HCL

LET'S TALK APPLICATIONS

BUSINESS ANALYTICS SERVICES
KNOW MORE. DO MORE.

THE BUSINESS CASE FOR MANAGING YOUR DATA AND ITS LIFECYCLE FROM END-TO-END

BIG DATA & BUSINESS ANALYTICS

AUTHOR:

COLLIN KLEPFER

DIRECTOR, ILM, BUSINESS ANALYTICS SERVICES

WHAT IS INFORMATION LIFECYCLE MANAGEMENT (ILM)?

ILM is a subtle concept. Corporations can grow storage, grow applications, grow data, and grow data governance need in every corner of the globe, over very long periods of time and never stop to take the time to properly plan for or predict the end-of-life of that data. Or the performance hit from accumulating too much data, or the exit strategy for cleaning up that data. Lack of planning and/or a cohesive program for data can be expensive and cumbersome for some. It might not be you at the moment, but someone will ultimately need to deal with old data and its ultimate resting place - when it is no longer considered 'active' data. Unfortunately, in the era of multi-year lawsuits, exploding regulations, and the need to dissect corporate history under scrutiny, it has become a big deal to establish and execute a proper ILM strategy that covers risk and provides a cost effective long-term strategy for retaining only the critical data. And that meets the companies legal, compliance, and reporting requirements.

Additionally, with the declining costs of storage over the years, the tendency has been to 'keep everything'. This, of course, becomes a problem that only pushes the problem out to a later data and will require even larger amounts of storage to continue to push the issue out into the future. The unfortunate result of the 'head in the sand' approach becomes system and application performance losing control of finding, accessing, or sorting through huge amounts of storage and data – or producing the data as evidence in a lawsuit or IRS inquiry, years after the fact.

In the IT application world, control over performance is often seen as a 'hardware problem', which is on a certain level. But hardware additions can also have their time and place as a solution to a performance issue. When the issue is 'too much data', throwing hardware at the problem for most times should not be the first choice. The first choice should be to determine where the bottlenecks originate (verify that large data sets and their processing are the problem) and then create effective strategies for reducing the size of the data and finding a resting place for the 'inactive data' that is no longer needed in the applications' environments (plus, retaining access to that data.)

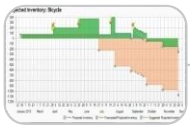
So, let's focus on the inactive data concept for a moment. What is data that is no longer 'active'? Typically, this is the data that no longer has any potential for update (from a system perspective). Keep in mind that not all "inactive data" is created equal, but this is the definition that drives the base concept of ILM – that there is plenty of data generated in a corporate environment (or at home for that matter) that needs to be dealt with because it now has either ZERO or reporting-only value to the company. The tricky part then becomes both determining what data is truly 'inactive' and what data needs to be kept for a reporting need (even though it may be considered inactive), and then what can truly be 'disposed' for the right reasons, at the right time. This quest takes the right technology, the right expertise, and the right processes in place to drive a program that sometimes, in the final analysis, may be as important as the effort to implement new systems and applications. Getting rid of old data and old application often proves to be a very difficult undertaking that requires the experts and skills to accomplish.

Key ILM Solutions

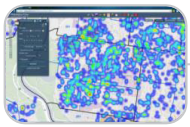
Database Archiving – A system of defining and extracting structured database data and transactions that preserves relational integrity and allows for the effect recall of the data when and where needed

Application Retirement – Transformation of entire OBSOLETE applications and its data into a form that will allow for reduced license and storage costs, while providing for long-term access and reporting on that data

Test Data Management – Integrated and coordinated transformation of live, production transactions into intelligent test data that adequately represents the needs and requirements of robust corporate test environments. The product suite includes the 'connectors' and intelligent technology components to select 'subsets' of key business data and the transforming and 'masking' that data in a coordinated fashion



Secured Data Privacy – A short implementation time technology that adds a high degree of control over presentation formats and access rights of sensitive and private personal or corporate data



Database Partitioning and Performance Control – The ability to smartly define how and where related application transactions are stored in storage partitions, thus giving the technical staff complete control over its movement and long-term access performance

Lastly, ILM does include the concept of 'unstructured data' and 'semi-structured' data. Here is the unstructured data definition from Wikipedia:

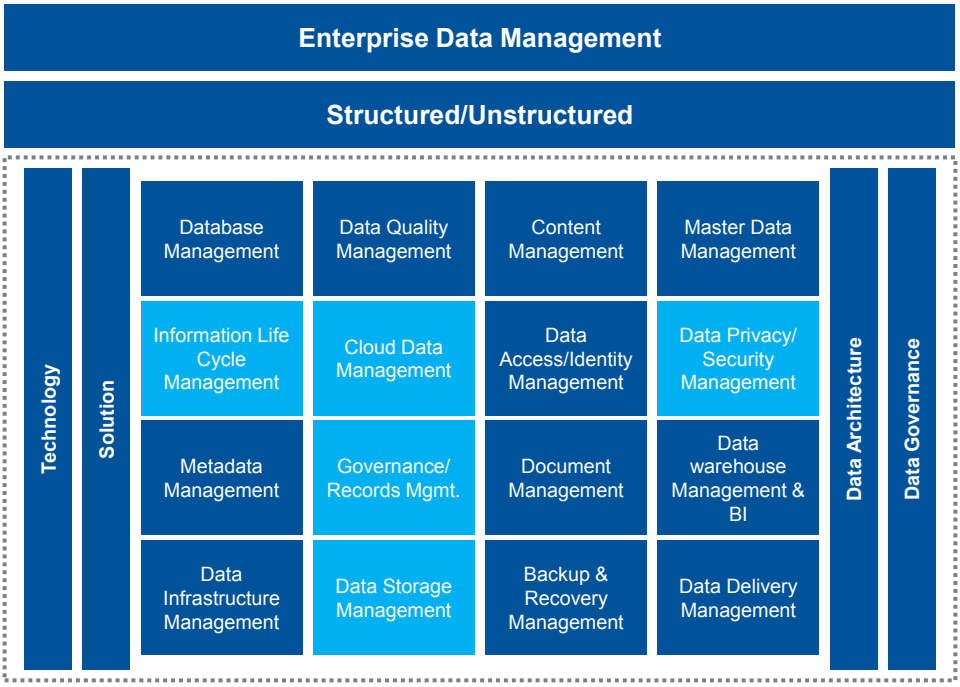
"Unstructured Data (or unstructured information) refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional computer programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents."

This is the area that is most easy to deal with from a lifecycle management perspective – because often it is just a matter of 'throwing money' at the problem and just finding the cheapest cost-of-storage. It also allows for the 'out-of-sight, out-of-mind' approach to archive old data. Ask any of the tape storage, data storage companies, and now Cloud storage vendors. When in doubt, just off-load the data (regardless of how many copies you might have), send it to its new location, and forget about it. And, hopefully, it will become someone else's problem. Unstructured data comes with its own challenges (mostly size and version related) and requires a different approach than its close cousin, 'semi-structured data'.

Semi-structured data is a form of structured data that does not conform to the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure.

Semi-structured data can be the most problematic due to the fact that it often requires an expert or someone with the knowledge of the history of data to interpret or access the data. The good news here is that most companies create the bulk of their critical data in structured or unstructured formats. And, most companies do not generally rely on semi-structured formats for their mission-critical applications or day-to-day operations.

Additionally, ILM touches on virtually all areas of 'Enterprise Data Management', which typically encompasses the entire spectrum of data creation and on-going management, but may not take a holistic approach to aging data and its movement (archival) within the IT infrastructure. The following chart illustrates several different touch-points for ILM within the EDM landscape.



Another subtly of the ILM capabilities and corporate need to control data and its age within the infrastructure is – the creation and dissemination of copies of sensitive production data and its protection to outside use and view. Obviously, there is imbedded sensitive data within the corporate databases and perhaps unstructured data. It is both legally and ethically required to keep this data safeguarded and under corporate governance.* (See additional HCL whitepaper for ILM and ILG – By Bill Tolson, recognized Data Governance expert, attached with this whitepaper.) ILM provides many good options to control or obfuscate that sensitive data and provide options for creating smaller 'subsets' of data under controlled conditions and the ability to modify that data to useful 'near copies' of it to provide valid test environments, but protect the data that should not be released.

So, to summarize, ILM contains the following key approaches to managing older, obsolete data that should be protected and removed on a controlled basis:

- Database Archival
- Application Retirement

- Test Data Management
- Data Privacy
- Application Performance Control
- Data Governance

SO WHAT IS THE STATE OF 'BEST PRACTICES' IN ILM?

ILM has been developing over the past decade or so as a solution to the growing problem of exploding data and no real, cohesive approach to dealing with the problem on a centralized, controlled basis. There are various companies that have massive problems in one or more of these areas. And even the smaller companies have the same issues on a smaller scale. So, thought leadership in this area is growing quickly and there are new solutions being tried and developed on a daily basis. And, the options to deal with the issues are vastly greater than just 10 years ago.

The concept of 'best practices' and 'best options' becomes a very interesting discussion, very quickly. Unfortunately, there is no 'one size fits all' answer to ILM challenges these days. And, as with most IT related things, the ILM program needs are mostly 'need' and requirements based, so the need to do 'exhaustive analysis' does not go away with ILM initiatives and option discussions. Additionally, because the (structured) ILM technology has grown up around specific technologies – and their issues – the tools in the ILM toolkit are generally not applicable across-the-board. When it comes to unstructured data, they are far more likely to handle 'most' generic forms of data and file types. But, in the structured world, far more services are required to get the true value out of the products. That is not necessarily a bad thing – since it gives you more control over the solution – but it also complicates the time-to-implement and has project cost ramifications – mostly due to the highly complex nature of structured databases and the applications built around them.

The state of 'Data Archive and Application Retirement Centers of Excellence' seems to be leaning toward creating a 'factory approach' for pumping large numbers of applications through the 'factory', while giving specific, defined users limited and restricted access to the data for specific business use-cases, mostly related to compliance and/or long-term reporting requirements. The state of 'Test Data Management' and creating secured and obfuscated test data is generally centered around tools that create consistently created/applied rules and policies. The test data world has evolved from the database archive world – since the logic that allows you to archive structured 'logical transaction'- that won't break the database is also the logic that allows you to create 'subsets of data' relational intact. The actual obfuscation of the data – based on rules and policies – is then easily applied as the data is moved from one DB to the next.

In the 'unstructured world' the management of files, email files, and miscellaneous data types typically centers round:

- De-duplication of redundant copies of data(files)
- Centralized management of the archive

- Compression
- Ability to recover specific versions or copies of files
- Lowering storage cost by controlling where the data is stored
- Ability to quickly search and find specific emails, files, and objects

There are some obvious benefits to all of these. But, the actual intersection of managing structured vs. unstructured data is not very extensive and usually limited to storage devices, at most. Therefore, companies still (generally) end up administering these as two separate technology stacks with different skillsets required. But, there are attempts to merge some of the two types of requirements together. At this point there are still relatively few options other than 'partnerships' between vendors that add peace-of-mind that someone knows and has thought thru the ramifications of both.

GOVERNANCE, REGULATORY COMPLIANCE, AND LEGAL COMPLIANCE (GRC)

One of the huge benefits of ILM and its ability to add sanity and an organized approach to the vast amounts of files, data, and information being created on a daily basis in corporate America is - the ability to govern and prove that you have made every effort necessary to keep and account for all critical data being created. AND PROVE THAT IN COURT. In some companies, this is not as critical as in others. Certainly in the realm of pharmaceuticals and healthcare, the ability to trace and produce a specific audit trail of files, emails, and communications becomes a very relevant discussion and requirement for specific court cases.

There is a need (and now products) that track not-only specific data traces, but also cross-references those traces against one-another to prevent duplication of effort. Another key feature is the ability to place specific 'legal hold' flags on data that is needed somewhere else. This prevents the resolution of one legal issue from releasing the same data to be deleted without 'all' requirements having been met. This, of course, can be huge since the volume of lawsuits in some companies can be daunting and requires massive 'eDiscovery' efforts to find and produce specific trails of data. (See the Bill Tolson Whitepaper attached related to eDiscovery and Data Governance.)

Some of the other issues related to Data Governance and the associated risks are best summarized in the following chart.

Common GRC Challenges in the Client Space

Need for more 'Standardized Policies and Procedures'

- No appropriate standard framework for audit and compliance activities
- Inconsistent audit plans, work paper methodologies, etc.

Lack of Real Time Visibility and Touch-points with Critical Data

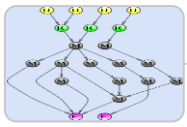
- Transactions occurring daily within the business
- Fields, columns, or configurations that are changed by users

Non-Standard Information

- Multiple legacy systems with disparate uses and different architectures
- No common platform for reporting and consolidation

Heavy Cost of Compliance Activities

- Cumbersome and manual process to audit
- Many man hours 'chasing paper'



Need for clearer 'Roles and Responsibilities'

- Roles within the business are unclear
- Responsibility for audit and accountability for system functions are blurred

Obviously, the need for GRC programs runs parallel to an ILM program - and is, from a legal stand-point, almost an absolute requirement. The need is for a holistic, end-to-end program and definition of where your data is, how it got there, and the history of how it was created. This implies you must have all three types of data covered, along with the ability to keep and reproduce those in a consistent and reliable program. (The 3 areas being: Structured, semi-structured, and unstructured data.)

DEFENSIBLE DISPOSITION

In recent years, the timing and 'defense-ability' of deleting data (eliminating it from the environment) has become an issue. The ability to prove that the regulatory and compliance aspects of data requires that companies both know – and can prove that the data was eligible to be deleted and that it was not encumbered by any outside events - such as, legal issues and litigation holds on the data. Therefore, it has become increasingly hard to justify pure 'deletion' of data. And fewer companies have taken the time effort to prove they can 'defensibly dispose' off that data. The default position in many cases is to 'keep it all', or at least move it to somewhere they may have long-term access to it (tape?). The sensitivity of the data – as it applies to deletion – also comes into play. If the data is not part of the mainstream flow of corporate applications and does not need to be retained, then it becomes eligible for deletion when it is no longer needed by the company. Transaction data – especially financial data – does not. It must adhere to the laws and regulations of the governing country/state.

Much of the disposal of data is driven by the 'General Retention Schedule' that some companies publish and as a result of exhaustive legal review and documented process – most times on a local-legal basis. Since laws and regulations related to data can

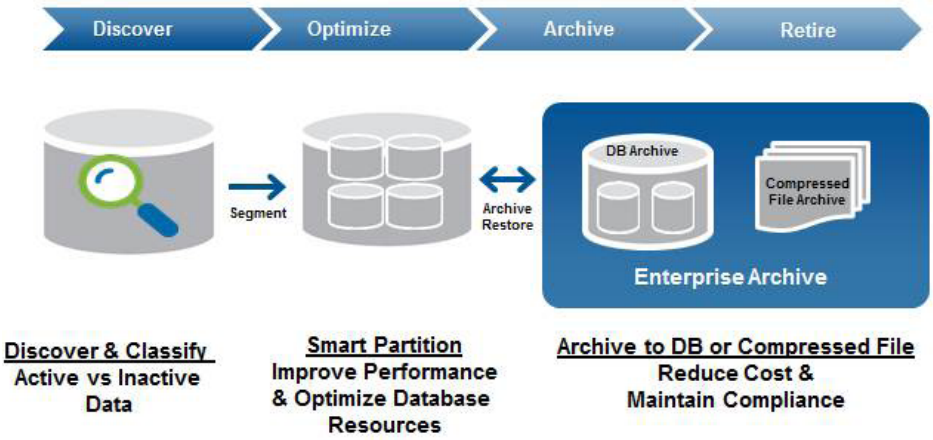
vary from country-to-country, the tracking and adherence to those laws/regulations becomes a corporate issue. And the disposal of data by the corporation also becomes a local-legal and corporate issue. If the company has taken the time to create retention policies that are solid, then once the data meets the policy, then it may be deleted (although many companies do not).

So, the defense-ability of getting rid of data falls back to the issue of 'deletion' or 'moving data to a cheaper tier of storage'. Needless to say, this can be an important undertaking and should not be taken lightly. A solid ILM strategy encompasses and works WITH governance, risk, and compliance rules and should not be seen as a stand-alone activity. One lawsuit that cannot produce the proper records or burden of proof could easily justify the cost of having a COMPLETE ILM and Governance program in place!

STORAGE AND STORAGE TIERS

At the heart of most Information Lifecycle Management programs is the issue of 'storage' and the cost of that storage. With many companies now in the PETA-byte storage capacity club, often there is both a need and, hopefully, a program in place to drive a cohesive strategy of where structured, semi-structured, and unstructured data resides…and the movement of that data between tiers of storage or storage environments. Additionally, application performance and query performance against data in the applications and the data in the storage architecture are key discussions points – from an ILM perspective and an end-user experience perspective. Obviously, if data is moved into an archive and that archive performances significantly differently from the end-users expectations, there is room for serious discussion around ILM governance of the data and technology options to help resolve any issues.

Some of the newer ILM components use the concept of storage 'partitions' to help manage and move large blocks of archived or retired data. The management of those blocks of data is absolutely at the fore-front of ILM and storage-tier discussions. So, when setting up ILM programs, tiered-storage architectures, disposal policies, and retention schedules, the discussion of storage and the limitless options that are key to the success of the program and its ultimate cost to the company.



INTERNAL ACCESS AND SEARCHES

The last important concept that becomes germane to any enterprise's Information Lifecycle Management discussion is the issue of access and how that data needs to be used internally to the corporation. Obviously, if data is not 'required' to be kept - nor 'relevant' to any future activity - then it should be deleted from the environment and not be kept as part of any program - ILM or otherwise. But, in modern corporations, that is almost NEVER true. It needs to be kept until it is no longer needed – from a legal perspective and from a usefulness perspective. And, probably not a day longer!

The whole point to of a cohesive ILM program is provide the right level of access, governance, storage, and risk reduction to meet the needs of the enterprise creating the ILM program. If any of those requirements are not met, then the program needs to be re-evaluated and probably re-designed. So, to summarize…

Key Objectives of a Successful ILM Program

- Holistic, end-to-end management of all company data
- Providing the RIGHT level of data classification and moving that data when it is time
- Reducing the risk that comes from NOT being able to provide timely and required access to older data
- Creation of the right LOCATION for data, based on its age and access requirements
- Covering ALL legal and compliance requirements for all company data
- Creating a 'best practices' environment that provides organized, fast access to any data, any time, or being able to defend why it is gone
- Archiving data when it is time to go
- Retiring old applications and its data to reduce hardware cost and maintenance cost
- Providing the right (and legally required) level of data governance and compliance for your enterprise
- Being able to track and defend why you are moving or eliminating data from your architecture
- Provide the peace-of-mind that one gets doing the 'right things, for the right reasons'

CONCLUSION AND NEXT STEPS

There are many options when it comes to information lifecycle management and managing your data in a way that covers risk and retains the right level of access – if needed. There are far fewer options when it comes down to creating an end-to-end ILM strategy that provides the peace-of-mind of knowing you have done 'everything' possible to optimize your approach to ILM and the data it governs. Unfortunately, the data management industry is still forming opinions and approaches for the vast number of ILM choices in the industry. What this means to you and your efforts to control the data growth in the environment is – you need to pick the right partner(s),

the right tools, and the right approach – and make sure that those choices work the way you expect and the way they are designed.

As next steps go, picking the right partner(s) to help you analyze your ILM choices is probably as critical to your success as picking the right tools and right approach. If you don't start with people that know and can assist the critical decisions around ILM strategy (and have the required experience), then your chances at succeeding in putting together an approach that covers all your risks and provides ultimate data access is greatly diminished. Start with a solid partner and a solid approach and you will sleep well knowing you have 'done the right things for the right reasons'!

HCL EXECUTIVE BIOGRAPHY

COLLIN KLEPFER

Director, ILM, Business Analytics Services

Collin is an Information Lifecycle Management (ILM) and Data Governance industry Subject Matter Expert (SME) on current best-practices around data archiving, application retirement, test data management, data governance, and data masking. With 25+ years of IT industry experience around the globe, he comes with deep background in delivering projects and consulting services to a long list of international clients and government agencies.

ABOUT HCL

About HCL Technologies

HCL Technologies is a leading global IT services company working with clients in the areas that impact and redefine the core of their businesses. Since its emergence on the global landscape, and after its IPO in 1999, HCL has focused on 'transformational outsourcing', underlined by innovation and value creation, offering an integrated portfolio of services including software-led IT solutions, remote infrastructure management, engineering and R&D services and business services. HCL leverages its extensive global offshore infrastructure and network of offices in 31 countries to provide holistic, multi-service delivery in key industry verticals including Financial Services, Manufacturing, Consumer Services, Public Services and Healthcare & Life Sciences. HCL takes pride in its philosophy of 'Employees First, Customers Second' which empowers its 91,691 transformers to create real value for customers. HCL Technologies, along with its subsidiaries, had consolidated revenues of US$ 5.4 billion, for the Financial Year ended as on 30th June 2014. For more information, please visit **www.hcltech.com**

About HCL Enterprise

HCL is a $6.5 billion leading global technology and IT enterprise comprising two companies listed in India – HCL Technologies and HCL Infosystems. Founded in 1976, HCL is one of India's original IT garage start-ups. A pioneer of modern computing, HCL is a global transformational enterprise today. Its range of offerings includes product engineering, custom & package applications, BPO, IT infrastructure services, IT hardware, systems integration, and distribution of information and communications technology (ICT) products across a wide range of focused industry verticals. The HCL team consists of over 95,000 professionals of diverse nationalities, who operate from 31 countries including over 505 points of presence in India. HCL has partnerships with several leading global 1000 firms, including leading IT and technology firms. For more information, please visit **www.hcl.com**

www.hcltech.com

Hello there! I am an Ideapreneur. I believe that sustainable business outcomes are driven by relationships nurtured through values like trust, transparency and flexibility. I respect the contract, but believe in going beyond through collaboration, applied innovation and new generation partnership models that put your interest above everything else. Right now 95,000 Ideapreneurs are in a Relationship Beyond the Contract™ with 500 customers in 31 countries. **How can I help you?**

Relationship™ BEYOND THE CONTRACT

HCL