WHITE PAPER
June 2006

Business continuity, disaster recovery, and information lifecycle management

More closely related than you thought

ABSTRACT
Information lifecycle management is not normally associated with business continuity and disaster recovery, but in fact they share many common features. This paper outlines these similarities and provides techniques for leveraging existing business continuity and disaster recovery investments for more efficient overall information management.

Sun microsystems

1.0 Executive summary

Although generally not considered to be related, information lifecycle management (ILM), business continuity (BC), and disaster recovery (DR) all have the same basic goal: to keep business working through proper information management. By understanding the relationship among these three disciplines, information managers can help ensure proper information safeguards while conserving resources.

It all begins with a clear picture of the data requirements of your data users, your corporate culture, and your stockholders. While they are each unique, each bears heavily on how you handle data. And, of course, evolving federal data management requirements — typified by the Sarbanes-Oxley Act — dictate baseline levels of compliance.

These drivers require that you to begin a detailed data classification process. When this process is combined with an audit of current and future supporting system requirements, you begin to see the close interrelationship between ILM and BC/DR concepts.

Planning is key, and keeping the ILM/BC/DR relationship in mind will allow you to implement more effective strategies for safeguarding existing systems as well as new applications and data loads.

2.0 Introduction — defining and scoping the problem

BC, ILM, and DR must coexist effectively for an organization to arrive at the appropriate level-of-availability solution.

While business continuity (BC) and information lifecycle management (ILM) are not generally considered related disciplines, they both have one common premise: making data available to users when they want it and when they need it. But how do these two relate to each other? And how do they relate to disaster recovery (DR), which is frequently used in the same context as BC?

BC, ILM, and DR must coexist to:

- Provide users the highest levels of application data accessibility
- Provide corporations with economical and government-compliant solutions
- Enhance shareholder value (prevent degradation of corporate value due to the perception of risk)

Your organization must take all of these factors into account to arrive at the solution providing the appropriate level of availability. The following three sections describe each of these drivers.

2.1 User application data access requirements

IT departments frequently view end-user demands as unreasonable — not because these departments can't satisfy user needs, but primarily because complex computing environments are intrinsically unpredictable. This unpredictability is the primary driver behind all business continuity and disaster recovery solutions; it's why they exist.

So what makes user demands unreasonable? All users ask is:

- No disruption to daily business operations
- Fast application performance
- No limits on storage regardless of whether information is valuable
- Constant access to information no matter how old

Not unreasonable, unless these demands are multiplied by 100, 1000, or 10,000 individuals. Anything beyond a handful of users becomes difficult to manage and problems become unpredictable. However, breaking down each demand into concrete requirements — defining the problem — can help determine what can be accomplished.

Users demand:

- *No disruptions*
- *Fast performance*
- *No storage limits*
- *Constant access to information*

User requirements can be summarized as follows: "I require this application and data be available during [this specific period], but I can utilize offline mechanisms as long as the data is synchronized by [this specific time]. At a minimum, I can tolerate a disruption for no more than [X hours] per [this period]."

The resulting solution implies load-balancing, resources available to process new tasks, and mechanisms to provide data that has been previously processed and stored. These features — combined with data policy definitions (such as the importance of the application, its data, and the time limits associated with data access) — form the basic tenets of business continuity and disaster recovery.

Notice that information lifecycle management requires yet another level of policy definition: retention. In other words, how long do you keep the data and what infrastructure is required to make it accessible and usable? This last policy definition must take into account both the user's requirements and the organization's requirements (as driven by the Sarbanes-Oxley Act, for example).

2.2 Corporate requirements

Similar to end-user requirements, corporations have certain applications and data that are more critical than others, based on relative business value. Applications and data, however, represent only one aspect of the computing environment. Corporations must also provide the infrastructure required to make these applications and their data useful to users. So, how do we go about classifying this data and planning the needed infrastructure?

Corporations must adhere to compliance regulations to:

- *Classify applications*
- *Manage data through archival storage*
- *Retain data for an extended period of time*

Many corporations have already defined the applications that are considered the most critical. They also typically rank customer-related information as one of the most important data sets. Both of these must be clearly defined in order to develop effective BC and DR plans.

Awareness of the need for this classification process was brought to new heights leading up to the year 2000 (Y2K). Faced with a potential IT disaster, many corporations did not have the methodology, disciplines, or mechanisms to prioritize application and data value. Nor could they determine the exact risks of not porting the applications and their associated data to compliant platforms. Eventually, corporations developed basic risk analysis techniques to address the Y2K "fire drill" and used the knowledge to form and/or refine useful long-term BC and DR plans.

Beyond information management required for basic business needs, Sarbanes Oxley regulations demand that each corporation be able to:

- Classify applications and data by level of importance
- Manage data in a disciplined fashion from daily use to archival storage
- Retain data for extremely long periods of time

There are very serious implications if data and the application used to access it are not available. Thus, the discipline of information lifecycle management (ILM) is born.

The final piece of the puzzle is the computing infrastructure. System administrators know that to use available applications and data, they must also have the correct operating system, computing platform, and data "extraction" gear.

This extraction process also requires workforce expertise and/or extensive documentation. It is likely that whole new industries designed to support these legacy applications and their data will soon emerge. Currently, IT outsourcing centers perform at least some of these functions. Such outsourcing may be the most economical way to meet corporate risk mitigation requirements.

Corporations must implement solid BC, ILM, and DR processes to meet stockholder requirements. Downtime results in:

- *Revenue loss*
- *Lack of confidence in service*
- *Lack of confidence in building shareholder value and wealth*

2.3 Stockholder requirements

Previous requirements focused on internal user and corporate needs for application, data, and infrastructure availability. But how might a lack of these availability mechanisms affect external customer perceptions? And how would a corporation's investors and shareholders react to the revelation that their investment may not be in compliance?

The stockholder requirement is simply the ability to report the relative "availability" of a corporation. The corporation is viewed as the sum of all of its processes, tools, and data. Downtime of any sort results in:

- Loss of revenue (directly and/or indirectly)
- Degradation of confidence in the corporation's ability to guarantee service
- Degradation of confidence in the corporation's ability to build shareholder value and wealth

To mitigate this loss of confidence, a corporation must enact effective BC, DR, and ILM processes and then share information about those processes. The reporting process should link IT expenditures on disaster prevention directly to the effective return on investment (ROI). Through sound implementation and reporting techniques, a corporation can meet and exceed shareholder expectations.

3.0 Business continuity today

The requirements described above do not identify whether a BC and/or a DR solution is needed. In fact, beyond defining their availability requirements to the IT department, most end users neither know nor care that a BC and/or a DR solution is in place. And while few understand the difference between these disciplines, it is important that corporate decision makers are familiar with these differences.

3.1 Business continuity vs. disaster recovery

The primary difference between BC and DR is the speed of recovery, often described as "availability."

The primary difference between BC and DR is the speed of recovery, often described as "availability." BC has a recovery speed that can be almost imperceptible. That is, disruption or interruption can be so brief that an end user may not even notice that an application, server, or storage array failed.

This speed of recovery is enabled through in-chassis operating system and application redundancy, server clusters, traffic management software, redundant paths and network components, redundant arrays and components, and tools that provide continuous backup. Continuous backup implies that data sets are frequently copied and can be restored in a seamless, automated fashion. Technologies that support continuous backup have only been available for the last few years. The most reliable and sophisticated of these products incorporate file system and archival techniques, are highly scalable, utilize multiple backup media and devices, and — most importantly — provide user and application data upon request, without special restore commands.

Typically, BC solution costs are rather high. In extreme cases where availability must be guaranteed, these infrastructure components are sometimes fully redundant and allowed only 50% peak utilization. In cases where human life can be at risk, typical configurations are triple-redundant and the relative peak utilization is closer to 33%. In addition, each of these components will also typically have redundant internal components.

BC solutions can be implemented locally (all components within one building or campus), or they can span vast geographic distances (metropolitan, national, or international dispersion of computing centers). Of course, the further these distances, the longer the recovery time, due to such fundamental limitations as the speed of light in fiber optic cabling. Recovery can be dependent upon an application's ability to sustain an interruption at one site and seamlessly turn over functions to another site. Applications may also exhibit limitations, such as synchronous processing in which an application typically has to wait for an operation or task to complete before it begins another process.

3.2 "Real time" vs. "down time"

Whereas BC strives to provide uninterrupted computing services, DR recovers these services within a planned downtime. Key to an effectively planned DR solution is the idea that users, applications, their data, and customers can wait a prescribed amount of time. Diverging from this plan can have business viability and longevity impact due to customer service disruption and shareholder loss of confidence.

DR typically focuses on providing an alternative computing environment that may not be a complete duplicate of the original. Frequently, DR plans make a clear distinction among applications and data that must be made available first, second, third, and so forth. In some cases, based upon cost considerations, a secondary site may only provide computing resources for the most critical applications. Additional acquisition plans are then implemented should the primary site and its data take longer to repair or restore than anticipated.

By its very nature, DR is typically implemented only in case of a "disaster." If each computing resource failure covered by BC is considered a "mini-disaster," it stands to reason that an organization would activate a DR plan as the result of something major.

Both BC and DR disciplines must have well-tested data management processes, procedures, and tools. Data sets change, expand, and can grow stale; therefore, a corporation must be prepared to handle these various conditions. BC, for example, requires continuous data availability, but even for this type of implementation, copies of this data must be made, at a minimum, to meet DR requirements. DR requires that data be stored and moved, plus be available at the remote computing site. Both of these disciplines require data handing schema that reflect access and availability requirements, and they both must sustain operations without negative external perception from the customer, the industry, and/or from a corporation's shareholders.

Scoping and planning are the keys to an effective combined BC-DR strategy. And while ILM is not BC or DR, the discipline is definitely molded by these requirements and, in fact, takes advantage of BC and DR data classifications.

ILM processes today must adhere to new corporate governance requirements that dramatically increase the useful lifetime of business data.

4.0 Major ILM characteristics

ILM is not a new concept; businesses perform basic ILM-related tasks every day, whether they are aware of it or not. Such tasks include organizing, saving, and using business data, as well as determining how data will be effectively and efficiently accessed. How ILM differs from five years ago stems from the new corporate governance requirements that dramatically increase the useful lifetime of this data. In addition, data now resides in many different locations — from laptops to enterprise storage infrastructures to vast archival centers — further compounding the problem of effective and efficient access. Lastly, preserving data for different retention periods (depending on the nature and value of the data) can become very expensive in terms of storage infrastructure and personnel.

4.1 Government-based ILM drivers and beyond

Sarbanes-Oxley regulations represent only one of several drivers associated with ILM. Others include:

- Leveraging investment in existing storage infrastructure (repurposing hardware as newer technology is procured)
- Utilizing "data profiling" to further understand how data is used and what data can be archived to secondary and tertiary storage media (beyond BC and DR classifications)
- Standardization of data types to reduce the potential complexity of restoration infrastructures, such as using Extensible Markup Language (XML) as opposed to proprietary e-mail or database formats

While addressing these drivers can provide corporations significant benefits (in addition to assisting them with compliance regulations), the fact remains that true compliance can only result from an educated workforce.

Regulations were put into place in part because of the emergence and changing nature of e-mail. Corporate e-mail exchanges that included unethical and/or illegal statements have forced the U.S. government to institute rigorous information retention and tracking rules. To fully adopt these rules and make information retention and tracking feasible, corporations must ensure that employees understand the issues, utilize effective data tagging tools and mechanisms, and lastly, are held accountable for the information they exchange.

Together, a robust infrastructure and an educated workforce allow corporations to fully gain the benefits mentioned above and meet government requirements.

4.2 Data classification

There is no question that corporate data is valuable, but there is so much of it that you must carefully organize this information asset to make it at all manageable. To do this, you begin by asking questions, such as:

- How do I determine the relative value of the data?
- How will the value change with time and/or classification?
- Will the value impact where the data is physically stored?

These are the questions organizations must address and resolve to manage information according to their ILM policies. Resolving these issues will help generate requirements for appropriate management tools. The following table represents a sample data classification and data placement profile according to corporate priorities. Business environments must define their own priorities and customize this table according to their own categories and available infrastructure.

Table 1. Data classification example

Data classification example

Priority	Category	Subcategory	Cached	Online	Nearline	Nearline archived	Offline archived	Offline vaulted
1	Business value							
		Customer applications and data	x	x	x	x	x	x
		Customer support (data produced internally/ externally, accessed, processed, reporting, updated)	x	x	x	x	x	x
		Operations (business, technical, operational data)		x	x	x	x	
		Legal (contracts, correspondence, communications, compliance)			x	x	x	x
		Internal (payroll, benefits, etc.)		x	x	x	x	x
2	Access frequency							
		Day-to-day	x	x	x	x		
		Monthly			x	x	x	x
		Periodic within or at yearly intervals				x	x	x
		Longer than 1 year						x
3	Cost and risk							
		Utilize current infrastructure (leverage investment in hardware, software, personnel) (risk is overtaxing resources)	x	x				x
		Outsourced infrastructure (includes overtaxing and access security)		x				x
		Shared responsibility (supply chain partners)		x				x
4	Retention period/ compliance policy							
		Day-to-day	x	x	x	x		
		Monthly		x	x	x		x
		Periodic within or at yearly intervals				x	x	x
		1, 3, 5, ... year increments					x	x

Prior to developing a specific ILM strategy for your organization, here are some things to do and consider:

- *Audit current data challenges*
- *Do a cost-benefit analysis*
- *Create a flexible ILM culture*
- *Leverage existing investments*
- *Ensure heterogeneous management*

5.0 Pre-Implementation ILM considerations

As evident by the variety of choices in Table 1, the development of corporate data classification schemes are unique for each corporation and even individual business units. How do you know what questions to ask? Before delving into a detailed approach for developing an ILM strategy, there are several general considerations to take into account.

These considerations include auditing to identify existing data-related challenges, performing a cost-benefit analysis to justify expenses, creating a flexible ILM culture to handle new applications and data types, leveraging existing infrastructure investments, and ensuring that management schemes are as robust and long-lived as the data you intend to manage.

5.1 Audit current data availability and management

The relative maturity of your current environment can impact the level of effort needed to become compliant. Look at user interactions and success in light of the current policies, processes, procedures, and tools used to protect data. Highlight the challenges and document mitigation methods. Unless the current environment is at extreme risk, delay changes until you can define an ILM strategy and plan for how requirements associated with this strategy can help fill in potential data protection holes.

Similarly, audit DR and BC plans and mechanisms, document current risks, and incorporate mitigation plans into the new ILM strategy. Note that ILM strategies provide focus for DR and BC — these strategies help improve these other disciplines while not replacing either of them.

5.2 Measure risk against compliance and expense

As was stated earlier, computing environments that require high levels of availability are generally very expensive. (Of course, system expense doesn't necessarily translate into higher availability.) However, adding ILM requirements can drive infrastructure cost to levels that would bankrupt most companies. Therefore, you must implement a reasonable approach that utilizes data classifications, accessibility, and methodology.

The three critical points to address in mitigating the risk associated with over-spending are ILM business requirements, ILM impact, and ILM maintenance. Requirements establish what data sets are the most critical (classification), impact scopes the problem in terms of "what if" scenarios (access, options, cause and effect), and maintenance mechanisms enforce the priorities that were initially established over prescribed retention timeframes.

In the extreme case in which all data must be kept for an indefinite period of time, the most important requirements are that the data be accessible and usable. And the only way to guarantee access and usability without retaining every piece of original hardware, software, and other operational component is data format normalization.

XML has emerged as a de facto standard for exchanging information between disparate processes, applications, and data types. In Table 1, for example, particular data was slated to be "online or offline" and finally "vaulted." This information can be encapsulated in an XML text-based data type that identifies the original source and level of priority. While normalizing data into text files does not represent the most efficient database usage, the intent was to be able to provide access and use without having to keep the original infrastructure (and associated cost) that originally produced the data.

5.3 Plan ahead

Thus far, this discussion has focused on implementing ILM "over" an existing infrastructure. Clearly, as new and modified applications and corresponding data emerge, organizations must apply additional ILM requirements proactively.

Typically, corporations use a combination of business, technical, and operational criteria to help define requirements for a new application. These criteria may be formally captured in documents, such as product requirements, request for proposals, or change requests, to name a few. Adding ILM requirements to these documents helps preserve the corporate culture and enforces development of education plans for users of these new applications and the data to be produced. This approach assumes that the corporation has already developed an ILM strategy and has made an investment in extending the data management environment.

5.4 Leverage customer investments

Causing as little disruption as possible to current data protection schema and mechanisms is key to a successful ILM strategy implementation. This requires organizations to fully audit current data protection processes (such as availability, backup, and restore), procedures, and tool implementations. But more importantly, organizations must understand the workflow of the current environment. Rather than create or aggravate single points of failure, ILM mechanisms should attempt to better utilize existing infrastructure components.

For example, once data value is known, retention periods identified, and metadata records kept, the next step is to place data according to these requirements. Access and availability requirements help to determine the "rating" that each storage media and device will be assigned. This begins to identify what may be called a "storage media cascade" that allows corporations to repurpose hardware according to its relative availability rating. This new classification then extends to how these various cascade groups and individual devices can be managed.

5.5 Ensure heterogeneous management

Management of the BC, DR, and ILM environment requires an end-to-end approach and tools that will last as long as the data. Although it would be ideal to have to manage only one vendor's products, many computing environments are designed so that vendors are redundant — that is, solutions that provide near-parity capabilities may be procured from two or more vendors. Given this, it becomes extremely important that BC, DR, and ILM technologies and products — and their relative management — be selected based upon adherence to industry standards. Exceptions, of course, are inevitable either because standards are not yet defined or the standard is not extensive enough to ensure the availability requirements stipulated above. For example, it is counter-productive to implement a fully redundant BC solution if the redundancy doesn't include management as well as the data.

Therefore, when faced with multiple-vendor environments, heterogeneous management becomes a base requirement. When vendors do not provide these types of end-to-end tools — or perhaps provide tools that manage only their own products — organizations may employ a "manager of managers" approach.

Enterprise management vendors have implemented this type of management schema for many years, based upon de-facto standards like the Simple Network Management Protocol (SNMP). However, although standards from organizations such as the Storage Network Industry Association (SNIA) are emerging, they have not yet incorporated U.S. government compliance requirements into their published standards.

As a last note, the idea that a vendor's tools must last as long as the data has two aspects:

- The longevity of the technology itself — The product must have an effective backwards compatibility capability for each new release and a fairly long-reaching road map.
- The viability of the vendor — The vendor must have plans for the future, both technically and financially, that form the basis for assessing the vendor's longevity.

6.0 Creating an effective BC/DR/ILM strategy

Most corporations have invested significant amounts of time and resources to BC and DR. This investment can now be leveraged for effective ILM deployment.

Time is the common link among our three disciplines: percentage uptime for BC, planned downtime for DR, and amount of retention time for each data set or type for ILM. These duration requirements form the basis for what needs to be done (and when) to ensure availability, access, and compliance.

Fortunately, most corporations have invested significant amounts of time, resources, and capital into BC and DR. Many, for example, have extended their business boundaries to include their suppliers and consumers, thereby ensuring that the entire supply-chain can sustain interruptions. All of this planning and investment can now be leveraged for ILM deployment — not only to shorten the time required to deploy and execute this strategy, but to help ensure ILM receives the same level of attention.

6.1 Leveraging BC and DR for ILM

So what are some specific ways in which organizations can leverage BC and DR investments for ILM? Here are summaries of key strategies:

- **Data classification** — Repurpose statements from BC and DR policies, such as "this application and data are: very important, important, not as important, and so on"
- **User interaction** — Understand the two possible scenarios for how users relate to ILM. Retention and retention procedures and tools may be:
  - Hidden from your users (this applies to typical corporate data)
  - Not hidden from your users because they must decide how important their data is and how quickly they need to access it, as dictated by emerging requirements
- **Infrastructure** — Identify investment in BC and DR infrastructures, then leverage it by first analyzing the gaps between availability, accessibility, and compliance
- **Data linkages** — Make certain that BC/DR strategies can also support ILM compliance requirements; for example, extend and expand "data labeling" techniques from BC and DR to ILM
- **Supply chain considerations** — If the BC and DR scope include business suppliers and consumers, include them in the ILM strategy development as well
- **Vendor compliance** — Because government regulations apply to all businesses in the U.S., ensure your vendors are compliant, have certifications, or are members of industry associations to create ILM standards. Ultimately, "how much skin in the game" have your vendors contributed?

Additional considerations:

- **Auditing** — Conduct independent audits
- **Testing** — Test the combination of BC/DR/ILM disciplines
- **Expertise** — ILM is still nascent; make sure your consultants are "experts"
- **Tools** — Ensure appropriate tools are available
- **Mitigating risk** — Treat this as the Y2K of the mid-2000s; form corporate user groups to better define requirements within and around your specific industry
- **Changing culture** — Educate all of your users that ILM is inevitable
- **Implementation** — It's not "all or none;" you can phase in an ILM strategy, but make sure it's government-compliant

The final word is that although waiting to implement is okay, waiting to plan is disastrous.

7.0 Conclusion

Information lifecycle management, business continuity, and disaster recovery all have the same basic goal: "keep the business running." How and when they are implemented varies, of course, but they have common factors.

By closely analyzing your needs at the user, corporate, and stockholder levels (as well as federal requirements), you can identify common ILM, BC, and DR factors that will allow for a more comprehensive, cost-effective management scheme. With planning, you can leverage existing time and material investments in BC and DR to more efficiently drive your ILM strategies.

8.0 Appendix: BC, DR, and ILM characteristics summary

Table 2 below summarizes various facets of BC, DR, and ILM disciplines. Of course, many details are not represented, such as the mechanisms used to support each driver. However, taken in light of the considerations described in this paper, this table provides a starting point for corporations that want to begin the journey towards defining a BC/DR/ILM strategy.

Table 2. Comparison of BC, DR, and ILM characteristics

Comparison of BC, DR, and ILM characteristics

Driver	BC characteristics	DR characteristics	ILM characteristics	How the characteristics relate
Availability	No downtime	Prescribed	Combination	Dependent on data valuation
Access	Defined, secure	Defined, secured	Not well-defined	May require security policy modifications
Economics	High-cost	Medium to high cost	High now	Requires analysis and tools
Risk	Risk-mitigated	Risk-mitigated	Not well-defined	Requires government standards
Value	Defined	Defined	No well-defined	BC & DR provide value basis; government requirements and education needed
Retention	Audits required	Audits required	New policies required	Modifications of current policies may be needed; minimum of key characteristics, such as metadata placement across BC, DR, and ILM, should be supplied

Driver	BC characteristics	DR characteristics	ILM characteristics	How the characteristics relate
Infrastructure	End-to-end redundancy	Most critical applications and data, then acquisition plans	First-generation data tagging will be manual, next-generation may use XML	Leverage investment with little modification of current implementation; use XML data types to comply
Government requirements	HIPAA and SEC requirements	HIPAA and SEC requirements	Add retention and data tagging	Privacy and financial reporting information sensitivity must be taken into account for new policies
Reports	Internal, real-time	Internal, near real-time, post DR	Roll-up reports on data location with brief description	Metadata reports must be generated on an ad-hoc basis
Testing	Testing of each component must be executed during off-time	Simulated DR failovers conducted periodically (quarterly)	Should be included as part of both BC and DR testing	Critical testing for ILM includes metadata reporting, data location, and physical storage cascade testing
Audits	Non-disruptive review of policies, processes, procedures, and reporting	Non-disruptive review of policies, processes, procedures, and reporting	Non-disruptive review of policies, processes, procedures, and reporting	Audits are driven by corporate policies; these policies must be kept up to date on what is business-critical
ROI	Survey customer and customer support organizations on relative impact vs. availability; break-even and productivity analysis required	Survey customer and customer support organizations on relative impact vs. availability; break-even and productivity analysis required	Leverage BC and DR investments; break-even vs. risk analysis for non-compliance required	ROI will ultimately be driven by revenue, cost, and profit analysis and comparisons; “best-in-class” implementations gain intangible benefit from good public relations that could become a competitive differentiator

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN **Web** sun.com

JT 0021 A 06/06