the Availability Digest

Unix Backup and Recovery

February 2007

Backing up is a pain. But it is the *restore* that counts.

This is the message that Curtis Preston delivers in his book, *Unix Backup and Recovery*.[1] Preston has been involved in backup and recovery for much of his professional career. In his book, aimed at heavy-duty business Unix systems and the databases they run, he passes on all of the knowledge that he wishes he had when he first started out as a System Administrator.

This book reviews in detail both commercial and freely-available file system and database backup and recovery utilities. It is applicable to the small shop with no money to spend and to large shops with hundreds of servers. It provides full examples of the use of each utility, with significant effort spent on the nuances of the syntax of each.

Every user dreads the inevitable system crash. The consequences of downtime and data loss range from inconvenient to catastrophic. Minimizing this impact requires serious planning and testing of backup and recovery procedures before catastrophe hits. This book is an invaluable reference for proper planning of these procedures.

Preston sprinkles his book with vignettes of actual recovery horror stories experienced by him and his cohorts. These stories are as entertaining as they are educational.

Planning

The author begins his book with an in-depth discussion of planning the backup and recovery process

The Disaster Recovery Plan

The place to start backup and recovery planning is the Disaster Recovery Plan. The elements of this plan should include:

- Decide what is an acceptable loss. Backup is like an insurance policy. You want to pay for the coverage you need but no more. Some data may be so critical to corporate survival that cost is not an issue. The loss of other data may be tolerable but will have a cost associated with its loss – a cost that should not be exceeded by overzealous backup policies.

[1] W. Curtis Preston, *Unix Backup and Recovery*, O'Reilly Media, Inc.; 1999.

- Back up everything. You never know what you might need. Not only should files and databases be backed up, but so should scripts, metadata, operating systems, and the instructions needed to get them back. In this regard, *exclude* lists are good, *include* lists are bad. It is too easy to forget to add something to an include list. This can result in the failure to ever back up that item.

- Organize everything. Mark every backup volume (e.g., tape) with a unique identifier. Keep an up-to-data inventory of the exact location of every file, database, and other object. This catalog will change with every backup. Keep this catalog not only online but also in a place safe from disaster.

- Protect against all disasters. There are disasters other than natural disasters, such as user error, system-staff error, hardware failure, disk drive failure, software failure, system-wide failure, electronic vandalism and theft, and the loss of archives. Each of these must be protected against.

- Document what you have done. The recovery plan should be documented so that any qualified person can follow it after an outage or a disaster. Don't count on the availability of certain staff members following an unplanned outage.

- Test, test, test. A Disaster Recovery Plan that has not been thoroughly tested is not a plan; it is a proposal. Make sure that the test covers every facet of the recovery process.[2]

The Backup and Recovery Plan

Once you have completed the Disaster Recovery Plan, you have a good feel for how important each data object is to your operation. Now is the time to decide exactly how you are going to back up different classes of data and how you will restore them following a loss. Issues to consider include:

- Decide what to back up. This is a direct result of the first step in the Disaster Recovery Plan – deciding what an acceptable loss is. It is a good idea to also include the details of the system hardware and software configurations in the event that one or more servers must be replaced.

- Decide when to back up. For each class of data to be backed up, when should full backups be made? Incremental backups? Snapshots?

- Decide how to back up. Is the intent to implement a high-availability solution to maintain user services consistently, a disaster-recovery solution to recover from a total system failure, or both? Choose the backup utilities to satisfy the protection and cost parameters for each type of data. Automate the backup procedures as much as possible within cost constraints. Plan for expansion – it is inevitable.

- Storing the backups. What will be used for onsite storage of backup media? Will backup volumes be sent to offsite storage? Will backup data sets be replicated to an offsite system?

- Test the backups. Don't wait for the disaster which requires that you restore data or the system only to find out that your backups don't work. Test everything often. This is probably the most important part of the Backup Plan.

[2] See this issue's (February, 2007) Never Again article, entitled Don't Wait for the Other Shoe to Drop, for a real-life story of what can happen if you only think that you have adequately tested your recovery plan.

- Monitor the backups. Make sure that someone other than those responsible for backup is assigned the task to review the backup logs.

- Follow proper development procedures. New backup procedures are as prone to errors as are new application programs. Test changes to these procedures thoroughly on a development system. Roll them out cautiously. Be prepared to revert to the previous backup procedure in case of problems.

Everything You Need to Know About Backup and Recovery Utilities

The meat of *Unix Backup and Recovery* follows the above discussion of proper planning. The book focuses on the most popular backup and recovery utilities. For each, the detailed syntax is explained, including discussions of the syntactical nuances between different versions of the same utility.

Native Backup and Recovery Utilities

These are the utilities found in a standard Unix distribution.

dump and *restore* are flexible and versatile commands with a simple interface.
cpio is an archive utility that stores backups contiguously on tape.
tar, probably the most popular utility of this type, can be used to selectively back up files.
dd copies raw bytes from a source to a destination. It has no knowledge of file structures.

Some of these utilities, such as *cpio* and *tar*, come as open source GNU distributions. These distributions are superior in many ways to the standard Unix distributions.

Free Backup Utilities

An overview is given of several free backup utilities, including:

hostdump.sh, a plug-and-play backup script.
infback.sh, an Informix backup utility.
oraback.sh, an Oracle backup utility.
syback.sh, a Sybase backup utility
star, a very fast implementation of *tar*.
SysAudit, a system configuration backup utility.
SysInfo, another system configuration backup utility.
queso, a program to determine the operating system used by a server.
nmap, a network probing tool.

AMANDA, the Advanced Maryland Automated Network Disk Archiver, is described in great detail. It is a public domain utility developed at the University of Maryland. AMANDA is easily the most popular free backup utility of its type.

Commercial Backup and Recovery Utilities

Author Curtis Preston does not review specific commercial offerings since there are many and since they change frequently. Instead, he suggests attributes to look for in a commercial product. When evaluating a backup and recovery product, questions to ask include:

- Does the product fully support your platforms?
- Does the product meet special needs such as raw partitions and very large files?

- Can it simultaneously back up many clients to one drive or one client to many drives?
- Does the product provide storage management?
- Is the product's backup format unique or standard?
- What network load does the product impose?
- How secure is the product? Does it provide encryption?
- How easy is the product to administer? Is it automated?
- How robust is the product? How well does it deal with backup problems?
- How easily does the product perform recoveries?
- How well does it protect the backup catalog?
- Can the product verify its backups?
- What does it cost?
- Is the vendor reliable and has good references?

Commercial High-Availability Solutions

This discussion focuses on cluster technology for achieving high availability.

As with commercial backup and recovery products, specific commercial high-availability products are not evaluated. Described instead are attributes that should be considered in the choice of a product.

Attributes include:

- the number of servers that can be clustered together.
- the load-balancing capability of the product.
- application recovery, including the applications that have been tested on the cluster.
- cluster monitoring facilities.
- application monitoring facilities.
- cost.
- customer support.

Bare-Metal Backup and Recovery

Bare-metal recovery is the case in which the system administrator has nothing on the system with which to start. The causes of such a failure range from the loss of the system root disk to a complete meltbdown of the system in a fire. The first requirement of bare-metal recovery is to recover the root disk from some sort of backup.

The usual solution to bare-metal recovery is to

- replace the defective boot disk (or the entire system if it has been destroyed).
- reinstall the operating system and its patches.
- reinstall the backup software.
- recover the current, backed-up operating system, overwriting the old operating system.

However, the author points out that there is a better way which does not require the operating system be installed twice. This procedure involves some preparation:

- Back up all appropriate metadata.
- Back up the operating system.

Should the root disk be lost, recover as follows:

- Boot the system into single-user mode with a CD-ROM.
- Set up the recovery disk to look the same as the old root disk and mount it.
- Recover the operating system to the mounted disk.
- Place the boot block on the mounted disk.
- Reboot.

Bare-metal recovery procedures are then detailed for several Unix operating systems, including:

- SunOS/Solaris
- Linux
- HP Tru64
- HP-UX
- IRIX
- AIX

Database Backup and Recovery

Database backup and recovery presents some of the greatest challenges to system administrators and to database administrators (DBAs) alike.

Preston has seen a serious disconnect between system administrators, responsible for backing up and restoring the file systems and other data objects, and DBAs, responsible for backing up and restoring databases. The technologies are substantially different, and one group typically does not have a deep understanding of what the other group does.

He attempts to bridge this gap in the book. The information described up to this point certainly gives DBAs a good feel for what a system administrator does. He now launches into some detail concerning the technology of databases and the DBAs' responsibility.

He then goes into great detail (50 to 75 pages each) on three database backup and recovery facilities:

- Informix backup and recovery using *ontape* and the newer *onbar*.
- Oracle backup and recovery using EBU or RMAN.
- Sybase backup and recovery using the Backup Server utility.

Detailed recovery flowcharts are provided for each facility.

Potpourri

The book concludes with commentary on several backup and recovery topics. These include ClearCase (an IBM revision control tool), choosing backup hardware, making transactionally-consistent backup copies of volatile data, and the applicability of Gigabyte Ethernet to the backup process.

Kudos

Unix Backup and Recovery has received many exceptionally complimentary comments from the backup and recovery community. Some samples:

> "Beyond good, this book is almost 'godlike'."
> "This is THE backup book."

> “If you think that you know enough about backups, WRONG. Read it and you will change your mind.”
> “Definitely the best backup book I’ve read.”
> “If you’re a sysadmin, buy it. It’ll serve you well for the rest of your career.”
> “It is fun reading and full of experience and cases.”
> “This is one book I would not like my competitors to have.”
> “This is a must read for any administrator that wants to do things right.”
> “It steered me clear of several bad choices.”

Based on these comments, this could be the book that saves your job – or your company.