



An Oracle White Paper
June 2012

Increase Data Access and Productivity with Oracle Optimized Solution for Lifecycle Content Management

Introduction	1
Challenges in Data Access and Longevity.....	2
Data Retention and Growing Capacity Requirements.....	2
The Storage Itself Is Not the Only Cost.....	2
Productivity for Content Users and IT Staff.....	3
The Two Stages of Content Lifecycle	4
Two Approaches to Content Management, One Goal.....	5
Overview of the Oracle Optimized Solution for Lifecycle Content Management	5
Reduce Business Risk Through Dependable Access to Information Throughout Its Complete Lifecycle	7
Protect Investments with Scalability and Longevity	8
Reduce Deployment Cost and Risk	12
Summary.....	12

Introduction

Data is the lifeblood of many businesses and organizations, so it must be managed and protected throughout its lifecycle. Some data is transient, needing to exist for only seconds, minutes or hours. Other types of data must be kept accessible for months, years, or even indefinitely, to meet compliance regulations as well as to support ongoing use of the data for product, business, and collaborative needs.

Content management is not just about managing data when it is created. It is about managing data throughout its lifecycle, making it easily searchable and accessible. The most important aspect of content management is providing reliable and dependable access, even over long retention periods, which in some cases could be forever. Delivering scalability and longevity to enterprise data requires a content management solution that is not only dynamic and sustainable but also cost effective and standards compliant.

Content storage provides the platform for digital content retention. Given retention requirements for access and the longevity of retention horizons, the proper mix of high performance and long-term storage systems will be required for content storage. A tiered storage solution that blends multiple types of storage with integrated management and automated data movement is critical to reducing total cost of ownership.

Oracle is the only solution provider that offers complete integration of enterprise content management, tiered storage, and tiered storage management. This integrated approach enables Oracle to deliver a solution that optimizes performance, cost, and capacity, while also delivering scalability, flexibility, and 100 percent data access.

This paper provides a discussion of the business needs for an integrated content management solution and the value of implementing such a strategy through the Oracle Optimized Solution for Lifecycle Content Management. Oracle's offering can help reduce costs, lower risk, and increase user productivity. By giving users easy access to the information they need, organizations have the opportunity to make better business decisions, foster collaboration both inside and outside the company, and bring products to market more quickly.

Challenges in Data Access and Longevity

There are many challenges related to managing TBs and even PBs of digital content that must be protected and retained for easy searchability and accessibility. Scott Rife from the Library of Congress stated it best when asked about the most important part of storing information. He stated, “Reliable retrieval is key: An archive that only stores content is indistinguishable from a landfill. We need technology that reliably delivers all the content whenever requested and tells us proactively if there are issues affecting the retrieval of archived content.”

Data Retention and Growing Capacity Requirements

In its 2011 annual digital universe study, IDC estimated that the total amount of digital information would surpass 1.8 zettabytes (1.8 trillion gigabytes) in 2011, representing 9x growth in the last five years.¹ Of the data that organizations maintain today, much of it is never used after 90 days and yet many companies require a 100-year archive. Why do customers want to keep the data if they never use it? Businesses are keeping data because compliance requires it or the rules are not clear enough, so they store everything with long retention periods to be sure all requirements are covered. The answer to why they ‘never use it’ is they do want to use it but do not have the tools to search and retrieve it. During a search for a solution to meet these growing data needs, the first problem to solve is how to increase the data access, which increases data value and its contribution to the growth of the company.

The Storage Itself Is Not the Only Cost

According to a tiered storage study by Horison Information Strategies, a three-tiered storage implementation using tape for tier 3 archival data can provide significantly lower acquisition cost compared to all disk implementations. For a 100-TB archive, Horison estimated the acquisition cost of a three-tier solution to be 2.1x lower than a two-tier disk solution and 3.7x lower than a single-tier disk solution.²

With respect to archive costs, it is also important to consider OPEX costs, which include the cost of power and cooling as well as the cost to manage the storage. All of these costs grow as a disk farm grows to petabytes of content. However, since operating cost for tape is significantly less than disk, a three-tier archive offers much better TCO. An independent study by The Clipper Group found that 12-year TCO for tape-based solutions was, on average, 15x less than TCO for disk-based solutions and

¹ “IDC Digital Universe Study, sponsored by EMC,” June 2011. Available at <http://idcdocserv.com/1142>.

² “Tiered Storage Takes Center Stage,” Horison, Inc. Available at <http://www.horison.com/OracleTieredStorageTakesCenterStage.pdf>.

that disk solutions used 238 times more energy over the 12-year period. In fact, the energy costs alone for a disk-based solution approached the total cost for tape.³

In addition to offering lower storage costs, the Oracle Optimized Solution for Lifecycle Content Management provides the option to archive up to four copies, local and remote, for data protection, eliminating the content backup requirement.

Productivity for Content Users and IT Staff

Two parts of organizations are affected by a content management system. One is the IT staff that implements the content management and tiered storage archive solution, and the second is the user community. In both cases, attention must be paid to lowering costs and lowering risks while giving users the data they need to grow the business.

- **IT productivity:** Within IT, there are many options for both hardware and software technology, and the technology changes frequently. Before placing a solution in production, the hardware and software must have been proven together with known sizing and performance data. As new servers and storage technologies are introduced over time, it is critical to move to these technologies with little or no disruption to daily activities, and with no data loss or corruption.
- **End user productivity:** User productivity is affected by the ease and speed with which data can be retrieved. In a survey of AIIM members in 2010, 72 percent stated it is harder to find information owned by their own organization than to find information not owned by them⁴. According to Network World in 2007, this results in millions in lost revenue due to time wasted related to data searches. An interesting statistic from June 2010 research by Anoto Group found that 86 percent of companies still do paper-based transactions⁵. This results in a lot of scanning to create digital images to be entered into a content management system.

Access, Longevity, and Scalability

Oracle Optimized Solution for Lifecycle Content Management provides longevity of data in a scalable, cost-effective solution, delivering access to 100 percent of the data throughout its complete lifecycle. Technology will change, users will change, departments will come and go, and data will grow to massive capacities—but the data still will be dynamically searchable and accessible.

³ “In Search of the Long-Term Archiving Solution—Tape Delivers Significant TCO Advantage over Disk,” The Clipper Group, December 23, 2010. Available at <http://www.oracle.com/us/products/servers-storage/storage/tape-storage/>.

⁴ <http://www.aiim.org/capture/leading-the-pack.aspx>

⁵ <http://www.ecmconnection.com/article.mvc/New-Research-Finds-86-Of-Companies-Still-0001>

The Two Stages of Content Lifecycle

There are two primary stages in the lifecycle of content. The first stage starts with content ingest and includes the workflow through content edits, versioning, approvals, and transformations. The second stage is content storage in an appropriate storage environment with the right level of data protection. Most content has a much longer retention time than the life of the hardware on which it is stored. Therefore, the storage stage of the content lifecycle must take into consideration the need to transition through technology changes, including both hardware and file format changes. Tax records may have a retention period of seven years; health records must be kept for the life of the patient plus a few years; and most books and journals must be kept forever. New health records will be accessed frequently while the patient is ill but may never again be accessed after the patient is well. Books and journals may be accessed 30, 40, or even 100 years after being created.

The content that is managed by a lifecycle content management system generally consists of unstructured data that is created digitally or reproduced digitally. It is typically processed through a workflow and then stored for a given length of time, possibly forever. The source of this content could be a person typing or scanning images or a machine that generates content 24 hours a day, seven days a week. Or, the source may be research vessels that scan the ocean floor 24/7, collecting large resolution images in order to map the ocean floor and searching for the perfect place for oil drilling. MRIs and x-rays now are born digital and can be stored on long-term storage immediately for data protection as well as saved on primary storage for frequent access shortly after being created.

Management of this content starts with the content's creation and includes management throughout a dynamic workflow that includes editing, versioning, approval, and other steps that result in a published version of the content. Saving the content to an appropriate computer storage environment and managing the digital files throughout the remainder of their lifecycle also is considered part of the content management task. Policy determines where the files will reside and how many copies will be kept based on data protection and disaster recovery requirements. It is important that the storage management be as dynamic as possible and that all content be copied to multiple storage devices with the ability to access the electronic files without human intervention. Retention times for content often far outlive the life of the storage devices and may outlive the format of the content. An important part of lifecycle content management is to migrate content through hardware technology changes, as well as transform content to current, readable formats.

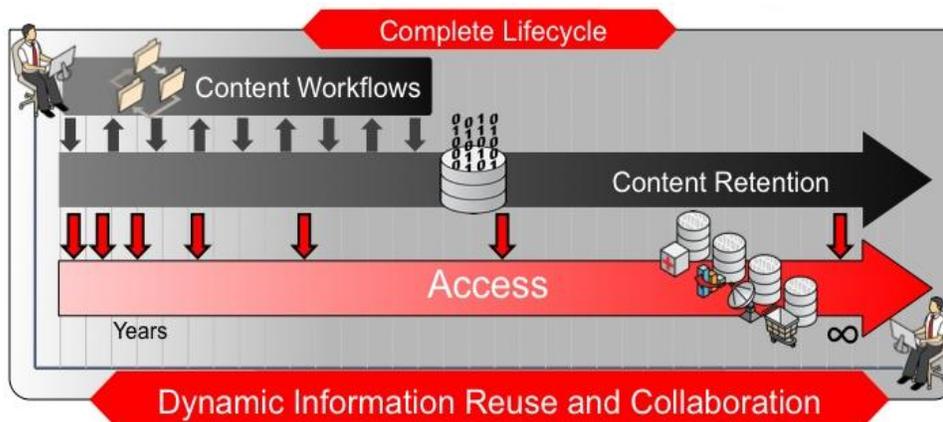


Figure 1. Lifecycle of information from ingest through its retention and access, both of which could be forever.

Two Approaches to Content Management, One Goal

Information is managed and touched by many people in an organization throughout its lifecycle. These people may have different approaches to handling the information, but all have the same goal—to make the content available. An IT manager is interested in system costs, power, cooling, device management, data protection, and other device-level concerns. People who manage and use the content are concerned with access, searches, workflow, versioning, data integrity, collaboration, and availability. Oracle brings both perspectives together in a solution that addresses the complete lifecycle.

The primary business benefits that the Oracle Optimized Solution for Lifecycle Content Management can help organizations achieve are described in the sections that follow.

Overview of the Oracle Optimized Solution for Lifecycle Content Management

The Oracle Optimized Solution for Lifecycle Content Management is designed to address content management from the perspectives of both a content user and an IT professional. It includes both a content management software application that manages the content workflow and a server and storage infrastructure designed to handle the preservation and protection of the data. By combining these two traditionally separate capabilities into a unified content management solution, Oracle offers a comprehensive approach to lifecycle content management that helps reduce business risk, protect investments, and reduce the cost and risk of solution deployment.

The Oracle Optimized Solution for Lifecycle Content Management is comprised of the following key components:

- **Oracle WebCenter Content:** A strategic content infrastructure suitable for managing documents, images, rich media files, and records.

- **Oracle's SPARC T-Series:** Running [Oracle Solaris](#), these server modules provide the most scalable, secure, and highly integrated platform for the optimized deployment of enterprise and mission-critical applications.
- **Oracle's Sun Storage Archive Manager:** Sun Storage Archive Manager software provides a configurable file system with storage, archive management, and retrieval capabilities. The software delivers dynamic management of data placement on appropriate storage based on policies and migrates data through technology changes.
- **Oracle's StorageTek tape libraries and tape drives:** Provide cost-effective long-term storage and data protection as parts of a tiered storage architecture.
- **Oracle's disk storage products:** Business-ready storage solutions provide a range of options for protecting and managing content files that are still in the active stages of the content lifecycle.
- **Oracle's Sun Storage 2500-M2:** A robust solution for small or midsize businesses or for integration into an existing enterprise SAN, which provides a NEBs-compliant, level 3 certified design.
- **Oracle's Pillar Axiom Quality of Service:** Provides and differentiates among the service levels of disk tiers. Pillar Axiom Quality of Service is used to enable the storage system to automatically deliver performance on demand and align storage resources based on the requirements of Sun Storage Archive Manager.

The following diagram describes how information flows from ingest into Oracle WebCenter Content, where metadata is created and stored in an Oracle Database for search and access purposes. The content is processed through the lifecycle of workflows, including edits, approvals, modifications, and transformations. Finally, the content stored in a tiered storage system where it is managed throughout the rest of its lifecycle, and possibly forever.

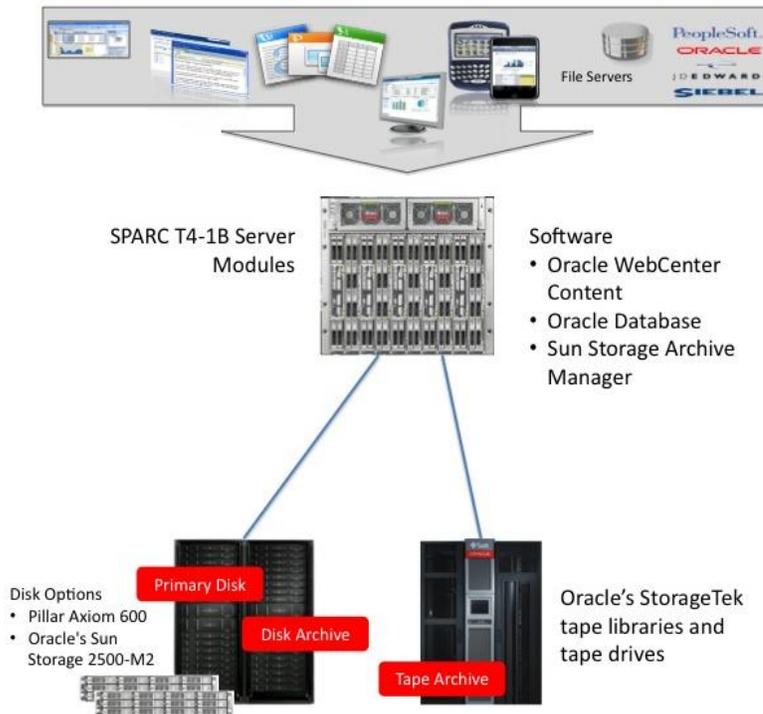


Figure 2. Oracle Optimized Solution for Lifecycle Content Management.

Reduce Business Risk Through Dependable Access to Information Throughout Its Complete Lifecycle

In order to preserve data in a cost-effective manner and ensure reliable access, tools are needed that will dynamically create multiple copies of information on low-cost storage. At the same time, the tools should provide assurance that when needed, the data will not only be available to be retrieved but also will have maintained its integrity. Data recall procedures are dynamic, requiring no human intervention to find and restore content.

- **Search 100 percent of content:** Reduce business risk through the ability to easily find information when it is needed. The Oracle Optimized Solution for Lifecycle Content Management includes intelligent metadata that describes the information stored within the system. This enables users to easily search and access content for cost-saving reuse and collaboration, resulting in better business decisions, more accurate product management, and faster time to market.
- **Access 100 percent of content:** Through the cost-effective use of tiered storage, information can be stored in an appropriate location based on policy. Multiple copies provide data protection and immediate, dynamic access when requested.
- **Preserve data integrity:** Oracle Optimized Solution for Lifecycle Content Management offers a means to easily preserve data and help assure its integrity. Oracle WebCenter Content makes it easy to search and find information, and the server and storage infrastructure helps maintain the data to provide long-term access while preserving data integrity.

Protect Investments with Scalability and Longevity

A solution that stores and accesses data successfully must be flexible and scalable in order to nondisruptively grow with the company and address technology changes. The Oracle Optimized Solution for Lifecycle Content Management provides a highly scalable system infrastructure that scales in terms of both capacity and performance throughput. It also simplifies migration of information to new storage technologies while enabling access to information on old technology.

- **Scalable storage capacity:** In addition to scalable performance, scalable capacity is required for content management and archive systems. The Pillar Axiom 600 storage system adds additional storage without downtime and can scale both the primary storage (often less than 10% of total capacity) and the archive disk storage. StorageTek libraries also enable nondisruptive growth by enabling multiple libraries to be connected as a single archive without taking the libraries offline. Sun Storage Archive Manager software is able to dynamically manage the addition of storage capacity nondisruptively to both the archive space as well as the primary storage area. The ability of Sun Storage Archive Manager to dynamically stage content from the archives greatly reduces the requirement for large costly primary storage capacity. The primary class of storage often is less than 10 percent of all capacity.
- **Longevity of storage systems:** The Pillar Axiom 600 storage system and StorageTek tape libraries support multiple generations of storage technologies. Migration of content from old to new technology platforms can be managed by Sun Storage Archive Manager software and can be done nondisruptively in the background. Users are unaware of the storage migration activity and once the migration is completed, the older storage can be removed.
- **Content format changes:** Oracle WebCenter Content features assure the content is not only retrievable but also readable. Content formats change over time and render themselves unreadable by current releases of software. Oracle WebCenter Content converts the content to open formats or to new formats, removing the need to maintain old releases of software for the only purpose of reading old data. If compliance requires it, the original copies are maintained in the archive.
- **New technology testing:** Oracle Optimized Solutions are refreshed as new hardware and software are released. An organization's IT is not the first to try new software releases or new hardware components. This brings longevity to the solution without running it on antiquated components or without an increased risk of downtime while migrating into the unknown. Sizing and performance of new components are a known factor, which reduces the impact to IT and to the users.
- **Solution scalability:** Figure 3 shows the scalability of Oracle Optimized Solution for Lifecycle Content Management. It compares the ingest throughput in terms of files per second using 100 K byte files on the small, medium, and large configurations.

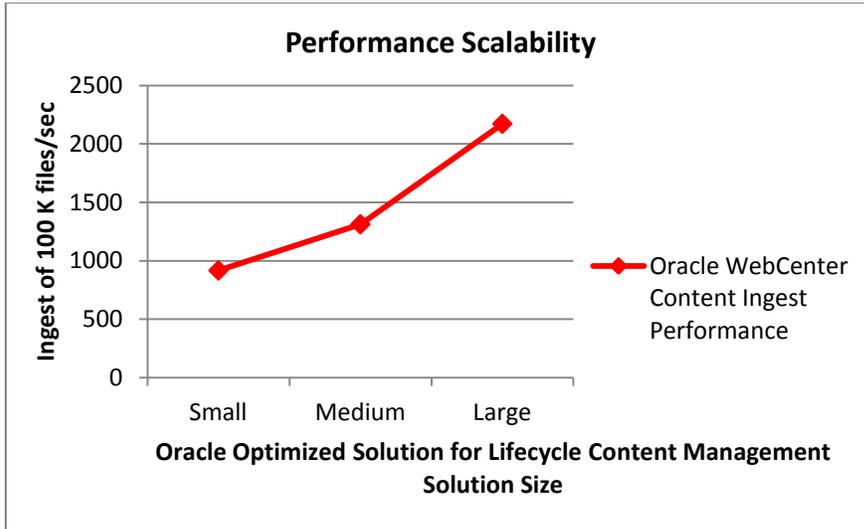


Figure 3. Test results show the scalability of the solution using a set configuration for small, medium, and large configurations.

Through testing, maximum ingest rates were determined for each of the configurations. These configurations have the potential to generate very large capacity repositories based on this ingest rate. Tables 1 and 2 show the capacity requirements for the small, medium, and large configurations assuming a continuous ingest rate for 24 hours per day, seven days per week.

Table 1 shows the capacity required for the first year, and Table 2 shows the capacity required for ingesting content at that same continuous rate for seven years. Table 2 also breaks out the seven-year capacity across the various tiers of storage using the rule of thumb that 10 percent of the total content will be kept on primary disk for frequent access, 30 percent will be kept on disk archive, and 200 percent will be kept on tape for data protection copies. Different percentages can be calculated for different use cases.

TABLE 1. ONE-YEAR CAPACITY REQUIREMENTS FOR SMALL, MEDIUM, AND LARGE CONFIGURATIONS

CONFIGURATION	INGEST RATE IN NUMBER OF 100 K RECORDS/SECOND	INGEST CAPACITY/SECOND (MB)	INGEST CAPACITY/YEAR (TB)
SMALL	500	50	1,576.8
MEDIUM	1,400	140	4,415.0
LARGE	2,200	220	6,937.9

TABLE 2. SEVEN-YEAR CAPACITY REQUIREMENTS FOR SMALL, MEDIUM, AND LARGE CONFIGURATIONS

CONFIGURATION	DATA CAPACITY FOR			
	SEVEN YEARS (PB)	10% PRIMARY DISK (PB)	30% DISK ARCHIVE (PB)	200% TAPE ARCHIVE (PB)
SMALL	11	1.1	3.3	22.1
MEDIUM	31	3.1	9.3	61.8
LARGE	49	4.9	14.6	97.1

Realistically, the ingest rates noted in Tables 1 and 2 will be experienced only at peak times. However, they must be processed without a negative impact to the users. The total ingest capacity will more likely be a much smaller percentage of the totals shown above. Selection of system size of small, medium, or large based on peak ingest rates, along with total capacity, expected growth, and the retention time, can be used to determine overall capacity requirements.

As an example of capacity requirements, Figures 4, 5, and 6 indicate the estimated capacity requirements for the small, medium and large configurations, respectively. The charts start with year one capacity and show an expected growth of 50 percent per year up to seven years for each of the different storage tiers. The charts use the same assumptions as Table 2 above—10 percent for primary disk, 30 percent for disk archive, and 200 percent for tape archive. Even though dramatic growth in overall capacity is shown, the primary disk storage tier, which is the most expensive tier, remains relatively small, thus keeping costs low. The most cost-effective storage—tape—carries the largest capacity.

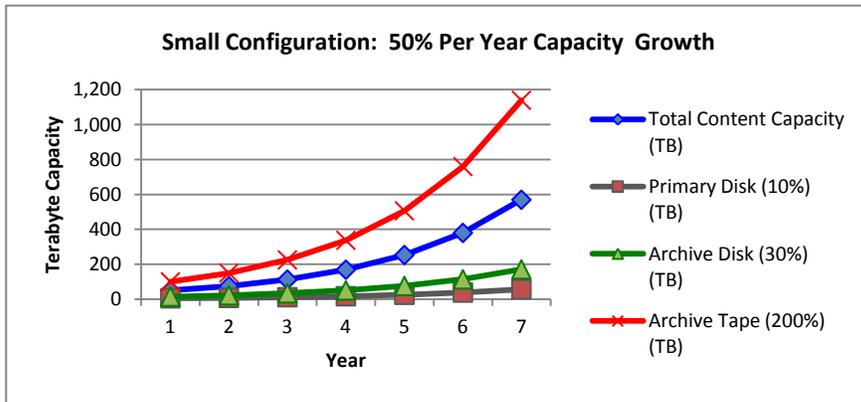


Figure 4. This table shows the total capacity increase over seven years based on 50 percent growth per year for a small configuration.

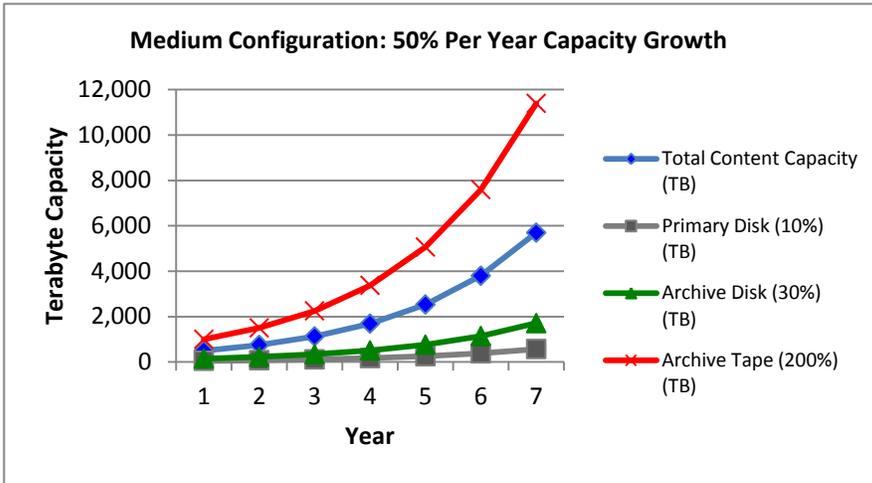


Figure 5. This table shows the total capacity increase over seven years based on 50 percent growth per year for a medium configuration.

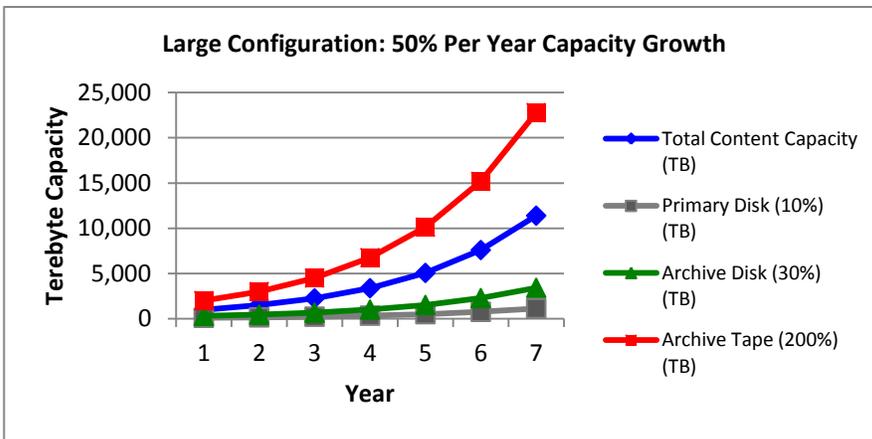


Figure 6. This table shows the total capacity increase over seven years based on 50 percent growth per year for a large configuration.

Businesses can only estimate capacity, ingest, and access requirements for the future when implementing a content management system. The Oracle Optimized Solution for Lifecycle Content Management scales nondisruptively in capacity and performance, meeting requirements during unexpected (but very welcomed) growth of the company. This growth is easily and quickly handled by the Oracle Optimized Solution for Lifecycle Content Management, leaving IT and users free to focus on important activities other than worrying about where the data is and whether it is safe.

Reduce Deployment Cost and Risk

Traditionally, IT has been faced with the task of ‘parts selection’ from a variety of sources with little or no knowledge if they work together. Standards in file types, access protocols, and computer languages have improved integration, but at the same time, add complexity to the selection, integration, and proof of concept. The Oracle Optimized Solutions team has completed the majority of this work, enabling customers to stand up a complete content management solution in much less time and with significantly less risk.

- **Server and storage selection:** The best server and storage components for the system have already been selected and tested together with the software.
- **Sizing guidelines:** Sizing is available for capacity and performance, enabling organizations to purchase what is required today while having good visibility into how the system will support future growth requirements. The pretested solution with sizing guidelines allows IT to spend more time on corporate IT strategy and planning than in the lab running proofs of concept.
- **Single vendor support:** The single vendor support for all components from application to disk and tape eliminates the finger pointing that can result from a multivendor solution.
- **Unexpected growth:** The solution offers flexibility for growth in both capacity and performance. Therefore, if unexpected growth in content ingest or retrieval access occurs, the solution can easily scale to meet the increased need.

Summary

Oracle delivers unique value through this Oracle Optimized Solution, which integrates standards-based Oracle hardware with Oracle software to deliver high performance, high scalability, and support for the complete information lifecycle. The Oracle Optimized Solution for Lifecycle Content Management helps organizations efficiently and effectively manage information from creation to archive to destruction. It is a flexible, scalable solution that grows with the organization and supports technology advancements.



Increase Data Access and Productivity with Oracle
Optimized Solution for Lifecycle Content
Management
May 2012
Author: Donna Harland

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2012, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. UNIX is a registered trademark licensed through X/Open Company, Ltd. 1010

Hardware and Software, Engineered to Work Together