Using Virtualization to Prepare Your Data Center for “Real-time Assurance of Business Continuity”

Rao Mikkilineni  
Kawa Objects Inc  
251 Yerba Santa Ave  
Los Altos, CA 94022, USA  
rao@kawaobjects.com

Gopal Kankanalli  
VirtualXL Inc.  
4620, Fortran Drive, Suite 206  
San Jose, CA 95134, USA  
gopal@virtualxl.com

Abstract— This paper describes our experience using the dynamic resource reallocation capabilities offered by virtualization technologies, in implementing:

1. Automation of end-to-end failover of mission critical virtualized application, a SAN network and EMC Clarion based storage to a remote site and

2. On-demand and scheduled assurance of failover and measurement of RPO and RTO (The Recovery Point Objective (RPO) is the point in time to which you must recover data as dictated by business needs. Recovery Time Objective (RTO) is the period of time after an outage in which the application and its data must be restored to a predetermined state defined by RPO.)

By creating an application to spindle resource utilization profile through various management systems, and utilizing a combination of server, network- and storage- virtualization technologies, application specific RTO and RPO objectives (defined based on workload profiles and business priorities) are met in a technology agnostic and multi-vendor environment.

Keywords—Real-time Business Continuity, Virtualization, Datacenter, High Availability, Disaster Recovery

I. INTRODUCTION

There are four ways in which business applications today can adversely impact business continuity (the ability to operate business without interruption to deliver quality service and generate revenue):

1. Drastic response time degradation of one or more applications causing customer frustration,

2. Planned down time for maintenance or upgrades,

3. System failures that cause short term service disruptions and

4. Application downtime during a disaster (such as an earthquake or a terror attack that can be hard to recover from without appropriate preparedness.)

Real-time assurance of business continuity (RABC) utilizes processes, technologies and people resources to implement:

1. Performance optimization by allocating the right resources (by tuning CPU, cache, throughput, bandwidth and server, network & storage I/O) to the chosen application based on workload demands and business priorities,

2. High Availability (HA) by providing instant failover and recovery of the application to its normal operation and

3. Instant disaster Recovery, from a remote secondary site when the primary site fails.

Fortune 100 companies have addressed these issues by throwing massive resources in the form of people, technology and processes. They overprovision their infrastructure and deploy armies of service consultants and systems engineers to operate and support multiple vendor systems and multiple technologies that often duplicate various functions. The fortune 100 company data centers are often filled with large number of consultants from server, storage and network equipment vendors and a lot of shelf-ware that was supposed to have reduced their Total Cost of Ownership and Return on Investment many times over. However, small and medium enterprises often cannot afford to throw resources to address these issues and struggle to make do with inadequate solutions, which increase their vulnerability in addressing business continuity.

In this paper, we will describe how various innovative virtualization technologies are leveraged to reallocate or restore server, network and storage resources to a critical application based on a critical application based on business priorities. These technologies offer a new paradigm in which various failover alternatives are integrated into the routine operational lifecycle of these applications. Using the dynamic resource reallocation capabilities offered by virtualization technologies, mission critical applications such as Exchange email and a transaction intensive database application are reconfigured and deployed for rapid recovery through dynamic reallocation of the server, network and storage resources. Leveraging virtualization technologies enables us to rapidly improve performance or restore the application and therefore minimize the interruption of the services they provide. We call this Real-time Assurance of Business Continuity (RABC) through virtualization where the “Real-time” can be designed to be non-disruptive to the application of interest or can be minutes to hours depending on business risk tolerance and affordability.

This paper describes the experiences from assisting VirtualXL, a Silicon Valley startup, to bring, at an affordable cost, enterprise-class business continuity solutions to small and medium businesses by leveraging emerging virtualization
technologies. By focusing on end-to-end (application to spindle) resource utilization in application planning, design and implementation phases, the right technology choices are made to optimize both operation & management with desired performance and availability objectives. By utilizing a combination of server, network- and storage- virtualization technologies, business objectives are met in a technology agnostic and multi-vendor environment. The business priorities are used to define the availability, performance and cost objectives during the configuration and deployment phases.

This paper presents our process, results and lessons learned in the planning, design and implementation of the “RABC” strategy at one of our customer’s site. Based on our experience, we are developing a checklist, an audit process and new set of tools to make the planning, design and implementation of mission critical applications more efficient and error-free, therefore reducing business risk without breaking one’s bank. In section II, we will discuss current Business Continuity practices. In section III, we will present our approach to Real-time Assurance of Business Continuity and in section IV, we present the conclusions drawn from our experience.

II. BUSINESS CONTINUITY IN PRACTICE

The “raison d’être” of the Enterprise Data Center is to provide right Information Technologies to develop & deploy business applications and assure application availability, performance and security during both service development and deployment phases. Figure 1 shows the evolution of today’s data center with myriad management systems from what originally evolved from a Mainframe/Server architecture.

While the productivity, flexibility and global connectivity made possible with this evolution have radically transformed the business economics of Information Systems, the complexity of heterogeneous and multi-vendor solutions have created high dependence on specialized training and service expertise to assure availability, performance and security of various business applications. Successful implementation must integrate various server, network and storage centric products with their local optimization best-practices with end-to-end optimization strategies. While each vendor attempts to assure their success with more software and services, the Small and Medium Enterprises often cannot afford the escalating software and service expenses associated with optimization strategies and become vulnerable.

In this section, we review current practices in Business Continuity Assurance to set the stage for Real-time Assurance of BC taking advantage of Virtualization.

A. Performance Optimization

Performance optimization is used to ensure that the computing, network and storage resources are optimally utilized by the applications that share them while meeting the changing business needs and workload variations. Various parameters can affect the response time and the user experience such as server CPU cycles, server cache memory size, network bandwidth, storage throughput, I/O per second, capacity etc. Currently, when the workloads vary and there is contention for resources from multiple business applications, performance experts gather various configuration and performance parameters from the server, the network (IP/SAN) and the storage systems, correlate them to various applications and tune various server, network and storage knobs using multiple management systems. This operation today is highly knowledge intensive and labor sensitive.
B. High Availability

There are two types of outages that are associated with information systems – planned and unplanned and High Availability solutions must address them both. A spectrum of solutions is used to implement HA from solutions that utilize custom and redundant hardware to ensure availability, to the other end to solutions that provide software solutions using off-the-shelf hardware components. Typical HA configurations using clustered servers with multipath SAN Network access and expensive shared storage devices with replication used by Fortune 100 companies tend to be expensive, require skilled work-force with deep knowledge about server, SAN and storage technologies and their management. Again the SMEs often cannot afford to implement these solutions either because they cannot afford or do not have the skilled work-force.

C. Disaster Recovery

In his book “Disaster Recovery Planning: Preparing for the Unthinkable”, John W Toigo [1] states: “as in the case of the 143 companies that simply disappeared in the months and years following the 1993 bombing of the WTC, many of the companies that endured the 9/11 tragedy without a continuity plan will likely not see the end of the decade. These companies learn their lessons about the importance of disaster recovery planning the hard way, adding further pain and anguish to the already sad memory of that awful event.”

On the other hand, Former Nasdaq CIO, Bailar, when asked by the CIO magazine [2] how Nasdaq systems were all up although some of their traders were in trouble, replied “Nasdaq is highly redundant. Every single one of our traders is connected to two different Nasdaq points of presence or connection centers, and there are four connection centers alone in downtown, Manhattan, none of which were really affected. Including Manhattan, there are twenty connections centers around the United States, and every single server connects to two of those centers through different paths, and often through two different vendors”.

More recently, Vincent Biddlecombe, CTO of Transplace explains [3] how new technologies such as server and storage virtualization, NetApp’s SnapMirror, and shared storage approach with Network Attached Storage (NAS) played a vital role in deploying a lower cost, real-time multi-site backup and recovery strategy.

The literature on Disaster Recovery is enormous. When you wade through various books, prescriptions, case studies and blogs, you learn that:

1. Balancing cost and benefits requires understanding your business priorities, risk of failure, impact of outage and application objectives such as mission criticality, availability, customer impact etc.

2. Planning, designing and implementing a right DR strategy can take from two months to two years depending on the size of your enterprise and each application Recovery Point Objective (RPO) and Recovery Time Objective (RTO). Planning must involve people, processes, technology and training

3. Successful implementation requires continuous supervision and control of DR operations & management which include periodic checks. Verification must take place to ensure recovery procedures and technologies function as expected. Many DR planning and design efforts end up as shelf-ware and even those backup strategies implemented are never verified that they really work in the event of a real disaster.

4. A successful strategy involves multi-site, multi-technology integration and a tiered DR architecture

5. Choosing the technology solutions from myriad options to match your cost constraints and meet your business objectives is a knowledge resource intensive task. Complex DR Systems often require expensive consultants with system integration and deployment expertise that spans across multiple technologies and multiple vendor products.

The complexity involved in choosing a DR solution, at a reasonable price, that meets the business needs and designing a manageable process that can be validated, presents a formidable problem to the IT staff in seamlessly integrating all the tasks involved. As a consequence, many DR plans, products and implementations end up as ineffective shelf-ware and provide no help when a real disaster strikes.

With the advent of newer technologies such as server, storage and network virtualization, it is now possible to design a “live” DR strategy that continuously assures that the various sites providing the DR solution are synchronized and the system is designed to failover correctly based on RTO and RPO requirements. This approach allows the DR planning to be incorporated into the application and IT resource allocation-planning, design and implementation phases and assures appropriate failovers occur to restore the application to its expected state when a disaster strikes.

In addition, application performance parameters such as bandwidth, storage capacity, throughput, and response time need to be monitored and tuned on an ongoing basis as part of the day-to-day operation & management of IT resources. The “Real-time Assurance of Business Continuity” strategy:

- Reduces the cost and time involved in performance tuning,
- Eliminates down-time during scheduled maintenance,
- Provides High Availability (through instant-recovery from local backup) and DR implementations involving a remote site and
- Reduces operational errors by incorporating automated DR operations and management into routine IT processes (e.g., routinely verifying that HA/DR strategy works and measuring the respective application RTO and RPO.)

In this paper we present an end-end (application–to–spindle) server, network and storage resource planning, design, implementation, operation and management to meet two mission critical application Business Continuity requirements. This document describes our experience in implementing RABC for two applications that are often found to be most critical to organizations today, namely, the email and a
transactional data base. We chose appropriate virtualization, networking and storage solutions to implement a Real-time assurance of HA and DR solution and integrated on-demand and scheduled monitoring and testing of RTO and RPO in routine operation and management processes in our customer’s data center. In Section III, we describe the RABC solution that consists of 4 steps:

1. Automate discovery of IT resources and their profiles to do RABC planning
2. Analyze application profile (configuration requirements, workloads and business priorities) for RABC and match the right computing, storage and network resources from discovery
3. Monitor application to spindle configuration and performance characteristics that affect Business Continuity
4. Automate:
   a. Failover in the case of trouble or disaster and
   b. Testing of RTO/RPO on-demand and on a schedule, using failover automation

Section IV concludes with lessons learned and future plans.

III. VIRTUALIZATION AND IMPLEMENTING REAL-TIME ASSURANCE OF BUSINESS CONTINUITY

Virtualization broadly refers to an abstraction layer for computer resources and allows a shared environment where multiple applications share computing, network and storage resources across multiple hardware, OSs, protocols, and vendor platforms.

Virtualization technologies are radically transforming the way in which business applications utilize computing, network and storage resources in such a shared environment. Resulting scales of economy and energy savings are driving the next generation Data Center consolidation and reshaping the IT infrastructure deployment and management. There are three major advances in virtualization technology:

1. Server virtualization has demonstrated significant energy savings and allows high availability, manageability and improved disaster recovery for application servers and desktops.
2. Network based virtualization shields the applications from heterogeneity of both network & storage resources and allows dynamic reallocation of network and storage resources to applications.
3. Storage virtualization allows multiple vendor storage subsystems to be grouped together into various types of tiered storage pools and improves the cost effectiveness of information life-cycle management. For example, rarely used data can be moved to the lower tier, low-cost storage pool while a mission critical performance intensive database application must be supported by the top tier high speed disk storage.

The addition of virtualization technologies enables the application to become transparent to physical computing, network, and storage resource pools that are allocated to it. The benefit of this transparency is that the resources can be switched dynamically without experiencing application downtime. This dynamism is used to design failover alternatives for the applications by switching the server computing, OS, IO resources, network protocols and storage resources without shutting down the application. This capability is referred to as “instant HA/DR”. By monitoring and knowing which resources are about to fail, failing or have failed, we can configure applications to dynamically fail over to utilize alternate resources.

A. Implementing End-to-end Application Failover Using Virtualization

To paraphrase Rentsch [4], "Virtualization will be in this decade, what object programming was in the 1980's and structured programming was in the 1970's. Everyone will be in favor of it. Every manufacturer will promote his products as supporting it. Every manager will pay lip service to it. Every system administrator will practice it (differently). And no one will know just what it is."

As with any evolution of technology, the acceptance will be gradual and with time, the solutions will mature to meet the market need. Industry leaders will adopt successfully those features that provide a competitive differentiation. Others will then follow to keep up with competition.

On one hand, server, network, and storage virtualization technologies offer unique capabilities to dynamically reallocate application computing, network and storage resources without shutting down the application. Using these capabilities, end-to-end failover alternatives can be implemented to improve the availability and recovery characteristics to meet the business requirement. On the other hand, the multitude of choices and the complexity of options require a deep understanding of an array of technology solutions from different vendors that often compete with each other and do not fully interoperate. Successful end-to-end optimal configuration often requires integration across multiple technologies and multiple vendor solutions. Until the technologies mature and are integrated, which may take a decade or more, no single vendor will have an end-to-end optimal solution. Pieces have to be integrated together often adopting best practices in various technology domains and vendor solutions. Choices and trade-offs have to be made between end-to-end design goals and local domain (server, network and storage) best practices. This has created a labor-intensive service business that increases the Total Cost of Ownership (TCO) in a data center.

In this paper we present our implementation of an end-to-end solution with an “instant DR” objective and present lessons learned. Using appropriate scripts to control various network, storage and virtualized servers, and their management systems, we have automated failover strategies. We hope that this will pave the way toward better cost effective solutions that transcend technology and vendor boundaries to meet growing market needs. This will also help in reducing the TCO through integration of multiple technology and vendor products and automation of end-to-end optimization using dynamic reallocation of resources.
B. Application Profiling

In order to optimize the use of IT infrastructure and deliver the requisite HA/DR support of each application, we need to carefully review the resources allocated. Computing is usually defined as Computer processor cycles, local hard disk storage and supporting memory to load the application into an operational state. Networking provides the avenue to other physical components and data on other devices. Storage devices, both large and small, can be local, network or fibre attached and these devices have various capabilities to support increased performance, availability and prevent data loss.

The goal of application profiling is to optimize resources and guarantee the business requirements regarding performance, availability and recoverability. We can assess and profile each application to better understand how it uses the physical components of the IT infrastructure that it runs on. As load is increased on the application, we understand how IT resources can be increased or tuned to deliver optimal functionality and data access.

Beyond the infrastructure requirements we must consider the impact on the business if the application is not available. The business requirements dictated by the Business Units that use them will be a significant factor in determining what tiers of hardware will be used in the supporting design. Hardware tiers are defined by the availability and redundancy of the associated hardware and software. The applications that are mission critical to the organization will usually be installed on the highest performance infrastructure tiers with the least possibility of failure. The lowest tiers are usually on commodity hardware which provides little or no redundancy internal to those devices. The highest performance infrastructure tiers usually have complete redundancy and can recover from nearly all internal hardware or software failures.

Different applications utilize IT infrastructure resources differently based on the services that they provide. For example, some applications can be I/O resource intensive or very throughput oriented while others serve relatively static data. With this in mind, the applications' computing, network and/or storage resource requirements will be dramatically different and their configuration must match their requirements, for optimal resource utilization. The example depicted below illustrates the differences between a transactional application and a relatively static application. Transactional applications that require fast disk and failover/recovery times in minutes will usually utilize faster SAN or NAS storage architectures. The additional issues surrounding fast growth will also benefit from the “Thin Provisioning” provided by storage devices. The environments that leverage “Thin Provisioning” technologies will benefit from dynamic allocation of infrastructure resources to adapt to capacity and performance requirements as the application changes over time. The business dictates whether the application must be highly available and recoverable to a specific point in time or recoverable within a certain time period. Applications providing services around static data and long RTO can be installed on commodity hardware and low cost local or iSCSI storage. Applications that are mapped to lower tiers of infrastructure will be subject to substantial downtime in the event of a hardware failure. The individual application profile combined with the availability requirements of the business determine the infrastructure resource mapping to each physical IT infrastructure component that the application will reside on.

Figure 2 shows various application requirements based on their business priority.

C. Automated Discovery and Application to Spindle Mapping

With today’s IT management infrastructure it is easy to discover available resource pools and their configuration, utilization, and performance characteristics. There are many open-source tools and management systems that collect and present this information. Figure 3 shows the discovery process using various management tools that are available.

Virtualizing computing, network and storage resources allows dynamic provisioning of these resources based on changing workload and business priorities and implement automated failover strategies.

D. Monitor and Automate Failure

Virtualization changes the conventional HA/DR model altogether. Passive node clustering solutions waiting to takeover in case of a DR event are not cost effective and are no longer necessary. Adopting innovative virtualization and existing replication technologies, an application can be failed over seamlessly and no passive node is needed.

Figure 4 shows implementation of RABC for a customer.
The configuration shows three applications (Active Directory, Exchange and Oracle database application) virtualized and replicated using EMC storage system Clarion. Both sites are active. The system monitors all relevant events from VMWare Operations Manager, Switch management systems and Clarion Management interface, correlates them and automates failover of the Virtual applications, switch connections and storage configurations as required to provide the HA/DR solution.

E. Test RTO & RPO on Demand and On-Schedule

The system also provides a push-button approach to actually cause fail-over on-demand or based on a scheduled event, as a routine operation process and measures and provides the actual RTO and RPO for each application. This feature assures that the HA/DR function is performing according to the expected SLAs and also adjust the SLAs as the workloads and business priorities change over time. Figure 5 shows the screenshot depicting the application and it’s RTO.

IV. CONCLUSION

Virtualization technology provides dynamic provisioning capabilities which are used to match application workload and business requirements to appropriate computing, network and storage resources (CPU, memory, bandwidth, throughput, IOPs, storage capacity etc.). Using virtualization, we can now implement real-time assurance of business continuity solutions to ensure application-specific availability, performance and security SLAs are met. In this paper we have presented the implementation of end-to-end RABC solution for business applications within an enterprise. Server virtualization is integrated with network and storage resource management systems to create an end-to-end application to spindle resource configuration and performance map. Based on continuously monitored server, network and storage events, from various sources, resources are dynamically re-provisioned to meet business requirements.

A side benefit of this approach is the push-button testing of HA/DR-function on-demand and measure the RTO and RPO to assure desired SLAs. System administrators can take comfort in knowing that the HA/DR plan is regularly scheduled and tested to verify that ongoing changes to the environment have not prevented the DR plan from functioning correctly.

The RABC solution incorporates failover alternatives in day to day operation and management to assure that the RTO and RPO objectives set for the application recovery are met when a disaster really occurs. In addition the solution provides the following benefits:

1. DR planning, design, implementation, operation and management are integrated with application planning, design, implementation, operation, management and optimization processes.
2. Dynamic provisioning of virtualized resources allows application FCAPS (Fault, Configuration, Accounting, Performance and Security) optimization to meet the changing workload and business demands using shared environment.
3. Multiple fail-over strategies are combined to provide “instant” High Availability/Disaster Recovery to meet Recovery Time Objective and Recovery Point Objective requirements.
4. Customer cost savings and benefits increase by eliminating the need for passive nodes.

REFERENCES

[4] Rentsch, T, Object Oriented Programming”; SIGPLAN Notices; Sept 1982; Vol.17 No.12; P51

1 FCAPS is the ISO Telecommunications Management Network model and framework for network management. FCAPS is an acronym for Fault, Configuration, Accounting, Performance, and Security which are the management categories into which the ISO model defines network management tasks.