

## **Maximizing Backup and Restore Performance of Large Databases**

## Forward (from Meta Group)

Most companies' critical data is being stored within relational databases. Over 90% of all mission critical systems, including financial, ERP and CRM systems store their data and maintain state within relational databases, i.e. DB2, Oracle, Sybase and SQL Server. The critical nature of these databases makes it necessary that IT organizations (ITOs) put in place high availability and backup and recovery processes and technologies to protect these systems and their data.

The dilemma that many companies are now facing is that the volume of data stored within critical database systems continues to grow exponentially while the allotted downtime for the backup and recovery process continues to shrink. Not only do the data volumes within existing business critical systems continue to grow, but also the number of databases that are now part of business and mission critical processes. Numerous factors, ranging from regulatory considerations to shifts in importance of analytical systems is contributing to this growth.

The overall Database Management System (DBMS) segment has evolved into one of higher business value, that used to only support On-line Transaction Processing (OLTP) systems, but now includes data warehousing and decision support. Most companies once viewed data warehouses and analytics as a "nice to have", now view them as a "must have" and in many cases a critical component of their overall operational processes. The underlying nature of data warehouses is one that lends itself to larger data sets contained within the DBMS and data sets that continue to grow in size over time. This overall shift in the importance of analytics is forcing ITOs to backup and recover extremely large databases at data volumes that were traditionally unheard of in the past (in the numerous terabyte range).

At the same time that companies are embracing the importance of decision support systems, government regulations are forcing many organizations to maintain data within their operational systems for longer periods of time. The data must not only be kept around for longer periods of time, but must also be protected from mishaps and disasters.

As ITOs face the problems of managing and protecting a growing number of extremely large databases, customer expectations surrounding the availability levels of the applications continues to grow. With the externalization of numerous business processes, customers now expect 24 by 7 uptimes or close to it. The traditional window for planned downtime, for database maintenance, backup and recovery operations, etc. is no longer numerous hours, and in many cases has moved to times ranging from 10 min to an hour per evening. This customer uptime expectation coupled with the rapid growth in the number of databases and the data volumes they contain is forcing many ITOs to take only partial backups and less frequent full backups. This process, brought on by the slow nature of numerous traditional backup products, as well as the complexity of their management, leaves many ITOs critical data exposed and vulnerable.

## Maximizing Large Database Backup and Restore Performance White Paper

The nature of the evolution of companies' database infrastructures means that this problem will only continue to get worse. By 2007-08, if backup and recovery solutions do not evolve to address this problem up to 50% of all companies critical data could be vulnerable, leaving many organizations legally and economically exposed.

**Bottom Line:** Companies must push existing back up and recovery vendors for solutions that enable them to easily back up extremely large amounts of data in shorter time frames.

## Study Overview

Database systems are among the most important and challenging environments to manage. Maintaining data availability, while simultaneously performing backup processing is not a trivial challenge. Even more importantly, when data must be restored, it must be done quickly. As data volumes grow and maintenance windows shrink the ability to rapidly backup and restore a large database is coming even more essential.

IBM and VERITAS recently conducted a joint study measuring the backup and restore performance of a large IBM DB2 database. The test was executed using the latest IBM and VERITAS technologies including: an IBM eServer pSeries 690 server, IBM DB2 Universal Database v8.1, the new IBM TotalStorage Enterprise Tape Drive 3592, the IBM TotalStorage Enterprise Storage Server Model 800 and VERITAS NetBackup™ 5.0 and the VERITAS NetBackup *for DB2* agent.

The goal of this exercise was to maximize the backup and restore performance of the environment as measured by the amount of DB2 data backed up and restored in period of time. Along the way, we learned a great deal about various tuning parameters available and some valuable lessons in maximizing large database backup and restore performance.

We found that it is possible to obtain high performance backup and restores of very large on-line databases! Using the new IBM 3592 tape drive technology, an pSeries server, ESS disk, DB2 UDB database software and VERITAS NetBackup, we were able to demonstrate a large database backup rate of over 2.5 TB/hour, and a restore rate, of the same database, of almost 1.7 TB/hour. This high level of performance was achieved by maximizing the degree of parallelism throughout the system infrastructure. This involved developing a database backup strategy and then utilizing the various hardware and software tuning parameters to optimize system throughput.

## Description of the Technical Environment

The tests were executed on a dedicated configuration that included the following:

Two IBM ESS Model 800's, each containing an 8x2GB Fibre channel connectors through a switch where the connections were concentrated down to 8x2GB Fibre Channel connectors in the p690. The tape drives were connected using 10x2GB Fibre channel adapters.

### Overall Layout

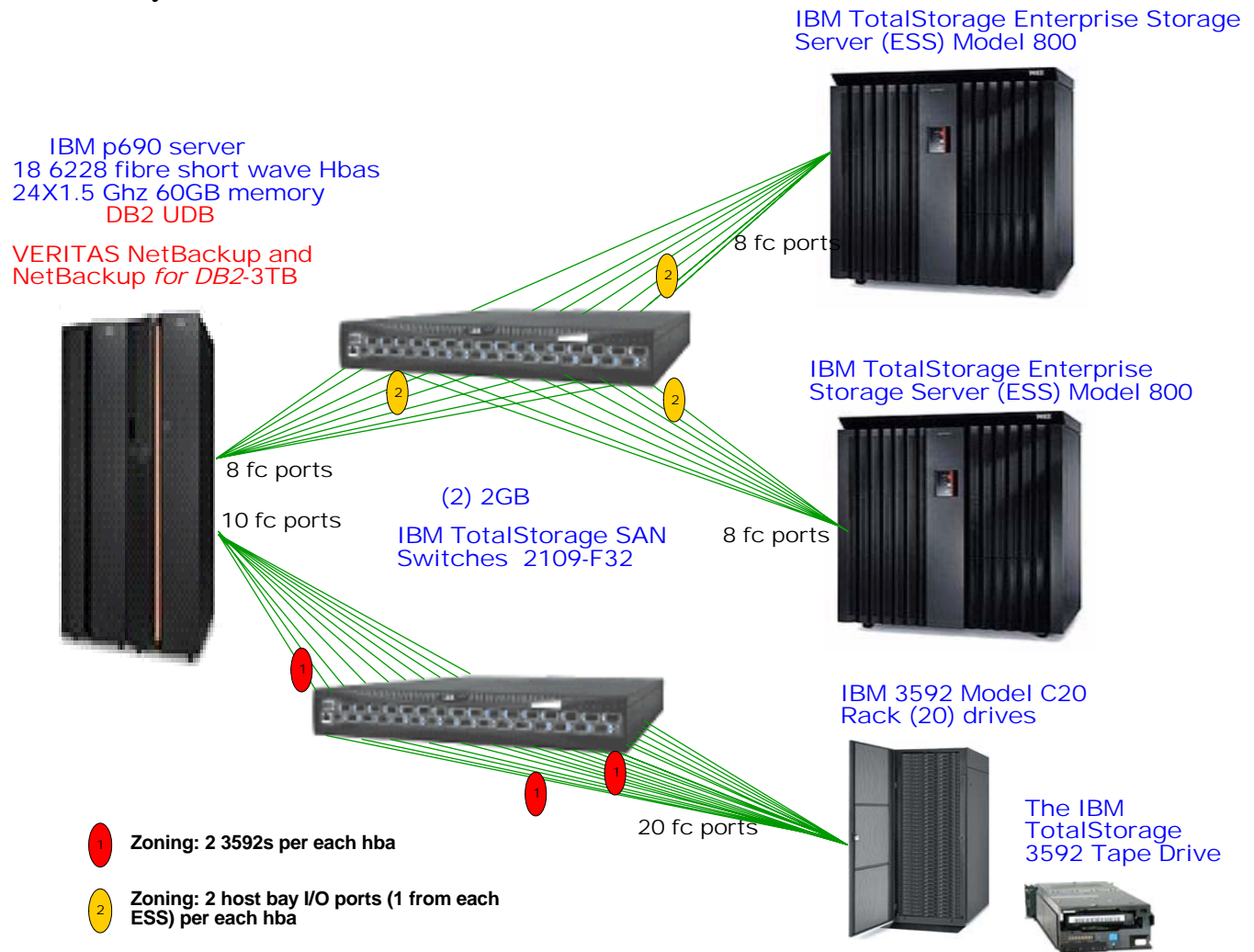


Figure 1

## Test Results

We measured the elapsed time associated with the backup of the large DB2 database. Just as importantly, we wanted to measure the elapsed time of restoring the same data.

The best results achieved with this configuration are shown in Table 1. The peak backup rate of 2.519 TB/Hr was achieved by backing up the database in about 57 ½ minutes for the 2.35 TB database. The associated restore of the same data set was completed at a rate of 2.13 TB/HR or 1hour and 26 minutes.

**Table 1**

Backup with 20 tape drivesRun #	Description	Run Time (H:M:S)	Amount of Data Transferred (TB)	Resulting Data Rate (TB/Hour)	Average Tape Drive Data Rate (KB/sec)	Peak Tape Drive Rate (KB/sec)
1	Online Backup	00:57:26	2.40	2.51	40956	46898
2	Full Restore	01:25:59	2.40	1.68	29045	41434

### Observations:

Greater throughput was possible! We were limited by the hardware on hand. We were able to sustain 980MB/sec doing sequential reads against the storage, server CPU, memory, SAN and tape bandwidth. In another case, we observed a single stream could occasionally achieve a sustained backup rate of 70MB/sec to a single 3592 tape drive. This equates to a total capacity of 1.4GB/sec tape I/O.

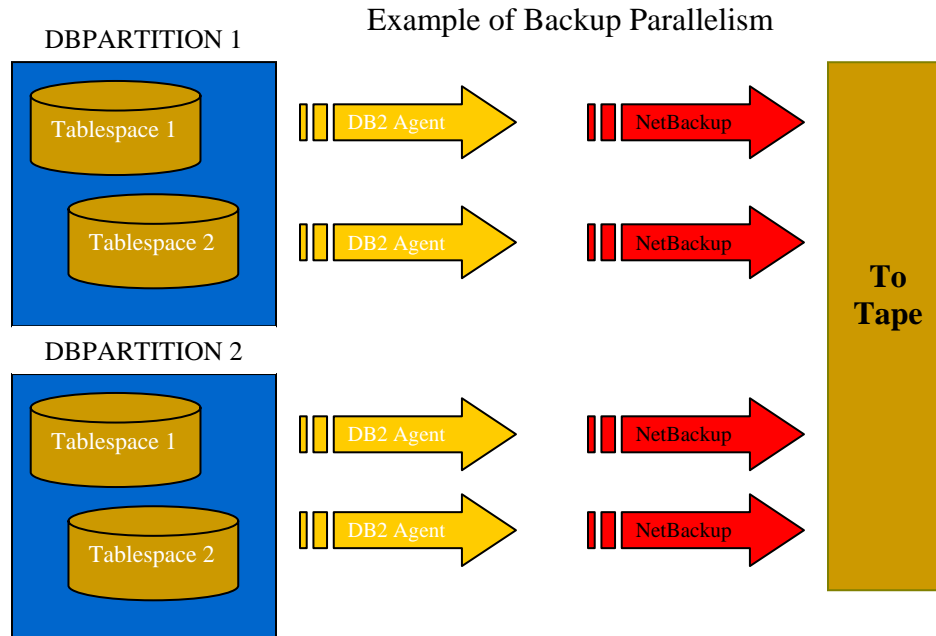
### ***Some lessons learned***

When designing a DB2 data warehouse it is important to consider the architecture of your backup scenario. The performance of a database backup is determined by many variables: data volume, tape speed, SAN/network speed, storage I/O rates and the database configuration. The following is a list of the top 3 suggestions for configuring your DB2 database for maximum backup performance.

1. Use Multiple Dbpartitions/Tablespaces
2. Create a separate catalog partition
3. Buffers, buffers, buffers

## Use Multiple DBpartitions/Tablespaces

When backing up a large database parallel operations are essential to performance. The number of tablespaces being backed up determines the parallelism required to perform a DB2 backup. DB2 backup, at most, supports one database agent per tablespace per dbpartition. VERITAS NetBackup can start one thread per DB2 backup agent. Therefore, to obtain the greatest degree of parallelism for a single table it should be spread across multiple dbpartitions. In this test the two largest tables were each 1.2TB in size and spread across 20 dbpartitions.



## Create a separate catalog partition

When backing up a DB2 database configured with multiple dbpartitions, DB2 backup starts with the catalog dbpartition first before starting to back up the other partitions. If the data are large on the catalog dbpartition this operation could greatly lengthen your total backup window. Placing no data on this dbpartition, except catalogs, allows this stage of the backup to be completed quickly allowing the parallel backup to begin much sooner. In this test the catalog backup took about 17 seconds, which had little impact on the overall throughput.

## Buffers, Buffers, Buffers

It is important to tune your buffer settings such that the backup device can keep streaming. The default of 1024 may work well for many situations but for this test we used 30 buffers with a buffer size of 4097 (4kb pages).

## **Configuring VERITAS NetBackup for Optimal Performance**

VERITAS NetBackup and the VERITAS NetBackup *for DB2* agent utilizes the DB2 backup API (including the “DB2 Backup” command). As long as the DB2 database is tuned as described in the paragraph above, only minimal tuning is required for NetBackup.

Care must be taken so that NetBackup has enough buffers to keep the data streaming. NetBackup processes use shared memory to move data between master and media server, when both the servers are on the same machine. During this test only 2 tunables were used: `NUMBER_DATA_BUFFERS`, set to 128 and `SIZE_DATA_BUFFERS`, set to 262,144 (256kb). After experimenting with various combinations, we found these settings to be optimum for our configuration.

### ***Summary***

Developing and implementing a backup strategy for large databases requires careful planning and design. But the rewards are clear - it is possible to very rapidly backup and restore very large databases, even while they are in use. The combination of a well laid out plan, with powerful, state of the art systems hardware can provide historic levels of performance for these critical systems.

## Appendices

### Detailed Hardware and Software Environment

<b>Component</b>	<b>Configuration</b>
System	IBM Server p690 24 processors 60 GB memory 1.5 GHz processor Speed 1 Logical Partition
Storage	2x IBM TotalStorage Enterprise Storage Servers Model 2105/800 2x IBM TotalStorage SAN Switch 2109-F32, 32-port 2GBs
Tape Devices	20x IBM TotalStorage Enterprise Tape Drive 3592 Model J1A
Software	AIX 5.2.0.0, Maintenance Level 2 DB2 Release level and Version: DB2 V8.1 FP4a VERITAS NetBackup 5.0 GA VERITAS NetBackup <i>for DB2</i> 5.0 GA agent for DB2 Universal DataBase
Database Configuration	20 dbpartitions 2 Tablespace Database Size 2,464,748 MB

## Maximizing Large Database Backup and Restore Performance White Paper

Copyright © 2004 by International Business Machines Corporation.

No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice.

This publication was developed for products and services offered in the United States. IBM may not offer the products, services or features discussed in this document in other countries. Information is subject to change without notice. Consult your local IBM representative for information on offerings available in your area.

The performance data concerning IBM products was obtained in a controlled, isolated environment. Many factors determine the sizing requirements and performance of a system's architecture. IBM assumes no liability for the methodology used for determining the configurations recommended in this document nor for the results it provides. Actual results that may be obtained in other operating environments may vary significantly. While IBM has reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere.

Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (*e.g.*, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein.

## Maximizing Large Database Backup and Restore Performance White Paper

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

Copyright © 2004 VERITAS Software Corporation. All rights reserved. VERITAS, VERITAS Software, the VERITAS logo, and all other VERITAS product names and slogans are trademarks or registered trademarks of VERITAS Software Corporation in the US and/or other countries. Other product names and/or slogans mentioned herein may be trademarks or registered trademarks of their respective companies. Specifications and product offerings subject to change without notice.

VERITAS Software Corporation  
Corporate Headquarters  
350 Ellis Street  
Mountain View, CA 94043  
U.S.A.  
650-527-8000 or 866-837-4827