White Paper

metagroup.com • 800-945-META [6382]

December 2003

Evolving Operations:

Simplification Through Coordinated Orchestration

A META Group White Paper
Created for IBM/Tivoli

"It has been a long-standing struggle to align the IT organization (ITO) culture and business entities. One of the biggest barriers to accomplishing the goal is that the ITO itself is not operated in a manner compatible with business entities. Rather, ITOs operate as a set of standalone groups that are technology-focused and often unaware of how their activities impact business goals. ITOs must get their house in order, addressing underlying practices, culture, and structures, as well as attacking the core of the "hero" mentality that currently is prevalent."

META GROUP

Contents

Current Situation

IT organizations (ITOs) are being presented with a great opportunity: they are being asked to play a larger role in driving the business. While no one questions the relevance of technology to any business, the influence and significance of the ITO are often questionable. With businesses having optimized many, if not most, key revenue-impacting parts of the organization (e.g., manufacturing, customer support), they now are turning their attention not only to using technology to further the optimization, but also to reinventing ITOs to run as an optimized entity, driven by business goals rather than technology ones. IT executives are so concerned with demonstration of business value that they consistently name business alignment as a key goal (see Figure 1).

Figure 1 — CIO Priorities

2003

1. Reduce cost
2. Reduce staff
3. Business alignment

2002

1. Metrics
2. Business alignment
3. Improve software process

Source: META Group Metricnet Survey

It has been a long-standing struggle to align the ITO culture and business entities. One of the biggest barriers to accomplishing the goal is that the ITO itself is not operated in a manner compatible with business entities. Rather, ITOs operate as a set of standalone groups that are technology-focused and often unaware of how their activities impact business goals. ITOs must get their house in order, addressing underlying practices, culture, and structures, as well as attacking the core of the "hero" mentality that currently is prevalent. Many IT organizations are now involved in internal analysis seeking optimization strategies, covering areas from infrastructure consolidation through implementation of operational processes. The common goal with all the analysis underway is determining how to get more out of people, processes, and technology, while still retaining the same or even lowering the investment and still retaining or improving service levels experienced by the business.

Successfully optimizing ITOs will require all people, process, and technology operating under four key goals:

- ***Efficient[1] Utilization[2]:*** Ensuring that the manner in which resources are used produce the desired effect and that all are consumed at optimal utilization. This does not necessarily mean that they are fully used, with no waste, but rather that they are consumed while being aligned with the

[1] Efficient: Acting directly to produce an effect; acting or producing effectively with a minimum of waste, expense, or unnecessary effort; or exhibiting a high ratio of output to input (Source: dictionary.com).
[2] Utilization: To put to use, especially to find a profitable or practical use for (Source: dictionary.com).

business goals being served. From efficient utilization of a server resource, to driving down device support ratios (e.g., more devices per administrator), there are many discrete metrics to track efficient utilization.

- ***Flexibility***[3]***:*** Ensuring that all resources can be used in whatever way deemed necessary — not locked to a single use when it can be prevented.
- ***Timeliness***[4]***:*** Ensuring not just that resources can be changed and adaptive, but completed in a suitable amount of time. Long delays in making change will hamper all efforts to become adaptive.
- ***Value***[5]***:*** Ensuring the price is right. Does the company believe it got what it paid for? Was a reasonable amount paid?

The Goal

Internal analysis within ITOs uncover an unfortunate reality; most ITOs exist in a non-optimized state. While the business itself tends to adapt to the ever-changing demands of the markets, IT is slower to change and cannot easily shift to meet new demands. Projects are long, expensive, and complex, with a high risk of failure or coming in overbudget for large projects. The goal for any ITO should be to become an adaptive organization. This will require maturity through several organizational stages, which will take time. Once organizations have moved through the stages (e.g., reactive, proactive), they will end up at the adaptive, or as IBM calls the "autonomic," state. The adaptive/autonomic state is characterized by intelligence embedded within the operational and infrastructure technology to provide automated responses to identifiable needs (e.g., outages, increased demand, pending problem). Although it is desirable to be adaptive/autonomic as an end state, reaching this state is a journey. Along the way, companies will need to invest in solving problems. META Group recommends initial investment in tactical automation, workflow, and virtualization technologies, with strategic implications. These investments address current challenges of increasing efficiency and flexibility, but still maintaining an eye on the future desired state.

No IT organization can become fully business-aligned without at minimum becoming proactive and striving to become adaptive/autonomic. A key element of proactive and adaptive is leveraging well-designed processes, complete with automation. While many companies have pushed to implement stronger IT processes, the operational side of IT is now in need of stronger automation. Automation technology, which allows companies to increase efficiency of the staff,

[3] Flexibility: Responsive to change; adaptable (Source: dictionary.com).
[4] Timely: Occurring at a suitable or opportune time; well-timed (Source: dictionary.com).
[5] Value: An amount, as of goods, services, or money, considered to be a fair and suitable equivalent for something else; a fair price or return. Worth in usefulness or importance to the possessor; utility or merit.

reduce costs, and even cut down on manual errors, is maturing and ready for companies to engage.

Impact of Complexity

While cutting cost is an obvious motivation to invest in automation, a critical emerging driver is the increase in IT complexity. IT infrastructure is getting more complex with every deployment, especially around Web-based applications. What used to be a simple three-tier application (Web server, Web application server, and database) is now exponentially more complex (firewalls, load balancers, volume managers, clusters, storage silos, multiple network devices). This complexity is matched with the desire to manage end-to-end in a consistent manner, at a reduced cost. The result is companies need to invest in automation.

There are elemental automation technologies (e.g., patch management) that provide value, but they are smaller in scope and attack only a single task. Elemental automation can be leveraged in a proactive manner. However, to become adaptive, companies must invest in coordinated orchestration as the brain behind automation, where intelligence coordinates the linking of multiple automated steps together to streamline a more complex workflow. Multiple tasks are tied together into a predefined workflow, triggered either manually or automatically, to accomplish a complex task such as deployment of a new server, reconfiguration of an application, or deployment of new application code.

Automation Impact

Automation is critical for increasing efficiency of staff and processes, as well as increasing the flexibility of IT, providing a greater ability to react (e.g., sense and respond) to business needs. The use of coordinated orchestration driving automation enables more significant automation, leading to a positive impact on companies in the three key areas of IT — people, processes, and technology:

- ***People:*** Automation technology allows companies to operate more complex systems with fewer people and fewer errors. While task-level automation will allow companies to do more with fewer people, companies will gain additional value leveraging coordinated automation, as it will cut down on manual steps and even eliminate people from having to string together a series of automation technologies manually.
- ***Processes:*** Automation technology allows companies to better define processes and ensure their consistent execution, providing maximum efficiency. Coordinated automation addresses more complex processes (e.g., 10- to 20-step change requests).
- ***Technology:*** Automation technology can allow the use of more complex infrastructure and application architectures. Complex infrastructure and applications present operational issues; they are harder to maintain.

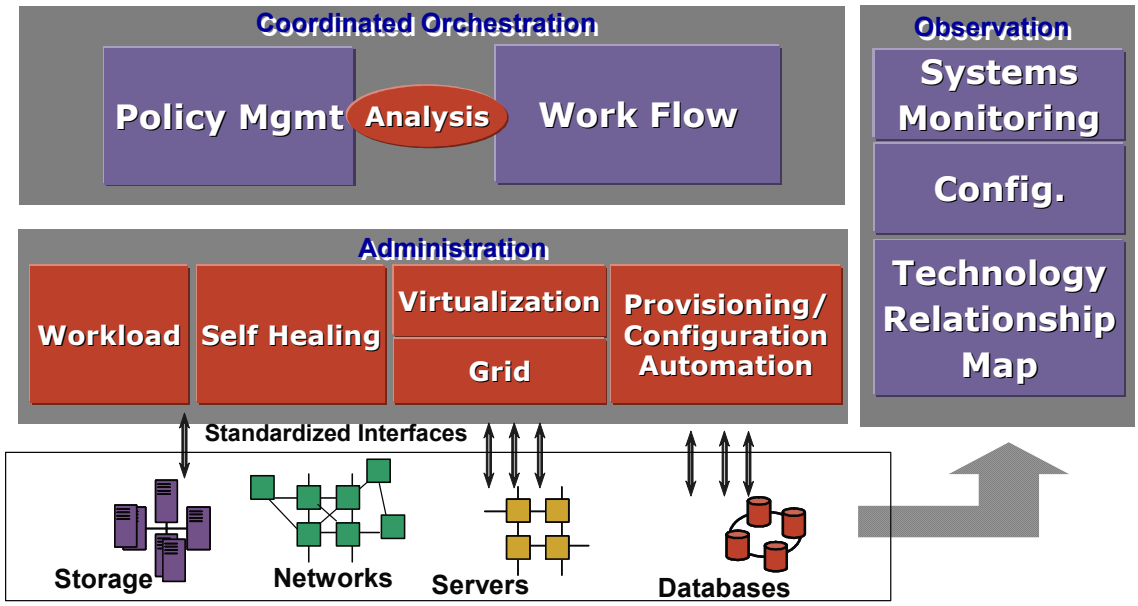
Coordinated automation makes that environment manageable by reducing confusion in maintenance and operational support, hiding complexity from the administrators. This opens new possibilities to a company. Technology for automation cannot require rip-and-replace strategies; rather, it must leverage as much of the installed capability as possible. Any rip-and-replace technology will substantially reduce the value of automation technology.

Adaptive Resource Management and Coordinated Orchestration

The success of coordinated orchestration as the core of automation will not be based on any single technology, but rather a group of technologies working together, known as “adaptive resource management” (see Figure 2). It is critical these technologies not only be easy to implement, but contain as much knowledge out of the box as is possible. Just purchasing a technology that is a shell of capability significantly reduces the value of coordinated orchestration and automation technology.

Adaptive resource management (ARM) is the use of operational technologies in concert to set policy, identify current state, and automatically take action required to keep infrastructure and applications operating within the defined policies, thereby enabling on demand infrastructure and applications. There are three key technology categories within the adaptive resource management model: coordinated orchestration (the brain of the operation), observation (determination of current state), and administration (technologies that take direct action). Many organizations will use these technologies in siloed areas initially (e.g., just for servers on a specific application), but a truly adaptive organization seeking on demand resources will use these technologies broadly across IT. At the heart of ARM is the brain — coordinated orchestration.

Figure 2 — Adaptive Resource Management



Source: META Group

Coordinated Orchestration

While organizations have invested in some level of automation in the past, what is new is the addition of intelligence to that automation, in the form of coordination[6]. Coordination is not a single technology, but a category of technologies used to establish the desired state of an environment and the actions to take to ensure the infrastructure and applications remain at that desired state.

The desired state is established via policy definition, as defined in one or many policy engines. Observational data will then be analyzed and compared against policies to determine what actions must be taken, thereby triggering a defined workflow of administrative technologies to alter the environment to keep it within the policy definition.

[6] To harmonize in a common action or effort Source: dictionary.com).

Policy Management
This is the embodiment of business rules, typically tied to service levels. The business rules become policies outlining technical expectations of how the infrastructure and applications should behave, while the desired service level is achieved. Policies are broad reaching, covering performance (e.g., application X requires response time Y when the application runs under Z capacity, no server in a cluster will exceed 75% CPU usage), security (e.g., no new identities can be deployed during a critical period), or even change policies (e.g., no application changes can be deployed during the holiday season for a retail application, no new servers can be deployed during business hours). Policies are a combination of detailed technical metrics (e.g., database and business goals). The policies are constantly changing and range in granularity. Most companies will end up with multiple tools containing policy management capability, leading to an eventual metapolicy engine, the policy manager above other policy mangers.

Analysis
This is the act of processing environmental information to determine if a policy is being violated, interfacing with many different observational technologies. Analysis is accepting data from many different tools across the environment as well as manual input and comparing that data against a policy. An example is analyzing monitoring data to determine that current performance is trending toward, or already exceeding, a defined performance policy. In most cases, analysis is more complex than just analyzing a single metric (e.g., threshold) and requires the input of multiple metrics, correlated against the policy. Analysis may also be of a configuration, not just operational metrics. Determination that the current environment is configured in a manner not within a current policy (e.g., missing a patch) should trigger an action. This will require definition of which underlying metrics are compared to which policy — a task that will be performed by users as well as predefined from vendors.

Workflow
This is the intelligence about what actions to take to bring the application or infrastructure back into compliance, and then establish the necessary workflow of these actions. This function interfaces with numerous technologies across the enterprise that are capable of taking action (e.g., deployment of a new server, reconfiguration of a network device, triggering of a virtualization change). The workflow engine itself does not take the action, but ensures the technologies that can take action are triggered; more importantly, if numerous steps are necessary to bring the infrastructure and application back into compliance with the policy, workflow will ensure the tasks are executed in the proper order and will not conflict with one another.

The Role of Coordinated Orchestration in Automation

Together, the three elements of coordinated orchestration act as the intelligence of an automated system, not necessarily taking action, but determining when action is necessary, what action is necessary, and what needs to be triggered to take the action. This is the most complex part of ARM and requires substantial thought on the part of a user to determine what the correct policies are, what actions they desire, and which actions should be automated and which manual. While coordinated orchestration is fed from other technologies (e.g., monitoring) or passes along commands to different technology (e.g., provisioning engines, virtualization engines), the logic will continue to live in the coordination. However, this "brain" is where the considerable long-term value is for an organization; investment in coordinated orchestration is key to operational on demand projects achieving large cost and time savings.

Although elemental automation can be achieved without coordination (e.g., server provisioning, configuration audit) with cost and time savings realized, the larger savings will come from further reducing manual effort. Elemental automation still leaves manual steps in the workflow of the tasks, doing analysis, and often is not based on a policy. This leaves the process open to manual errors (e.g., triggered incorrect step) or slow reactions (e.g., user did not get to trigger the automation fast enough). When actions can be automatically analyzed and a broad workflow created and executed as a singular flow, a more efficient environment will result.

Adaptive Resource Management: Observation and Administration

It is important to understand the other categories of technology (observation and administration) encompassed within ARM in order to fully understand the breadth of potential.

Observation

This is the act of observing infrastructure and applications to determine the current state. It is among the most mature areas of ARM, with most companies having deployed numerous monitoring technologies. The goal is not to replace the current tools, but rather to have them pass information to tools in the coordinated orchestration layer. Technologies in the observation include the following:

Systems Monitoring

This involves the health, availability, and performance monitoring of infrastructure and applications. While most infrastructure is currently monitored (e.g., servers, network devices), many organizations have gaps in monitoring of new application architectures (e.g., J2EE, .Net) and end-user response time. End-user response

time is a critical piece of information every organization must collect. Response time represents the user's perception of the service, which will be a key data point to compare against a policy.

Configuration
This includes the monitoring of the current configuration of devices (e.g., for a server). This is an emerging area in many organizations, with initiatives led by a security team. While companies have captured configuration information in the past, it was primarily used for backup purposes. Now, in many cases, that data is being compared to security policies to identify if a system is out of compliance.

Technology Relationship Maps
Technology relationship maps (TRMs) are composed of an inventory of what software and infrastructure resides in the environment, combined with the relationships among those entities. For example, it concerns what database resides on what server, as well as what database talks to what Web application server via what network router. The TRMs are a combination of inventory, asset, topology, and connectivity information. They expand on traditional topology by relating the software elements and software/application connectivity. TRMs are created through a combination of automated effort and manual relationship drawing. This data is critical to "the brain" of coordinated orchestration. It is used to assess what the impact of change is, as well as identify dependencies that may also require change to comply with the change being made. The TRMs are accessed by nearly every other technology involved in automation, but it is not taking action, which is why it lives in the category of observation.

Administration

Administration is where action actually takes place. These are the collection of technologies that interface with infrastructure and applications to make configuration or other changes, deploy new technology, redeploy workload, etc. There are many different kinds of administrative technologies, and currently, they also vary for different types of infrastructure or applications (e.g., different configuration management tool for storage devices versus servers). Over the next two to three years, the tools will converge and allow more heterogeneous access to interfacing with devices. One critical factor will be the standardization of interfaces to infrastructure. Currently, it is complex for an administrative tool to interface with a piece of technology, as every vendor has different interfaces, exposed actions, proper behaviors, nomenclature, etc. For broad coordinated automation to occur, these interfaces must be based on standards, allowing the administrative technologies to more easily interface with numerous infrastructure elements and all behave in the same manner. When sourcing administration technologies, ensure the broadest platform coverage (or at least coverage that

meets all the organizations needs). Technologies in the administration layer include the following:

Provisioning Automation
This involves the act of automating the deployment of a new device or application and its associated software layers. Provisioning has long been sought by organizations, with many efforts resulting in some form of software distribution. The good news for companies is that this technology is maturing beyond the basic capabilities to include additional logic and automated steps. Provisioning automation tools can best be described in three functional categories:

- ***Basic push:*** Creating a package of software to push to a device (may be sending the bare metal operating image) or the basic initialization of that device (e.g., router). It involves pushing the software package to that device, and executing an installation script. This is a basic capability.
- ***Operationalize:*** Building on the basic push category, but also adding instructions on how the environment must be configured to recognize the device. Next, it is pushing software to the target device, but also pushing configuration changes to other elements (e.g., load balancer, DNS server, cluster manager) necessary to have the device operational once the installation is complete.
- ***Self-directed:*** Building on the operationalize category and adding the ability for the technology to determine on its own when new resources are necessary and then automatically provision those. This requires tight integration between coordination, observation, and administration. This is achieved when a provisioning tool is integrated with coordinated orchestration capability

While most organizations are comfortable with basic push, and many even have some of that technology in place, basic push is not enough. It lacks intelligence. Technology in the operationalize and self-directed categories not only are more sophisticated, but are beginning to appear on the market. META Group recommends exploration of these technologies because they provide broader automation by elimination of manual steps. Effectively, this is leveraging coordination technology at a lower level, tying together low-level discrete tasks. While it is possible to “script” this level of automation and feed it to a scheduler, that method is not recommended, because it is complex and lacks strong process controls. META Group recommends technologies that have closely linked related discrete tasks out of the box, to ensure the most efficient automation.

Configuration Automation

This encompasses automating the change of a hardware or software configuration, beyond the initialization of a device — including activities such as altering a server configuration file or directory entry and changing runtime parameters in a database. Also, configuration automation includes automatically stopping and restarting the device or software if necessary to reinstantiate the new settings. Every automation does not require new technology to be deployed, but it may require changes to the environment. When the observation function identifies systems behaving poorly, the policy and workflow functions may determine that the action required is to reconfigure technology, not add.

Virtualization

To virtualize a set of diverse concrete resources is to access them through a uniform interface that, from the user's perspective, enables them to behave as one "virtual" resource that can be shared by diverse users with varying degrees of dynamic behavior. Virtualization creates a "single-system illusion" for certain users; while administrative-level staff members are still aware of the physical environment, to the rest of the world the physical configuration becomes irrelevant. Basically, it is making many look like one or one appear as many, but across the organization, the user is not aware of the change. This can be applied to numerous technologies (e.g., servers, databases, networks, storage arrays). When action needs to be taken to bring an environment into policy compliance, workflow may choose to virtualize a device or alter the virtualization configuration to bring operational performance back into policy compliance:

- ***Grid:*** In practice since the mid-1990s[7], this is a form of virtualization, but for specific workloads. Grid computing is a distributed workload in parallel across a collection of computers that act as a single resource. This sharing is based on resource availability, capability, performance, cost, and ability to meet quality-of-service requirements. This specific form of virtualization typically leverages desktop computers and is best designed for computational applications. Grid is yet another option of automation technology that workflow has at its disposal when determining what actions to take.

Self-Healing

This is technology embedded within an infrastructure element or application with the purpose of self-managing and self-healing. While self-healing is commonly represented by IBM (as part of the autonomic vision) at this point[8], META Group believes that most infrastructure and application vendors will seek to add self-

[7] Source: www.golbus.com — anatomy of a grid (Foster, Kesselman, & Tuecke).
[8] Source: http://www-3.ibm.com/autonomic/index.shtml.

healing capability. Also, it is anticipated that standards will be agreed upon, offering standard interfaces to access self-healing technology from a workflow capability.

Many organizations are concerned about automation technology, as it takes a degree of control out of their hands. META Group believes, however, that companies must explore automation technology or risk increased operational costs, as embodied by escalating manual error (due to complexity of change) as well as ever-eroding admin-to-infrastructure ratios. Provisioning automation and virtualization technologies are maturing rapidly and are now worth investigation by end users, with configuration automation close on its heels. For specific workloads, grid holds potential, but it is not for everyone.

Bringing It Together: Case Study

Thus far, we have discussed a great deal of technology and some complex concepts of technology collaboration and automation. Now it is time to take a step back and discuss some examples. The first will be an example of how a flow will operate in an environment containing coordinated orchestration, observation, and administration. The second will be a case study of a financial institution that has a business case to be more adaptive to business needs.

Example

Let us walk through a simple example of how these technologies will interact and share data to create an adaptive/autonomic environment:

1. Business policies are created within the coordinated orchestration layer. One of the policies is that, for a particular application, response time cannot fall below five seconds when traffic is below 1,000 transactions an hour. This response time becomes the key performance indicator for this policy. This application is a basic three-tier Web application with Web server, Web application server, and database tiers. The entire application comprises about 15 different components (e.g., DNS, load balancers, cluster management) necessary for the three tiers to interface properly.
2. Systems monitoring, from the observation layer, has thresholds set to match key performance indicators within policies. The correct response time threshold is set and monitoring begins. While monitoring the application, response time degrades. Systems monitoring triggers additional data collection from components related to the transaction (by looking up relationships in the technology relationship map and gathering all required monitoring data). This information is then passed to the policy engine for analysis.

3. The policy engine analyzes the data and determines that the servers within the application server tier are all running at a high CPU level. The goal is to keep the application operating within policy definition, so the workflow engine is called to trigger a deployment of a new device to the application server tier.
4. The workflow engine creates a workflow of all the necessary steps to deploy a new device; a key is referencing the TRM and identifying all dependencies. Also, the combination of analysis and workflow validates all the administrative capability at its disposal (e.g., is the provisioning engine present, are there servers available to be deployed). The workflow engine may have also determined not to provision a new server, but rather create a virtual machine within an underutilized device to temporarily get the application operating within desired policy parameters. This capability is the end goal (to broadly integrate all the necessary steps to alter an environment into a single logical workflow, regardless of platform, application, or environment), and while vendor efforts are underway to deliver this level of sophistication, it will be several years before users should expect this broad level of capability to be mainstream — though several instances of it are available (e.g., IBM Tivoli ThinkDynamic Orchestrator) for particular environments (e.g., servers, Web applications).
5. The workflow engine enacts and manages the workflow. Administrative technology provisions a new server (or adds a virtual server within another), updates the load balancers, updates DNS, and takes all other necessary steps to make the new application server operational.
6. Systems monitoring continues to monitor and finds the response time is now within the policy definition — service has remained at the proper level.
7. While service levels have been maintained, the problem has not necessarily been solved. Administrators are notified of the situation, what metrics have been collected, and what action has been taken. After analysis of the data, an application administrator comes to the conclusion that the reason the application server tier was overloaded was a queuing problem between the application server and database, causing long database transactions and excessive waits at the application server tier.
8. Developers create a new “fix” for the application server/database interface problem.
9. The workflow engine is called, automatically or manually, to trigger a workflow that will distribute the fix to the application servers, as well as de-allocate the newly deployed server, as it should no longer be necessary.

This simplistic example illustrates how these complex systems can be more easily analyzed and administered with the use of coordinated orchestration technology as the brain of the integrated system.

Case Study: Financial Services

Now, we will take a more complex example. Let us look at Acme bank. Acme has been analyzing how it interacts with its customers and has come to the conclusion that there are three primary manners of client interface for a single account: ATM machines, online banking, and point-of-sale charging. Each of these manners of interfacing involves its own application logic. However, because it is for a single account, they all come together at a back-end database. Each region of the bank has its own database for its clients.

Currently, Acme has three distinct applications built, which then reference a central database for a given region. Each application enters Acme's environment via a partner network, with three different partners. The applications each have a distinctly unique usage pattern. During lunch hours, Acme has found that users interface primarily via ATM machines, leading that application to peak at an 80% usage during that period of time. After work hours, while people are commuting home, point-of-sale swipes increase dramatically, with that application peaking at 75% usage. Finally, in the evenings, when customers are at home, they do online banking, peaking those applications at 60%-70% utilization.

Something is troubling Acme. When it looks at utilization in the off-peak hours, it is determined that it averages 10%-20% utilization, leading to an average utilization in the 20%-25% range. This means that all three applications are overbuilt on an aggregate level, because each is built for peak utilization. So how can Acme become more adaptive/autonomic?

Acme can become more adaptive/autonomic by leveraging a pool of servers, which can be automatically reprovisioned, either in full or by creating virtual instances on each server and reallocating system resources for each partition based on which application requires the most resources at a specific time. Each application will retain a fixed set of resources, which will average 70%-80% utilization during non-peak hours, but for peak hours, server resources from the shared pool will be provisioned (either fully or via virtualization) to the application to meet its demand.

Acme has choices. It could take partial approaches (e.g., invest in server virtualization, invest in automated provisioning) to allow minimal initial investment. While limited, this path can be a start. However, the bank will ultimately need a combination of automation, virtualization, workflow, etc. to ensure the optimal use of the infrastructure, the desired speed, and reduced cost of support.

Acme would save in multiple ways. First, it would not have to pay for the same amount of hardware, software, and operational labor, a key goal of most organizations' server consolidation projects. Another benefit is that Acme can make better use out of what it already has, even if it keeps it all so that it can increase workload without as much increase in capital. In addition, Acme would not have to maintain as many devices, reducing the operating costs of the environment. Also, the automation technology will be leveraged to maintain the environment (e.g., deploy new versions, patches, and upgrades), thereby further reducing the operating effort required. Overall, it is a savings of time and money.

Barriers to Adoption

What has been described is a very useful set of technology that, when in place, can save companies significant time and money. So why haven't companies been falling over themselves to deploy these solutions? Here are a few barriers META Group has determined and how we believe companies should handle them.

Lack of Understanding

META Group has found that many companies are not even aware that solutions exist that can automate provisioning, let alone solutions that can orchestrate a comprehensive workflow of complex tasks. The year 2003 has acted as a year of education, with organizations now becoming aware that some solutions exist, but still seeking more proof. Organizations are seeking references, as well as proof points on scale and product success (e.g., low rate of failure in users). The good news is that these successes are beginning to mature, with investments that have been made over the past year. META Group predicts 2004 will begin mainstream understanding and investigation, with some initial deployments. By 2005, the majority of Global 2000 companies will be seeking tools to automate key operational tasks, and over 50% will be seeking advanced workflow technology.

Trust in Technology Maturity

Companies, rightly so, have a hard time trusting complex new technology. Solutions that offer coordinated automation, or adaptive resource management, are viewed as mostly new by companies. While 2003 has been a year of education, it is not the first year of technology. The tools have been built and tested over the past 2-2.5 years, with early customer deployments as much as 1.5-2 years ago. The tools still have some maturing to do, but we are not talking about totally untested, untried, greenfield technology.

Companies also fear technology that takes action. Monitoring tools have for years had the ability to take basic corrective action (e.g., restart a process), and companies have done a poor job of leveraging this capability. META Group

believes there are differences with these new technologies. First, monitoring tools gave a general action capability, rarely containing specific details of a task and never tested by the vendor itself. The new adaptive tools not only have predefined the tasks at hand, but have tested many scenarios to ensure the automation is smooth. Also, the tools are seeking to automate well-known tasks — ones that companies have expected for years — so it is easier to ensure the proper execution when the process is understood. Finally, if companies do not become comfortable with automation, they risk increasing support costs, a proposition that does not sit well with most IT executives. Automation is one of the untapped areas to leverage in the drive to cut IT costs.

Political issues

As is normal, success of technology is only partly based on the technology itself. The success is also largely predicated on an organization's readiness to accept the technology and willingness to make it successful. This industry has witnessed many winning technologies fail based on poor deployments and internal political sabotage. META Group believes there are a few identifiable areas of political impact that will be specific roadblocks:

- ***Loss of control:*** Administrators are used to controlling everything around a device they own. From making ad hoc configuration changes, through making complex updates and new deployments, administrators have their own way of operating and succeeding with these tasks. There is a fear among these users that bringing in a tool will not "do it my way" and that the tools cannot be trusted to execute the tasks just the way the administrator seeks. Companies can get around these issues by deploying technology that takes incremental steps, allowing administrators to interface with the automation technology along the way, and ensure that the job is being executed properly. Over time, the administrator will tune the automation tools and reduce the time they spend managing an automation workflow, and allow the tools to take more and more action on their own.
- ***Pooling of resources across groups:*** For broad automation schemes to be successful, organizations must be willing to shift IT resources across internal groups, making the devices multipurpose and resulting in less device ownership by a specific group or business unit. Organizations are often unwilling to do this, as they feel ownership is important to be sure their specific application needs are met. A way to mitigate this barrier is to first do pooling within a discrete group within a company (e.g., pool all servers that serve a single business unit, pool all servers that serve corporate applications). Over time, as groups become comfortable that the automation technology will allow them to obtain the necessary resources when required, we expect there to be an expansion of resource pooling to include other groups.

User Actions

So what should users do? META Group recommends companies begin with ensuring the internal operational processes are well established. When internal processes are understood, they are then candidates for automation. Once automation candidates are clear, organizations can begin to investigate technology. But what technology should be sought? What technology is the most mature?

Organizations should start with virtualization and provisioning automation technology. Server and storage virtualization are the most mature, with storage virtualization ready for production environments, and server virtualization well suited for test and development environments. By 2005, server virtualization will be ready for a production-level environment. Provisioning automation is mature enough for a production environment. It has great potential for most companies, for Unix and Windows servers. Configuration automation has begun to mature, with targeted point tools around patch management and configuration audit now being adopted by the mainstream. While we currently see many discrete point tools, by 2005 provisioning automation, configuration audit, and patch management will all exist within a single tool, mostly accomplished with provisioning automation tools adding that capability, as opposed to the point tools growing up to the new functions.

Organizations should seek provisioning automation technology that has elements of coordinated orchestration already present (e.g., a brain of prebuilt workflows and policy capability). Administrative tools will have to be called by workflow, and with few standardized interfaces between administration tools and coordination tools, it is important to seek integration out of the box, pre-created by the vendor.

Policy management is the most immature of all elements in adaptive resource management. META Group anticipates sophisticated policy management will begin broad exploration during 2004/05 and mature over the next three to four years. It is logical to invest in tools that leverage a policy foundation to prepare for policy usage, and then learn from that experience.

The final key to success is to start small. Businesses should not attempt to automate the entire organization in tool deployment. They should select a target area and build an internal success, using it as a reference point for selling the use of this technology to other groups.

Bottom Line

Universally, organizations are seeking ways to offer the business more flexibility from IT, control costs, increase resource utilization, and increase overall IT efficiency. META Group believes automation, workflow, and virtualization technologies are ready for adoption and are critical to addressing high-level needs. Although these are newer technologies in many cases, they can clearly add great benefit to organizations now. Companies must seek tools that not only cover the broadest array of platforms (allowing the best coverage for the environment), but also embed intelligence. Embedded intelligence means a faster time to value as well as a broader use available. By 2005, it will not be optional to understand or engage adaptive resource management technology; rather, it will be a necessary part of any successful IT organization.

Corey Ferengul is vice president and principal analyst with Operations Strategies, a service of META Group, Inc. For additional information on this topic or other services, contact info@metagroup.com.

About META Group

Return On Intelligence[SM]

META Group is a leading provider of information technology research, advisory services, and strategic consulting. Delivering objective and actionable guidance, META Group's experienced analysts and consultants are trusted advisors to IT and business executives around the world. Our unique collaborative models and dedicated customer service help clients be more efficient, effective, and timely in their use of IT to achieve their business goals. Visit metagroup.com for more details on our high-value approach.