

Distributed Content Server

How To Deploy Content Server in a Distributed Environment

TABLE OF CONTENTS

Executive Summary	2
The Problem	3
A Project Driven Business	3
The Case for Distributed Content Server.....	3
Geo-replication	4
Content Server Extranet	5
Real-Time Active-Active Disaster Recovery	5
Content Aggregation and Syndication.....	6
The Alternatives	6
The Challenge	6
Remote Cache	7
Uniting Portal.....	7
Replication.....	7
The Solution	8
The Approach.....	9
Syntergy Replicator for Content Server.....	9
Distribution Matrix.....	9
Flexibility and Control	10
Enterprise Searching	10
Modular approach	10
Administration	11
Benefits	11
Special Features.....	11



Confidentiality

This document contains confidential information of www.syntergy.com, Inc., which is provided for the sole purpose of permitting the client to evaluate the proposal submitted and any information provided subject to this project. In consideration of the receipt of this document, the recipient agrees to maintain such information in confidence. In preparing this document, www.syntergy.com, Inc. may have been provided with information by the client about its systems, business or strategy and agrees to maintain this information in confidence.

Each party agrees not to reproduce or otherwise disclose this information to any person outside the group directly responsible for evaluation of its contents unless this information is publicly known through no fault of either party or ownership of the intellectual property contained in this document has passed through payment or other agreed consideration by the recipient to www.syntergy.com, Inc.

Content Server is a registered trademark of OpenText Corporation.

EXECUTIVE SUMMARY

The challenge for a modern, global business is to provide consistent information to support the organization across boundaries imposed by geography, while minimizing the cost of telecommunications and IT infrastructure.

Units of any business need efficient access to tools so that they can compete locally, and through collaboration, effectively leverage resources of the organization globally. This is the case regardless of whether they are operational or project driven. The modern knowledge organization is also a competitive one. To be successful it must not only apply its resources more effectively than the competition, it must do so by utilizing them wherever they exist within the business.

Conventional Enterprise Document Management Systems (EDMS) are an essential element of the infrastructure required for organizations to manage information across global boundaries. These systems offer a highly efficient way to store and re-use business information as well as an excellent environment for collaboration in project delivery. However, they are demonstrably less successful in businesses with a distributed organizational structure – those with a network of locations rather than one central head office structure that controls and manages information services.

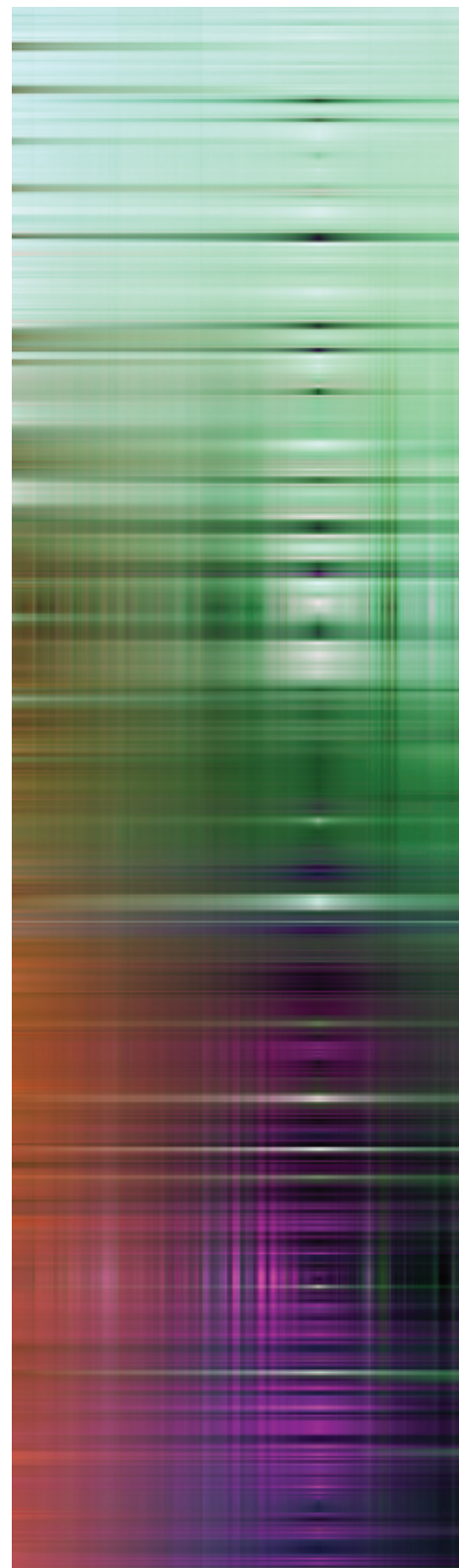
Few purely centralized organizations exist – the global economy is forcing the change. As important as global collaboration is, organizations succeed only if performance at the local level is effective. Any solution addressing global availability of documents must also continue to perform well at the point of service delivery.

The solution outlined in this document recognizes and builds on the inherent value of the document management architecture offered by Content Server and the powerful collaboration capability it provides.

The solution works with Content Server's inherent central server architecture to provide a cost-effective method for supporting a distributed network of Content Server servers with selective replication of data. Unlike the alternate solutions of remote caching, uniting portals or regional replication, the solution optimises the use of bandwidth and IT infrastructure while providing local access at local Area Network (LAN) speed.

It does so while supporting and leveraging all of Content Server's capabilities, thereby providing a truly powerful, globally capable, knowledge management and collaboration tool.

This document defines the scope of the solution as “Distributed Content Server”.



THE PROBLEM

A PROJECT DRIVEN BUSINESS

In examining Distributed Content Server solutions it is perhaps important to discuss the nature of the distributed business, which necessitated the development of the solution.



The business initiating the “Distributed Content Server” project was a global consulting company with 3000+ employees located in 50 regional project offices across Australia, South East Asia, South America and the United Kingdom with developing interests in Eastern Europe. Its central business function is the application of knowledge and development of documents for delivery subject to specific project needs. At any one time 10,000 projects are active with 100 new projects started each day. The projects share resources with between 1 and 300 staff per project and may vary between \$5,000 and \$50 Million. Projects cross times zones, geographical areas, and national boundaries with stakeholders from different companies. Commonly projects are composed as “virtual teams” that include internal resources with disciplines from different locations and external resources from project partners and customers. Document content may include word processing documents, spreadsheets, CAD, maps, pictures, graphics or models. The organization’s data exceeds five terabytes in cumulative size.

Although initially applied to a project driven business in the engineering field, the “Distributed Content Server” solution is not limited to this model and will successfully support organizations wishing to maintain regional offices or strategic business units in any business structure.

THE CASE FOR DISTRIBUTED Content Server

Distributed Content Server refers to OpenText Content Server environments that span multiple instances – located in either a single location or spanning multiple locations. An instance is defined as one set of database tables used by one implementation of Content Server. This article describes the common application scenarios where a Distributed Content Server solution is warranted and why a Distributed Content Server solution is the answer.

Distributed Content Server | THE PROBLEM

The easiest way to understand where a Distributed Content Server solution is the best answer is to consider the following application scenarios.

Application Scenario	Description
Geo-Replication	Local, national or global deployments of multiple Content Server instances with a single server or multiple load-balanced servers in each implementation.
Extranet with same Domain or Cross-Domain Synchronization	Multiple Content Server instances deployed in the same location to support extranet or other application-specific scenarios.
Real-Time Active-Active Disaster Recovery	Secondary Content Server instances deployed to support disaster recovery and COOP (Continuity of Operation) scenarios
Content Aggregation and Syndication	Collection and/or distribution of Content Server content between two or more web applications (including Content Server workflows and distributed web content management).

It is common for most OpenText customers to have a need for one or more of these scenarios in their organization. Let's look at each of these in turn.

GEO-REPLICATION

Why is geo-replication an important application scenario? The business driver for implementing a geo-replication solution is the need for organizations with multiple Content Server instances to synchronize all or part of their Content Server content across some or all of their Content Server instances. This turns out to be a common requirement. The requirement for multiple local Content Server instances can result from several scenarios:

- a merger or acquisition and the need to consolidate or synchronize Content Server content
- a requirement to address WAN bandwidth, latency and reliability issues by deploying additional local or regional Content Server instances
- Smaller deployments of Content Server (departmental or project oriented) that need to be synchronized with or migrated to a company's primary Content Server implementation

Distributed Content Server | THE PROBLEM

GEO-REPLICATION CONTINUED

The need to address WAN bandwidth, latency and reliability issues deserves special mention. This is the most frequent driver for a Distributed Content Server environment. Here are some examples:

- Lawyers in a global firm will want immediate access to case files
- An engineering team at a remote construction site needs fast access to a large variety of business process documents and potentially very large drawing files
- The crew of a cruise ship with slow or unreliable network connections needs fast, reliable access to up-to-date procedures and information needed to run the ship
- A battlefield scenario where communications between command head quarters and remote Content Server instances may be limited by slow or unreliable WAN connections

For this Distributed Content Server scenario to be effective, it needs to efficiently support near-real-time (NRT) document-level replication.

Content Server EXTRANET

For Content Server Extranet scenarios, a common requirement is the ability to publish content authored internally using the corporate intranet to a Content Server extranet that is isolated from the internal network using separate web applications, Forms Based Authentication (FBA) and/or different Active Directory domains. For this scenario, the ability to replicate Content Server content across different authentication protocols (Windows NTLM and FBA) or multiple Active Directory domains are the key requirements.

REAL-TIME ACTIVE-ACTIVE DISASTER RECOVERY

There are numerous Content Server disaster recovery strategies and technology solutions, but one of the simplest to deploy and support is near-real-time (NRT) replication of a Distributed Content Server implementation to an online, active disaster recovery instance. This is easily enabled using a NRT replication of document or item-level changes to a disaster recovery instance that can act as a full or partial replica of the source environment. The replication requirements are similar to the previous application scenarios:

- NRT replication.
- A Distributed Content Server replication solution is ideal when the hit disaster recovery instance is located in a different city or a different part of the world where WAN efficiency is an important factor.



CONTENT AGGREGATION AND SYNDICATION

Content aggregation in a Content Server context refers to the collection and replication of Content Server content (almost all Content Server object types are supported including work-flows) from one or more Content Server instances to another instance. A common example is the replication of work products produced in a remote or regional office to a central headquarters farm. Content syndication refers to the distribution or broadcasting of content from a central Content Server instance to one or more remote instances. In addition, there are hybrid scenarios where multiple source instances broadcast content to one or more target instances. The key replication requirements for these scenarios include the above (NRT document-level replication) plus the ability to replicate objects to locations in a target environment that is structured differently from the first.

THE ALTERNATIVES

The challenges outlined above remain a problem that forces distributed enterprises to make a choice between two unsatisfactory alternatives. The first limits the Document Management System to centralized corporate information only. This means a loss of the substantial value of business knowledge that comes from day-to-day work.

The second is to distribute the Document Management System to various locations, causing fragmentation. Each site only sees the information in its slice of the Document Management System and teams working collaboratively across the organization lose a coherent picture of their projects.



There is a real need to build a decentralized Document Management System without fragmenting business knowledge and while still maintaining coherency in a collaborative team environment. Three responses to this challenge exist – Remote Caches, Uniting Portals or Replication – while each has advantages, each approach also has limitations.

Distributed Content Server | THE ALTERNATIVES

REMOTE CACHE

For Content Server users the Remote Cache Server provides some support for a distributed environment deploying presentation elements of Content Server to remote locations to minimize data transmission to only content items. Copies of documents held on the central server may be selectively forwarded to the cache server to optimize the use of bandwidth for distribution of documents from the central server.

However, the solution provides only performance advantages to the remote server maintaining documents control as other remote sites will still access the central server for read-only copies. Additionally, document control between remote locations must still pass through the central server.

In short, the solution partially solves the problem but still does not address the needs of a distributed group accessing a single virtual collection while minimizing performance overheads and devalues the business effectiveness of the Enterprise Document Management System.

UNITING PORTAL

Recently the emergence of portal technologies has offered an approach using a uniting portal that spans every distributed installation of the Document Management System across the organization. Information that is in fact dispersed appears to be located centrally. Gathering information in real time from every site supports the illusion of a collective enterprise-wide view. However, the portal is only ever as fast as its weakest link.

Having many WAN links involved simultaneously leaves the uniting portal highly susceptible to communications line performance issues. These are commonly line errors, poor latency and overloaded links, all of which are typically experienced on every Wide Area Network.

REPLICATION

Replication emerged as a method for managing data across distributed organizations over a decade ago. It promotes installation across all locations of a copy of the Document Management System with a background process that replicates transactions between systems. Whenever a document at one location is updated or changed, the corresponding document at every other location is also updated.

This traditional approach ensures every site has effective access to common information across the enterprise as well as a fully functional collaborative environment. However, the enormous amount of background processing required to keep all locations up to date significantly increases bandwidth needs. Scheduling batch updates of changes in windows of opportunity may be

a compromise.



REPLICATION CONTINUED

However, in a large organization the window of opportunity may be limited and not allow all changes to be effected within the time allowed. Opportunity for conflicts is also greatly increased as numerous changes are held for after hours processing.

Substantial resources are also required to maintain such an environment minimizing the benefits delivered. Storage problems emerge, as every location needs the capacity to store all documents.

One solution might be partial replication where only selected content is replicated from a central server, but this leads to an inevitable compromise - as bandwidth solutions are sought, advantages of information access are lost. The necessity to collaborate between all offices and not just with a central location means a move toward full replication again.

THE SOLUTION

“Distributed Content Server” creates a virtual Content Server environment that offers a greatly enhanced architecture to Content Server users.

It selectively replicates areas of the Content Server Document Management System structure; including security, metadata, categories, document information, task lists, workflows, news channels, discussions, and so on – but not the document files themselves.

The solution exploits the fact that the document files constitute a very high proportion of the total Document Management System storage requirements.

The documents are still available enterprise-wide but in a form that minimises storage and bandwidth requirements. When configuring the environment the administrator indicates, based on use dynamics, what content is to be distributed and in what form.



Distributed Content Server | THE SOLUTION

THE APPROACH

Synergy Replicator for Content Server

Synergy Replicator for Content Server is an enterprise Content Server replication solution that supports near real-time, cross-domain, document-level synchronization of any network of Distributed Content Server instances. Replicator installs and is managed in the same way that other Content Server modules are installed and managed. Replicator detects changes in the Content Server content as they occur and batches these changes into replication packages. A Java based Transfer Service (Java Assistant) is employed by Replicator as the highly-reliable, re-startable transport for downloading replication packages over fast, slow and unreliable network connections. Replicator uses Java compression capability to compress packages making for efficient use of the WAN bandwidth.

Distribution Matrix

The decision to distribute projects or folders between locations is defined by business use and bandwidth capabilities. A distribution agreement can be initiated from any participating server and target any or all other servers in the virtual Content Server environment.

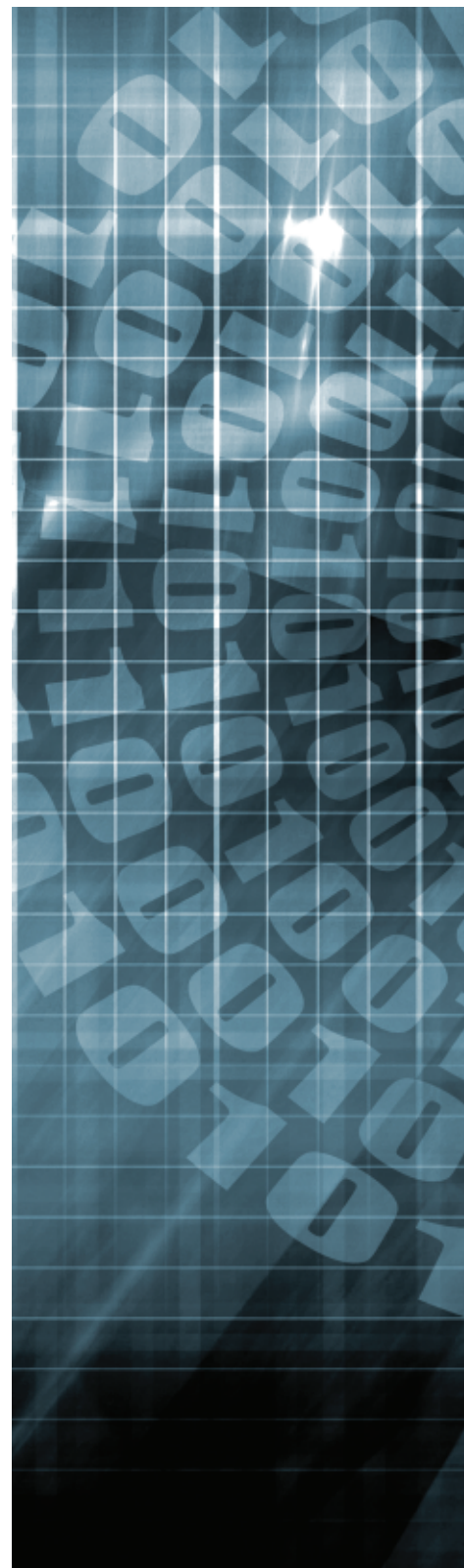
Global Users

The user environment including workspace, notifications, assignments, project lists and so on is managed through a user's home server so that user information is not duplicated. All user accounts are fully replicated to ensure uniqueness. As users logon to any Content Server server in the virtual environment their home server address is resolved and they are automatically redirected to their personal work-space or home server location. Administration functions allow home server locations to be relocated for users moving between sites in the virtual Content Server environment.

Flexibility and Control

Distribution agreements between locations are not limited to on/off. These agreements are also used to determine the form of replication for the selected container. Three levels of content distribution are provided:

- 1) **Shell** – This level will only distribute a copy of the selected project or folder itself to the targeted location(s). It is typically used to ensure sites are aware of the existence of projects and to make them visible and searchable. However, when a user selects the distributed shell object, they are redirected to its home server so that all child content can be accessed according to their permission rights.



Distributed Content Server | THE SOLUTION

Flexibility and Control continued

2) **Sub-item** – This level is used to unlock the real power of the Distributed Content Server environment. Projects and folders are replicated with all child content including task lists, workflows, categories, discussions, news, sub-folders and sub-projects. This does not however, include the actual document files themselves. Documents are only replicated as a type of shadow object.

A shadow document works like a Content Server alias: it marks the place of the real document across the virtual Content Server environment, knows where to find it and pretends to be it, but it uses up only a fraction of the space. Whenever a real document is created or amended, the original is stored locally and what is updated across the distributed network is its shadow.

Where users wish to view documents they have access like any normal Content Server document based upon its permissions. An efficient XML file transfer between servers is used to action document fetch requests. Document views generate HTML renditions from original documents on home servers minimizing bandwidth needs for document access and transfer.

3) **Sub-items and Content** – This level is used to enable full replication of all content within the project or folder including the document files. This is typically used to provide local speed access to content that is reasonably static or as a methodology for disaster recovery and redundancy.

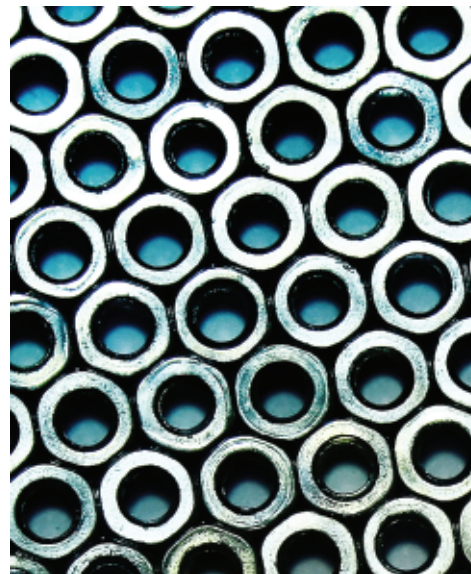
Enterprise Searching

With the distribution of Content Server servers comes the potential for fragmentation of the search indexes. This problem has been addressed in two ways: Administrators can choose between indexing all available content that is distributed to a location, or just the content that is held locally. In addition a federated search engine is provided to allow enterprise-wide searches to be performed efficiently, directly from within the standard advance search screens. Snapshots and queries are also available in this federated mode.

Modular approach

The Distributed Content Server solution is designed to integrate with Content Server and support all standard objects including Content Server optional modules conforming to standard development strategies. The solution creates a virtual Content Server environment and is achieved through using standard Content Server module functionality and a Content Server server communication application.

The strategy has been implemented to minimised cost to both hardware and administration while maximising performance in different business models. We believe that this approach also supports compatibility with new Content Server modules and customisations conforming to standard Content Server development strategies.



Distributed Content Server | THE SOLUTION

Administration

Any environment conflicts are reported through an administration layer that allows resolution. A reconciliation engine has been added to facilitate disaster recovery scenarios. Administrators are provided with a re-homing tool that allows migration of content from one server to another, or if required a distributed project can be rolled back to one location.

The approach outlined above allows maximization of limited bandwidth, while still retaining all the advantages of a Content Server as an Enterprise Document Management and Collaboration System. It gives users an enterprise-wide, fully collaborative environment at local LAN speed.

BENEFITS

- Supports all standard Content Server functions.
- Enables a Content Server project portal to provide a fully synthesised view of a multi-site project.
- Lets users navigate the Enterprise Document Management System at LAN speed.
- Stores 80-90% of the documents users regularly need to access at the local site.
- Provides a lightweight HTML view option for quick previewing of distributed documents.
- Offers an efficient server side compressed XML transfer to optimise downloading remote documents in full.
- Allows users to perform federated searches from their local servers, across the entire enterprise or to limit a search to specific locations of their choice.
- Requires only one sign-on for access across the enterprise.
- Gives the users the capability to replicate their project to other sites as required.
- Maintains data integrity by allocating a business-wide unique identifier to every document.
- Replicate selective or the entire content of the system
- Create read-only replicas
- Compressed payload transfers to minimize network traffic
- Avoid a single point of catastrophic failure
- Hot-backup of the entire contents of a production server for failover
- For partner access, physically separate secure partitioning of selected Content Server® information
- Supports all base Content Server® functions, modules and customizations
- Push common folders and documents to regional Content Server® servers

SPECIAL FEATURES

- Integrates third party and OEM modules without any additional customization and enables every site to run them at LAN speed.
- Allows easy project archiving by enabling re-homing of distributed content to a single selected location.
- Facilitates disaster recovery procedures by enabling full content replication as required.
- Provides faster extranet access for clients.

Synergy, Inc. has offices located in La Jolla, CA, USA | Vaughan, ON, Canada

For inquiries: Phone: 858-964-3243 • Fax: 858-729-0429 • Email: info@synergy.com • www.synergy.com