Industry Trends and Technology Perspective White Paper

Data Protection Management (DPM)

"A look at the benefits of DPM for timely and effective data protection management"

By Greg Schulz

Founder and Senior Analyst, the StorageIO Group

September 27, 2006

Data protection management (DPM) provides timely insight and analysis of data protection activities, including cross domain event correlation and what-if analysis for planning purposes. This paper looks at the industry landscape, market adoption trends, benefits, and desired attributes of a DPM solution.

www.storageio.com P.O. Box 2026 Stillwater, MN 55082 651-275-1563 info@storageio.com

Introduction

To manage data effectively, it is essential to ensure that stored data is safe, secure, and accessible. Good data protection management (DPM) means having solid processes and methodologies in place to maintain data integrity, which involves processes that cut across different IT technologies and business domains. Sound DPM also involves having rules and policies that enable effective data management decisions.

The changing face of data protection

Timely, relevant information combined with historical perspective and policy rules form the basis of effective DPM and decision making. With more technology components, processes, and procedures than ever before, today's applications depend on multiple systems that need to be protected. With such heavy dependence on information systems, the reliable accessibility of their data is critical.

Disk based data protection solutions (disk to disk-D2D, continuous data protection-CDP, virtual tape library-VTL, snapshots, replication, and so forth) are generally pieced together from multiple vendors and lack a single cohesive management strategy. Without a cohesive strategy, backups fail due to lack of timely event correlation and analysis. The lack of timely event resolution results in data integrity degradation complicating compliance and other reporting. For example, backup failures are often caused by problems with other components in different technology domains, such as network configuration or server availability. When these technology domains are owned by separate groups, it can be difficult to determine the true root cause of problems let alone in a timely fashion.

A goal of effective DPM is to identify what is really happening and anticipate problems rather than chasing down a false trigger after problems occur and trying to prevent problems from recurring. As more servers (physical or virtual) are deployed, the number of backup and data protection tasks increases, resulting in mountains of event and activity data to sift through. Manual management techniques cannot respond to the flood of information from multiple sources as the number of servers and backup jobs increase. Event and resource correlation may be done manually for a couple of servers using common desktop tools like Microsoft Excel. However, as the number of servers and the complexity of backup jobs increase, it becomes impossible to manage hundreds of servers manually or in a semi automated manner.

Data protection trends

There is a growing awareness of the need for reliable data protection to meet varying threats as well as regulatory compliance. Data is being retained and needs to be preserved for longer periods of time to meet regulatory compliance and to meet internal business requirements. Also, more data is being stored outside traditional data centers with more distributed and layered systems (think virtualization), making it that much harder to know if an application is protected, much less find and fix problems. Even with initiatives like information lifecycle management (ILM) and data pruning or reduction efforts, there is a growing need for more data to be stored and protected. The backup windows that still exist are few and far between, so they must be used effectively to avoid missed opportunities to protect data. Decisions that impact the effective protection of data must be made in real time using current, active data along with historical perspective for event correlation.

Data protection management landscape

With this new need for information insight into data and storage infrastructures that is not available in traditional, after-the-fact reports and problem logs, organizations are migrating from two-dimensional, technical success/failure reporting to 360 degree views of all business metrics (exposure, time-to-restore, SLAs, compliance). Real-time and historical views are being used for trending and analysis to enable timely data management decisions and to aid in the ability to catch errors before they grow into problems. The resulting complexity and cost of data protection are being addressed with automation for more timely and informed data management. The shift is from vendor-specific backup reporting tools to third-party DPM tools from leading storage vendors, such as EMC's Backup Advisor solution based on WysDM.

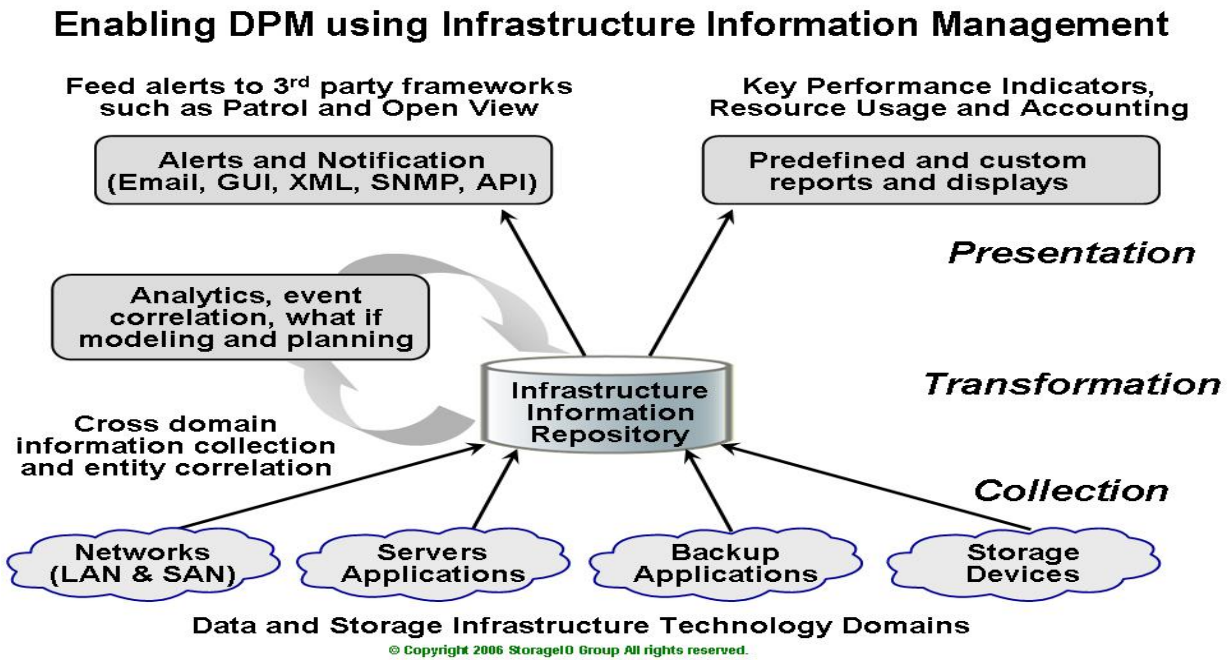

Figure-1: Cross-domain correlation and analysis

The continued shift from vendor- and backup-specific management tools includes the adoption of cross-domain data collection, along with automated resource and event correlation. In Figure-1, information is collected from across different technology domains, including backup applications, servers, applications such as databases, networks (LANs and SANs), and storage devices (SAN, NAS, disk, and tape). Leveraging a robust repository of correlated events and combining powerful analytics, DPM tools enable timely and descriptive alerts to be sent that provide insight into how backup and data protection tasks are performing to address errors and problems in real-time.

Data protection management benefits

At its most basic level, data protection management means ensuring that backup and data protection tasks run on schedule and that the intended data is, in fact, saved and safe. However, there is more to DPM. The benefits of DPM include viewing timely information crucial to making quick decisions and determining problems, along with improving overall data protection and IT resource utilization. Good DPM enables organizations to fully leverage and maximize existing IT resources instead of just solving problems with

more hardware. DPM makes it possible to do more with fewer resources in less time, while improving service delivery and data protection for increasing numbers of servers, applications, and storage systems.

A key feature of DPM is its ability to provide the information to support reactive and proactive decision making instead of simply reporting the success or failure of backup operations. This ability to make timely interventions to correct problems as they occur makes it possible to meet higher service-level objectives as well as recovery time objective (RTO) and recovery point objective (RPO) commitments. An important benefit of maintaining RTO, RPO, and other service commitments is meeting regulatory compliance needs for data protection and preservation purposes.

One of the worst data protection nightmares is attempting to recover data only to discover that the restoration failed due to bad, inconsistent, or incomplete backups. For example, protection of a billing system can require that different backups on various servers and databases all complete at a specified time. To ensure successful restoration and restart, all of the interdependent systems must be restored with application and transactional integrity preserved to the same point in time or recovery point objective. Timely and coherent data protection across multiple interdependent systems requires correlation of many events across multiple sources, including databases, servers, networks, and storage. Consequently as the size and complexity of any environment increases, so too does the volume of events and activity information that must be collected, correlated, and analyzed in a timely manner.

The major benefits of effective DPM include the ability to:

- Correct data protection errors when they occur, eliminating the surprise of failed backups
- Identify weak links in the end-to-end data protection chain for timely problem resolution
- Proactively manage data by using trends, analytics, and cross domain event correlation
- Analyze multi-dependent backup environments with more precision and in less time
- Measure and meet business and technical goals or service objectives
- Complement vendor-specific element managers and tools

Therefore, the best DPM tool is one that can do many things besides simply reporting on the success or failures of backups or which tapes were used. A product that includes the ability to collect activity and event information from devices in different technology domains and then present it in a cohesive manner is ideal. Thus an effective DPM tool should perform multiple functions, such as performance and capacity planning, modeling, event correlation, analysis, reporting on backup success or failure, as well as trigger alerts that contain insightful information to expedite problem resolution.

The more a tool can do, the better. However, it's important to find a balance between a monolithic framework-type tool that tries to do and be everything to everyone and a vendor-specific tool focused on specific technology. For smaller environments, a robust tool can take on functionality that might otherwise be performed with other monitoring and alert tools. By tailoring alerts that are sent to monitoring teams to include specific backup vendor error codes and other pertinent information, support staff can spend less time on diagnostics and triage moving directly to faster problem resolution. In addition, the ideal DPM tool would be able to use the reporting capabilities to display information pertaining to infrastructure resource utilization for budget planning and accounting purposes.

An ideal DPM tool is one that is focused on the key strengths of DPM (cross-domain correlation, cause-event analysis, reporting, and timely notification) yet also integrates cleanly with the rest of a company's IT infrastructure. This means that a DPM tool needs to co-exist and be able to install without disruption so it can collect data across different technology domains while seamlessly interfacing with other tools, such as assessment management input and event management (notification) systems, for output.

For Orange Business Systems, using DPM-based technology from WysDM has enabled key performance indicators (KPIs) that measure the effective service delivery, including data protection, to increase by double digits to near 100% success. This means that by having more insight into the data and storage infrastructure being backed up including understanding where errors were occurring and what was causing them, Orange was able to successfully focus on and repair those problematic areas. For example, tapes may have appeared to be the root cause of a data protection problem when in fact greater insight from DPM tools shows the real problem to be a slow network or perhaps a server off-line or un-available for backup. This increased access to reliable information may show that backup to, and rapid restoration from disk is more effective than using tape for daily backup, and that works best for longer term retention.

From a reporting standpoint, the more robust the data repository and correlation tools are, the more effectively you will be equipped to do root-cause event analysis and correlation. For example, you could go back several months and determine that a backup job failed because a particular server was not available to the network at that particular time of the day. Similarly, you could produce a report showing where and how a particular tape has been used over the past year to proactively determine if it is a candidate for replacement before it fails while in use.

What to look for in a DPM solution

An ideal DPM solution will support gathering information from multiple backup and data protection applications, platforms, and components in a timely and unobtrusive manner. Real-time information on the current status of executed, recently completed, and scheduled backup operations from all system components should be readily available in a single aggregated view. Additionally, reports should be available about resource consumption of storage and performance components of the backup path, such as servers, storage interfaces, and networks.

Reports should include the amount of CPU and memory used on servers, I/O activity, bandwidth consumption, and how much disk capacity or how many tapes were used. Event and error information should also be collected for real-time notification and problem-solving. It may also be necessary to incorporate data from nonstandard data protection systems, such as manual backups, mainframes, and midrange systems, to support comprehensive analysis and event correlation.

The result of an effective DPM solution is extensive information delivered in real time with a historical perspective in front of you in a single view. For example, an effective DPM tool should generate informative, to-the-point real-time alert messages for problem resolution as well as provide detail on why a backup job failed six months ago. As such, DPM requires intelligent tools to sift through the mountains of information and perform automated analysis. A key capability for a robust and scalable DPM solution is the ability to gather information from different technology domains.

This means collecting information from different hardware and software sources that is then correlated with events and associated backup jobs. For example, some DPM solutions allow for integration of varying business and technical data, such as directory services, asset management databases, cost code information, business ownership, and application dependencies. The result is a repository that is used for sending alerts, monitoring and reporting status, along with performing analytics for planning purposes. In addition to real-time monitoring, DPM technologies can enable planning and review of configurations to proactively find problems before they disrupt IT operations and assist with planning for future service improvement or IT infrastructure upgrades.

DPM tools that support predictive analysis can automatically keep a complicated environment performing optimally by automatically detecting performance or service-level degradation even if no apparent problem is being reported. This includes the ability to detect and send an alert for trending problems, such as servers that are backing up more slowly than normal or an application that is in danger of missing its RTO due to normal data growth. Noticing important service level degradation conditions in a timely manner leads to a consistently (and automatically) well-tuned data protection environment.

DPM tools that support predicative analysis like those from WysDM automatically detect situations such as RTO service-level degradation and alert customers to problems with their backups before it's too late. A byproduct of DPM solutions can be to provide performance and capacity planning reports that facilitate budget planning and monitoring of resource use by business units and applications. Effective DPM also will support multiple vendors' technologies across different domains while being easy to deploy.

Here's a checklist of features to look for in a scalable DPM solution:

- Support for onsite and offsite electronic and physical vaulting of data
- Analyze configuration settings and backup jobs to enforce best practices and polices
- Predefined and custom reporting with timely event notification via SNMP and email
- Flexible interfaces for information access, including GUI, XML, and custom export
- What-if analysis and event correlation using real-time and historical data
- Real-time activity and event monitoring, along with event cause-and-effect correlation
- The ability to enable performance and capacity planning across various infrastructure components
- Breadth of coverage of collected information from servers, storage, and network
- The ability to import or enter rules for policy enforcement and validation
- Cross technology domain and application-context-aware monitoring
- Support for applications with multi-system, multi-vendor, and multi-job interdependencies

Conclusion

Effective data protection management requires timely information to enable informed decision-making. Avoid tools that only report on what happened yesterday instead look for tools that proactively manage the process of promoting accurate and timely data protection leveraging intelligent analytics and cross domain event correlation capabilities. Data protection management tools such as those from WysDM provide operations and monitoring staff, engineers, designers, and managers with real-time and historical insight into how storage infrastructure and data protection tasks are functioning.

About the author:

Greg Schulz is founder and Sr. Analyst of the StorageIO group as well as the author of the book "*Resilient Storage Networks - Designing Flexible Scalable Data Infrastructures"* (Elsevier).