



EXECUTIVE WHITEPAPER

**How to Harness the “Cheap
Computing” Revolution:
Utility Computing—The
Fundamental Shift in Data
Center Economics**

IT Trends, Their Effects, and a Strategy to Exploit Their Value

IT costs are escalating, data centers are multiplying, utilization is way too low, power and space requirements are way too high, complexity is increasing, and businesses are locked into static, often out-of-date applications that are making them less, not more, competitive. This is the state of IT today and something obviously has to change. Well, the “Cheap Computing” revolution is dawning and it is only the beginning of a transformation that will unleash the next wave of change, freeing organizations from the costs and bonds of static IT and delivering competitive advantage for 21st century’s world of increasingly dynamic global competition.

Companies today are buying smaller, less expensive, more commodity-based systems to run their business applications. At the same time, their most important software is being created and delivered in smaller, more discrete components to assist with flexibility. New technologies such as IT automation, grid computing, and server virtualization are being adopted to help harness this growth in volume, cost, and complexity. Finally, free open source software is playing a larger and larger role because of its cost and time-to-market benefits. What do these shifts mean to companies with large, complex data centers running their core business processes?

This commoditization and virtualization of IT, which is being effectively deployed to get control of capital costs, has the downside of causing a veritable explosion of data center complexity and operational expenses. The sheer number of components to manage in the data center is exceeding what humans can manage; in fact, operational costs (administration, monitoring, maintenance, etc.) are outpacing the cost of hardware and software. Estimates from Gartner, Forrester, and IDC analysts place operations costs somewhere between 50 and 70% of IT costs and they are growing at twice the rate of overall IT expenses. Of even more concern is that none of these capabilities address the most important goal of IT—to deliver the dynamic business process automation required to compete effectively in the 21st century.

“There is a fundamental shift occurring in data center economics. The ability to drastically cut the capital expenses that have been burdening data centers is finally possible. What you’re seeing is only the first step of the Cheap Revolution.”

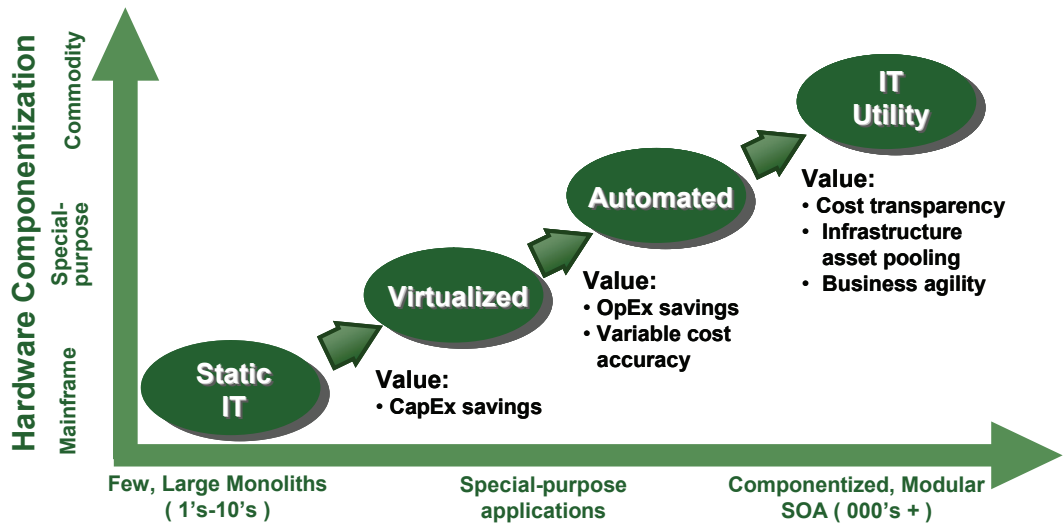
Bill Coleman, Cassatt CEO

Well, what if businesses could turn their current data centers into utilities that automatically adapted to their business goals while utilizing data centers to their optimum capacity and dramatically cutting costs? This is the next revolution in IT.

This paper addresses just how this can be done today, beginning with current systems and applications. The overriding approach to achieve this goal is called utility computing. Businesses that transition their IT infrastructure to operate as a utility will yield benefits far beyond the capital and operational costs they avoid. No longer will they need to treat their IT as an inert monolith that slows business agility. Instead, IT will become an asset that dynamically adapts to the needs of the business.

Steps Toward Utility Computing

At Cassatt, we believe the move toward utility computing is being driven by two major industry trends that take advantage of the economic leverage of the Internet. These trends represent the commoditization of computing; they are: (1) Server scale-out computing—the use of multiple inexpensive servers to replace large, expensive ones and (2) Service-Oriented Architectures (SOA)—the componentization of software applications that can be easily assembled into larger, composite applications. Although both of these trends promise dramatic improvements in IT's operation, their value is being overshadowed by the dramatic increase in operational costs and complexity that they also introduce.



The solution to this dilemma is the transformation of IT from a static to a utility computing model. This transformation is proceeding in three stages, each of which adds value and can be implemented with technology available today.

1. Virtualization

Today, software is well on the way to componentization, and hardware is moving from special-purpose to commodity. Static IT “silos” in which software is permanently associated with dedicated banks of under-utilized servers are rapidly becoming a thing of the past. The technology accelerating this is virtualization, in which software is allowed to be independent of specific hardware. This decoupling of the bond between software and hardware allows data centers to efficiently pair the two, increasing server utilization levels by packing more software onto existing hardware. The result of this virtualization movement is an opportunity to save large amounts of capital expenses.

2. Automation

As organizations move to adopt commoditization and virtualization, IT departments are finding that the level of complexity in dealing with the sheer number of things to manage (e.g. hardware, software, networks, and storage) is driving operational costs to unsustainable levels. However, by introducing automation into the data center, much of the complexity of day-to-day operations can be masked. What is being automated are tasks like maintaining software and hardware, adjusting both supplies to meet user demands, and performing maintenance. The benefits of automation are many, but primarily include dramatically lower operational costs.

3. Utility Computing

At Cassatt, we believe that the final and most important step in this transformation is to pool resources into a common utility that can be dynamically adjusted to meet business objectives and demands in the most efficient and cost effective manner. This is accomplished by treating the software and hardware as a “pool” of assets to supply services and continuously optimizing the combinations of software and hardware based upon customer-defined business policy. By knowing the IT “supply” as well as user “demand,” and by correlating them with the costs of the data center, enterprises can turn IT into a variable cost of business that is generated based upon the value to the business. This is what we at Cassatt call *Service Level Automation*, the ability to deliver dynamic, policy-based application workflow at desired service levels. Service level automation is the ultimate benefit of utility computing and the true goal of IT.

“The goal of Service Level Automation is to deliver the guaranteed quantity and quality of application service flow as efficiently and cost effectively as possible, limited only by total capacity available... the purpose of IT.”

Bill Coleman, Cassatt CEO

What Utility Computing Means to Business: A Transformation Beyond Technology

The move to utility computing is a transformational enabler of business for the 21st century. It will drive not only cost savings, but also dramatic changes in how technology is managed, accounted for, and used to enable businesses to improve their efficiency and their effectiveness. What follows is a subset of the hard and soft benefits of implementing utility computing as a business enabler. At Cassatt, we believe that these benefits fall into three categories:

1. Reduced total cost of ownership
2. Increased reliability and availability of business services
3. Improved business agility and value generation

Utility Computing Reduces Total Cost of Ownership

Hard Cost Savings

Consolidating servers down to as few as is necessary is becoming the norm with virtualization. With automation driving a company’s IT infrastructure, Cassatt believes that the current 5–8% utilization rate companies experience can be expanded to reach 50% or higher. This will allow for substantial reductions in capital expenditures as well as savings in key areas such as power consumption and even real estate that houses data centers. The utility model makes the most efficient use of these systems and also allows IT to make the most efficient use of the software licenses by allocating software based on demand.

True Commodity Acquisition Model

The nature of the utility model reduces the ability of hardware vendors to lock in their customers. Inside the utility, incremental compute power can be added on demand with little concern as to vendor-specific hardware or even many software issues. This eliminates vendor lock-in and leaves customers free to choose the best hardware at any given time based on factors such as price/performance, power consumption, form factor, and reliability.

Labor Cost Savings

Many sources of cost associated with human operators can be reduced by moving to this new model. Part of the promise of utility computing is the realization of the autonomic properties of self-configuring, self-healing, and self-optimizing IT operations. For example, the following operational efficiencies can be automated or at least substantially improved:

“Given the proper control, a company’s data center can use commodity hardware and software in a way that maps exactly to business needs at any moment in time.”

“CIOs that I talk to tell me their jobs can be boiled down to one thing: a CIO must deliver the quality and quantity of service for their key applications to his or her business as efficiently and effectively as possible. The underlying technology exists only to make sure the business runs.”

“In fact, separating the technology infrastructure from the applications enables IT departments to shed the restrictions that drive up costs and hinder responsiveness. As the demands on your business applications change by the month or by the minute, your computing infrastructure can respond, and IT becomes a utility, available on demand.”

Bill Coleman, Cassatt CEO

- New application service roll-outs and updates
- Server additions or removals
- Failure/disaster recovery for hardware and/or software
- Scheduled re-configuration of hardware systems
- Load balancing between different machines and software services
- Re-prioritizing software, hardware, and services to meet demands

Decreased Power Consumption

Reducing the total number of servers required to host an organization’s workloads will decrease power costs. In addition, a utility can automatically power servers off and on as required based upon demand and/or user-defined policy (e.g. shutting them down during off-peak hours), thus dramatically reducing power consumption.

Cost Transparency

The ability to effectively measure and meter key variables in the system can drive true cost transparency in the system. This will allow IT to effectively bill the line-of-business for the amount of service they’re receiving based upon the true cost of that service. This will also allow management of an organization’s business to better drive share-holder value by making improved ROI calculations on various services. This also gives the IT organization a way to much more effectively track assets (both hardware and software), making it possible to quickly generate reports that list the average and maximum number of licenses for a given piece of middleware or application in use over a given month. This can prevent companies from overbuying software and lead to dramatic efficiencies (both in acquisition costs and on-going maintenance costs).

Utility Computing Improves Application Reliability and Availability

Quality of Service Improvement

Service outages in mission-critical systems are highly costly. A utility infrastructure, by its very nature, is self-healing, thus reducing system downtime as well as system performance degradation. The utility measures, analyzes, and responds to situations in the data center environment, making any application highly available—allowing the system to automatically correct software failures (e.g. by restarting or rolling-back failed applications) or to correct serious hardware faults (e.g. by automatically replacing a server with bad RAM and then diagnosing the problem). If system capacity falls below required levels, the utility will gracefully degrade performance on lower priority systems in favor of more important ones, based on user-defined priorities and policies. In addition, the system measures customer-defined performance metrics and can automatically assign capacity to applications that are under-performing.

Improved Security

By reducing custom configurations and moving to a more standardized hardware and software model, IT departments can drive out custom one-off configurations that may lead to unforeseen security issues. In addition, because software and hardware are centrally managed in the utility, all actions (e.g. software updates, hardware replacements, etc.) are automatically logged and tracked. This can assist in compliance reporting needs.

Utility Computing Increases Business Agility and Value

Improved Time to Market

Software development and test cycles are critical to deliver new services in customer IT environments. In addition, the act of moving software from development to staging and finally into production can be almost totally automated in a fashion that ensures consistency and eliminates many causes of human error that can lead to failed test cycles or deployments. New versions of custom applications can be quickly rolled out. Finally, new applications do not require dedicated servers; they can be added when available and run at the priority required.

Service Level Automation

This is the capability to dynamically produce the quantity and quality of application services required based upon automated and dynamic user-defined and user-managed policies. This will help align IT with the business, reduce overbuying and increase efficiency. By combining many of the advantages outlined above, an organization can move to a true pay-as-you-go, just-in-time model for IT service.

A Driver for Standardization

The utility model can be an enabler to an IT program to drive standardized software combinations for various services. It would be possible to quickly deliver preconfigured sets of software to a development or deployment team. Constructing these custom stacks, which previously might have taken weeks, can be reduced to mere minutes. In addition, automation technology enabling a utility like this makes it possible to quickly roll out software changes to those approved stacks quickly.

Utility Computing: What it is, and What is Needed to Make it Work

Service Level Automation is the goal of utility computing—to deliver the quantity and quality of application service flow required by the consumer as efficiently and cost effectively as possible, limited only by total capacity available. This goal cannot be achieved through grid computing, virtualization, system management, provisioning, or IT automation.

In order to automate and dynamically deliver software application service levels, the underlying infrastructure components must be decoupled, breaking the traditional interdependence between physical and logical infrastructure facilities for computers, storage, communication hardware, application software, application containers, and other enabling services. It is only by decoupling these elements that the computing infrastructure can be automated as a utility to regulate and respond to the three dynamics of utility computing service flow:

1. Consumer-defined and consumer-managed dynamic business policies, and priorities
2. Data center(s) capacity and capabilities
3. Dynamics of the demand flow

A New Service Delivery Model

Because of the need to decouple the computing infrastructure, a new service delivery model implemented by a management service that is independent of the computing infrastructure is required for utility computing. The purpose of this new service delivery model is to provide service level automation independent of the computing infrastructure. The management service is the independent facility that guarantees the delivery of required application services at the time and location required, limited only by the total capacity and capability of computing infrastructure available. The basis of this service delivery model is that it manages three fundamental capabilities that meet five basic requirements as follows:

Fundamental Capabilities

1. Application Containers

Application containers are the software services upon which application services are generated. Examples of containers include, but are not limited to, virtual machine managers, operating systems, and web application servers; containers may also contain other containers.

2. Application Service Flow

The application or multiple related applications in combination that implement a service flow.

3. Service Level Automation

The capability to dynamically deliver required application services based upon automated and dynamic user-defined and user-managed policies (including, but not limited to, customer, location, quantity and quality of service, service type, priority, duration, pricing, date and time of day, business demands and conditions).

Basic Requirements

1. Scale Independence

The incremental network and server overhead introduced by the management service must be independent of the number of servers or services being managed. Without this capability, utility computing will not be commercially viable for large-scale operations.

2. Data Center Independence

The management service must operate as a “metaware” control function independent of the computing infrastructure and, therefore, not require any changes to the applications, hardware, system software, or system management tools. This requirement provides the necessary decoupling of service levels generated, metered, and delivered from the capacity and capabilities of the data center. Without this level of independence, it will not be possible to achieve that goal of utility computing. This also enables service level automation to span data centers making disaster recovery dynamic and policy based while enabling global utilization trade-offs (e.g. “follow-the-moon” server utilization and power management).

3. Dynamic Provisioning

The management service must support the ability to dynamically provision at start-up and run-time all software to support installation, capacity management, service levels, business policies, and software upgrades. This requires near-real-time server activation without the overhead and time required by current provisioning systems to build and activate servers.

4. Policy Automation

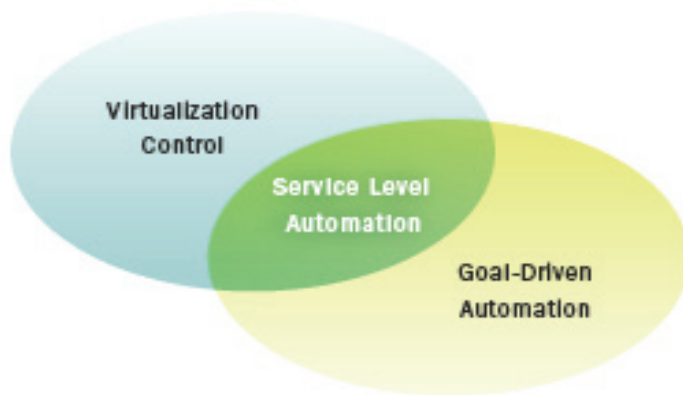
The purpose of the management service is to dynamically adjust the application service levels to implement user-defined policies limited only by the capacity and capability of the available computing infrastructure in order to provide guaranteed service levels for single and/or multiple application service flow(s) in response to changes in the environment. The policy facility must also allow the service provider to manage the delivery and meter the capacity that they deliver to individual customers independent of the customers' ability to, in turn, set and manage their own policy for how their capacity is utilized.

5. Data Capture Capability

The management service must acquire the data necessary to enable the service provider to implement the metering and auditing facilities necessary to control, audit, bill, and account for services provided in complete detail. The ability to capture this data must also be used to educate the system about the performance characteristics of combinations of resources to allow it to improve the quality of its scheduling decisions over time, enhancing its ability to provide the requisite services with high efficiency.

How Cassatt Helps IT Move Toward Utility Computing Today

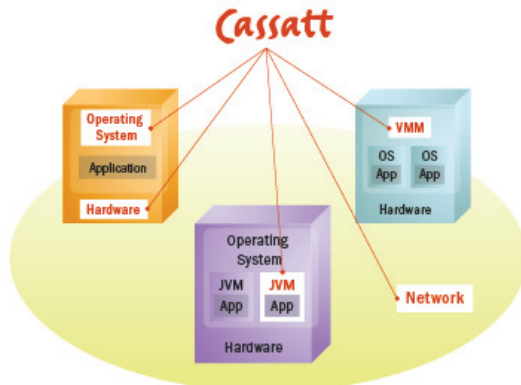
At Cassatt, we believe that our Collage software is the first complete management service that implements utility computing. The organizing principle of Collage is service level automation. The service delivery model incorporates all of the capabilities and basic requirements of utility computing. Collage implements service level automation as follows:



Virtualization Control, Goal-Driven Automation, and Utility Metering

Virtualization Control

Virtualization exists at many layers of the system including hardware, operating systems, networks, and various kinds of virtual machines such as Java Virtual Machines (JVMs) and general purpose virtual machines such as VMware and Xen.

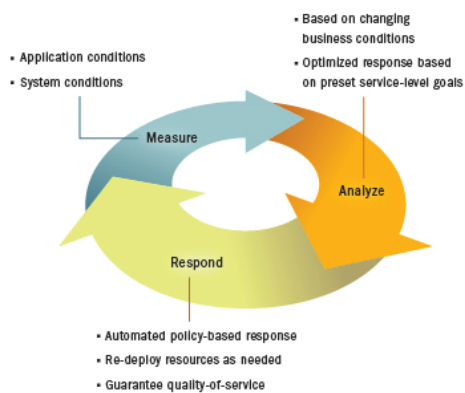


Cassatt Virtualization Points of Control

All of these levels of virtualization are elements of a truly dynamic, fluid data center. However, with all of this dynamic behavior comes a huge increase in complexity. Cassatt begins by becoming the single control-point for all of these virtualization technologies, radically simplifying their operations.

Goal-Driven Automation

The next step required to effectively drive cost-effective service levels is automation. This is the capability to deliver required application services based upon automated and dynamic user-defined and user-managed policies (including but not limited to customer, location, quantity and quality of service, service type, priority, duration, pricing, date and time of day, business demands and conditions). More than half the cost of IT is not in the hardware and software itself, but in the operations of those resources. Cassatt takes a goal-driven approach, where each application (whether single- or multi-tier) in the system is assigned a set of well-defined, quantitative service level goals and constraints. Collage continuously monitors all application service levels to ensure they continue to meet their assigned policy levels; if they are not meeting these requirements, the software can take automatic corrective action. We call this cycle, *Measure, Analyze, Respond*, as shown in the following diagram.



Cassatt's Measure, Analyze, and Respond Loop

Utility Metering

The goal-driven nature of Cassatt's automation engine is a key in the overall scalability of the solution. Other "automation" products are built around meticulously produced code or scripts that specify each scenario to be automated. In environments with many resources and applications, this quickly becomes unmanageable as rules must be continuously edited and changed whenever the environment or business goals change.

The key differentiator between an automated data center and an IT utility is the ability to generate metered activity. Because Cassatt Collage inventories assets, monitors services, and controls instances of hardware and applications, it is an ideal tool to measure granular resource consumption. The resulting data is available to control, audit, bill, verify compliance, and account for services provided in complete detail.

This data provides highly accurate measures of license use and infrastructure utilization and is easily applied toward pay-as-you-go and/or charge-backs, variable cost accounting, or resource planning systems. Collage even enables true business metrics (such as variable cost information) to be leveraged as infrastructure control mechanisms. This is much the same model whereby businesses curtail electrical consumption during periods when energy rates are highest. Use of some chargeback and/or pay-as-you-go pricing in IT environments is forecasted to be as high as 40 percent of enterprises by 2010 (Saugatech Technology, 11/23/2005). Collage records each change to the infrastructure, planned and unplanned. The derived infrastructure measurements include, in part:

- Physical and Virtual Servers: time powered-on or created, time and number of CPUs and/or blades assigned to applications, amount of memory and CPU bandwidth assigned to applications
- Application/OS/payloads: number of concurrent licenses in use, including duration each licensed application is active
- Network: devices, topology configuration, traffic, and bandwidth use
- Inventoried (but idle) resources: such as powered-down infrastructure residing in spare, failover, and maintenance pools
- Service levels: such as availability and responsiveness, which can be translated into cost "premiums" or pro-rated

Collage records all activity in a format that is easily exported for analysis (from a simple Excel spreadsheet to SQL tables dumped directly into a data warehouse). Data can then be easily analyzed and mined for relevant information. There are currently dozens of IT-specific costing and accounting packages that will easily import this data, cross-reference it with capital expense, depreciation data, budget data, and services costs. They can then pro-rate costs per any number of conventions and generate on-demand reports, such as:

- Fee-for-services and transfer costing
- Flat fees and simple resource accounting
- For-profit zero-based chargeback
- Business unit/functional charge-backs
- Break-even estimates & periodic adjustments
- Standard rates & negotiated prices

Vendor Neutrality

A key aspect of Cassatt's product philosophy is vendor-neutrality. Customers have many different types of hardware, middleware, and applications in their IT portfolio. IT must maintain the right to choose the best solution for any given need without fear of vendor lock-in. Because the Cassatt Collage architecture is independent of the computing infrastructure, it works with any vendor's hardware, software, system management, and provisioning systems without change, thus ensuring the customers can choose the best option for those components. Collage is also designed to work with existing applications, operating systems, system management, and provisioning systems without modification.

Conclusion—Taking Advantage of These Shifts Today for Competitive Advantage

Today's hardware and software trends are driving a fundamental shift in the economics of IT infrastructure and the data center. This is creating a new opportunity for nimble organizations to take steps now that will let them outpace their competitors in their ability to respond to changes in their business, while doing so at drastically reduced expense levels. This move—from a static data center to one operated as a utility—is one that large organizations in industries like finance, pharmaceuticals, telecommunications, government, and others, are starting today.

A first step is to address the capital and operational expenses with virtualization and automation. This is where Cassatt can help. Cassatt customers begin the steps toward a utility in as small or large of an effort as they prefer. No matter what the scope, the result is a dynamic, real-time infrastructure for key applications that provides the service levels that are required.

For more information on ways to get started and the type of returns that Cassatt customers are getting on their investments, go to www.cassatt.com and contact a Cassatt representative. Detailed technical information and demonstrations of Cassatt Collage are available on request.



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 t: 408.437.7600
1740 Technology Drive f: 408.437.1364
San Jose, CA 95110 www.cassatt.com