



## The Critical Role of Data Consolidation for Your Data Protection Systems

***BOCADA***<sup>®</sup>

**Technology Brief**

*March 2005*

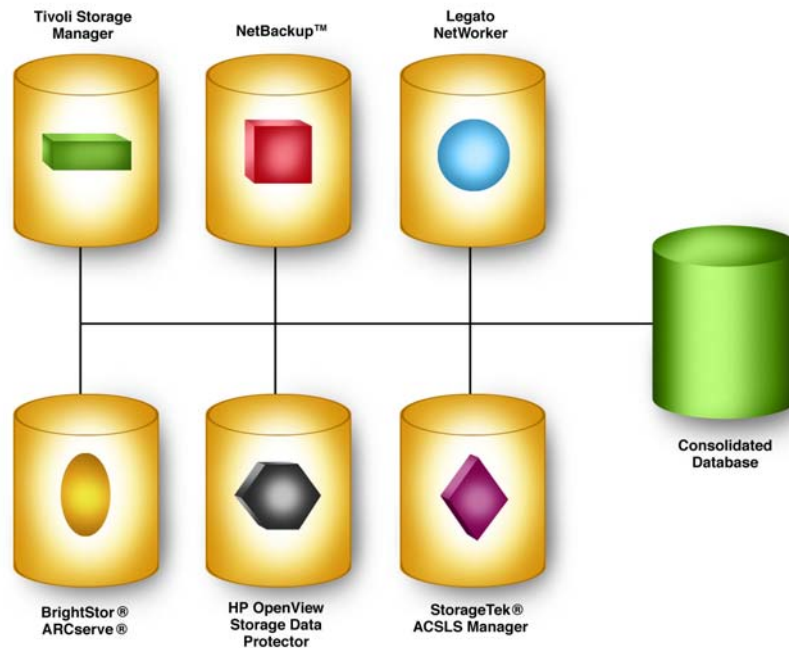
## Abstract

To present an enterprise-wide view of data protection measured against business objectives, a management application must be able to consolidate data. Consolidation is the process of collecting data on activities and outcomes from disparate backup applications across multiple servers and then storing it in a uniform database that enables business intelligence controls. Only consolidated data delivers a high-level, strategic view of data protection, ensuring mission-critical data is protected and recoverable. This strategic view is the proverbial “gold” that can be mined from your data protection systems; it is the asset that will enable IT management to prioritize activity to bring activities in line with business objectives and communicate policy compliance to corporate officers and data owners.

## Data Consolidation: The Key to Enabling Business Intelligence

Most enterprise applications provide operational reports rather than presenting strategic business information. These reports are commonly designed only to support the application or improve component performance, such as network throughput. Data collection therefore mirrors the operational aspects of the application instead of providing insight into—and benchmarks against—business objectives.

Presenting higher-level business-relevant information requires the additional step of data consolidation: the process of mining data from multiple sources and then storing it in a uniform database with a consistent structure. Data consolidation is more than merging databases by importing tables. It also requires canonizing data to create consistency across data elements so that all



**Figure 1. Consolidated Data for Data Protection.** Only through consolidation can data in disparate formats and from different sources be canonized to create a consistent, enterprise-wide view of data protection.

fields and tables contain the same class of information. The structure of a consolidated database specifically supports business intelligence operations, while still maintaining “hooks” to native data sources. These hooks are critical to support troubleshooting and auditing activities common in data protection.

In the data protection business, only consolidated data enables IT to move beyond a simplistic reporting of backup application status to delivering information about SLA performance, asset utilization, audit and bill-for-service functions. Consolidated data yields accurate, strategic and actionable information about the operational and systemic aspects of data protection.

## The Challenge of Consolidation for Data Management Protection

As previously noted, consolidation for data protection management involves more than merging fields and records. Restructuring data for business intelligence requires thorough operational knowledge of each backup application. It also requires an understanding of the underlying workflow issues important to the enterprise as well as the ultimate business value of the data itself.

Backup applications are not standards-based and, as a result, often take different approaches to solving the same fundamental problem. For example, not every vendor tracks full and incremental backups the same way. Tape backup rotation strategies and terminologies vary. Error codes are unique—in number, format and content—to each vendor. Even data structures differ substantially between products.

For example, VERITAS NetBackup™ uses the term “policy,” but Legato NetWorker®

refers to “save groups” and “save sets.” While these terms reflect approximately the same concept, this concept is nonexistent in Computer Associates BrightStor® ARCserve®. Moreover, all three of these applications apply and report on both full and incremental backups, but IBM® Tivoli Storage Manager does not. Even when applications use exactly the same terms, they often have completely different meanings.

Canonizing this data requires extensive knowledge of each application. For example: the time since the last backup of a critical server is essential to its recoverability. In one application, this data may be found in a table with the retrieval based on the name of either the server or the tape library. In another application, the time may be calculated by parsing the log file and extracting the timestamps. If the backup failed, information about the causes for failure may be reported as text strings, error codes or both. Because no two vendors share a common methodology, unconsolidated data derived from disparate sources cannot offer insight into data recoverability.

A data protection management application that consolidates dissimilar data and then delivers information in a format that makes business sense is a quantum leap beyond reporting “problems with technology.” Consolidated data propels the leap. It is the difference between knowing that an individual backup failed on a single server and knowing that the company’s critical data assets can be recovered with a 98% success rate.

## The Operational Limits of Vendor Solutions

Fundamentally, large enterprises rely on distributed, network-based systems consisting of hardware and software applications purchased from multiple vendors. Any reporting capability provided

by these applications will be tightly linked to the vendor's particular technology and methodology—strictly operational by nature. These applications may disclose the function of individual servers, but they will not offer insight into how data protection activities—or the absence of data protection activities—impact the overall business across a heterogeneous environment. While some vendors try to report against competing products in addition to their own, the results are mixed. These vendors either design a product that imports the other application's entire database in native form, or they force-fit their competitor's data into their own data structure. However without first consolidating the data, the differences in data structures result in distortion. The ultimate consequence for the enterprise of unconsolidated data is substandard reporting for all applications.

Consolidated data must be completely restructured so that it does not reflect the underlying technology or bias of any one vendor to avoid this distortion. However, it

must still be retain its hook to its origin and to its fundamental components. More importantly, the data structure must support the underlying business processes rather than vendor operational biases to provide true business intelligence and insight into the data protection process.

## The Power of Consolidated Data

Only consolidated data delivers true visibility: a high-level view of activity at multiple levels of abstraction across multiple servers. This visibility has value for IT management and staff, but its uses also exceed IT operations by helping the entire business meet its goals for data protection, cost control and policy compliance.

### Improved Data Protection

Backup is more than copying data; it is protecting extremely valuable business assets. Yet, it is estimated that 40-60% of

## Recognizing Consolidated Data

Consolidated data will typically have these five characteristics:

1. **Its structure does not reveal the personality of the application(s) from which the data is mined.** If the data reflects one vendor's structure, it is evident the other vendors' dissimilar data is being "shoehorned" into the database. If the data contains multiple formats each reflecting a different application, it is evident the data is being largely imported, with no attempt to consolidate it.
2. **The same set of data elements are extracted from the different applications supported.** If message numbers are stored with NetBackup messages, then messages coming from NetWorker will also have numbers.
3. **Data is rationalized at the earliest point in the mining process to minimize performance impact.** The data is converted to the common format as soon as possible, preferably as it is mined. In very large environments, this is a key requirement for performance reasons.
4. **The structure of the data reflects its purpose.** If the purpose of the data cannot be discerned, too much data has likely been stored and it and must be significantly "scrubbed" to extract meaningful reports.
5. **It includes a means to reconcile the consolidated data with the original source.** This is not only a requirement for "drill-down" troubleshooting, it also verifies the integrity of the data for compliance reporting and maintaining an audit trail.

backups fail<sup>1</sup>, and most often these failures have nothing to do with backup hardware or software. Failures are caused by network problems, inappropriately configured servers, too much data for the backup window, insufficient tape capacity or forgotten tape swaps. In fact, most enterprises that rely on conventional backup operational reporting do not know their true backup success and failure rates. They overestimate their backup success—often by a wide margin—and cannot prove they are meeting their Service Level Agreements (SLAs). However, enterprises that move to data protection management solutions relying on consolidated data structures quickly identify chronic sources of error and can double their success rates as a result of a realistic, strategic view of their data protection outcomes.

Without consolidation, chronic failures are buried within myriad details of log files and embedded databases, blocking access and concealing or distorting information on trends and projected futures. With consolidated data, failure patterns can quickly be identified and corrected because the causes, rather than the symptoms, of continual errors with specific servers, clients, jobs, files, tape devices and libraries can be addressed. Another advantage is that consolidated data retains the character of the original data to facilitate troubleshooting and to prove compliance, which is often necessary for an audit.

Consolidated data also allows IT staff to prioritize their efforts rather than responding to alarms. IT staff viewing reports of backup failures will know where they must first focus their attention on the failures with the most impact on the business. Examples: a failed backup of a database that contains current customer orders demands a higher priority than a failed backup of a parts catalog that could be largely restored from an older backup. From an overall business perspective, restarting the backup of the

finance server is more important than protecting individual laptops. Compare this approach to the flood of arbitrary alerts typically generated by multiple backup applications that can cause IT staff to descend into a frenzy of non-strategic troubleshooting. Consolidated data provides insight and perspective to focus troubleshooting efforts and promote the prioritization of activities.

## Reduced Costs

Global enterprises spend hundreds of millions of dollars on backup systems and management to handle exponential data growth and the corresponding need for protection. In the process, these enterprises are often forced to overspend on backup systems to accommodate unnecessary backups. By understanding utilization, enterprises can lower their total data protection costs.

Without consolidated data, backup can become a “black hole,” devouring time and resources. Data owners and business units can only hope the critical data that supports their business needs is being backed up and that the cost for the service is fair. With consolidated data, it is possible—for the first time—to see and share the actual cost of data protection on a per-asset or per-department basis. Often, enterprises will find that data of comparatively low value to the organization is overprotected through redundant backups. This redundancy results in high costs in tape, time and staff resources. However, with visibility, business units can be charged for the costs of providing data protection services, including the degree to which the data is protected. Data owners can then see if the cost of protection coincides with the information asset’s value and then adjust their data protection requirements accordingly. Simply by notifying data owners of their share in the cost of protecting data and thereby encouraging a reprioritization of asset production, many companies have

---

<sup>1</sup> Enterprise Strategy Group

recovered the cost of data protection management solutions within six months.

### **Better Utilization and Throughput**

A consolidated database captures detailed information about backup utilization and throughput across the enterprise and over time. Enterprises gaining this higher-level view have often been surprised to find only a fraction of the capacity their tape libraries is used, or that long-expired data is being retained on tapes. They have also realized that network collisions are often the cause of backup problems—a problem they might otherwise have attempted to solve by adding new servers and tape libraries. With the cost of new assets in the tens of thousands of dollars, the ability to fully utilize existing resources in lieu of additional purchases alone often justifies the investment in data consolidation reporting tools.

### **More Efficient Use of Staff Time**

Without consolidated data on data protection activities, IT staff spends the bulk of their time hunting for problems rather than solving them. Data consolidation replaces the time spent tracking down causes of failures with proactive management and solutions.

Most importantly, training and staff burnout costs can be lowered. With consolidated data members of the IT staff do not need specialized knowledge to interpret cryptic data from each of backup vendor—each with its own formats, codes and vocabularies—to find errors. Nor do they need to spend time manually collating information and creating reports for management and data owners.

For example, creating a 30-day success and failures report for a group of SQL servers required a 45-employee-hour investment for one company. A consolidated data application generated the same report with a few keystrokes. In fact,

this company was able to build custom queries to meet their business-specific needs, saving time and money as a result.

### **Proof of Compliance**

Consolidated data allows IT staff to share high-level reports depicting SLA performance and data protection with stakeholders across the enterprise. By returning information at multiple levels of abstraction, a consolidated reporting application reveals the recoverability of critical data in light of policy mandated benchmarks. Consolidated data enables the enterprise to prove compliance with policies based upon Sarbanes-Oxley, HIPAA, SEC 17a, BASEL II, FDA 21 CFR Part 11, and other regulations because it provides validation and tracking through an independent third-party. This realistic view of data recoverability not only allows CIOs to ensure process consistency, it also allows senior IT management to sleep better at night, as they have a clear understanding of the recoverability of their data assets.

### **Summary**

Data consolidation is essential to data protection management success. Only through this complex yet necessary step can business intelligence be delivered across large heterogeneous backup environments. Through consolidated data enterprises can reduce risk and cost, discover opportunities and prove data protection processes and results align with organizational objectives.

## About Bocada

Bocada is dedicated to ensuring that data protection services, systems and processes meet organizational objectives for quality, cost and compliance. Our flagship product, BackupReport, provides objective insight into SLA performance, helping companies to improve data recoverability, reduce the cost of managing data protection operations, ensure compliance and communicate results. Bocada solutions have been deployed in more than 165 market-leading customer and partner environments worldwide, including Amgen, BankOne, Cap Gemini Ernst & Young, SBC, Sprint, Unilever and Xerox. Bocada is a private company headquartered in Bellevue, Washington.

For more information about Bocada and BackupReport, please contact us or visit [www.bocada.com](http://www.bocada.com).

Bocada  
10500 NE 8<sup>th</sup> Street  
Bellevue, WA 98004  
Tel: +1.425.818.4400  
Fax: +1.425.818.4455  
[sales@bocada.com](mailto:sales@bocada.com)  
[www.bocada.com](http://www.bocada.com)