CHAPTER 1

Managing a Backup and Restore Initiative

Sun and the Evolving Datacenter Model

Organizations today are reaching beyond the traditional glass house boundaries to interact directly with customers, partners, suppliers, and employees. In this new extended enterprise model, organizations seek to provide universal access to information and services via business portals on the World Wide Web. This shift is born out of the pressure of global competition in a networked world, and places new availability, manageability and performance requirements on a datacenter. Today's IT priorities include: improving business processes, improving customer data quality, developing better customer relationship management, implementing high availability networks, boosting networking bandwidth, and building e-commerce infrastructures.

In this evolving business paradigm, new data is generated at exponential rates, and plays a more vital and publicly visible role than ever before. Consequently, organizations must place a high priority on safeguarding mission-critical data and ensure it remains continuously available. A reliable, flexible, and highly-scalable network-based backup and restore solution is an essential part of this endeavor.

Sun Microsystems, Inc. supports the new business model with enterprise servers, storage products, networking hardware, and related software and services, including leading backup and restore solutions. For up-to-date information on Sun datacenter initiatives, see `http://www.sun.com/datacenter`. For information on services available from Sun, including professional consulting services, see `http://www.sun.com/service`.

Sun offers a line of enterprise servers, from the Sun Enterprise Ultra™ 5S up to the mainframe-class Sun Enterprise™ 10000, also known as the Starfire™ server. At the time of this writing, Sun has shipped well over 1000 Starfire servers, which are now operating in 46 countries worldwide.

FIGURE 1-1 Today, many datacenters are powered entirely by Sun Enterprise servers. Starfire servers are deployed in the datacenter shown above.

High availability is a major component of Sun datacenter initiatives. SunUP™ is a collaborative program between Sun, customers, and third parties to analyze, develop, implement and manage services, infrastructure, and products that improve availability. For additional information, see `http:`

`//www.sun.com/availability`

Genesys is the Sun code name for the platform architecture of datacenter.com, the new datacenter model. Genesys is also a program, consisting of products and services, aimed at helping IT organizations move their datacenters into the dot com era. For additional information about Genesys, see `http:`

`//www.sun.com/datacenter/genesys.html`

Sun has received the attention and enthusiasm of today's software developers. Many of the new Internet-enabled applications are written for the Solaris™ operating environment and optimized for Sun systems. Sun is the platform of choice for enterprise resource planning (ERP), electronic commerce, and database/data warehousing applications, as well as server consolidation initiatives.

The Sun-Netscape Alliance, known as iPlanet (`http://www.iplanet.com`), also supports the portal computing model with software and services—as well as content and audience reach through the AOL/Netscape end of the alliance. The Sun-Netscape alliance offers a full line of software products including the Netdynamics™ application server, the highest performance application server on market.

To support the new datacenter model, large organizations must employ a backup and restore solution that can scale massively in a widely distributed environment. To help customers meet this demand, Sun offers OEM versions of the two most popular and powerful enterprise backup and restore tools on the market. They are: VERITAS NetBackup as Sun StorEdge Enterprise NetBackUp™, and Legato Networker™ as Solstice Backup.™

There are many other backup and restore products on the market, but these two are among the most scalable and robust products available. Choosing a backup and restore tool is one of the most important decisions an IT manager must make. For guidance on this issue, see "Backup Tool Selection Criteria" on page 6.

Note – This book uses the name NetBackup when referring to the VERITAS NetBackup product.

Modern developments in backup technology require significant processing power and I/O bandwidth. Sun Enterprise servers provide scalable symmetric multi-processing, from one to 64 high-performance UltraSPARC™ II processors, with up to 64 Gbytes of memory, and up to 20 Tbytes of disk storage. The advent of scalable I/O platforms such as these enables a database to be configured for the optimal balance of processing power and I/O bandwidth, enabling online backups to proceed with minimal impact on database performance.

Note – The Sun StorEdge Instant Image and Sun StorEdge Network Data Replicator products are not covered in this book. For information on these products, see the BluePrints book: *Business Continuity Planning for Sun Microsystems Technologies*.

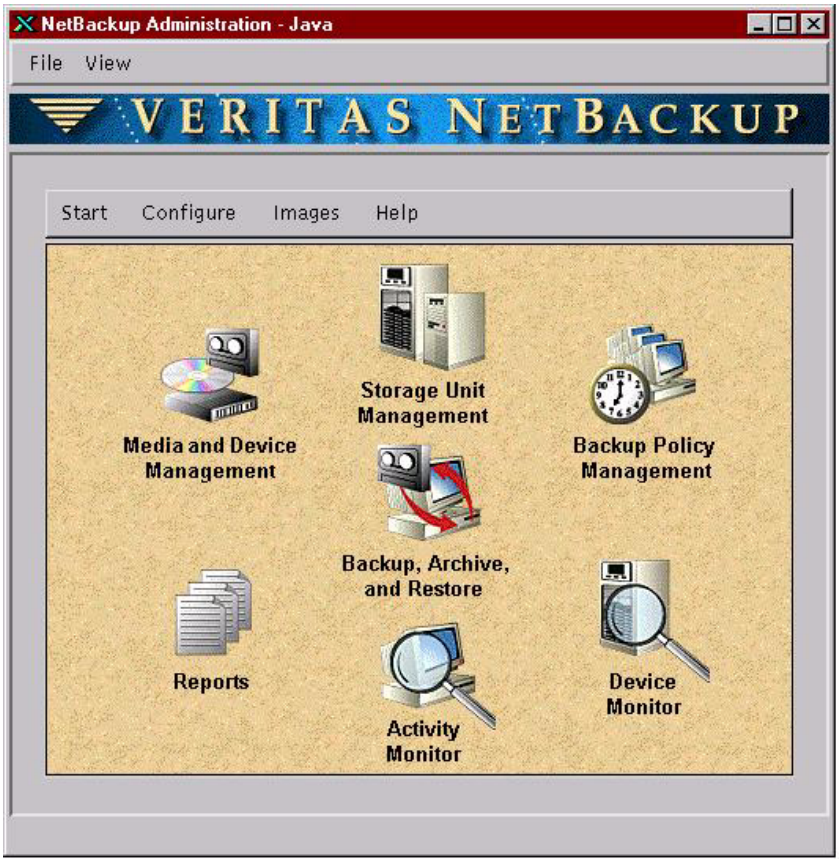

FIGURE 1-2 The NetBackup GUI supports centralized administration of a backup and restore architecture that may include widely distributed nodes located throughout a global enterprise. The Java™ technology version of the GUI is shown.

Managerial Issues

Implementing a backup and restore architecture scalable enough to meet current and future business needs, requires broad support for the project. Depending on the organizational structure of the business, a diverse set of people and interests will need to be brought onboard. The cooperation and approval from different VPs, business units, and operational groups may be required.

Some of the groups that could be involved are:

- Systems Administrators—The persons responsible for planning and implementing backup and restore solutions.
- Database Administrators—Including DBAs inside and outside the datacenter.
- Application Development Groups—Those responsible for developing internal applications and customizing purchased enterprise applications.
- Operations—Those responsible for day-to-day operations of the backup and restore tools.

The system administrators and the database administrators must work together to define responsibilities. The DBAs will want to retain control over the database backup and restore procedures. The system administrators will define the ways that file systems are backed up, including files that are created when databases are exported for backup purposes. These two groups should work together to define the backup and restore architecture, evaluate and test the backup and restore tool that will be used, customize the backup and restore tool, and define the operational lines of responsibility.

The mainframe and client-server system administrators need to work together. It is important that these two groups be integrated as much as possible, rather than separated into distinct realms. Integration can increase trust and cooperation.

Negotiations can be carried out to determine service levels for each business unit. Decisions on such things as how long backed up data must be kept online, which data is critical and which is not, what the backup window will be, and other questions, need to be addressed. (See TechEvolve Corp. Case Study" on page 28, for more information on negotiations.) Service levels need to be set up with the various business units and formalized into service level agreements. The customers of the IT services need to pay for what they are receiving, since service levels imply human, software, and hardware resource costs.

In a rapidly growing enterprise, there may already be an established backup and restore infrastructure and set of procedures, however, the current architecture may no longer be suitable. Several different backup and restore tools may be in use,

either purchased or developed in-house. Eventually these tools could reach their limitations in terms of scalability, or may become too unwieldy to handle the growing needs of the enterprise.

The solution is to consolidate under a highly scalable toolset and architecture. All groups using the legacy tools will need to agree on their use to make this consolidation successful. This could involve tradeoffs, therefore, everyone involved should be informed of the overall goals. For example, a backup and restore tool developed in-house might serve some purposes well, and might be easy to customize for certain groups. But the tool may need to be replaced with a standardized tool in the interest of overall efficiency. Everyone involved should understand the larger goals, and be prepared to make some concessions if necessary.

To successfully implement a new backup and restore architecture, a plan should be developed based on guidelines in Chapter 4, "Methodology: Planning a Backup Architecture" on page 63. All parties affected by the transition should be involved in developing, approving, and implementing the plan.

Backup Tool Selection Criteria

Choosing a backup and restore tool is one of the most important decisions you will have to make. The entire backup and restore architecture will be built around that tool. The features, and development direction of the tool should be evaluated in light of your current and future business requirements. Consideration of the stability of the tool vendor, quality of their service, and level of technical support should be included in the evaluation.

The following section covers a wide range of selection criteria that should be taken into consideration when purchasing a backup tool.

Architectural Issues

The architecture of a backup tool is extremely important. The entire backup and restore infrastructure can be enhanced or limited by the architecture of the underlying tool.

Ask the following questions:

- *Does the architecture scale to support your current and future needs?*

  NetBackup and Solstice Backup use hierarchical architecture. Hierarchical architecture simplifies the function of adding nodes to a network of backup servers, and in structuring backup architecture appropriately for a particular organization. For example, a global enterprise may have several datacenters

around the world in which master backup servers can be located. With hierarchical architecture, it is easy to add and delete slave backup servers beneath each master. This architecture can therefore be scaled to a global level, while still providing required flexibility.

- *Is SAN support provided?*

  A storage area network (SAN), is a high-speed dedicated network that establishes a direct connection between storage devices and servers. This approach allows storage subsystems, including tape subsystems, to be connected remotely. Tape SANs enable the sharing of tape resources efficiently among many servers. Both the backup and restore tool and tape library must provide SAN support to make this possible.

  With a SAN, information can be consolidated from increasingly remote departments and business units than was previously possible. This approach enables the creation of centrally managed pools of enterprise storage resources. Tape resources can be migrated from one system on a SAN to another, across different platforms.

  SANs also make it possible to increase the distance between the servers that host data and tape devices. In the legacy model, tape devices that are attached via a SCSI interface are limited to 25 meters. With fibre channel technology, distances of up to 10 kilometers can be supported. This makes it possible to use storage subsystems, including tape devices, in local or remote locations to improve the storage management scheme, and to offer increased security and disaster protection.

Note – At the time of this writing, tape SANs are not a viable solution for production environments. However, planning for a tape SAN will ensure your backup and restore architecture is well positioned to transition to this technology as it becomes production-ready.

- *Can backups to remote devices be made*?

  If a server hosts a small amount of data, (less than 20 Gbytes) it can be more convenient to back up over the standard network. Traditional network backups may be chosen in some cases.

Remote and Global Administration

Any widely distributed organization, needs to centrally manage and remotely administer the backup and restore architecture.

The following questions should be asked:

- *Does the tool support centralized administration*?

  The VERITAS Global Data Manager (GDM) utility supports the concept of a global data master. This master-of-masters server enables central control of a set of master backup servers located anywhere in the world.

- *Does the tool support remote administration*?

  The tool should support all capabilities from any location including dial-up or low bandwidth networks.

- *Is electronic client installation available?*

  Fast, easy software distribution of backup client agents should be supported.

- *Is backup progress status available*?

  The completion time of a backup should be available, including the amount of data backed up so far and the remaining data to be backed up.

- *Can historical reporting logs be browsed*?

  The tool should support an in-depth analysis of prior activity.

- *Does the tool provide disaster recovery support?*

  It should be possible to recover data remotely across the network.

- *Are unattended restore operations supported?*

  The unattended restore of individual files, complete file systems, or partitions should be supported.

- *Are unattended backups supported?*

  Does the tool have the ability to schedule and run unattended backups. A backup tool generally has a built-in scheduler, or a third-party scheduler can be chosen. Large organizations commonly use a third-party scheduler, since many jobs, not just backups need to be scheduled. A greater level of control is offered by the script-based scheduling approach. If using a third-party tool, ensure the backup tool has a robust command-line interface, and the vendor is committed to backward compatibility in future versions of the commands that control the execution of the backup tool.

Automation

Backup process automation is essential in any large organization as it is impractical to run backup jobs manually. The effectiveness of the entire backup and restore architecture is dependent upon the automated support provided by the backup tool.

Ask the following questions:

- *Does the tool support automation of system administration*?

  The tool should provide a robust set of APIs that enable customizing and automation of system administration. The API should allow customization by using standard or commonly accepted scripting language such as bourne shell, perl, or python.

- *Is there a GUI-based scheduler?*

  It should be easy to define schedules, set backup windows, and identify backups with meaningful names.

High Availability

If the data source must be highly available, then the backup and restore tool needs to support that requirement. This means both the tool, and the data it manages must be highly available.

Ask the following questions:

- *Is the backup tool, itself, highly available*?

  This involves not only the backup and restore tool, but also the servers on which the tool runs. In a master-slave architecture, the master and slave software and hardware servers may need to be designed using redundant systems with failover capabilities. The availability requirements of the desktop systems and backup clients should also be considered.

- *What are backup retention requirements?*

  Determine how long tape backups need to be retained. If backing up to disk files, determine the length of time backup files need to be retained on disk. The media resources needed to satisfy these requirements depends on the retention times and the volume of data being generated by the business unit.

- *Does the tool ensure media reliability*?

  The backup and restore tool should ensure media reliability, and reliability of online backups.

- *Does the tool provide alternate backup server and tape device support?*

  A failure on a backup server or tape device should cause an automatic switch to a different backup server or device.

- *Does the tool restart failed backup and restore jobs for single and multiple jobs?*

  A backup or restore job could fail mid stream for any number of reasons. The backup tool should automatically restart the job from the point it left off.

Performance

The performance of the backup architecture is critical to its success, and involves more than just the performance of the backup tool itself. For additional information on this topic, see Chapter 4, "Methodology: Planning a Backup Architecture" on page 63.

Ask the following questions:

- *Will the backup tool performance meet your requirements?*

  The efficiency of the backup tool—for example, the speed at which it sends data to the tape devices—varies from product to product.

- *Does the tool's restore performance meet your requirements?*

  The efficiency of the backup tool—for example, the speed which it sends data to tape devices—varies from product to product.

- *Does the performance of a full system recovery meet Business Continuity Planning requirements?*

  If the tool will be used in disaster recovery procedures or business continuity planning, it must meet those BCP requirements. For example, many BCP requirements specify a maximum amount of time for the restore of all data files and rebuilding of any backup catalogs or indices.

- *Does the tool provide multiplexed backup and restore*?

  To achieve optimum performance, the backup and restore tool should read and write multiple data streams to one or more tapes from one or more clients or servers in parallel. For additional information on multiplexing, see Section "Multiplexing" on page 22.

- *Does the tool enable control of network bandwidth usage*?

  The backup and restore tool should have the option of controlling network bandwidth usage.

- *Is raw backup support provided?*

  The backup and restore tool should be able to backup raw partitions. Under some conditions raw backups can be faster than filesystem backups. (See "Physical and Logical Backups" on page 17.) Also, determine if an individual file can be restored from a raw backup. (See "Raw Backups With File-Level Restores" on page 24.)

- *Is database table-level backup support provided?*

  If there are situations where the individual tables in the environment can be backed up, rather than always having to backup entire databases, it could significantly increase the performance of the backup architecture. The backup tool must support this option.

- *Does the tool provide incremental database backup?*

  This is important, since it is impractical to backup an entire database every hour. Incremental backups significantly increase the performance of the backup architecture.

Ease-of-Use

Ask the following questions:

- *Is it easy to install and configure the backup tool?*

  For a large corporation this may not be a major consideration, since it is possible to use the vendor's consulting services during product installation and configuration. For smaller organizations, ease of installation and configuration could be more important.

- *Does the tool provide backward compatibility*?

  Backup tool versions should be compatible with earlier versions of the tool. This makes it possible to recover data backed up with earlier versions of the tool. This also enables upgrading without having to change the backup architecture.

- *Are error messages are clear and concise?*

  If this is not the case, delays or difficulties could occur when attempting to recover data in an emergency situation.

- *Is message log categorization and identification provided?*

  This function will make it easier to diagnose problems.

- *Is the tool's documentation clear and complete?*

  Good documentation is fundamental to proficient use of the tool.

- *Does the tool's vendor provide training*?

  A training package should be included with the purchase of any backup tool. The vendor should be available for on-site training of operations staff, and to supply documentation about the specifics of your configuration.

- *Does the vendor provide worldwide customer support*?

  Technical support should available around the clock from anywhere in the world.

Ease-of-Customization

The backup and restore architecture must be flexible and customizable if it is to serve the growing needs of a dynamic organization. Any efforts to design flexibility into the architecture can either be enhanced or limited by the backup tool chosen.

Ask the following questions:

- *Is it easy to customize the tool?*

  No two environments are the same. Highly customized backup and restore infrastructure may be needed to fully support business needs for a specific environment. It should be possible to modify the backup and restore tool to fit any requirements. For example, an environment may require a customized vaulting procedure. Or, an API may be needed that makes it possible to add and delete information from the file history database. This feature could be used to customize the backup and restore tool to interface with legacy disaster recovery scripts that need to be inserted into the file history database.

- *Does the tool provide state information from before and after a backup job is run*?

  This function provides the ability to place a wrapper around the backup tool. This is useful if a script needs to be executed prior to running a database backup, for example, to shut down the database and perform related functions. Or, if after a full parallel export, to run another script to bring the database up.

- *Does the tool provide the ability to add and delete servers*?

  Hierarchical architecture enables servers to be added, deleted, and managed separately, but still be encompassed into a single unified master management interface. The hierarchical design allows for easy scaling of the entire backup and restore infrastructure.

Compatibility With Platforms and Protocols

It is important that the backup tool supports the platforms and protocols specific to a business.

Ask the following questions:

- *Is the tool compatible with your past, present, and future operating systems*?

  Many different operating systems may need to be supported in a heterogeneous enterprise environment. These could include; Solaris software, UNIX, Microsoft Windows, Novell Netware, OS2, NetApp, and others. The tool should backup and restore data from all these sources, and should run on any server computer.

- *Does the tool support Network Data Management Protocol (NDMP)*?

  NDMP is a disk-to-tape backup protocol used to backup storage devices on a network. NDMP supports a serverless backup model, which makes it possible to dump data directly to tape without running a backup agent on the server. The

backup tool should support NDMP if running small network appliances which do not have the resources to run backup agents. For further information on NDMP, go to:

`http://www.ndmp.org`

Compatibility With Business Processes and Requirements

The backup tool should support real business needs. These include the technology resources currently in place, as well as the day-to-day business processes within an organization.

Ask the following questions:

- *Does the tool support leading databases and applications?*

  Support should be provided for all leading databases and applications such as Oracle, Microsoft SQL Server, Sybase, Informix, Microsoft Exchange, and SAP R/3.

- *Are user-initiated backups and restores available?*

  In some environments, a backup policy may be in place to provide easy-to-use interfaces for end-users that reduces system administrator intervention. In other environments, user-initiated backups and restores may be prohibited. If user-oriented features are required, ensure the tool provides them.

- *Is vaulting support provided?*

  Vaulting can involve managing tapes, moving tapes out of libraries after backups are completed, processing tapes, and transporting them offsite to external disaster recovery facilities.

  For example, NetBackUp's BP Vault facility automates the logistics for offsite media management. Multiple retention periods can be set for duplicate tapes, which will enable greater flexibility of tape vaulting. It supports two types of tape duplication—tape images can be identical to the original backup, or they can be non-interleaved to speed up the recovery process for selected file restores.

- *Can data be restored in a flexible manner, consistent with business needs?*

  Depending on the different situations that arise from day-to-day, it may be necessary to restore different types of data, such as a single file, a complete directory, or an entire file system. The tool should make it easy to perform these kinds of operations.

- *Does the tool enable the exclusion of file systems?*

  There are situations when this feature is crucial. For example, when using the Andrew File System (AFS) as a caching file system. To the operating system, AFS looks like a local filesystem. But AFS is actually in a network "cloud", similar to

NFS. It may not be desirable to backup AFS partitions (or NFS partitions) that are mounted on an AFS or NFS client. For example, if backing up a desktop machine with different partitions mounted from other servers, you would not want to backup the other servers.

With NFS, it is possible to tell when traversing into NFS space, however AFS is seamless and therefore any file systems that don't need to be backed up should be excluded.

- *Does the tool support the security needs of a business?*

  The tool should support the security required by the operating system. If added data protection by encryption is required, the tool should support it.

- *Can jobs be prioritized according to business priorities?*

  Priorities for backups should be based on importance. For example, a critical database should take priority over less important desktop data.

- *Does the tool support internationalization and localization*?

  The backup tool should provide the ability to run under a localized operating environment.

- *Does the tool support Hierarchical Storage Management (HSM)?*

  Will the tool support HSM directly or integrate with an HSM solution?

Backup Catalog Features

The backup catalog lists historical backups, along with files and other forms of data that have been backed up. This features of the backup catalog can be important to the performance and effectiveness of the architecture.

Ask the following questions:

- *Is an online catalog of backed up files provided?*

  A file history catalog that resides in a database will enable the user to report out of the database, perhaps using different types of tools. For example, the file history catalog may reside in an Oracle database. However, the user may want to report with different reporting tools such as e.Report from Actuate Corporation, or Crystal Reports from Seagate. If the backup catalog resides in the database, the vendor should publish the data model. On the other hand, if the backup catalog resides in a flat file, no special database is required to read the catalog.

- *Does the tool provide the ability to quickly locate files in a backup database*?

  It is important to quickly locate files or groups of files in the backup database. Tools that take a long time can adversely affect recovery times.

- *Does the tool provide the ability to modify the backup database through an API*?

  If the backup catalog needs to be programmatically modified, an API published by the vendor should be used. If a standardized API is not available, it is not advisable to modify the backup database programmatically.
- *Does the tool provide historical views of backups*?

  It should be easy to determine which historical backups are available
- *Does the tool provide a true image restore*?

  Restores should be able to recreate data based on current allocations, negating the recovery of obsolete data. (see "True Image Restore" on page 24)
- *Can the backup catalog be recovered quickly?*

  If a catastrophic failure occurs, the tool should allow the backup catalog to be quickly restored. This may involve retrieving the catalog and indices from multiple tapes.

Tape and Library Support

Ask the following questions:

- *Does the media (volume) database provide required features?*

  Indexing, tape labelling, customizing labels, creating tape libraries, initializing remote media, adding and deleting media to and from libraries, or using bar codes in the media database are functions that may be required. It is important to be able to integrate the file database with the media database. Additionally, the library will need to be partitioned, for example, to allocate slots in the library to certain hosts.
- *Is tape library sharing supported?*

  Lower tape robotic costs can be achieved by sharing tape libraries between multiple backup servers, including servers running different operating systems
- *Is tape management support provided?*

  The backup tool should enable management of the entire tape lifecycle.
- *Does the tool support your tape libraries?*

  Support should be provided for all leading robotic tape devices.
- *Does the tool support commonly used tape devices?*

  Support should be provided for all leading tape devices.
- *Can tape volumes, drives, and libraries be viewed*?

  The tool should report on tape usage, drive configuration, and so forth.

Cost

Backup and restore costs can be complex, ask the following questions:

- *What are the software licensing costs?*

  Are software licensing costs based on; number of clients, number of tape drives, number of servers, or the size of the robotics unit? These costs will impact the backup architecture and implementation details.

- *What are the hardware costs?*

  The architecture of a backup solution may require the purchase of additional tape drives, disks, or complete servers. Additionally, the backup architecture may require, or drive changes to your network architecture.

- *What are the media costs?*