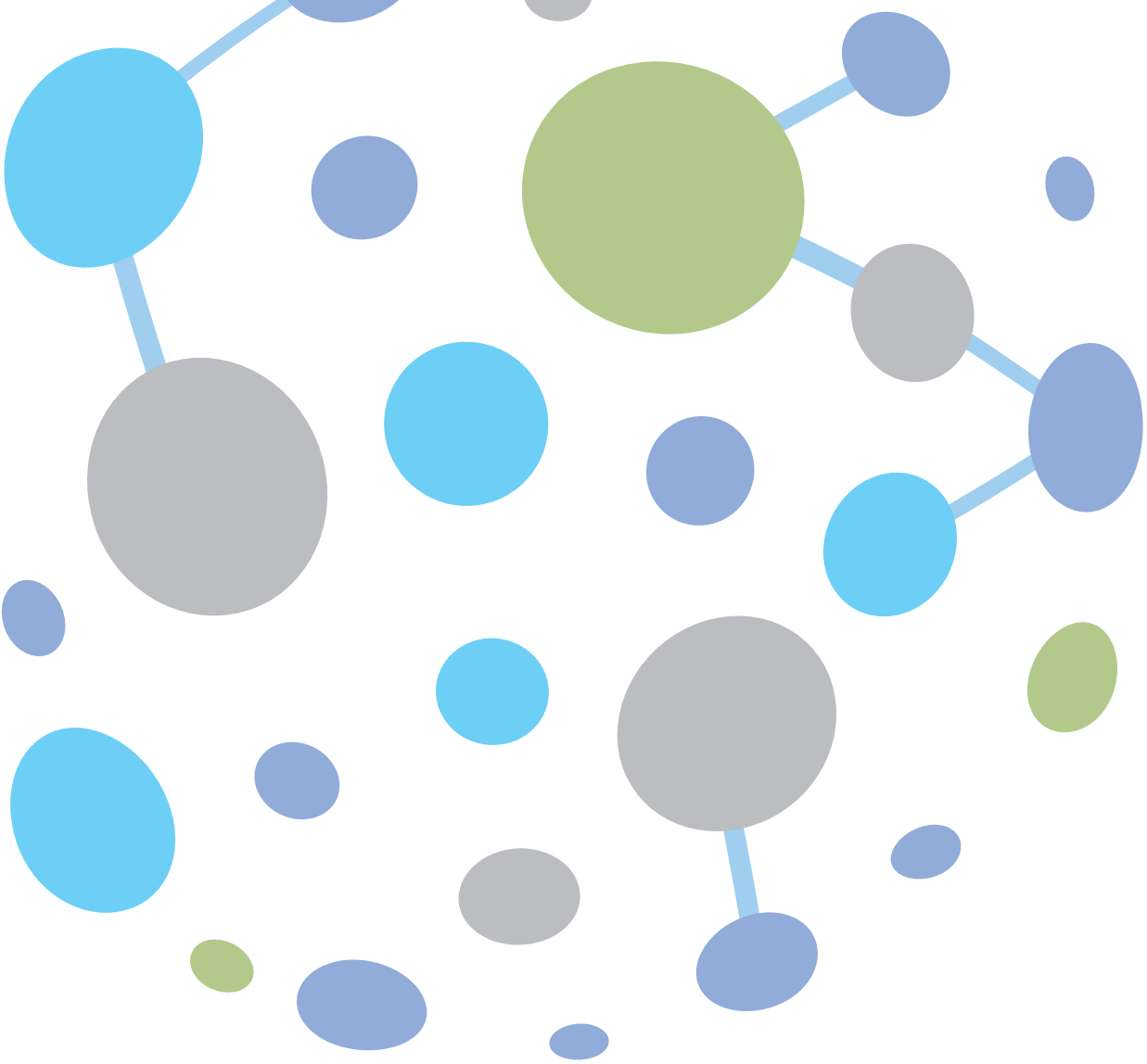


Building a foundation for a high-performance, low cost private cloud

Author: Gary Barnett
Published: November 2011

bathwick group

Contents

A growing number of organizations are building private clouds

Cloud computing is the next evolution of systems architecture that promises to deliver computing infrastructure at lower cost, greater flexibility and significantly higher levels of scalability than ever before. The benefits of cloud computing are primarily built on the notion that cloud environments allow you to achieve these results by virtualizing the different work-loads that you run, and deploying them within an environment that can automatically scale up, or down, as work-load demands change over time.

Many organizations are looking at public cloud services offered by companies like Amazon and Google, but a growing number of organizations are achieving the benefits of public cloud computing by building private clouds.

If you're interested in any of the following, you ought to read this paper;

- Would you like to use a platform that enables you to reduce the running costs of your web server estate by 50%?
- Would you like to use a platform that can enable you to reduce your software license costs by as much as 90%?
- Would you like to reduce the time it takes to deploy new servers to minutes rather than hours or days?
- Do you want the flexibility of cloud computing, but can't move all of your infrastructure and data onto public cloud platforms?
- Can you see a point where you run out of physical space in which to run all of the servers that your organization needs?

This paper describes the way a number of organizations have used IBM's high-end server and software technology to build private cloud environments that have helped them save money, improve agility and deliver better quality of service, while maintaining control of their applications and data.

> "Performance, reliability, disaster recovery, server provisioning and cost efficiency have all seen dramatic improvements—helping BCBSM deliver better service and better value to its members across the state.
>
> Ted Mansk, Director of Infrastructure Engineering and Databases at BCBSM

Key Messages

- **Cloud computing represents the future of our computing infrastructure**

  Cloud computing represents an approach to managing server infrastructure that is a logical evolution of a number of long-standing trends in automation, provisioning and virtualization.

- **Private clouds represent a pragmatic alternative to public cloud for many organizations**

  Very few organizations can place all of their server infrastructure into the public cloud, many will develop private cloud infrastructure in order to achieve many of the benefits of public cloud computing, while mitigating some of its drawbacks.

- **IBM's Enterprise Linux Server (ELS) platform offers a compelling alternative to x86/VMWare infrastructure for private cloud deployment**

  The virtualization and automation capabilities of IBM's ELS technology are based on 40 years of research and development by IBM, and are unmatched in the industry today in terms of reliability and performance, and in many cases offer a significant benefit in terms of cost as well.

- **If you have over 75 separate Linux server machines it is worth looking at consolidating them onto an IBM Enterprise Linux Server**

  A single ELS can support over 300 virtual Linux servers (and can potentially host considerably more depending on workload). At between 50 and 75 servers the ELS is worth examining from a price performance point of view, and for over 100 servers the economics become even more compelling.

Private vs public cloud

Cloud computing represents a very natural evolution of a trend that has been underway since the birth of distributed computing. At its heart cloud computing is based on the idea that computing resources should be managed as a pool, with management tools being used to ensure that the resources in that pool are used in the most effective way possible.

While the headline grabbing stories tend to center on public cloud providers like Amazon, Microsoft and Google, a growing number of organizations are making their first move to cloud computing by creating internal, private, cloud environments, even if some "orthodox" cloud pundits continue to insist that *private cloud isn't really cloud*.

Public cloud

In the public cloud model the entire physical infrastructure is procured, owned and operated by the cloud service provider. They are responsible for the capital expenditure associated with building the infrastructure, with the idea being that they make their margin by capitalising on the economies of scale that they can achieve and the efficiencies they can derive by pooling many organizations' computing requirements.

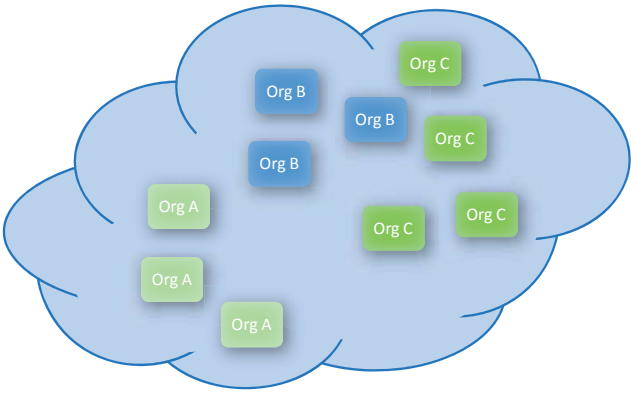


Figure 1. Public cloud

Public clouds are multi-tenanted environments in that the physical resources are shared with many different organizations (or tenants). Different cloud providers offer different levels of isolation between tenants, in some cases it is possible to ensure that your virtual servers will always be physically isolated from other tenants' virtual servers (effectively guaranteeing that your workloads will be the only ones running on a particular server), but this is likely to come at significant additional cost.

Data storage is another issue that needs to be considered. Where will you store your organization's data, on your own infrastructure, or on the cloud, or both? Are there regulatory restrictions that govern how, and where, you store your organization's data? Again, some cloud providers offer to address this challenge by giving clients assurances about where their data is stored, and how it is secured, but in some industries (notably healthcare and financial services) it is still not completely clear to what extent these solutions meet the relevant data storage standards and regulations.

Public cloud infrastructure is extremely well suited to workloads that are subject to very large fluctuations in workload demand. One good example is the UK-based television company Channel 4. Channel 4 uses Amazon's EC2 infrastructure to host websites for a number of its key shows. Channel 4's experience is that traffic to a show's website peaks while the particular show is on air, and then trails off between shows. The peak demand can be as much as 300 times typical levels. Channel 4 has managed to achieve cost savings of over 90% for these particular applications by making use of the ability to rapidly scale up and then down according to fluctuations in demand.

But, in practice, relatively few organizations have a requirement to be able to scale by a factor of 300 for short periods of time, and in Channel 4's case, the vast majority of the company's infrastructure remains in-house.

Organizations that do have specific applications that need very high levels of elasticity can also take a hybrid approach, creating an internal cloud for the bulk of their computing requirements and making use of public cloud infrastructure for those few applications that require massive levels of scalability.

Private cloud

The private delivery model for cloud involves an organization building its own internal cloud. The organization, or the organization's partner in the context of an outsourced arrangement, buys the hardware, installs and operates it. The organization puts in place the virtualization, provisioning and management software and carries all of the costs associated with procurement, datacenter, power, cooling and management.

While this approach does place some limits on the extent to which you can scale up in a short period of time, organizations can still obtain many of the benefits of public cloud computing within a private environment.

There are still significant potential cost savings to be achieved by implementing a private cloud, for example, and it is possible to achieve a very high level of virtualization as a result of developing a private cloud infrastructure, if you choose the right platform on which to base your private cloud.

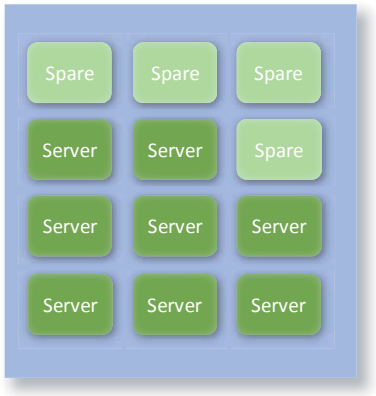


Figure 2. Private cloud

Case Study - Blue Cross and Blue Shield of Minnesota (BCBSM)

Blue Cross and Blue Shield of Minnesota (BCBSM) is the largest health plan provider in the state, with more than 2 million clients. After conducting a detailed evaluation, the organization decided to consolidate 140 Intel-based servers onto a single IBM zEnterprise server with the ELS Linux processors, called Integrated Facility for Linux, or IFL's. The key outcomes of the project were;

- 30-50% projected cost savings over five years compared with a distributed Intel-based environment
- 99% reduction in the time it takes to provision new servers
- 97% reduction in the time it takes to perform a full disaster recovery

In common with many organizations, BCBSM had come to the conclusion that its existing Microsoft Windows and Intel processor-based landscape was becoming increasingly inflexible and costly to operate and maintain.

"*Since Microsoft releases patches for Windows about once a month, we needed to invest a sizeable amount of time to keep the operating systems current. This caused downtime for the business as well. We decided to investigate some other options and see if we could find a cost-effective solution that would avoid these issues*." Ted Mansk, Director of Infrastructure Engineering and Databases at BCBSM.

BCBSM was already running one zEnterprise server to host its DB2 database system, but had never considered consolidating its distributed work-loads onto the platform. After an extensive evaluation process, which involved talking to other organizations to learn from their experiences, BCBSM decided to activate just six new Integrated Facility for Linux (IFL) processors to its existing System z server.

"*Performance, reliability, disaster recovery, server provisioning and cost efficiency have all seen dramatic improvements—helping BCBSM deliver better service and better value to its members across the state.*" Ted Mansk, Director of Infrastructure Engineering and Databases at BCBSM

Having taken the decision, the project was given an extremely short timescale. BCBSM had two months in which to complete the consolidation of 140 physical servers onto a single machine. In partnership with IBM the consolidation was completed within the timescale, ensuring that key projects were able to proceed according to plan.

Cont.

For BCBSM the benefits are not limited to cost savings. The increased flexibility of the platform, and the greater overall reliability of the solution, while harder to quantify in terms of money, makes a significant difference to the business.

"*We've received some really positive feedback from the business. Users really appreciate the fact that when they need something, we can now get to work on it at once, instead of having to wait weeks for new hardware to arrive. Equally, because the System z platform is so much more reliable than our previous infrastructure, we no longer get any complaints about performance and availability issues*." Ted Mansk, Director of Infrastructure Engineering and Databases at BCBSM

Organizations can still benefit from the standardisation of server images, and can still take advantage of some elasticity within the constraints of their total investment in capacity. It is also possible to make use of *capacity on demand*, in which additional capacity is shipped with the platform that customers can turn on, and off, when necessary.

It's also worth remembering, in the context of the differences between Public and Private cloud, that massive levels of elasticity may not be a priority for your organization when it comes to the availability of computing resources.

Private clouds also have some distinct advantages over the public cloud. The most often cited benefit is security. While it is certainly technically possible to use public cloud infrastructure in a highly secure fashion, it is still the case that when you use public cloud infrastructure you are handing your applications and data over to a third-party, and that it will be deployed to an environment that is, by definition, accessible via the internet. Whatever the technical solution, this still represents an additional level of exposure to risk that some organizations are unwilling, or unable (sometimes for regulatory reasons), to accept.

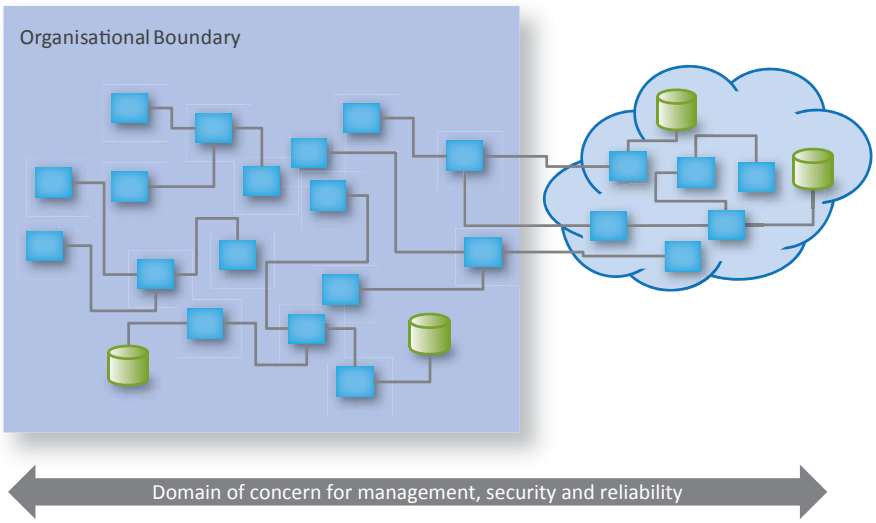


Figure 3. Public cloud extends your management, security and reliability concerns beyond the traditional borders of your organization

Case Study - Yi Lian Zhong Information Technology (YLZ), China

Yi Lian Zhong Information Technology (YLZ) is a leading provider of information services in China focusing on the delivery of social services to citizens.

YLZ is expanding its existing services available in Fujian Province to seven more Chinese provinces, using IBM's zEnterprise hardware as the foundation for a citizen livelihood service information network that will serve a population of 300 million. YLZ selected System z for its reliability, security and cost effectiveness. Specifically, the mainframe's ability to scale easily will help YLZ manage the unpredictable nature of transaction peaks and valleys without incurring additional costs.

The new cloud platform will connect a number of networks of urban and rural residents, government departments, social service providers, medical institutions, corporate and public organizations, and vocational schools. The project is aimed to help meet the government's 12th five-year-plan focused on developing programs to improve the citizen's overall well-being.

Kiosks will be installed across selected communities at civic centers, hospitals, banks, and shopping malls. Participants will be able to swipe their social insurance identity cards, which are already connected to banks, to access the additional services offered by various government departments.

Cloud and the enterprise

There are undoubtedly cases where public cloud is well suited to enterprise computing, particularly where the demand for elasticity is very significant. IBM and others already offer cloud-based development environments for example, which allow development teams to quickly ramp up (or down) as their needs change.

On the other hand Public cloud is less suited to complex applications where some of the processing or data can't be run in the cloud. Imagine you have an ERP system running on your own infrastructure, and you want to deploy a web application that is available to your customers. It is likely that you won't be able to simply deploy your ERP application in the cloud, so that will have to remain on your infrastructure.

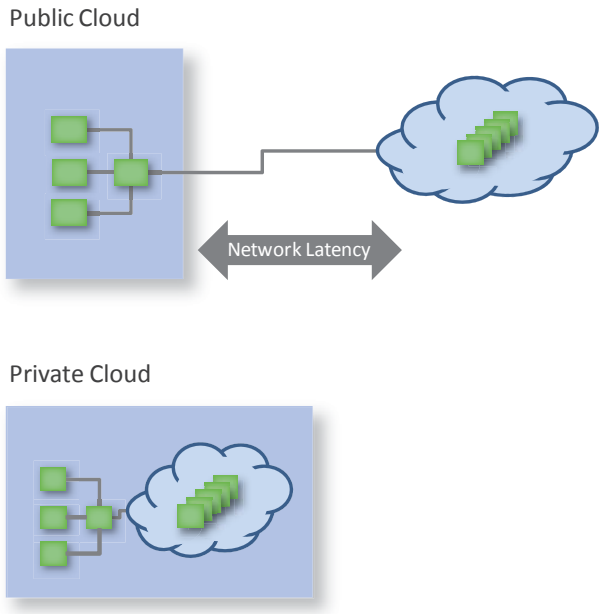


Figure 4. Network latency across the internet can harm application performance

If you deploy the web components of the solution in a public cloud environment, you will have to manage the link between the cloud-based application and your ERP system. While it is possible to do this, and you can do this securely by creating a VPN that reaches beyond your infrastructure into the cloud environment, there is a real likelihood that you'll encounter significant performance headaches as transactions have to cross from the cloud environment, over the network, and into your infrastructure.

If you are able derive the same benefits of scalability, ease of management, and automation by deploying cloud technology on your own infrastructure, you can avoid the problems that network latency creates – or at least you'll be in a much better position to manage them, as the whole application will be running on your network.

Building your own private cloud

Many pundits will say that the only way to build a private cloud is to deploy a virtualization technology, like VMware, and install a number of x86 blades that will provide the processing horsepower for your cloud.

But there is an alternative to the conventional belief that scale-out is the way to deliver cloud computing, it is based on a technology platform that has benefitted from five decades of research and development and which continues to evolve as a platform that supports multiple operating systems, and provides the most reliable and secure computing platform in the world; it is based on IBM's ELS technology.

Introducing the Enterprise Linux Server– cloud in a box

At its foundation, the Enterprise Linux Server (ELS) is an IBM zEnterprise server that ships with the hardware and software necessary to create, and manage, a private cloud of Linux servers supporting industry standard Linux distributions (SuSE and RedHat). The system takes advantage of the server's built-in support for virtualization, combined with IBM's Tivoli Service Automation Manager, which provides the management and provisioning framework for the machine.

zEnterprise : a strong foundation for ELS

The zEnterprise system enables clients to combine IBM's multiple processor architectures into one integrated pool of virtual resources, all managed under a single management framework. The zEnterprise system is comprised of three key elements: 1) the core server that comes in two flavours (z196 with 80 cores and z114 with 10 cores), 2) The zBX which provides integration with IBM blade and accelerator technology, and 3) the Unified Resource Manager (zManager) which is firmware that integrates and manages the system as a unified set of virtual resources. The release of the core z196 represents a technology milestone for IBM, and the industry as a whole.

The platform boasts the fastest CPU in the world, running at 5.2GHz, and introduces some completely new capabilities, most notably the introduction of RAIM (Redundant Array of Independent Memory) which offers the same support for reliability for RAM as RAID does for disk storage.

The most significant feature of this release, however, is the integration the platform offers between mainframe and non-mainframe hardware. The zEnterprise can be attached to a BladeCenter Extension or zBX, which makes it possible to couple up to 112 Power and System x blades to the mainframe and manage all of the elements as a single integrated set of virtual resources.

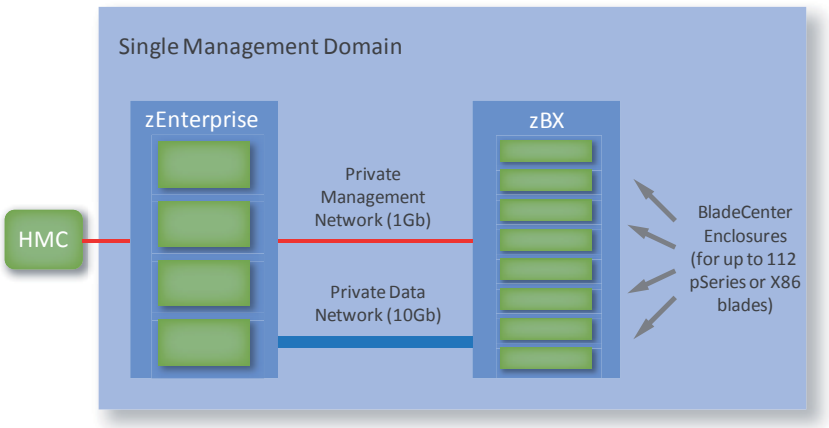


Figure 5. zEnterprise and zBX

The zBX is connected to the zEnterprise machine via two private network connections: a 10Gb ethernet connection is as a private data network that connects all of the blades with the z196 or z114 processors, and a second management network (running at 1Gb/s) is used by the machine's management software to control the different hardware devices within the system.

The zBX can also host IBM's Datapower and DB2 Analytics Accelerator (IDAA) appliances. This support for appliances is an important step and IBM plans to support further appliances in the future.

Appliances are designed to deliver specific additional functionality in order to help improve application performance. For example, the IBM DB2 Analytics Accelerator enables the intelligent and automated routing of SQL queries that are compute intensive to an in-memory datamart that can deliver improved query response time by orders of magnitude (ranging from 10x – 1000x in improved response time depending on the query).

The zBX, and its contained devices are built with levels of hardware redundancy comparable to the zEnterprise itself, so key components like power supplies, power distribution units, network switches and fans are all designed to support mission critical applications that require a common set of business policies and SLA's across heterogeneous architectures.

The zBX provides two key benefits:

First, it provides a very high performance link between the zEnterprise applications, services, and the blade applications that use them, by effectively giving the blades a private 10Gb/s network connection directly into the server, latency and data transfer rates are dramatically reduced. The use of this private network along with its corresponding management network allows for a dramatic reduction in the use of the external firewalls, and hence drives further simplification of the environment.

Second, it provides a single point of monitoring and management. It is not unusual for systems administrators to have to run several monitoring tools at once in order to oversee the end-to-end functioning of a multi-tiered web application. The provision of a single integrated management framework will make it much easier to monitor, and administer, large or complex applications and is likely to reduce the number of configuration/management errors that occur. This alone is significant, as we estimate that over 65% of all application outages are the result of administration/configuration errors rather than hardware failures.

This integration of multiple architectures and appliances is achieved and managed by the zEnterprise Unified Resource Manager (zManager), which oversees the provisioning and management of this federated set of computing resources under a common set of business and governance policies.

Enterprise Linux Server : a Linux cloud in a box

The Enterprise Linux Server (ELS) is made up of an IBM zEnterprise server that ships with the hardware and software necessary to create and manage a private cloud of Linux servers. The system takes advantage of the server's built-in support for virtualization, combined with IBM's Tivoli Service Automation Manager, which provides the management and provisioning framework for the machine.

The key features of the ELS are that it is built on mature technology, and makes extensive use of IBM's software portfolio (Tivoli, Rational, Systems Director) to provide all of the core infrastructure that you need in order to deploy an internal cloud – all within a single hardware footprint.

The system allows users to define their own standard images, based on either RedHat or SuSe linux, that can dramatically reduce the time it takes to deploy new servers. This approach can have a significant impact on hardware and software costs (IBM clients have seen savings of up to 90% in software costs as a result of consolidating onto the ELS) as well as reducing the cost and time associated with systems management. Another by-product of using standardised images is the reduction in faults arising from configuration errors (which in some environments can account for the vast majority of application failures).

The key advantages of using ELS as a cloud platform for consolidating Linux workloads are;

Much faster provisioning than either physical or virtualized environments

When compared to physical environments the comparison is one of minutes compared to hours, days, or even weeks.

Much faster inter-server communication

Network communication between servers running on the platform takes place over an internal virtual network, operating at bus speed, rather than the speed of the network. This has a major impact on application performance, especially where you have a number of application servers accessing a common database.

Consolidated management

A single management interface allows you to manage both native and virtualized workloads, with the zBX add-on to the zEnterprise server, this management interface also relates to the different blades (be they Linux, Unix, Windows or Appliances) running within the zBX.

Mainframe –class reliability/availability features

In a distributed environment reliability and availability are achieved by setting up redundant servers, network and storage infrastructure. On both zEnterprise and the ELS these reliability features are built into the machine itself. The platform doubles up on key infrastructure components (such as network hardware) and makes use of built-in reliability features like RAIM to provide an extremely high level of fault tolerance.

Much higher densities of virtual machines per processor core

This has a very significant impact on software costs, particularly where software is licensed per processor core) and is the driving factor behind the savings in license fees (of up to 90% in the context of Oracle consolidation) that a number of organizations have achieved.

Best x86 practice currently limits the number of virtual machines per core to around 3 and recommends that server utilization should never exceed 50%, so if you were consolidating 100 Linux servers you'd still be looking at purchasing 10 relatively high-end blades or server machines. These factors limit the effectiveness of x86-based virtualization in comparison with zEnterprise or ELS servers where the target utilization is typically well over 70% and the number of virtual machines regularly exceeds 30 per processor core.

The right private cloud infrastructure will deliver the scalability and resilience you need, and enable you to save money

A number of organizations have chosen IBM's enterprise server technology to form the basis of their own private cloud. One example is Transzap. Transzap is a SaaS provider that offers a software suite called Oildex that provides Internet-based ePayable, digital data, and expenditure analysis solutions in the energy industry. Today, it serves more than 6,000 companies/clients and 170,000 registered users. Clients include all of the major oil companies, hundreds of independent oil and gas producers, the world's largest banks, and tens of thousands of royalty and working interest owners.

As a service provider, Transzap has to be able to offer the highest levels of availability to its customers, and be confident that it can scale its operations smoothly as the company grows. The company had begun to encounter reliability issues with its existing x86 distributed infrastructure. After two significant disruptions in service, Transzap took the decision to look for an infrastructure solution that would enable them to support current, and future, demand at levels of availability that its customers expect.

> "At Tranzap we don't have a very large staff, it's clear that this platform will help us to avoid hiring people that we would have otherwise had to hire as our business grows.
>
> We're going to be able to deliver on the guarantees we make to our customers, in fact were going to be able to improve on them and get a competitive edge on our competition.
>
> Peter Flanagan, President, Tranzap

One of the options that the company looked at was IBM's zEnterprise platform, Transzap worked with IBM to conduct a formal total cost of ownership study and concluded that a migration to System z would deliver high levels of reliability and availability while also saving the organization money, notably on hardware costs and Oracle license fees.

Case Study – Atos

Atos is a major international IT services provider, employing 50,000 professionals in 40 countries. The company offers integrated design, build and operate solutions to large multi-national clients in targeted industry sectors, and has annual revenues of around £5.8 billion. An important part of Atos' UK business is managed services – from basic server hosting through to full lifecycle management for both hardware and applications.

"*We have been using IFL processors for several years now and, in our opinion, the platform has proved itself as an ideal enterprise Linux platform.*" Colin Clews, Technologies Manager, Atos

Atos recently developed a new Linux-based application that would help rail transport companies deliver travel information to staff and passengers via their mobile phones, and wanted to deliver it as a shared service for several major rail companies. The company also wanted to enhance its service to customers in the insurance industry by upgrading from MQSeries Integrator (MQSI) middleware to a new IBM WebSphere Message Broker solution, which would provide support for Web Services integration for a number of core insurance applications.

"*The IFL engines in the new ELS not only run Linux environments approximately twice as fast as the previous generation – they also offer around 40% more capacity, enabling us to expand our Linux footprint without increasing our costs. With the ELS, we can run hundreds of environments within a single physical footprint, and easily deliver the 24x7 availability that our customers demand.*"

"*With a traditional distributed approach, even using the latest virtualization-enabled x86 processors, you end up with a lot of unreliable boxes to manage.*" Colin Clews, Technologies Manager, Atos

Cont.

Running the mobile application under Linux on the ELS proved so successful that Atos also decided to deploy the WebSphere Message Broker solution in a similar environment. The company has invested in a second IFL to handle this new workload.

To provide disaster recovery for their environment, the production ELS machine is mirrored to a smaller server at a secondary data center. The backup server features IBM Capacity Backup Upgrade (CBU), which enables its capacity to be dramatically increased if and when it is required to take over the production workload from the production server. With CBU, this spare capacity attracts no maintenance or software licensing costs until it is actually activated.

Linux on ELS is also helping Atos develop an on-demand computing model for its customers, enabling them to pay for computing power on a transactional basis.

"*We can provision new Linux servers very rapidly, and measure capacity usage very precisely. If customers need more power, we can just turn it on almost instantly, and bill them accordingly*." Colin Clews, Technologies Manager, Atos

Conclusion

The world of computing infrastructure is changing and, in the course of that evolution, many of the truths that we have taken for granted for some time are set to be challenged. The perception that the future belongs to x86 blades with VMWare and horizontal scaling is one of those widely held beliefs that will face a significant challenge, from above with technologies like IBM zEnterprise and the Enterprise Linux server and from below with micro technologies like ARM-based processors.

When it comes to a scale-up alternative, a choice has emerged and a number of organizations have elected to take that choice. While the creation of a private cloud based on IBM's high-end server technology may not be right for you, you should at least be watching this space, and also plan to stay informed, as the economics of computing go through another shift that will place the impact of the emergence of client-server computing in the 1990's into the shade.

About this Paper

This paper was sponsored by IBM, however the views contained within it are those of Bathwick Group. We value our independence, and it is our policy never to write papers for clients unless we are permitted to retain full editorial control.

All of the case studies referenced in this paper have been verified both by IBM and Bathwick Group, and where necessary we have interviewed the clients concerned.

If you have any questions regarding the content of this paper, please feel free to contact its author, Gary Barnett by emailing him at gary@bathwick.com.

About Bathwick Group

Bathwick Group has three divisions: research and consulting, powerful survey and marketing software, and a strategic think tank.

Bathwick: Research and consulting

Research and consulting in four core practice areas:

- ***IT industry*** Customer analysis and unique marketing programmes for vendors targeting mid-market and enterprise markets; strategic planning and go-to-market execution support
- ***Enterprise IT*** Strategic support for enterprise IT leaders; productivity and infrastructure agility benchmarking and best practice
- ***Sustainability*** Benchmarking and planning support for organisations wishing to embed sustainable practices and mitigate strategic risks. Future threat and opportunity identification
- ***Innovation*** Rapid collaborative engagement services combining your key people with external experts to expedite solutions to major corporate challenges. IP protection frameworks

STORRM Software

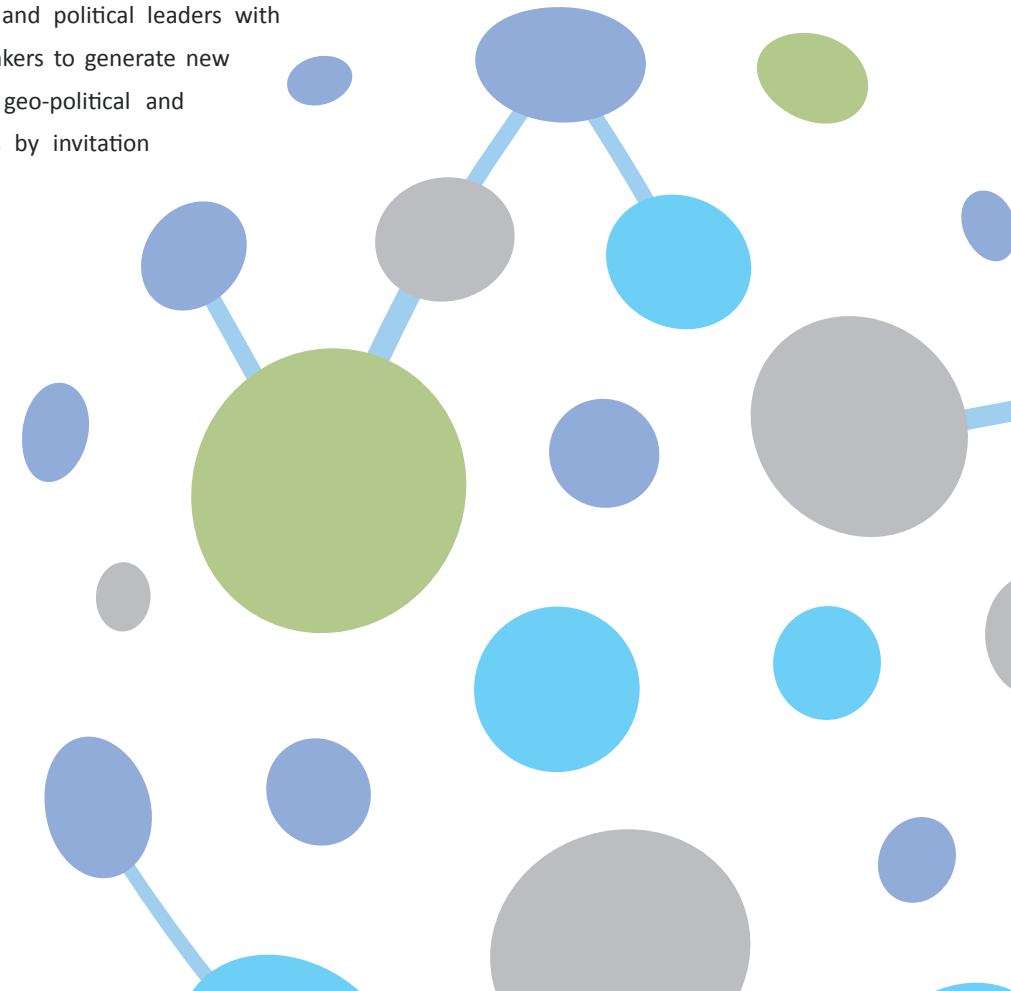
An enterprise-wide data collation, analysis and distribution platform

A flexible multi-language software platform for surveys, benchmarks and assessments, market intelligence, lead generation and nurturing, sales enablement, marketing collateral distribution and control, offline surveys, games and multiple intelligence applications.

ThinkAgain

A global collaborative research network

ThinkAgain brings together business and political leaders with investors, academics, and leading thinkers to generate new insights into business performance, geo-political and sustainability issues. Membership is by invitation only.



Document number: BG-INF-W-00083UK-EN-01