THE BIG DATA READINESS STUDY

HOW DATA LIFECYCLE MANAGEMENT AND STORAGE ARE ENABLING BIG DATA

Commissioned by Vodafone

customer experience foundation

vodafone

ABOUT THIS REPORT

To understand the role that Data Lifecycle Management and Storage can play with respect to harnessing the benefits of Big Data, Vodafone commissioned the Customer Experience Foundation and Research Now to conduct some research.

In August and September 2012 Research Now's panel was used to survey 150 IT Professionals working within the UK's largest organisations.

In order to fully understand the results of the online survey and to add depth and insight to the results, the Customer Experience Foundation conducted in-depth interviews during September 2012 with 25 Senior IT Directors in large enterprises.

ABOUT THE AUTHOR

Morris Pentel, Chairman, Customer Experience Foundation

Morris Pentel runs one of the most respected consulting businesses in the UK, having completed successful work for some of the largest institutions in the world. He has been involved in the delivery of many significant innovations in technology over the last 30 years, winning a number of prizes. He has also conducted a number of major research programmes looking at best practice in the use of technology.

ABOUT VODAFONE

Vodafone is a leading global communications service provider, delivering a wide range of high-quality managed voice, data, hosting and IP-based services and applications.

This report was commissioned as part of our storage infrastructure proposition. We offer products for shared primary storage, cloud storage, data protection and dedicated storage; each adapted to your specific needs, with a choice of service packages.

An optimised storage infrastructure created by Vodafone allows you to effectively:

- manage your data lifecycle
- reduce costs
- improve business agility
- safeguard business continuity
- ensure compliance
- help your employees be more efficient by giving them easy, timely access to the data they need

RESEARCH RATIONALE

UNDERSTANDING THE ROLE OF DATA LIFECYCLE MANAGEMENT AND STORAGE WITH RESPECT TO BIG DATA

There is a great deal of talk in the Boardrooms of Britain about Big Data. It's a subject that has also captured the business headlines more and more over the last few years and large organisations have got the bug. It is probably one of the fastest growing areas of IT spend currently. Highly visible successful results from early adopters mean that there is real momentum in the race to integrate Big Data into a core platform for large organisations.

Big Data is seen as a key enabler in challenging market conditions and as a result, most organisations are now investing. It provides a missing link between data and insight that did not exist before. It is about delivering game changing insight to organisations, that can be applied on a case-by-case basis to produce real impact. It adds a new level of insight to decision making processes that had not been possible before and that drives business value in new ways.

The amount of data created globally in 2011 was 1.8 zettabytes (according to IDC). That is 1.8 trillion gigabytes of data. It is also widely reported that the volume of data is doubling every two years. These market forecasts are on such a grand scale that they are difficult for most of us to take in.

This research report considers the increasing need to focus on the efficient management of data as a Big Data enabler. Specifically, it considers how optimisation strategies such as Data Lifecycle Management (DLM) and alternative storage solutions can be used to create a data efficient organisation, ready to accommodate the ever-increasing demands for rapid storage processing and retrieval of massive volumes of data.

METHODOLOGY

This report is based on three research streams:

1. **Desk research** – an extensive desk research programme. Where forecasts and trends have been provided, multiple sources of corroboration were sought.

2. **Quantitative research** –an online survey amongst 125 IT decision-makers working in large organisations (defined as organisations having at least 500 servers). This provides a broadly representative picture of results for the UK's largest organisations. The profile of respondents according to size of the organisation they work for and their age is shown below:

Number of employees	
500–999	**4%**
1,000–4,999	**21%**
5,000 or more	**75%**

Age of respondent	
18–29	**9%**
30–39	**29%**
40–49	**31%**
50–59	**27%**
60+	**4%**

3. **In-depth interviews** – detailed interviews lasting an average of 50 minutes held with 25 IT Directors from major organisations across a number of sectors including financial services, media, retail, travel and utilities.

The topics covered in the online survey and the in-depth interviews included:

- Big Data and its impact on data management
- the value and cost of data
- the importance of DLM
- levels of maturity with respect to DLM
- use of DLM technologies

EXECUTIVE SUMMARY

This report looks at growth of data and how DLM and alternative storage solutions can help to enable organisations to derive value from Big Data.

It highlights some of the key challenges and opportunities associated with data growth. It examines the way in which IT organisations within large enterprises are benefitting from DLM as an optimisation strategy.

With the size of data within organisations expected to grow by factors rather than percentages, data efficiency is a key cost driver providing operational performance and cost benefits. We identify three of the top cost-saving storage solutions and share some best practice tips from those at the sharp end of DLM.

This report is based on a research programme comprising a significant desk research exercise, an online survey amongst 125 senior IT professionals and a series of in-depth interviews with 25 IT Directors.

RESEARCH HIGHLIGHTS

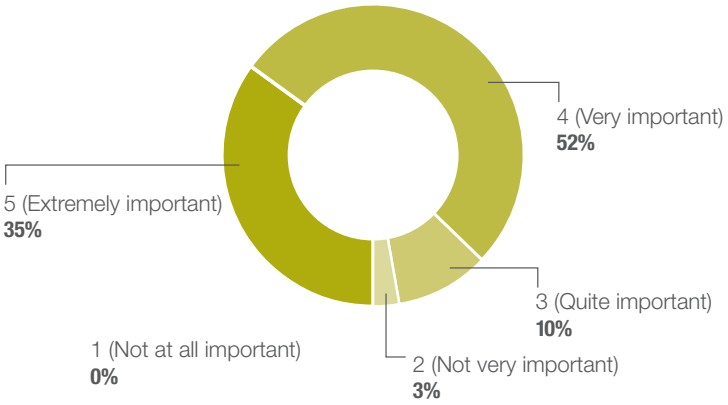
- IT Professionals are acutely aware of the growth of data and the significant challenge of managing data cost efficiently whilst ensuring their organisation optimises the value of its data.
- DLM means different things to different people; however it is considered very or extremely important by 87% of our online respondents and widely accepted as a good framework for developing a consistent approach to the management of data.
- Many large organisations have fairly immature DLM strategies. Only a fifth of the organisations in our online survey are fully realising the benefits of DLM, with a quarter to a third either only at pilot stage or doing nothing.
- Valuing data remains a complex issue. Organisations continue to struggle with assigning a 'positive' value, which is derived from processing data and turning Big Data into insight. For now, it is the 'negative' value associated with data storage and data loss, including the cost of rebuilding, non-compliance, lost sales and the impact on reputation, which is more straightforward to value.
- Cost control is the focus for many organisations, although relatively few (less than a third) are taking full advantage of data management technologies such as tiering, deduplication and compression.
- Amongst those organisations with mature DLM strategies with whom we conducted in-depth interviews, not only are they deploying these technologies, but they are also moving to automated approaches.
- Big Data is driving the need for new and different skill sets amongst data management professionals.
- Some key learning emerged in terms of best practices employed by the mature DLM organisations we interviewed. These can be found later in this report.

RESEARCH FINDINGS

WHAT IS DATA LIFECYLE MANAGEMENT (DLM)?

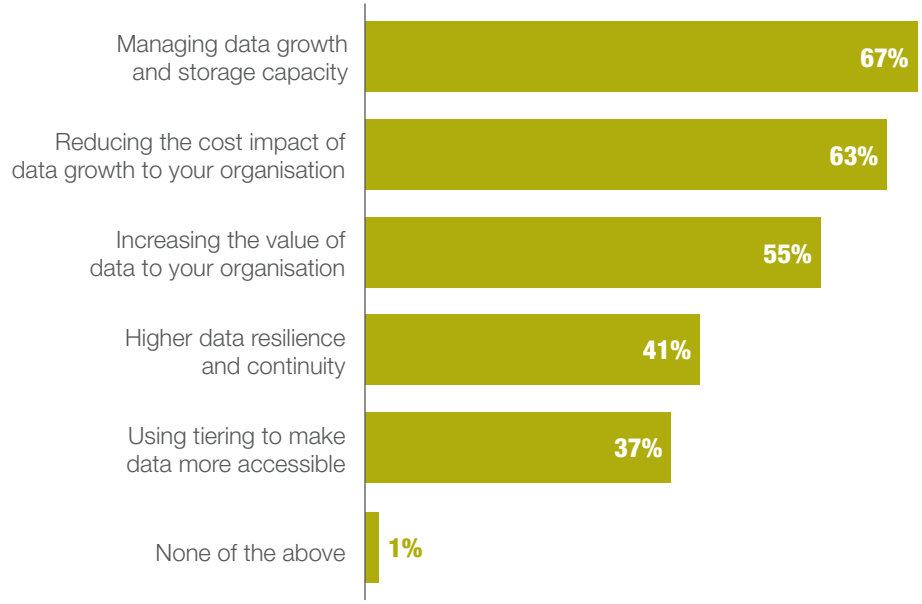
DLM is a model for the latest strategies and technologies for managing data. It includes the strategies for where data is hosted and how it is managed. It focuses on the most cost effective approach to managing data while ensuring its availability, disaster recovery provision and compliance meets the requirements of an enterprise.

Figure 1: The Importance of DLM



Base: 150 IT decision-makers working in large organisations
How important is DLM to your future strategey?

Figure 2: The meaning of DLM



Base: 150 IT decision-makers working in large organisations
What does DLM mean to you? (multi-response)

87% of our online respondents considered DLM extremely or very important to their organisations' future strategy. When asked what it means, it is clear that it has multiple dimensions, with managing data growth and storage capacity being cited most frequently.

One of our depth interviewees characterised DLM as follows:

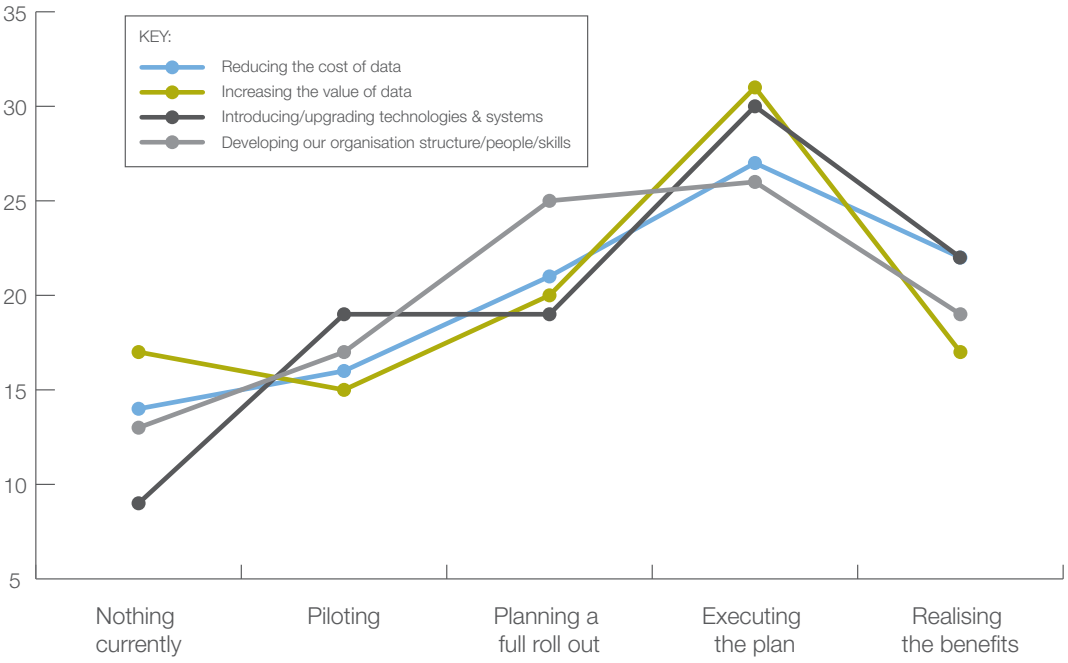
> "DLM is a great way to look at how you store your data and why. The programme has already helped us to remove a raft of legacy systems that were costing us millions in service costs and to reduce the size of the data we have to keep. However, we still have a major issue around duplication that we are looking at. I think that I would describe DLM as an iteration of the general trend towards data efficiency rather than a completely new approach."
> **IT Director – Financial Services**

It is worth acknowledging that many aspects of DLM are not new, however what DLM and other associated terms such as ILM (Information Lifecycle Management) recognise is that data management is a highly professional and strategic discipline which is becoming an essential pre-requisite in the Big Data era.

DLM STRATEGIES ARE RELATIVELY IMMATURE

We were interested to understand the overall maturity of large organisations with respect to DLM.

Figure 3: DLM Maturity



Base: 150 IT decision-makers working in large organisations
What do you do today?

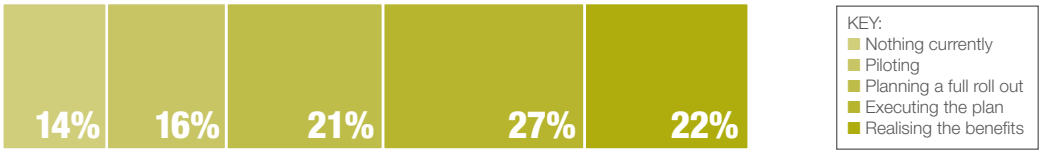
We asked our online respondents to assess at what stage of maturity they considered their organisations to be, with respect to their DLM strategies.

Recognising that the organisations within the survey are some of the largest in the UK, we were surprised to find that only a fifth consider themselves to be fully realising the benefits of DLM, with a quarter to a third either only at pilot stage or doing nothing across our four dimensions.

Our in-depth interviews helped to better understand the challenges faced with respect to developing mature DLM organisations.

REDUCING THE COST OF DATA

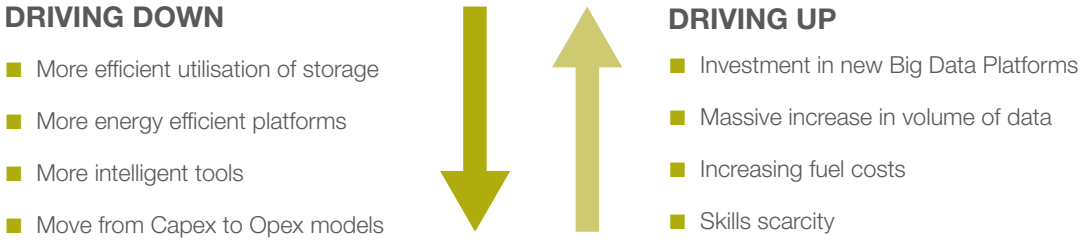
Figure 4: Current state of maturity with respect to reducing the cost of data



Base: 150 IT decision-makers working in large organisations
What do you do today?

There are a number of trends which impact the cost of managing data that have competing effects. It is also increasingly difficult to measure costs accurately.

Figure 5: The drivers of the cost of data



The trend towards Big Data means higher cost because there is more data to be managed and stored. The cost of storing data in data centres is also increasing and is likely to grow further as power costs continue to rise over the next few years.

However, to some extent these drivers are being cancelled out by trends which are driving down costs. Technology improvements are having an impact on the bottom line cost of managing data, as are managed solutions, which are based on utility pricing enabling a move from Capex to Opex-based financial models

More efficient management and storage of data meant that in 2011, the cost of creating, capturing, managing, and storing information was down to 1/6th of where it was in 2005. **(Source: EMC)**

However, as one of our depth interviewees succinctly expressed it:

"I expect my DLM programme to reduce my storage and data management costs by around 25% year-on-year for the next three years and that would be great, except I expect my data warehouse to grow by a factor of ten over the same period."
IT Director – Retail

As a result of these competing trends, it is difficult for organisations to establish budgets and forecasts with a degree of confidence.

INCREASING THE VALUE OF DATA

Figure 6: Current state of maturity with respect to increasing the value of data

Base: 150 IT decision-makers working in large organisations
What do you do today?

With a third of our online respondents either doing nothing at all or only at pilot stage with respect to increasing the value of their data, we were keen to understand the reasons why increasing the value appears to be a challenge.

IT professionals we interviewed felt that their organisations are struggling to understand how to value their data and, without being able to assign a value, it is difficult to assess if that value is increasing.

Many that we spoke to described two elements of value – negative and positive.

- **Negative value** is the cost of ownership or the cost of building or using the data or rebuilding it in the event of some catastrophic loss. It can include the run rate of not operating a service, or the revenue not collected, or the products or services not sold or delivered. It is also the operational cost of hosting and delivering the data within the organisation. It may be considered to include any compliance exposure for the loss of data. Negative value was relatively easy for our interviewees to understand.

- **Positive value** is what might be called the book value, or notional value, of the data. As organisations invest in Big Data there should be some definable increase in the value of that data. If you take some clay and turn it into a vase, then the vase will have some tangible increased value over the raw clay. As organisations process data into insight through Big Data applications, there should also be some tangible increase in value.

The first problem is that they struggle to value the raw data as anything other than a cost. The process that turns the data into insight is where the value is created and the value of a specific insight cannot be done in isolation. An insight may build upon something that exists within an organisation, so measuring the increase in value is almost impossible. Some institutions are working on creating different finance models with varying degrees of success, but the most successful organisations appear to have set aside significant budgets against the idea that they will come to understand the value side of the equation better as they gain experience.

> "We are able to identify the positive value of certain features such as the data required to upsell or the data around previous orders, but mostly we use the notional value of a customer and our primary focus is service interruption which forms part of our reporting."
> **IT Director – Retail**

Data has always been regarded as an intangible asset. Historically this kind of value was included in what used to be called 'good will' on corporate balance sheets. However, the importance of data is so significant today that the use of this approach is no longer seen as valid. The other challenge must be that there are now two additional objects of value for financial consideration. The first is the actionable insight that is derived from the data and the second is the application of that actionable insight. Accounting practices will be wrestling with this challenge for some time to come.

> "I blame the accountants – we have had this issue for years and as the professionals involved they should have been working on providing an industry set of guidelines that are workable."
> **IT Director – Financial Services**

DEVELOPING THE DLM ORGANISATION STRUCTURE, PEOPLE AND SKILLS

Figure 7: Current state of maturity with respect to organisation structure, people and skills

Base: 150 IT decision-makers working in large organisations
What do you do today?

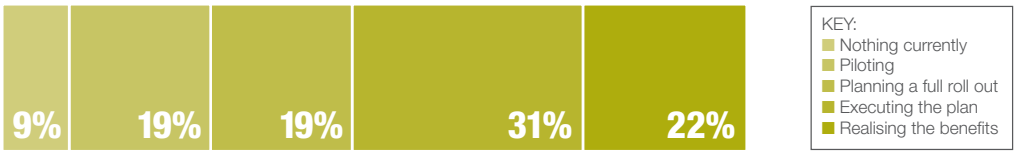
Only one in five of our online survey respondents consider that their organisations are benefitting from a mature DLM organisation structure.

One of the issues that emerged in many of our in-depth interviews was a concern around available skills. It was clear that all of the interviewees had a concern around resources for the new disciplines that are developing to leverage Big Data. Some spoke of experiencing challenges for existing programmes and some of a potential concern over future programmes. Organisations will see a premium for those with experience and a need to invest in developing those skills as well. It is also worth noting that there have been several reports from respected organisations that highlight this issue. These reports have received significant press and will fuel the concerns of organisations as they build their programmes.

By 2015, Big Data demand will reach 4.4 million jobs globally, but only one third of those jobs will be filled. **(Source: Gartner)**

DEPLOYING DLM TECHNOLOGIES AND STORAGE SOLUTIONS

Figure 8: Current state of maturity with respect to deployment of DLM technologies and storage solutions



Base: 150 IT decision-makers working in large organisations
What do you do today?

Relative to the other dimensions of our DLM maturity framework, deployment of DLM technologies and storage solutions is a little more mature.

Figure 9. The use of DLM technologies

Base: 150 IT decision-makers working in large organisations
Which of the following do you use?

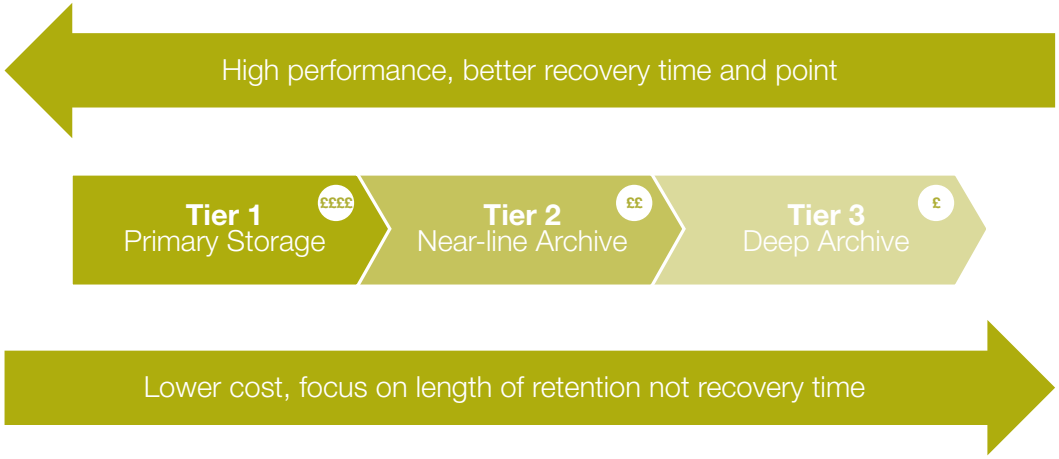
Of the five technology areas we included in our online survey, centralised document management systems, collaboration technology and data warehousing are relatively well established and mature DLM technologies. Somewhat less widely used and mature, are storage solutions such as tiering and data duplication.

We asked our depth interviewees about the benefits of these technologies. Along with compression, we found that mature DLM organisations are experiencing real tactical wins from these solutions.

TIERING

Tiering is about assigning different levels of storage to different types of data to manage cost. The costs associated with having your data available live has been something that organisations have been looking at for many years. As the automated management of tiers becomes more sophisticated, the number of tiers is increasing.

Figure 10. A simple tiering model



- **Tier 1** – is about using the most robust, fastest, highly available platforms (which tend to be more expensive) to manage the most frequently used mission critical data.
- **Tier 2** – storage platforms are lower cost and are used for information that is required less frequently but is still needed to support live processes. The cost of Tier 2 storage is typically half of that of Tier 1.
- **Tier 3** – tends to be used for large data sets that are required for storage in case of complaints or for compliance but are called upon as an exception. This data may be held on CDs or tapes. For many organisations the most common type of data would be call recordings that need to be held for long-term compliance. The cost of Tier 3 storage is typically a half of that of Tier 2.

The use of Tier 1 and Tier 3 media is mature – many organisations have used tape to store data that could be off-line for years for Tier 3. Several organisations we interviewed are still using video tapes from VHS systems for data that is now coming to the end of its valuable life.

The role of Tier 2 storage has increased in importance as SANs (Storage Area Networks) have improved in performance. However, findings from our online research suggest that Tier 2 storage is not as well used and this was borne out by our in-depth interviews.

> "Today we operate with two classes of data – live and off-line. Our off-line media is primarily tape but we started using CDs three years ago. Our compliance requirement is seven years so we will have finished with tape in a few years. We had never really invested in Tier 2 but are in the process of redefining our use of data though our DLM programme. Currently we are looking at moving about 35% of our Tier 1 data to Tier 2 and we should see some cost benefits from this part of the programme."
> **IT Director – Financial Services**

This suggests that many organisations are missing out on the opportunity to benefit from significant cost savings which can be achieved through tiering and deployment of Tier 2 storage.

Those organisations we in-depth interviewed and judged to be in the top half of the maturity scale for DLM, all said that they were exploring or delivering benefits from automated tiering. Automated tiering involves the automated progression or demotion of data across different tiers according to performance and capacity requirements. In most process design cycles, the move to automation is often an indicator of maturity and it appears that the move towards automated tiering may be an indicator of DLM maturity.

Only 27% of our online respondents reported realising the benefits of tiering and 15% are doing no tiering at all. As Big Data becomes more of an issue for organisations, it is likely that tiering will become an obvious choice.

DEDUPLICATION

Our online survey findings suggest that a quarter of large organisations are delivering the benefits of de-duping. However, as our in-depth interviews revealed, for some organisations deduplication is not simply a question of identifying extraneous duplicate copies of data sets that aren't needed, it is far more complex than this. Some organisations have multiple record sets, supporting different silos with duplicated names but different data, and for these organisations, deduplication requires a more fundamental redesign process.

> "We are now rolling out DLM as a set of practices across the whole of the group and providing training and support. As a complex organisation with mergers in our recent past it has been difficult to get buy-in from the senior team to spend money on tidying up our data and infrastructure into a strategic model. We have been focused on the tactical. DLM provides the opportunity to express our requirements in a consistent way. Our biggest challenge is duplication. For us it is very complex as we have lots of different partial records stored in different systems and our strategy is about having a single record rather than 20 partial records."
> **IT Director – Financial Services**

COMPRESSION

Compression was not a technology included in the online research but emerged from our in-depth interviews as an important tactic.

For those with large audio and video requirements, compression holds some real opportunities to manage costs. Compression technologies and algorithms are dramatically improving. Those interviewees with large recording requirements were very focused on the wins from compression. Their views can be best summed up by the following two quotes.

"With more than 8,000 hours of call recordings every day to be held for compliance, we get a huge cost benefit from data compression. We have been actively watching innovations in compression for years and spend money in every budget on ensuring that we have the best possible compression technology."
IT Director – Retail

"DLM is all about cost of storage for us. As one of the world's leading media organisations, we are producing nearly 1,000 hours of high definition video to store every 24 hours. Elsewhere we have significant content to store. Tiering is important to us but we are actively funding research into compression because of our streaming propositions and data housing needs."
IT Director – Media

WHAT CAN BE LEARNT FROM MATURE DLM ORGANISATIONS?

Based on the experiences of the most mature DLM organisations we interviewed, the following important learning emerged:

- It is vital to classify your data and involve the business in the process. Set goals based on understanding what proportion of your data is currently on the correct platform.
- Review your data sets every 6–9 months based on actual usage. Consider the use of age tracking software and the use of data producing alarms at set frequencies.
- Focus on closely monitoring storage and platform utilisation to identify opportunities to cut costs through server virtualisation and additional tiering opportunities.
- Consider the use of third party benchmarking to assess the strengths and weaknesses of your DLM approach.
- Develop a clear model of how your data relates to process in order to detect and address overly complex and inefficient data flows.
- Make an objective assessment of the skills you currently have within your DLM organisation and those required to support it in the future. Take steps to close any gaps which exist.
- Ensure your data design and operational processes are clearly defined and appropriate SLAs and mechanisms are in place for testing adherence.
- Do not skimp on security, disaster recovery or compliance. It is vital to ensure you stay up to date with the latest innovations in crime and data theft. Ensure that your disaster recovery is neither over nor under-engineered. If you have a large customer base, recognise that the changing boundaries of customer-owned information mean that ensuring compliance is a continuous task.

CONCLUSIONS AND RECOMMENDATIONS

- Big Data is driving many organisations' data management strategies and the expectation is unprecedented growth in data used and accessed across their organisations.
  - ► Traditional technologies are unlikely to meet your future needs for storage, processing and retrieval. If it isn't already, make Big Data your number one infrastructure challenge.
- Organisations like DLM as a methodology for developing a consistent approach to the management of data.
  - ► Recognise the different value data has through its lifecycle and manage it appropriately to maximise its value and minimise costs.
- Many large organisations still have fairly immature DLM strategies. Those which are more mature are gaining benefits from sophisticated DLM programmes that adopt a single strategy for the storage of data across the whole enterprise. This is producing measurable improvements in the availability of data that impacts the bottom line.
  - ► Evaluate the benefits of an outsourced DLM strategy compared with building and running your own data storage infrastructure.
- If assigning positive values to data remains challenging, use a 'negative' valuation approach to support data tiering.
  - ► Consider business criticality and performance needs in order to assign value tiers to your data and align your storage resources accordingly.
- For the majority of organisations the greatest focus of DLM programmes is on cost management and reduction.
  - ► In addition to tiering, deploy deduplication and compression technologies to reduce your storage costs.

HOW VODAFONE CAN HELP

By understanding your needs and identifying opportunities for data tiering and de-duplication, we will recommend the most appropriate storage strategy for your primary, near-line and deep archived data, ensuring appropriate accessibility and helping to significantly reduce your costs..

Ensuring compliance and security
Tiered data storage reduces costs by placing data on the most appropriate platform or storage tier. At the same time, our platforms will ensure your data is secure at every level and ready to support your business continuity and disaster recovery plans. You'll still have access to vital data even if you suffer systems failures, physical incidents such as fire, or human errors such as accidental deletion.

Simplifying management
We can make data simpler to manage, too, by introducing applications that increase functionality and allow you to automate tasks, so it becomes easier to fulfil your business strategy.

Great connectivity
Our storage platforms are connected via the best network options with our Multi-Service Platform Network meaning that your environment becomes a core component of our secure Vodafone Wide Area Network and is built on the Internet backbone with other connectivity services available.

FIND OUT MORE

For more information about how we can help you create and implement a Data Lifecyle Management strategy, based on storage infrastructure that provides the optimum balance of cost and performance, and to arrange to meet with a storage specialist please contact us.

T: +44 (0) 0800 096 5838

E: HostingEnquiries@vodafone.com

NOTES

NOTES