
Database Application Replication and Subsetting

Abstract

This paper discusses techniques for replicating a database application when used in conjunction with EMC's DatabaseXtender Subsetter technology. A sample use case is reviewed, comparing and contrasting complimentary replication techniques.

October 2005

Copyright © 2005 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, EMC ControlCenter, AlphaStor, ApplicationXtender, Catalog Solution, Celerra, CentraStar, CLARAlert, CLARiiON, ClientPak, Connectrix, Co-StandbyServer, Dantz, Direct Matrix Architecture, DiskXtender, Documentum, EmailXtender, EmailXtract, HighRoad, Legato, Legato NetWorker, Navisphere, OpenScale, PowerPath, RepliStor, ResourcePak, Retrospect, Smarts, SnapShotServer, SnapView /IP, SRDF, Symmetrix, TimeFinder, VisualSAN, VSAM Assist, Xtender, Xtender Solutions, and where information lives are registered trademarks and EMC Developers Program, EMC OnCourse, EMC Proven, EMC Snap, EMC Storage Administrator, Access Logix, ArchiveXtender, Authentic Problems, Automated Resource Manager, AutoStart, AutoSwap, AVALONidm, C-Clip, Celerra Replicator, Centera, CLARevent, Codebook Correlation Technology, Common Information Model, CopyCross, CopyPoint, DatabaseXtender, Direct Matrix, DiskXtender 2000, EDM, E-Lab, EmailXaminer, Engenuity, eRoom, FarPoint, FLARE, FullTime, Graphic Visualization, InfoMover, Invista, MirrorView, NetWin, NetWorker, OnAlert, Powerlink, PowerSnap, RepliCare, SafeLine, SAN Advisor, SAN Copy, SAN Manager, SDMS, SnapImage, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, UltraPoint, Viewlets, VisualSRM, and WebXtender are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

Part Number H1520

Table of Contents

Executive Summary	4
Introduction to Application Replication	4
Replication Technologies Overview	5
Full-Volume Replicas	5
Pointer-Based Replicas	5
SAN Copy.....	5
Network-Based Copy	6
DatabaseXtender Subsetter Overview	6
Drop Object Policies	7
Tablespace Size and Datafile Policies.....	7
Move Tablespace Storage Policy	7
Rebuild Segment Policies	7
Sample Configuration	8
Planning and Implementation Considerations	8
Number of Clones Required	8
Performance: Read and Write Response Requirements	8
Percent Change	9
Summary	9

Executive Summary

IT organizations responsible for supporting database applications are typically required to make multiple copies, or replicas, of their production installations for backup, recovery, reporting, test, and development purposes. As the data volumes grow in the production database, the replication task becomes more challenging to manage. There are several methods for creating a replica of the application environment. EMC provides disk-based replication functionality for those customers who have their application residing on EMC storage arrays. When making the decision on what replication method to use, it is important to understand how the replica will be used. For example, the requirements for a replica that will be used as a backup and recovery instance are very different from the requirements of one that is used in a test and development environment.

In cases where the purpose of the replica requires a full-volume copy, as the data volumes grow, the amount of time it takes to make the replica increases. This limits the number of test cycles if the copies take too long. Taking advantage of array-based replication technologies, this process can be streamlined reducing the amount of time it takes to create the replica. In other cases, where the purpose of the replica does not require all of the application data, these copies, or instances, may be subset by removing unnecessary data for that particular purpose. Examples of these activities may be new application module development, testing of functionality for just one operating unit in a company (i.e., Finance), training, and reporting. The purpose of subsetting the copies is to reduce the storage footprint of the instance, lowering cost and streamlining test and support operations.

This paper begins by providing a brief discussion of complimentary replication technologies followed by an overview of EMC[®] DatabaseXtender[™] Subsetter software technology.

Introduction to Application Replication

Replicating a database application environment requires a coordinated effort among the infrastructure and database support teams. Storage administrators need to be involved to make sure enough disk is allocated and visible to the target hosts. Network administrators need to be involved to make sure enough network bandwidth will be available. System administrators need to be involved when mounting the replicated environment to the target host. Business application owners and database administrators (DBAs) need to be involved to make sure the application starts up properly in the new environment. Each step in the process may vary depending on the use of the replicated environment and the application itself.

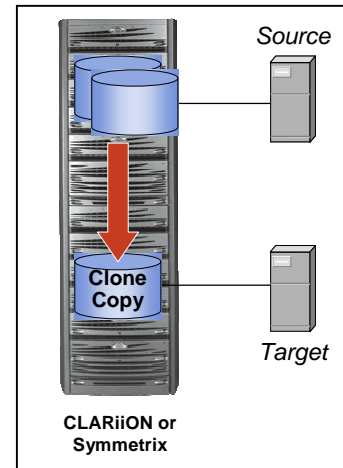
The process of creating a replica of an Oracle E-business Suite environment is called *cloning*. Once an identical copy of the database and corresponding application data files are replicated and can be presented to a secondary host, Oracle provides a set of cloning utility scripts, called RapidClone, that modifies the identical copy within the application and database, and points the replicated database and files to the secondary host. The process of actually copying the files has traditionally been done via FTP or RCP and sometimes requires taking the application database temporarily offline. Depending on the size of the application database and file system, using FTP to transfer the database files and application binaries can be as long as 20 hours or more over a local area network (LAN). If the application database and file system resides on a storage area network (SAN), disk-base replication technologies can be used to streamline the process, allowing the replication to take place while keeping the production application online and also reducing the amount of time required to make the copy.

Replication Technologies Overview

The Oracle E-Business Suite cloning process can be streamlined and automated when the replication process can be done within the same storage array or even within the same SAN. EMC offers two options for creating array-based replicas: full-volume replicas and pointer-based replicas. When array-based replication may not be the right solution, users can replicate over a network. If the target is on a separate array, EMC offers a block-level SAN Copy™ capability. EMC also offers a comprehensive solution for replicating data in a disaster recovery scenario; these configurations are not discussed in this paper.

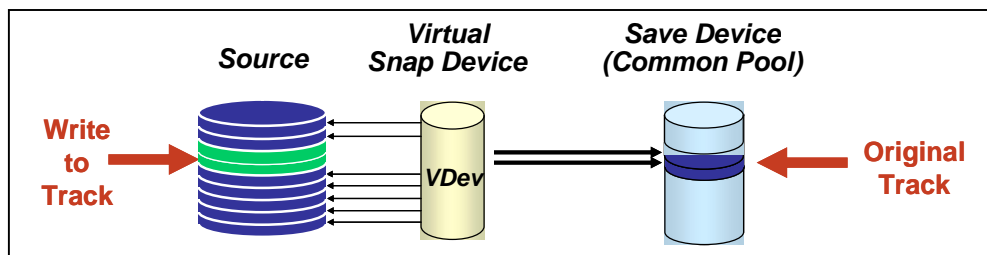
Full-Volume Replicas

EMC TimeFinder® business continuance volumes (BCVs) or CLARiiON® SnapView™ clones allow users to create full-volume replicas of production applications reducing the amount of time it takes to create a clone of an E-Business Suite implementation from hours to minutes. The BCV or clone requires the same amount of designated storage as the production instance, and is an identical block-for-block copy. If a client requires multiple application clones, the required storage that needs to be allocated for the clones is a linear multiplier. EMC's Replication Manager can be used to automate the entire process, eliminating the need for writing scripts manually. An EMC Proven Solution™ streamlines the replication process with Oracle's Rapidclone technology.



Pointer-Based Replicas

EMC TimeFinder snaps and SnapView snapshots replication technology also create pointer-based copies, or virtual replicas. These snaps, or snapshots, require less than 20 percent of disk space to make the copy and deploy copy-on-first-write technology. The primary benefit of using a snap, or snapshot, is that it requires less disk space to make a copy.



This replication technology is best suited for making point-in-time copies that are less write intensive, such as a source for backup or for recovery, and was designed for such purposes. The drawback for using this approach in a scenario where the database will be modified significantly and performance requirements are high, full volume copies are recommended. Write response times on pointer-based copies are typically three times as long as access directly to the source volumes. If the copy is going to be used for a test and development cycle that requires acceptable response times, the response times are typically three times that of production. The recommendation for using snaps is when the changes to the source or copy are less than 30 percent.

EMC SAN Copy

SAN-based copying can be done between classes of EMC storage arrays on the same SAN. In some cases, EMC customers may have their production instances residing on Symmetrix® arrays, but would like to create replicated environments on CLARiiON arrays, lowering total cost of ownership for their non-

production support replicas. EMC SAN Copy technology facilitates a block-for-block copy between the arrays. EMC's SAN Copy is also available with EMC's Replication Manager technology.

Network-Based Copy

Network-based copying can be done independent of the storage array, network topology, and replication technology. This is the method that has been traditionally used by the support DBAs because they are not required to have the storage or SAN configured in order to make a copy of a database. The files to be copied are selected and transferred via FTP or RCP to a target host across the network. This technique requires the most amount of network bandwidth, and typically has the longest duration when compared to the other options.

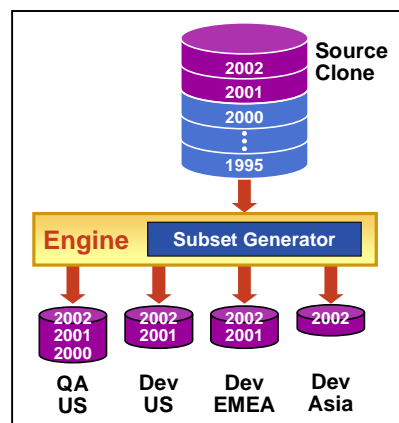
IT organizations that make several copies of their database applications for test and development purposes may recognize that as the production databases grow in size, the storage requirements for each copy increases. In many cases, the increased rate of the data growth is faster than that of the rate of budget increases. An alternative solution to this problem is application database subsetting. EMC's DatabaseXtender Subsetter can reduce the storage footprint of test and development environments, thereby shrinking the size of the database based on user-defined policies.

DatabaseXtender Subsetter Overview

DatabaseXtender Subsetter reduces the test and development instances by removing sets of transactional data based on predefined business application data subsetting policies. The resulting subset instance is a fully operational and functional application environment. Because the test and development instances are smaller, they require less server and storage resources. IT organizations that back up their test and development instances to tape realize significant cost savings based on these reduced datasets. More test and development instances can be created in the reclaimed disk space and replications can be completed faster, reducing the duration of each test cycle. By enabling more testing, confidence in the quality of the application improves dramatically. All of these operations can be automated, streamlining operations.

DatabaseXtender Subsetter uses policies for subsetting the transactional data for each individual application module in an Oracle E-Business Suite application (i.e., General Ledger, Accounts Payables, and Receivables). For example, users can define policies to subset aged sets of books within a General Ledger module of an Oracle E-Business Suite implementation. This means that just those transactions related to the aged set of books will be removed from the database and the General Ledger application continues to work as a fully functional application:

- **Age of data**
Users can create policies to keep transactions that are of a certain age or less when subsetting application data by the age of the transaction.
- **Application Module**
Users can create subsets by E-Business Suite Module, such as General Ledger, Inventory, or Accounts Payable.
- **Organization / Geography**
Policies can be defined to subset an entire organization, such as the US business units.



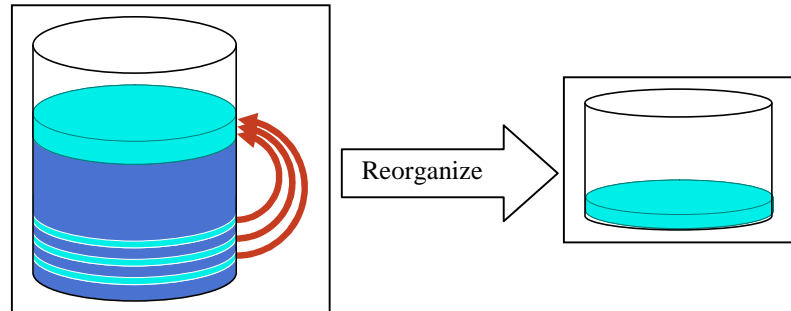
For each replicated instance, DatabaseXtender Subsetter can run the same or a different set of user-defined subsetting policies to reduce the instance copies either the same way each time or differently, based on the use of the copy. The user-defined policy also determines how the database should be reorganized and restructured from a database storage perspective. For example, the user can specify which tablespaces to drop and what percentage to shrink existing tablespaces.

Once the data has been selected and relocated and the unwanted objects are dropped, the resulting instance size is reduced. Examples of the types of storage reorganization policies are described next.

Drop Object Policies

Drop Object policies drop an object from the database. Use the Drop Object policies with caution, as dropped objects are not recoverable.

For example, use the Drop Object policy when a self-contained custom user not relevant in the subsetted database needs to be removed.



Tablespace Size and Datafile Policies

Tablespace size and datafile policies include Create/Alter Tablespace storage policy and Datafile storage policy.

Move Tablespace Storage Policy

The Move Tablespace storage policy moves all the objects from the specified source tablespace to target tablespace using the segment's source attributes or the attributes you specify in the Rebuild Segment policies. The Move Tablespace storage policy provides significant space reductions by consolidating several tablespaces into fewer, more compact tablespaces. The source tablespaces that are empty after moving out objects are dropped. If the source tablespace has any unsupported objects, they are left behind and the tablespace is made smaller.

Rebuild Segment Policies

You can use the following Rebuild Segment policies on their own or with the Move Tablespace storage policy:

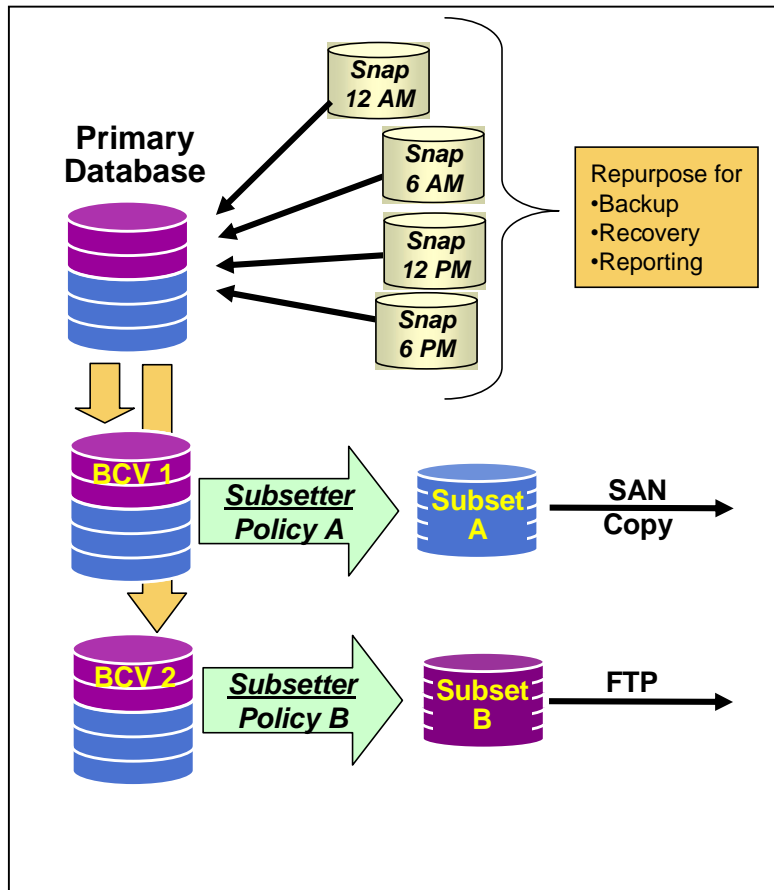
- Alter Table storage policy
- Alter Table Partition storage policy
- Rebuild Index storage policy
- Rebuild Index Partition storage policy
- Alter Table LOB storage policy

For more information on Subsetting policies, refer to the *DatabaseXtender Subsetter Operations Manual*.

Sample Configuration

The service level agreements for the test and development environments should drive the topology for each instance. In the figure at right, the assumption is that the production instance resides on a Symmetrix or CLARiiON array, with snap technology used to create replicas for the purpose of backup, recovery, and some reporting functions. These uses require less performance and will have a small percentage in change.

In this scenario, if the array is a Symmetrix, two BCVs are created to support test and development efforts. DatabaseXtender Subsetter is then applied to these clones residing on BCVs, reducing the physical size of the database on the BCVs. At this point, DBAs can be used directly, or can be copied to another lower cost environment by using a SAN Copy tool or by using FTP to transfer the files over the network. Because the database is now smaller, the time to complete the copy is significantly reduced. At this point, the BCVs could be refreshed back to a full-volume clone.



Planning and Implementation Considerations

When planning and implementing an application database subsetting process, there are several issues to consider before making your selection for storage replication technology. The following metrics on the test and development instances should help guide your decisions.

Number of Clones Required

On average, IT organizations make between eight and ten copies of production to support production and non-production activities. If this number exceeds four, then an automated subsetting technology, such as DatabaseXtender Subsetter should be considered.

Performance: Read and Write Response Requirements

Before deciding on which replication technology to use, it is a good idea to understand the targeted test and development community's response time requirements. For example, if a standard Oracle report takes 10 seconds to return in production, this same report could potentially take 30 seconds if the development instance was created using snap replication technology.

Percent Change

If the replication environment is intended for read only, then snap replication would be recommended to reduce the required storage to support that initiative. If deploying DatabaseXtender Subsetter to the instance, the percentage change from the storage perspective could be as much as 100 percent; full-volume replicas or file-based replication would be the preferred method. If the expected change is less than 30 percent, snaps could be used, but response time analysis would need to be considered.

Summary

When implementing an application database subsetting process, it is important to understand the purpose and requirements of each copy that is made. The requirements should be considered when selecting the appropriate subsetting and replication technology. EMC's DatabaseXtender Subsetter can be used with existing EMC disk-based replication technology, such as TimeFinder BCVs and CLARiiON SnapView clones. TimeFinder snaps and SnapView snapshots provide additional replication functionality for those copies that may be used for backup, recovery, and reporting purposes.

For more information about these and other EMC products and solutions, go to www.EMC.com