IDC
Analyze the Future

WHITE PAPER

An Overview of Compute-Intensive Utility Computing

Sponsored by: IBM

Christopher G. Willard, Ph.D.   Earl Joseph, Ph.D.
Lionel Lamy
January 2007

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

Executive Summary

Utility computing is a usage model in which customers pay for computational resources through an established fee-per-time schedule, as if they were utilities. These fees can cover hardware, software, or storage usage, and fees for associated services rendered. Typically the utility computing customer logs into a remote computer hosted by the provider, possibly uploading data pertinent to the calculation. The provider then logs the time and resources used. Utility computing as a service has existed for many years, but with the increased demand for high-performance computing (HPC) resources, it is emerging as a worldwide trend.

As HPC customers strive to maximize their investments, other related trends have also emerged, such as grid computing and on-demand computing. Grid computing is a model in which networked computing systems are virtualized such that the customer does not know (or care) which part of a system is providing the answer, just as most residential customers using electricity on a power grid do not know what type of facility provides the electricity that reaches their homes.

Offerings such as IBM's Deep Computing Capacity on Demand (DCCoD) combine both utility computing and grid concepts and thus might be called "utility grids." The business model is based on utility computing approaches of providing customers with computing resources or capacity on a pay-for-use basis. This business model is enabled by grid concepts and technology.

Although various forms of utility computing have existed for some time, IDC believes that a confluence of market factors is making it more prevalent than before. More companies than ever are seeking competitive advantage through high-performance computing, but many lack the facilities or budgets to expand their own infrastructures to meet the internal demands of their scientists, engineers, or analysts. Furthermore, customers who are seeking to become more responsive to their own customers' needs might wish to have immediate flexibility in their computational capabilities. Lower cost and increased flexibility are two of the main goals of utility computing.

Computationally intensive utility computing is a complex model. Not all equipment providers can successfully deliver it. The required technology, processes, and skilled staff need to be carefully managed *as a whole* in order to make the utility model a success. The utility computing model provides a solution to many compute-intensive problems and has great potential for addressing mainstream IT challenges.

In This White Paper

This white paper addresses how demand for HPC capabilities is beyond the reach of many potential customers that do not have the means to acquire and manage an HPC infrastructure. As such, the demand for HPC capabilities via a utility computing model is expanding on a global basis and potentially leading to mainstream usage. Utility computing is especially useful to customer organizations that have peak workload demands and those that have limited requirements for specialized resources. Broadly speaking, there are two current utility models for addressing computationally intensive workloads.

- **Variable utility.** In the variable utility model, the customer contracts for an augmentation of its own IT infrastructure for a specific amount of time, or a specific project (which can be cyclical). This model is usually more expensive on a CPU basis, but the customer only pays for capacity they require and reserve, making it less costly than traditional purchase models. Customers wishing to pursue a variable utility model should consider what their level of commitment is and how much they will pay for incremental levels of service. The better they can predict their variability, the more efficient this model will be.

- **Dedicated utility.** When constrained by available floor space, compute capacity, power, or other major issues, customers can gain access to third-party IT assets configured and managed to meet their specific requirements. In the dedicated utility model, customers use the compute-intensive equipment as an extension of their IT infrastructure. This type of utility model typically involves longer-term contracts.

Utility computing can be viewed as an additional set of resources that organizations can add to their overall computing portfolio. Similar to different investment types in a financial portfolio, utility services provide insurance/hedges against workload risks and access to tools that have previously only been available to those able to make large investments. We see the variable and dedicated utility models as representing two points on a spectrum that trade-off customer commitment with cost and service levels. It is important to note that customers may look to combinations of service levels to meet their ongoing requirements. This IDC white paper examines organization approaches to utility computing and looks at potential benefits of utility strategies through the use of a case study focusing on IBM Deep Computing Capacity on Demand's utility computing offerings.

Types of Utility Computing Offerings

Grid-based utility products range from a basic offering of raw computer cycles up to products approaching hosted services or outsourced computer services. For purposes of definition IDC considers four reference product levels:

- **Cycles for money.** A strict dollars-per-processor hour-based product with the vendor supplying the computing environment (i.e., servers, storage, systems management, power and cooling, etc.) and the customers supplying the software stack. In addition, customers are responsible for managing and running their workloads. In this case the customer assumes responsibility for all applications licenses and/or provides internally developed codes. Customers who wish to retain full control of their application environment may prefer this model.

- **Applications development environment.** In addition to raw cycles, the vendor provides the software middleware layers to allow customers to develop and implement applications on the providers' systems. This model appeals to customers who need complete development environments, not just compute cycles.

- **Applications services.** A service-based around providing access to a specific applications suite that is optimized for the service providers' hardware. The goal in this model is to provide a turnkey solution to the end customer. The Exa PowerFLOW case study in this white paper is an example of an application services utility grid.

- **Multi-applications workload support.** A service that provides an operational environment to match all or part of the customer's environment with multiple applications supported. This model could apply to customers who are testing new markets, providing ease of entry and exit with minimal investment.

The reference levels provide points along a spectrum of services where vendors also have the ability to "mix and match" offerings from different parts of the spectrum.

It is important to note that just as there are different models for providing utility computing resources, there are also alternative channels for providing them. Major system vendors are now offering utility computing services, but for various workloads these services could also come from independent software vendors, value-added resellers, selling partners, or standalone utility providers.

The Worldwide Market for Utility Computing

The worldwide HPC market has undergone three straight years of double-digit growth, growing to $9.2 billion in 2005, its highest-ever revenue mark. This growth has been driven by such economic/geopolitical factors as strengthening economies, world demand for petroleum, increased competitive requirements for product design quality, faster time to market, and increasing requirements in national security segments. In addition, improvements in price/performance have attracted increased spending in existing HPC accounts as well as new customers.

Utility computing is a generic term, which means different things to different organizations. Loosely defined, utility computing is the provisioning of computing power on a metered, pay-per-usage basis. The customer effectively pays for the hardware, software, storage and services associated with the computational resources actually used or reserved.

Utility computing can incorporate grid computing or on-demand computing. Both models are related yet distinct. Grid computing entails the virtualization of networked hardware to provide the relevant level of computational power required by the customer. Effectively, customers of grid-computing models do not know which part of the system actually provides the answer. The system is made up of separate hardware, working as a seamless and coherent unit for the purpose of computationally intensive problem solving.

Offerings such as IBM's Deep Computing Capacity on Demand (DCCoD) combine both utility computing and grid concepts and thus might be called "utility grids." The business model is based on utility computing approaches of providing customers with computing resources or capacity on a pay-for-use basis. This business model is enabled by grid concepts and technology.

Customer Purchase Strategies

In purchasing utility services, customers need to decide the best pricing approach. If they are uncertain as to their requirements, then a pay-as-you-go type approach will work well. If the customer has a clearer understanding of their requirements over time, they can save money by making a longer-term commitment. This leads to a spectrum of strategies, for the purposes of definition IDC considers three reference points along this spectrum:

- **Pay as you go.** Customers purchase services only when they need it, on a metered basis, paying for resources in small increments and only when used. This strategy provides vendors with no information for capacity planning or assurance of longer-term sales, thus it commands a premium price as the vendor has to absorb the risks associated with demand fluctuations.

- **Variable utility.** In the variable utility model, customers purchase a specific amount of compute capability to be used within a given amount of time. This purchase often includes an initial stop of setting up a "base presence," which includes establishing a VPN connection, configuring management nodes with software, and so on. In this case customers pay for capacity they reserve in advance; allowing vendors to provide better pricing than no commitment, pay-as-you-go models. However, customers often find it difficult to anticipate usage levels, which tends to push up cost per a cycle if they over estimate requirements or forces the purchase of extra cycles at pay-as-you-go rates if they underestimate.

- **Dedicated utility.** This type of utility model typically involves long-term contracts, with 100% of specified physical resources dedicated to the customer's workload. The customer spells out the details of the computer system, including software, storage, etc. Communication between the utility and customer sites is usually via a point-to-point dedicated network. This approach may be driven by constraints on available floor space, compute capacity, power, or other major issues. In the dedicated utility model, customers use the compute-intensive equipment as an extension of their IT infrastructures.

Similar to short-, medium-, and long-term bonds, these approaches are all variations on accessing the same basic product that provide different levels and types of returns and require different levels of commitment from their buyers. Like bonds, the above strategies represent a broad range of options, and customers may opt for some combination of these strategies to meet their requirements. The important point is that the better the customer is able define their requirements, the more the customer can expect with regard to flexibility on pricing from the utility vendor.

Except for the most basic pay-as-you-go approaches, customers are investing more than the cost per hour of processor time, they are also investing their time and effort to set up and manage the operational relations and, more important, they are investing their trust in that the service provider will be able meet their requirements. Thus, in addition to pricing, there are a number of other factors that customers need to consider when determining with which provider to partner. These factors include provider experience — both overall and at the industry/applications level — total resources available (e.g. server types, storage device, software), provider ability to meet customers short- and long-term scale-up requirements, additional services (e.g., initial consulting and setup support, system management support, applications support), security policies, location, provider-specific software tools, and facility/system-level high-availability features.

Technical Utility Grid Market Overview

The technical computing market has often played the role of entry point and proving ground for new technologies and product strategies. It has held this role for two major reasons: it is "cycles hungry" and "technology independent." Scientific researchers, engineers, analysts, product designers are always in need of more detailed, complex, and accurate analyses within their field, as well as the ability to consider as many options in a given amount of time as possible. Subsequently, computer models and demand for cycles expand to fill the available resources.

For example, a major factor driving requirements for increased capacity is time pressure for product development. Pressure to bring drug compounds to market as quickly as possible, pressure to assess the impact of aerodynamic changes to a Formula 1 car in time for the next grand prix, pressure to calculate the relative impact of market fluctuation on share prices, pressure to reduce cycle time in order to bring new automobile models to market faster. High-performance computing can play a pivotal role in speeding up all these cases by allowing scientists, engineers, and analysts to understand the fundamental aspects of their problems more quickly and with less use of physical experimentation, prototyping, and acceptance of risk.

Thus the technical IT manager is faced with the problem of providing as much computing power as possible, rather than trying to consolidate applications from underutilized machines.

The research and development processes continually generate new and more demanding problems, which in turn generate requirements for more powerful tools including computers and applications software. The combination of increasing applications requirements and expectations for regular technology upgrades has led technical customers to maintain a high degree of flexibility in their technology choices. Technical customers also desire to maintain as much technology independence as possible in order to take advantage of new technologies or computing strategies.

Utility computing has been used within the technical computing market for many years, a trend that has been driven by companies providing a combination of applications software and computing resources. This trend has largely been centered in the petroleum industry. In 2004 the market entered a transition phase as major vendors such as IBM began offering generalized utility computing products. This transitional growth continued into 2005, with the market expanding by an order of magnitude over a two-year period to about $25.5 million.

IDC expects the technical utility grid market to continue to grow at double-digit rates through 2010, with a compound annual growth rate of over 50% reaching about $162 million by the end of the forecast period. IDC feels that the market has hit an inflection point and is in a high-growth or "free rise" section of its life cycle. At the fundamental level, IDC sees the market as exploiting green field opportunities made available by a combination of high-bandwidth access to faculties via the Internet, HPC computers being built around more commodity components and processors, and significant investments in hardware and support infrastructures made by major systems vendors. IDC also sees the potential for further market expansion driven by the entry of applications-based services into the market, and such companies can act either as channels to larger general-purpose utility grid suppliers or as independent "boutiques."

IDC believes growth was initiated in the United States because the majority of the technical computing market (about 50% of revenue in 2005) resides in the United States and continues to grow, plus customers have experience with the model. That said, IDC believes that this has always been a worldwide market that continues to show signs of growing international interest.

In particular, the Asia/Pacific (AP) market, while small, has the opportunity to exhibit substantial utility computing growth. Customers in the market are rapidly being exposed to the potential benefits of computation, but budgets could be slow to respond. Utility computing could give growth-seeking customers in AP an opportunity to see some of the benefits of scalable computing without a steep initial investment.

Within the European market, customers have traditionally been more reluctant to adopt hosting or work with external services providers when compared to the U.S. and Japanese markets. This relative cultural reticence, while hindering the development of the market, is bound to give way to the economic pressure in which European organizations find themselves: cutting costs, providing competitive advantage through innovative use of IT, and aligning business and IT requirements. This cultural reticence is decreasing for outsourcing, but remains stronger for the adoption of new delivery models. Within the past two years, IDC has witnessed concrete examples of utility computing being implemented within the European marketplace, such as the recent announcement that BNP Paribas is using the IBM London DCCoD center to deliver its offsite computing requirements to the support the bank's expanding derivative operations. As the benefits of such solutions become known within the market, IDC expects a snowball effect to take place. In other words, as the early adopters are deriving competitive advantage via utility computing, the rest of the market will take notice and inquire about the potential applications of this model within their organization. At the same time, service providers delivering utility computing to the market will gradually become more proficient — as more and more customers adopt their solutions — and replicate these solutions faster and more effectively.

IDC Analysis: Utility Computing Market Drivers and Inhibitors

Market Drivers and Opportunities

Drivers for the utility model can be viewed as first providing an "instant infrastructure" for customers, which is allowing customers to bypass system acquisition, installation, and start-up phases, and instead move directly to operational phases. A corollary to this strategy is that organizations can just as quickly dissolve that infrastructure when a contract or project comes to an end, thus avoiding ongoing expenses. Examples of how this approach can be used include:

- **Supporting peak demand problems.** This approach occurs when the customers computing infrastructure is sufficient to meet normal demand but cannot meet demand on all occasions. Variations in demand can be due to cyclical requirements, fluctuations in the number of projects or contracts, unexpected design problems requiring extraordinary support, and so on. Organizational strategies for these situations include: purchasing excess capacity that will be underutilized the majority of the time, attempting to quickly expand capacity as needs arise, or purchasing the capacity through a utility provider.

- **Supporting start-up operations.** New ventures with significant computing requirements may choose to use a utility service in order to shorten their time to market and also reduce short-term costs for capital equipment

- **Supporting company growth where datacenter space becomes a limiting factor.** In addition to the actual floor space consumed by server racks, considerations such as power consumption, cooling, and staffing are all factors that can drive an investment in utility computing versus systems purchases.

A second major driver for utility computing is a vendor's ability to provide access to "specialty resources." In cases where organizations either cannot justify or simply afford a special resource based on amount of use or total cost, a utility-based solution can provide the resources for short periods of time at acceptable costs. (This is essentially a rent versus buy decision.) Resources may range from access to high-end supercomputers to specialized software packages. In this case the utility vendor can spread the underlying resource costs over a large number of clients, but also assumes the risk of not getting adequate return on investment for the resources.

The opportunities for computationally intensive technical utility computing customers essentially fall into four categories, summarized below:

- **Extensibility.** First and foremost, the utility model provides customers an extension of their own IT infrastructure without the drawback of having to own, manage, maintain, and upgrade the technology.

- **Cost savings.** The cost advantages of a utility model can be compelling, as customers only pay for the compute capacity they require while gaining access to the capacity of state-of-the-art supercomputers they may not otherwise be able to access or afford.

- **Flexibility.** The utility model enables clients to be extremely flexible and react to market changes rapidly. The compute-intensive utility model can cut the time to market of a new product dramatically, by avoiding all the delays associated with acquiring and implementing the required IT infrastructure.

- **Availability.** Finally, while bounded by business conditions and scheduling, the utility model is always available, and the client can rely on the utility provider to deliver the required compute capacity in order to run business-critical compute-intensive workloads.

From a business perspective, IDC believes computationally intensive technical utility computing is not only a business enabler, but it also allows companies to focus on their core competencies — unlocking potential effectiveness and efficiencies within their organization instead of dealing with IT issues.

Inhibitors and Challenges

On the other hand, it is undeniable that key inhibitors remain in place for the market creating challenges to vendors of utility computing offerings.

- Security is a key concern for potential customers. Industries are extremely competitive, and security of the data is priceless to the corporations concerned. However, IDC believes security is neither a technical issue nor an intrinsic problem in itself, but the perception of security associated with hosting applications and data on an infrastructure managed off their premises. In a utility computing environment, security can be addressed through planning and careful selection of the external IT provider.

- Another challenge to the adoption of compute-intensive utility computing is the difficulty associated with demonstrating ROI. Indeed, as potential clients investigate the model, they often have very little to which they can compare. In addition, there are many hidden cost elements that need to be factored into the ROI beyond just the servers and the resources to manage them, including space, power and cooling costs, and network costs, among others. Educating the market is crucial in overcoming this hurdle.

- A third inhibitor is dealing with the complexity of many HPC problems and environments, along with a concern over level and consistency of performance that they will receive. Some HPC customers (and applications) are focused on the bandwidth and latency of systems in order to extract the highest performance possible, and it is hard to prove to them that a utility computing approach will consistently deliver the required performance level.

- Finally, some workloads simply are not conducive to being run over the Internet. This can be due to bandwidth limitations or the amount of interactivity required. If the customer needs to continually evaluate results of many short runs with large data sets that need to be upload at each iteration, utility computing might not be an efficient model.

IDC believes that once the utility computing model has proved itself in the technical computing arena it can expand into more commercial applications, which rely less on high-performance computing. Within this scenario, the benefits of utility computing will be the same, but the importance attached to each benefit will vary. In terms of inhibitors, security will remain a prime concern and the demonstrability of the ROI case for utility computing will also be crucial within the commercial applications side.

Case Study: Exa Corporation Mission, Products, and Utility Computing Strategy

Founded in 1991 and employing over 100 people, Exa supports a worldwide customer base in industries ranging from automotive, aerospace, petroleum, chemical processing, environmental engineering, material processing, power generation, and HVAC (heating, ventilation, and air conditioning). Exa Corporation provides computer-aided engineering software and consulting services for computational fluid dynamics problems, including airflow over automobile and airplane surfaces, flows in plastic injection molding, the movement of fluids and heat transfer in nuclear reactors, and the effects of spray cooling and humidification in climate control systems. Exa has partnered with IBM to offer engineers on-demand access to Exa's flagship PowerFLOW simulation capabilities via a secure Internet connection.

Stephen Remondi, Exa's President, CEO and cofounder, describes the company's value proposition as "providing software and services for simulation of physical environments to help customers avoid physical testing, and be able to make time-critical decisions based on data and not only opinion."

Why Utility Computing and Partnering with IBM?

Exa began offering its software and services via a utility grid model in the 1999–2000 time frame. At that point the company purchased computers, installed its software environments, and managed all aspects of the computer operations. This provided the company with a fixed capacity infrastructure for utility-based software and services products. This approach was successful because the company understood the problem set, the applications, the computational requirements, and how to best fit the applications to the computer hardware.

However, Exa learned that the greatest limiting factor for its customers' use of computational methods, and thus its own growth, was access to HPC computing capacity. Exa tried to provide the resources to maintain and grow an HPC infrastructure for customer use, but it was constrained by capital limitations. The solution to what was essentially an infrastructure scaling problem came from IBM's DCCoD program. Mr. Remondi notes that "as soon as we heard what IBM was providing, we knew this was a perfect match for us, because of IBM's capital infrastructure and ability to scale-up the environment. So now, instead of having fixed capacity, we have from our point of view and our customer's point of view unlimited capacity."

The partnership with IBM provided a way for Exa to get customers access to its software running on high-performance computers, and it allowed Exa to get out of the capital equipment and systems management business and to focus its energies on its core competencies as a software development and consulting company.

Getting Answers on Time: Opinions Versus Data

Exa's goal is to deliver computational capabilities that allow its customers to avoid building prototypes and doing physical testing (e.g., wind tunnel testing, climactic tunnel testing, road testing, etc.). Mr. Remondi notes that "Our customers want answers about how their products perform upfront in the design process. Problems that we are solving are very, very computationally intensive, which means a single simulation can consume from hundreds to, more typically, thousands of CPU hours."

Given such demands the issue for customers then becomes one of turnaround time versus system cost. For example, using "back of the envelope" calculations, a company might buy an eight-processor cluster to run a simulation that takes 2,000 hours for each design iteration. Such a run would take roughly 250 hours (or about 10 days) to complete. In contrast if the engineers could run that same simulation on a 128-processor cluster, it would take about 16 hours or be done roughly over night. "Engineers really start getting addicted to this ability to turn problems around," says Mr. Remondi. "It allows them to spend much more time thinking about the design and not just waiting."

Mr. Remondi explains why turnaround time is so important to the product design process, "What happens, for example, in an automotive development process is the design studio releases the data on Thursday and the designer review meeting with the VP of engineering and the VP of product development is on Monday morning. They need to know the direction they should take with the product. People can come to the meeting with opinions, or they can come with data. The people who win in those meetings, the people who have impacts on the vehicle design, are the people that show up with data — not just an opinion of what they think will happen or what they think is better."

Working with Exa and IBM: Customer-Use Models

Customers have a range of strategies in working the Exa and IBM DCCoD services. The strategies are based on what parts of the process are assigned to Exa and what parts are kept in house by the customer and then on the amount of computer time that the customer needs. Example use models include:

- **Exa as one-stop shop.** The customer provides the data, model specifications, etc. to Exa, which in turn is responsible for pre- and post-processing, the full problem setup, running the simulation, and so on. Exa then delivers a simulation results report to the customer. This approach is often an initial strategy for customers, as it allows them to determine where PowerFLOW fits into their design process and to get up and running while learning the operational details.

- **Exa as administrator.** The customer handles most or all of the pre- and post-processing, problem setup and so on, but hands-off the job to Exa to run. In this case, Exa is managing the system configuration use and is able to optimize costs by scheduling and sharing its base infrastructure at IBM.

- **Exa as software and setup provider.** Customers do the pre- and post-processing, set up their own job runs, and run the simulation. In this case, the customers purchase their own dedicated base presence at IBM, including storage and time for additional analysis. Exa is responsible for account enabling, account management, and making sure the software is available and properly configured on the IBM systems at run time. In this case Exa does not manage any customer data. This approach is effective for customers intending to use large amounts of machine time in a given year.

Individual customer organizations may use a variety of these models based on whether they use utility computing strategies to support peak workload problems, ongoing requirements, or special case studies.

Exa believes that it is able to price its solutions competitively based on total cost of ownership and/or internal chargeback models. In addition, its customers do not need to be concerned about long-term capital expenditures. Customers can budget on a per-project basis, and when the project is complete the cost of infrastructure goes away. A final cost advantage is that Exa has over 20 years of experience with the PowerFLOW application, and is able to assure that the computers running the code are optimally configured thus providing a high level of efficiency to minimize computing costs.

How the Utility Grid Model Works for Exa

Exa views its utility grid operations as an integral part of its overall business model. The utility model provides two major strategic advantages: elimination of capital equipment management and costs and expanded opportunities in its core business.

A Computerless Computer Company

Exa's utility grid model revolves around the company's ability to deliver computer cycles to its customers. However, its relationship with IBM allows the company to avoid investing in or owning those cycles. Mr. Remondi explains the ROI advantages of this approach as "a win because we can match our computer use to our customers' immediate requirements, and there is no upfront cost, so there is no investment to make. There is no 'I' in the ROI equation." For example, Exa might have thousands of nodes in its environment at IBM running its software for customers. And then, the next day, pare back to a 100 nodes as projects are completed. Thus the company is able to avoid expenses for systems that are not in use, while assuring a return on systems when they are in use.

Expanded Opportunity

The ability to scale up capacity virtually overnight allows Exa to respond to a very broad range of requests. Mr. Remondi explains, "When a customer says, 'I need the results by Monday,' we are always able to say 'yes'. If we say 'yes' some of the time and 'no, you have to wait in line' some of the time, we do not get the business. Either it's on demand, it's available, or it's not."

Mr. Remondi estimates that about 75% of Exa business supported by DCCoD would not be possible with a model based on in-house computer systems because customers would not accept the turnaround time. "They would have said 'Too late. I'm going to make a decision anyway, I'll build the model. I'll go to the wind tunnel. I'll guess. I have to have the answer on time. I don't have a choice.'"

Conclusion

Computationally intensive utility computing is a complex model requiring not only the right level of hardware, software, storage, and associated services, but also the right blend of skills and robust processes to deliver the anticipated benefits. It is therefore not surprising that very few vendors are in a position to successfully deliver utility computing. Technology, processes, and people need to be carefully managed as a whole in order to make the utility model a success. Even with these ingredients at hand, successful vendors will be those who have the experience in delivering leading-edge, complex IT solutions. Service levels are going to be crucial within this market. In conclusion, the utility computing model provides a solution to many compute-intensive problems, one that has great potential for addressing mainstream IT challenges.

IDC contends that utility computing is coming to the fore, due to the combination of two significant market factors: first and foremost the lowering of the utility computing price point, making it financially viable and attractive to CIOs around the world, and second the increasing demand for computing power that cannot easily be fulfilled by capital expenditure alone. As such, utility computing is a natural answer to many customers' needs.

IDC believes utility computing is a model for the future, with strong benefits to its customers. This model can fulfill an important role within the IT market. However, IDC has identified three areas of focus that will need to be addressed before the market can enter the mainstream:

- **Technical.** The technical side of utility computing will need to be addressed. While the solutions are available and working today, operational and practice standards have not yet appeared, and only a very small number of companies can deliver large utility computing engagements.

- **Cultural.** Overcoming cultural reticence is vital. This is especially true within European organizations. Indeed, many companies do not feel entirely comfortable with their data being located not only at an external service provider's site, but even outside of their IT department. Utility computing is a good concept, but it is also one that highlights resistance to change within organizations, especially when data security is concerned. To overcome this challenge, IDC believes vendors will need to take great care educating and reassuring the market that their data is safe and that service levels achieved are inline with or higher than what can be done using internal resources.

- **Financial.** Demonstrating ROI will be crucial in removing hurdles to market growth. The financial benefit of utility computing can be blurred as companies have many hidden costs, which aren't taken into account when comparing the in-house model with the utility model. Again, education of the market will be crucial in addressing this point.