

Using EMC Documentum to Improve Content Responsiveness in Distributed Environments

Applied Technology

Abstract

In a global, distributed environment, constricted bandwidth and network latency can significantly undermine business efficiency. The EMC[®] Documentum[®] platform supports a modern caching solution that optimizes content access and responsiveness for geographically dispersed task teams.

January 2008

Copyright © 2008 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

Part number H3359

Table of Contents

Executive summary	4
Introduction	4
Audience	5
Documentum solutions for optimizing content responsiveness.....	5
The foundation: The Documentum repository	5
EMC Documentum Branch Office Caching Services	6
Proactively push content to remote offices	7
Control IT costs with centralized deployment, monitoring, and administration	8
Read <i>and</i> write quickly through the cache from remote locations	8
A word about remote content servers	8
Conclusion	9

Executive summary

The prospect of poor WAN performance—unpredictable or slow data transfer times across vital wide area networks—has given pause to many organizations seeking to leverage the benefits of their content management systems enterprise-wide.

Several factors affect content responsiveness:

- Bandwidth
- Network latency
- File size
- Frequency of remote fetches and updates

These mechanical challenges impact business. Access delays cause engineers to miss project deadlines. Business units may embark on new sourcing initiatives to reduce operational costs and accelerate product innovations, but executives and line managers alike are unhappy with the missed deadlines and IT's inability to fully support the company's game-changing strategies.

Say, for example, an engineering team in New York City wishes to collaborate over an extranet with designers in Tokyo, sharing frequent product specification and design revisions with their Japanese colleagues.

Rather than immediately accessing the latest document versions, the remote design teams must wait for the documents to download from the central file store, which is located in the New York corporate data center. American engineers, in turn, cannot rapidly review, revise, and approve design changes made by the Japanese designers. The teams are thwarted by the “mechanical” limitations of content access.

Furthermore, because file transfer bandwidth and network latency vary widely between New York City and Tokyo, accessing, fetching, and uploading large files across the network is unpredictable. And local IT administrators are unsure how to best support the individual design teams.

IT organizations can regain the upper hand by using a distributed repository that flexibly determines how content is accessed and stored across multiple servers and systems within an enterprise. When content files are hosted at multiple network locations, closer to the end user, the impact of network latency is mitigated. Network connections automatically transfer content among servers, rapidly delivering files where needed. In short, a distributed repository addresses the key factors that adversely impact content responsiveness across a global enterprise.

Introduction

This white paper describes the way many organizations use the EMC® Documentum® enterprise content management platform to achieve the goals of centralized control with local access and responsiveness. Documentum allows these organizations to manage content systematically and ensure enterprise-wide security and access controls, consistent categorization, and end-to-end business process management capabilities.

This white paper also explains how EMC Documentum Branch Office Caching Services can help IT ensure every local experience includes the best possible content access and delivery. The paper shows how organizations can use BOCS to host content files at multiple network locations, putting them closer to the end users and mitigating the impact of network latency.

Audience

This white paper is intended for IT administrators in organizations with remote users, particularly those in branch offices, who access content stored in a central repository, typically on a server at headquarters. The subject matter of this paper is particularly relevant to organizations with users who are dispersed throughout a large region or across the world, and where improving the speed and efficiency of information collaboration and production across their enterprise has become an important objective.

Documentum solutions for optimizing content responsiveness

The Documentum platform supports a solution that optimizes global content access and ensures content responsiveness for distributed task teams.

The foundation: The Documentum repository

The Documentum repository comprises the following:

- Metadata—stored in a relational database
- Content files—usually stored in the file system

Repository content is accessed through EMC Documentum Content Server, which uses its many features to enhance content transfers to remote locations while managing overall access to the metadata and the content files. EMC Documentum WDK/Webtop Server, a J2EE-based web application framework, provides an application interface to the metadata and content files.

The Documentum repository is often hosted at a single location. Multiple workgroups within a global enterprise then connect over the network to access and retrieve content.

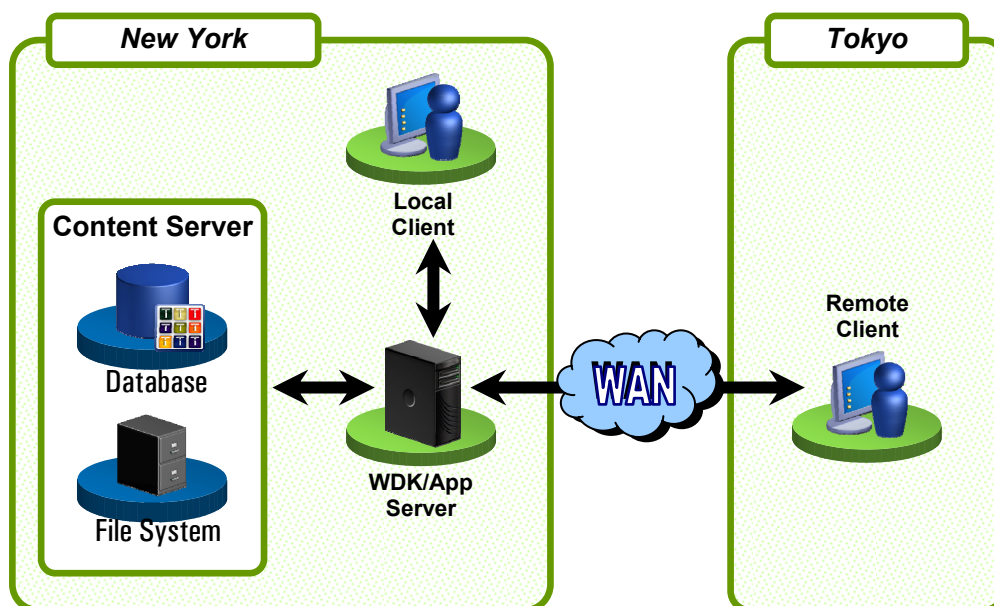


Figure 1. EMC Documentum Web Development Kit/Webtop Server provides an application interface to the metadata and content files.

EMC Documentum Branch Office Caching Services

EMC Documentum Branch Office Caching Services provide distributed workgroups with local access to content—even in business units that do not have a local Documentum Content Server. BOCS supports a single, enterprise-wide view of all content, dynamically caching and storing content, on demand, on a local file system located within a branch office. The metadata is stored and managed centrally. Because such metadata is very small, typically only a kilobyte or two, the WAN latency involved in its transfer is insignificant.

BOCS functions as an information appliance, running on an organization's existing hardware and software and requiring virtually zero IT administrative effort to move needed content close to users. BOCS systems are also self-cleaning and virtually maintenance-free: Automated housekeeping procedures periodically cull old content based on age or cache size; these parameters can be specified upon installation and configured remotely through Documentum Administrator.

It should be noted that BOCS cached content “behaves” the same as all other repository content. That is because the cached content's metadata, permissions, and entitlements still apply, exactly as they would if the content came directly from the repository.

To access the content, a user queries the database at the central location to determine the metadata. The EMC Documentum WDK/Webtop Server detects the user's network location and redirects the request to a local file server.

BOCS then delivers content by:

- Determining whether the content currently resides on the local file system
- Automatically fetching the content from an adjacent content server and caching the file in the local file system (if the content does not already reside on the local file system)
- Presenting the content to users through the Documentum Webtop user interface

To continue our use-case scenario, let's suppose the designers in Tokyo open a branch office in Osaka, and need to ensure their engineers can rapidly access large design documents. When a Japanese engineer requests a document, it is automatically transferred from the content server in Tokyo and stored in a local cache running on a server in Osaka. The next time someone in the Osaka office requests the document again, it is automatically fetched from the local cache on the server.

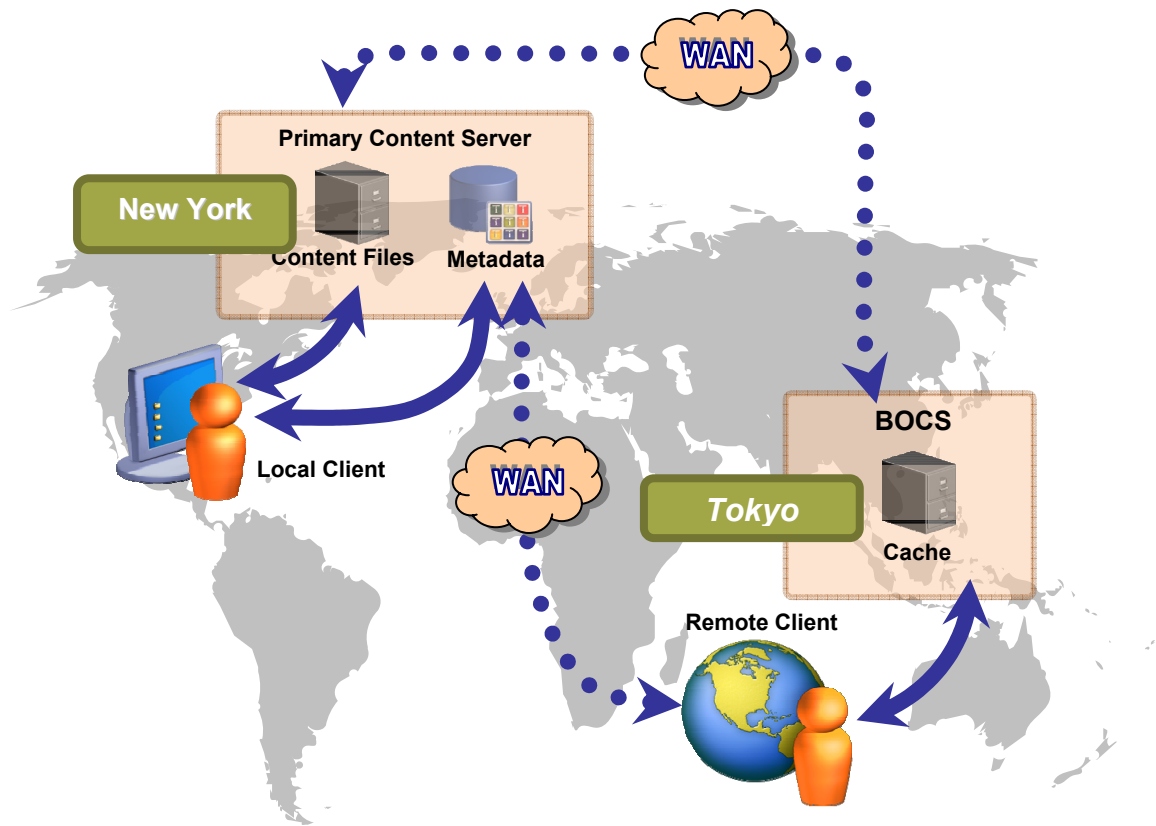


Figure 2. BOCS content files are cached remotely while the metadata resides on the primary content server.

Proactively push content to remote offices

With BOCS, the first user requesting content from a remote location may experience a fetching delay due to the same latency issues and bandwidth constraints affecting other network users. But in many cases, such first-time fetching delays can be eliminated thanks to the BOCS system’s precaching capability.

Precaching enables organizations to “push” predetermined collections of content files from a content server to the BOCS cache. Such content can be precached based on any query that can be captured in the metadata—for example, BOCS can query for evidence of particular events, such as a change in workflow status or lifecycle state, or for content in a specific folder, content modified within a certain date range, content created by a particular author, and so on. The precaching process can be initiated as frequently as every few seconds or, if preferred, it can be regularly scheduled for off-hours, when the networks are little used. The content can be pushed to multiple BOCS machines or only specific remote machines, and the process can be implemented by an administrator or an application; that is, an application built on top of WDK can programmatically execute a query to precache content.

The BOCS system can be easily configured, for example, to push the Tokyo team’s latest design files to the local Osaka server whenever the team lead in Tokyo posts a new approved version.

The result is greater content responsiveness for all users who initially request content files from across the network.

Control IT costs with centralized deployment, monitoring, and administration

Since BOCS can be deployed, monitored, and supported from a single, central location, organizations can increase remote user performance without the cost of dispersed, onsite IT support.

BOCS servers are lightweight Java applications, which can be deployed and undeployed remotely. Just as important, BOCS can be deployed in heterogeneous environments; that is, the BOCS cache can be hosted on any type of supported server, whether or not it matches the server on which the corresponding content server is installed. BOCS can also be configured and monitored remotely using Documentum Administrator, which takes advantage of the new EMC Documentum platform unified JMX framework.

Because there is no replicated content server or database to maintain, there is no need for onsite IT or other administrative support. Everything can be easily handled from a central location. With BOCS, the metadata (as well as permissions and entitlements) is accessed from the content server through the WDK on the application server, enabling administrators to maintain central control over all content.

Read *and* write quickly through the cache from remote locations

In addition to its distributed read capabilities, BOCS handles distributed write duties. Users can write back to the central repository from cache using two different modes: synchronous and asynchronous.

In synchronous mode, content written to BOCS by remote users is immediately saved to the central or primary repository, and thus, becomes immediately available for use by others; remote users will gain immediate access to the content from the nearby cache. At the same time, the associated metadata is written to the database through the central WDK server. The obvious advantage of synchronous write is, once the content transfer is complete, the new content is immediately available to anyone accessing the primary content server; additionally, the content is cached at the remote site to provide remote users with quick access as well.

The synchronous mode user, however, gives up control of the application until the writing process has been completed. This is still better than writing directly to the content server as the content is also cached on the BOCS machine and available for quick download by anyone accessing that remote cache.

In asynchronous mode, content written to BOCS is saved to the central repository *after* the user has been given back application control. The repository upload can begin immediately after content has been uploaded to the cache, or it can be scheduled as appropriate—for example, late at night when network bandwidth is generally more available—by new Documentum Messaging Services (DMS). Remote users can write content to BOCS, which can then optimize network traffic by writing all new content to the primary repository at a later time. The advantage of asynchronous write is it provides the best performance for remote users, who upload locally and immediately get back application control. Content is also immediately available to other remote cache users, whether or not the content has been transferred to the primary repository.

In the asynchronous mode, however, the new content will not be available to others on the primary repository until the upload is scheduled and completed by DMS.

In both modes, metadata is immediately written to the database via WDK, where it is indexed and can be searched. All permissions and entitlements apply and are enforced on cached content.

A word about remote content servers

Prior to the release of Branch Office Caching Services, many organizations chose to use remote content servers (RCS) to replicate the entire content of their primary repository to remote locations. While RCS provided these organizations with remote site responsiveness, the lack of IT support and hardware resources that hamper remote locations usually made it impractical to deploy RCS where it was needed.

BOCS generally offers all the RCS functionality with substantially lower costs. Organizations currently running remote content servers, however, may choose to leave them in place, or implement BOCS around them, depending on the situation.

An organization wishing to maintain its RCS implementation can still gain the BOCS benefits for WDK-based clients simply by adding BOCS nodes. BOCS is network-aware and will automatically download and upload content to the nearest content server, whether it's a remote content server or the primary content server. Adding BOCS to an existing RCS system in a hub-and-spoke model enables organizations to quickly extend the distributed environment. Remember, no onsite IT staff is required for its installation or configuration, and BOCS runs on existing hardware and software.

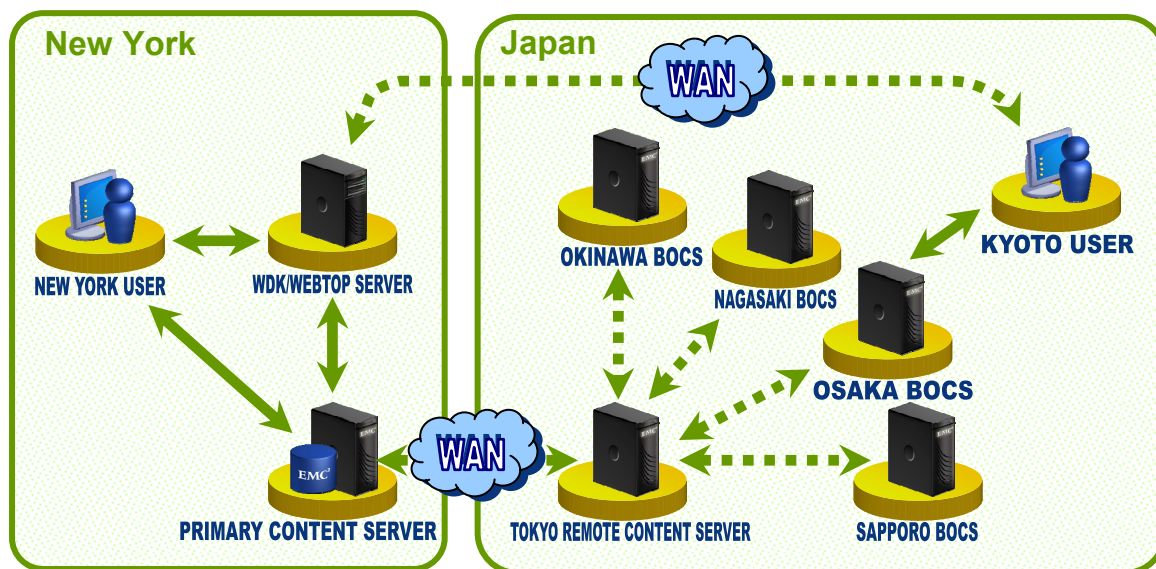


Figure 3. BOCS in a hub-and-spoke configuration with RCS.

For organizations using only WDK-based clients with remote content servers, the recommended approach is to undeploy RCS and replace it with BOCS. BOCS provides the functionality of RCS but with a smaller resource footprint. BOCS can also be deployed at locations where a lack of onsite IT support prevented using RCS.

Conclusion

In today's global business environment, performance matters more than ever. The success of any enterprise content management system depends on remote users getting the performance they need when working with content from far away. Branch Office Caching Services offer a lightweight, cost-effective, and easily supported solution. BOCS can extend content management to any location, regardless of its size or distance from headquarters, while sidestepping the effects of WAN latency. The result is fewer information silos and increased employee productivity.