Content Server Planning and Implementation Guide
10*g* Release 3 (10.1.3.3.0)

July 2009

ORACLE

Content Server Planning and Implementation Guide, 10*g* Release 3 (10.1.3.3.0)

Contributing Authors: Contributing Author, Contributing Author, Contributing Author

Contributors: Contributor, Contributor, Contributor

Table of Contents

Chapter 4: Intranet Sites

Chapter 5: Extranet Sites

Chapter 6: Internet Sites

Appendix A: Planning

Chapter

1

INTRODUCTION

OVERVIEW

This section covers the following topics:

ABOUT CONTENT SERVER

The Content Server offers a comprehensive, end-to-end solution that quickly deploys content-centric websites and easily integrates with a company's existing infrastructure. It can be implemented as a stand-alone application—enabling rapid deployment of content-centric websites—or as an integrated application—supplying websites and others e-business applications with consistent, managed content. Regardless of the implementation scenario, the Java-based, modular architecture offers flexible and scalable configurations to meet your organization’s content management needs.

The Content Management suite of products can be deployed at all organization levels, in a wide variety of environments and scenarios. It is fully scalable, which means it can grow with any organization’s information and content management needs.

ABOUT THIS GUIDE

This *Planning and Implementation Guide* is intended for integrators of the Content Server family of products. They install and set up the products, and make sure it integrates with the existing system and infrastructure. This document aims to provide guidelines for successful implementation and integration of Content Server in site-centric applications. It provides information about a number of common scenarios, and explains important considerations for each of them.

For general information on the architecture and functionality of the Oracle family of products, as well as a number of key concepts, refer to the *Getting Started With Content Server* guide.

Important: The recommendations in this guide are not “hard and fast” rules. Rather, they are guidelines which can help companies meet their particular content management needs using the Content Server suite of products. However, every implementation is different, with its own characteristics and requirements. This means they may require their own, very specific approach.

Symbols

The following symbols are used throughout this document:

Symbols	Description
	This is a **note**. It is used to bring special attention to information.
	This is a **technical tip**. It is used to identify information that can be used to make your tasks easier.
	This is an **important** notice. It is used to identify a required step or required information.
	This is a **caution**. It is used to identify information that might cause loss of data or serious system problems.

Chapter

2

CONTENT MANAGEMENT

OVERVIEW

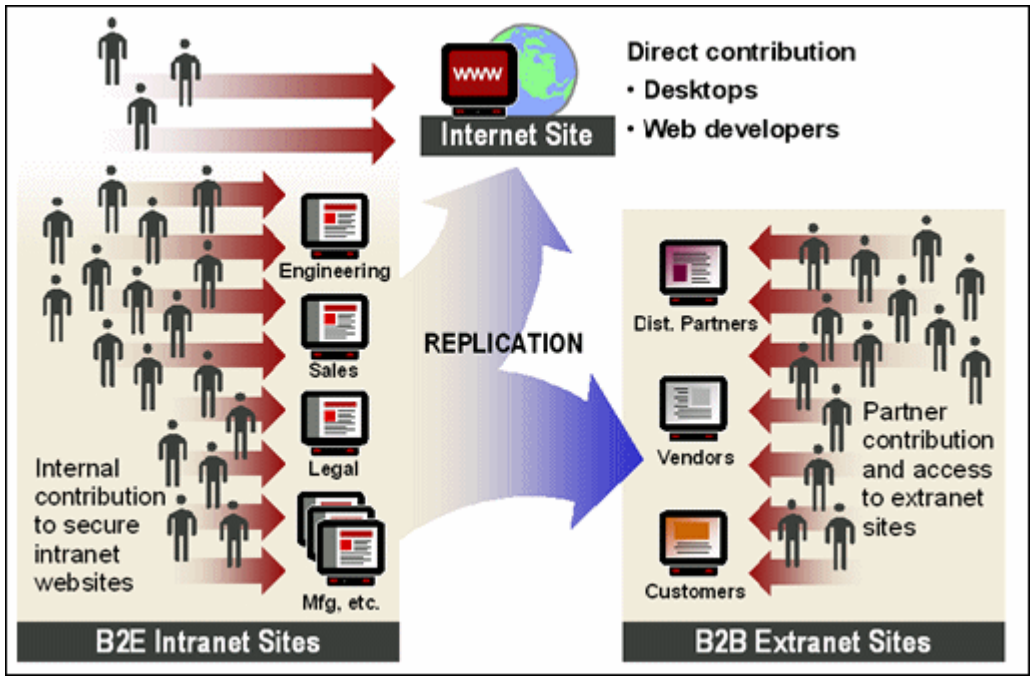
Content management is a very broad field which comprises many possible implementation scenarios, each with its own requirements and considerations. Content management is a multifaceted area, and several topics are important when implementing a content management application based on the Content Server, including:

- Type of website: intranet, extranet, or Internet (see page 2-1)
- Management model: documents or website (see page 2-5)
- Type of managed content: business, web, or structured content (see page 2-8)

INTRANET, EXTRANET, AND INTERNET WEBSITES

Content Server can manage content-centric websites. These websites can either be published on an intranet, extranet, or the Internet, as shown in Figure 2-1.the figure below. Intranets, extranets, and public Internet applications all run on the same network infrastructure and all use the same basic web technology. They are all publication models of the same content, and their differences primarily lie in the fact that their content resources are administered for different levels of accessibility, presentation, and security.

Figure 2-1 Intranet, extranet, and Internet websites



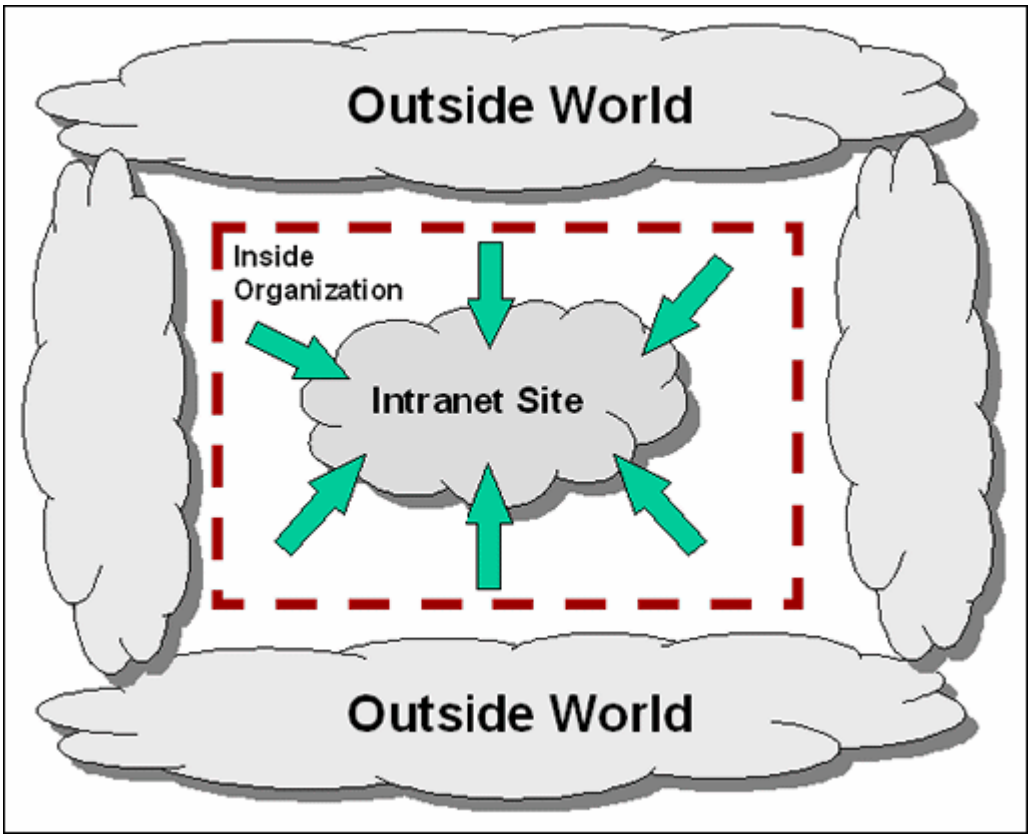
Intranet Sites

An intranet is a private network that uses common web technology for use within an enterprise or organization. Access to the network is restricted. Intranets may serve anything from small workgroups sharing the same office space to entire corporations with locations around the globe. They may be entirely disconnected from the public Internet, but are usually linked to it and protected from unauthorized access by security firewall systems or virtual private networks (VPNs).

Intranet applications are typically used in a “Business-to-Employee” (B2E) context, which means they are used to communicate with employees and share information within the organization.

Figure 2-2 The figure below shows the basic access model of a typical enterprise intranet.

Figure 2-2 Access to enterprise intranet site



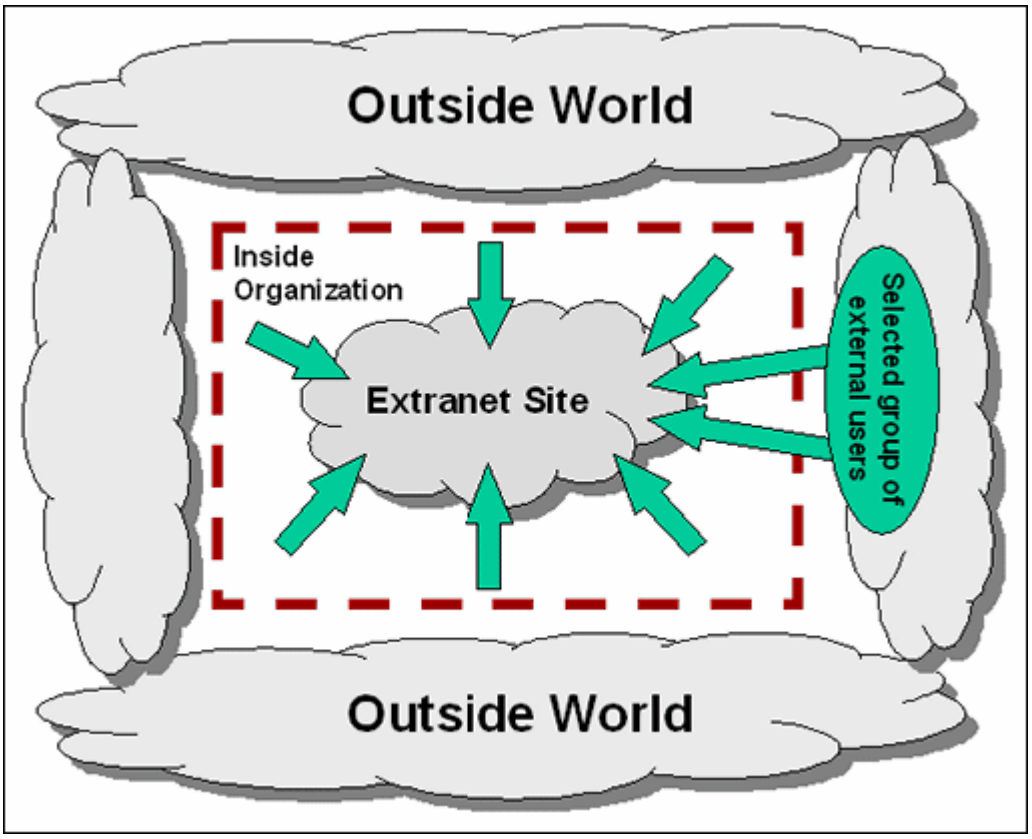
Extranet Sites

An extranet basically is an intranet that also offers limited, controlled access to a select group of users outside the organization. As such, extranets serve “extended organizations”— i.e., the organizations themselves as well as defined sets of customers, suppliers, or other partners. An extranet offers more secured access than an intranet site. As such, it is often inside a firewall, but does not need to be. An extranet is closed to the general public, but, unlike a pure intranet, it is open to the selected partners (a “closed user group”).

Extranet applications are typically used in a “Business-to-Business” (B2B) context, which means they are used to communicate with external partners, suppliers, etc. to share information and conduct business.

Figure 2-3 The figure below shows the basic access model of a typical enterprise extranet.

Figure 2-3 Access to enterprise extranet site



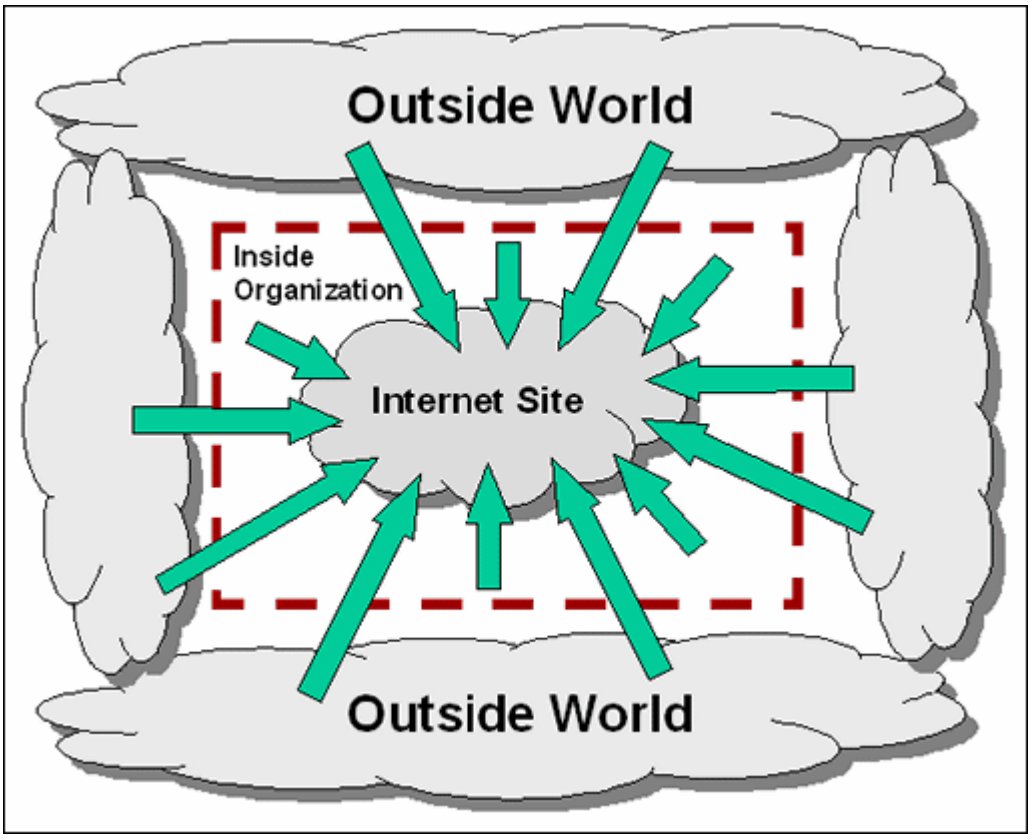
Internet Sites

The Internet is the public, global network of networks that is based on the Internet Protocol (IP) and related standards. It offers universal accessibility, which is one of its biggest strengths but also poses important security risks. Internet sites may be secured, but in most cases they are open to public consumption.

Internet applications managed by Content Server are typically used in a “Business-to-Customer” (B2C) context, which means they are used to allow any external users access to information on the website.

Figure 2-4 The figure below shows the basic access model of a typical enterprise Internet site.

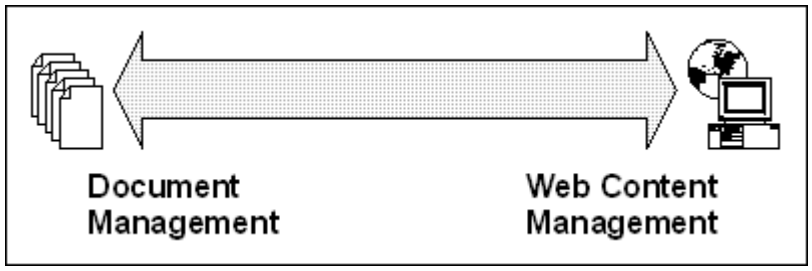
Figure 2-4 Access to enterprise Internet site



DOCUMENT MANAGEMENT AND WEBSITE MANAGEMENT

Content Server has a wide variety of content management applications, varying from pure document management on one end of the spectrum to pure web content management on the other (see Figure 2-5the figure below). In practice, however, applications are typically somewhere in between, with both document and web content management characteristics.

Figure 2-5 Content management spectrum



Document Management

Effective management of everyday business documents is crucial to any organization. There is a flood of information that needs to be managed and harnessed.

If used for strict document management, Content Server will typically serve as a central file repository with “library service” capabilities. Users can check content items in and out, creating a new revision each time an item is checked in. All revisions of a content item are stored, which enables easy tracking of changes. Figure 2-6 The figure below shows part of a revision history of a content item managed by Content Server.

Figure 2-6 Revision control in Content Server

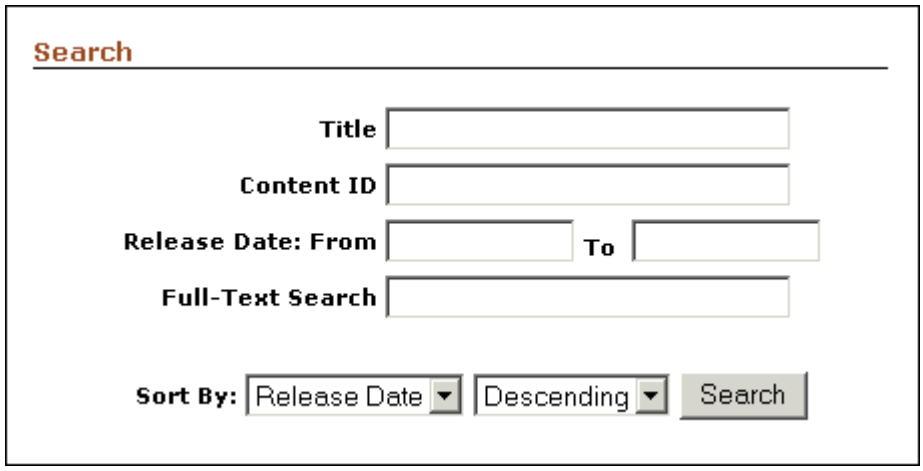
Revision History

Revision	Release Date	Expiration Date	Status	Actions
[5]	7/10/06 3:49 PM	None	Released	Delete
4	7/6/06 10:40 AM	None	Released	Delete
3	6/23/06 11:27 AM	None	Released	Delete
2	9/6/05 3:32 PM	None	Released	Delete
1	8/24/05 3:28 PM	None	Released	Delete

Content consumers use standard web browsers to access the content managed by Content Server. They will typically search for content using keyword-based queries. Figure 2-7 The figure below shows one of Content Server’s out-of-the-box search forms.

In pure document management scenarios, no content is published to a website for public consumption. A web interface is used to find and access the content, but no content managed by Content Server is used to populate an external website.

Figure 2-7 Searching for content



Web Content Management

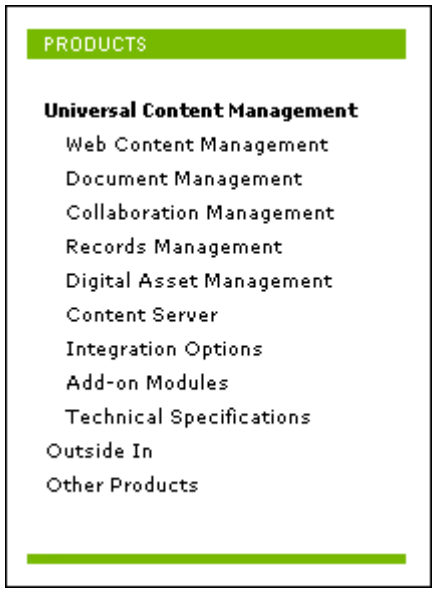
Websites provide faster access to information than print or other traditional communication tools. To realize the benefits of web communication, businesses need tools that increase the manageability and reduce the cost of publishing content.

Traditionally, content is published by a webmaster whose main responsibility is the technology, not the content of the site. This may create a serious bottleneck in the publishing process, as all content needs to pass the same point before being put on line. A web content management solution puts content publishing in the hands of the content experts, and enables webmasters to focus on their core tasks. This is accomplished by offering tools that automate the publishing process and providing business users with the ability to create web content from their own workplaces. Web content management applications put business users in control of creation, contribution, and updates of content published to a website. This enables non-technical users to create material with desktop applications, such as Microsoft Word, and post it to the website with little effort. Workflows can be created to automatically route updates through the approval process.

In web content management scenarios, the content managed by Content Server is used to populate a website. Content Server then basically serves as a dynamic content repository, which uses template-based publishing models to decide what content is published to the website, how, when, and in what form.

Content consumers will typically navigate to content using a tree-like hierarchy structure to “drill down” to the content they need. Figure 2-8 The figure below shows a simple example of such a hierarchy.

Figure 2-8 Navigating to content



BUSINESS, WEB, AND STRUCTURED CONTENT

Content managed by Content Server falls into one of three main categories:

- Business Content (page 2-8)
- Web Content (page 2-9)
- Structured Content (page 2-10)

Business, web, and structured content is typically stored in multiple repositories throughout an organization.

Business Content

Business content is generated by an organization in the course of doing everyday business. It is generally unstructured, and comprises about 85% of the data within a typical large company or organization. It can be published to the web, but is usually not created specifically for a website.

Examples of business content include:

- Word-processing documents
- Presentations

- Requests for proposal (RFPs)
- Requests for quote (RFQs)
- Product information
- Specifications
- Manuals
- Policies and procedures
- CAD files
- Contracts
- Faxes
- E-mail and attachments
- Transaction “hardcopies”
- Financial reports
- Consulting and analyst reports
- Video / audio

Web Content

Web content is generated primarily for publishing to a website. It is typically developed by web developers or creative organizations using graphic design and web authoring tools. Examples of web content include:

- HTML pages
- XML-based content
- HCSP and HCST pages
- Flash objects
- Images
- Active server pages (ASPs)
- Java server pages (JSPs)
- Multimedia files

Structured Content

Structured content is set up according to strictly defined data formats, and is stored in structured environments. It is typically created through templates and forms, and accessed from enterprise applications. Examples of structured content include:

- Web content submitted using forms and templates
- Database-driven content
- XML-based content
- Content from business applications such as enterprise resource planning (ERP) and customer relationship management (CRM) systems
- Product catalogs

APPLICATION SCENARIOS

The Oracle content management system can be used for a wide variety of applications, each with its own publishing model and content requirements.

Business-to-Employee (B2E) Applications

B2E applications are mostly used in an intranet context. They are primarily used to make information available to people within the organization. There can be tens or even hundreds of separate intranet sites within one organization, each serving a particular segment (geographical, organizational, functional, or otherwise). The content typically consists of unstructured, business-related information from multiple sources that needs to be shared within the organization. Employees continuously add and update content in real time, which means the number of contributors is typically high. Intranet sites are usually fully part of an enterprise application infrastructure.

Business-to-Business (B2B) Applications

B2B applications are mostly used in an extranet context. They share most characteristics with B2E (intranet) applications. The main difference is that the content in B2B applications is made available, in whole or partly, to a known group of users outside the organization. As such, a B2B application serves the “extended organization”—i.e., the organization itself as well as defined sets of customers, suppliers, or other partners.

Business-to-Customer (B2C) Applications

B2C applications are mostly used in an Internet context. Companies typically set up a single presence on the World Wide Web, which anyone in the world can access. The number of users is therefore limitless, and users typically do not need to make themselves known to enter the website. The content is web-related and often structured, especially if the website is database-driven. Strict business content is less relevant in B2C applications. There are usually few direct contributors and many consumers. The number of developers is typically also relatively high. For security reasons, Internet sites are often stand-alone applications, entirely disconnected from the organization's infrastructure.

The table below summarizes some important differences between B2E/B2B (intranet/extranet) applications on the one hand and B2C (Internet) applications on the other. Each of the characteristics is addressed in more detail in the next few sections.

Characteristic	**B2E/B2B (intranet/extranet)**	**B2C (Internet)**
Sites per company	Tens to hundreds	One to few
Source content	Multiple business sources	Templates, web developer conversion
Users	Known	Anonymous
Developers per site	Few	Many
Direct contributors	Many	Few
Content updates	Frequent, real time	Planned
Authentication	LDAP, NT, NDS	None
Personalization	Security, rules	Rules, inference
Nature	Incorporated in organization's infrastructure	Stand-alone, disconnected from organization

IMPLEMENTATION EXAMPLES

In addition to providing general guidelines, this *Content Server Planning and Implementation Guide* also addresses a number of implementation scenarios, each with a "basic" and an "extended" example. They are centered around the three main website types (see Intranet, Extranet, and Internet Websites on page 2-1):

- Intranet Sites (B2E) (see Chapter 4)
  - Basic scenario: workgroup intranet
  - Extended scenario: enterprise-wide intranet
- Extranet Sites (B2B) (see Chapter 5)
  - Basic scenario: partner site
  - Extended scenario: support site
- Internet Sites (B2C) (see Chapter 6)
  - Basic scenario: "stand-alone" Internet site
  - Extended scenario: Internet site with portal integration

Important: It is important to note that the above scenarios are not mutually exclusive in any way. A single Content Server-based application may well address the full content management spectrum, with content being published to an intranet, extranet, and/or Internet site.

Chapter

3

BASIC CONSIDERATIONS

OVERVIEW

Regardless of the implementation scenario, there are a number of basic considerations that need to be taken into account. They are related to the following areas:

- **Search Solution** (page 3-50)
  What search solution is used in conjunction with Content Server? Who will search for what?
- **Conversion to PDF** (page 3-51)
  How can content managed by Oracle be made available to content consumers as PDF files, and what conversion modules can be used to accomplish that?
- **Conversion to HTML** (page 3-53)
  How can content managed by Oracle be made available to content consumers as HTML, and what conversion modules can be used to accomplish that?
- **Conversion to XML** (page 3-56)
  How can content managed by Oracle be made available to content consumers as XML, and what conversion modules can be used to accomplish that?
- **Conversion to WML or cHTML** (page 3-59)
  How can content managed by Oracle be made available to content consumers as WML or cHTML, and what conversion modules can be used to accomplish that?
- **Other Conversions** (page 3-63)
  How can content managed by Oracle be made available to content consumers as WML or cHTML, and what conversion modules can be used to accomplish that?
- **Workflows** (page 3-63)
  Are workflows used to guide the review and approval process of content, and if so, how are they set up?
- **Metadata** (page 3-66)
  How is content managed by Oracle primarily categorized and organized, and how is it stored?
- **Backup Strategy** (page 3-72)
  What additional feature sets, if any, are needed to meet the organization's content management requirements, and what is their impact?
- **Backup Strategy** (page 3-72)
  What backup strategy should be followed to ensure continuous operation and efficient disaster recovery?

It is important to realize that none of the areas above stands alone. Choices in one area may affect others as well. For example, the security strategy impacts the infrastructure and communication requirements as well. Similarly, the type and quantity of managed content also affect the hardware setup. This means you should always consider all of the basic

areas above as a whole when implementing a content management solution based on the Oracle family of products.

Note: The above list is not necessarily exhaustive. Depending on the specific implementation, other areas not mentioned above may be important and may need to be taken into account.

Note: Appendix A (*Planning*) provides Refer to Appendix A (*Planning*) for details on some of the key questions that need to be addressed when implementing a Oracle-based content management solution.

MANAGED CONTENT

The following are important factors for the implementation of Oracle content management:

- Types of Content (page 3-4)
- Quantity of Content (page 3-4)
- Autogenerating Content IDs (page 3-5)

Note: Appendix A (*Planning*) provides Refer to Appendix A (*Planning*) for more questions to consider when it comes to the types and quantities of content managed by Oracle.

See also:

- *Users (page 3-5)*
- *Authentication Types (page 3-12)*
- *Hardware Setup (page 3-14)*
- *Infrastructure (page 3-27)*
- *Publication and Data Exchange (page 3-35)*
- *Security (page 3-41)*
- *Search Solution (page 3-50)*
- *Conversion to PDF (page 3-51)*
- *Conversion to HTML (page 3-53)*
- *Conversion to XML (page 3-56)*
- *Conversion to WML or cHTML (page 3-59)*

– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*
– *Backup Strategy (page 3-72)*

Types of Content

Content Server can handle any file format that can be stored on a file system. All checked-in files are stored in their native file formats, i.e. the formats they were originally created in (for example, Microsoft Word). It is recommended that you assess beforehand what types of content will be managed by Oracle. This simplifies setting up the content management environment as all required file formats and content types can be defined correctly.

Depending on the types of content, conversion modules may be available to present the content in a web-viewable format (for example, PDF, HTML, WML, cHTML, XML) or convert them otherwise.

In addition to knowing what types of content will be managed, it is also important to know what software applications they are created in and what file extensions are used (defaults or aliases?).

See also:

– *Quantity of Content (page 3-4)*
– *Autogenerating Content IDs (page 3-5)*

Quantity of Content

Several factors play a role when determining the expected quantity of managed content:

- Number of files
- Number of pages
- File sizes (byte counts)

If there is a high file volume to be managed (both regarding number and size), this obviously has an impact on the required hard disk storage space on the content server(s). Also, if a lot of content is accessed at the same time, this puts certain load handling requirements on the content server(s), and some type of load balancing may be required.

It is therefore useful to assess beforehand the expected quantity of content to be managed by Oracle.

Note: It is wise to assume that file storage and traffic will go up after the content management system has been put into place, since it is then much easier for users to find, access, and add information.

See also:

– *Types of Content (page 3-4)*

– *Autogenerating Content IDs (page 3-5)*

Autogenerating Content IDs

Each content item checked into Content Server is uniquely identified by its content ID, which is assigned upon check-in. The system may be set up to allow users to provide their own content ID with each check-in. This enables them to create “meaningful” content IDs for every item they check in. Alternatively, the system may be set up to generate content IDs automatically as items are checked in. This will be a numerical sequence that is incremented by one with each check-in, preceded by a user-definable prefix if required. This method does not produce “meaningful” content IDs, but does have the advantage that all content IDs are formatted in a uniform manner (to fit the corporate standard) and there is a clear sequential order, for example id_00001, id_00002, etc.

See also:

– *Types of Content (page 3-4)*

– *Quantity of Content (page 3-4)*

USERS

This section covers the following topics:

- Integrators (page 3-10)
- Web Developers and Webmasters (page 3-11)
- Administrators and Sub-Administrators (page 3-11)

Note: Appendix A (*Planning*) provides Refer to Appendix A (*Planning*) for more questions to consider when it comes to the types and numbers of users.

See also:

- *Managed Content (page 3-3)*
- *Authentication Types (page 3-12)*
- *Hardware Setup (page 3-14)*
- *Infrastructure (page 3-27)*
- *Publication and Data Exchange (page 3-35)*
- *Security (page 3-41)*
- *Search Solution (page 3-50)*
- *Conversion to PDF (page 3-51)*
- *Conversion to HTML (page 3-53)*
- *Conversion to XML (page 3-56)*
- *Conversion to WML or cHTML (page 3-59)*
- *Other Conversions (page 3-63)*
- *Workflows (page 3-63)*
- *Metadata (page 3-66)*
- *Backup Strategy (page 3-72)*
- *Backup Strategy (page 3-72)*

User Base

It is very important to know how many users will be involved in the content management process, and who and where they are. This plays a major role in setting up the infrastructure and security of the system.

The target groups of each of the main website types (see Intranet, Extranet, and Internet Websites on page 2-1) are typically different (although there may be overlap):

- Intranet site: employees (internal)

- Extranet site: selected partners/suppliers (external) and employees (internal)
- Internet site: customers and other external users

This has a big impact on how the user base needs to be defined and set up. Access and security requirements for a strictly internal user base are different than for groups of users that also include people outside the organization.

See also:

– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*
– *Contributors (page 3-9)*
– *Integrators (page 3-10)*
– *Web Developers and Webmasters (page 3-11)*
– *Administrators and Sub-Administrators (page 3-11)*

User Authentication

Users—especially if they are internal to the organization—are normally authenticated by the system. This means they need to make themselves known, typically using a login name and password. Based on these credentials and the corresponding assigned rights and permissions, the system decides what the users are allowed to do. All login names and passwords are stored in the Content Server system itself and/or an external security database.

Note: For details on user authentication and security refer to the *Managing Security and User Access* guide.

See also:

– *User Base (page 3-6)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*
– *Contributors (page 3-9)*
– *Integrators (page 3-10)*
– *Web Developers and Webmasters (page 3-11)*
– *Administrators and Sub-Administrators (page 3-11)*

User Types

From an implementation point of view, five types of users are involved in a content management application:

- Consumers (page 3-8)
- Contributors (page 3-9)
- Integrators (page 3-10)
- Web Developers and Webmasters (page 3-11)
- Administrators and Sub-Administrators (page 3-11)

Each of these user types has a different role in the content management process. Individuals within an organization may perform multiple roles, and each role may be performed by multiple individuals.

See also:

- *User Base (page 3-6)*
- *User Authentication (page 3-7)*
- *Consumers (page 3-8)*
- *Contributors (page 3-9)*
- *Integrators (page 3-10)*
- *Web Developers and Webmasters (page 3-11)*
- *Administrators and Sub-Administrators (page 3-11)*

Consumers

Consumers are users who access the content server through their web browser to search, find, and view files. They cannot check new content into Content Server. Typically, the majority of users are consumers. Consumers do not need a user name and password to access files, unless a specific security configuration has been set up. If that is the case, they can only access files which they are permitted to see, based on their login credentials. Many systems are set up to prevent access to anyone without a login.

There are two types of content consumers:

- Consumers with Logins (Internal)
- Public Consumers (External)

Consumers with Logins (Internal)

Consumers with logins are internal to an organization. They log into the system and can access content based on their user credentials. Consumers with logins can be authenticated by the network. This user type is primarily relevant to intranet and extranet sites, where a named group of users is granted access to the intranet of an enterprise.

Public Consumers (External)

Public consumers are external to an organization. They do not need to make themselves known to Content Server to have access to the content. This user type is primarily relevant to Internet sites, where any external user can access the website. Public consumers cannot be authenticated by the network; they typically log into a different security provider (for example, LDAP or an RDBMS database provider).

See also:

– *User Base (page 3-6)*
– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Contributors (page 3-9)*
– *Integrators (page 3-10)*
– *Web Developers and Webmasters (page 3-11)*
– *Administrators and Sub-Administrators (page 3-11)*

Contributors

Contributors are users who—like consumers—can find, view, and print files, but can also check new content into Content Server or revise existing content. To safeguard the integrity of the files, contributors typically require a user name and password to check files in and out of the system.

Knowing who needs to contribute content is important for two main reasons:

- It determines where and how the contribution site(s) is/are set up and how security is organized.
- It plays a key role in the pricing structure of the Content Server content management software.

Contributors log into the system, which authenticates them. After their user credentials have been verified, they can access and add content based on the rights and privileges assigned to them. This user type is primarily relevant to intranet and extranet sites, where a named group of internal or external users can contribute content to the intranet of an enterprise. Internet site scenarios do not usually allow visitors to contribute content directly to the website.

See also:

– *User Base (page 3-6)*
– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*
– *Integrators (page 3-10)*
– *Web Developers and Webmasters (page 3-11)*
– *Administrators and Sub-Administrators (page 3-11)*

Integrators

Integrators deploy the Content Server system both during the initial installation and the maintenance updates. They typically install the software and ensure that it functions properly with the surrounding infrastructure elements such as hardware, third-party software products, communication protocols, and security.

See also:

– *User Base (page 3-6)*
– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*
– *Contributors (page 3-9)*
– *Web Developers and Webmasters (page 3-11)*
– *Administrators and Sub-Administrators (page 3-11)*

Web Developers and Webmasters

Web developers and webmasters customize Content Server's out-of-the-box functionality and look-and-feel of the website to suit the specific needs of their organization. They create and enable custom components to modify Content Server's standard functionality. They may do this using design tools such as Oracle Content Publisher, Oracle Dynamic Converter templates, Oracle's proprietary scripting language Idoc Script, or other, third-party tools.

See also:

– *User Base (page 3-6)*
– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*
– *Contributors (page 3-9)*
– *Integrators (page 3-10)*
– *Administrators and Sub-Administrators (page 3-11)*

Administrators and Sub-Administrators

Administrators set up, maintain, and modify the configuration of the content management system and its user logins. To safeguard the integrity of the system, administrators require a user name and password to access the system. Common tasks for an administrator include configuring the system to manage and index files, archiving and replicating information, working with content server security, adjusting system properties, reviewing log files, etc.

Enterprise administrators can set up sub-level administrators to perform a subset of administrative tasks within applications, specific departments, or security groups. These sub-administrators maintain a portion of a management system and its user logins.

See also:

– *User Base (page 3-6)*
– *User Authentication (page 3-7)*
– *User Types (page 3-8)*
– *Consumers (page 3-8)*

– *Contributors (page 3-9)*
– *Integrators (page 3-10)*
– *Web Developers and Webmasters (page 3-11)*

AUTHENTICATION TYPES

Users belong to a certain authorization type, which defines how they are authenticated by the system. There are three authorization types:

❖ Local Users (page 3-13)
❖ Global Users (page 3-13)
❖ External Users (page 3-14)

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*
– *Backup Strategy (page 3-72)*

Local Users

Local users are defined by an administrator or sub-administrator within a content server. The credentials of local users may extend to multiple content servers (for example, master and proxied servers). Administrators assign these users one or more roles, which provide them with access to security groups.

Local users can change their password, full name, e-mail address, and user type. The names of local users are included in the author drop-down lists in the content server user interface.

This type of user is recommended if there is a maximum of 1,000 users. If there are more, performance issues may become a problem, and global users should be used.

See also:

– *Global Users (page 3-13)*

– *External Users (page 3-14)*

Global Users

Global users are recommended for enterprise situations with more than 1,000 users. They are lightly-managed users whose credentials extend to multiple content servers. The validation of global users is always performed dynamically. The user profiles are created and stored on a master server, and retrieved by proxied servers. The content server roles are set on the master server, which provides access to security groups across multiple content server instances.

Global user credentials are not published to the web server security filter, so the master server always validates the credentials by querying the database tables. Because of this, the master server must be set up to log in as a global user.

Global users can only log in if the master server is running. They can change their full name, e-mail address, and user type. The names of global users do not appear in the author drop-down lists in the content server user interface.

See also:

– *Local Users (page 3-13)*

– *External Users (page 3-14)*

External Users

External users are users who are automatically registered in the system but are not manually set up by an administrator. These users might use a Microsoft login or some other type of provider login (for example, LDAP or Active Directory). Generally, these are users in a trusted domain to whom you grant access. These users cannot set their own password. Their passwords are owned by the Microsoft network domain or other type of provider.

External users can only log in if the content server is running. They cannot change their password. The names of external users do not appear in the author drop-down lists of the content server, but they can participate in workflows and use subscriptions.

External users are defined outside the Content Server system and authenticated through external security. External users who are automatically registered in the system but are not manually set up by an administrator might use a Microsoft login or some other type of provider (such as LDAP) login.

Generally, these are users in a trusted domain to whom you grant access and do not manage through Content Server. Their password is owned by the Microsoft network domain or other type of provider.

See also:

– *Local Users (page 3-13)*

– *Global Users (page 3-13)*

HARDWARE SETUP

The hardware setup of a Oracle-based content management system is determined by the following factors:

❖ Number of Servers (page 3-15)

❖ Functional Distribution (Server Types) (page 3-16)

❖ System Availability (Server Redundancy) (page 3-18)

❖ Separation of Content Server Environments (page 3-19)

❖ Organizational Considerations (page 3-21)

❖ Master Servers and Proxied Servers (page 3-22)

❖ Multiple Masters on the Same Computer (page 3-25)

❖ Scaling Options (page 3-26)

See also:

– *Managed Content (page 3-3)*

– *Users (page 3-5)*

– *Authentication Types (page 3-12)*

– *Infrastructure (page 3-27)*

– *Publication and Data Exchange (page 3-35)*

– *Security (page 3-41)*

– *Search Solution (page 3-50)*

– *Conversion to PDF (page 3-51)*

– *Conversion to HTML (page 3-53)*

– *Conversion to XML (page 3-56)*

– *Conversion to WML or cHTML (page 3-59)*

– *Other Conversions (page 3-63)*

– *Workflows (page 3-63)*

– *Metadata (page 3-66)*

– *Backup Strategy (page 3-72)*

– *Backup Strategy (page 3-72)*

Number of Servers

The number of servers required for an application depends on four main factors:

❖ The degree of “functional distribution” (server types) (see page 3-16)

❖ The required system availability (server redundancy) (see page 3-18)

❖ Separation of Content Server environments (see page 3-19)

❖ Organizational and application-specific considerations (see page 3-21)

See also:

– *Functional Distribution (Server Types) (page 3-16)*

– *System Availability (Server Redundancy) (page 3-18)*

– Separation of Content Server Environments *(page 3-19)*

Functional Distribution (Server Types)

"Functional distribution" means that separate physical machines are used to perform different tasks within the content management application. Although, in principle, all content management tasks could be performed on a single server, it is often better to divide them over multiple servers.

This is good for a number of reasons:

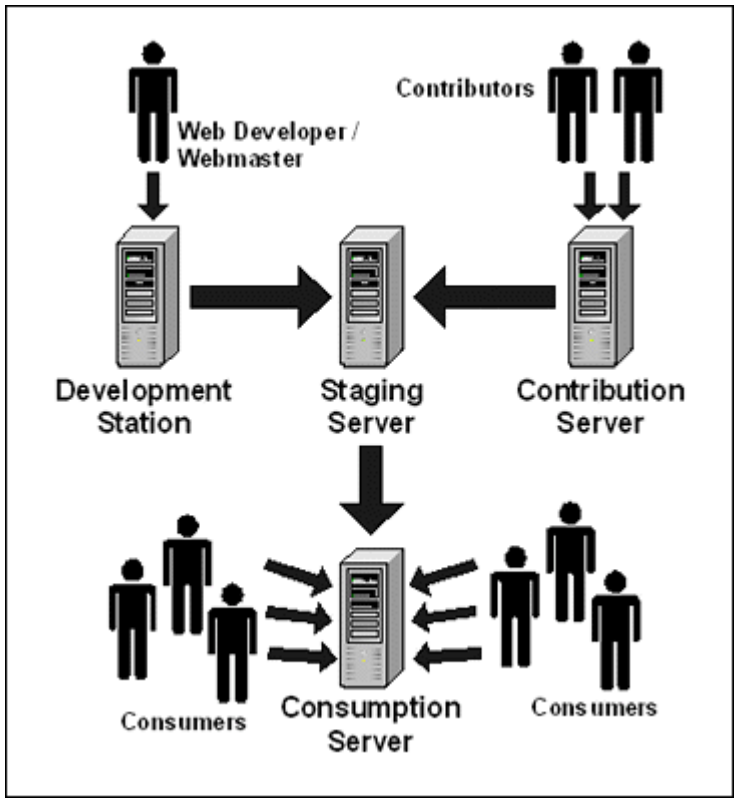
- It helps spread the overall system load over multiple servers.
- It allows security policies to be optimized.
- It enables servers to be used as functional "backups" of others—see System Availability (Server Redundancy) (page 3-18).
- It allows different hardware and software to be used for the various functions (for example, Microsoft Windows for the contribution function and UNIX for the consumption function).

The following main server functions can be distinguished in a content management system (as shown in Figure 3-1the figure below):

- **Consumption servers or deployment servers**
  These are servers whose only function it is to deliver web content to content consumers. They are the servers that content consumers access directly when they want to view content. These servers are often located outside a firewall in the DMZ (see page 3-33), and their web content is often replicated from contribution or staging servers.
- **Contribution servers**
  These are servers where users with the appropriate permissions check in new content or revise existing content. These servers are usually located inside a firewall.
- **Development stations**
  These are computers where webmasters and developers can perform all kinds of site-related tasks without actually having to touch the live site. This may include creating and testing new functionality or customizations, defining and testing interactions with other applications, etc.

- **Staging servers**
  These are servers where website content is temporarily stored for verification and approval purposes. This enables webmasters and others with the appropriate permissions to double-check web content before it is put on the “live” site, which all authorized users can access. Staging servers are typically used for Internet websites in conjunction with Oracle Content Publisher.

Figure 3-1 Server types



Depending on the size and complexity of the application, you may also choose to separate other content management tasks and assign separate servers to each of them. For example, you can use a separate database server to handle all of Oracle’s database storage and transactions, or a refinery server to handle all conversions. This helps spread the system load, which is especially important in high-volume applications.

It may not be necessary to split up all four of the main functions into separate servers. The consumption and contribution functions are often combined on one server, especially for simple document management scenarios. In web content management applications, the consumption server(s) will often be separate and located outside the firewall, whereas the contribution server(s) will be inside the firewall. Staging servers are most relevant in

Internet site applications, where web content may need to be verified before it goes "live"—for literally the whole world to see.

See also:

– *Number of Servers (page 3-15)*
– *System Availability (Server Redundancy) (page 3-18)*
– *Separation of Content Server Environments (page 3-19)*
– *Organizational Considerations (page 3-21)*
– *Master Servers and Proxied Servers (page 3-22)*
– *Multiple Masters on the Same Computer (page 3-25)*
– *Scaling Options (page 3-26)*

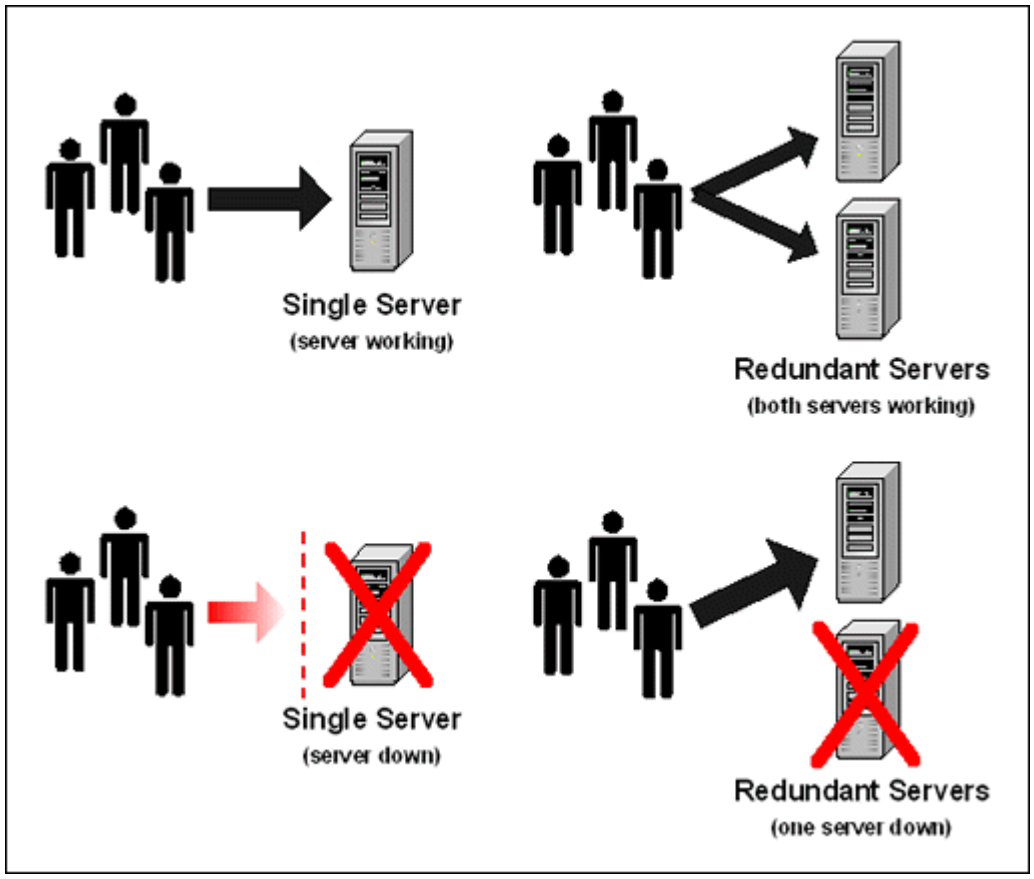
System Availability (Server Redundancy)

In certain applications, it may be crucial that users can access a website at all times. Under no circumstance may availability of the website be interrupted. In other words, there must be 100% "uptime." If only one server is used, this may be a problem. If the server fails due to hardware defect, for example, this would render the entire site inaccessible. To counteract this relative weakness, server redundancy may be set up (see Figure 3-2 the figure below). This means multiple servers are used to perform the same task.

Even though all server types—see Functional Distribution (Server Types) (page 3-16)—can be redundant in principle, it is typically consumption servers that are set up redundantly. This is done to ensure that content consumers (for example, visitors to an Internet site) can still access a website even if one site server goes down, for example due to hardware failure. Whether redundant servers are needed depends on the required system availability (or "uptime"). This may be a more important consideration for some applications than others, and it is up to the website operator to lay down the requirements in that area.

Another advantage of redundant servers is that the system load can be spread over all available servers ("load balancing"). This is especially useful in situations of high-volume traffic, when many users access the website at the same time. It prevents one server from being overloaded while another stands idle and could easily take some of the load.

Figure 3-2 Server redundancy



See also:

– *Number of Servers* *(page 3-15)*
– *Functional Distribution (Server Types)* *(page 3-16)*
– *Separation of Content Server Environments* *(page 3-19)*
– *Organizational Considerations* *(page 3-21)*
– *Master Servers and Proxied Servers* *(page 3-22)*
– *Multiple Masters on the Same Computer* *(page 3-25)*
– *Scaling Options* *(page 3-26)*

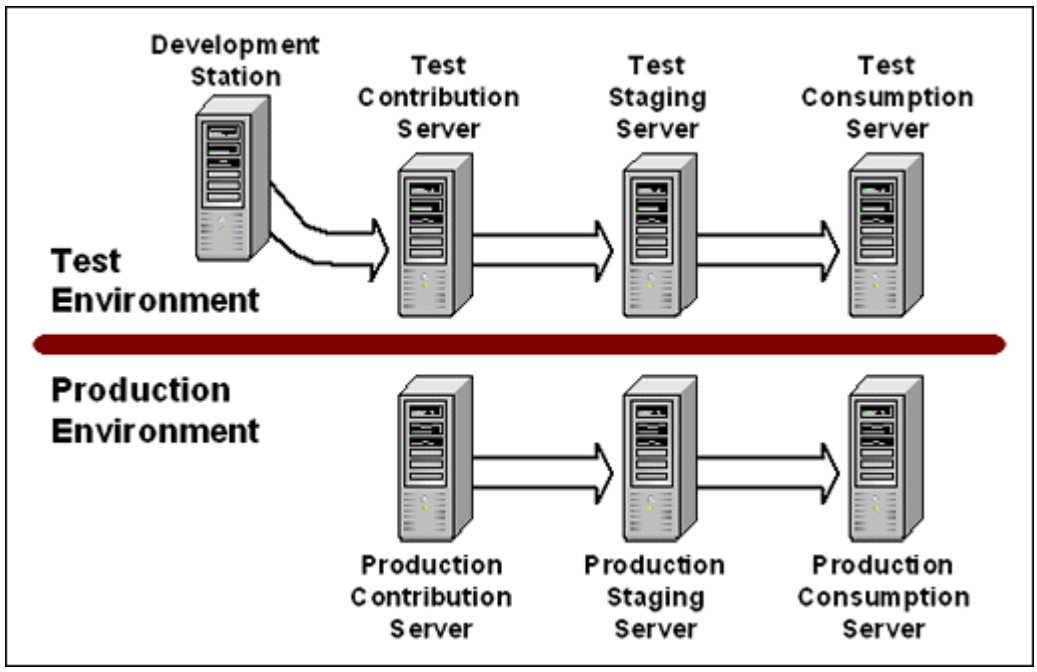
Separation of Content Server Environments

Companies may choose to use two entirely separate Content Server environments, one for development and/or testing purposes and one for “production” (that is, used in the real-life

business processes, including a “live” site that users really access). Figure 3-3 The figure below shows this separation of environments.

The main advantage of this setup is that updates, customizations, and other modifications to the Oracle content management application can be developed and tested in an off-line, unconnected environment that mimics the “live” environment, but is entirely separate from it. Only after everything has been thoroughly tested and verified will the production environment be updated. The test environment can then be used for new customizations or updates.

Figure 3-3 Separation of environments



A test environment can also be very useful for “stress testing”—i.e., assessing how the application holds up under high loads. Depending on the application, it may be important to determine the system’s performance and ability to handle various load levels. However, you probably do not want to perform such tests on a “live” production environment. For optimum testing, you then need an exact copy of the production environment, with all its hardware and software configurations. Only then can you perform accurate, “real-life” tests.

See also:

– *Number of Servers (page 3-15)*
– *Functional Distribution (Server Types) (page 3-16)*

Organizational Considerations

The number of servers required for an application also depends on a number of organizational and application-specific considerations:

Number of Content Items

Content Server can manage large numbers of content items. The recommended maximum for a single content server is about one million items. If the number of content items to be managed exceeds that number, the application should preferably be split up into multiple content servers.

Site Usage

Site usage comprises more than just the number of users or visitors to a website. It also includes the intensity of use (i.e., the frequency and duration of site visits as well as the level of content access). The combination of visitor number and access intensity determines the overall site traffic. If this traffic is high, multiple servers may be needed to handle all system loads correctly.

Metadata Model

The metadata model of a content management application may have an impact on the number of servers required. For example, individual departments may want to use their own metadata models specifically geared to their needs. If separate servers are then used

for each of the departments, this prevents other departments from having to deal with metadata which is totally irrelevant to them.

Security Model

The security model may affect the number of servers required. If, for example, departments want to control the security of their documents at the department level, multiple servers are required to accomplish this.

Organizational Structure

Finally, the number of servers needed may be influenced by the structure of the organization or enterprise served by the content management system. For example, an organization may be set up in such a way that it consists of a number of semi-independent entities. Each entity may then have its own content servers or set of servers.

The accountability structure within an organization may also be a reason to set up multiple servers, as well as other organizational or business-related considerations.

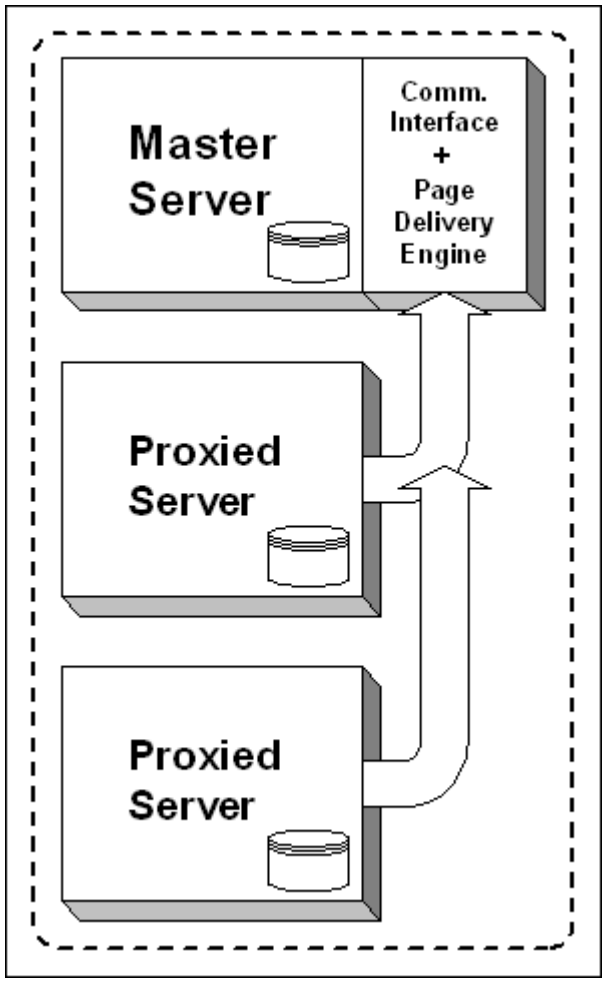
See also:

– *Number of Servers (page 3-15)*
– *Functional Distribution (Server Types) (page 3-16)*
– *System Availability (Server Redundancy) (page 3-18)*
– *Separation of Content Server Environments (page 3-19)*
– *Master Servers and Proxied Servers (page 3-22)*
– *Multiple Masters on the Same Computer (page 3-25)*
– *Scaling Options (page 3-26)*

Master Servers and Proxied Servers

Content Server supports the use of multiple servers on the same computer, with one server designated as the “master” and one or more others as the “proxied server(s).” Proxied servers are separate content server instances which have their own content database, native-file repository (“Vault”), and web-viewable file repository (“Web Layout”), but share the main communication interface and page delivery engine with the master server (see Figure 3-4the figure below).

Figure 3-4 Master server and proxied servers



Master/proxied server configurations can be useful in situations where you want to create distributed content management environments:

- You want to set up a Content Server environment with distributed user management. You may, for example, want to set up a separate content server for each department in a company, each with its own user definitions and setup, but with a shared security model.
- You want to set up a Content Server environment with distributed management of corporate functions, while maintaining centralized user management. You may, for example, want to set up multiple content servers, each serving a specific corporate function, such as document management, website management, and customer relationship management. If users are defined on the master server, they will

automatically also have access to the other proxied servers (within the bounds of their privileges). It does not matter which server they use to log in.

- You want to set up a Content Server environment with distributed content management. This will typically be the case if the application needs to manage huge numbers of files (more than one million). You then need to split up the content management system into a master server and one or more proxied servers, each taking care of part of the volume.
- A master and proxied server can also be useful if you want Content Server's search index to handle documents in both Western European languages (for example, English, French, and German) and an Asian language (for example, Japanese). Since there can be only one search engine locale per instance, you need multiple instances to achieve this. You could, for example, set up a master server to handle the Western European documents, and a proxied server to handle the Asian-language documents.

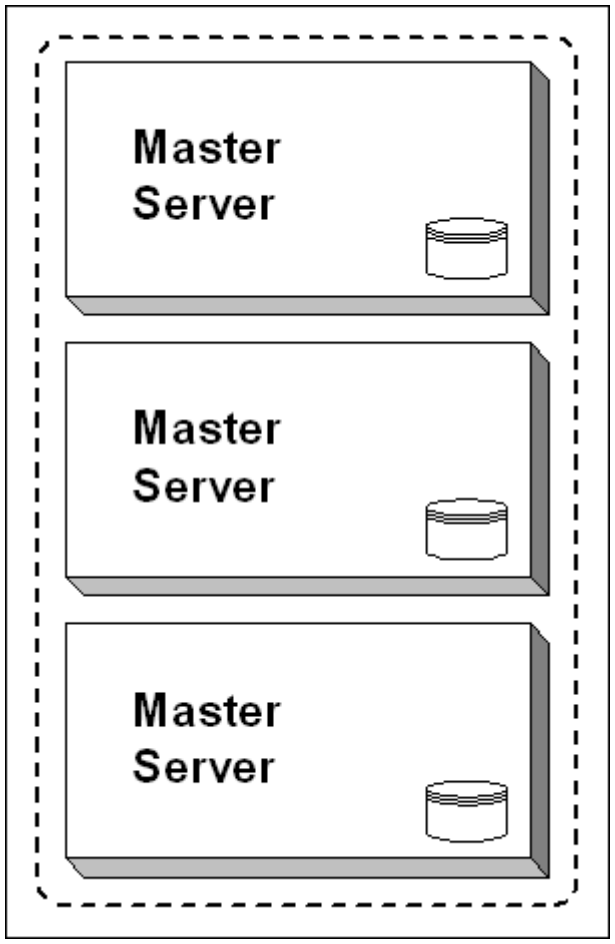
See also:

– *Number of Servers (page 3-15)*
– *Functional Distribution (Server Types) (page 3-16)*
– *System Availability (Server Redundancy) (page 3-18)*
– *Separation of Content Server Environments (page 3-19)*
– *Organizational Considerations (page 3-21)*
– *Multiple Masters on the Same Computer (page 3-25)*
– *Scaling Options (page 3-26)*

Multiple Masters on the Same Computer

Content Server supports the use of multiple master servers on the same computer. Each master is basically an entirely separate entity, with its own metadata and security model, content database, native-file repository (“Vault”), and web-viewable file repository (“Web Layout”) (see Figure 3-5 the figure below). Multiple-master server configurations can be useful if you want to use a single computer and web server for your content management solution to run several, independent content servers (see also Scaling Options on page 3-26).

Figure 3-5 Multiple masters on the same computer



See also:

– *Number of Servers (page 3-15)*

– *Functional Distribution (Server Types) (page 3-16)*

Scaling Options

Oracle-based content management solutions are fully scalable. This means they can grow with an organization's needs. Scaling usually means that additional content server instances are added as required.

This may be for any number of reasons including:

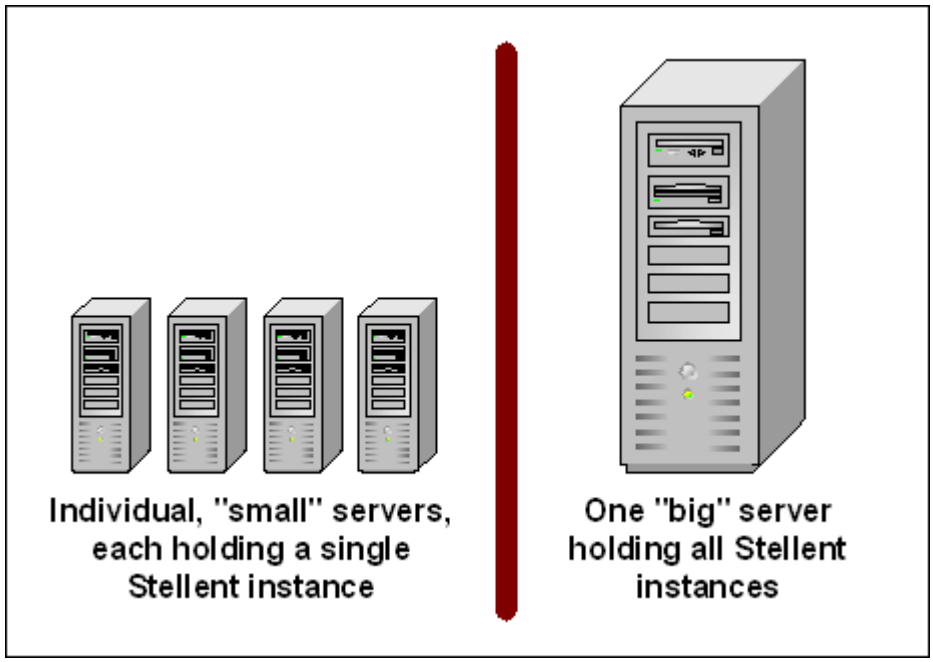
- The number of users increases.
- The content volume increases.
- System usage intensifies.
- The site scope broadens.

If several content servers instances are required, there are essentially two options (as shown in Figure 3-6the figure below):

- Use separate computers for each instance.
- Use a single computer (e.g. mainframe) with multiple instances.

Each of these scenarios has its own characteristics and benefits, depending on the application. Contact your Oracle representative for more information.

Figure 3-6 Scaling options



See also:

– *Number of Servers (page 3-15)*
– *Functional Distribution (Server Types) (page 3-16)*
– *System Availability (Server Redundancy) (page 3-18)*
– *Separation of Content Server Environments (page 3-19)*
– *Organizational Considerations (page 3-21)*
– *Master Servers and Proxied Servers (page 3-22)*
– *Multiple Masters on the Same Computer (page 3-25)*

INFRASTRUCTURE

A content management environment consists of a number of (hardware) components that all need to be integrated into the enterprise network. They include the following:

❖ Firewall (page 3-28)
❖ Application Server (page 3-30)
❖ Web Server (page 3-31)

- Database (page 3-31)
- Mail Server (page 3-32)
- DMZ (page 3-33)
- Load Balancing (page 3-34)
- Clustering (page 3-34)

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*
– *Backup Strategy (page 3-72)*

Firewall

In today's highly interconnected, global business environment, firewalls are indispensable tools to protect a company's network resources and control access to them. Without firewalls, network security may easily be jeopardized. The challenge is to find a strategy which allows the right data to pass through the firewall without compromising network security, but at the same time not opening up the network too much.

In the context of Oracle content management, the main question is how to push (published) content through the firewall, often to a consumption server in the DMZ (see page 3-33) in order to provide consumer access to a website.

There are basically four ways to push data through a firewall:

- Direct File Share Access
- File Transfer Protocol (FTP)
- Secure Socket
- HyperText Transfer Protocol (HTTP)

Depending on the communication method selected (see also Publication and Data Exchange on page 3-35), certain ports may need to be opened for data to pass through the firewall. Company-internal security policies and considerations may narrow down the choice of methods.

Direct File Share Access

With this method, data is typically pushed through a firewall using mounted drives and written directly to a file system. It is important to note that this method may pose potential security risks. It is therefore only recommended if you are on a secure network publishing a website through Oracle Content Publisher.

File Transfer Protocol (FTP)

With this method, data is pushed through a firewall using the standard TCP/IP protocol and FTP ports 20/21. This method is particularly useful if you only allow writing to a web server's file system through FTP. It is primarily used to publish a site onto a consumption server using Oracle Content Publisher.

Secure Socket

With this method, data is pushed through a firewall using a secure socket (port 4444 by default). This method is primarily used for archiving or replicating content between two content servers using the Archiver utility.

HyperText Transfer Protocol (HTTP)

With this method, data is pushed through a firewall using the standard HTTP protocol (port 80 by default). The data exchange is entirely controlled by the web server. This method can be used for archiving or replicating content between two content servers using

the Archiver application, or for transferring content from Oracle Content Publisher to Content Server, where the content is subsequently checked in.

See also:

– *Application Server (page 3-30)*
– *Web Server (page 3-31)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Load Balancing (page 3-34)*
– *Clustering (page 3-34)*

Application Server

An application server is basically the middleware glue that enables browser-based applications to link to multiple sources of information (one of which could be Content Server). It performs the data processing necessary to deliver up-to-date information as well as process information for web clients.

In the context of Oracle content management, application servers are typically used in conjunction with portal technologies using portlets, where they retrieve information (content) from Content Server and use that content to populate one or more portlets on the website.

Note: For further information on Content Server and portlets refer to the *Getting Started With Content Server* guide. For details on the application servers supported by Content Server refer to the Content Server installation guides.

See also:

– *Firewall (page 3-28)*
– *Web Server (page 3-31)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Load Balancing (page 3-34)*
– *Clustering (page 3-34)*

Web Server

A web server manages web page requests from the browser and delivers HTML documents (web pages) in response. The web server also executes server-side scripts (CGI scripts, JSPs, ASPs, etc.) that provide functions such as database searching.

In the context of Oracle content management, the web server is located in between the web client (browser) and the Oracle content server. It handles all requests from the web browser, passes them on to the content server, and delivers the results it gets back from the content server to the web browser. The web server is often located in the DMZ of the network (see page 3-33).

Note: For further information on the role of the web server in the Content Server architecture refer to the *Getting Started With Content Server* guide. For details on the web servers supported by Content Server refer to the Content Server installation guides.

See also:

– *Firewall (page 3-28)*
– *Application Server (page 3-30)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Load Balancing (page 3-34)*
– *Clustering (page 3-34)*

Database

Databases are used to store information in a structured manner. In the context of Oracle content management, a database is used to store the metadata of all revisions of all the content managed by Content Server. It is used to describe and access the content.

In addition to the metadata, the database also stores a wide variety of other Oracle-related data, including the user profiles (with the logon and security information for each individual user), workflow definitions, subscriptions, and historical data.

By default, new installations of Content Server are set up to use the database to provide search capabilities. The out-of-the-box database configuration provides metadata-only searching, but you can modify the default configuration for some of the supported

databases to enable full-text searching. For further details refer to the Content Server installation guides.

Depending on the search solution used, the full text of content may also be stored in the database.

Note: For further information on the role of the database in the Content Server architecture refer to the *Getting Started With Content Server* guide.

Note: For details on the databases supported by Content Server refer to the Content Server installation guides.

See also:

– *Firewall (page 3-28)*
– *Application Server (page 3-30)*
– *Web Server (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Load Balancing (page 3-34)*
– *Clustering (page 3-34)*

Mail Server

An e-mail server provides “post office” facilities in a network. It stores incoming mail for distribution to users and forwards outgoing mail through the appropriate channel.

In the context of Oracle content management, e-mail is primarily used in workflows and subscriptions. In workflows, users receive e-mail to notify them of something they need to do with the content item in the workflow. With subscriptions, users are notified by e-mail if changes are made to a content item they are subscribed to.

Content Server supports any mail server that is SMTP-compliant. The mail server must have SMTP mail routing enabled. In addition, the server hosting Content Server must have the SMTP gateway open.

Note: For further information on workflows and subscriptions refer to the *Getting Started With Content Server* guide.

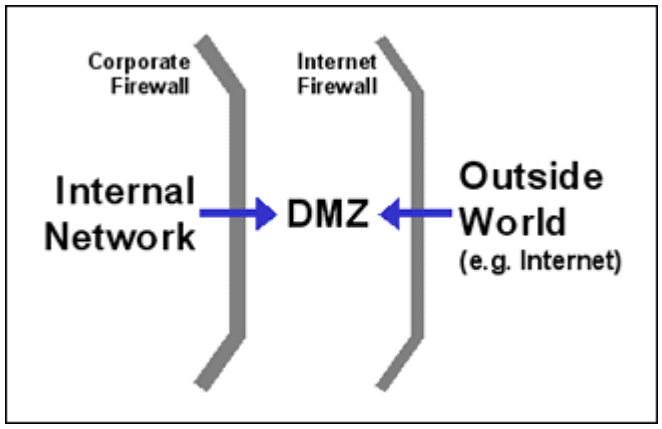
See also:

– *Firewall (page 3-28)*

DMZ

The “demilitarized zone” (DMZ) of a network is the area which is outside the main firewall but part of the internal network (see Figure 3-7the figure below). As such, it serves as a “buffer” between an organization’s trusted internal network and an untrusted, external network such as the Internet. It is more secure than the Internet, but not as secure as the internal network.

Figure 3-7 DMZ



The main purpose of a DMZ is to prohibit any direct data transmission between the Internet and the secure core network. All data must first go through servers in the DMZ, which means there is an extra line of defense against unauthorized access.

The DMZ gives access to services required from both the secure, internal network and the insecure, external network. In Oracle-based content management, the DMZ will typically accommodate all “inward and outward facing” servers—that is, servers that need to be accessed from both the internal network and any external networks (for example, the Internet). These servers may include one or more web servers, consumption servers, mail servers, and authentication servers.

See also:

– *Firewall (page 3-28)*
– *Application Server (page 3-30)*
– *Web Server (page 3-31)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *Load Balancing (page 3-34)*
– *Clustering (page 3-34)*

Load Balancing

Load balancing is the fine-tuning of a network in order to distribute the data and/or processing more evenly across available resources. For example, load balancing might distribute the incoming transactions evenly to all servers, or it might redirect them to the next available server.

In Oracle-based content management, load balancing is done by a custom shared disk installation with a load balancer directing requests to the appropriate web server.

Note: For more information on load balancing refer to the *Clustering Concepts Guide*.

See also:

– *Firewall (page 3-28)*
– *Application Server (page 3-30)*
– *Web Server (page 3-31)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Clustering (page 3-34)*

Clustering

Clustering is the act of connecting two or more computers together in such a way that they behave like a single computer. In Oracle-based content management, a clustered,

multi-server configuration uses a common content server file system, database, and search index collection. A load balancer can be used to provide high availability for consumption.

Note: For more information on clustering refer to the *Clustering Concepts Guide*.

See also:

– *Firewall (page 3-28)*
– *Application Server (page 3-30)*
– *Web Server (page 3-31)*
– *Database (page 3-31)*
– *Mail Server (page 3-32)*
– *DMZ (page 3-33)*
– *Load Balancing (page 3-34)*

PUBLICATION AND DATA EXCHANGE

There is a lot of data exchange within a Oracle-based content management application. Data is copied and moved within the Oracle content server, but it may also be transferred to a location outside the content server (often in web-viewable form). There are various methods to transfer all this data. The available methods depend on the source and target of the data exchange as well as its purpose. Security considerations may also play a role.

Note: For details on security refer to the *Managing Security and User Access* guide.

The following data exchange scenarios are typical for Content Server:

- **Content Publisher to File System** (page 3-36)
  Oracle Content Publisher extracts content from Content Server and publishes the generated web content to a file system.
- **Content Publisher to Content Server** (page 3-38)
  Oracle Content Publisher extracts content from Content Server and transfers the generated web content to another content server, where the web content is subsequently checked in.
- **Content Server to Content Server** (page 3-39)
  Content Server archives or replicates checked-in content to another content server.

❖ **Content Server to File System** (page 3-40)
Content Server backs up checked-in content to a file system.

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*

Content Publisher to File System

Content Publisher is a powerful optional feature set to Content Server, which provides advanced template-based technology to automatically publish documents or content as well-designed, fully linked websites, including full site navigation.

Oracle Content Publisher periodically extracts content from a Oracle content server and uses advanced templates to convert it to web resources such as HTML, WML, cHTML, or XML pages and image files. It can then publish the generated web content to a directory hierarchy on a file system, typically on a web server. The web server can process the content further, or make it available on the website for content consumers to access. Oracle Content Publisher is mostly used in conjunction with extranet or Internet sites.

This publication method can be useful in a number of situations, including:

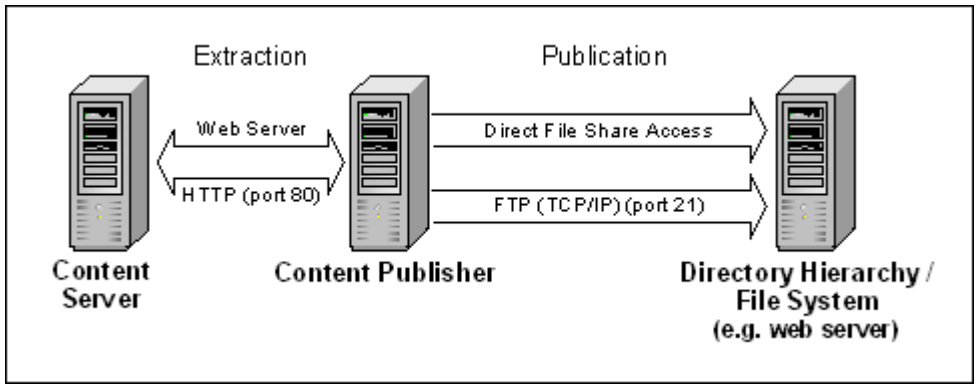
❖ You already have a website structure in place, and you want to use Content Server to add to that structure.

- You want to retain full control over the website. If required, the web server or other backend systems can process the content further, supply search functionality, optimize security, etc.

As shown in Figure 3-8the figure below, Content Publisher uses the standard HTTP protocol (through the web server) to extract data from the content server. It can use either of two methods to publish the generated web content to the file system:

- **Direct file share access**
  With this method, web content is directly written to a shared network location, typically using mounted drives. It is important to note that this may pose potential security risks, especially if the data is to be pushed through a firewall. It is therefore only recommended if you are publishing a website on a secure network.
- **TCP/IP (FTP ports 20/21)**
  With this method, web content is published using the standard TCP/IP protocol and FTP ports 20/21. In most cases, this will be the preferred method as it is much more secure than direct file share access.

Figure 3-8 Using Content Publisher to publish content to a file system



See also:

– *Content Publisher to Content Server (page 3-38)*
– *Content Server to Content Server (page 3-39)*
– *Content Server to File System (page 3-40)*

Content Publisher to Content Server

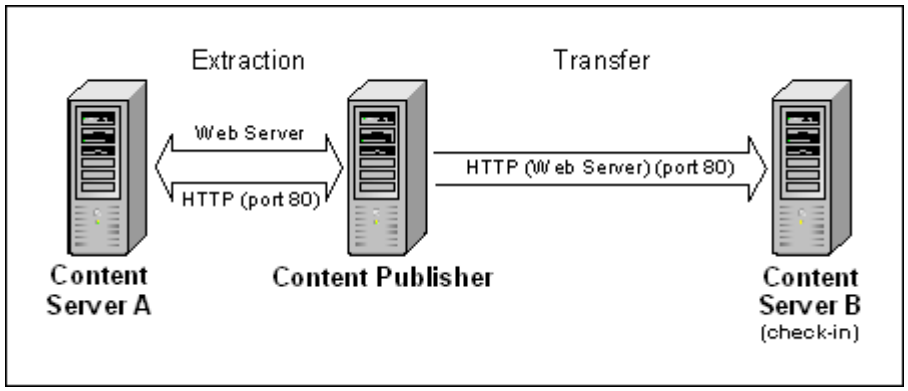
In addition to publishing web content directly to a file system, Oracle Content Publisher can also transfer generated web content to a content server, where it is subsequently checked in. The checked-in web content can then be processed by Content Server, if required, and made available to content consumers.

This has a number of advantages.

- The generated web content can be HCSP, HCST, JSP or ASP pages with Idoc Script in them. These pages are processed by Content Server, which can then output highly dynamic HTML pages. This enhances personalization of the website and can provide context-sensitive web content.
- You can use Oracle's built-in search capability; no external search engine is required.
- You can use Oracle's built-in security features; no external security provider is required.
- Content Server provides secure management and version control of the web resources used to build the site.

As shown in Figure 3-9the figure below, Content Publisher uses the standard HTTP protocol (through the web server) both for extracting data from one content server and transferring it to the other.

Figure 3-9 Using Content Publisher to transfer web content to a content server



See also:

– *Content Publisher to File System (page 3-36)*
– *Content Server to Content Server (page 3-39)*
– *Content Server to File System (page 3-40)*

Content Server to Content Server

It is possible to transfer checked-in content between two content servers. This is done using the transfer functionality of the Archiver utility. This can be useful in a number of situations:

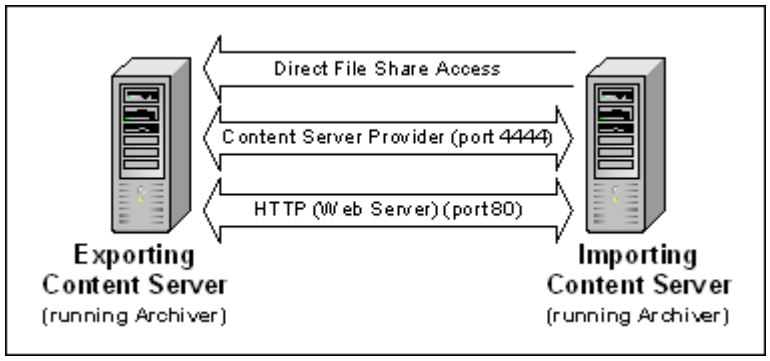
- You can use Archiver to synchronize two content servers.
- In applications without Oracle Content Publisher, Archiver can be used to transfer data from a contribution server to a consumption server. This may be useful in situations where not all the content checked into the contribution server should be available to content consumers. The Archiver utility can then be used to transfer only a subset of the content from the contribution server to the consumption server.

The data transfer process can be automated. Content items are then transferred to the other content server automatically, either at regular intervals or whenever certain metadata criteria are met (for example, all new content items checked into particular security groups).

As shown in Figure 3-10the figure below, there are three methods that can be used to archive or replicate content between content servers:

- **Direct file share access**
  With this method, the importing content server copies data directly from the file system of the exporting content server. It is important to note that this may pose potential security risks, especially if the data is to be pushed through a firewall. It is therefore only recommended if you are transferring data on a secure network.
- **Content Server provider**
  With this method, data is transferred between content servers using a secure socket (port 4444 by default).
- **HTTP (web server)**
  With this method, data is transferred between content servers using the standard HTTP protocol (port 80 by default). The data exchange is entirely controlled by the web server.

Figure 3-10 Transferring data between content servers



See also:

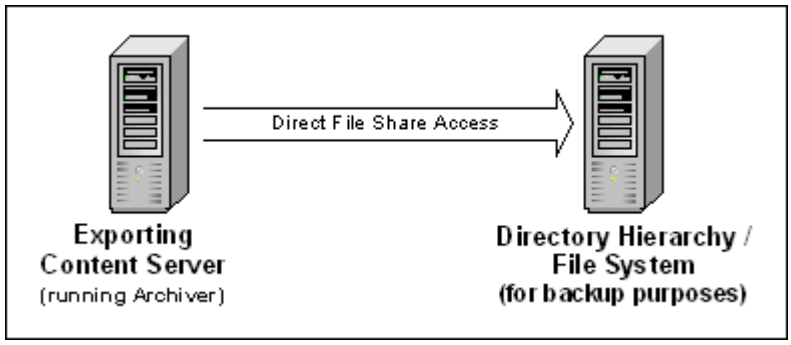
– *Content Publisher to File System* *(page 3-36)*
– *Content Publisher to Content Server* *(page 3-38)*
– *Content Server to File System* *(page 3-40)*

Content Server to File System

Content Server can also transfer checked-in content to a file system for backup purposes. This is done using the Archiver utility. In addition to backing up all checked-in content, you can also back up subsets of content. This enables you, for example, to back up obsolete content and remove it from the main repository to save disk space or improve performance.

As shown in Figure 3-11the figure below, Content Server (with Archiver) uses direct file share access to back up content, which means the files are written directly to a shared network drive or other shared location. It is important to note that this may pose potential security risks, especially if the data is to be pushed through a firewall. It is therefore recommended to back up content within a secure network only.

Figure 3-11 Backing up data from a content servers to a file system



See also:

– *Content Publisher to File System (page 3-36)*

– *Content Publisher to Content Server (page 3-38)*

– *Content Server to Content Server (page 3-39)*

SECURITY

Security plays a crucial role in Oracle-based content management applications. The following topics are important with regard to security:

- Security Concepts (page 3-42)
- Security Models (page 3-43)
- Creating a Security Model (page 3-43)
- Determining if You Require Accounts (page 3-44)
- Standard Security Model (page 3-45)
- Accounts-Based Security Model (page 3-47)

See also:

– *Managed Content (page 3-3)*

– *Users (page 3-5)*

– *Authentication Types (page 3-12)*

– *Hardware Setup (page 3-14)*

– *Infrastructure (page 3-27)*

– *Publication and Data Exchange (page 3-35)*

Security Concepts

Security in Oracle-based content management solutions centers around three main concepts:

- **Content server security**
  Users are authenticated through the content server itself, either locally or externally. For details on content server security refer to the *Managing Security and User Access* guide.
- **Content-level security**
  Users are authenticated at a project level, folder level, or content-item level, through the content server itself, either locally or externally. For details on content-level security refer to the Collaboration Manager documentation.
- **External security**
  Users are authenticated through the network by a network protocol (Microsoft challenge response), Active Server Directory, LDAP, or an RDBMS database provider. For details on external security refer to the *Managing Security and User Access* guide.

See also:

Security Models

If the content server's predefined security elements cannot provide the level of security that is required for your application, then you must design a custom security configuration.

Important: Proper planning and design of the security model cannot be overemphasized!

It is extremely important that your users understand the security model and metadata fields. If contributors do not assign the correct security and metadata to content items, other users might not be able to find or view the files, or users might be able to view files they are not supposed to have access to. A good rule of thumb is to evaluate whether the security and metadata model can be explained to a user in 10 minutes, or in one written page. If the model is more complicated than this, there is a risk that users will not use the system correctly, which will result in content access issues. Make sure that your implementation plan includes thorough security and metadata model development, end-user training, and documentation.

To demonstrate how a security configuration is developed, we use two examples: the first one, Standard Security Model (page 3-45), applies the security group, role, and user features to a security configuration. The next example, Accounts-Based Security Model (page 3-47), introduces accounts.

See also:

– *Security Concepts (page 3-42)*
– *Creating a Security Model (page 3-43)*
– *Determining if You Require Accounts (page 3-44)*
– *Standard Security Model (page 3-45)*
– *Accounts-Based Security Model (page 3-47)*

Creating a Security Model

To create your security model, you need to do the following:

1. Determine the number of users.
2. Determine if the users will be local, global, or external. For details refer to Authentication Types (page 3-12).
3. Determine how the users will be authenticated, either in Content Server or by an external system (for example, Microsoft security or LDAP).

4. Create the security groups.
5. Define the roles.
6. Set the permissions for both security groups and roles.

See also:

– *Security Concepts (page 3-42)*
– *Security Models (page 3-43)*
– *Determining if You Require Accounts (page 3-44)*
– *Standard Security Model (page 3-45)*
– *Accounts-Based Security Model (page 3-47)*

Determining if You Require Accounts

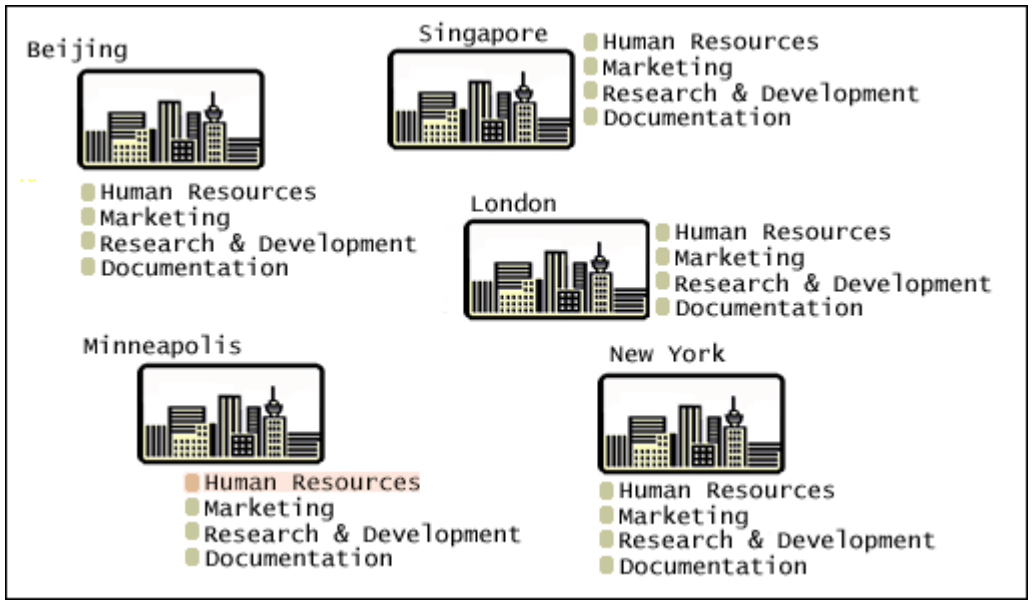
Content Server provides two levels of content security: security groups (required) and accounts (optional). Generally, use accounts when you require more than 50 security groups. For example, if you have 1,000 vendors, and you want each vendor to access content that is only for them, then you need accounts.

Consider the following scenario (as shown in Figure 3-12 in the figure below):

- Company ABC has five offices in Minneapolis, Singapore, Beijing, London, and New York.
- Each office has four main departments: Human Resources, Marketing, Research & Development, and Documentation.
- Some content is available only to Human Resource users.
- Some content is available only to Human Resources at the Minneapolis location.

Because the scenario requires a significant number of security groups, accounts would be used in the security configuration.

Figure 3-12 Security scenario



See also:

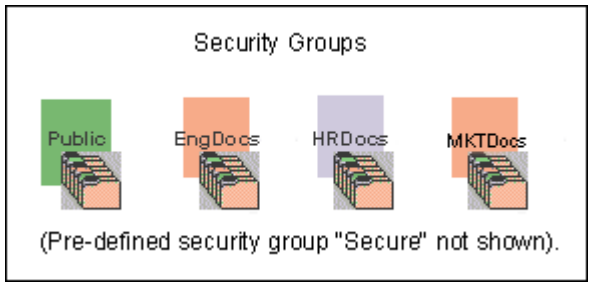
- *Security Concepts (page 3-42)*
- *Security Models (page 3-43)*
- *Creating a Security Model (page 3-43)*
- *Standard Security Model (page 3-45)*
- *Accounts-Based Security Model (page 3-47)*

Standard Security Model

In this example, a company wants to set up product security for three functional areas: HR, Engineering, and Marketing. The company requires some documents to be accessed among all functional areas, while others must be accessed only by personnel within their own department. Also, management wants to be able to view all documents. The guideline below demonstrates a typical sequence that configures the system to meet these requirements:

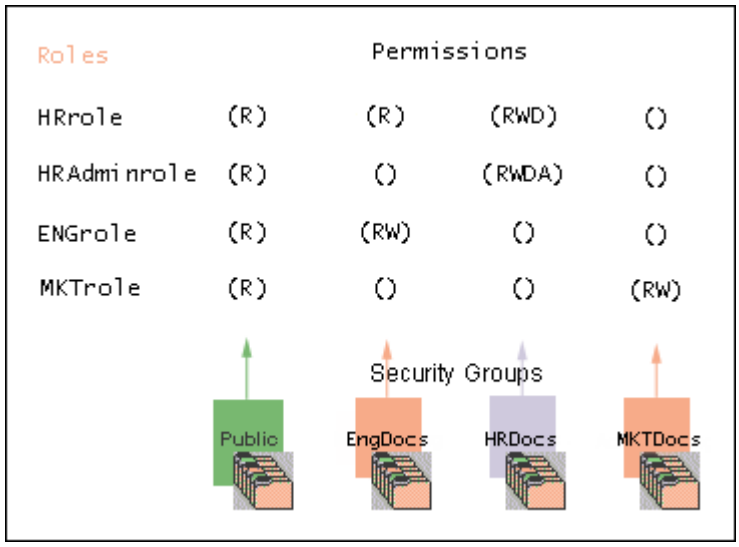
1. A security group is created for each department's files in User Admin (Security—Permissions By Group).

Figure 3-13 Security—creating security groups



2. Roles are created that control access to the security groups:

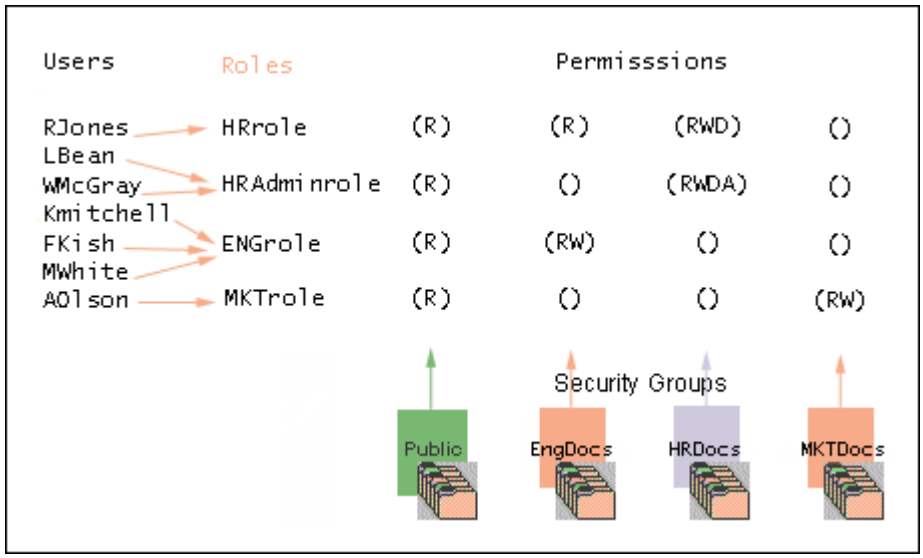
Figure 3-14 Security—creating roles



Note: Documents in the security groups that are accessible by the *guest* role can be read without a login.

3. Each user is assigned to one or many roles:

Figure 3-15 Security—assigning roles



This simple model demonstrates that:

- Initial planning is essential to achieve a properly-functioning security configuration.
- Many roles can be created for the set of security groups.
- Many users can be assigned one role.
- Many roles can be assigned to one user.

See also:

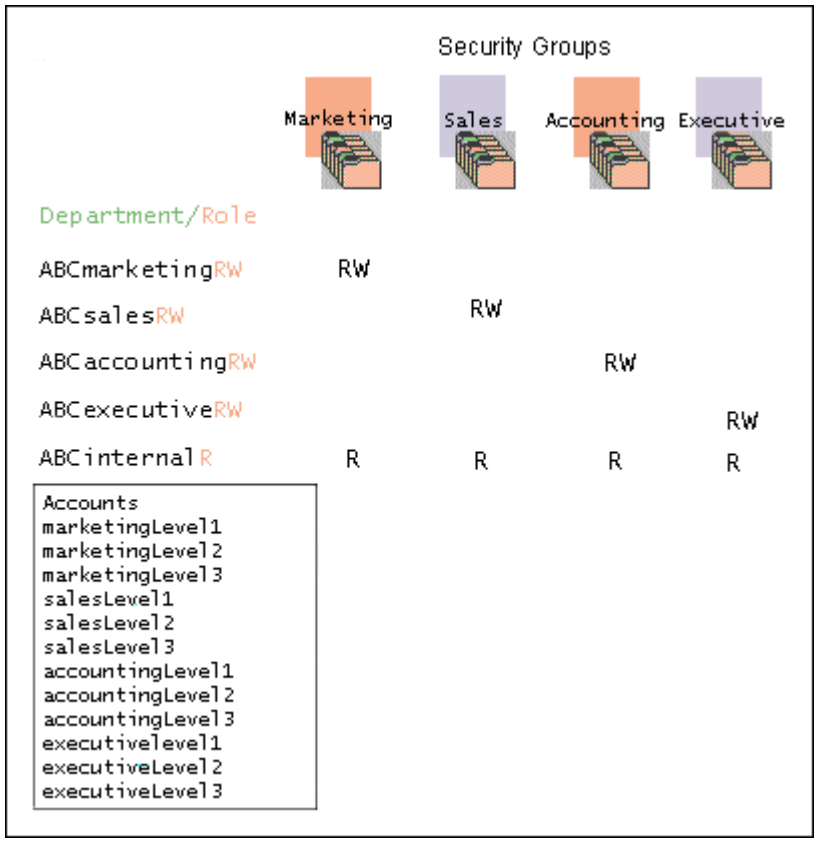
– *Security Concepts (page 3-42)*
– *Security Models (page 3-43)*
– *Creating a Security Model (page 3-43)*
– *Determining if You Require Accounts (page 3-44)*
– *Accounts-Based Security Model (page 3-47)*

Accounts-Based Security Model

In this example, company XYZ is providing a new content management system for their intranet. XYZ has content that needs to be secured at multiple levels throughout the company, as well as multiple levels within each department. This scenario will implement the use of Content Server accounts. In this example there is no need for the use of the Public security group because departments will have a company wide read-only area

within their departmental security group. The security matrix would be outlined as follows:

Figure 3-16 Accounts-based security



In this scenario, there is no true public content. All content is controlled either by the contributor of the content or by a member in the department. In this model, all of the security groups are assigned Read (R) access by the ABCinternal role. When content is checked in, the contributor associates the security group that they want with the content so they can control who can edit the content. In addition, they provide an account attribute that specifies levels so employees external to the department can secure the content within the security group for read access. Sample uses could be as follows:

User 1

User1 is a user in the Marketing department and has Read (R) and Write (W) permissions to Level 1, Level 2, and Level 3 content inside of the Marketing security group. User1 has permission to read any Level 1 content across the company, but has no permissions to any content that is not allowed past Level 1 in any department external to Marketing.

Roles:

- ABCmarketing(RW)
- ABCinternal(R)

Accounts:

- marketingLevel1(RW)
- marketingLevel2(RW)
- marketingLevel3(RW)
- salesLevel1(R)
- accountingLevel1(R)
- executiveLevel(R)

User 2

User2 is in the Sales department and has Read and Write permissions to Level 1 and Level 2 content inside of the Sales security group. User2 has permission to read any Level 1 content across the company as well as read Level 2 executive content.

Roles:

- ABCsales(RW)
- ABCinternal(R)

Accounts:

- salesLevel1(RW)
- salesLevel2(RW)
- marketingLevel1(R)
- accountingLevel1(R)
- executiveLevel1(R)
- executiveLevel2(R)

Note: The added complexity of this model is that contributors need to be trained to specify the correct account when checking in content. If a contributor does not specify an account, then anyone with Read access to the security group associated with the document will have Read access to the document itself.

See also:

– *Security Concepts (page 3-42)*

SEARCH SOLUTION

By default, new installations of Content Server are set up to use the database to provide search capabilities. The out-of-the-box database configuration provides metadata-only searching, but you can modify the default configuration for some of the supported databases to enable full-text searching. For further details refer to the Content Server installation guides.

You can also use an external search engine in place of the content database to provide full-text search functionality to the content server. Integration modules are available for the following external search solution:

- Verity VDK

Note: The distribution media for UCM 10.1.3.3.1 contained the VDK6 component; however, UCM versions after 10.1.3.3.1 do not include VDK. VDK-based indexing for UCM is an option only if:

- You have purchased and installed UCM prior to version 10.1.3.3.1.
- You are upgrading to a UCM 10gR3 version from a prior version (such as UCM 7.5.2), which included VDK.

The table below outlines the various search options that can be used with Content Server:

Metadata Searching Handled By	Full-Text Searching Handled By
Database*	— (no full-text searching available)*
Database	Database**
Database	External search engine (VDK)
External search engine (Verity or FAST)	External search engine (VDK)

* Default, out-of-the-box configuration for new installations.

** See the Content Server installation guides for configuration instructions.

It is recommended that you carefully consider what search solution you will be using with your content server. A useful document in this respect is *Choosing a Search Solution*, which discusses the various search options that can be used with Content Server, as well as the factors to consider when selecting a solution. It is available as a PDF file on the Content Server documentation CD (*\integrator\search_solutions_80en.pdf*), and is also included in Content Server's online help system.

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*

CONVERSION TO PDF

The Adobe Portable Document Format (PDF) is a standard for electronic document distribution worldwide. It is a universal file format that preserves all the fonts, formatting, graphics, and color of any source document, regardless of the application and platform used to create it. Adobe PDF files are compact and can be shared, viewed, navigated, and printed exactly as intended by anyone with the Adobe Acrobat® Reader® software. Using a special plug-in, which is installed with Acrobat Reader, PDF files can also be viewed

from any standard web browser without the need of the application the document was originally created in. This makes PDF files very suitable for Oracle-based content management solutions, which use standard web technology and browsers.

Note: Further details on the Portable Document Format (PDF), as well as downloadable copies of Adobe Acrobat Reader, can be found on Adobe's Internet website at http://www.adobe.com.

The Content Server suite of products includes an optional feature set to Content Server called PDF Converter, which converts checked-in content to the web-viewable PDF format. PDF Converter uses Adobe Distiller Server and a PostScript printer to convert files to PDF.

Implementation Considerations

The following implementation considerations are important for content conversion to PDF through PDF Converter:

- PDF Converter needs Inbound Refinery to run properly.
- Files can only be converted to PDF if their native applications (i.e., the applications used to create the files) are actually installed on the computer running PDF Converter (with Inbound Refinery). PDF Converter needs the native applications to perform the conversion.
- Content items are converted when they are checked into the content server. Both the native and converted files are stored, which means the PDF files are immediately available to content consumers as soon as they request a content item.

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*

– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*

CONVERSION TO HTML

The HyperText Markup Language (HTML) is used to display web pages in browsers. The Content Server suite of products includes two optional feature sets which can convert content to the web-viewable HTML format:

❖ Dynamic Converter (page 3-54)

❖ Content Publisher (page 3-55)

Each of these optional feature sets has its own characteristics and considerations. The combination of both provides optimum conversion in all situations.

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*

Dynamic Converter

Dynamic Converter is an optional feature set to Content Server. It provides on-demand conversion of content from over 225 native file formats to the web-viewable HTML format. Dynamic Converter uses rules-based templates to deliver a consistent look-and-feel for converted content. Metadata rules are used to determine which template should be applied to convert the native source file to HTML. The HTML renditions of content items are generated “on the fly” when a user clicks special HTML links on the search results pages or an item’s content information page.

Implementation Considerations

The following implementation considerations are important for content conversion to HTML through Dynamic Converter:

- Dynamic Converter provides on-demand conversion to HTML, i.e. only when content is specifically requested. This means that the HTML rendition always reflects the most recent version of the content.
- Dynamic Converter enables users to view content without having its native application installed on their computers.
- Dynamic Converter outputs HTML renditions that can retain all formatting of the original documents, if required.
- Dynamic Converter provides an easy, fast way to make content web-viewable, especially in situations where precise formatting is not the top priority.
- The conversion to HTML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).
- Dynamic Converter is particularly useful for conversion of content that changes a lot or that is not accessed very often (for example, large documents).
- Dynamic Converter is particularly good at handling large amounts of documents. This is because it does not use polling to detect new content, and it only converts content on demand (i.e., when a content item is requested for display).
- Dynamic Converter puts some load on the content server, because the conversion to HTML is done on the server itself (with Dynamic Converter running on it).
- Dynamic Converter relies on caching technology to reduce the load on the server and ensure that documents are not re-translated unnecessarily.

- Dynamic Converter is best able to effectively service the requests for conversion if the number of users is not too large. This is due to the fact that Dynamic Converter must run a service every time a user wants to view a document. If the document is not in the Dynamic Converter conversion cache, the conversion must be done again. The conversion process is very resource-intensive. If there are large numbers of users, it is important to tune the system to use as much disk cache as possible for Dynamic Converter to prevent repeated conversions of the same document.
- In addition to HTML, Dynamic Converter can also output WML and cHTML for wireless devices (see page 3-60).
- Dynamic Converter can convert XML to HTML, but it cannot output XML. If you need XML output, you should use a different conversion module (see Conversion to XML on page 3-56).

See also:

– *Content Publisher (page 3-55)*

Content Publisher

Oracle Content Publisher provides scheduled publishing of content to HTML using advanced template-based technology. It extracts content from the Content Server repository, intelligently translates the various elements into HTML (using abstraction), and publishes the resulting web pages to a web server or back into the content server.

Implementation Considerations

The following implementation considerations are important for content conversion to HTML through Content Publisher:

- Content Publisher runs on Windows 95, Windows 98, Windows NT, Windows 2000, and Windows XP.
- Content Publisher provides scheduled conversions—that is, at set times or regular intervals. This means that if content changes, there may be some lag time before the new version is available as HTML (depending on how the conversion schedule has been set up).
- Content Publisher enables users to view content without having its native application installed on their computers.
- Content Publisher outputs HTML renditions that can retain all formatting of the original documents, if required.

- Conversion to HTML using Content Publisher can be extremely granular. There is enormous flexibility when it comes to defining what content elements should be converted to what web elements. Detailed formatting characteristics can be used to define the conversion.
- The conversion to HTML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).
- Content Publisher is particularly useful for conversion of content that is accessed very regularly by large numbers of users or that does not change very much.
- Content Publisher enables the creation of entire websites, including hierarchy, navigation, and layout designs.
- Content Publisher is not sensitive to the number of users since all conversions are done on a separate machine and produce a static file that is available from the web server without any service to determine if the file needs to be converted.
- Content Publisher needs to maintain a translation area that replicates the repository that is on the server. It also uses a polling method to detect new content that is on the server. Polling for content becomes less practical as the number of documents increases.
- Content Publisher can assist web developers in creating websites that comply with requirements of Section 508 (a U.S. government initiative designed to help disabled individuals operate and consume electronic and information technologies).
- In addition to HTML, Content Publisher can also output XML (see page 3-58) as well as WML and cHTML for wireless devices (see page 3-62).
- Content Publisher can leverage multi-processor machines for the scheduled publishing process.

See also:

– *Dynamic Converter (page 3-54)*

CONVERSION TO XML

The eXtensible Markup Language (XML) is an open standard for identifying and describing data on web pages and in documents. It uses a similar tag structure as HTML; however, whereas HTML defines how elements are displayed, XML defines what those elements contain. XML is “extensible” because, unlike HTML, the markup symbols are

customizable and self-defining. XML tags are unlimited and descriptive, making them more meaningful, whereas HTML tags are a limited set and do not supply information about what the data is, only how it should appear. Thus, virtually any data items can be identified, allowing web pages to function like database records. This enables the definition, transmission, validation, and interpretation of data between applications and between organizations.

Document type definitions (DTDs) are used to define how the XML markup tags should be interpreted by the application presenting the document.

The Content Server suite of products includes two optional feature sets which can convert content to the XML format:

- XML Converter (page 3-58)
- Content Publisher (page 3-58)

Each of these optional feature sets has its own characteristics and considerations. The combination of both provides optimum conversion in all situations.

See also:

- *Managed Content (page 3-3)*
- *Users (page 3-5)*
- *Authentication Types (page 3-12)*
- *Hardware Setup (page 3-14)*
- *Infrastructure (page 3-27)*
- *Publication and Data Exchange (page 3-35)*
- *Security (page 3-41)*
- *Search Solution (page 3-50)*
- *Conversion to PDF (page 3-51)*
- *Conversion to HTML (page 3-53)*
- *Conversion to WML or cHTML (page 3-59)*
- *Other Conversions (page 3-63)*
- *Workflows (page 3-63)*
- *Metadata (page 3-66)*
- *Backup Strategy (page 3-72)*

XML Converter

Oracle XML Converter is an optional feature set to Oracle Inbound Refinery. It converts more than 225 file formats from their native format to the XML format. The XML output files can then be used by other applications or engines for further processing (for example, display in a web browser or inclusion in a database).

Implementation Considerations

The following implementation considerations are important for content conversion to XML through XML Converter:

- XML Converter uses Oracle's OutsideIn technology and either the FlexionDoc or SearchML schema.
- The XSLT (eXstensible Style Language Transformation) component can post-process the resulting XML document, transforming it to a different XML document based on your XSL (eXstensible File Language) file and validated against your DTD (Document Type Definition) or schema.
- Content items are converted at the moment they are checked into the content server. Both the native and converted files are stored, which means the XML renditions are immediately available as soon as they are needed.
- The conversion to XML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).

See also:

– *XML Converter (page 3-58)*

Content Publisher

In addition to converting content to HTML (see page 3-55), Content Publisher can also output XML.

Implementation Considerations

The following implementation considerations are important for content conversion to XML through Content Publisher:

- Content Publisher runs on Windows 95, Windows 98, Windows NT, Windows 2000, and Windows XP.
- Content Publisher provides scheduled conversions—that is, at set times or regular intervals. This means that if content changes, there may be some lag time before the new version is available as XML (depending on how the conversion schedule has been set up).
- Content Publisher produces output based on fully customizable document type definitions (DTDs). This means you have complete control over the XML output tags.
- The conversion to XML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).
- Content Publisher is not sensitive to the number of users since all conversions are done on a separate machine and produce a static file that is available from the web server without any service to determine if the file needs to be converted.
- Content Publisher needs to maintain a translation area that replicates the repository that is on the server. It also uses a polling method to detect new content that is on the server. Polling for content becomes less practical as the number of documents increases.
- In addition to XML, Content Publisher can also output HTML (see page 3-55) as well as WML and cHTML (see below).

See also:

– *Content Publisher (page 3-58)*

CONVERSION TO WML OR CHTML

WML (Wireless Markup Language) is a tag-based language used to specify content and user interface for devices that support the Wireless Application Protocol (WAP). WML is supported by most mobile phone browsers.

cHTML (Compact HTML) is a subset of HTML, which is used for small information devices such as cellular phones and personal digital assistants (PDAs). It is essentially a pared down version of HTML in order to accommodate for the hardware restrictions that these devices have (limited memories, small mono-color displays, restricted input methods, etc.)

The Content Server suite of products includes two optional feature sets which can convert content to the WML and cHTML format:

- Dynamic Converter (page 3-60)
- Content Publisher (page 3-62)

Each of these optional feature sets has its own characteristics and considerations. The combination of both provides optimum conversion in all situations.

See also:

- *Managed Content (page 3-3)*
- *Users (page 3-5)*
- *Authentication Types (page 3-12)*
- *Hardware Setup (page 3-14)*
- *Infrastructure (page 3-27)*
- *Publication and Data Exchange (page 3-35)*
- *Security (page 3-41)*
- *Search Solution (page 3-50)*
- *Conversion to PDF (page 3-51)*
- *Conversion to HTML (page 3-53)*
- *Conversion to XML (page 3-56)*
- *Other Conversions (page 3-63)*
- *Workflows (page 3-63)*
- *Metadata (page 3-66)*
- *Backup Strategy (page 3-72)*

Dynamic Converter

In addition to converting content to HTML (see page 3-54), Oracle Dynamic Converter can also output WML or cHTML for wireless devices.

Implementation Considerations

The following implementation considerations are important for content conversion to WML or cHTML through Dynamic Converter:

- Dynamic Converter provides on-demand conversion to WML or cHTML, i.e. only when content is specifically requested. This means that the WML or cHTML rendition always reflects the most recent version of the content.
- Dynamic Converter enables users to view content without having its native application installed on their computers.
- Dynamic Converter outputs WML or cHTML renditions that can retain all formatting of the original documents, if required.
- Dynamic Converter provides an easy, fast way to make content web-viewable, especially in situations where precise formatting is not the top priority.
- The conversion to WML or cHTML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).
- Dynamic Converter is particularly useful for conversion of content that changes a lot or that is not accessed very often (for example, large documents).
- Dynamic Converter is particularly good at handling large amounts of documents. This is because it does not use polling to detect new content, and it only converts content on demand (i.e., when a content item is requested for display).
- Dynamic Converter puts some load on the content server, because the conversion to WML or cHTML is done on the server itself (with Dynamic Converter running on it).
- Dynamic Converter relies on caching technology to reduce the load on the server and ensure that documents are not re-translated unnecessarily.
- Dynamic Converter is best able to effectively service the requests for conversion if the number of users is not too large. This is due to the fact that Dynamic Converter must run a service every time a user wants to view a document. If the document is not in the Dynamic Converter conversion cache, the conversion must be done again. The conversion process is very resource-intensive. If there are large numbers of users, it is important to tune the system to use as much disk cache as possible for Dynamic Converter to prevent repeated conversions of the same document.
- In addition to WML and cHTML, Dynamic Converter can also output HTML (see page 3-54).
- Dynamic Converter can convert XML to WML or cHTML, but it cannot output XML. If you need XML output, you should use a different conversion module (see Conversion to XML on page 3-56).

See also:

– *Dynamic Converter (page 3-60)*

Content Publisher

In addition to converting content to HTML and XML (see pages 3-55 and 3-58), Oracle Content Publisher can also output WML or cHTML for wireless devices.

Implementation Considerations

The following implementation considerations are important for content conversion to WML or cHTML through Content Publisher:

- Content Publisher runs on Windows 95, Windows 98, Windows NT, Windows 2000, and Windows XP.
- Content Publisher provides scheduled conversions—that is, at set times or regular intervals. This means that if content changes, there may be some lag time before the new version is available as WML or cHTML (depending on how the conversion schedule has been set up).
- The conversion to WML or cHTML works best if the source documents are set up and formatted in a uniform manner (for example, all headings share the same formatting characteristics, etc.).
- Content Publisher is not sensitive to the number of users since all conversions are done on a separate machine and produce a static file that is available from the web server without any service to determine if the file needs to be converted.
- Content Publisher needs to maintain a translation area that replicates the repository that is on the server. It also uses a polling method to detect new content that is on the server. Polling for content becomes less practical as the number of documents increases.
- In addition to WML and cHTML, Content Publisher can also output HTML (see page 3-55) and XML (see page 3-58).

See also:

– *Content Publisher (page 3-62)*

OTHER CONVERSIONS

In addition to “general” conversions to PDF, HTML, XML, and WML/cHTML, there are also other, more specialized conversion options. Each of these is handled by specific add-ons to Content Server:

- **Tiff Converter**
  This feature set enables conversion of TIFF (Tagged Image File Format) image files to PDF. Upon check-in of single or multi-page TIFF files, Tiff Converter will convert single TIFF images to a single PDF file, and multi-page TIFF images to a single, multi-page PDF file.

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Workflows (page 3-63)*
– *Metadata (page 3-66)*
– *Backup Strategy (page 3-72)*

WORKFLOWS

This section covers the following topics:

See also:

– *Managed Content (page 3-3)*

– *Users (page 3-5)*

– *Authentication Types (page 3-12)*

– *Hardware Setup (page 3-14)*

– *Infrastructure (page 3-27)*

– *Publication and Data Exchange (page 3-35)*

– *Security (page 3-41)*

– *Search Solution (page 3-50)*

– *Conversion to PDF (page 3-51)*

– *Conversion to HTML (page 3-53)*

– *Conversion to XML (page 3-56)*

– *Conversion to WML or cHTML (page 3-59)*

– *Other Conversions (page 3-63)*

– *Metadata (page 3-66)*

– *Backup Strategy (page 3-72)*

About Workflows

Workflows can be used in the process of reviewing and approving content before it is released and published. This is particularly important in Business-to-Customer (B2C) web content management scenarios, where published content will be available to anyone who accesses the website, usually on the Internet. Adequate reviewing and approval processes are then crucial to ensure that only final, authorized content is actually published on the website. Workflows are also very useful in document management implementations, where they provide a good tool to ensure that all appropriate people review documents relevant to them.

See also:

– *Type of Workflows (page 3-65)*

– *Workflow Process (page 3-65)*

– *Workflow Considerations (page 3-66)*

Type of Workflows

There are three types of workflow:

- **Basic workflows**
  Basic workflows are workflows in which content is specifically assigned to the workflow. This type of workflow requires someone to initiate the process. They are particularly useful when setting up review processes for a group of related content items.
- **Criteria workflows**
  Criteria workflows are workflows in which any content matching predefined metadata criteria enters the workflow automatically upon check-in. This type of workflow is particularly useful when setting up standardized review processes for individual documents.
- **Sub-workflows**
  A sub-workflow is a workflow that does not have an initial contribution step. Sub-workflows are useful for splitting large, complex workflows into manageable, reusable pieces.

See also:

– *About Workflows (page 3-64)*
– *Workflow Process (page 3-65)*
– *Workflow Considerations (page 3-66)*

Workflow Process

After a content item enters a workflow (see above), it goes through all the steps defined for that workflow. When a file is approved by the defined minimum number of reviewers for a particular step, it goes to the next step in the workflow. If any reviewer rejects a file, it goes back to the most recent contribution step for editing. When a file is approved at the last step in the workflow, it exits the workflow and the content item is released to the system. If a workflow contains multiple files, none of them are released to the system until all of the files have completed the workflow.

See also:

– *About Workflows (page 3-64)*

– *Type of Workflows (page 3-65)*
– *Workflow Considerations (page 3-66)*

Workflow Considerations

If you decide to set up workflows, you need to take the following implementation considerations into account:

- Users do not need Workflow rights for their files to enter a criteria workflow.
- Files can enter a basic workflow only if a user with Workflow rights starts the workflow.
- If a file is checked in with the wrong security group or wrong metadata value, it can enter a criteria workflow accidentally.
- If users are frequently processing files through the same basic workflow, consider setting up a criteria workflow to automate the process.
- Users cannot use the same criteria for multiple workflows.

See also:

– *About Workflows (page 3-64)*
– *Type of Workflows (page 3-65)*
– *Workflow Process (page 3-65)*

METADATA

The content managed by Content Server is categorized using metadata, which is basically information about the content. There are a number of predefined metadata fields (such as author, title, and check-in date), but it is also possible to add user-defined fields. What customized metadata fields are useful to add depends on the implementation and security requirements.

The following topics are important with regard to metadata:

- Predefined Metadata Fields (page 3-67)
- Custom Metadata Fields (page 3-69)
- Metadata Models (page 3-69)
- Metadata Considerations (page 3-71)

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*
– *Search Solution (page 3-50)*
– *Conversion to PDF (page 3-51)*
– *Conversion to HTML (page 3-53)*
– *Conversion to XML (page 3-56)*
– *Conversion to WML or cHTML (page 3-59)*
– *Other Conversions (page 3-63)*
– *Workflows (page 3-63)*
– *Backup Strategy (page 3-72)*

Predefined Metadata Fields

Content Server comes with a number of predefined metadata fields. They cannot be deleted and their definition cannot be edited. (Fields below marked with an asterisk are required for a content item to be checked in successfully.)

- **Content type***
  This field allows for broad categorization of content items. When checking in content, contributors assign the content to a certain type, for example “Financial Statement” or “Sales Report.” This enables clustering of similar content into functional groups. The content types are fully user-definable, which means they can be geared to any business situation. Content types are defined and managed using Configuration Manager, which is one of the tools accessed from the Administration page of Content Server.

- **Title***
  This field provides is a descriptive name identifying a content item. It is typically what is displayed on the search results screen.

- **Author***
  This field identifies the person who created or revised the content item and checked it into the content server.
- **Security group***
  This field sets a container around the content item, so that searching and access of the content item can be restricted. When checking in content, contributors assign the content to a security group. These are sets of content grouped under a unique name and are used to restrict access to the content. Security groups are defined and managed using User Admin (Security—Permissions By Group), which is one of the tools accessed from the Administration page.
- **Account**
  This is an optional, predefined field that, once turned on, sets more granular security.
- **Primary file***
  This field refers to the original file being checked in.
- **Alternate file**
  This field refers to a secondary file to be checked in along with the primary file. It is especially useful when the primary file is not web-viewable, such as a zip file or an executable.
- **Content ID***
  This field provides a unique identification for each content item checked into the content server. The system may be set to generate the content ID automatically.
- **Revision***
  This field identifies the current revision of the content item. It identifies the number of updates (life cycles / revisions) a content item has gone through. The revision sequence can be customized to meet the unique revisioning requirements of each application (for example, 1-2-3 or a-b-c).
- **Comments**
  This is an optional field that for writing “memo”-type comments concerning the content item being checked in.
- **Release date***
  This is usually the date that the document is checked in. Alternatively, it can allow for the timed release of a content item.
- **Expiration date**
  This is the date that a content item should automatically expire from the system (i.e., no longer be available for viewing).

See also:

– *Custom Metadata Fields (page 3-69)*

– *Metadata Models (page 3-69)*

– *Metadata Considerations (page 3-71)*

Custom Metadata Fields

In addition to the predefined metadata fields, custom metadata field may also be created and defined to suit particular content management needs. Some examples of commonly used custom metadata fields are:

- Product
- Part Number
- Reference Number
- Department
- Category
- Subcategory
- Customer
- Project
- Location

See also:

– *Predefined Metadata Fields (page 3-67)*

– *Metadata Models (page 3-69)*

– *Metadata Considerations (page 3-71)*

Metadata Models

There are various common metadata strategies:

❖ Departmental Model (page 3-70)

❖ Geographical Model (page 3-70)

❖ Content Type Model (page 3-70)

❖ Website Navigation Model (page 3-70)

Note: The metadata models listed above are not mutually exclusive. In practice, there will often be a combination of metadata models, for example a departmental metadata model which includes content type metadata fields.

Departmental Model

In this metadata model, content managed by Content Server is primarily categorized based on functional units within an enterprise. Content items are organized by the division, department, or workgroup they originate from—for example, Sales, Marketing, R&D, Human Resources, Service, Training, and Support.
This metadata model is particularly useful in situations where departments want to control the security of their documents at the department level.

Geographical Model

In this metadata model, content managed by Content Server is primarily categorized based on geographical units within an enterprise. Content items are organized by the location they originate from—for example, North America, South America, Europe, Middle East, Asia Pacific, and Australia. This metadata model is particularly useful in situations where the infrastructure is distributed and "enterprise-searchable" (i.e. search queries include various content servers).

Content Type Model

In this metadata model, content managed by Content Server is primarily categorized based on the type of content—for example, Reports, Quotes, Invoices, Articles, Forms, Graphics, Memos, and Policies. This metadata model is particularly useful in situations where a generic model can be established that all business units can agree on. More generic type definitions provide for reusable metadata structure between business units.

Website Navigation Model

In this metadata model, content managed by Content Server is primarily categorized by their navigational position in a website—for example, Top Menu, Level1, Level2, Header, etc.

This metadata model is particularly useful in web content management scenarios, where it simplifies populating the website with content and site elements. The metadata information then helps to determine where each content item goes on the website.
A number of methods can be used to help contributors check in files with the correct metadata, including dependent choice lists, accounts, and web folders.

See also:

– *Predefined Metadata Fields (page 3-67)*
– *Custom Metadata Fields (page 3-69)*

– *Metadata Considerations (page 3-71)*

Metadata Considerations

The following are some “best practices” when it comes to defining and setting up metadata:

- Try to define only as much metadata as needed. Metadata fields that are not used serve no purpose whatsoever, and should be avoided. It is recommended that you only define metadata fields that users will actually search on, or that are used otherwise.
- Metadata should describe the content item (“What is it *about*?”). It should generally *not* describe what will happen to the content item (“Where will it go?”, “Who should see it?”).
- Avoid redundancy. If a value appears in one field, it is not needed in another field.
- Avoid metadata value combinations in a single field. For example, you do not want to define a field called “Category” with the options “Department A Form,” “Department B Form,” “Department A Policy,” “Department B Policy,” etc. Instead, define two fields: one called “Department” with the value “A” and “B”, and one called “Type” with the values “Form” and “Policy.”
- Avoid single-use values. Make sure optional metadata values apply to more than a single content item. The content ID is the only unique identifier that Content Server requires.
- Remember that metadata models should describe a unique subset of content through the intersection of metadata values. Where large corporate taxonomies are available, the first step in merging such a taxonomy into a metadata model should be data entity relationship modeling.
- Use different field labels on the same metadata field where differences in presentation (user interface, look and feel) may use the same field types. For example: Presentation 1 may have a field labeled “Comments” while presentation 2 may use the same field labeled “Summary” and presentation 3 may use the field but label it “Keywords.” In all cases the xComments field is used, but the label and the purpose changes for each presentation. This cuts down on the need for multiple metadata fields that contain the same kind of metadata while preserving the different ways and contexts in which that metadata can be used.
- Avoid irrelevancy. Do not display metadata fields and/or values that are not relevant or not viable candidates for the content item being contributed. Remember that if all

content items in a set have the same metadata values, then that metadata does not help distinguish content items within that set.

- There may be fields that are dependent on each other such that certain values in option lists are displayed or hidden based on what values have been selected in an earlier option list. Such dependencies should be identified. These are implemented in Content Server using Dependent Choice List components.
- Make sure there are as few value options as possible for any metadata field while maintaining all requisite categorization or labeling schemas.
- Check whether option lists can be broken down into smaller components or broken out into other metadata fields.
- When defining a metadata strategy, try to take known or planned future requirements or developments into account.

See also:

– *Predefined Metadata Fields (page 3-67)*
– *Custom Metadata Fields (page 3-69)*
– *Metadata Models (page 3-69)*

BACKUP STRATEGY

This section covers the following topics:

See also:

– *Managed Content (page 3-3)*
– *Users (page 3-5)*
– *Authentication Types (page 3-12)*
– *Hardware Setup (page 3-14)*
– *Infrastructure (page 3-27)*
– *Publication and Data Exchange (page 3-35)*
– *Security (page 3-41)*

Using Backups

It is important that you back up your files on a regular basis to make disaster recovery as efficient as possible. Use standard backup principles when setting up a back-up strategy, taking the following considerations into account:

- There are basically four storage areas in Content Server:
  - the native file repository (“Vault”)
  - the web-viewable file repository (“Web Layout”)
  - the content information database
  - the search index

  Of these four storage areas, the two file repositories and the content information database are the most important—a Oracle application can never be successfully recovered without them. Without the actual files, there is no content. Without the content information database, Content Server knows nothing about the content, which effectively means it does not exist.
- The backed-up versions of the file repositories and the content information database should be synchronized as closely as possible. Ideally, the content server and database should be shut down before making a backup. This achieves maximum synchronization, since no content can be checked in or out while the backup is being made. If this is not an option, make sure that the backup is made during periods of low system activity (for example, at nights or on weekends).
- It is recommended that you include the entire Content Server directory structure in your backup strategy. This ensures that you include all configuration files and other important application-specific files (for example, custom components).

See also:

– *Backup/Recovery Methods (page 3-74)*

– *Disaster Recovery (page 3-75)*

Backup/Recovery Methods

Content Server is created to utilize both an RDBMS (Oracle, SQL Server, etc.) and an operating system's file structure. With this in mind, the options vary with each customer's implementation model. When encountering a disaster recovery, the first question that should be asked is what is an acceptable risk that an organization can handle in the event a data loss occurs and what will be your recovery protocol. Here are two options to consider:

- Nightly backups
- Incremental backups

With this understanding, this guide provides backup/recovery methods that can be used as guidelines for customer strategies.

Incremental Archive

Content Server's Archiver utility can perform incremental updates. These updates are triggered anytime an object is checked in or modified in the content server. You can keep an archive on a separate server in another location. Recovery is performed by importing the archive into either the same server or another server. By importing the archive, Content Server returns to the state of the last object modification. This is a method suggested for organizations that have an active contributor base.

Manual Archive

Content Server's Archiver utility can perform "on-demand updates." Recovery is performed by importing the archive into either the same server or another instance. This method is suggested for organization that have a small contribution community. Optionally, a component can be built to allow a third-party scheduling application to initiated the Archiver server on an automated bases.

Replication

You can set Content Server's Archiver utility to perform to replicate the content in the content server. Replication can be implemented using either the Incremental or Manual

Archive functionality. These replications can be used to create a mirrored instance of a content server. If the incremental option is used, an archive can be automatically transferred and imported to the mirrored server on any object modification on the main content server. This method is suggested for organizations that want to implement the content server in an automatic fail-over environment.

Full Database/File System Restore

Oracle's open architecture leverages industry-standard database and file systems. Therefore, Content Server can take advantage of the backup/restore capabilities of each. Most databases have the capability to do full back-up and incremental logging. Combined with a built-in or third-party file backup application, this solution can provide an organization the option to completely restore the state of a database and file system. This option should be used in conjunction with one of the archive or replication solutions as a secondary disaster recovery plan.

Tech Tip: It is important to back up your files on a regular basis to make disaster recovery as efficient as possible. We recommend that you back up the following:

– All changes to the content server's directory structure on a daily basis
– All files in the content server's directory structure on a monthly basis
– The software after you install it and before you update it with a new version.

See also:

– *Using Backups (page 3-73)*
– *Disaster Recovery (page 3-75)*

Disaster Recovery

Disasters are unusual; however, in the event of a total machine failure, it is best to have strong disaster recovery procedures in place. Following a total machine failure, perform the following tasks:

1. Reinstall all required third-party software.
2. Configure the platform to be the same as it was when the last backup was made.
3. Reinstall the database software and the Content Server software.
4. Copy the backed up Content Server software directory structure over the fresh installation.

5. If the native file repository (“Vault”) and web-viewable file repository (“Web Layout”) are not on the same computer as the content server, copy them to their respective original locations.
6. Copy the backed-up database over the empty database that was created when the software was reinstalled.
7. Restart the software.
8. Run the Content Server Analyzer to confirm the integrity of the content server repository components and clear up any discrepancies between the file system and the database.

See also:

– *Using Backups (page 3-73)*
– *Backup/Recovery Methods (page 3-74)*

Chapter

4

INTRANET SITES

OVERVIEW

Intranets are private networks that use common web technology within an enterprise or organization. They can be used for anything from small workgroup collaborations on a single location to large enterprise-wide information networks which serve as dynamic communication vehicles providing up-to-the-minute (business) information to employees worldwide.

This section addresses two main intranet scenarios:

- Basic scenario: workgroup intranet
- Extended scenario: enterprise-wide Business-to-Employee (B2E) site

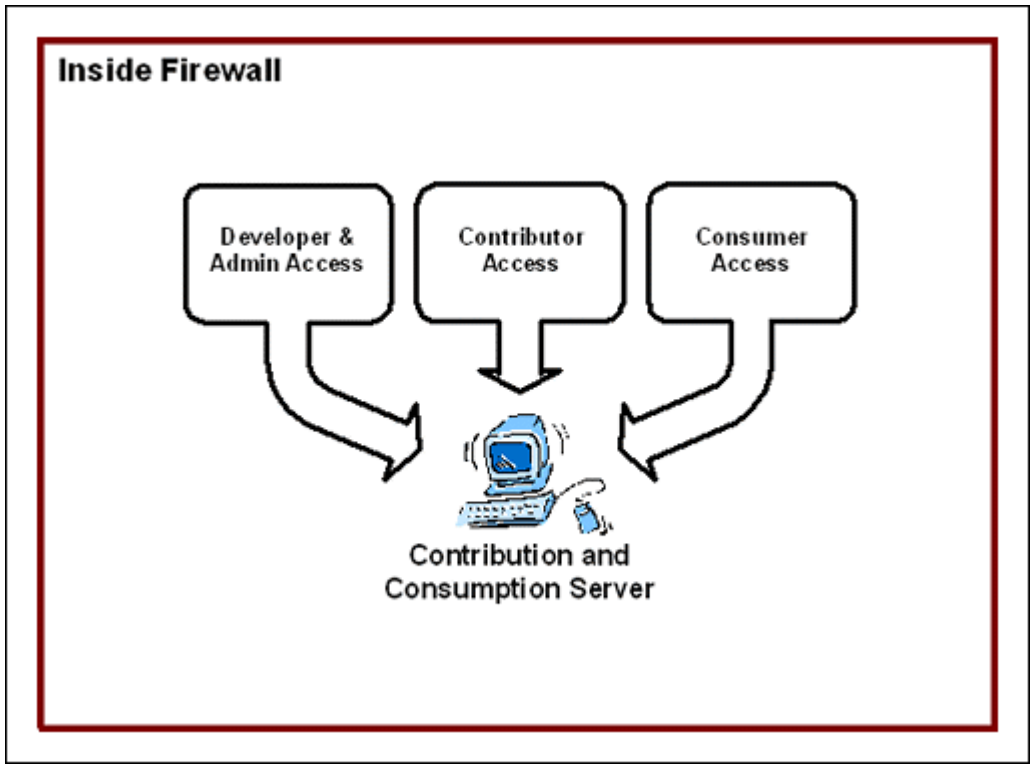
The following sections contain implementation guidelines for each of these scenarios, along with other considerations.

Important: The guidelines provided in this section are by no means hard and fast rules. Every application is different and may require its own, very specific considerations.

WORKGROUP INTRANET SITE

Figure 4-1 The figure below shows an organization's intranet in its most basic form: a single content server inside the firewall-protected IT environment. The content server acts as the central point of access for everybody involved in the content management process: developers, administrators, contributors, and consumers.

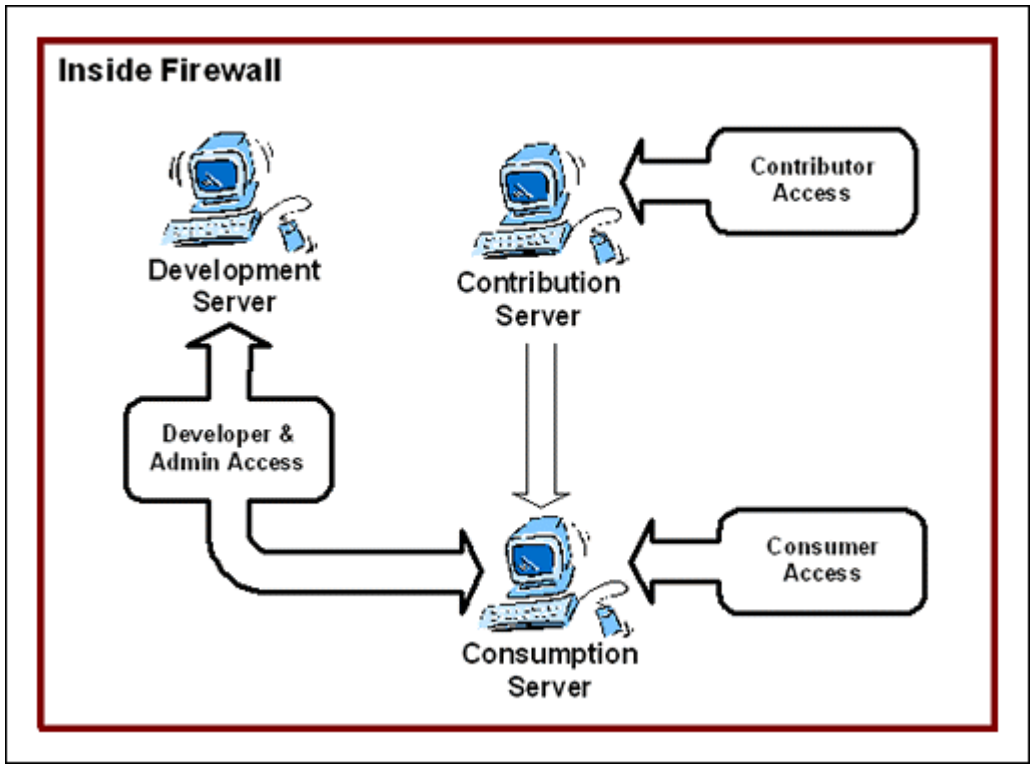
Figure 4-1 Basic intranet scenario: workgroup site (basic)



In a typical basic intranet scenario, Content Server is mostly used for document management purposes, as a revision/versioning tool. The website is used for presentation purposes only. Also, there is no interfacing with other systems.

The basic architecture shown in Figure 4-1 the figure above may become a bit more complicated if there is "functional distribution" (see page 3-16) and/or redundancy (see page 3-18), as shown in Figure 4-2the figure below. In such scenarios, several servers are used to perform the content management tasks, but the application still serves a single organizational unit (which can be a workgroup, department, division, or entire organization). Each of the (redundant) servers provides access for a specific group of users, depending on their role in the content management process.

Figure 4-2 Basic intranet scenario: workgroup site (with functional distribution)



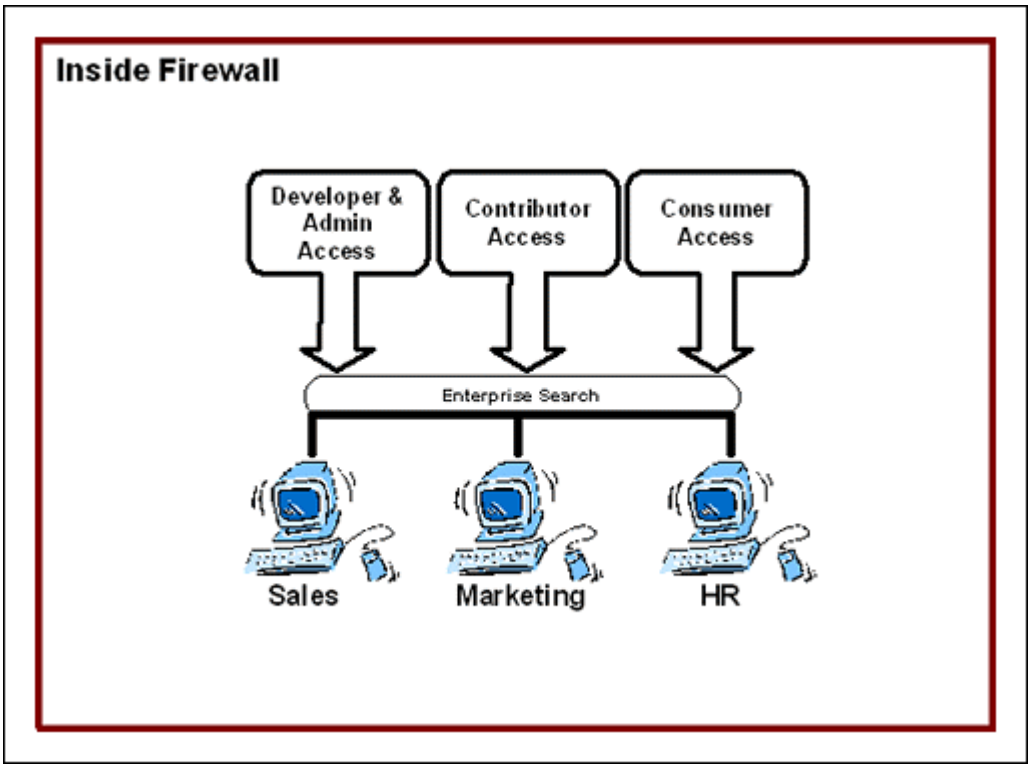
See also:

– *Enterprise-Wide B2E Intranet Site (page 4-3)*

ENTERPRISE-WIDE B2E INTRANET SITE

Many companies use separate content servers to serve specific organizational units (for example, departments). Figure 4-3 The figure below shows the most basic form of this setup: separate departmental content servers within one firewall-protected IT environment. Each of the servers is the central point of access for everybody involved in the content management process in their respective department.

Figure 4-3 Extended intranet scenario: enterprise-wide B2E site (basic)



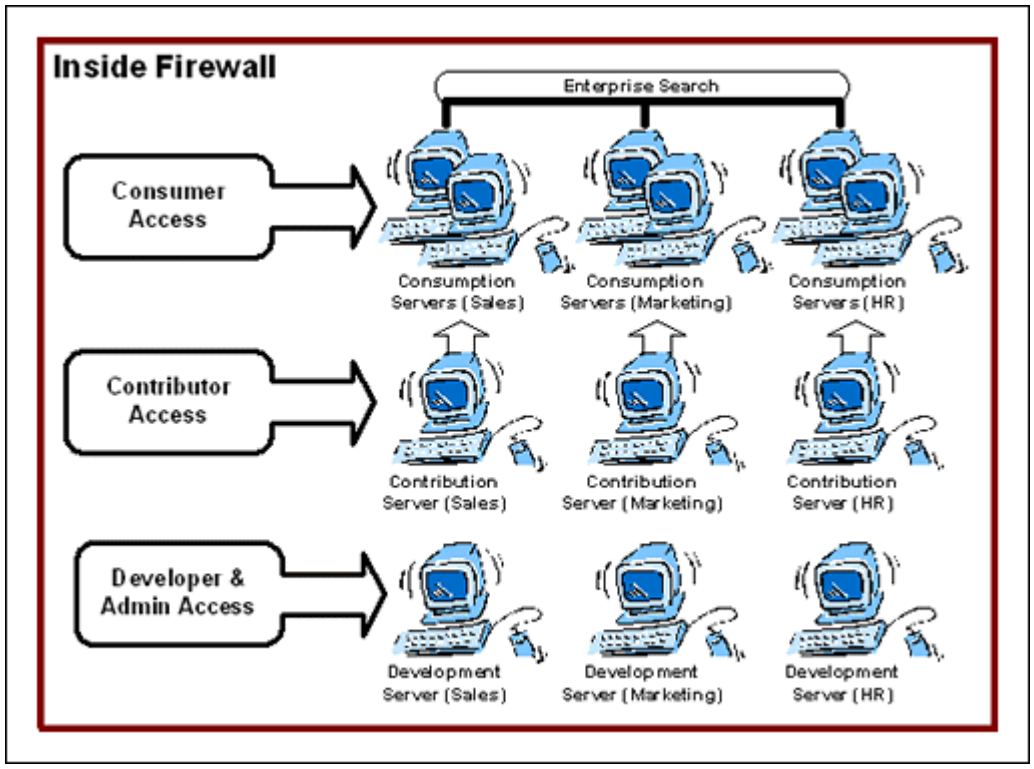
There are various reasons for companies to use a setup like this:

- It allows departments to control the security of their documents at the department level.
- It allows individual departments to use their own metadata models. This prevents other departments from having to deal with metadata which is totally irrelevant to them.
- If the departments or divisions are spread out geographically, this setup can be used to provide each department or division with its own server. This can help eliminate long-distance, slow network links. As such, it serves as a form of “load balancing.”

As in the basic intranet scenario outlined in the previous section, the basic architecture shown in Figure 4-3 the figure above may become a bit more complicated if there is “functional distribution” (see page 3-16) and/or redundancy (see page 3-18), as shown in Figure 4-4the figure below.

Each of the (redundant) servers provides access for a specific group of users, depending on their location and role in the content management process.

Figure 4-4 Extended intranet scenario: enterprise-wide B2E site (with redundancy)



See also:

– *Workgroup Intranet Site (page 4-2)*

MANAGED CONTENT

Intranets typically serve corporate/organization environments. They are primarily used to make business information accessible to all company employees. This means that content managed in an intranet application will mostly be business-related content from multiple business sources. Most of the content is unstructured. The content updates are usually frequent and are done in real time.

Intranet applications are used to make business information accessible to all company employees, often on many different locations around the globe. This means intranet sites often have an international orientation, which means content may be in several different languages. As a result, the Content Server system may need to be configured to handle multilingual content. For example, the search index needs to handle special characters (for example, ä, ß, ê, ñ, ó) in documents correctly.

Workflows may be used to ensure content is reviewed and approved by the appropriate people before being released on the intranet site.

For more general implementation considerations on the subject of content refer to Business, Web, and Structured Content (page 2-8) and Managed Content (page 3-3).

See also:

– *Users (page 4-6)*
– *Hardware Setup (page 4-7)*
– *Infrastructure (page 4-9)*
– *Publication and Data Exchange (page 4-9)*
– *Security (page 4-10)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

USERS

An intranet application generally has a defined number of users, which are all internal to the organization. Even though the user base is a defined group, the number of users as such may be quite high. This is especially true if the intranet site serves a globally operating company or organization, with locations around the world. Whatever the number of users, all users are known to the system. They typically need to identify themselves using a username and password to gain access to the intranet site (often through a regular network login). Only after the system has verified the user credentials will a user have access to the site. The general public (i.e., anyone outside the organization) has no access to the site.

All employees are typically content consumers. This means that everybody in the organization can access and view content within the bounds of their assigned privileges. In addition, they may also be content contributors, which means they can add content to the intranet. Depending on the size and complexity of the intranet application, not all employees may be contributors. This role may, for example, be limited to a number of key employees in each department or location.

Since many users may be both content contributors and consumers, the number of direct contributors is relatively high. Users constantly check content in and out. The number of developers per site, on the other hand, is typically low.

Large intranet sites may serve an organization's global workforce. This means users may have different nationalities and speak different languages. Content Server may then need to be set up to handle content in different languages correctly and provide a multilingual content management environment to users.

For more general implementation considerations on the subject of users refer to Users (page 3-5).

See also:

– *Managed Content (page 4-5)*
– *Hardware Setup (page 4-7)*
– *Infrastructure (page 4-9)*
– *Publication and Data Exchange (page 4-9)*
– *Security (page 4-10)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

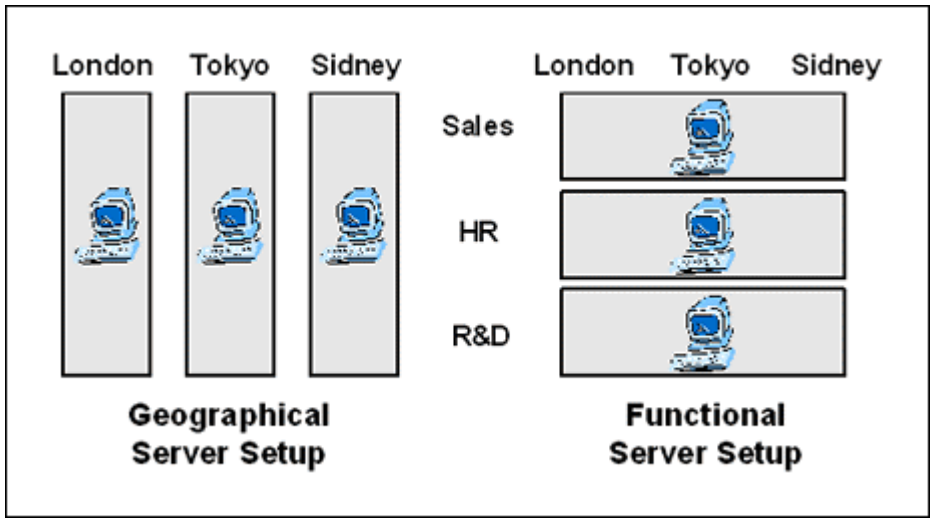
HARDWARE SETUP

In its most basic form, a small workgroup-oriented intranet only has one content server, which handles all content management tasks. In enterprise-wide B2E intranets, the number of required content servers depends on the number of organizational units served (for example, departments or locations). Depending on the size and complexity of the application, the number of servers may increase if multiple servers are used to perform different tasks in the content management process or to improve system availability. If there is a lot of conversion, a separate server may be dedicated to that task in order to minimize the system load on the "main" content server(s).

If the intranet needs to handle both European- and Asian-language documents, separate servers need to be set up, since a single server cannot control both at the same time.

In multiple-server intranet environments, the servers may be set up vertically (serving geographical units) or horizontally (serving functional units), as shown in Figure 4-5the figure below.

Figure 4-5 Horizontal vs. vertical server setup



Several factors determine which setup is best for a particular implementation, including the level of autonomy of the locations and departments, the company's organizational structure, security considerations, and the availability and speed of network links.

If Enterprise Search is to be used in a multiple-server intranet environment, all servers need to be able to communicate with each other. Only then can users search for content throughout the organization.

For more general implementation considerations on the subject of hardware setup refer to Hardware Setup (page 3-14).

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*
– *Infrastructure (page 4-9)*
– *Publication and Data Exchange (page 4-9)*
– *Security (page 4-10)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

INFRASTRUCTURE

In intranet scenarios, all computers that are part of the intranet are located inside the firewall-protected IT environment. This ensures optimum protection of the network, and prevents non-authorized persons from accessing data. Intranets are typically an integral part of the overall infrastructure of an organization. There may be dozens of separate, self-contained intranets within an organization, each serving its own segment.

For more general implementation considerations on the subject of infrastructure refer to Infrastructure (page 3-27).

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*
– *Hardware Setup (page 4-7)*
– *Publication and Data Exchange (page 4-9)*
– *Security (page 4-10)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

PUBLICATION AND DATA EXCHANGE

If separate contribution and consumption servers are used, content is typically moved between them using the transfer functionality of the Archiver application. This process can be automated. Data is then automatically transferred periodically or whenever certain metadata criteria are met. Since the data exchange takes place within the main firewall on a secure network, content can normally be transferred between the various servers using direct file share access (mounted drives). If data needs to be transferred internationally or between separate firewall-protected environments, this may pose security risks, which need to be addressed (see Infrastructure (page 3-27) and further for more details).

If Enterprise Search is used, the search queries are all handled by the web server, which uses standard Content Server sockets to communicate with all content servers configured to be “enterprise-searchable.”

For more general implementation considerations on the subject of publication and data exchange refer to Publication and Data Exchange (page 3-35).

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*
– *Hardware Setup (page 4-7)*
– *Infrastructure (page 4-9)*
– *Security (page 4-10)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

SECURITY

Not all information in an intranet application may be intended for everyone in the organization. Some of the content may not be public, which means its access is restricted. To see non-public content, intranet users need to make themselves known to the system by means of a username and password. The system determines what content users can access based on their user credentials. Authentication typically takes place through LDAP, NT Security, or NDS.

Depending on the number and location of the organizational units, user accounts may be needed to implement security efficiently.

For more information on Oracle-related security refer to the *Security Integration Guide*.

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*
– *Hardware Setup (page 4-7)*
– *Infrastructure (page 4-9)*
– *Publication and Data Exchange (page 4-9)*
– *Conversion Options (page 4-11)*
– *Metadata Model (page 4-11)*

CONVERSION OPTIONS

An intranet may serve a large number of users in a multitude of business and IT environments, possibly in different countries. This means the content managed by Content Server is typically in many different file formats and possibly various different languages. To ensure consumers can view all content, it is often converted to PDF or HTML before being presented to them. This enables them to view the content, even if they do not have the native applications on their computers (i.e., the applications the content item were originally created in). Also, to ensure consumers can search for all documents regardless of their language, the Content Server environment needs to be set up correctly for multilingual use.

For more general implementation considerations on Oracle's conversion options refer to Chapter 3 (*Basic Considerations)*.

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*
– *Hardware Setup (page 4-7)*
– *Infrastructure (page 4-9)*
– *Publication and Data Exchange (page 4-9)*
– *Security (page 4-10)*
– *Metadata Model (page 4-11)*

METADATA MODEL

Any metadata model can basically be used in intranet scenarios. Which one works best all depends on very application-specific considerations, such as the organizational structure of the environment in which Content Server runs, the server setup, and security considerations.

For more general implementation considerations on the subject of metadata models refer to Metadata (page 3-66).

See also:

– *Managed Content (page 4-5)*
– *Users (page 4-6)*

- *Hardware Setup* *(page 4-7)*
- *Infrastructure* *(page 4-9)*
- *Publication and Data Exchange* *(page 4-9)*
- *Security* *(page 4-10)*
- *Conversion Options* *(page 4-11)*

Chapter

5

EXTRANET SITES

OVERVIEW

Extranets are basically intranets that also offer limited, controlled access to a select group of users outside the organization. They serve “extended enterprises”—i.e., the enterprise itself as well as defined sets of customers, suppliers, or other partners.

This section addresses two main extranet scenarios:

- Basic scenario: partner site
- Extended scenario: support site

The following sections contain implementation guidelines for each of these scenarios, along with other considerations.

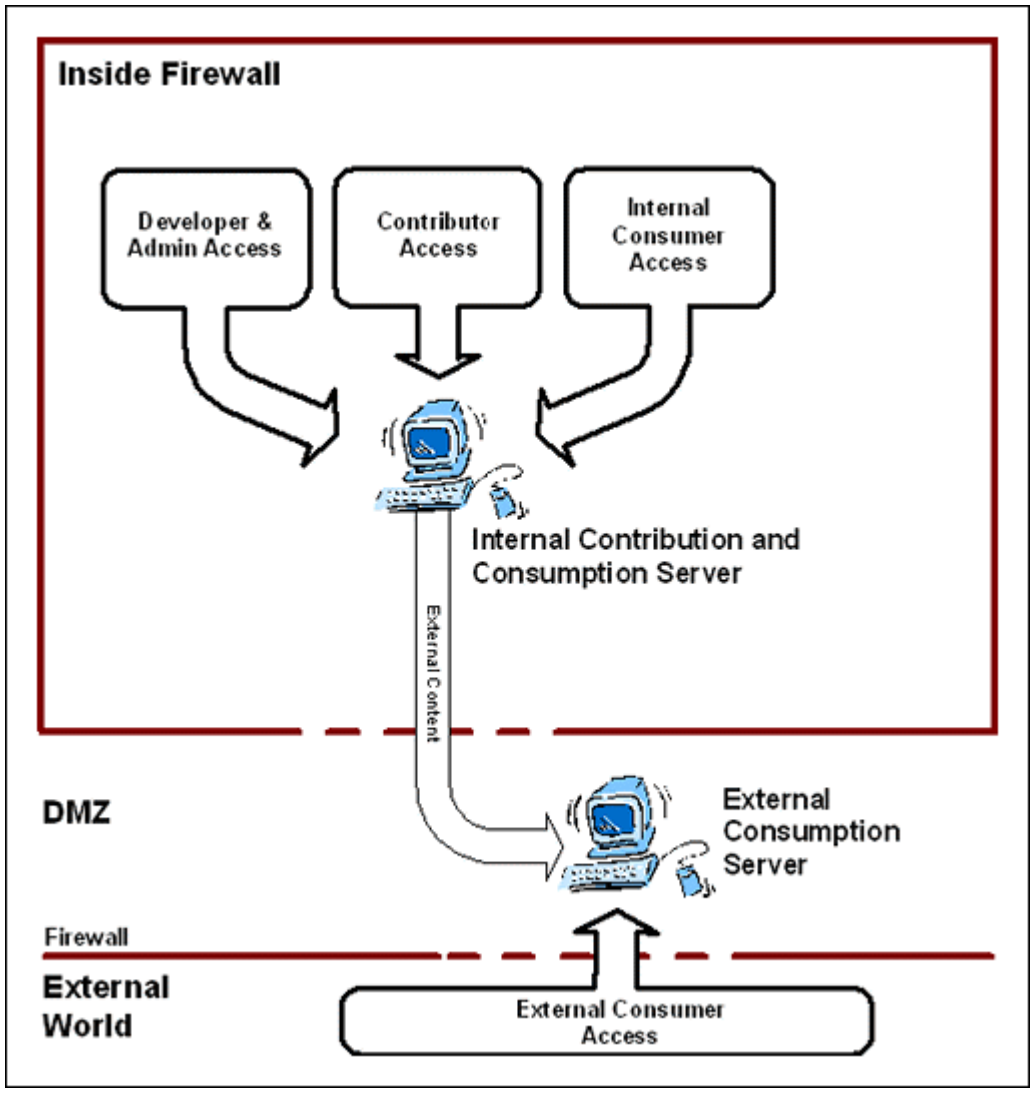
Important: The guidelines provided in this section are by no means hard and fast rules. Every application is different and may require its own, very specific considerations.

Note: The main difference between intranet applications on the one hand and extranet/Internet applications on the other is the fact that with the latter the site needs to be accessible by (some) people outside the organization. This means that extra security considerations play a role.

PARTNER EXTRANET SITE

Organizations may want to share part of their internal business information with a select group of partners outside the organization. This means these partners need to view content managed by Content Server. This is typically done by setting up one or more consumption servers outside the main firewall, which the external consumers (partners) can access. Figure 5-1 The figure below shows what a basic partner site application could look like.

Figure 5-1 Basic extranet scenario: partner site



As only persons within the organization contribute content to the Content Server system, the contribution server is located within the firewall-protected environment. No external partner needs to access this server. The server is also used for accessing content intended for internal use only. All content for external use is transferred to a consumption server in the DMZ. This server can be accessed by a defined group of users outside the organization who have been granted the appropriate rights.

The partner content can be a subset of the internal content, but can also be entirely separate from it. It may also be that all content is shared by both internal users and external partners, which means they all have access to the exact same information. Everyone can then use the same consumption server in the DMZ, and no separation of internal and external consumption servers is required.

Note: Partner extranet sites as described in this section are not intrinsically as simple as shown in Figure 5-1the figure below. Depending on the application, they may be a lot more complex and may require a setup more resembling that of the support site outlined in Support Extranet Site (page 5-3).

See also:

– *Support Extranet Site (page 5-3)*

SUPPORT EXTRANET SITE

Organizations may want to make service and support information available to a select group of their customers. This means these customers need to view content managed by Content Server. This is typically done by setting up one or more consumption servers outside the main firewall, which the external consumers (customers visiting the support site) can access. Figure 5-2 The figure below shows what a support site application could look like. Its setup is basically identical to that of the partner site described in the previous section. The main difference with the partner site is not so much related to the basic infrastructure, but more to accessibility and availability.

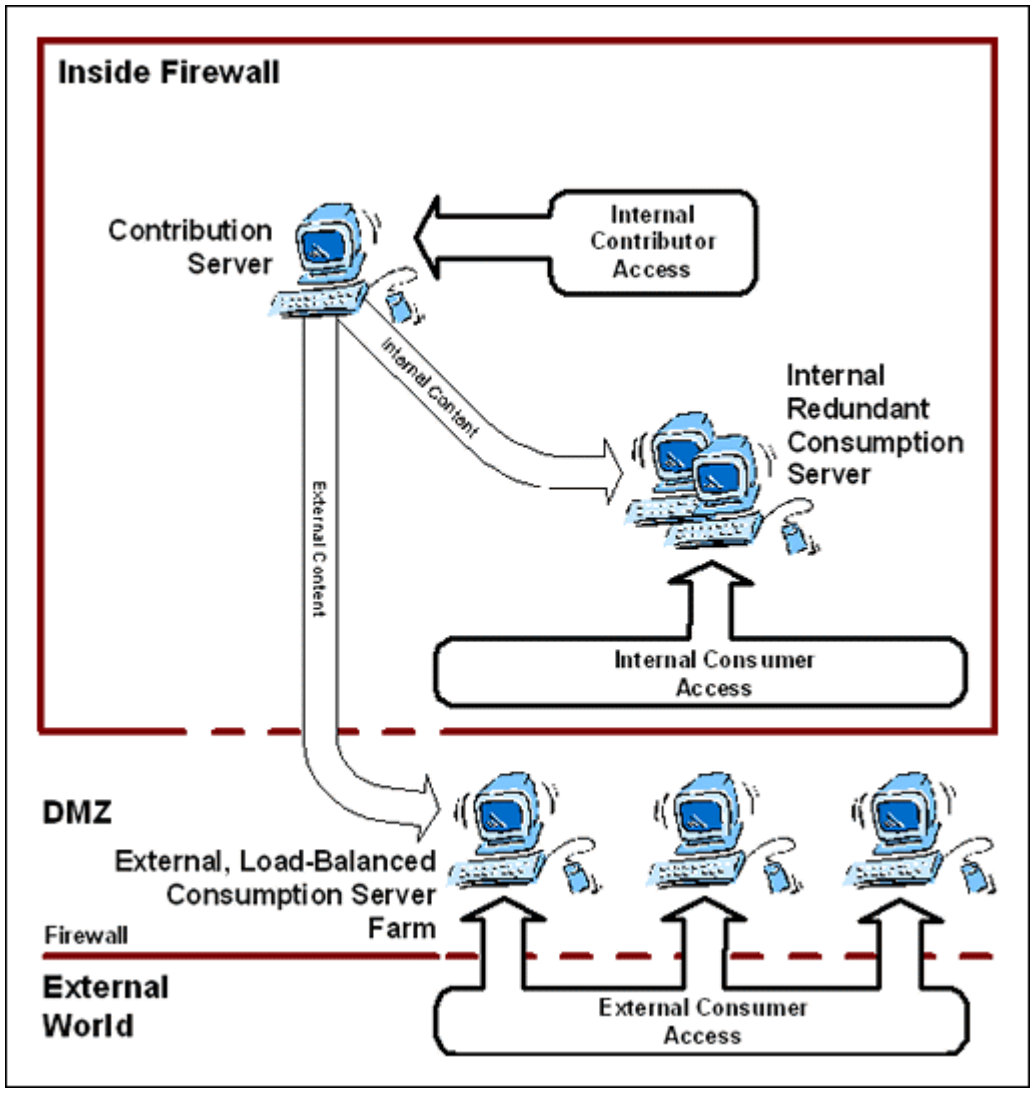
Contrary to the partner site, potential visitors to the support site may not be exactly known beforehand. A self-registration process may be set up to verify if the user is eligible to access the extranet site. For example, the prospective visitor to the site might be asked for a valid product serial number, which is then validated and linked to the visitor's name. The next time that person visits the extranet website, this name-number pair is used to authenticate the user.

It may be important to add functional distribution and redundancy to the support site for a number of reasons:

- High availability of the extranet website may be crucial, especially if support is a service that customers pay for.
- The user base may be substantially larger than that of a partner site, which may have a more limited target audience. This means that traffic may be much higher, and redundant, load-balanced consumption servers may be in order.

Note: Support extranet sites as described in this section are not necessarily as complex as shown in Figure 5-2the figure below. Depending on the application, they may be a lot simpler and their setup might look something like that of the partner site outlined in Partner Extranet Site (page 5-2).

Figure 5-2 Extended extranet scenario: support site



See also:

– *Partner Extranet Site (page 5-2)*

MANAGED CONTENT

Extranets serve “extended organizations,” which include defined sets of customers, suppliers, or other external partners. The majority of content in an extranet scenario is typically unstructured business content from multiple internal sources. The nature of the content may vary widely between applications and largely depends on the target audience.

In the case of a partner site, the content is most likely business and operational information such as marketing reports, sales forecasts, and presentations. Support sites, on the other hand, may contain entirely different types of content, including technical publications, data sheets, Q&A documents, and the like. Other applications may yet have their own specific content that is made available to a select group of people outside the organization.

Since the content is usually published to a full-fledged website, some of the content may be web content—related to creating the site rather than populating it (for example, Flash objects, images, server pages, or HCSP/HCST pages).

The internal content is usually updated frequently and in real time. Updates of the external content may take place on a scheduled basis, for example once a day.

The content will be viewed by people outside the organization. As a result, it may be useful to implement workflows. This helps to ensure content is reviewed and approved by the appropriate people before being released on the website for the external users to see.

For more general implementation considerations on the subject of content refer to Business, Web, and Structured Content (page 2-8) and Managed Content (page 3-3).

See also:

– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*

USERS

Contrary to intranets, extranets have both internal and external users. The internal users can both be consumers and contributors, whereas the external users are typically only consumers. Since at least part of the user base is external, security is very important. It is especially important that all external users are authenticated before being allowed access to non-public content. This is typically done through usernames and passwords. This enables Content Server to authenticate each user against a database of known, authorized users, and either allow or block access to the website. Once accepted, users have access to content based on their assigned privileges.

The number of internal and external users of extranet sites may vary considerably, depending on the exact application. An organization's intranet may be accessible to a very limited group of key partners, but may also be opened up to large groups of customers.

If the extranet site is targeted at an international audience, users may have different nationalities and speak different languages. Content Server may then need to be set up to handle content in different languages correctly and provide a multilingual content management environment to users.

For more general implementation considerations on the subject of users refer to Users (page 3-5).

See also:

– *Managed Content (page 5-5)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*

HARDWARE SETUP

In its most basic form, an extranet consists of two servers: one inside the main firewall, which handles all internal content management tasks, and one outside the main firewall (in the DMZ), which is accessed by all external content consumers. Redundant consumption servers may be set up to improve site availability and reliability (uptime). This may be especially important in sites that provide paid or business-critical services. In addition, depending on the application size and complexity, separate contribution and consumption servers may be used within the firewall. If there is a lot of conversion, a separate server may be dedicated to that task in order to minimize the system load on the "main" content server.

For more general implementation considerations on the subject of hardware setup refer to Hardware Setup (page 3-14).

See also:

– *Managed Content (page 5-5)*

– *Users (page 5-6)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*

INFRASTRUCTURE

In extranet scenarios, external users access content managed by Content Server. This means consumption servers need to be accessible from outside the organization. In order to minimize the security risks, the "outward facing" consumption servers are placed in the DMZ, which serves as a buffer between the organization's secure, internal network and the insecure, external world. Typically, only servers that need to be accessed by both internal and external users are located in the DMZ. This provides external users with access to the content (providing they have the appropriate privileges), while not unnecessarily compromising security.

Depending on the application, a master content server could be set up, with one or more proxied servers. The master server could then be located inside the main firewall, automatically transferring the appropriate content to the proxied consumption server(s) in the DMZ.

Note: If Content Publisher is used to publish the content, the consumption server in the DMZ can either be a Oracle content server or a "normal" web server, depending on how the site was published (see Publication and Data Exchange on page 3-35). Otherwise, it can only be a Oracle content server.

For more general implementation considerations on the subject of infrastructure refer to Infrastructure (page 3-27).

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*

– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*

PUBLICATION AND DATA EXCHANGE

The extranet website resources need to be moved from a (staging) server inside the firewall-protected environment to the consumption server(s) in the DMZ. There are several options to transfer all required data to the consumption server(s):

- Use Oracle Content Publisher to publish to a file system (web server)
- Use Oracle Content Publisher to publish to a content server
- Use Oracle Content Server to replicate content to a content server

Each of these methods has its own characteristics. Which one to choose depends on the existing infrastructure, security, and other considerations. For more information on the various publication methods, including their main characteristics and options, refer to Publication and Data Exchange (page 3-35).

The data transfer to the consumption server(s) in the DMZ takes place through the main firewall. You should therefore generally avoid using direct file share access (mounted drives) to transfer the data, as this may pose considerable security risks (see Firewall on page 3-28).

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*
– *Feature Sets (page 5-12)*

SECURITY

Effective security is extremely important in extranet scenarios since they include servers that are accessible from outside the organization.

All external users need to make themselves known to the system by means of a username and password. Only after the name-password combination is authenticated do external users have access to the extranet site. They can then view content based on their user credentials. Authentication may take place through Active Directories or LDAP.

In addition to the external consumption server(s), there may also be one or more internal consumption servers, which are not accessible by people outside the organization. Internal users need to identify themselves if not all internal content is public—in other words, if access restrictions have been implemented, for example because not all content is intended for everyone in the organization. Users can then access internal content based on their user credentials.

For more information on Oracle-related security refer to the *Security Integration Guide*.

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*
– *Feature Sets (page 5-12)*

CONVERSION OPTIONS

Content on an extranet typically originates from a multitude of diverse sources. It may therefore consist of many different file formats. To ensure consumers can view all content, it is often converted to PDF or HTML before being presented to them. This allows them to view the content, even if they do not have the native applications on their computers (i.e., the applications the content item were originally created in). Content may also be converted to XML for further processing by enterprise applications.

For more general implementation considerations on Oracle's conversion options refer to Chapter 3 (*Basic Considerations)*.

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Metadata Model (page 5-11)*
– *Feature Sets (page 5-12)*

METADATA MODEL

Any metadata model can basically be used in extranet scenarios. Which one works best all depends on very application-specific considerations, such as the organizational structure of the environment in which Content Server runs, the server setup, and security considerations.

Since part of the content is likely used to build the website (for example, image files, Flash objects, etc.), it may be useful to use metadata based on website navigation for this content. This makes it easier to specify where exactly web content items should go on the website.

For more general implementation considerations on the subject of metadata models refer to Metadata (page 3-66).

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*

– *Feature Sets (page 5-12)*

FEATURE SETS

Content Server offers a number of optional feature sets, which provide enhanced functionality to the Content Server core. The following Content Server feature sets are commonly used in extranet scenarios:

- Inbound Refinery
- PDF Converter
- Dynamic Converter
- XML Converter
- Content Tracker
- Content Categorizer
- Report Parser

See also:

– *Managed Content (page 5-5)*
– *Users (page 5-6)*
– *Hardware Setup (page 5-7)*
– *Infrastructure (page 5-8)*
– *Publication and Data Exchange (page 5-9)*
– *Security (page 5-10)*
– *Conversion Options (page 5-10)*
– *Metadata Model (page 5-11)*

Chapter

6

INTERNET SITES

OVERVIEW

Internet sites provide universal accessibility, which means that, in principle, anyone can access it. There is normally no user authentication.

This section addresses two main Internet scenarios:

- Basic scenario: “stand-alone” Internet site
- Extended scenario: Internet site with portal integration

The following sections contain implementation guidelines for each of these scenarios, along with other considerations.

Important: The guidelines provided in this section are by no means hard and fast rules. Every application is different and may require its own, very specific considerations.

Note: The main difference between intranet applications on the one hand and extranet/Internet applications on the other is the fact that with the latter the site needs to be accessible by (some) people outside the organization. This means that extra security considerations play a role.

"STAND-ALONE" INTERNET WEBSITE

Organizations very often set up a presence on the Internet. They may do this for a number of reasons:

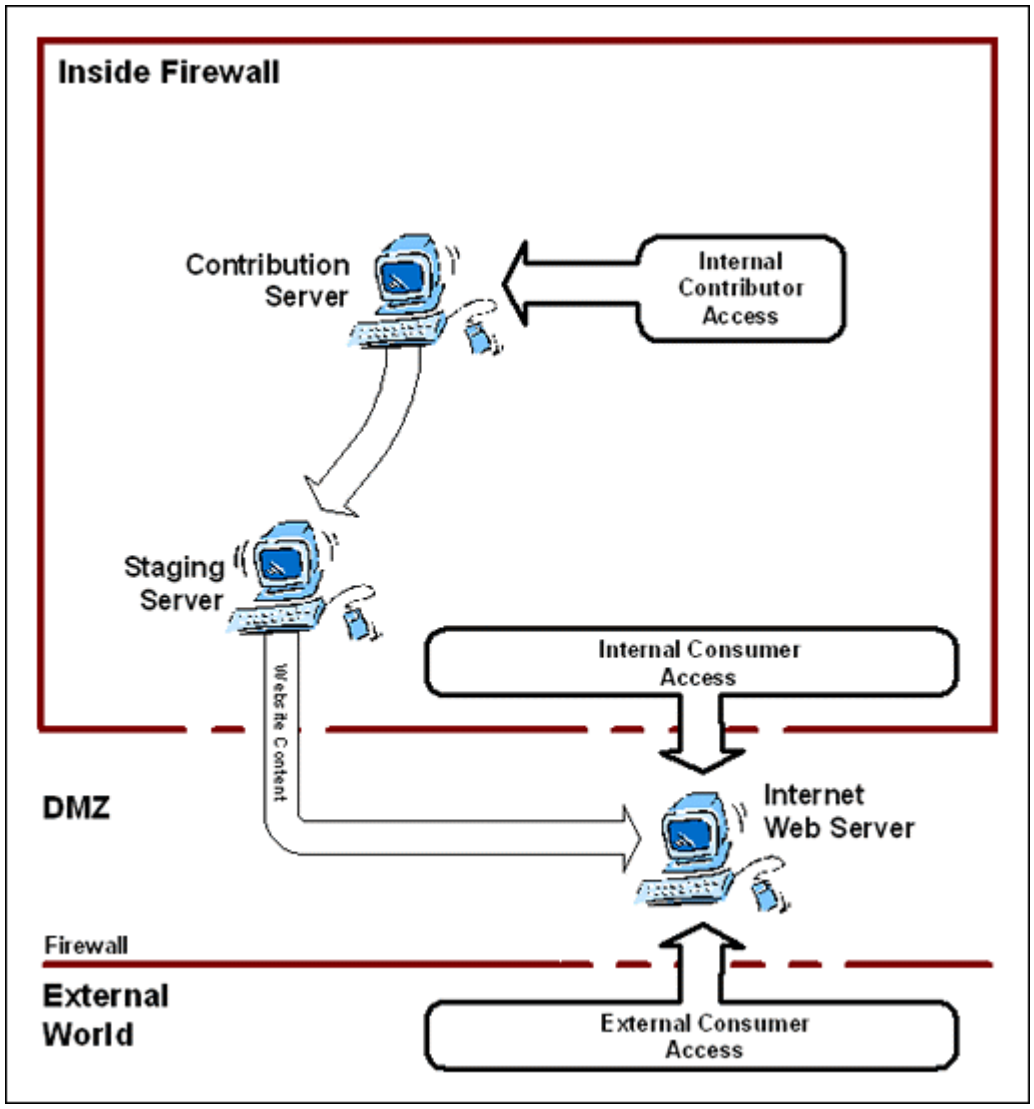
- The Internet site may serve as a "showcase" for an organization's capabilities, including its products and services.
- The Internet site may be used to provide information to the general public on a wide variety of topics.
- The Internet site may be used as a virtual store, where customers can buy items on line.
- The Internet site may be a direct extension—or even an intrinsic part—of a company's core business, especially if it operates in the services or information industry.

As with extranet sites, information is made accessible to people outside the organization. This means these people need to view content managed by Content Server. This is typically done by setting up one or more consumption servers outside the main firewall, which the external consumers (visitors to the website) can access. Figure 6-1 The figure below shows what a basic Internet website application could look like.

Individuals within the organization can contribute directly to the Internet website—providing, of course, they were given the appropriate privileges. As all content contribution is done internally, the contribution server is located within the firewall-protected environment. No-one outside the organization needs to access this server. A staging server is often used as an intermediary step between the contribution server and the consumption server. This enables webmasters and others with the appropriate privileges to verify all content and website elements before they are published on the actual consumption server for literally the whole world to see.

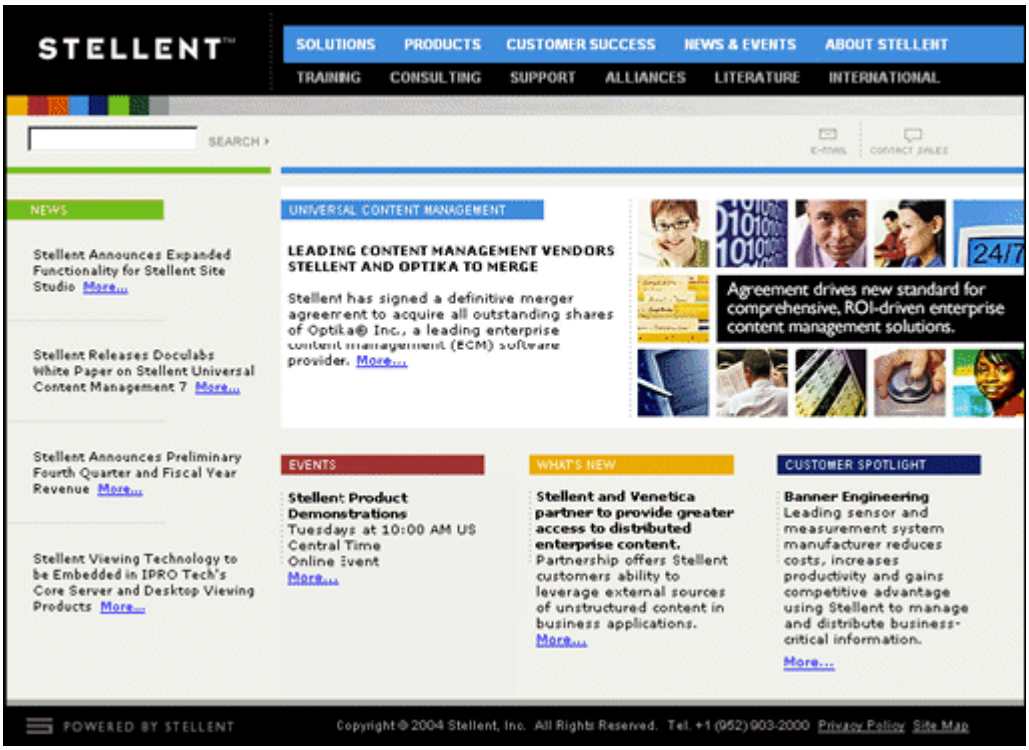
Once everything has been approved on the staging server, the website data is pushed through the main firewall to a consumption server in the DMZ. This is the server visitors to the website actually access when viewing the website.

Figure 6-1 Basic Internet scenario: “stand-alone” website



In many cases, visitors to the Internet website are not even aware they are visiting a site built and managed using the Content Server family of products. Oracle provides products that are capable of running entirely in the background, creating well-designed, fully linked websites including all navigation elements. Figure 6-2 The figure below shows a corporate website, which is entirely powered by the Content Server family of products.

Figure 6-2 Corporate Internet site (powered by Content Server)



See also:

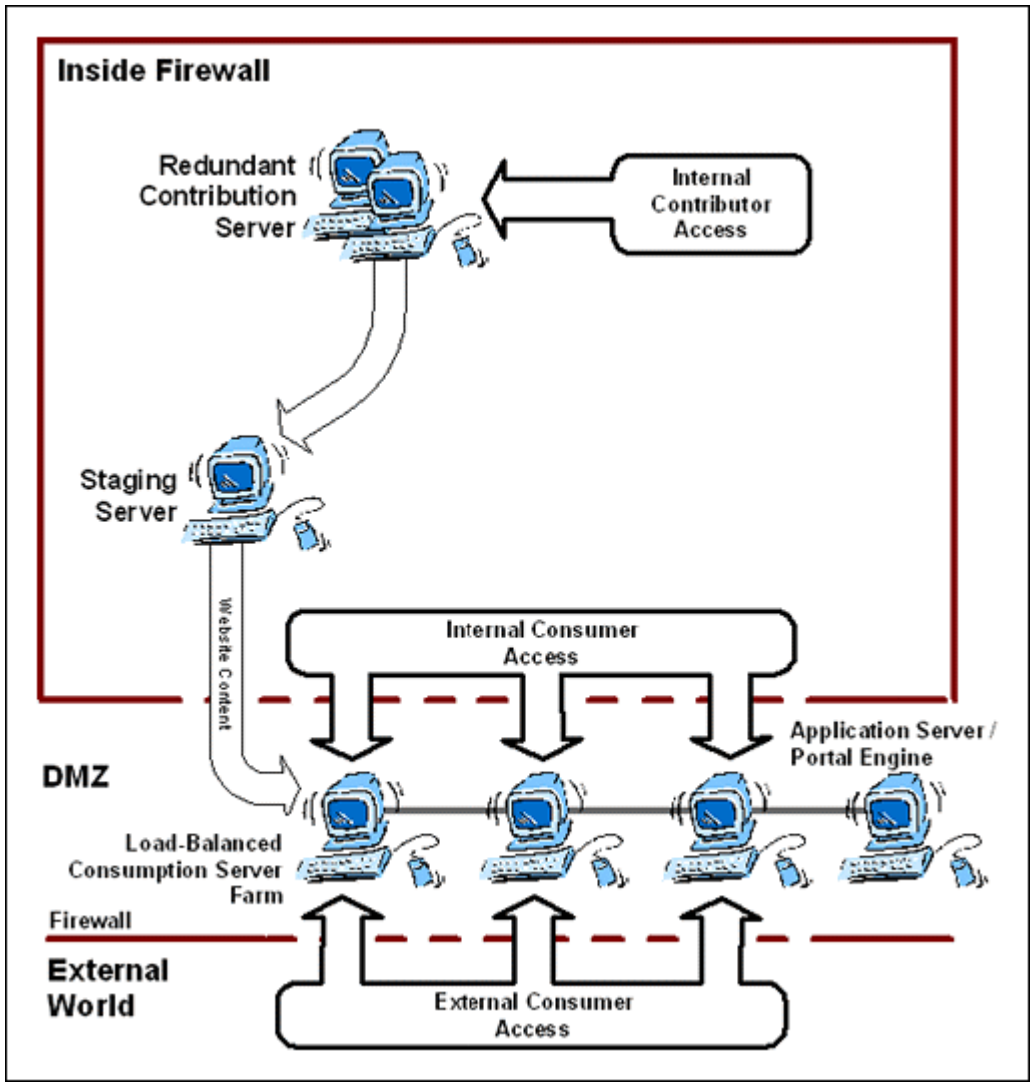
– *Internet Website with Portal Integration* *(page 6-4)*

INTERNET WEBSITE WITH PORTAL INTEGRATION

In addition to creating “stand-alone” Internet websites, Content Server can also be integrated with other enterprise applications such as application servers or portal engines. This may be useful in a wide variety of situations. In an Internet context, for example, it enables organizations to create small, Oracle-based “windows” on their websites that serve as content channels. Such “windows” are called portlets or portal applications. These portlets are populated with content managed by Content Server using an interface with an enterprise application. Content Server is then a content provider for the portal presentation layer.

Figure 6-3 The figure below shows what an Internet website application with portal integration could look like. Its setup is basically identical to that of the Internet website outlined in the previous section. The main difference is that there is an application or portal engine in the DMZ, which controls the portlet functionality of the website.

Figure 6-3 Extended Internet scenario: website with portal integration



It is important that Internet site implementations are scalable. The number of hits may increase dramatically over a relatively short period of time, and the infrastructure should be able to handle this. This is why there are often redundant, load-balanced consumption servers, which can handle large numbers of site visits.

A staging server is mostly used to enable verification and approval of content before it is published on the “live” website. This helps prevent avoidable errors in the website design elements or content.

Figure 6-4 The figure below shows an example of a website with a number of individual portlets, each providing different, personalized pieces of information. A portlet could, for example, show the latest headlines or provide personalized search capability.

Figure 6-4 Content Server-controlled portlets in a website

See also:

– *"Stand-Alone" Internet Website (page 6-2)*

MANAGED CONTENT

Content on an Internet site can be extremely diverse. The nature of the content varies widely between applications, and depends on the purpose of the website as well as its target audience. Since the content is published to a full-fledged website, some of the content may be web content—related to creating the site rather than populating it (for example, Flash objects, images, server pages, or HCSP/HCST pages).

The website data is typically stored in a variety of places, for example databases, ERP systems, content servers, etc. Content for Internet websites is usually not updated in real time, but on a scheduled basis (for example, once a day). The content is typically all public, which means no security needs to be set up to access the content on the website.

The content is primarily intended for people outside the organization. As a result, it may be useful to implement workflows. This helps to ensure content is reviewed and approved by the appropriate people before being released on the website for the external users to see.

For more general implementation considerations on the subject of content refer to Business, Web, and Structured Content (page 2-8) and Managed Content (page 3-3).

See also:

– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Infrastructure (page 6-9)*
– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*
– *Metadata Model (page 6-12)*

USERS

Internet sites have mostly external users, which are all consumers. Since an Internet site is usually public and accessible to everyone, users do not normally have to identify themselves or register to view content on the Internet website. If users do need to register, it is typically not so much for security reasons, but more to log visitor statistics and monitor site usage.

Contribution to the website takes place within the organization by a limited number of people. Content contributors are typically authenticated by the network before they can add new content.

Depending on the application, the number of visitors to an Internet site can be very high and fluctuate dramatically. The Internet application should to be able to handle all this.

For more general implementation considerations on the subject of users refer to Users (page 3-5).

See also:

– *Managed Content (page 6-6)*
– *Hardware Setup (page 6-8)*

– *Infrastructure (page 6-9)*
– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*
– *Metadata Model (page 6-12)*

HARDWARE SETUP

The hardware setup of Internet sites is basically identical to that of extranet sites (see Chapter 5 (*Extranet Sites)*). In both scenarios, the website needs to be accessible to external users—either a limited group (extranets) or the entire online community worldwide (Internet).

A Oracle-based Internet website application typically consists of at least three components:

❖ One or more contribution servers inside the main firewall

❖ A staging server inside the main firewall

❖ One or more consumption servers outside the main firewall, in the DMZ.

Redundant consumption servers may be set up to improve site availability and reliability (uptime). This may be especially important for organizations whose core business relies on the Internet website.

If there is a lot of conversion, a separate server may be dedicated to that task in order to minimize the system load on the “main” content server(s).

For more general implementation considerations on the subject of hardware setup refer to Hardware Setup (page 3-14).

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Infrastructure (page 6-9)*
– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*
– *Metadata Model (page 6-12)*

INFRASTRUCTURE

In Internet scenarios, external users access content managed by Content Server. This means consumption servers need to be accessible from outside the organization. In order to minimize the security risks, the “outward facing” consumption servers are placed in the DMZ, which serves as a buffer between the organization’s secure, internal network and the insecure Internet. Typically, only servers that need to be accessed by both internal and external users are located in the DMZ. This provides all external users with access to the Internet website, while not unnecessarily compromising security.

The number of visitors to an Internet site can be very high, and may fluctuate greatly. It may therefore be important to ensure the application is scalable and can adapt quickly to varying load demands. Load balancing may be important to evenly spread the traffic across all available consumption servers. This helps the website handle peaks in the number of site visits successfully.

Depending on the application, a master content server could be set up, with one or more proxied servers. The master server could then be located inside the main firewall, automatically transferring the appropriate content to the proxied consumption server(s) in the DMZ.

Internet website applications often use staging servers in conjunction with Oracle Content Publisher. At regular intervals, Content Publisher generates the website based on the content in the content server, and publishes it to the staging server. There the site can be checked and verified before it is released to the consumption server(s).

Note: In practice, some content may be intended for internal use only, while other content is for the Internet site. In that case, there may be one or more consumption servers inside the firewall in addition to the ones in the DMZ.

For more general implementation considerations on the subject of infrastructure refer to Infrastructure (page 3-27).

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*

– *Metadata Model (page 6-12)*

PUBLICATION AND DATA EXCHANGE

The Internet website resources need to be moved from a (staging) server inside the firewall-protected environment to the consumption servers in the DMZ. There are several options to transfer all required data to the consumption server(s):

- Use Oracle Content Publisher to publish to a file system (web server)
- Use Oracle Content Publisher to publish to a content server
- Use Oracle Content Server to replicate content to a content server

Each of these methods has its own characteristics. Which one to choose depends on the existing infrastructure, security, and other considerations. For more information on the various publication methods, including their main characteristics and options, refer to Publication and Data Exchange (page 3-35).

The data transfer to the consumption server(s) in the DMZ takes place through the main firewall. You should therefore generally avoid using direct file share access (mounted drives) to transfer the data, as this may pose considerable security risks, especially if the entire online community worldwide can access the website (see Firewall on page 3-28).

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Infrastructure (page 6-9)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*
– *Metadata Model (page 6-12)*

SECURITY

Internet websites are generally intended to be accessed by anyone who wants to. This means there is no user authentication for content consumption. Of course this does not mean security is not an issue with Internet scenarios. The challenge is to provide optimum accessibility to the Internet site, while not compromising the security of the organization's

resources behind it. This is especially important because anyone may potentially try to break through the firewalls. On the consumption side, it may be the application or portal server that handles security.

Contrary to content consumers, contributors typically do need to be authenticated before they are allowed to add or change content. This can, for example, be done through Content Server security or LDAP.

For more information on Oracle-related security refer to the *Managing Security and User Access* guide.

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Infrastructure (page 6-9)*
– *Publication and Data Exchange (page 6-10)*
– *Conversion Options (page 6-11)*
– *Metadata Model (page 6-12)*

CONVERSION OPTIONS

As with extranet, content on an Internet website typically originates from a multitude of diverse sources. It may therefore consist of many different file formats. To ensure content consumers can view all content, it is often converted to PDF or HTML before being presented to the consumers. This allows them to view the content, even if they do not have the native applications on their computers (i.e., the applications the content item were originally created in). Content may also be converted to XML for further processing by enterprise applications.

For more general implementation considerations on Oracle's conversion options refer to Chapter 3 (*Basic Considerations)*.

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Infrastructure (page 6-9)*

– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Metadata Model (page 6-12)*

METADATA MODEL

Any metadata model can basically be used in extranet scenarios. Which one works best all depends on very application-specific considerations, such as the organizational structure of the environment in which Content Server runs, the server setup, and security considerations.

Since part of the content is likely used to build the website (for example, image files, Flash objects, etc.), it may be useful to use metadata based on website navigation for this content. This makes it easier to specify where exactly web content items should go on the website.

For more general implementation considerations on the subject of metadata models refer to Metadata (page 3-66).

See also:

– *Managed Content (page 6-6)*
– *Users (page 6-7)*
– *Hardware Setup (page 6-8)*
– *Infrastructure (page 6-9)*
– *Publication and Data Exchange (page 6-10)*
– *Security (page 6-10)*
– *Conversion Options (page 6-11)*

Appendix

A

PLANNING

OVERVIEW

To ensure a content management solution meets all needs of the organization, adequate planning is crucial. It is essential that a number of key questions and issues are addressed before starting with the actual implementation of the system.

The questions and issues are related to the following main areas:

This appendix section aims to help you plan a Oracle-based content management solution and ask the right questions before starting its implementation.

CONTENT

Key questions to address in the area of content include the following:

1. What file formats will be managed by Content Server?
2. What are the file extensions of each file format? Will an application's default extensions (for example, xls for Microsoft Excel files) be used or customized aliases?
3. How can the content be categorized? In other words, what content types can be defined? Content types are typically characterized by purpose, product type, or group responsible for creating or managing them (for example, AutoCAD Engineering Drawings, Sales Proposals, etc.). Also, who creates each content type?
4. What is the expected number of content items for each content type that will be managed by Content Server?
5. What are the average file sizes of the content? This refers not only to the byte counts (number of kilobytes), but also to average page numbers.
6. Is there any content with special file structures? This may include compound documents or files, documents or files with embedded or linked graphics, etc.
7. Is there any custom content information (metadata) that needs to be defined?
8. Is there any existing content that needs to be "imported" into the content management system? If so, what file formats, how much, and from where?

The tables on the next few pages address the questions above. When answering them, make sure that you not only consider the current situation, but also take into account any known or planned future requirements.

Questions 1 and 2: File Formats and File Extensions

Software Application	File Extension(s)	Extension Aliases (if any)
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		

Question 3: Content Types

Content Type Name (max. 30 characters)	Description	Created By
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		

Questions 4 and 5: Number and Size of Content Items

Content Type	Average Size (Pages)	Average Size (KB)	Total No. of Content Items	Anticipated Storage Requirements
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				

Question 6: Special File Structures

Content Type	Special File Structures
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	

Question 7: Custom Content Information (Metadata)

Custom Content Information	Description / Comments
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	

Question 8: Existing Content

File Formats of Content to be Imported	Number of Files to be Imported	Current Location
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		

See also:

– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Search Solution (page A-23)*
– *Publication and Conversion (page A-24)*
– *Workflows (page A-26)*
– *Security (page A-28)*
– *Additional Considerations (page A-29)*
– *Customization Requirement Definition (page A-30)*

USERS

Knowing what the user base will be for a content management application is very important. Key questions to ask in this area include the following:

1. Where are all users located? Are they on a single location or on different locations? If there are multiple locations, where are they (local, national, global)?

2. How many content consumers will there be?

3. How many content contributors will there be?

4. How many (sub-)administrators will there be?

5. Who will be able to view the content managed by Content Server?
   - ❑ People inside the organization.
   - ❑ A defined group of people outside the organization.
   - ❑ Everybody outside the organization (no access restrictions).
   - ❑ Other: _____

6. Who creates each content type and who accesses it? Using the table on page A-4 as a reference, list by content type the functional groups that will be contributing or accessing content.

Content Type	**Group Creating**	**Group Accessing**
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		

Content Type	Group Creating	Group Accessing
10.		
11.		
12.		

See also:

– *Content (page A-2)*

– *Hardware (page A-11)*

– *Infrastructure (page A-16)*

– *Search Solution (page A-23)*

– *Publication and Conversion (page A-24)*

– *Workflows (page A-26)*

– *Security (page A-28)*

– *Additional Considerations (page A-29)*

– *Customization Requirement Definition (page A-30)*

HARDWARE

There are a number of important issues related to hardware that need to be addressed before implementing a content management solution. They are related to the following:

❖ Server Systems (page A-12)

❖ Client Stations (page A-15)

See also:

– *Content (page A-2)*

– *Users (page A-9)*

– *Infrastructure (page A-16)*

– *Search Solution (page A-23)*

– *Publication and Conversion (page A-24)*

Server Systems

Key questions to ask for the server systems include the following:

1. How many servers are required? (For details refer to Number of Servers on page 3-15 and further.)
   - Number of consumption/deployment servers: _____
   - Number of contribution servers: _____
   - Number of staging servers: _____
   - Number of test/development stations: _____
2. Where will the servers reside (inside/outside firewall, etc.)?
   - Consumption/deployment server(s):
   - Contribution server(s):
   - Staging server(s):
3. Will there be separate testing/development and production environments? (For details refer to Separation of Content Server Environments on page 3-19.)

   Separation of environments: ❑ Yes ❑ No
4. What hardware and software configuration will each server have? This refers to the combination of operating system, database, and web server. Consider all computers involved, including those running Content Server, Content Publisher, Inbound Refinery, the database, and/or the web server. Also, make sure that you provide detailed version information, including the service pack level.

Attribute	Configuration 1	Configuration 2	Configuration 3
Server type (e.g. Content Server 2)			
RAM (in MB)			
CPU (type and model)			
Hard disk space			
Operating system			
Web server			
Database			
Network connectivity			
Devices currently connected (if any)			
_____			

Attribute	Configuration 4	Configuration 5	Configuration 6
Server type (e.g. Content Server 2)			
RAM (in MB)			
CPU (type and model)			
Hard disk space			
Operating system			
Web server			
Database			
Network connectivity			
Devices currently connected (if any)			
_____			

5. Based on the information provided on pages A-2 and further, Based on the content requirements, what are the expected storage requirements?

See also:

– *Client Stations (page A-15)*

Client Stations

Key questions to ask for the client stations include the following:

1. How many client stations will there be and where are they located?

2. What hardware and software configurations will the client stations typically have? This includes the platform (for example, Windows, Macintosh, UNIX) as well as operating systems (for example, Windows 98, Windows 2000 Professional, Windows XP Professional, Linux). Also describe network access requirements for these systems (for example, Ethernet, dial-up, etc.).

3. Which web browser(s) will content consumers and contributors use?
   - ❑ Internet Explorer 5.5 SP2
   - ❑ Internet Explorer 6.0 SP2
   - ❑ Internet Explorer 7.0
   - ❑ Mozilla 1.7
   - ❑ Firefox 1.5
   - ❑ Firefox 2.0
   - ❑ Safari 2.*x*
   - ❑ Other:
   - ❑ Other:
   - ❑ Could be any web browser

Note: See the Content Server installation guides for details on the supported web browsers for use with Content Server.

4. Which web browser(s) will (sub-)administrators use?
   - ❑ Internet Explorer 5.5 SP2
   - ❑ Internet Explorer 6.0 SP2
   - ❑ Internet Explorer 7.0
   - ❑ Mozilla 1.7
   - ❑ Firefox 1.5
   - ❑ Firefox 2.0
   - ❑ Other:

Note: See the Content Server installation guides for details on the supported web browsers for use with Content Server.

See also:

– *Server Systems (page A-12)*

INFRASTRUCTURE

There are a number of important issues related to infrastructure that need to be addressed before implementing a content management solution. They are related to the following:

❖ Network Considerations (page A-17)

❖ Data Integration Requirements (page A-20)

❖ Data Distribution Requirements (page A-21)

See also:

– *Content (page A-2)*

– *Users (page A-9)*

– *Hardware (page A-11)*

– *Search Solution (page A-23)*

– *Publication and Conversion (page A-24)*

– *Workflows (page A-26)*

– *Security (page A-28)*

– *Additional Considerations (page A-29)*

– *Customization Requirement Definition (page A-30)*

Network Considerations

Network Infrastructure

What are your existing network infrastructure support needs? Describe in as much detail as possible your existing network infrastructure support needs. This includes the existing network security model, server environment in place or planned, existing data storage for files you wish to manage, and storage space requirements.

Backup and Recovery Systems

What are the backup and recovery systems in place for the hardware and file system environment? Describe the backup and recovery systems in place for the hardware and file system environment:

Note: See Backup Strategy (page 3-72) for more information on backup and recovery guidelines.

E-Mail Systems

Is e-mail in place for all users? What are the characteristics of the e-mail system (server address, administrator e-mail address, etc.)? Describe the e-mail system (server address, administrator e-mail address, etc.):

External Server Database

If the database will not reside on the Content Server content server, what server hardware, software environment, and network access exists or will be used? describe the server hardware, software environment, and network access that exists or will be used:

Is the database currently installed and in use? ❑ Yes ❑ No
If so, what other software applications (if any) access this database?

If so, describe the other software applications (if any) that access this database:

Client Access Systems

What are the characteristics of the client systems (for example, Intel-based CPU with Windows 98, NT Workstation, Macintosh, etc.)? This includes network access requirements for these system (for example, Ethernet, dial-up, etc.).

Describe the client systems (for example, Intel-based CPU with Windows 98, NT Workstation, Macintosh, etc.). Also describe network access requirements for these systems (for example, Ethernet, dial-up, etc.).

Client Software for Viewing Content

What client software is available to view the content (for example, Acrobat Reader, Microsoft Word, etc.)? Which releases?

Describe all client software available to view the content (for example, Acrobat Reader, Microsoft Word, etc.). Describe the release levels of all software used.

See also:

– *Data Integration Requirements (page A-20)*

– *Data Distribution Requirements (page A-21)*

Data Integration Requirements

Once the Oracle content management solution is in place, will the managed data need to be exported to any other existing data management systems? For example, material requirements planning (MRP) systems, billing or invoicing systems, product data management systems, mainframe reporting systems, etc.

If so, what systems are currently in place or planned that will need data exported to from the Oracle system?

If so, describe any systems currently in place or planned that will need data exported to from the Oracle system:

What systems are currently in place or planned that data will be migrated from and into the Oracle system? What are the data types, anticipated file sizes, and purpose or objectives of integration?

Describe any systems currently in place or planned that data will be migrated from and into the Oracle system. Describe data type, anticipated file sizes, and purpose or objectives of integration (see also the table on page A-8):

See also:

– *Network Considerations (page A-17)*

– *Data Distribution Requirements (page A-21)*

Data Distribution Requirements

What are the distribution needs, if any, of content that will be managed by the Oracle system? This could include different view format requirements for internal or external customers of the server installation, release timing requirements, replication of the data to external or remote servers, etc.

Document Distribution

What is the the current content distribution model in place?

Describe the current content distribution model in place:

What are the access or replication requirements for intended viewers of the data? (Consider remote users who may not have network access, partial data replication needs depending on location, intra- or interdepartmental needs, etc.)

Describe the access or replication requirements for intended viewers of the data (consider remote users who may not have network access, partial data replication needs depending on location, intra- or interdepartmental needs, etc.):

What security concerns are there for access to the distributed data, if any?

Describe any security concerns for access to the distributed data:

What special access speed requirements are there due to data type or use, if any?

Describe any special access speed requirements due to data type or use:

Content Information Definition

Are there any special content search and retrieval requirements that exist with data access? Describe any special content search and retrieval requirements that may exist with data access. Are users currently using specific data descriptions (for example, titles, file naming conventions, project numbers, location, or any other data description elements that aid in the location of content)?

Viewing Distribution Definition

Are there any special viewing integration needs for the content distribution? Describe any special viewing integration needs for the content distribution. Consider intranet or Internet viewing currently in place for your existing web environment. Will the Oracle system need to be integrated with an existing web environment, etc.?

See also:

– *Network Considerations (page A-17)*

– *Data Integration Requirements (page A-20)*

SEARCH SOLUTION

Key questions to address in the area of search capabilities include the following:

1. What search solution will be used for **metadata** searching and indexing?
   - ❑ Database (SQL Server)
   - ❑ Database (Oracle)
   - ❑ Database (DB2)
   - ❑ Database (Sybase)
   - ❑ Database (PostgreSQL)
   - ❑ Verity
   - ❑ FAST
2. What search solution will be used for **full-text** searching and indexing?
   - ❑ None (full-text search not available)
   - ❑ Database (SQL Server)
   - ❑ Database (Oracle)
   - ❑ Database (DB2)
   - ❑ Verity
   - ❑ FAST

Tech Tip: A useful document in this respect is *Choosing a Search Solution*, which discusses the factors to consider when selecting a search solution for use with Content Server. It is available as a PDF file called *search_solution_80en.pdf* on the Content Server documentation CD (in the *integrator* directory).

Note: There are integration documents available for both the Verity and FAST search solutions.

See also:

– *Content (page A-2)*
– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Publication and Conversion (page A-24)*

PUBLICATION AND CONVERSION

Key questions to address in the area of publication and conversion include the following:

1. Do you want to publish to a file system, portal server, or content server?
   - ❑ File system (directory structure)
   - ❑ Content Server
   - ❑ Portal server:
   - ❑ Other:

2. Do you want to publish dynamically (i.e., as the need arises), on a scheduled basis (i.e., with regular intervals or as certain criteria are met), or both?
   - ❑ Only dynamic publication:
   - ❑ Only scheduled publication:
   - ❑ Both dynamic and scheduled publication:

3. How do want to push data through the firewall? (See Firewall on page 3-28 for more details.)
   - ❑ Direct file share access [not recommended]
   - ❑ File Transfer Protocol (FTP)
   - ❑ Secure socket
   - ❑ HyperText Transfer Protocol (HTTP)
4. What should the library hierarchy be? (For further details, refer to the *Getting Started With Content Server* guide.) Describe the current or planned hierarchical structure for content distribution with which users are familiar. Is content organized by department,

information type, file location, product type, internal company users, or external customers? For example:

DIVISION

Engineering
Manufacturing Process Control
Procurement

Manufacturing

Research
Engineering
Testing/QA

HUMAN RESOURCES

5. How will content consumers use the published files?
   - ❑ Viewing
   - ❑ Searching
   - ❑ Downloading
   - ❑ Combination:

6. How would the users retrieve information, what do they want, and how would they typically search for it?

See also:

– *Content (page A-2)*
– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Search Solution (page A-23)*
– *Workflows (page A-26)*
– *Security (page A-28)*
– *Additional Considerations (page A-29)*
– *Customization Requirement Definition (page A-30)*

WORKFLOWS

For certain content types (see the table on page A-4), there may be required review and approval steps or processes for managing the content creation or revision process. Consider also file or document content requirements such as embedded graphics or objects that may be external to the source file or document, links to other areas within or external to the files or documents, data import requirements, etc.

❖ **Content Type:**

Creation Process:

Review and Approval Process:

Release and Information Distribution:

- **Content Type:**

  Creation Process:

  Review and Approval Process:

  Release and Information Distribution:

- **Content Type:**

  Creation Process:

  Review and Approval Process:

  Release and Information Distribution:

- **Content Type:**

  Creation Process:

  Review and Approval Process:

  Release and Information Distribution:

- **Content Type:**

  Creation Process:

  Review and Approval Process:

  Release and Information Distribution:

❖ **Content Type:**

Creation Process:

Review and Approval Process:

Release and Information Distribution:

See also:

– *Content (page A-2)*
– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Search Solution (page A-23)*
– *Publication and Conversion (page A-24)*
– *Security (page A-28)*
– *Additional Considerations (page A-29)*
– *Customization Requirement Definition (page A-30)*

SECURITY

Key questions to address in the area of security include the following:

1. How is security-related information about users stored?
   - ❑ Content Server Security
   - ❑ Microsoft Network
   - ❑ LDAP
   - ❑ ADSI
   - ❑ Other:

2. What security model will be used?

See also:

– *Content (page A-2)*
– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Search Solution (page A-23)*
– *Publication and Conversion (page A-24)*
– *Workflows (page A-26)*
– *Additional Considerations (page A-29)*
– *Customization Requirement Definition (page A-30)*

ADDITIONAL CONSIDERATIONS

As you plan for your system, keep the following additional considerations in mind:

- Expect an increase in the information that you manage.
- Expect an increase in consumer and contributor access.
- Maximize search performance.
- Maximize browser performance.
- Maximize refining and indexing performance.

- Coordinate system maintenance and customization.
- Perform regular system backups (see page 3-72).

See also:

- *Content (page A-2)*
- *Users (page A-9)*
- *Hardware (page A-11)*
- *Infrastructure (page A-16)*
- *Search Solution (page A-23)*
- *Publication and Conversion (page A-24)*
- *Workflows (page A-26)*
- *Security (page A-28)*
- *Customization Requirement Definition (page A-30)*

CUSTOMIZATION REQUIREMENT DEFINITION

There may be additional customizations that have not been covered in other sections of this Appendix this section. Examples of what could be included in this section would be:

- Custom web page formatting or content for the initial server home page
- Custom search screen formatting specific to your site needs
- Data viewing format conversions for files or document types not currently supported by the content refinery system
- Mixed format viewing requirements dependent on viewer or user location, security level, etc.
- Custom document check-in screen modifications specific to users or location
- Data import or export requirements to or from the Oracle content server

See also:

– *Content (page A-2)*
– *Users (page A-9)*
– *Hardware (page A-11)*
– *Infrastructure (page A-16)*
– *Search Solution (page A-23)*
– *Publication and Conversion (page A-24)*
– *Workflows (page A-26)*
– *Security (page A-28)*
– *Additional Considerations (page A-29)*

Index

D

E

F

G

H

I

T

U

V

W

X