

VERITAS High Availability Solutions for DB2 Universal Database

TABLE OF CONTENTS

TABLE OF CONTENTS	2
Introduction	3
The Causes and Impact of Downtime	3
The Different Types of Downtime	4
Planned Downtime	4
Calculating the Cost of Planned Downtime	4
Unplanned Downtime	5
Calculating the Cost of Unplanned Downtime	5
High Availability Strategies	5
Strategies to Minimize Planned Downtime	5
Online Administration	5
Point in Time Copy	6
Performance	6
Clustering	6
Service Groups	7
Cluster Scalability and Manageability	7
Manageability	8
Replication	8
Global Clustering	8
Reliability	9
Backup – Your Insurance Policy	9
Conclusion	9

INTRODUCTION

Today, the IBM DB2 Universal Database (UDB) is widely used to support mission critical enterprise level applications such as: Enterprise Resource Planning, Supply Chain Management, and Customer Relationship Management solutions. In addition, DB2 is frequently used as the foundation to support high-end data warehousing applications.

For many enterprises, these critical applications and business intelligence systems create the backbone of their business – managing business critical data, supporting key business operations such as order entry and inventory management, and decision-making exercises. The enterprise's productivity, performance, and its ability to serve its customers and employees will be impacted if the database is unavailable or under performing.

For this reason, enterprises that use DB2 must invest in a computing solution that allows them to deliver excellence of service in the area of database availability – and still maintain the levels of performance, manageability, and reliability that their users and customers have come to expect.

This paper will examine:

- The causes and impact of downtime on an enterprise
- The different types of downtime
- Strategies for minimizing downtime
- VERITAS solutions that ensure availability while maintaining quality of service in the areas of performance, manageability, and reliability

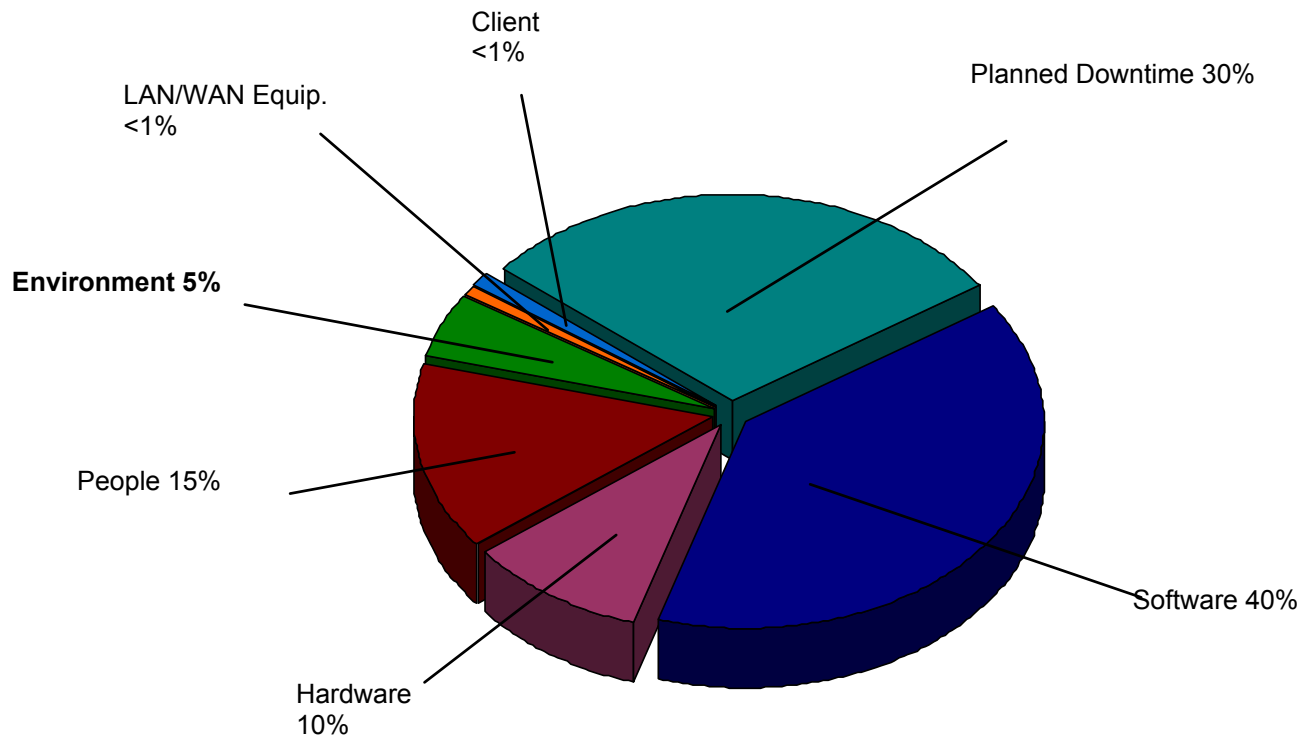
THE CAUSES AND IMPACT OF DOWNTIME

Unplanned, excessive downtime can dramatically impact the user experience and productivity of your internal and external customers. Downtime for internal users can impact employee productivity, product development cycles, and time to revenue. It can also affect the perceived value that the IT department provides to users. For external users and customers, downtime impacts customer perception of quality of service and the level of customer support, it may also impact their ability to purchase goods and services. For investors, downtime can translate directly into a decline in stock price.

In data centers worldwide, much attention and investment has been made to ensure redundancy and high availability of hardware system components that house corporate data. However, it has been demonstrated that hardware failures are the cause of downtime only 10% of the time. The leading causes of downtime are software, people, and planned downtime itself. In fact, as much as 30% of all downtime is pre-scheduled. Another 40% of downtime is due to software errors – something as simple as a DB2 running out of disk space and halting operations as a result.

Availability can also be impacted by external factors such as the environment (e.g. power, cooling, and network connectivity), the client, and users. The availability of DB2 and the applications that it supports is determined by the frequency with which any one of these factors occurs – and the time it takes to fix the problem, and get the application back online.

THE CAUSES OF DOWNTIME



Source: IEEE Computer

THE DIFFERENT TYPES OF DOWNTIME

Companies that plan for and proactively manage downtime can significantly minimize the impact of downtime on their businesses. To understand the best strategy for managing downtime, it is important to understand that there are two fundamental types of downtime: planned and unplanned downtime.

PLANNED DOWNTIME

Planned Downtime is generally associated with maintenance windows that are necessary to maintain and upgrade systems.

There isn't an end user out there who hasn't witnessed the IT announcements about unavailable servers or websites. If a print server is unavailable, an individual may be impacted but the corporation can continue to operate without incident. If, however, a global corporation's business critical application cannot be accessed, or if an e-commerce website is taken down for servicing, major consequences can occur.

There are an increasing number of new technology solutions that help to address planned downtime including technologies that allow for online database maintenance. Clustering technologies allow for the downtime (planned and unplanned) of one or more particular nodes while the rest of the nodes in the cluster continue to function normally – with little or no impact on end users.

Calculating the Cost of Planned Downtime

Common causes of planned downtime in clustered and non-clustered environments include:

- New server installation and integration
- Storage maintenance and reconfiguration time

The annual cost associated with new server installation and integration is comprised of the number of new servers introduced, the labor cost per installation, the total number of hours during which the server and associated storage is unavailable, and finally the lost revenue due to lost transactions during planned downtime.

The annual cost associated with storage maintenance and reconfiguration time is made up of the number of storage reconfigurations per year, amount of data stored, projected annual downtime for storage reconfigurations, and the total annual labor costs required for reconfiguration. Revenues lost due to planned downtime must also be added to the mix.

High availability solutions like those provided by VERITAS benefit enterprises directly by decreasing annual server downtime and labor costs associated with new server integration as well as storage maintenance and reconfiguration, resulting in minimized revenue loss due to missed opportunities.

UNPLANNED DOWNTIME

Unplanned downtime can be the result of something as simple as a human error or a surge in usage resulting in unacceptable system performance. It can be caused by an incident as dramatic as a natural disaster (flood, fire, or earthquake), or an act of war. Responsible enterprises have made a conscious effort to architect and plan for unexpected downtime. These companies are using technologies that actively monitor the health of mission critical systems for events that can result in downtime – and deploy recovery solutions that have been predefined and tested. With these measures put in place, issues are identified faster, minimizing actual downtime and enabling users to quickly get back to work.

Calculating the Cost of Unplanned Downtime

As with planned downtime, the cost of unplanned downtime is comprised of: total server downtime and the lost revenue associated with this downtime. The most critical factor in recovering from unplanned downtime is recovery time.

In planning an availability strategy, recovery time is an important factor to take into consideration. Recovery time is the period of time from when an application can no longer perform its function to the time when the application is back online again and end users are once again productive. The investment in an enterprise wide availability strategy for a DB2 environment must be proportional to the impact that downtime of the database and the applications it supports has on the enterprise. For some organizations, the time to recovery must be within minutes or hours. For others, recovery time can take as much as a day or longer without significantly impacting the business.

HIGH AVAILABILITY STRATEGIES

The majority of applications supported by DB2 UDB are considered to be mission critical. These applications are core to running the business, supporting customers, developing products, and making key decisions. For this reason, your DB2 high availability strategy should reflect the critical nature of these applications.

STRATEGIES TO MINIMIZE PLANNED DOWNTIME

There are a number of solutions available for managing planned downtime and minimizing its impact.

Online Administration

The ability to perform online maintenance and reconfiguration of the base environment is an important factor in minimizing DB2 downtime. This is particularly true in dynamic DB2 environments where high growth requires frequent tuning to maintain performance as the database grows, or, in large DB2 environments where the impact of downtime is widely felt across the enterprise.

With VERITAS, IT managers and DBAs can conduct common administrative tasks such as tuning the underlying storage infrastructure to occur without shutting down the database. VERITAS also allows administrators to

manage volumes and file systems online permitting tasks such as file system resizing and storage reconfiguration to be accomplished while users continue to access data.

Point in Time Copy

Point in time copy minimizes the planned downtime and performance issues caused by maintenance and backup activities. Using point in time copy, administrators create “snapshots” of their data that can be easily exported to another host. VERITAS allows administrators to create a duplicate database or volume snapshot of data and use them to conduct resource intensive processes like backup, decision support, and reporting – off host. This allows DB2 to run without impacted performance or extended downtime.

Additionally, DB2 can be quickly recovered by restoring the database files from a volume snapshot of by recovering logical objects (tables, etc.) from duplicate database created from volume snapshots. This protects DB2 from logical errors that may be caused by an administrator or an application – and is much faster than a full restoration from tape or other backup media.

PERFORMANCE

Availability solutions from VERITAS allow you to maintain performance levels that your users expect. VERITAS solutions:

- Support faster decision making
- Improve user productivity
- Ultimately allow you be more responsive to your customer and end user

Clustering

Clustering is a widely accepted solution for managing planned downtime and minimizing its impact. Historically, clustering was perceived as a costly solution that was very complex to implement and manage. It was used for only the most high-end, mission critical applications. Today, clustering is a widely adopted as an effective solution for centrally managing and monitoring a group, or, cluster of servers across a variety of applications – especially for databases.

A typical cluster is comprised of two or more servers (nodes). In the event that maintenance is required on the primary node, the primary node simply migrates over its entire service group to the secondary node. Users experience little or no disruption in their work; the primary node can be serviced, and reintroduced to the cluster when the maintenance work has been completed. Administrators have the ability to add or delete resources while maintaining full service. Local clustering is an excellent solution for enterprises that want to maximize availability and minimize downtime of their mission critical databases and their attendant applications.

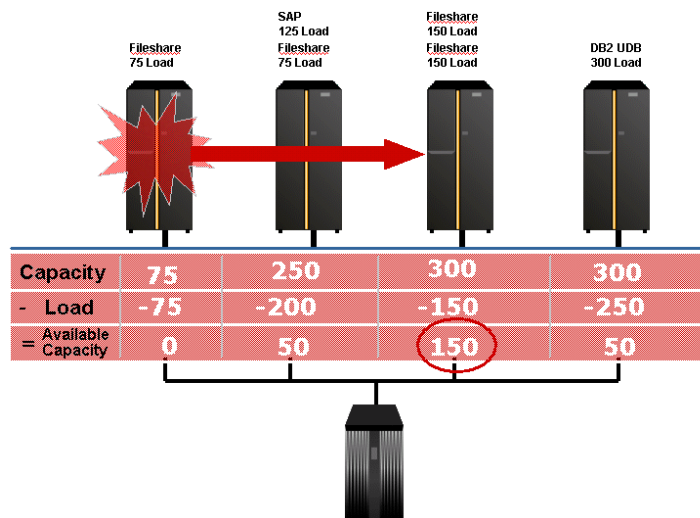
SERVICE GROUPS

There is a tendency to think of business applications and databases as independently functioning elements. Although these major applications and databases are often managed separately, they are intrinsically connected and are supported by a series of common components including: server hardware, data storage devices, operating systems and the network. Database availability can be directly affected if there is an interruption in the service provided by any of these elements.

A Service Group is the collection of resources that are required for a service or application to be available or online. An application service group is typically composed of multiple resources, some hardware-based and some software-based, all cooperating to produce a single service. For example, a DB2 service group may be composed of a logical network (IP) addresses, the database management system software, underlying file systems, logical volumes and a set of physical disks being managed by the volume manager. If this service group needs to be migrated to another node for recovery purposes, all of its resources must migrate together to recreate the service on another node. This can be a complex and time consuming task that begs the need for automation.

Cluster Scalability and Manageability

Clustered environments have proven to be an excellent way to manage enterprise application environments. Clusters allow for overall increased capacity because they scale readily. VERITAS provides support for very large clusters in SAN and traditional client/server environments, providing the flexibility to add or remove servers in the cluster as needed without bringing the cluster, enterprise applications, or DB2 offline. In addition, VERITAS clustering software is intelligent enough to provide load balancing and to dynamically determine the optimal target node in a failover scenario. In the event of a fault, VERITAS makes intelligent decisions regarding application failover by monitoring server capacity thresholds and available resources on alternate servers – a powerful method for ensuring that DB2 and users have access to limited computing resources.



Intelligent workload management

With their centralized architecture, clusters simplify the management of multiple servers and service groups. This allows administrators to consolidate highly available application environments onto a centrally managed platform, lowering administrative overhead and costs.

MANAGEABILITY

With availability solutions from VERITAS, you can improve the manageability of your DB2 environment by:

- Centralizing management of clustered environments
- Managing a rapidly growing DB2 environment while maintaining high levels of performance
- Reducing administration and training costs

VERITAS provides failover support for entire service groups and multiple enterprise applications and databases, including DB2. VERITAS has the ability to monitor DB2 and all of the hardware and software components associated with the DB2 server, ensuring automatic detection and recovery for all types of node failures that impact the availability of DB2. VERITAS also supports a manual fail-over mechanism to enable server maintenance or upgrades that only result in a momentary disruption in the data service as opposed to time-consuming disruption.

VERITAS Consulting Services has created multiple customizable agents that provide business continuity for core specialized and custom application services. These include a variety of ERP, CRM, and other enterprise applications including solutions from Siebel, SAP, and PeopleSoft. Continuity is ensured through comprehensive management of each application service and by performing indepth monitoring to ensure the cluster is fully aware of the state of each application service.

VERITAS clustering solutions are developed to provide support for UNIX and Windows platforms, as well as a wide range of heterogeneous hardware configurations — including servers, storage and SAN interconnect devices. Because VERITAS technology works with a wide range of hardware, servers and storage, it can be re-purposed easily as an organization changes and grows.

STRATEGIES TO MINIMIZE UNPLANNED DOWNTIME

While it is not possible to foresee the event that would cause unplanned downtime for an enterprise, it is possible to prepare, plan, and test scenarios to address unplanned downtime. Unplanned downtime can constitute a disaster situation for an enterprise. For those mission critical applications that must be recovered in minute or hours, it is essential to have not only a recovery plan, but also one that has been thoroughly tested and that can be executed at the touch of a button with little or no expertise.

In a disaster recovery scenario, enterprises must create plan protection and failover plans for mission critical application groups – not only from node to node in a local cluster, but from data center to remote data centers. A natural disaster or even a more benign rolling blackout may cause the complete shutdown of a company's data center.

Replication

The first step in extending availability beyond a single data center is ensuring that a duplicate copy of the data is available at a remote site. Replication accomplishes this.

Replication is an automated and rules-based method for the geographical distribution of identical data. This second data set is necessary in the event the data at the original location cannot be accessed. Replication transfers data in real-time over a wide area network (WAN) to the secondary location. This allows for a current copy of the data at a remote location should it be required.

Global Clustering

Data availability needs in today's information driven economy require all associated resource groups, applications, and accompanying data to cleanly fail over to another site in order to provide uninterrupted access to

the information managed at the affected site. These geographically dispersed take-over sites add yet another dimension to the enterprise's availability infrastructure.

Global Clustering extends the concept of local clustering to the wide area by providing automated application and resource migration from a primary data center to a remote geographically dispersed location or multiple geographically dispersed locations. Your critical DB2 database can now continue functioning even in the event of a site outage.

Global clustering management tools also allow IT teams to develop and test disaster recovery protocol. They provide the ability to centrally monitor and manage the entire clustered environment on an ongoing basis, and the analysis tools to determine whether global failover is indeed required.

RELIABILITY

The investment in a comprehensive high availability solution for your DB2 environment allows you to:

- Have peace of mind, knowing that you have predictable, pre-tested system actions in place in the event of a disaster
- Meet service level agreements more effectively
- Improve your decision making based on current, quality data

BACKUP – YOUR INSURANCE POLICY

Backup continues to be an important part of any availability strategy and is the most basic technique for protecting against database disaster. Backup protects your DB2 data center from logical errors caused by an administrator or an application program that can compromise the integrity of the database, as well as system component and environmental failures. Backup is the only technique that allows recovery of data from any point in time, thus protecting the corporation's data integrity. For this reason, backup remains the last line of defense against total data loss.

Some backup technologies can actually contribute to unnecessary server downtime. When selecting a backup solution, it is critical to identify a solution that allows you to conduct online backups or that can minimize the time required to execute off-line backups. This minimizes planned downtime associated with backup.

VERITAS backup technology interfaces with DB2 through a DB2 backup API that allows for the online "hot" backup of DB2 – delivering comprehensive data protection for DB2.

CONCLUSION

VERITAS Software provides solutions for a highly available DB2 environment. These solutions minimize the impact of planned and unplanned downtime on your enterprise data center and the mission critical applications it supports. VERITAS ensures a highly available environment for your enterprise while maintaining the levels of performance, manageability, and reliability that end users expect.

Visit www.veritas.com where you can learn more about VERITAS solutions for DB2 and find detailed product information.