Best Practice in Designing and Operating Infrastructure for Microsoft .NET

Date published 24.08.01

Microsoft®

Attenda

Contents

Foreword and Executive Summary

Managing sophisticated multi-tier Internet infrastructure is hard. The technologies are relatively new and the pace of change is staggering. There are limited sources of information on how to do it well. This paper has been developed from three years' first hand experience of delivering outsourced operations for business critical Internet infrastructures on the Microsoft® platform. Attenda is Europe's largest provider of outsourced Internet operations and we are learning from our experiences every day.

The paper is intended to share some of the knowledge that we have built up. It is designed for a technical audience with an interest in the platform and service architectures required to create an enterprise quality operating environment for Internet applications based on a Microsoft platform. Much of it is as valid for an in house environment as for one that is outsourced. A common theme throughout the report is the importance of designing in scalability, availability, manageability and security from the ground up. Detailed examples are given of how to do this in practice.

The paper covers the technical platform required including a route map for designing an optimised network. It then describes best practice for project management, enabling accurate deployment of applications into a fully supported operating environment. Finally, it describes the service management processes needed to support the applications to ITIL and Microsoft Operations Framework (MOF) standards. This section provides a particularly detailed view of best practice in monitoring the enterprise Microsoft operating environment.

The learning contained in this paper has been hard earned, but it has been a technically exciting process. And it continues. For example we have recently delivered successful deployments of sophisticated, multi-tier applications based on Microsoft BizTalk™ Server 2000, Microsoft Application Center 2000 Server and Microsoft Commerce Server 2000. We welcome any feedback and the opportunity to meet and exchange knowledge with people and organisations facing similar challenges.

1 Microsoft In the Internet Operations Infrastructure

1.1 Microsoft .NET Platform

As a result of the changes in how businesses and consumers use the Web, the industry is converging on a new computing model that enables a standard way of building applications and processes to connect and exchange information over the Web using the eXtensible Markup Language (XML). Internet-based integration using XML Web services enables applications, machines and business processes to work together in a way never previously possible.

Microsoft has transformed the world of computing twice, first with MS-DOS which powered the first personal computers and then with Windows® which made them accessible to everyone. The next stage in this evolution is the introduction of the .NET XML Web services platform for building, deploying, operating and integrating XML Web services. Microsoft .NET utilises industry standard technology to bring the greatest amount of speed, interoperability, scalability and accessibility to making the next generation Internet a part of everyday operations. Microsoft .NET products and services include developer tools and technologies, a scalable, reliable server infrastructure, client support for a broad user experience and a set of core XML Web services.

Unlike some revolutions, this one won't demand a totally new look and feel for the computing environment although the user experience will improve dramatically. Users will still be able to use the tried and trusted environments of Windows and Office to access published XML Web services. Indeed the services will be accessible from a wide variety of new platforms including mobile phones, PDA's and now embedded computing devices as well as the PC.

The business opportunities around providing these services are far more appealing than the hype that surrounded the first wave of web companies. Providing free services and going for volume is not a sound business model as we have seen. With the .NET platform, providers of XML Web services will be able to charge for them as they will provide a real tangible benefit to the client.

1.2 Infrastructure for XML Web Services

XML Web services should be built on a next-generation infrastructure that offers the benefits of modular architecture, economical and linear scaling, security, reliability, manageability and high availability. Microsoft has invested heavily in developing the .NET platform to address these demands.

1.2.1 Windows 2000 Servers

The Windows 2000 family provides a secure, scalable foundation for running the .NET Enterprise Servers and the next generation of business applications. Windows 2000 Server includes built-in Internet services for developing, deploying and managing Web applications that are scalable, reliable and secure. These include Internet Information Services (IIS), component services, data access services, message queuing services, indexing services, security services and support for XML.

1.2.2 .NET Enterprise Servers

The .NET Enterprise Servers and the Windows 2000 Server family make up the Microsoft .NET platform for deploying, managing and orchestrating XML Web services. The .NET Enterprise Servers are: -

Application Centre 2000 to deploy and manage highly available and scalable Web applications;

1 Microsoft In the Internet Operations Infrastructure

BizTalk™ Server 2000 to build XML-based business processes across applications and organisations;
Commerce Server 2000 to quickly build scalable e-commerce solutions;
Content Management Server 2001 to manage content for dynamic e-business Web sites;
Exchange 2000 Server to enable messaging and collaboration, anytime, anywhere;
Host Integration Server 2000 to bridge data and applications on legacy systems;
Internet Security and Acceleration Server 2000 to provide secure, fast Internet connectivity;
Mobile Information 2001 Server to enable application support by mobile devices;
Sharepoint™ Portal Server 2001 to find, share and publish business information;
SQL Server™ 2000 to store, retrieve and analyse structured XML data.

For more information visit http://www.microsoft.com/uk/servers/

1.3 Managing the Infrastructure for Best Service Delivery

In order to achieve the true mission critical availability needed in an Internet Operations Infrastructure, this software needs to be managed and supported in the same way that traditional large scale systems have been. This requires addressing the software, hardware and management framework as a holistic whole the aim of which is to provide a true enterprise management environment.

The team that looks after these systems needs in depth knowledge of all the elements required to deploy and manage highly available systems. This requires a structured, proactive approach to system management and an integrated framework of people, processes and technologies to ensure that these services are taken on, provided and maintained at the correct level. This level of specialisation is difficult to achieve and needs an ongoing commitment to maintain a best practice environment. The benefits of developing this approach are the ability to run highly available and scalable environments. The Microsoft Operations Framework (MOF) provides comprehensive technical guidance for achieving mission critical system performance and addresses the issues pertaining to the people, processes and technologies to effectively manage systems within today's complex, distributed, heterogeneous IT environment.

2 Best Practice Technical Platform

2.1 Goals for Design and Operation of an Infrastructure for .NET

Web based solutions operate in a competitive and volatile business environment. Volumes of traffic and transactions are unpredictable and the customer expects information and services to be quickly and easily available for use. On the Web the competition is only ever one click away. Lower cost of entry allows new players to easily enter a market and the cost of down time can be measured in lost business relationships not just lost transactions. As the market drives towards orchestration of XML Web services to deliver a complete business solution, dependency on the availability of these services increases considerably. The systems that deliver Web based solutions must be able to meet peaks in demand without degrading performance, provide constant availability of the solution to the customer and be able to respond quickly to change. The objective of an Internet operational infrastructure is to provide a secure and robust framework that supports the needs for growth and change, optimises the ability of the customer to compete and minimises the risk of downtime from physical component failures, malicious activity or simple human error.

The design and operation of this infrastructure require a combination of technical elements to work together harmoniously to provide a complete solution. In order for this to happen, the overall design parameters of scalability, availability and manageability have to be addressed throughout the design and operation of the infrastructure. This section focuses on the technical aspects of designing and operating the environment. Section six then examines the service management procedures needed to support an enterprise class infrastructure for the .NET platform.

2.2 Scalability

A system is said to be scalable when it has the ability to meet increasing requests from users and still maintain an acceptable performance level. An Internet application is often distributed over a number of tiers. This provides the ability to improve scalability by splitting the load over multiple front-end and back-end servers. This multi-tiered approach allows a greater degree of scalability.

Growing demand in a typical Internet data centre can come from an increase in the number of simultaneous sessions, duration of sessions, number of transactions, complexity of transactions, volume of traffic, or data size. These demands have different impacts on each component of the system. For example, increasing the number of transactions results in more CPU utilisation at the Web tier, greater network load on the network component, and an increased number of queries at the data tier.

To scale a solution effectively, it is necessary to identify the nature of the increasing demand and its impact on the various components. It is then possible to identify the component that becomes a bottleneck and choose either a scale-up or a scale-out strategy to meet these demands.

2.2.1 Scaling-Up

Scaling-up is a strategy that increases the capacity of a component to handle load. For example, adding a more powerful CPU to scale a Web server's throughput, or replacing a smaller disk with a larger one to increase storage capacity.

2 Best Practice Technical Platform / cont

2.2.2 Scaling-Out

Scaling-out is the approach by which the number of similar components is increased, thereby increasing the overall capacity of the system. Adding more servers can scale the Web tier. Network bandwidth can be scaled by partitioning different types of traffic to different virtual local area networks (VLANs.) Spreading storage across a number of small disk devices can scale data access.

2.3 Availability

Availability in an Internet environment is often, and erroneously, defined in terms of availability of the Internet to the application. The correct definition must be that the full functionality of the application is available to the end user.

The availability of a solution to an end user depends on the continuous running of the supporting platform, including hardware, operating system, applications, storage, networks and security systems. The complete infrastructure must be robust in its assembly and integrated in a way that does not allow a single failure to halt delivery of the solution. The ability physically to build a high availability platform by designing in redundancy to the entire physical infrastructure utilising dustering technology and automated fail over procedures is well proven. However, the distributed nature of today's modern application environment does call for much more rigid and proactive management if the goal of continuous availability is to be achieved.
In supporting enterprise level systems such as major Internet sites it is crucial that the supporting environment is managed to the same level as traditional mainframe type systems. Availability is improved by having rigorous system management and change control facilities to maintain the integrity of the system, automated and proactive monitoring of system and network behaviour to prevent problems occurring and a rigorously applied security policy to ensure that all operations are correctly authorised.

2.4 Security

Security is a critical component of operating a modern operational infrastructure. A compromised solution can result in loss of image, brand value and customer confidence resulting in a total failure of the original business objectives.

To facilitate a fully secure infrastructure, security needs to be an organisational "ethos" where everyone thinks, breathes and lives security in everything they do. This applies to the temporary secretary who works in the organisation for a day through to the technician who, if they made one mistake, could compromise the entire organisation. But what is security? What are the elements that enable an organisation to become secure?
Security can be broken down into eight major risk areas:-

Hardware failure
Software failure
Software copyright
Data Protection Act and confidentiality
Hacking, sabotage and espionage
Viruses
Theft
Business continuity / Disaster recovery

2 Best Practice Technical Platform / cont

Policies, procedures and standards must be available and complied with across all the areas above to ensure every risk area is minimised. Users need to be authenticated to ensure they are who they say they are. There must be authorisation procedures to ensure only those who should have access do have access and only to the permitted areas. Access must be secure so the level of encryption needs to be ascertained to provide the correct protection from unauthorised access. Regular repeatable risk assessments must be carried out to ensure the correct controls are in place to minimise the present and future risks and possible impact. It only takes one risk area to be missed and a system can become compromised.

So, a comprehensive security policy must be in place to create the infrastructure to remove or defend known vulnerabilities, make it as hard as possible to exploit new vulnerabilities, and limit the damage that can be caused by any security compromise. But the key security controls must always be commensurate with the business needs to ensure buy-in and support from customers and senior management.

BS7799 is the most widely recognised security standard and provides a powerful framework for all the organisation's security requirements ensuring everyone gets it right, first time - every time – every where.

Manageability has a profound effect on the security of a system. Lack of procedures can result in non-standard solutions that are difficult to manage. Similarly, overly complicated procedures will result in technicians "Doing their own thing" as the procedures are too difficult to follow.

Centralised management and policy-based security allow more efficient control over the security environment and improved accuracy of security settings while reducing the work required to establish and maintain configurations.

2.5 Manageability

Management and operations broadly refer to the infrastructure, technologies, and processes needed to maintain the health of an Internet operations environment and its services. Manageable Internet Operations Infrastructures allow the use of centralised, automated procedures to minimise the risk of operator errors and ensure the correct implementation of policies. They support both reactive (e.g. monitoring of all relevant information in the infrastructure for fault identification) and pro-active actions (e.g. using that information in the analysis of trends so that, for example, potential capacity problems can be identified and avoided before they occur, the causes of recurrent problems are identified, and opportunities for optimisation noted and acted upon). This manageable infrastructure needs to be accompanied by sound management policies and procedures, including tight change control procedures to minimise the chance of unexpected or unwanted results and a well designed authorisation policy e.g. restricting any operator's permissions to only those needed to accomplish the tasks they are responsible for.

There is a great deal of overlap between the elements described above and the functions required to carry out the service and delivery management functions. Consequently these elements are described more fully in the sections covering the service and delivery management functions.

2 Best Practice Technical Platform / cont

2.6 Designing a Microsoft .NET based Infrastructure

2.6.1 Physical environment

Setting up a secure data centre requires a large capital outlay to provide the level and quality of infrastructure needed. Many enterprises' own data centres, as opposed to commercial data centres, will not gain the economies of scale to warrant such an outlay.

Power failure, fire, flood, and theft are major risks in an enterprise's own Data Centre. It is critical to delivering a high availability solution that the physical environment of the Internet data centre is fully resilient. This will include having redundant air conditioning systems and redundant dry fire suppression capability. Consistent power supply must be ensured through redundant UPS backed up by multiple standby power generators. These generators should have at least 24 hours fuel and come with re-fuelling contracts. Best practice includes installing connections to two power grids taken down to individual rack level. Multiple layers of physical security are a must. For more detail on physical security please see the item in section 4.5.2.1.5.5.

2.6.2 Network

This section outlines the design for a network to support a best in class infrastructure for .NET. A large amount of effort, time and expertise is typically required in the design phase to ensure that the network will provide a scalable, available, secure and manageable solution.

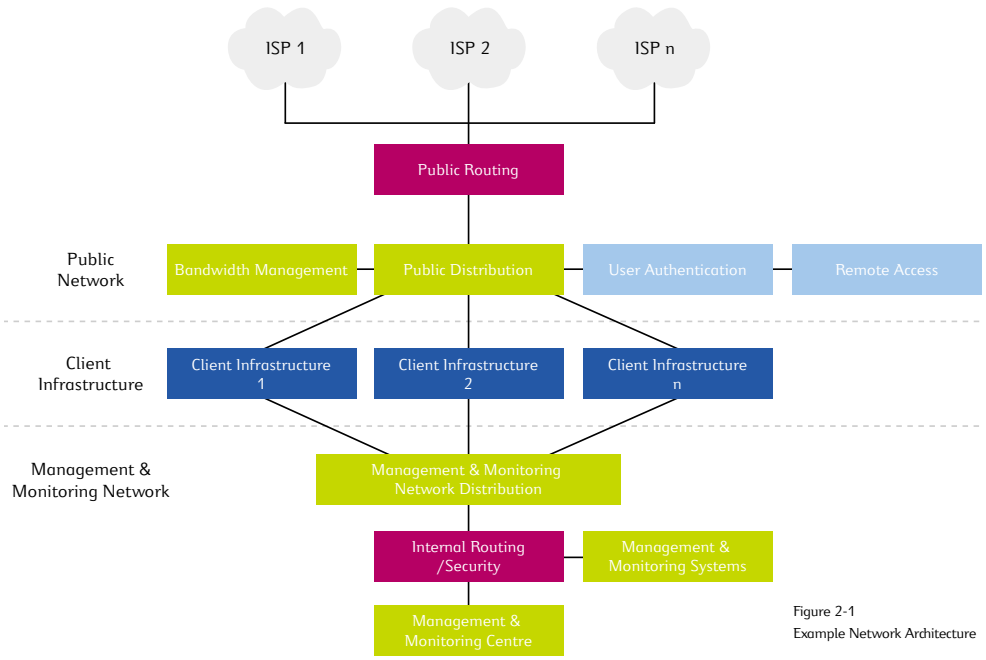

Figure 2-1
Example Network Architecture

2 Best Practice Technical Platform / cont

2.6.2.1 Public Routing and ISPs

An Internet data centre must have connections to at least two Tier 1 networks for redundancy. Having two connections to the same network is not an acceptable alternative.
Connecting to the Internet via multiple ISPs has the following benefits: -

- Provides a resilient connection to the Internet if one ISP fails or degrades
- Peering with multiple ISPs gives improved performance as the number of hops to any given destination will be minimised through best path routing.

In order to achieve connection to multiple ISPs the Internet infrastructure must have its own Autonomous System (AS) number. This will allow use of the Border Gateway protocol to optimise the traffic with the ISPs.

2.6.2.2 Public Distribution

The Public Distribution layer distributes incoming traffic to individual customer cells and aggregates outbound traffic on its way to the Internet. Bandwidth management and Remote Access functions are connected at the Public Distribution layer.

The connectivity within the Internet infrastructure needs to be fault tolerant with no single point of failure to ensure that very high levels of network reliability can be maintained throughout the network from the customer's servers through to the redundant connections on to the Internet.

This resilience should be both at the architectural level (i.e. n routers, n switches) but also at the component level (i.e. two power supplies within the router, redundant management cards within the router, etc).

2.6.2.3 Bandwidth Management

In order to provide customers with a measurable and definable level of service it is important that the facility exists to manage the bandwidth available to the customer application. Without this facility, another application or customer could use up all the bandwidth available leaving a properly functioning application without adequate connectivity. This management should take the following form: -

- The ability to specify a guaranteed bandwidth limit on a per customer system basis
- Enabling capped burst limits above this guaranteed level, allowing an application to exceed its bandwidth limit provided spare bandwidth is available without impacting other systems
- The ability to report on application usage of guaranteed and burstable bandwidth for capacity planning purposes

This type of bandwidth shaping approach also has security benefits, limiting the impact of distributed denial of service attacks to one system.

2.6.2.4 User Authentication and Remote Access

Centralised analogue and ISDN dial-in access to customer equipment will allow new customer cells to be created without deploying bespoke hardware for access each time. This speeds up and simplifies deployment, making it a more consistent process. In addition to this type of remote access, customers will often require secure access to their systems using a Virtual Private Network (VPN). All dial-in customer sessions on their equipment should be authorised and their actions audited.

2 Best Practice Technical Platform / cont

2.6.2.5 Customer Cell

A customer cell is the term for the section of the network infrastructure that contains unique customer systems. Customers' requirements can vary from single servers through to multi-tier deployments spanning multiple server racks and multiple application layers.

The design criteria in setting up the Internet infrastructure should be flexible enough to cater for these requirements, thus allowing customers to deploy the optimum solution whilst maintaining the overall design consistency of the infrastructure. Each customer cell needs secure connection to both the public network for external access and to the Management and Monitoring network.

2.6.2.6 Management and Monitoring

The Management and Monitoring network is completely segmented from the public network. This is essential for the following reasons: -

- A wide range of functions and hence types of traffic need to be performed within a network (for instance http, ftp, media streaming, monitoring, management and backup) but only a small part of this should be publidy visible
- Segmenting this traffic into two separate networks improves security, performance and scalability.
- The public network allows only a few, strictly defined types of traffic (such as http). Hence it is highly secure and fast because it is not contending with other internal traffic

Deep management requires high levels of administrative access. This should not be possible in the part of the network that faces the public Internet. So, without a separate management network, either security or manageability is compromised. The public network is used only for customers' traffic, maximising performance.

The management network should be both logically and physically distinct from the rest of the network to maintain its integrity. This management network should also be configured to be highly available, as it is the means by which the applications are monitored.

2.6.2.7 Internal Routing and Security

A separate layer of Internal Routing and Security functionality sits behind the Management and Monitoring Distribution Layer and provides access to Management and Monitoring systems housed on the network. They, of course, need to be designed and operated so as to achieve high availability and security.

2.6.2.8 Management and Monitoring Centre

This is where the centralised monitoring and management functions are carried out for the infrastructure. It is manned and operational on a 24 x 365 basis and provides a complete suite of monitoring functionality.

With many servers in a typical data centre, thousands of events will be raised each day, many of which do not require action. A centralised management tool such as NetIQ AppManager or Microsoft Operations Manager is the best way of managing these events, ensuring the critical events are detected and acted upon.

2 Best Practice Technical Platform / cont

Centralised management tools also enable collection of statistics in a consistent manner, allowing trends of performance and capacity to be analysed and potential capacity limits to be dealt with before they affect the customer systems.

Normally, these need to be augmented with specialist tools for managing particular devices (e.g. firewalls, switches, etc). Where possible, the use of a single vendor for network devices will allow a common approach to management as many centralised network device management tools are vendor specific.

All of the above facilities are carried out in the Management and Monitoring Centre and more details of the specific activities can be found in section 4.

2.6.3 Physical Devices

2.6.3.1 Servers

Enterprise class infrastructure requires servers that have been designed for high availability with advanced manageability features. Utilities such as Compaq SmartStart and Compaq Insight Manager add significant value and greatly aid in achieving these goals. Lack of these features will remove from consideration most non-brand equipment or servers priced and positioned at the desktop or under desktop market.

To further improve availability, single points of failure need to be identified at both a hardware and application level. Best practice design would assume that anything that can go wrong will do so and that any single point of failure will fail first. Thus not only do physical architectures need to be redundant, but all components within a server should also be. This necessitates features such as: -

- Hardware RAID (providing high protection against disk failures)
- "Hot swappable" disk drives, power supply units, network cards and fans, allowing part replacement without interrupting service

To further improve performance requires integral hardware disk controllers, multiple dual port network cards, dual processors and high-speed buses.

To effectively deploy the servers within a data centre environment, manageability is critical and getting the basics right is vital. The servers must be rack mounted to ensure proper cooling and efficient cable management and a documented racking process is a sensible measure.

Compaq servers are particularly strong for running Microsoft solutions as they provide internal monitoring agents (Compaq Insight Manager) that integrate effectively with NetIQ AppManager, the monitoring toolset recommended for the Microsoft platform. Indeed Compaq offers pre failure warranty based upon alerts from these agents

2.6.3.2 Firewalls

To reduce single points of failure, best practice would again dictate that firewalls are deployed in a fail over configuration, enabling downtime due to firewall failure to be reduced to nearly zero in a well designed and configured deployment.

2 Best Practice Technical Platform / cont

In a multi-tiered architecture, an example being Compaq's Distributed Internet Server Array (DISA), firewalls might also be deployed at multiple levels within the architecture – not only in front of the web servers but also between the application and database layers. This will improve security but requires extensive experience in designing suitable monitoring and management solutions.

Firewalls generate significant amounts of data in a busy website and large log files need to be stored, interpreted and alarms raised when suspicious activity occurs. Frequent firewall IOS upgrades and patches need to be rapidly deployed across potentially large numbers of devices to react to security vulnerabilities. Challenges such as these mean that careful selection and implementation of centralised management tools will greatly reduce the management overhead and aid in accurate management of change.

2.6.4 Windows

Microsoft Windows 2000 is clearly preferable as an enterprise platform over Windows NT Server 4.0.

It offers significantly improved manageability via the Microsoft Management Console (MMC) giving administrators a common user interface presentation tool and the Windows Management Instrumentation (WMI) provides updated access and event services. This allows better control and monitoring of Windows based environments. Windows 2000 also exhibits better scalability than Windows NT Server 4.0, improved performance and better support for the latest hardware.

There are a number of variants of Windows 2000. Windows 2000 Advanced Server can be used to take advantage of Network Load Balancing services - a significant improvement over Windows NT load balancing. Clustering is a must for high availability so it is advisable to deploy Advanced Server for all back end systems. Windows 2000 Server is a lower cost alternative for hardware load balanced Web servers. Windows 2000 Datacenter Server is designed for enterprises that demand the highest levels of availability and scalability. It is the right operating system for running mission critical databases, enterprise resource planning software, and high volume, real-time transaction processing.

Clear best practice is to develop and deploy a scripted automated build and configuration procedure for Windows 2000. This delivers a number of benefits including accuracy, consistency and improved reliability. Moreover, it can reduce training and recruitment costs as lower skilled staff can be employed.

A formal process should be in place to upgrade these automated builds on a regular basis. New versions incorporate service packs and required hot fixes that have been fully tested.

Microsoft Windows 2000 is a multi-purpose operating system. When used in an enterprise Internet environment effective security requires that Windows be "hardened" by removing superfluous components of the operating system - such as IIS for back end servers and by changing default settings. As simple examples, best practice will be to choose a different user name for "administrator" accounts and disable default last username display on "ctrl-alt-delete." This "hardening" is a highly complex procedure that is impossible to carry out cost effectively without automated configurations. While the exact nature of hardening varies by application to some extent, best practice will typically involve more than 300 configuration changes and the removal of up to 50 superfluous components.

2 Best Practice Technical Platform / cont

2.6.5 System Software

System software is defined here to mean applications, in addition to Windows, that are well understood, monitored and managed with a high level of support and available expertise. The Microsoft .NET Enterprise Server products fall into this category.

The Microsoft .NET platform includes a comprehensive family of products that provide for each aspect of developing, managing, using, and experiencing XML Web services. The feature set and functionality of the products is both broad and deep and to effectively support the applications a Service Description must exist for each of the .NET Enterprise Servers. This Service Description must specify precisely and in detail all tasks to be undertaken and roles and responsibilities.

Experience of deploying the .NET products, with their many configuration files and settings, reduces deployment times but more importantly directly affects the long-term availability and manageability of the platform. Although standardised builds assist in this process, the deployment team requires extensive experience and training to ensure successful secure builds.

2.6.6 Applications

In addition to the system software in most deployments additional applications, both packaged and custom developed, are required. As there are a vast number of commercial and custom developed applications, it is not feasible to offer the same level of support to each of them.

Hence these applications will not have their own unique Service Descriptions but a few basic principles will ensure that they can be monitored, managed and supported to a defined level

- The provider of the infrastructure should work with the application owner to ensure that relevant and appropriate monitoring is in place. Experienced monitoring personnel add significant value and ensure that error states are quickly and accurately captured using both custom and packaged tools such as NetIQ.

- Specific management tasks that are required by software supported by third parties should be identified and documented within an Operations Manual. These should be codified into procedures that support personnel can accurately and repeatably follow

- When problems do occur, clear escalation paths should be agreed and documented so that personnel with expert application knowledge are available. This might mean that support is escalated back to the application vendor

2.6.7 Content/ Data

Due to the wide range of possible content stored on a site it is important to fully understand and document how it will be managed, monitored and supported.

During deployment, experienced staff need to set up content specific monitors. Support and escalation needs to be clearly documented within the Operations Manual.

2 Best Practice Technical Platform / cont

Although in a well designed site content will rarely affect the sites availability or security, it may have a large effect on its manageability. How content is uploaded, the frequency with which this occurs and how change control is managed can incur a significant support overhead. Use of content replication tools can significantly reduce this workload. Similarly the functionality within the .NET server products such as Commerce Server 2000, Application Center 2000 and Content Management Server 2001 greatly aids content and application distribution.

2.6.7.1 Backup and Restore

The overall availability of a client's solution depends on both the supporting platform (hardware, operating system, network, security systems, etc) and also the availability of the content, data and applications. This might be static content such as HTML or dynamic data from SQL Server 2000 databases or .NET applications.

When data has been lost or corrupted the reliable, rapid restoration of content, data and business-critical applications is crucial to solution availability.

Because of the importance of frequent backups, the effect on normal operations must be considered, so as not to have a serious impact on availability and performance. As the sizes of databases grow and content continues to change, it is critical that backups can be performed without disrupting work and after a catastrophic failure it must be possible to restore both data and databases in the minimum amount of time.

Understanding of the business risks and the value of lost data thus plays an important part in understanding the requirements of a backup solution. In general, there are three backup techniques of interest.

Offline (cold) backup

The database and applications are shut down cleanly and taken offline. Separate backup software then copies the files to the backup devices. When the copy completes, the applications may be brought online. The data is unavailable from the time that the applications begin shutting down until it is brought back online.

Online (hot) backup

The database management system is running and the database is online. However, the database itself is not being accessed, and is therefore unavailable for use by applications during backup.

Active Online Backup

The database and applications are online and being used actively. The backup runs during normal transaction processing. No backup window is required.

Active online backup during normal operations is the preferred solution and in the 24x7 environment of the Internet often the only viable solution.

3 Best Practice Project Management

3.1 Managing Effective Implementation of the Internet infrastructure as a Project

The adoption of a formalised approach to managing projects is well established in many industries including IT. The demanding nature of the new e-business paradigm makes strong project management even more important for successfully delivering projects. In such a turbulent environment having a clear, efficient and flexible approach to Internet project delivery is a characteristic that differentiates successful projects from the others.

Many organisations have an ideal of how they would like their Internet infrastructure to be deployed but this is often not delivered. Effective project management provides the framework - essential processes, skills, tools and techniques - to bridge the gap between ideal and delivery.

Judicious application of project management principles will underpin the realisation of the following key business requirements:

- Establish clear objectives, deliverables and measures of success
- Manage risk
- Manage change
- Enable customer focus and alignment
- Optimise the use of organisational resources
- Control costs
- Incorporate quality principles
- Manage communication

3.1.1 Best Practice Project Management

The foundation of the best practice observations that follow is the use of a project management method. The professional discipline that flows from using a method allied with the common understanding throughout the organisation of the way projects will be executed is invaluable in delivering a successful Internet infrastructure.

There are a number of methods to choose from. However, one of the most widely accepted and well established methods is PRINCE2. **PRINCE**®, which stands for **Pr**ojects **in C**ontrolled **E**nvironments, is a project management method covering the organisation, management and control of projects. PRINCE was first developed by the **Central Computer and Telecommunications Agency (CCTA)**, now part of the **Office of Government Commerce (OGC)**, in 1989 as a UK government standard for IT project management.

Since its introduction, PRINCE has become widely used in both the public and private sectors and is now the UK's de facto standard for project management. Although PRINCE was originally developed for the needs of IT projects, the method has also been used on many non-IT projects. The latest version of the method, PRINCE2, is designed to incorporate the requirements of existing users and to enhance the method towards a generic, best practice approach for the management of all types of projects.

3 Best Practice Project Management / cont

3.2 Examples of Best Practice Project Management

3.2.1 Enable customer focus and alignment

One of the key requirements for successful project implementation is to have an appropriate Project Sponsor. This should be a senior manager who is the key stakeholder in the business objectives supported by the delivery of the project. Numerous studies have shown that having the right Project Sponsor backing is one of the key requirements for delivering projects to time and budget. Another factor is having a Project Manager who has experience of implementing this type of solution. Obtaining the appropriate seniority of people for these two roles will greatly improve the likelihood of success.

In some organisations it is common for project ownership to be vested in the Project Manager. This approach does not lend itself to tight alignment with business imperatives. The Project Manager is clearly key to the successful delivery of any project but should be viewed as the facilitator. Initial and ongoing ownership of a project should rest with the Project Sponsor. It is necessary to create an environment that ensures the Project Manager is suitably empowered to manage the vast majority of project issues on a day-to-day basis but takes direction from the Project Sponsor on matters of a fundamental nature.

3.2.2 Establish clear objectives, deliverables and measures of success

At the outset of each project it is imperative that a clear understanding is arrived at for the key objectives of the project and main deliverables or products that will be produced. This understanding should be formally documented and signed off by at least the Project Sponsor and the Project Manager. The resulting project initiation documentation should also include, amongst other things, agreed measures of success that are objective and concise plus relevant timescales, budgets and clearly stated role descriptions for project resources.

3.2.3 Manage risk

One of the first tasks for the Project Manager after the initiation documentation has been completed is to pull together the Project Team and other affected parties to identify the risks pertinent to the successful delivery of all objectives. Common sense needs to be applied to the depth of the risk assessment, which should be commensurate with the size of the project. The identified risks should be prioritised in terms of probability and impact. Effort should then be applied to managing and mitigating the high probability, high impact events.

Having invested time in developing a risk management approach it is vital that the subject is reviewed regularly and refreshed to ensure that changes within the project or external factors are reflected in the current risk assessment.

3.2.4 Manage change

The management of change is an objective of most companies' IT organisations, with the emphasis on assessment of the impact of change on the operational infrastructure. It is equally important to have a defined approach for controlling change within the project environment. Changes to project deliverables or the environment into which they are to be introduced should be formally assessed and the likely impact determined. An informed decision can then be taken by the Project Sponsor or Project Manager on whether to proceed with the change.

3 Best Practice Project Management / cont

3.2.5 Optimise the use of organisational resources

It is vital that well thought through project plans are developed (typically in a Gantt chart format) and task resource estimates (in people-days) produced for prioritisation discussions with senior management. It is inefficient for people to be moved about between projects regularly but organisational priorities can change swiftly and the information on which to base resource re-assignment should be maintained.

3.2.6 Control costs

The control of cost is a priority for any business. Project costs can be a significant contributor to an organisation's overall spend and it is highly important to have processes in place to track both hardware and software procurement costs and people days devoted to each project. The capture and assessment of this data is a key feed into the ongoing review of business case viability. It should not be considered a failure to stop a project when the business case is no longer valid; this should be viewed as sound business practice enabled by effective project management.

3.2.7 Incorporate quality principles

Quality is subjective and the important project discipline is to remove this subjectivity by determining what is to be produced and to what standards. Quality can be defined by delivering against the agreed specification. The project method must embrace a formal approach to deciding which project deliverables require tight quality management and how that delivery against specification is to be controlled.

3.2.8 Manage communication

Good communication both within project teams and between projects and the rest of the business is vital. Each project should develop a communication strategy that identifies all stakeholders, the type of information that needs to be conveyed, the optimum medium for the communication and the required frequency of communication.

Regular project team meetings and updates with the Project Sponsor would normally figure within the project communication strategy. These sessions ensure all relevant parties hold a common view of the project progress and the issues experienced.

4 Best Practice Service Management

4.1 IT Service Management

IT Service Management is concerned with delivering and supporting IT services that are appropriate to the business requirements of the customer. The principles of IT Service Management have been in use for many years in traditional environments and have been proven to improve the quality of service delivered to the end user.

The principles of Service Management are not based on technology but on management functions that are required for the successful operational management of any IT system.

The essence of successful IT Operations is a disciplined, controlled approach based on processes and procedures that ensure reliable, repeatable and consistent service delivery.

Service Management is highly relevant to the provision of Internet operations services and Attenda has identified more than 100 processes that are required to support Service Management functions in an Internet operations environment. This requires significant investment to develop, implement and maintain the processes.

Implementation of proven Service Management functions will result in IT services that have:

- High levels of service availability and reliability delivered through well controlled operational environments
- Effective incident reporting, resolution and management processes
- Effective plans for managing future capacity needs and business continuity
- Regular reporting of achievement against agreed Service Level measures

4.2 ITIL - A Best Practice Framework for Service Management

The **IT I**nfrastructure **L**ibrary **(ITIL)** was originally developed in the late 1980s by the Central Computer and Telecommunications Agency (CCTA) and has become a global standard in Service Management. It began as a guide for UK government, but the framework has proved to be useful to organisations in all sectors through its adoption by many Service Management companies as the basis for consultancy, education and software tools support.

ITIL provides a comprehensive and consistent set of best practice processes for IT Service Management, which promotes a quality approach to achieving business effectiveness and efficiency. ITIL processes are intended to be implemented so that they underpin the existing business processes of an organisation.

ITIL provides a framework in which to place existing methods and activities in a structured context. ITIL does not dictate every action that needs to be taken on a day-to-day basis; instead it focuses on best practice that can be utilised in different ways according to need. By emphasising the relationships between the processes, any possible lack of communication and co-operation between various entities should be minimised.

4.3 The Microsoft Operations Framework and ITIL

The Microsoft Operations Framework (MOF) combines the collaborative industry standards outlined by ITIL with specific guidelines for using Microsoft products and technologies. MOF also extends the ITIL code of practice to support distributed IT environments and industry trends such as application hosting and Web-based transactional and e-commerce systems.

4 Best Practice Service Management / cont

MOF is a collection of best practice processes and models. The framework supports solutions and services built with Microsoft technologies by providing comprehensive technical guidance. It focuses on achieving mission-critical system reliability, availability, supportability, and manageability.

MOF has been designed with the following principles in mind:

- Use ideas that have been proven in action and leverage industry best practice
- Incorporate input from customers, partners, and Microsoft product and service organisations
- Integrate with frameworks that manage other parts of the IT lifecycle, such as planning and deployment
- Support managing end-to-end services, including processes and procedures, rather than just managing servers and technology

MOF is divided into four quadrants as illustrated below.

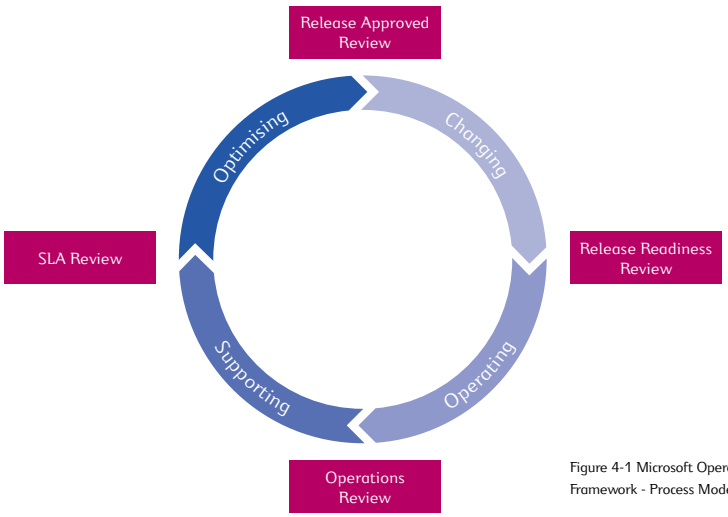

Figure 4-1 Microsoft Operations Framework - Process Model

Each quadrant is based upon the fundamental ITIL processes, although in the Operating quadrant MOF provides operations guidance specific to Microsoft products and technologies. The following sections outline each of the phases in the model, giving an overview of the Service Management functions within the phase and their supporting processes.

4.4 Changing the Environment

The changing quadrant of the Microsoft Operations Framework comprises three key Service Management Functions:

- Change Management
- Configuration Management
- Release Management

4.4.1 What is Change Management?

Rapid and ongoing change is a fundamental feature of operating an Internet infrastructure. This constant change is necessary to deliver new business functionality and has a direct impact on customer facing operations. Rigorous change control procedures are essential to maintain site integrity as the smallest change can affect stability and operation.

4 Best Practice Service Management / cont

Change Management is the application of management processes to ensure that all changes to the environment supporting the Internet infrastructure are effective, undertaken in a controlled manner and completed with the least possible impact on site availability.

Change Management covers all assets that exist within the environment and are necessary for meeting the service level requirements of the solution. Examples of such assets are hardware, communications equipment and software, system software, applications software, processes, procedures, roles, responsibilities and documentation.

4.4.1.1 Best Practice Change Management

Best practice Change Management ensures that changes are rigorously tested and introduced to the live environment without disruption. A best practice approach has the following features.

Changes are initiated via a Request For Change process that describes the change, describes the components affected and provides a risk assessment. Changes are implemented following change request authorisation by a Change Advisory Board, a group that evaluates change requests for business need, priority and risk with relation to the system that is being changed and its potential impact on other systems.

For each type of change there is an agreed list of approvers. There should be a standard list of approvers for each type of change e.g. security changes will need to be agreed by a different specialist from storage changes. Management team approvers should be consistent for all changes.

Conflicting, high risk and high impact changes must be identified. Conflicts for use of the same people or IT resources must be identified and resolved either by prioritisation, rescheduling or provision of additional resource. High risk changes due to the nature of the change, the difficulty in providing an effective regression plan or any other factor must be identified and mitigated. Similarly, changes with a high impact in the event of failure should be thoroughly evaluated before implementation.

All changes are planned including the provision of back out plans so that failing changes can be regressed without causing disruption.

To ensure effective communication between those making the change and those affected by the change, all relevant parties must collaborate in authorising the implementation of changes. It is important to ensure that all parties have bought in to the change process, since a single change implemented outside the process, without appropriate evaluation and back out procedures could cause significant downtime. Encouraging the customer and other third parties to follow the same Change Management process ensures that there are no weaknesses or unexpected interruptions in service delivery.

Most customers, whilst wanting their Internet sites to be monitored and managed, expect that no changes, beyond agreed regular housekeeping operations, will be made to their sites without their knowledge. To meet this need, processes have to be in place that require full customer agreement before any software updates, configuration changes or hardware replacements are undertaken.

4 Best Practice Service Management / cont

There should be a process for implementing emergency changes and providing documentation retrospectively. Emergencies will result in changes that have not gone through the formal process before they are implemented but documentation must be provided retrospectively to ensure that communication is complete and all records are updated.

The impact of change on security and Capacity Management should always be considered. Changes must conform to security policies and any risk of compromising security must be considered as well as the effect of the change on resource utilisation, shared resources and capacity plans.

Following a successful change, configuration and inventory data should be updated on all systems to reflect the new state.

Achievements should be reported and measured and an audit trail kept, providing a feedback loop to improve the implementation of future changes. The information generated by the Change Management process is used to drive internal and external improvements.

4.4.2 What is Configuration Management?

Configuration Management is the function that records, tracks and reports on key components of the Internet infrastructure. This should include a description of each item, its relationship to other configuration items, version numbers and locations.

Configuration Management is a crucial element of Service Management for complex Internet operations. It provides the process for recording all changes and updates to the critical components of the system. An environment cannot be competently managed unless the provider understands what exists and with the high rates of change required to satisfy customer requirements, knowing what is in the environment at any point in time is more challenging than in traditional IT environments

4.4.2.1 Best Practice Configuration Management

Best practice Configuration Management requires a Configuration Management Database (CMDB) that is a single logical data repository for configuration information. The information within the CMDB should, whenever possible be self-maintaining with automated updates and should be an up-to-date record of the physical and logical relationships between service components. The content of the CMDB is readily accessible to everyone who is authorised to use it. Regular audits of the CMDB should be carried out to verify the accuracy of configuration data and initiate any necessary action to correct discrepancies.

Configuration Management should be integrated with Change Management and configuration data should be used to determine the risk and impact of changes. Accurate information is essential, as incorrect configuration data will result in failure to execute changes correctly. Following successful change configuration items should be updated.

A well-documented process should control the introduction of new configuration items, updates to them during their lifetime and their eventual disposal. Clear standards should define the detailed configuration information that is required. Processes should be in place to control the use of software licences. Regular audits are performed to ensure that processes have been followed and to identify any required corrective action.

4 Best Practice Service Management / cont

4.4.3 What is Release Management?

Release Management is the process of co-ordinating and managing the planning, testing and implementation of all releases in to the live environment. Release Management ensures that releases are implemented in the live environment as quickly as possible to meet business requirements, yet in an extremely controlled and systematic process that limits impacts to the existing environment.

Release Management is especially important to Internet infrastructure and operations due to the high volume of changes, short turn around times expected for changes, and the high impact and visibility of errors.
A formal Release Management process ensures quality releases from the development teams to the production site. Such a process provides a managed environment for the development and testing teams for the release of new features or fixing known issues with the existing infrastructure. Release Management also provides a timely and accurate mechanism for releasing new content, therefore creating a better experience for the customer with minimal downtime.

Release Management should be applied to both the application environment and the hardware and system software environment.

4.4.3.1 Best Practice Release Management

Best practice Release Management successfully controls the release into the live environment, whilst maintaining system integrity and availability.

A central location, typically called the Definitive Software Library, holds master copies of all software for each release. This creates a known, single source in which definitive, authorised software configurations are stored.

The Release Management function should co-ordinate comprehensive user acceptance testing and pilot staging that includes verification of rollout and back-out procedures. Release Management should be integrated with Change Management and Configuration Management for the reasons outlined in the previous sections e.g. recording the updated level of software and licence in the configuration database. Release versions define product version numbers, service packs, patches and configuration settings for hardware and software. There should be a process for evaluating and approving new versions of existing products and additions to the approved product list.

Release Management should be responsible for creating and implementing release rollout plans for use in new solutions and updates to existing solutions. A well-documented process should be used for building a release to ensure consistency and accuracy. The process for applying a release should be automated. This ensures consistency, accuracy and assists in the capture of best practice. The Release Management function should notify the Problem Management process of errors that occur during the application of a new release.

4 Best Practice Service Management / cont

Example: Changing the Environment in Practice

Attenda has developed a Change Management application based on Microsoft Exchange forms that is used to provide Request For Change information to change authorisers in preparation for the daily meetings of the Change Advisory Board.

Attenda has developed processes to ensure that new and updated versions of hardware and software products that are used within the Internet infrastructure are evaluated, tested and implemented in a controlled manner. This process includes an initial paper based evaluation, as well as full lab testing. Attenda has developed a list of standard evaluation criteria to assess each and every product. This ensures that all new hardware and software products are properly evaluated and consistent quality is achieved.

Attenda has implemented access control procedures to ensure that developers cannot access the Web environment to make changes to, for example, new releases of applications or content. If access is required a request with a specific objective and start and end time must be made. Once this is agreed access is enabled. Following completion of the agreed change access is removed again.

4 Best Practice Service Management / cont

4.5 Operating the Environment

4.5.1 What is Service Monitoring and Control?

Service Monitoring and Control provides a centralised environment from which to monitor and operate the Internet infrastructure. The service provider should be in a position to take immediate corrective action to minimise the impact on users of any deviations from the planned service delivery according to customer requirements.

4.5.1.1 Best Practice Service Monitoring and Control

Best practice Service Monitoring should:

- Identify and analyse areas of weakness and other trends, so that remedial action can be taken to improve future service quality
- Illustrate where either customer or user actions are causing faults and identify where working efficiency and/or training may be improved
- Focus on the key business requirements
- Provide reporting against Service Level Agreements

4.5.1.2 Monitoring Internet Applications

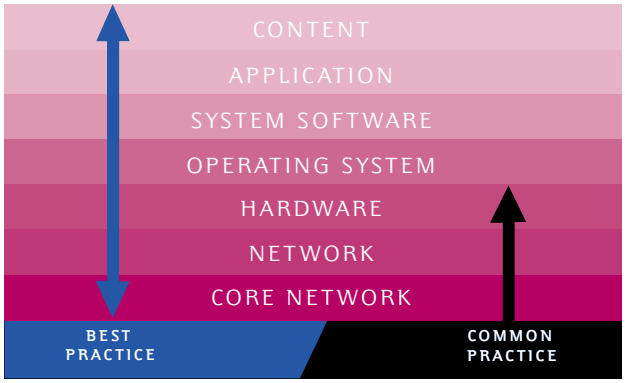

Figure 4-2 Extent of monitoring required for Best Practice

The diagram above illustrates the layers within the Internet infrastructure. The monitoring function should focus on all levels of this model. Failure to monitor the many levels of technology can lead to deterioration of the end-user experience. However, excessive monitoring may lead to performance degradation of the network, the servers or the application. Monitoring of Internet applications should be much more than just monitoring network connectivity and server availability.

User experience monitoring is concerned with the performance and availability of the user interface. This is monitored by automatically testing the connectivity to key pages within the Web site. This kind of testing can be performed either remotely, or close to the Web application.

Testing the system remotely gives a reasonable representation of the remote users experience but is also subject to the vagaries of the Internet. Whilst this kind of data can provide useful statistics on performance, particularly if it is carried out from many geographical and logically diverse locations, it does not provide useful diagnostic or application-specific performance information.

4 Best Practice Service Management / cont

Placing the monitoring systems closer to the monitored application gives a more accurate measure of the performance of the application itself. With bandwidth becoming almost a commodity item, the performance bottleneck of most business-critical applications is in the performance of the application and server platform itself.

There are many tools available for testing the response from a Web page. Most of them operate by targeting a provided URL and checking that a valid response is returned. With the majority of Internet applications being built on dynamic content, however, it is vital to also provide mechanisms for checking the content of the returned pages. Web servers such as IIS may return perfectly valid html with nothing but errors displayed within the page. By checking for specific known content on the pages the probability that the correct content is reaching your customers is increased.

For pages that change infrequently it is also useful to monitor for content changes. If you do not expect the content to change regularly, then a sudden change may indicate that the site has been compromised and unauthorised content has been uploaded.

The user experience is not just about the ability to connect. The content must also be delivered in a consistently timely fashion. Response time monitors should be set to download full content from a page (i.e. including all graphics) and should be set with a reasonable timeout (this will often be fine-tuned after the site has been operated for a period of time). In particular download times should be logged over time and examined to look for trends. For example, if the download times are seen to increase significantly during peak times, this may be indicative of a bandwidth or server performance bottleneck.

It is also important to bear in mind that when designing connectivity and download time monitors, different pages on the site may be generated differently and each kind needs to be represented in the monitor collection. For example, flat html files may be delivered promptly even though dynamic asp pages are failing completely. Similarly a page containing data driven content may fail because of a failed database connection while all other pages on the site appear normal.

4.5.1.3 Load Balanced Sites

Web applications using some form of load balancing at the Web layer present particular difficulties to monitoring. When monitoring a load-balanced site it may not always be possible to tell which specific Web server is providing the content. Wherever possible, the individual Web servers should be monitored directly using one or more monitors per server. Some load-balanced configurations do not allow direct connections to the individual servers and it may become a question of interpretation.

In these cases it is important to understand the type of load balancing in use. A round-robin or load driven balancing mechanism is likely to share the requests from the monitoring systems across Web servers, therefore a failing server appears as an intermittent error, showing up only when the request is directed to the failing server.

More difficult to spot are errors in a system that uses some form of affinity to maintain session state. It is perfectly possible in this scenario that the monitoring server be directed to the same Web server each time, so unless that happens to be the failing server, errors may not be spotted. In these configurations it is vital to find another way of testing the functionality on the individual servers.

4 Best Practice Service Management / cont

4.5.1.4 Transaction Monitoring

The primary purpose of many Web applications is to lead the user into completing an online transaction and some errors may only become apparent when the user reaches this point. To test transactional functionality it is necessary to have a tool that can reproduce a user transaction on the site. Such tools should also have the capability of testing content and response times at each step to provide accurate fault isolation and service availability tests.

It is important to carefully consider both the maintainability and impact of such a monitor. Very complex transaction monitors may break whenever any of the site content or data is changed. This causes a considerable maintenance overhead and reduces the usefulness of the monitor because it causes too much monitoring 'noise'. Also consider carefully whether it is sensible to commit transactions to the database as part of your monitor. It may seem simple to identify and ignore orders or transactions from a test user, but they may impact on others who have to produce usage statistics or control inventory.

It is important to consider other trackable consequences of committing a transaction on the Web interface. As an example placing an order on a site may simultaneously cause an email to be sent and a database entry to be made. Deploying a tool that can test for these actions as well as checking the user's view provides true end-to-end transaction monitoring.

4.5.1.5 Platform Monitoring

Servers in an Internet environment need to be considered as much more than just a collection of hardware. Each server provides vital services to the Web application and these services must be constantly monitored to ensure that the application stays up and running.

When planning monitoring for an application, it is vital to consider the application architecture as well as the physical architecture. The application architecture describes the interaction between the services and the flow of data around the solution.

Once the application architecture has been fully described it is possible to identify which key elements of the application need to be monitored. This provides much greater granularity in monitoring. As an example, a Web interface transactional monitor such as completing and posting a requisition may fail due to an error in any one of several key services. The transactional monitor may not always indicate the exact source of the error but a monitor applied directly to the service would isolate the error immediately.

Having identified the key services it may also be possible to monitor specific metrics associated with the service and consequently proactively identify potential problems before they cause a transactional failure or outage. An example of this might be a Message Queue server. Having identified this as a key service it would be possible to monitor the number of messages waiting in the queue. A growing number of messages may indicate a failure at one end of the queue that can be addressed before any transactions are lost.

4 Best Practice Service Management / cont

4.5.1.6 Server Performance

Today's transactional Internet applications tend to place a heavy load on the servers and it is vital to monitor and collect server performance data in order to provide performance alerts and trending data to help with future capacity planning.

Windows itself provides tools for doing this in the Performance Counter utility, but specialised monitoring tools can provide greater flexibility and control, and more open data storage. Thresholds should be set on the measured metrics to alert when the server reaches an unacceptable load. This may be due to a sudden increase in usage of the site, but it may also indicate an application error or bottleneck. In the latter case additional monitors should be set up to try to isolate the cause of the problem. This is often an iterative process, gradually zeroing in on the problem, so a tool that allows the flexibility to retrieve data in real time and deploy new monitors quickly is vital to this process.

Excessive server load should never be allowed to result from a gradual increase in usage. Monitors should be put in place to save historical data so trend analysis can be carried out. Preferably the trending itself should be an automated process, raising an alert if the average load on the server is consistently increasing. It should then be possible to scale the solution up or out before the performance becomes an issue.

4.5.1.7 Hardware Monitoring

Highlighting a hardware problem before it causes an actual failure allows the repair to be undertaken in a controlled manner whilst keeping attention focused on the health of the remaining hardware.

Compaq Insight Agents provide unparalleled hardware monitoring on the Compaq platform, and provide pre-failure alerting on most aspects of hardware. For example, a failing disk drive should continue to function for at least 72 hours before it fails completely.

Hardware alerts should be integrated into the main monitoring systems; tools that can receive SNMP traps or which use SNMP to monitor the status of devices enable this level of integration.

4.5.1.8 Network Monitoring

The basic functionality of network devices can be simply tested by pinging through them to the system end-point, usually the Web server. However, this does not give a complete picture and does not pick up on potential problems.

As an example, firewalls and switches should always be configured as fail over pairs, so that connectivity is not lost even if a device fails completely. In this scenario, the simple ping test would continue to work even though the risk of an outage has increased significantly due to one device being in a failed state.

It is essential therefore that the status of devices be tested on an individual basis. The most common mechanism for this, supported by the majority of devices, is through SNMP and SNMP Traps. SNMP can be used to poll a device for its status, while the device itself can be configured to send an SNMP Trap under certain conditions, such as a fail over occurring.

It should be noted, however, that SNMP is an inherently insecure protocol and should only be used on secure private networks such as the management network described earlier.

4 Best Practice Service Management / cont

4.5.1.9 Alert Mechanisms

With all the monitoring in place it is vital that appropriate alerting mechanisms are in place to ensure that problems are identified and dealt with. Alerting systems should offer guaranteed alert delivery and automatic escalation.

Guaranteed alert delivery ensures that 100% of alerts are received in real time or are retrievable in the event of a monitoring system failure. No alerts should ever be completely lost or critical failures may be overlooked. Guaranteed delivery requires multiple resilient alerting routes. Agents running on servers or other devices should be capable of functioning autonomously and sending alerts by redundant routes. Saving alerts in a local event log ensures that any incident can be recreated later even if all connectivity to the server is lost.

Automatic escalation ensures that the alerts are always forwarded to the right person or that they are escalated if they are not dealt with in a timely manner. Escalation should provide the capability to take different alerting actions depending on the severity or number of occurrences of a particular alert.

4.5.1.10 Integrating Monitoring Into The Application

A successful monitor and control function will bring together the application developers and solution architects early on in the project to:

- Ensure that the application architecture is understood, so that the monitoring solution is most useful and effective
- Ensure that the Monitoring and Management Centre has a proactive approach, with tools available to 'auto-fix' known problems
- Provide total solution monitoring including all aspects of the infrastructure, particularly the application, where approximately 30% of all Internet infrastructure failures occur
- Feedback requirements to other functional areas such as capacity planning

4.5.2 What is Security Management?

Security Management is responsible for protecting the Internet infrastructure and application from unauthorised access. Connecting systems to a public network such as the Internet is inherently a very high risk. Therefore, security must be the primary concern of all businesses operating a complex Internet infrastructure.

Security is always measured in Confidentiality, Integrity and Availability (CIA). Customers need to know that their Internet infrastructure has the most appropriate and robust security policy in place using the latest in best practice. They expect their personal information and financial transactions to be totally secure and accurate whilst the site remains available at all times.

4 Best Practice Service Management / cont

4.5.2.1 Best Practice Security

4.5.2.1.1 IT Security Management Standard

A widely accepted and recognised standard for best practice in Information Security Management is BS7799. This is organised into ten major sections covering:

- Business continuity planning
- System access control
- System development and maintenance
- Physical and environmental security
- Compliance
- Personnel security
- Security organisation
- Computer and network management
- Asset classification and control
- Security policy

Within each section are detailed statements that make up the standard. BS7799 requires regular risk assessments that become the driving force behind the controls that are introduced to minimise the risks. Obtaining BS7799 certification is a major undertaking and an organisation that has achieved this certification has made a major commitment to providing customers with industry leading Security Management

4.5.2.1.2 Adopt a Multi-Layered Approach to Security

Some infrastructure designs can be likened to an eggshell; well defended on the exterior (often by the use of firewalls) but once that outer defence is breached everything is vulnerable. Best practice in security is to have multiple layers, like an onion, with consideration being given to minimising the impact of a breach in any one layer. Many of the other best practices described in this paper contribute to this end, for example the separation of layers of a customer solution, the minimisation of permissions for an account. Other features should include the isolation of one customer system from another. If one is compromised, this should not affect the security of the others.

Best Practice Examples

- Bandwidth management should be deployed (to limit the effect of a distributed denial of service attack)
- Firewalls must be configured so that only essential traffic at each tier of the infrastructure is permitted
- Servers should be hardened (see section 2.6.4) to minimise the chance of there being vulnerabilities to exploit
- Restrictive traffic policies should be implemented so that a successful attack on one server does not expose other servers to additional risk
- Extensive monitoring must be used to maximise the likelihood of detecting attempted or successful attacks
- Security should be considered within every Request For Change as part of the Change Management process to ensure that as necessary changes are made to the Internet infrastructure security is not compromised

4.5.2.1.3 Specialist Security Management Team

In addition to obtaining BS7799 it is important that an organisation has a supporting team with in depth expertise and a clear focus on security. This requires a dedicated security manager backed by the appropriate specialists.

4 Best Practice Service Management / cont

Specifics responsibilities of the security organisation would include the following:

- Pre-live penetration tests
- Regular production penetration tests
- Pro-actively looking for new vulnerabilities (See below)
- Installing security patches in a controlled manner as soon as possible after their release
- A server hardening process built by security architects
- A separate monitoring and management LAN to the Internet facing customer LAN
- A controlled separate access route for 3rd party access to customers' sites
- A protocol control policy across the tiers to ensure the application developer works within secure boundaries
- Stringent change control processes for firewall change requests

In addition to the above, the overall approach of the security organisation should be to pro-actively work with customers to address security before it becomes an issue and ensure that regular dialogue minimises emerging concerns.

4.5.2.1.4 Taking A Proactive Approach Reduces Vulnerability To Attack

A proactive approach to Security Management will help to minimise the effects of security attacks. The security team should proactively research the latest vulnerabilities on a daily basis, ensuring that the security controls are updated with the latest information. Providing customers with weekly reports of security incidents and a regular newsletter covering all aspects of Internet infrastructure security ensures that customers are kept aware of the latest developments and can adapt their applications as necessary.

Management should ensure that all staff are fully briefed and aware of Internet security issues and actively encourage their customers to audit their security solutions.

4.5.2.1.5 Specific Security Measures

This section describes some of the specific security measures that should be implemented.

4.5.2.1.5.1 Deploy Individual Accounts With Minimal Permissions

Generic "Admin" accounts have far too many privileges and are often shared by different members of the support organisation. It is better to provide individual accounts which in turn provide an audit trail of who has done what. This provision also means that accounts can be restricted to the access needed by an individual. Adopting this approach means that if an account is compromised, the damage inflicted is minimised.

This best practice should be in place for customer access to their systems, as well as accounts for operations purposes.

4.5.2.1.5.2 Centralised Account Management, Authentication and Authorisation

Without central authentication, accounts and passwords are maintained in many places, and the problems with managing them mean effective password update polices tend not to be enforced. Centralised management also eases the introduction of higher levels of authentication as required, such as smart cards or various biometric methods.

4 Best Practice Service Management / cont

All accounts should be authorised through a centralised management system. This includes accounts for administrative access as well as those used by customers when remotely accessing their own systems. This allows passwords and account changes to be implemented once only, and to be instantly enforced.

4.5.2.1.5.3 Privileged Customer Access Should Be Kept Separate From Core Network

Best practice is to have a separate route for customers to access their servers. They will require access that should not be possible over the public Internet, in particular to back-end tiers that should have no access to the public Internet. The implementation of this might include dial-up access, leased lines, or the use of VPNs. Whatever method is used it should be resilient and, if it involves traffic passing over the public Internet, that traffic should be encrypted.

Access should be permitted based on individual accounts and only to the servers strictly required for that customer.

4.5.2.1.5.4 Security Management Process

The security management process should consider both infrastructure and application vulnerabilities.

Security should be a key consideration during the design phase of any solution. Separating the core Internet infrastructure network and the monitoring and management network to ensure maximum security and flexibility is recommended as described in Section 2.6.2.6.

The Internet infrastructure should be protected by at least one firewall and all servers should be hardened, including Web, application and database servers. Prior to going live, penetration testing should be performed to ensure that the loading of the application has not compromised security.

Although it is not possible to completely eliminate the risk from distributed denial of service attacks, use of bandwidth management devices will ensure that one client solution is not affected by denial of service attacks on another client. Multiple connections to different ISPs will also reduce the impact of such attacks.

Continuous risk assessments should be performed to ensure that the Security Management processes remain appropriate for the business requirements.

4.5.2.1.5.5 Physical Security

The Security Management processes become redundant if unauthorised physical access can be gained to the Internet infrastructure. Therefore physical security of the data centre is absolutely fundamental and the following factors should be considered in order to achieve physical security:

- Strict access controls to limit access to authorised personnel only with use of access cards and/or biometric controls
- Onsite 24x7 security personnel
- Constant CCTV surveillance
- Locked server racks

4 Best Practice Service Management / cont

Example: Operating the Environment in Practice

Attenda possesses in depth expertise on Internet solution monitoring tools and has leveraged this to provide a consolidated alerts portal in its Management and Monitoring Centre. This ensures that support staff have a complete, accurate and easy to use view of all alerts at all times.

Attenda employs personnel with a strong background in Microsoft solution engineering and application development to manage monitoring. This gives us the capability to deliver customised monitoring scripts that add significant value to the client solution.

In Attenda's experience it is essential that Internet security is built upon a strong technical architecture but this must be supported by policies, processes and procedures that conform to BS7799 and that are followed with meticulous attention to detail. Attenda has a specialist Security team that ensures that this happens.

4 Best Practice Service Management / cont

4.6 Supporting the Environment

The supporting phase is all about resolving problems in the most efficient manner and keeping the customer fully informed. The following Service Management functions are integral to this phase:

- Service Desk
- Incident Management
- Problem Management

The supporting phase is of fundamental importance to the achievement of Service Level Agreement commitments and as such the operation is continually assessed against service level performance.

4.6.1 What is a Service Desk?

The Service Desk is a collection of people, processes and technology that enables a company to offer a single point of contact for all customer problems and queries. The customer communication is two way with the Service Desk both receiving all customer service related enquiries and then ensuring that progress updates are provided at agreed intervals to agreed parties.

4.6.1.1 Best Practice

To support Service Desk operation, companies must invest time and money in dedicated people, processes and technology. The Service Desk should have the following characteristics: -

Contactable – The Service Desk should be contactable 24 hours a day, 365 days a year. The means of contact must suit the customer and could be via telephone, email or via the Web.
Responsive - the Service Desk must have a published SLA for responding to all customer issues, and define the frequency of updates and escalations.
Knowledgeable - Having got through to the Service Desk it is vital that the person answering the call can understand the problem or query. This first contact is all-important and should be conducted in an efficient, professional and caring manner.
Technically Capable - Not all problems will be straightforward and some will necessitate additional technical support to resolve. Access to second and third line technical specialists must be immediate when the need arises.
Responsible for progress reporting - It is likely that a large proportion of problems will not be resolved during the initial customer contact. It is imperative that agreement is reached on when the customer will be called back for progress updates and that these update calls are made on time every time.
Accountable for escalation - It is likely that from time to time a problem will arise that proves to be very difficult to resolve. When this occurs there should be defined escalation processes to ensure that appropriate levels of management are informed and are able to support the problem resolution effort.
Disciplined - It is vital that defined processes and procedures are followed. True efficiency comes from harnessing the collective experiences and knowledge of the people in the company and distilling these into concise yet flexible support documentation that is easily and reliably accessible to the Service Desk team.
Focused on priority problems - This could be considered another facet of 'Knowledgeable' but it is crucial that the priority of a problem is understood and agreed with the customer from the outset and appropriate levels of response are then delivered in line with the Service Level Agreement. Clearly the urgency of a technical enquiry is somewhat different to that of a major service-affecting incident.

4 Best Practice Service Management / cont

4.6.2 What is Problem Management?

Problem Management is responsible for both reactive and proactive problem handling. From a reactive perspective, Problem Management provides assistance to the Incident Management process in the form of diagnostic expertise and the implementation of workarounds. From a proactive perspective, Problem Management works to identify the underlying causes of problems and has a key focus on problem prevention.

4.6.2.1 Best Practice

A best practice approach to Problem Management embraces the following:

- Operating to a documented Problem Management process that has a clearly identified owner
- Ensuring appropriate levels of technical resource are always available to resolve problems in line with the Service Level Agreement; this is both in terms of number of engineers and depth and breadth of skills.
- Providing an appropriate toolset to rapidly identify and resolve all problems

Another facet of the Problem Management function is to maintain open problem records and tracking until the underlying cause has been identified and resolved. This should include managing multiple incidents with the same underlying cause as a single problem. The Incident Management tool should be integrated into the Problem Management process.

As part of the problem resolution process the Problem Management function should use previous incident data to support efficient problem diagnosis and resolution. This should also include root cause analysis on all incidents where the impact is significant or where multiple incidents exhibit common symptoms.

The root cause analysis findings should be used to identify areas of potential service improvement. The Problem Management function should maintain a known errors database that can be exploited by the Incident Management process. All incident data should be analysed and trend analysis performed to identify consistently occurring problems and action taken to address them.

As part of the ongoing management of the service the Problem Management function should review the Problem Management process and apply any continuous improvement opportunities.

4.6.3 What is Incident Management?

Incident Management is responsible for the efficient handling of service affecting incidents and the restoration of service in a timely manner within the terms of the Service Level Agreement. The incidents should have been reported via the Service Desk acting as the single point of contact for all service affecting problems and queries

4.6.3.1 Best Practice Incident Management

A best practice approach to Incident Management embraces the following: -

- Operating to a documented Incident Management process
- Ensuring that all incidents are logged and managed until they have been closed with the agreement of the customer. A single tool should be used for the logging and management of all incident data
- Providing a clear and simple method allowing customers to log incidents
- Delivering the Incident Management service to meet measures recorded and committed to within the Service Level Agreement

4 Best Practice Service Management / cont

- Logging all events that require further investigation or action on the Incident Management system
- Prioritising all incidents in line with the definition within the Service Level Agreement and progressing service restoration in line with this prioritisation
- Providing an appropriate tool set to rapidly identify and resolve all incidents
- Delivering a capability to match incidents to known causes (with known resolutions)
- Ensuring appropriate levels of technical resource are always available to resolve incidents in line with the Service Level Agreement; this is both number of engineers and level of skill
- Defining escalations procedures and ensuring they are adhered to
- Updating all incident records to ensure the current status is always reflected
- Monitoring all open incidents and taking action to ensure all Service Level Agreement commitments are honoured
- Recording sufficient and accurate incident data to enable management reporting, root cause analysis and trend analysis
- Reporting on Incident Management performance versus Service Level Agreements and taking action to address any areas of concern

Example: Supporting the Environment in Practice.
The Service Management functions in this quadrant of MOF are closely related and all need to use data that is held in the system used to log and progress client requests for support. Attenda uses the Remedy Help Desk system to log client problems and service requests, to escalate these as and when necessary and to provide the data for reporting on call response times, call volumes, trend analysis and root cause analysis.

Attenda has also developed detailed processes and procedures to facilitate the consistent and effective support of its clients' Internet infrastructure. Some of these are generic and are therefore used for all clients but these are supplemented by client specific support instructions that are detailed in the client Operations Manual. Both the generic and client specific processes and procedures are essential. One example of a generic process would be how to action a change to DNS whereas an example of a client specific process could be how to deal with the failure of a communications connection to a third party business partner of the client.

In addition to its investment in systems, processes and technology Attenda has also invested in personal skills training for its support staff so that they are client focused and can interact effectively with clients

4 Best Practice Service Management / cont

4.7 Optimising the Environment

4.7.1 What is Service Level Management?

Service Level Management includes the processes of planning, co-ordinating, drafting, agreeing, monitoring and reporting on Service Level Agreements (SLAs). Service Level Management is responsible for ensuring that achievable levels of service are formally agreed and consistency achieved. Achievable, cost-justified, and measurable service requirements and goals should be negotiated and documented within the SLA. Parties responsible for meeting the SLA will have been involved in its negotiation and agreement. The service goals must be regularly monitored and any required improvement activities undertaken.

4.7.1.1 Best Practice Service Level Management

A best practice approach to Service Level Management requires operating a documented Service Level Management process that defines all services to the customer in Service Level Agreements (SLAs). These may be generic SLAs for a particular offering or customer specific SLAs. The Service Level Management function ensures that all SLAs have been reviewed and agreed to by all supporting parties and Operational Level Agreements (OLAs) established. OLAs are simple agreements with internal support groups setting out specific targets that underpin targets included in the SLA.

As part of the ongoing management of SLAs they should be reviewed periodically, at least once a year.
An SLA should include at least:

- Hours of service availability
- Service availability goals
- Costs, charges and penalties.

It is the responsibility of the Service Level Management group to specify service level targets that can be measured and are achievable. Additionally they should be responsible for measurement and reporting of adherence to SLAs. The Service Level Management function should implement mechanisms for soliciting regular feedback from customers and ensuring action is taken on issues identified.

4.7.2 What is Capacity Management?

Capacity Management is responsible for the cost effective provision of sufficient delivery capacity to satisfy current SLAs and foreseeable business requirements.

Capacity Management ensures optimal use of resources meaning that resources are consumed at the best place, time, quantity, and price.

Capacity Management defines thresholds that have to be monitored so that the performance stays within the defined levels allowing time to take preventative measures (such as tuning and/or acquisition of new resources) when required. Capacity Management also defines actions to be undertaken once the thresholds are exceeded.

Capacity Management is responsible for delivering the performance related information to the business support system.

4 Best Practice Service Management / cont

4.7.2.1 Best Practice Capacity Management

A best practice approach to Capacity Management embraces the following:

- Operating to a documented Capacity Management process, that encompasses the full end-to-end service including environment, hardware, system software, application and database, network and IT management processes
- Identifying and implementing monitoring tools to regularly provide information on performance, workload, throughput and resource utilisation
- Defining achievable performance and capacity goals within the SLA and ensuring that these are regularly monitored and recorded. Regular analysis of the recorded data should be used to drive any improvement activities through the Change Management process
- Establishing workload profiles, or other capacity measures, to be used in conjunction with trend analysis and business plans to proactively manage future capacity requirements. These forecasts should be used to produce a regular capacity plan that incorporates business forecasts. The solution provider should then work dosely with dients to ensure that any such capacity increases are understood and implemented, without detrimentally affecting service delivery

4.7.3 What is Availability Management?

The importance of availability of IT to the success of a business has never been more apparent. The interdependency between the business and the IT operation has developed to a point where quite simply if the IT stops, then the business stops. This is evidenced by the emergence of Internet based online business-to-business and business-to-consumer services. Trends such as the global economy, 24 hour economy, e-commerce and flexible working are now viewed as essential if a business is not only to attract new customers but also retain existing ones.

The single goal of Availability Management is to ensure that the customer can use a given IT service at any time.

Availability Management is responsible for providing the levels of availability required by the business and specified in the SLA. Availability Management ensures that appropriate availability goals are established and that the corresponding cost of downtime is understood. All risks to availability can then be cost effectively addressed, whether arising from the underlying technology, or from people-oriented process and procedure.

4.7.3.1 Best Practice Availability Management

A best practice approach to Availability Management embraces the following:-

- Operating to a well-documented Availability Management process under the management of a single or individual group.
- Negotiating with the customer to define achievable, cost-justified, and measurable availability goals, to be documented within the SLA.
- Ensuring that the configuration of the system, the design of the services, and the operational management are capable of meeting the required availability goals. High availability for a service solution begins with the requirements phase of the project. High availability cannot be achieved without being planned into the technical architecture and systems design

4 Best Practice Service Management / cont

- Identifying logical and physical components that make up the service, designing appropriate countermeasures to address any inherent availability risks and ensuring that availability goals can be met
- Identifying high availability risks such as single points of failure or critical personnel and ensuring that actions to manage these risks are documented. All opportunities to minimise planned downtime need to be identified and plans documented
- Verifying that serviceability requirements within both internal and external SLAs are capable of underpinning the defined availability goals. The performance of external suppliers should be monitored and evaluated; reporting on their performance and holding regular review sessions with them
- Regularly monitoring availability goals and undertaking improvement activities as required
- Integrating Availability Management with Change Management to ensure that the availability impacts of any proposed changes are appropriately considered
- Employing management reporting to review the Service Continuity Management process for efficiency and effectiveness

4.7.4 What is Service Continuity Management?

Service Continuity Management is responsible for ensuring that services continue to be provided in accordance with the SLA in the event of a major failure or disaster. Service Continuity Management works closely with Availability Management to ensure that all risks to service availability are appropriately handled.

4.7.4.1 Best Practice Service Continuity Management

A best practice approach to Service Continuity Management embraces operation to a documented Service Continuity Management process that covers all aspects of service continuity in the event of a loss of any part of the supporting infrastructure including premises, power supply, or computing or telecommunications services.

A designated manager should be responsible for Service Continuity Management. The Service Continuity Plan should identify full details of how to execute the plan, priorities for service restoration, the expected status of the service following restoration, the roles, responsibilities and out-of-hours contact details of all personnel named in the plan and an expected timeline of key decision points and expected recovery events.

The contingency plans should be tested on a regular basis and during this testing it should be verified that all the items needed to restore normal operation are available in accordance with the plan. For example, it should be assumed that no access will be possible to the affected premises and therefore all documentation, data and software must be available immediately elsewhere.

It is important to work with Change Management to ensure contingency plans are kept up to date with technology and business developments. Service Continuity Management should be integrated with the overall business continuity planning process. The Service Continuity Management process should be reviewed regularly for efficiency and effectiveness.

4 Best Practice Service Management / cont

Example: Optimising the Environment in Practice

Attenda has a standard Service Level Agreement that is agreed with clients as part of the project delivery process. Experience has shown that client expectations must be matched against the SLA as early as possible in the lifecycle of the implementation project in order that the implications of any client specific changes can be identified and agreement reached on how to deal with them.

Attenda has identified all of the key components in the shared network and server infrastructure used by clients and identified usage thresholds and lead times for provision of additional capacity. Regular measurement of resource usage has been implemented across the complete shared infrastructure.

Attenda has a well-defined Service Level Agreement for its core service offering. The Attenda network infrastructure has been designed with no single point of failure and client solutions can also be designed to this objective.

Attenda has developed Business Continuity Plans for its key locations and there is an ongoing programme of further development and testing. Attenda is also able to work with clients to identify performance and capacity issues in their solutions and to recommend and implement improvements.

5 About Attenda

Established in 1999, Attenda is Europe's leading provider of outsourced Internet operations. It designs and operates reliable, scalable and secure Internet infrastructures for some of the world's largest companies. The company specialises in supporting Internet applications built on the Microsoft .NET platform.

Attenda's flagship solution, Attenda M.O. is designed to deliver 100% outsourced operations, higher availability and reduced costs for business critical .NET infrastructure. Attenda M.O. includes solution design, deployment, security, ITIL compliant Service Management, monitoring deep into the application layer, 24*365 support, backup, reporting and ongoing account management.

Attenda's clients comprise established large enterprises such as Shell, Debenhams, De La Rue, Jordan Grand Prix team, Datamonitor, Asite and Emetra.

The company operates across Europe from offices in England, Germany, and France, and has substantial financial backing from Phoenix Equity Partners (part of Credit Suisse First Boston), UBS Capital, Texas Pacific Group and Compaq Computer Corporation.

For further information :

Attenda Ltd
One London Road, Staines
Middlesex TW18 4EX
United Kingdom

Tel : +44 (0) 1784 211100
Fax : +44 (0) 1784 211200
E-mail : enquiries@attenda.net
www.attenda.net

6 About the Authors

Paul Morris, Director of Service Architecture, Attenda

Paul joined Attenda in March 2000 from the AA (Automobile Association) where he was General Manager of IT Strategy and Operations.

Stuart Bonell, Director of Platform Architecture, Attenda

Stuart joined Attenda in August 2000 from marchFirst where he was their E-Commerce Practice Leader.

Date published 24.08.01