Sun StorEdge™ Instant Image 3.0 and Oracle8i Database Best Practices

Art Licht

Sun BluePrints™ OnLine—August 2002
Sun StorEdge Instant Image 3.0 and Oracle8i Database Best Practices

As databases grow in size and backup windows shrink, Oracle8i database customers need methods for performing faster and less intrusive backups and restores. Customers are also looking to maximize the use of their valuable data contained in these databases.

Current Methodology

Currently, system and database administrators clone an existing database to test their application. Cloning requires the existing database to be backed up to tape and restored to another location. A cold backup of the database requires the database be unavailable to production users for a significant period of time. Given the average size of databases and slow performance of tape technology, the backup could take hours. For this reason, cold backups are not frequently done because the data used for testing is stale.

One method used today is to perform hot backups. This method allows online activity to continue, but it requires the database to be in backup mode for hours. This method might also require various transactions to be rolled forward, and any uncommitted transactions must be rolled back. Additionally, online performance is affected during the backup. After the backup is completed and restored to the test area, it has to be recovered. Various transactions also have to be rolled back.

Products such as VERITAS NetBackup or Legato NetWorker are used to back up data. Traditionally, as the product names imply, NetBackup and NetWorker were designed to back up data across the network; however, these products can also be used with directly accessible storage.
Another method used today is to create a new database. The new database would be similar to the online database; however, it would contain different data. This has little impact on online activity. The downside to this method is that the application is being tested on a different database; thus, it is not real-world data.

New Methodology With the Instant Image 3.0 Software

Backups, reporting, data mining, and new application development can be accelerated by using point-in-time copy utilities to capture frequent snapshots of production data for use as input for the new applications or backup utility. This methodology offers the opportunity to keep critical online transactions going while satisfying the background use of the snapshot. The result is that development cycles can be accelerated, applications can be deployed more quickly, and backups done less intrusively.

Configuration Details

This paper discusses using the Sun StorEdge™ Instant Image 3.0 software to make a copy of an online database and to export that copy for application and version testing purposes. The following table contains a description of the tested environment.

<table>
<thead>
<tr>
<th>TABLE 1 Test Environment Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Database</strong></td>
</tr>
<tr>
<td>Oracle8i Database</td>
</tr>
</tbody>
</table>

Terminology

This section contains commonly used words and phrases in this document.
- **Master volume**—A physical disk device or volume that contains the original information for a particular application or suite of applications.

- **Shadow volume**—A physical disk device or volume that contains the changed point-in-time data or full image of the original master volume.

- **Bitmap**—A file, physical disk partition, or volume used by the Instant Image software to maintain the current state of the data between the master and shadow volumes.

- **Independent copy**—An identical copy of the point-in-time master volume (that is, a copy of the master volume when the image was taken) that is stored on the shadow volume. After it is established, an independent Instant Image operation starts a block copy operation from the master volume to the shadow volume as a background process. While the copy is in progress, the Instant Image software treats accesses to the shadow volume as if the shadow volume is a dependent image. An independent copy can be exported to another Solaris Operating Environment (Solaris OE) server.

- **Dependent copy**—A point-in-time view of the master volume. After it is established, a dependent copy is immediately available to the backup application. The shadow volume maintains the original data from modified blocks on the master volume since the copy was established. All access to blocks not changed after the point-in-time copy is taken come directly from the master volume. The shadow volume is accessed only when the bitmap indicates that the master volume has been changed for the accessed data block. A compressed dependent copy has the advantage of not requiring a duplicate amount of storage. A dependent copy must remain within the server.

- **Volume Pair**—A volume pair consists of the master volume and its related shadow volume. Each volume pair is supported by one bitmap. You can create as many volume pairs as necessary. The Instant Image software is limited only by the amount of available storage.

---

**General Operational Issues**

Operational procedures, automated or otherwise, need to be established to employ the point-in-time copy technology. Deciding when the snapshots will occur, the frequency in which they will occur, and the method of notifying the secondary application when the snapshot is available are examples of operational issues that need to be addressed.

In most cases, a point-of-data consistency needs to be reached before a point-in-time can be established. This ensures that the snapshot image is logically consistent (that is, you have not captured any incomplete transactions).
To Create a Point-in-Time Copy

1. Ensure that the master data has reached a point of consistency (that is, all of the data has been flushed to the disk).
2. Create a point-in-time copy of the data.
3. Export the snapshot to a second server.
4. Run the test application against the point-in-time copy.
5. Rejoin and resynchronize the point-in-time copy with the master data, if necessary.

Procedural Details

This section describes a method for making a point-in-time copy of the online database with minimal impact to current online transactions. Before getting started, make sure you are logged in as superuser. The following steps assume that the Instant Image 3.0 software is installed on all of the computers that have access to either the master and/or point-in-time copy.

1. Calculate the maximum amount of raw volume to be duplicated.

   For an independent copy, you need a duplicate amount of usable storage to store the point-in-time copy. Take your existing disk group and volumes and duplicate them to another disk group. The online or master volumes can be any RAID level supported by the Solaris OE or volume manager. The point-in-time or shadow volumes do not have to be the same RAID level as the master. The independent shadow volumes must be at least as large as the master volumes.

2. Make the master and shadow volumes available to the Instant Image software.

   Any volume that is available to the Solaris OE is eligible to be used with the Instant Image software. If the shadow volumes are to be exported to a second server, the shadow volumes should be in their own disk group. This enables the shadow disk group to be exported by VxVM as a single entity.

   a. Create two disk groups (one for the online production volumes and another for the snapshots).

      The master volumes are the volumes where the production data files, control files, redo logs, and archived redo logs reside.
b. Create the bitmap volumes.

For simplicity, you should create the bitmap volumes within the master disk group and create the secondary bitmaps within the shadow disk group. This enables the shadow volumes and the secondary bitmaps to be moved as a single entity.

3. Allocate space for the Instant Image software to keep track of the differences between the master and the shadow volumes.

The Instant Image software requires 8 Kbytes, plus 8 Kbytes for each gigabyte of raw disk space, with a minimum of 24 Kbytes. One bitmap is required for each Instant Image volume group (a group contains one master and one shadow volume). It is recommended that raw volumes be used for the bitmap volumes. In this procedure, the following bitmaps are used:

```
/dev/vx/rdsks/masterdg/warehouse.bmp
/dev/vx/rdsks/masterdg/orders.bmp
/dev/vx/rdsks/masterdg/stocks.bmp
/dev/vx/rdsks/masterdg/orders.bmp
/dev/vx/rdsks/masterdg/neww_order.bmp
/dev/vx/rdsks/masterdg/orderdata.bmp
```

Similar bitmap volumes should be created in the shadow disk group. These bitmaps are used on the secondary server.

4. Put the database and/or file system into a known state.

The simplest way to do this is to shut down the database. Alternatively, the database can be put into backup mode for a few seconds. The method varies based on the database vendor.

For the Oracle8i database, you should put the database into backup mode. For the Informix database, you should use the `onmode -k` command.

For UFS file systems, you should use the `lockfs(1M)` command. For example, the `lockfs -w /mnt_point` command flushes the cache and suspends future writes. The `lockfs -u /mnt_point` command clears the write lock. Without options, the `lockfs(1M)` command displays the existing locks. Refer to `lockfs(1M)` for more information. Depending on how the database was set up, you might have to flush the file system to cache.
In the Oracle8i database test, the database was put into backup mode by using the following SQL commands:

```
SQLplus system/manager <<!
SELECT tablespace_name, file_name
FROM sys.dba_data_files;
!
svrmgrl systems/manager<<!
Connect internal;
ALTER TABLESPACE WAREHOUSE BEGIN BACKUP;
ALTER TABLESPACE ORDERS BEGIN BACKUP;
ALTER TABLESPACE STOCK BEGIN BACKUP;
ALTER TABLESPACE CUSTOMERS BEGIN BACKUP;
ALTER TABLESPACE NEW_ORDER BEGIN BACKUP;
!
```

For UFS, if you need to flush the file system cache to the storage arrays, use the `lockfs(1M)` command, as in the following example:

```
# lockfs -w /oradata
# lockfs -w /warehouse
# lockfs -w /orders
# lockfs -w /customers
# lockfs -w /stock
# lockfs -w /new_order
```

If the file system or the application is set to direct I/O, the above steps are not necessary, nor desired.

5. Start the point-in-time copy operation.
For managability, all of the volumes that belong to the database should be grouped in a Instant Image I/O group; however, this is not required. The following commands create an I/O group called OracleDB1 and insert each master and shadow volume set into that group.

```
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/warehouse
/dev/vx/rds/shadowdg/warehouse
/dev/vx/rds/masterdg/warehouse.bmp
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/orders
/dev/vx/rds/shadowdg/orders /dev/vx/rds/masterdg/orders.bmp
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/stock
/dev/vx/rds/shadowdg/stock /dev/vx/rds/masterdg/stock.bmp
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/customers
/dev/vx/rds/shadowdg/customers
/dev/vx/rds/masterdg/customers.bmp
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/new_order
/dev/vx/rds/shadowdg/new_order
/dev/vx/rds/masterdg/new_order.bmp
# iiadm -g OracleDB1 -e ind /dev/vx/rds/masterdg/oradata
/dev/vx/rds/shadowdg/oradata /dev/vx/rds/masterdg/oradata.bmp
```

You can now manage the database snapshot as a single entity.

In previous releases, you had to insert the volumes into the storage volume software. This is no longer a needed step because the `iiadm -e` command performs all of the needed storage volume insertions.

6. Copy any necessary database ASCII files.

You should make a duplicate copy of all of the needed database ASCII files (for example, `init.ora` or `config.ora`). One copy is for the master, and one is for the shadows.

```
# cp /Oracle/db1/cntlfile1 /Oracle/shadow/cntfile1
# cp /Oracle/db1/cntfile2 /Oracle/shadow/cntfile2
```

7. Bring the database online.

The original database can now be brought back online, and transactions can resume. It should only have been shut down, placed in backup mode, or locked for less than a minute.
For the hot backup, use the following commands to resume normal file system operation if it was previously locked:

```bash
# lockfs -u /oradata
# lockfs -u /warehouse
# lockfs -u /orders
# lockfs -u /customers
# lockfs -u /stock
# lockfs -u /new_order
```

If you placed the database in backup mode, use the following commands to bring the database back online:

```bash
svmgrl systems/manager<<!
Connect internal;
ALTER TABLESPACE WAREHOUSE END BACKUP;
ALTER TABLESPACE ORDERS END BACKUP;
ALTER TABLESPACE STOCK END BACKUP;
ALTER TABLESPACE CUSTOMERS END BACKUP;
ALTER TABLESPACE NEW_ORDER END BACKUP;
Archive log list;
```

All of the above commands should be placed in a script. The above commands need to be done only once. After it is completed, the shadow database then can be exported to another server. For future updates of the shadow, you would import the shadow then update it with the changes from the master. The database then can be exported.

8. Export the shadow to another server.

After the initial copy has been completed (only done once), you can export the shadow volumes to a secondary server. You can monitor the progress by using the `iiadm` command. If you used scripts, the `iiadm -w OracleDB1` command enables the script to wait until the copy is completely established before it is exported. Use the following commands to export the shadow volume:

```bash
# cp /Oracle/db1/cntfile2 /Oracle/shadow/cntfile2
# iiadm -g OracleDB1 -E
# svadm -d shadow_volumes
# vxdg -g shadowdg stopall
# vxdg -g shadowdg deport (This is necessary only if you are using VxVM.)
```
You must remove the volumes from the storage volume software if the volumes are going to be deported. This is volume manager specific. If you are not using a volume manager, you may not have to remove the volumes from the storage volume software. Logical volume managers that require volumes to be deported should have no open processes.

9. Import the shadow disk group and snapshot volumes to the second server, and bring up the shadow database.

Use the following commands to import the shadow disk group and snapshot:

```bash
# vxdg -g shadowdg import
# vxdg -g shadowdg startall
# iiadm -I /dev/vx/rdsk/shadowdg/warehouse /dev/vx/rdsk/shadowdg/warehouse.bmp
# iiadm -I /dev/vx/rdsk/shadowdg/orders /dev/vx/rdsk/shadowdg/orders.bmp
# iiadm -I /dev/vx/rdsk/shadowdg/stock /dev/vx/rdsk/shadowdg/stock.bmp
# iiadm -I /dev/vx/rdsk/shadowdg/customers /dev/vx/rdsk/shadowdg/customers.bmp
# iiadm -I /dev/vx/rdsk/shadowdg/new_order /dev/vx/rdsk/shadowdg/new_order.bmp
# iiadm -I /dev/vx/rdsk/shadowdg/oradata /dev/vx/rdsk/shadowdg/oradata.bmp
# svadm –e shadow_volumes
```

The next step is to mount the file systems, if any, that reside on the above volumes. The snapshot volumes can be put back into the Instant Image I/O group, if desired, by using the following command:

```bash
# iiadm -g OracleDB1 -m shadow_1, shadow_2, shadow_n
```

It is necessary to insert the volumes into the storage volume software to track all of the changes that occur to the shadow volumes while they are being used by the secondary server.

Modify any scripts used to bring up the database to use the new control files and the path names of the new table spaces. You can do this with links. Alternatively, the shadow volumes can be mounted on the second server with the same mount point names. Script modification should not be required.
To bring the shadow database online, use the following commands:

```sql
Srvmgrl system/manager <<!
CONNECT INTERNAL;
STARTUP NOMOUNT;
CREATE CONTROLFILE REUSE DATABASE "II" RESETLOGS
MAXLOGFILES 15
MAXLOGMEMBERS 2
MAXDATAFILES 30
MAXINSTANCES 1
MAXLOGHISTORY 7196
LOGFILE
  GROUP1 '/oradata/oradata/ii/redo02.log' SIZE 1024K,
  GROUP2 '/oradata/oadata/ii/redo01.log' SIZE 1024K
DATAFILE
  '/warehouse/warehouse.dbf'
  '/stock/stock.dbf'
  '/new_order/new_order.dbf'
  '/customers/customers.dbf'
  '/order/order.dbf'
  CHARACTER SET US7ASCII
RECOVER DATABASE USING BACKUP CONTROLFILE;
ALTER DATABASE OPEN RESETLOGS;
```

The recover command is required if any of the files are restored backups or if the last shutdown was not normal or immediate. The following statement assumes that the data files, the archive log files, and the redo logs are consistent (that is, they are part of the same Instant Image snapshot and/or group).

```
RECOVER DATABASE USING BACKUP CONTROLFILE;
```

You can now open the database, which also resets the redo logs, by using the following command:

```
ALTER DATABASE OPEN RESETLOGS;
```

The `RESETLOGS` option is required after an incomplete media recovery or, as in this case, recovery using a backup control file. This causes any information that was not previously applied to be discarded, ensuring that it will never be applied. It also re-initializes the control file information for the online redo logs and redo threads, and it clears the contents of the online redo logs and resets the log sequence number to 1.

10. Re-import and resynchronize the volumes.
The Instant Image software logs the changes that occur on the master volumes and the tracks on the shadow volume. When you are ready to update, import the bitmap from the second server back to where the master database resides. The Instant Image software compares the bitmaps to determine which blocks need to be updated. You do not need to recopy the entire database to update the shadow volumes. Only the modified tracks need to be copied, along with a current control file.

a. Ready the snapshot to be brought back to the master on the second server by stopping the application and removing the volumes from storage volume software.

b. Export the volumes, as in Step 8.
   - To reattach the snapshot to the master, use the following commands:

```
# vxdg -g shadowdg import
# vxdg -g shadowdg startall
# iiadm -J /dev/vx/rdsk/shadowdg/warehouse /dev/vx/rdsk/shadowdg/warehouse.bmp
# iiadm -J /dev/vx/rdsk/shadowdg/orders /dev/vx/rdsk/shadowdg/orders.bmp
# iiadm -J /dev/vx/rdsk/shadowdg/stock /dev/vx/rdsk/shadowdg/stock.bmp
# iiadm -J /dev/vx/rdsk/shadowdg/customers /dev/vx/rdsk/shadowdg/customers.bmp
# iiadm -J /dev/vx/rdsk/shadowdg/new_order /dev/vx/rdsk/shadowdg/new_order.bmp
# iiadm -J /dev/vx/rdsk/shadowdg/oradata /dev/vx/rdsk/shadowdg/oradata.bmp
# svadm –e shadow_volumes
```

The above commands take the changes done on the second server, recorded in the bitmaps volumes, and merge them with the changes done on the master (production) server.
   - To resynchronize the volumes, use the following steps:
      1. Put the database into a known state.
      1. Copy the needed ASCII files (such as control or log files).
c. Execute the Instant Image software commands to update the shadow volumes, as in the following example:

```
# iiadm -g OracleDB1 -u s
```

or

```
# iiadm -u s /dev/vx/rdsk/shadowdg/warehouse
# iiadm -u s /dev/vx/rdsk/shadowdg/orders
# iiadm -u s /dev/vx/rdsk/shadowdg/stock
# iiadm -u s /dev/vx/rdsk/shadowdg/customers
# iiadm -u s /dev/vx/rdsk/shadowdg/new_order
```

If you use a script to perform these commands, the Instant Image software waits for the updates to complete before it allows any other commands to be executed. Use the `iiadm -w` command after the script is completed. You can repeat this process as often as needed.

d. Apply the same scheme to applications that are on file systems because the Instant Image software works at the raw device level. Therefore, if the application resides on a file system, use the following steps:

i. Quiesce the application.

ii. Flush the page cache, if necessary.

   For example, use `lockfs -w /mount_point` on UFS file systems.

e. Establish and/or update the point-in-time copies.

i. Unlock the file systems, if previously locked, by using the following command:

```
# lockfs -u /mount_point
```

ii. Mount and export the shadow database.

   It is not necessary to stop the application or unmount the production file systems. If you add table spaces, file systems, or raw volumes in the future, follow the steps described in this document to enable the Instant Image software.

You must modify the proceeding examples for each database layout. For repetition and ease-of-use, you should put the sample commands used in this document into a script. It is highly recommended that error checking be inserted as well. If an error should occur, the script should stop, and the error should be addressed.
Related Documentation

iiadm(1M)
lockfs(1M)

Sun StorEdge Instant Image Release Notes, 805-7953
Sun StorEdge Instant Image System Administrators Guide, 805-7951
VERITAS Volume Manager User Manual

About the Author

Art Licht is a Storage Technology Engineer for Sun Microsystems, Inc. Art works with customers to build storage and SAN architectures, as well as architecting storage solutions such as disaster recovery, data-sharing applications, and backup and restore. Previously, Art worked in engineering on various storage hardware and software products.

Ordering Sun Documents

The SunDocsSM program provides more than 250 manuals from Sun Microsystems, Inc. If you live in the United States, Canada, Europe, or Japan, you can purchase documentation sets or individual manuals through this program.

Accessing Sun Documentation Online

The docs.sun.com web site enables you to access Sun technical documentation online. You can browse the docs.sun.com archive or search for a specific book title or subject. The URL is http://docs.sun.com/

To reference Sun BluePrints OnLine articles, visit the Sun BluePrints OnLine Web site at: http://www.sun.com/blueprints/online.html