

Next Generation Network Series

Utility Computing and Business Value

Introduction

The telecommunications network has traditionally been seen as a passive infrastructure, designed only for moving data from point A to point B. A new, more expansive view considers how recent developments are taking the network in a new direction. This view looks at the convergence of voice and data, the integration of wireline and wireless and the growth in IP-based services. It sees the next generation network not only as a transport mechanism, but as the foundation for distributed computing architectures.

An important step toward this distributed architecture is the introduction of new, more flexible pricing options. The problem with a traditional, dedicated-server lease is the same problem faced when leasing a car: you're committed to the car for the length of the lease, even if your needs change. Utility computing can address this issue. It enables new business models for procurement and billing that provide CIOs with enhanced flexibility at reduced cost.

New Pricing Models

The first step toward increased flexibility is utilization-based pricing. This model still involves a fixed number of dedicated servers, but instead of a flat rate, customers pay a variable rate based on how much they actually use these servers. Run them lightly, and pay less; then pay incrementally for increasing loads. An even more flexible approach is true utility pricing, where resources are dynamically allocated according to demand, and customers are only charged for the number of resources allocated for a given amount of time.

IT managers tend to build for peak demand, on the principle that underestimating requirements can lead to performance problems, which can then lead to unhappy customers. They prefer separate dedicated resources in their environment to handle eCommerce, payroll, year-end accounting, email, ERP and so forth. But the truth is that these servers rarely run at full capacity, even in extreme conditions. Most studies show that typical average server utilization is around 10%. That means most companies have far more dedicated resources than they need, because they've been acquired and deployed to support a worst-case scenario.

In fact, a blended resource environment makes the most sense.

Dedicated servers on flat-rate leases are still the most cost-effective way to handle daily baseline demand. Usage-sensitive servers can be added to the mix to support typical variations in demand due to normal statistical fluctuations. Pay-per-use servers might seem to be the most expensive option when considered by per-minute price, but if they're part of the portfolio, they can be used to handle unexpected spikes or other unusual situations. They actually save money, since a premium is paid for the short time those resources are used, but nothing is paid when they're not used!

AT&T's studies have shown that most customers benefit from a hybrid application of all three pricing models: dedicated resources with flat rate pricing, dedicated resources with variable rate pricing and dynamically-allocated resources with variable rate pricing. It's similar to how people sort out their transportation requirements. Sometimes you buy a car, sometimes you rent and sometimes you take a taxi. It's a matter of finding the right balance between owned and on-demand resources.

The Value Proposition

How do these new models generate real economic value? At first glance it isn't clear why they would.

In any market, one typically pays a premium for increased flexibility. A refundable airline ticket costs more than a non-refundable one, because it grants the flexibility to choose not to travel on that particular day or flight. Another example can be seen in car leasing. Take a car that can be leased for a flat rate of \$300 a month, which

"Almost every market eventually evolves to offer usage-sensitive pricing in addition to traditional models such as ownership and flat rate. This is as true of computing, networks, storage, and applications as it is for rental cars and hotels."

Joe Weinman, Emerging Services VP, AT&T



works out to about \$10 a day. Compare that rate to a pay-per-use model in the same industry: a rental car. It's very hard to get a similar model in similar condition for a rental rate of only \$10. In fact, that car might rent for \$40 a day. That means a premium is being paid for being able to provision a car dynamically in a distant city without a long-term commitment. In this case, that flexibility and pay-per-use privilege is worth a premium of about 300% (\$40 rather than \$10).

This is the ultimate win-win. Customers may pay a little bit more per unit of time, but the long-term savings can be enormous. In fact, AT&T studies have shown that even if a premium of 50% is assumed for utility resources, savings can be as large as 70% to 80% depending on the usage profile, by utilizing a mix of traditional and hybrid pricing models. When there is no premium for flexibility, then a utility model is even more obviously compelling. AT&T is currently looking to price some utility network services, such as AT&T's Optical Mesh Service, without a utility premium.

Potential Applications

Almost any business can benefit from a utility approach to computing. Biomedical research, for example, involves complex scientific analysis for drug discovery, protein folding, genomics, and the like, which requires secure and cost-effective computing resources that can be ramped up or down according to changing requirements. New research findings can mandate a re-analysis of past results, with serious bottom-line impacts for every day of delay. Utility computing is a perfect solution, since it provides the necessary flexibility and power, while saving money by avoiding an investment in excess infrastructure. It supports a company's need for simpler data management, faster analysis of results and shorter time to market.

Given the global energy crisis, efficient energy exploration is of importance to everyone. The analysis of seismic survey data to support mapping and drilling work is a demanding, number-crunching process that's a natural candidate for utility computing. Companies in this business need a quickly scalable infrastructure that will let them avoid capital outlays and stranded investments, by providing computing power on demand.

"I believe AT&T is the first in the industry to offer usage sensitive pricing for both dedicated and dynamically allocated environments not just for the physical servers but for the application middleware as well."

Joe Weinman, Emerging Services VP, AT&T

Looking Forward

Flexibility in server pricing is a good thing, but server costs are often a small fraction of the total expense. A larger fraction is in application licensing. To a degree, this is a problem that can be addressed by using open source or freeware, but that's a restriction that may not be acceptable to many enterprises. Most companies want to run applications of their own choosing with full support.

The good news is that AT&T will soon be applying the same pricing model it offers for servers to an expanding family of applications, such as web server and application server middleware. This change will enable true utility pricing for computing power, with a huge potential for saving costs and increasing platform efficiency.

It will bring big-company computing power to smaller companies, and could change the way the world does business.

For more information contact your AT&T Representative or visit us at www.att.com/business.

