the Availability Digest

www.availabilitydigest.com

Business Continuity Today

March 2009

Business Continuity Today is a freely-available living eBook covering a broad range of topics relating to business continuity. We say "living" because the book is a perpetual work-in-progress. Published by Vision Solutions, a new chapter is added every few months. Six chapters have been published since February, 2007; and the seventh is due out soon. The book is downloadable from www.businesscontinuitytoday.com.

Each of the chapters is an overview of some pertinent field. Some of the chapters are written by Vision Solutions staff, and others are written by guest experts in their fields. As overviews, they are short on specifics but long on the broad coverage of their subjects. Though Vision Solutions is an IBM Premier Business Partner focusing on business continuity solutions for IBM systems, the book is wonderfully non-vendor specific.

To date, the Business Continuity Today eBook has published the following chapters:

Chapter 1: Planning for Business Resilience
Chapter 2: Security Threats, Trends, and Risks
Chapter 3: Optimizing Data Availability in Information Hungry Organizations
Chapter 4: Compliance and the New Reality for IT Resiliency
Chapter 5: Maintaining Availability with Virtualization Technologies
Chapter 6: Data Replication: Making Sense of the Options

Its forthcoming Chapter 7 is entitled "Meeting the Recovery Imperative."

Chapter 1: Planning for Business Resilience

Published in February, 2007, and authored by Bill Hammond, director of Vision Solutions' product strategy for the MIMIX family of information availability software solutions, "Planning for Information Resilience" is an overview of the needs for business continuity. It is targeted to those trying to familiarize themselves with the field.

The author notes that business resilience encompasses what is commonly known as business continuity, continuous availability, high availability, and data protection and recovery. All of these terms relate to the protection of the continuous availability of an enterprise's IT services and data.

System downtime can be caused by unplanned outages as well as planned downtime for scheduled maintenance. Surprisingly, studies have shown that planned downtime creates 70% to 90% of all IT interruptions.

Depending upon the application, acceptable downtime can range from multiple hours per year to absolute continuous availability with no data loss. High-availability solutions can be extended if desired to disaster-recovery solutions that protect against major data-center disasters.

The first step in determining an organization's optimum availability strategy is to determine the allowable amount of downtime (RTO – recovery time objective) and data loss (RPO – recovery point objective) for each application. Next, service level agreements (SLAs) must be defined to determine whether these objectives are being met based on meaningful measurements.

Ideally, an application will have an RTO and an RPO of zero (no downtime, no data loss). However, as RTO and RPO approach zero, the cost of the availability solutions significantly increase. Therefore, the availability requirements for each application must be mapped onto a continuum of availability solutions that will provide an acceptable balance between availability and cost.

The author discusses several approaches to resilience along the availability/cost curve. Off-site hot sites provide remote backup of entire systems but can take hours to days to recover and may lose hours to days of data. The same is true for magnetic tape backup and recovery. Disk-based backup and recovery can reduce these times significantly, but objectives are still measured in hours. Continuous data protection (CDP) that captures data modifications as they occur can reduce RPO and RTO to near zero.

High-availability software solutions use a backup server with a current replica of the application environment that can take over in the event of a primary system failure in minutes or seconds with zero loss of data. These solutions are becoming affordable enough to be considered by small to mid-sized businesses.

Chapter 1 ends with a pair of brief case studies.

Eliminating downtime via business resilience can reap many rewards, including increased profitability, customer satisfaction, market share, and revenue. It can solve the availability issues of working with offshore customers and ensure that the enterprise can meet new and existing regulatory requirements.

Chapter 2: Security Threats, Trends, and Risks

Viruses, worms, denial-of-service attacks, and other malware can cause outages to Internet-based applications just as system failures can. IBM's Internet Security Systems (ISS) X-Force has been monitoring the presence and growth of malware for over a decade. Chapter 2, published in April, 2007, and authored by Gunter Ollmann, director of security strategy at ISS, is an in-depth look at what X-Force has found.

During 2006, identified vulnerabilities in commercially available applications increased at an annual rate of 40%, totaling over 7,000; and this rate is exponentially increasing. The largest originators of spam were the U.S., Spain, and France, with 93% of spam being in English (German is the next most popular language). U.S.-based businesses are most often targeted by phishing emails. Traditionally, the weeks before Thanksgiving and Christmas are the busiest.

Software vulnerabilities are usually divulged by the vendor. The top ten vendors disclosing vulnerabilities accounted for 14% of the total vulnerabilities. 14% of these vulnerabilities remained unpatched, whereas 65% of the other vulnerabilities remained unpatched.

The most highly prized vulnerabilities are those than can be exploited remotely, providing an attacker with the greatest opportunity for host compromise. In 2000, 44% of all vulnerabilities

were remotely exploitable. This proportion grew to over 80% by 2006. The consequences of exploitation include:

- gaining access
- denial of service
- bypassing security
- informational
- data manipulation
- obtaining information
- gaining privileges
- file manipulation

Spam is particularly intrusive and increased by 100% in 2006. In an average day, X-Force analyzed over 150,000 spam messages. Half of all spam messages pass through no email servers because most spam messages are sent by botnets directly to the recipient.

Image-based spam is exceptionally difficult to detect, and the outlook is that this is going to get even worse as spammers improve further on their technology. For example, animated GIF images and multicolor embedded text in images are being increasingly used to propagate spam. Image-based spam increased from a little over zero in 2005 to over 40% by 2006.

Modern malware is now the digital equivalent of the Swiss army knife. It includes, in order of occurrence, downloaders, Trojans, worms, backdoors, dialers, password stealers, viruses, and key loggers.

Internet Explorer reported 34 vulnerabilities in 2006, and Firefox reported 64.

Internet security breaches are costing businesses billions of dollars a year. Businesses need to aggressively address Internet malware to preclude downtime, loss of customers, decreased employee productivity, lost revenue, and the risk of impending lawsuits.

Chapter 3: Optimizing Data Availability in Information Hungry Organizations

Chapter 3, published in June, 2007, and written jointly by IBM and Vision Solutions, deals with the differentiation between companies in how they use information. At the core of the problem is the distribution of a company's data over a variety of systems and databases due to evolution and acquisition. Data availability is the ability for every organization to consolidate this data and to ensure that the right information reaches the right people at the right time.

Three characteristics of data stand in the way of full data availability:

- Complexity – more applications, more nontraditional data, and more nontraditional formats (digital video, IM, email, wireless, etc.).
- Volume – studies show that data volumes are doubling every six to eighteen months.
- Latency – the single biggest impediment to real-time decision making.

Classically, data integration has been achieved with periodic ETL (extract, transform, and load) utilities that run in off-hours. Real-time integration is more closely approached with application integration or message-oriented middleware products. However, these approaches typically require significant application modification and, in some cases, additional intermediate servers.

Coming to the rescue are software-based replication products. These products work at the database level, not the application level, and therefore are generally nonintrusive to the applications. They are capable of on-the-fly data transformation to support replication between heterogeneous databases.

An ultimate goal is the dynamic warehouse. Rather than receiving data in periodic batches, the dynamic warehouse receives data changes in real time from applications throughout the enterprise. It is available to any business process or user on demand. A powerful approach to achieve the dynamic warehouse is today's open-source standards. Using SOA (service-oriented architecture) to package legacy services and XML (extensible markup language) to communicate with the services, applications can be easily integrated as a set of services available to other applications.

Coupled with powerful real-time analytics, users can be provided with the current information they need to optimize the business processes of the enterprise. Dynamic warehousing today is not a product. It is an approach requiring:

- information integration to aggregate and transform information from disparate data sources.
- data replication to acquire data from transaction databases and to support secondary replicas.
- data management to ensure a common view of customers, partners, and products.
- search and text analytics to extract useful knowledge from unstructured information.
- publishing to notify applications of actionable events.

The benefits of information integration include speeding the time to market for new applications, extending the value of existing investments, gaining further insights into the business, controlling costs, and increasing developer productivity.

The chapter concludes with several case studies showing the benefits of dynamic warehousing.

Chapter 4: Compliance and the New Reality for IT Resiliency

This chapter, published in February, 2008, is authored by Thomas M. Stockwell, an independent industry analyst. It deals with the modern-day challenges of regulatory compliance.

Following the collapse of Enron Corporation in 2001 and the collapse of its accounting firm Arthur Anderson for nefarious practices, Congress passed the Public Company Accountability Reform and Investor Protection Act of 2002, better known as the Sarbanes-Oxley Act, or SOX. This act was followed by a number of other regulatory acts, which together form a compliance burden on all companies.

Compliance describes any set of rules, regulations, standards, or agreements that create a framework for fulfilling an organization's commitment to its customers, business partners, shareholders, and government entities. As IT looks at how to be compliant, it should also be looking for ways to make its infrastructure more resilient. The more resilient the IT infrastructure is, the more efficiently it will meet its compliance requirements.

There are several common types of compliance requirements that IT must meet:

- internal business standards
- governmental regulations
- industry regulations
- international standards
- service-level agreements

All of these compliance requirements impose two burdens on the IT infrastructure – reliability and availability. It is the Business Impact Analysis (BIA) that provides the process to get from the identification of compliance requirements to the implementation of these requirements. BIA is an information-gathering exercise that identifies the functions performed by the organization, the

resources required to support each function, the impact of failing to perform a function, a recovery time objective (RTO) for each function (the time that it can be out of service), and the recovery point objective (RPO) for each function (how much data can be lost following a failure).

The goal of the BIA is to give the company the information it needs to move toward a resilient infrastructure. It should also include specific recommendations for security, information retention, data and service consolidation, service-level agreements, and the data and application resilience of the current IT infrastructure.

The author provides a table of some of the larger compliance regulations and the impact that they have on an organization. He targets five areas for improved resiliency:

- Reducing backup time to reduce or eliminate the amount of nonproductive time required to back up systems. This can range from the use of improved tape technologies to maintaining a replicated backup system. The backup system can be kept synchronized with data replication, switched storage, or cross-site disk mirroring.

- Eliminating planned downtime for maintenance and upgrade activities. This requires the existence of a backup system that can be put into service while the primary system is taken offline.

- Disaster recovery for critical applications so that they can survive a catastrophic interruption over an extended period of time. This also requires a remotely located backup system that may or may not be kept synchronized with the primary system depending upon the RTO and RPO objectives.

- High availability for unplanned outages so that critical applications may continue, also achieved with a synchronized backup system.

- Balancing workloads to comply with service-level agreements.

The author then uses a series of RTO/RPO diagrams to develop the various resiliency options available to a company for single-system and multisystem infrastructures.

A major requirement of the new regulations is that of audit. The IT infrastructure must be audited for compliance frequently and independently. From a resiliency viewpoint, each resiliency technique should have a documented audit procedure. For instance, if a backup system is available, a role swap may be periodically demonstrated.

The recognition of the role of IT in compliance has taken on a new meaning since SOX. This is particularly true when it comes to business-continuity planning. Studies have shown that 40% of all companies that experience a disaster go out of business within two years, yet 40% of companies do not consider business-continuity planning a priority. Consequently, a well-formulated business-continuity plan can actually serve as a strategic competitive driver.

Chapter 5: Maintaining Availability with Virtualization Technologies

Chapter 5 is an excellent overview of the benefits and pitfalls of server virtualization. Virtualization allows a company to consolidate several older, slower, or underutilized physical servers onto a single server as “virtual machines” (VMs). The move to virtualization is fueled by the rapidly increasing power of today’s servers and the consequent benefits of reduced data-center costs. The major pitfall is availability – the failure of a single physical system running many VMs can take down many critical applications in one fell swoop.

Virtualization can also be extended to multiple storage systems. Virtualized storage allows applications to view diverse storage systems as a single storage system.

Consolidation via server virtualization can reduce the total cost of ownership of the IT infrastructure by requiring the acquisition of fewer (albeit larger) servers, reducing license and energy costs (virtualization is green), and downsizing the data-center facility. Labor costs can be reduced since system administration is simpler. In fact, data centers can be consolidated, eliminating entire facilities. Legacy applications can be moved from legacy systems and migrated to virtual machines running on modern, open industry-standard servers.

Storage virtualization consolidates multiple storage systems into a unified group that appears as a single storage facility that can be managed, backed up, and replicated with synchronized automation and a reduced administrative force. Though no physical storage systems may be saved, storage is more efficient, flexible, and manageable.

One pitfall of virtualization is the “all of your eggs in one basket” syndrome. If the physical server fails, it takes many virtual machines with it. Therefore, backup facilities are required to take over the support of all of the failed VMs. Furthermore, all of the failover software facilities required for single servers are required for virtualized servers.

Another pitfall is “server sprawl.” Companies expect that virtualization will reduce the number of physical servers that they must provide. This is true. However, as soon as departments realize that new virtual machines can be obtained at the cost of a few clicks, controls must be in place to prevent virtual-server sprawl.

Good VM management tools are a necessity. VMs can be easily migrated from one physical server to another for load balancing and fault recovery. Unless the impact of these movements is monitored, severe performance problems may suddenly occur due to the overloading of a physical server.

A company must evaluate what applications to virtualize. Good candidates include test and development environments, smaller applications, applications whose utilization does not vary widely, and perhaps some legacy applications. Poor candidates include compute-intensive applications, graphics and CAD systems, and applications requiring specialized hardware. Applications that should be held off until the virtualized infrastructure is working well include mission-critical applications using many servers as well as new, untested applications.

Best practices dictate that virtualization should be approached cautiously. Start out with a few test and other noncritical applications in a safe sandbox. Practice with these until all the kinks are worked out, and then proceed carefully with expanding virtualization to the rest of the data center.

Chapter 6: Data Replication: Making Sense of the Options

Chapter 6 was published in February, 2009, and deals with data replication – a subject close to our heart. Data replication is used to create a copy of a source database onto a remote target database. Data-replication engines move data changes from the source to the target over high-bandwidth communication lines.

The source and target databases may not be exact copies of each other because the databases may be heterogeneous (different schema or even different vendors) or because they may be satisfying different needs (the source may be doing transaction processing while the target is being used for complex queries).

Replication may be bidirectional. In this case, each system acts both as a source of data and a target for data.

There are many reasons to replicate:

- Application Integration – Data applications can be integrated, often seamlessly, at the data level. With replication, when an application makes a change to a data item, that change immediately appears in the databases of other applications.

- Data Sharing – Common data can be shared by multiple users at once without interference. This is especially useful if users are geographically dispersed by replicating to remote systems that are local to concentrated user populations.

- Business Intelligence – Enterprise-wide data can be replicated to a real-time data warehouse for extensive data mining, for running ad hoc queries against current data, and for generating real-time events to trigger corporate actions.

- System Upgrades and Migrations – The target system can be upgraded and fully tested while its database is kept synchronized with the source database. When ready, it can be put into production relatively risk-free by switching users over to it.

- Disaster Recovery –Data can be replicated to an offsite system that can be put into operation if the primary system is taken down by a disaster.

- High Availability – Data can be replicated to an onsite backup system that is prepared to take over should the primary system fail.

- Data Vaulting – Data can be replicated to a secure disk facility for later restoration if needed.

There are several technologies today for data replication. One is disk mirroring, in which data is written to two separate disks attached to the same system. Disk mirroring protects against a disk failure. However, the mirrors cannot always be separated by the large distances required for disaster protection. In addition, the system cannot be taken down for maintenance without taking down all applications.

Software-based replication solves these problems. These products use some sort of transaction log to replicate data changes to remote systems. Software replication is more flexible and often more efficient than replicating at the storage-system level. It can provide all of the replication benefits discussed above. However, due to subtle faults, some administrative monitoring and maintenance of the target system is often required.

Data replication can support *continuous data protection*, or CDP. With CDP, corrupted data can be reconstructed and restored.

Data replication is an important component of high availability. However, high availability requires much more. For instance, it must provide for the role-swapping function should a system fail.

Replication features that should be considered when choosing a replication engine include platform coverage (does it cover the required platforms in the data center), data collision detection and resolution if bidirectional replication is used, flexible topology support, guaranteed data delivery, monitoring and control, and data movement methods (batch versus real time).

Chapter 7: Meeting the Recovery Imperative

The next chapter to be published is described by *Business Continuity Today*:

”Recovery is no longer just about disasters. Just having servers or network connections down for a few minutes, or even a few seconds, can be an operational and financial catastrophe. You must be able to recover quickly from any event, big or small, that threatens business continuity.”

You can register for this chapter and others by following the link given at the beginning of this article.