Cassatt

TECHNICAL WHITEPAPER

How to Use Automation and Virtualization to Improve Service Levels of Your Applications

IT in Crisis

Across the globe, IT organizations are realizing that their application infrastructures are in a state of crisis.IT is working hard to respond as rapidly as it can to the demands of their business users. However, IT is being hindered by applications that were built in silos with dedicated hardware and software that are very tightly coupled; this restricts their use to supporting only those specific applications. IT systems are too brittle to rapidly respond to changes and new directives, despite the huge positive potential those changes could bring. IT departments are instead forced to focus most of their time and resources on just keeping the applications running. This means spiraling operations costs (even when using lower cost "commodity" computing systems) and a very restricted ability to respond to new business requirements. All this is happening in an environment in which application service level expectations are increasing.

Today's reality is:

- On average, 90% of IT resources in major data centers sit idle, which translates into millions of dollars in unused assets.
- Data center space and power requirements are running at maximum capacity or pushing beyond because so much hardware is needed to maintain required service levels.
- As much as 90% of IT budgets are spent on "keeping the lights on," leaving very little money for implementing new ideas that could help with competitive differentiation.
- IT organizations are constantly playing catch-up with the ever-increasing demand for improved service levels from their business applications.
- Even with the dramatic capital expenditure savings realized by low-cost hardware, and moving from RISC/SMP to small commodity servers, IT organizations are not able to make substantial cost reductions.
- Capital savings are completely off set by increases in operational costs.

The IT mantra to "do more with less" requires a different approach—one that is uniquely realized with Cassatt Collage.

Today's Static Data Centers: IT "Waste"-lands

A central problem facing IT organizations is the static data center. Because applications are tightly coupled to the machines on which they run, dedicated hardware must be set aside each time a new application is put into production. The amount of hardware dedicated to an application is based primarily on projections of peak-usage for the application. Thus, over-provisioning is rampant in today's static environment and dramatically limits the return on investment on resources.

Why are IT departments unable to make full use of their systems? Consider an Order Entry system. For many businesses, the vast majority of orders for goods are placed in the last few days of a fiscal quarter. This means that for two or three days of each quarter the system is at peak usage. For the other 87 days of the quarter the system may sit almost totally idle—but during this time the servers are still under-utilizing software licenses, burning power and Heating, Ventilation, and Air-Conditioning (HVAC) and consuming real estate. Even though these servers have horsepower to spare on most days, there is no way to share this power with other applications that might need additional resources to meet their service level requirements.

The only way out of these problems is to create a more dynamic, real-time IT infrastructure—one that allows resources to be quickly repurposed to the most critical tasks and applications. And, one that's not labor-intensive, so savings from consolidation aren't off set by increasing operational costs.

What Collage Does: Service Level Automation

Collage is a single software solution that combines Goal-Driven Automation with Virtualization Control in a real-time infrastructure to deliver Service Level Automation.

- **Goal-Driven Automation**—optimizes resource allocation across the enterprise by taking existing policies, priorities, and Service Level Agreements and mapping them to Collage goals without using scripts.
- **Virtualization Control**—delivers unified control of all the application components required to deliver increased service levels. Collage makes use of all types of virtualization technologies, including operating systems (OSs), Virtual Machine Managers (VMMs), Javan Virtual Machines (JVMs), Virtual Local Area Networks (VLANs), and services.

While traditional management software focuses on physical resources such as servers and networks, Collage focuses on application service levels, treating both physical and virtual entities as a dynamic pool from which resources are drawn based on priority and current demand. While the system is running, the software continuously monitors applications and resources to ensure goals are met. As conditions change, the Cassatt software automatically makes decisions, and without human intervention, reallocates resources between applications (taking into account the customer's priorities and policies). This doesn't just change IT operations, it transforms them by a quantum leap by safely eliminating over-provisioning and realizing dramatic cost savings.

More importantly, Collage achieves this using existing, off -the-shelf hardware, operating systems, network switches and applications—allowing IT organizations to gradually introduce this solution into their infrastructure without radical change.

How Collage Works

Service Level Automation is achieved in Collage by:

- Breaking the chains between applications and physical servers
- Creating pools of applications and servers
- Dynamically allocating resources to applications

Let's take a closer look at the two components that enable Collage to help IT departments deliver a real-time infrastructure—and to meet the demands of constantly changing business requirements and ever-increasing application service level requirements.

Goal-Driven Automation: Measure, Analyze, Respond

To achieve goal-driven automation, Collage takes on the task of measuring, analyzing, and responding to events in the enterprise, a task currently handled in most cases by a human being. With Collage, all customers have to do is translate their IT policies, business priorities, and Service Level Agreements into goals—which are continuously and automatically managed by the software. The beauty? It's all done without a single script. This automation strategy is far more scalable and robust than competing script-based solutions.

With Collage's Goal-Driven Automation, each application is assigned:

- A set of targets for availability and performance
- A priority relative to other applications in the environment

Automating resource allocation

Server management, maintenance, provisioning, and application deployment are time-consuming activities. When these activities take days to weeks to complete, they are neither meeting the goals of reducing cost of ownership nor of increasing response to the business.

Collage automation features a feedback loop that constantly measures, analyzes, and responds to changes in application and server conditions and matches an application with available resources based on user-defined policies. In this way, Collage enables lights-out, hands-off operations. As a result, Collage simplifies planned upgrades, automates application deployments, improves availability, and intelligently responds to hardware and software failures by dynamically allocating or deallocating compute resources. Collage monitors both applications and resources while continuously comparing the state of the system against customer-defined goals. For example, when there's an increase in demand or a hardware failure, and an application cannot maintain these goals, Collage's Optimization Engine computes a new optimal distribution of resources—taking into account current conditions and inter-application priorities. Collage then quickly redeploys resources where they're most needed—without human intervention

The Collage Goal-Driven Automation eliminates the need to provide scripts that detail how to respond to different types of events—resulting in a highly robust system in a constantly-changing environment. Examples of events that are automatically handled by Collage include:

- Load spikes
- Hardware failures (including server, network, power, etc.)
- Hardware additions
- Hardware retirements
- Application crashes
- Application rollouts
- Service priority changes

Virtualization Control

A true automation product should affect change—not just monitor what's going on. Collage's ability to autonomously make changes differentiates it from simple management products. This is accomplished by Collage's Virtualization Control technology, and by leveraging other virtualization technologies.

Virtualization, in its most general definition, is about abstraction. Virtualization technology adds layers of abstraction that hide the details of one part of the overall system stack from another part. These abstractions enable rapid change in the overall configuration of the data center and allow resource silos to be broken down. Collage provides a set of abstractions that allows for software flexibility and can optionally leverage additional virtualization technologies to enable additional agility and efficiencies.

Cassatt's Image Matrix technology provides an abstraction layer that decouples the hardware and software stacks. Where software (including operating systems and applications) is normally installed on a server's hard disk, the Image Matrix allows the software to be hosted on the network where it can quickly move from machine to machine. It also allows many physical systems to operate from the same software image, dramatically improving maintenance costs.

There are many virtualization technologies today, and more and more are being moved into production. Several types of virtual resources can be optionally leveraged by Collage—enabling more flexibility and agility while driving even greater cost savings.

Java Virtual Machines

Java Virtual Machines (JVMs) underlie the application servers that power the world's data centers. JVMs provide an abstraction layer to host applications, isolating J2EE applications from the operating system and hardware beneath them in the stack. As part of a server consolidation strategy, JVMs also allow a single operating system instance to host multiple applications. The Collage Web Automation Module (WAM) extends Collage to control Java Application Servers, such as BEA WebLogic Server. This allows customers to radically improve the flexibility, utilization, and availability of their existing application server farms.

Virtual Machine Managers

Virtual Machine Managers (VMMs) such as VMware create a complete abstraction of the hardware and allow multiple operating systems (and the applications that ride on top of them) to be hosted on a single physical machine. This enables customers to increase server utilization. However, without an automation strategy for these virtual resources, the resulting virtual machine sprawl creates management issues that more than off set the cost savings from the utilization improvements. According to eWeek, "Attendees at Gartner Inc.'s recent Symposium/ITXpo griped about virtualization's introduction of an untenable mountain of virtualized servers that make for management headaches."1 Collage can help customers manage virtual machine sprawl by automating the management of these resources.

Virtual Local Area Networks

Virtual Local Area Networks (VLANs) create an additional abstraction inside Internet Protocol (IP) networks. VLAN technology allows switches and routers to control multiple sub-networks and to quickly move hardware into and out of these networks. Collage allows physical server resources to be automatically shared between applications on different parts of the network.

How Collage Works

The following examples describe how Collage handles several common IT events.

Handling Load Spikes

Let's assume a sample environment where a collection of J2EE applications is being managed by Collage with the Web Automation Module. Let's also assume Collage is managing several non-J2EE applications. All of these applications are being hosted from a shared pool of 100 servers.

Each J2EE application has a set of service level goals created by the application administrator. Each service level goal is defined in terms of a formula using data gathered from the Application Server's Java Management Extension (JMX) Agent. The goal defines how the system will know that each application is meeting its expected performance and availability targets.

Collage continuously monitors all applications. During the day, an event occurs that drives many more users than average to one of the J2EE applications. Collage automatically detects that the application is running below its service level goal and may be providing degraded services to its users. Collage's

1 http://www.eweek.com/article2/0,1895,1883680,00.asp

Optimization Engine then examines each of the nodes in the resource pool that is already running an appropriate J2EE server. It selects a server that has spare compute capacity and deploys an additional copy of the application to that server—in effect, increasing the amount of compute power available to that application. This action takes approximately one minute.

The system again measures the application against its goals. The criteria have improved, but are still below target. The Optimization Engine again evaluates the available application servers and finds that none have significant free compute power. The engine decides to create a new J2EE server. It selects a bare-metal server from the global free pool and Collage uses the server's remote management interface to turn on the server. As the server starts, Collage causes it to boot using a software image that contains an appropriate operating system and application server. After the server starts, the engine deploys an additional copy of the overloaded J2EE application to the server. The complete act of taking the server from bare metal to providing service takes about seven minutes. Now, the application has sufficient capacity and it is meeting its service level goals. When the load spike diminishes, the Optimization Engine may decide to power down the extra server and return it to the global free pool for use by other applications in the system.

Ensuring Applications Run Despite Hardware Failures

Continuing with our sample environment, let's assume another event occurs: a Power Distribution Unit (PDU) in the data center fails, which in turn causes eight servers to fail. Collage detects this in seconds and the Optimization Engine begins to calculate a response. Each of the goals and priorities in the system is evaluated and the engine finds that three high priority applications have fallen below their targets. The system goes to the global free pool, finds an available server, and assigns it to the highest priority application. The engine then evaluates the service levels and resources assigned to lower priority applications, and decides to repurpose three servers to higher priority applications. After realigning the resources in the system, the engine finds that two low priority applications are still below their targets, but are still operational. While all this is occurring, Collage automatically pages a system administrator.

After the PDU is repaired by a technician, Collage automatically puts these offline resources back into use and realigns the system so that all applications can again meet their service level goals.

Making Application Roll-Outs Easier

In most environments, bringing a new application into production requires the acquisition of dedicated hardware. In a Collage environment, you can leverage what you already have by creating goals for a new application and asking Collage to activate it. Collage then evaluates the status of the existing physical resources and takes one or more of the following actions, depending on the situation:

- Assigns one or more physical nodes to run the application (selecting them from the global free pool or borrowing from lower priority applications).
- For a J2EE application, selects one or more running application servers and deploys the application on hardware with spare capacity.
- Finds one or more Virtual Machines on existing physical servers and deploys the application on them.

Collage Helps Meet and Exceed Application Service Level Expectations—Automatically

Collage provides a unified software solution that provides Service Level Automation, eliminating wasteful over-provisioning and improving the flexibility and overall usage of an organization's IT infrastructure. This results in dramatic cost savings and major improvements in an IT organization's ability to react to changing business demands.

Collage's unique Goal-Driven Automation is far more scalable, robust, and maintainable than script-based IT automation systems. Virtualization Control provides unified control of virtualization resources and the ability to interface to external virtualization software like VMMs and JVMs. With these capabilities, Cassatt Collage helps customers change their static infrastructure into a dynamic, real-time infrastructure, providing automated management and a single control point for all physical and virtual IT resources. All this is possible using existing, off -the-shelf hardware, operating systems, network switches and applications. Moreover, IT organizations can gradually introduce a Collage solution into their infrastructures without radical change.

Instead of chasing sprawl and other operational problems, IT organizations that implement Collage can focus on innovation, giving their organizations a powerful weapon to stay ahead of the competition.

Solution Centre Limited
Vickers House
Priestley Road
Basingstoke, RG24 9NP.
Tel: 01256 818600
Fax: 01256 819600
E-Mail: sales@solutioncentre.co.uk
www.solutioncentre.co.uk

Cassatt Corporation
1740 Technology Drive
San Jose, CA 95110

t: 408.437.7600
f: 408.437.1364
www.cassatt.com