to N^2 * *($200) = $20,000*. The total lifetime cost of our 10-node P2P integration network is *$90,000 + $10,000 = $110,000*.

This doesn't seem too scary, but then again, we limited our IT infrastructure to just ten components. It is doubtful that a company could react to change in real time with only ten high-value business functions at its disposal. The optimal size of the network will vary from company to company but it is probably safe to assume that it includes at least 100 nodes, especially when considering the range of valuable HR, CRM, contract, billing, security, logistics, loan origination, analytics, and other functions that may be present in any given IT environment. Running the cost figures above for a network of 100 fully connected nodes yields a total build-out cost of $9,900,000 and lifetime maintenance of $2,000,000, for a total cost of $11,900,000. Taken a step further, a very large IT shop may need to network 200 or more nodes to be agile, at a total cost of more than $47,800,000. Frighteningly, as the number of nodes doubled from 100 to 200, the total cost more than quadrupled.

The P2P model does do an excellent job of keeping integration costs low at the project, or microeconomic, level. Since this is the scale at which IT historically evaluates success it is no wonder that this is the predominant approach today. The ugly truth behind this approach is that big problems begin to hatch at the enterprise, or macroeconomic, level very quickly. Each new low cost, tightly coupled P2P connection added to the enterprise has a non-linear, compounding effect on the total lifetime cost of the overall IT infrastructure. The P2P cost model increases exponentially as the size of the network increases linearly, as illustrated by the cost curve in Figure 3. But this is only half the story. What does IT *gain* by investing in point-to-point architecture?

Every connection in the P2P network has exactly two endpoints: one information source, or provider, and one information sink, or consumer. Since the connection exposes the functionality of one provider node to a single consumer node, the business gains one unit of value from the investment in the connection. As a rule then, IT can expect a return on investment of one unit of value (whatever it defines that to be) per connection in the integration network regardless of the size of the network. Therefore, the total value of the complete network of connections increases linearly as nodes are linked together. The value of a P2P network is equal to the sum of the individual values of its physical connections. To put it another way, the value of a P2P network is equal to the sum of its parts.

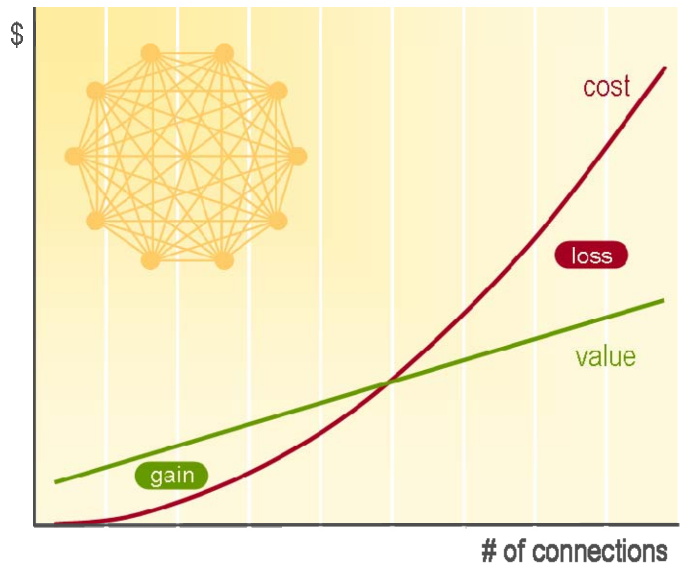

Figure 3: Exponential P2P Costs Yield Linear Value

Looking at the cost curve and value curve together (Figure 3) reveals a striking characteristic of the P2P network: the value of the network increases linearly over time while its costs increase exponentially. With the addition of each component, more money must be invested to eek out the same rate of return as the component before it. As a result of this phenomenon, short-term gains realized at the project level can explode into massive losses at the enterprise level. Indeed, this analysis paints an ominous portrait of traditional IT spending, but let's see if the service-oriented approach can do any better.

The ROI of SOA

Service-oriented architecture differs from the point-to-point approach in that it avoids dedicated, hard-wired connections between components. Instead, components exchange information indirectly via shared middleware services, which act together as a distribution "hub" for messages traveling between isolated nodes. Components in a SOA network always interact with the service hub, never directly with each other. Moreover, the link any component has with another component is a *virtual* link, in that the two are connected implicitly via the hub, not explicitly through a physical connection to each other. This architecture shields any node in the network from the technical details of all other nodes, resulting in a web of loosely coupled integration relationships that spans the IT infrastructure. This characteristic of SOA is extremely powerful because it significantly reduces the number of dedicated connections, and therefore overall costs, required to build and operate a fully integrated, agile network.

Let's apply the SOA approach to our 10-node infrastructure. Just as with the P2P approach, we begin by linking two nodes together. This time, however, we must incorporate the service hub as an intermediary between the two functional nodes. We see right away that this requires more architectural complexity than creating a point-to-point link between two nodes. To begin with, the service hub represents an additional piece of infrastructure that must be added to IT's inventory. Next, two physical connections (one from each node into the hub), instead of one, must be built to complete the integration. Unfortunately, this increase in complexity naturally begets a corresponding increase in spending.

Our first two nodes have yet to be linked and there are already clear signs of cost increases compared to the P2P approach. Project-oriented, date-driven IT organizations have a big problem with how SOA presents itself at this level. Managers are apprehensive about taking on the burden of delivering this extra layer of complexity and project controllers are reluctant to approve funding for it. At this microeconomic level, it is easy for SOA to be stifled because it inflates scope and budget precisely where scope and budget are scrutinized most heavily. However, welcoming a bit more investment at the project level can pay off handsomely at the enterprise level.

Continuing with the example, things begin to improve as soon as we get beyond integrating the first two nodes. (Figure 4) In the P2P model, linking the third node into the network required two additional physical connections. In the SOA model, one is sufficient. Because the first two components are already members of the network, node three gains access to them implicitly via virtual connections through the hub. Similarly, the fourth node gains virtual access to all existing nodes when it is attached to the hub. By contrast, linking the fourth node into the P2P network required *three* additional physical connections.

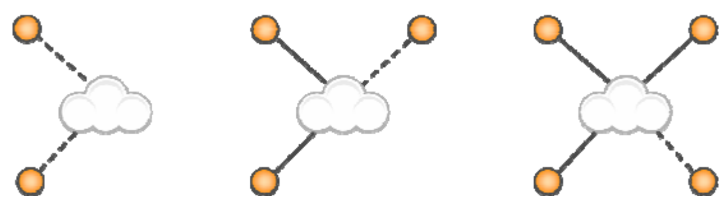

Figure 4: Evolution of a SOA Network

When all ten nodes have been fully integrated into the network, that is, when all ten nodes have the ability to share data with all other nodes, the network will be composed of just 10, or *N*, physical connections – less than a quarter of those required by the P2P network. Only one additional physical connection is required to fully link one additional node into a service oriented network, regardless of the size of the network. (Figure 5) Now that we know how many connections there will be, let's examine what it will cost to build and maintain them.

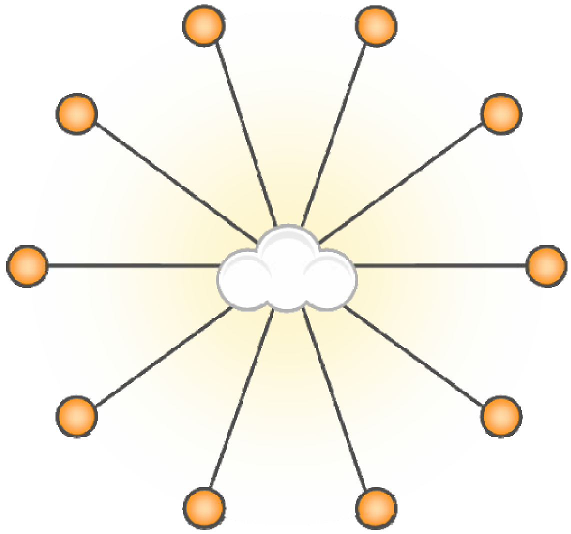

Figure 5: ***N*** **Dedicated SOA Connections**

Again, each connection has a total cost comprised of initial build-out and lifetime maintenance expenses. We will assume that connecting a node to the service hub entails roughly the same complexity as linking it to another node, so the build-out cost of each physical connection will remain $2000. However, the additional cost of the service hub itself must be accounted for.

For simplicity, we will spread hub expenses evenly over the network as a “tax” on each connection. The complexity of the service hub should grow with the complexity of the network so that hub costs are kept commensurate with connection costs. We will tax each connection at a very generous flat rate of 100%, meaning that each integration project will allocate equal resources to building connections and enhancing the service hub. This effectively doubles the build-out cost of each connection to $4000. The resulting build-out cost of our 10-node, fully connected service oriented network comes to *N * ($4000) = $40,000*.

We will assume again that maintenance costs equate to 10% of the build-out cost per connection – in this case $400. This time, though, we do not have to account for tight coupling between nodes. When a node changes materially in the SOA network the change is insulated from all other nodes. At most, the service hub will need to be brought into compliance. We will conservatively assume that this will always be the case; a change to one node will always require a corresponding change on the hub. This will produce a lifetime node maintenance bill of *N * 2 * ($400) = $8000*.

In addition, the hub itself represents a new source of ongoing maintenance. When it changes, the effect will likely reverberate through the network requiring corresponding changes to all nodes. Again, let’s say this happens once during the life of the network and equates to *(N+1) * ($400) = $4400*. Summing the two maintenance figures gives a total SOA network maintenance cost of $12,400, yielding a total lifetime cost of $52,400 –

roughly half the total cost of the P2P network. A 100-node SOA network would cost $120,400 (1% of the cost of a comparable P2P network). A 200-node SOA network would cost $240,400 (0.5% the cost of a comparable P2P network). In the case of SOA, doubling the size of the network from 100 to 200 nodes also doubles its total cost, producing a linear effect.

Let's now look at return on investment. Recall that, in general, the return on one connection is a function of the number of nodes that can leverage the connection. In the P2P network, each connection links one provider to one and only one consumer. Thus, the return on a single P2P connection is always one unit of business value. On the other hand, each physical connection in the SOA network links one node to *all* other nodes via virtual connections through the hub. The return on a single SOA connection then is proportional to the size of the network. This means that the value of the SOA network increases exponentially as its size, or number of nodes, increases. The value of the SOA network is *greater* than the sum of its parts.

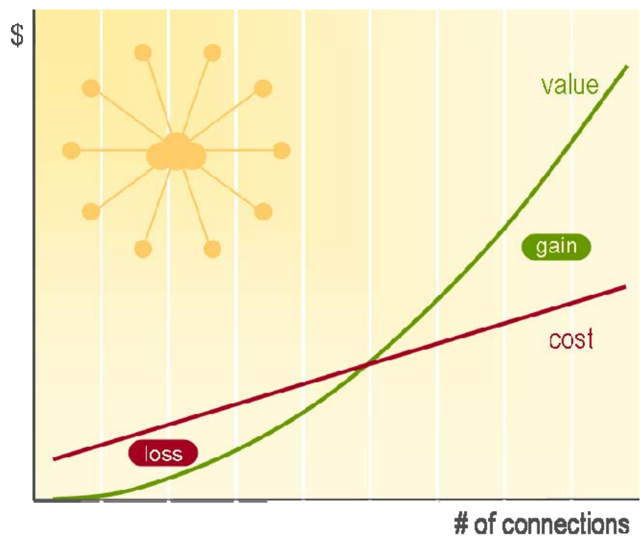

Figure 6: Linear SOA Costs Yield Exponential Gains

Examining the SOA cost and value curves together in Figure 6, we see that the graph mimics the P2P graph almost exactly: each contains one curve with a linear progression that intersects a curve with a non-linear progression. The fundamental difference, however, is that the curves' labels have been swapped. The exponential P2P cost curve has evolved into a linear SOA cost curve and the linear P2P value curve has evolved into an exponential SOA value curve.

The value of the SOA network increases exponentially over time while its costs remain linear and predictable. Up-front costs will likely be higher for SOA than P2P producing temporary deficits at the project level, but they pave the way for disproportionately high returns at the enterprise level. Once the break-even (intersecting) point is reached, SOA delivers runaway value without negatively impacting its cost structure.

The Bottom Line

Agility is the Holy Grail anxiously sought by today's IT organizations, and the quest begins at a fork in the road. One stretch cuts through familiar P2P territory and is easily navigable by intuition; the other winds through the mysterious, uncharted regions of SOA. Many companies have been enchanted by the folklore about the SOA route but have ultimately turned away from it (or strayed from it) out of fear and uncertainty, choosing to travel the well-beaten path of P2P instead. They have chosen poorly.

The journey to agility comes at a price and IT must be ready to endure both the short-term and long-term financial burdens of its trek. Point-to-point architecture carries seductively low up-front costs, resulting in little impact at the project level. The ability to create integrated solutions with no impact on prevailing business practices makes P2P the instinctive favorite. However, microeconomic stability can quickly fester into macroeconomic chaos. As more and more one-off solutions are piled atop one another, IT maintenance costs swell exponentially and connection sprawl strangles the infrastructure. This often results in full dedication of IT resources toward simply "keeping the lights on," leaving little left over to support new, strategic business initiatives. Companies that adopt the P2P approach must be prepared to experience increasingly negative returns on their IT investments.

SOA presents exactly the opposite model. Higher up-front expenditures are required in order to ensure loose coupling between linked components, which inflates project budgets. This translates into initial financial losses at the microeconomic level and often catapults projects beyond conventional thresholds of risk tolerance. However, if this aversion to early-stage risk can be overcome, investments in SOA at the microeconomic level can blossom into colossal benefits at the macroeconomic level. By dramatically reducing the number of physical connections in the infrastructure and leveraging no-cost, virtual connections to achieve everything-to-everything connectivity, SOA's return on investment increases exponentially as the infrastructure grows. Furthermore, since SOA keeps infrastructure costs linear and scalable, IT always has resources free to innovate and respond to changing business needs. SOA may be the road less traveled by, but for companies seeking agility it will make all the difference.