White Paper

EMC® Documentum® Interactive Delivery Services Accelerated Overview

A Detailed Review

Abstract

This white paper presents an overview of EMC Documentum Interactive Delivery Services Accelerated (IDSx). IDSx publishes content directly from a Documentum repository to a target machine or to multiple server machines. IDSx can provide single-source publishing support using secure and efficient file transfer technology over any IP network.

July 2011

EMC²

Table of Contents

Executive summary

This paper provides an insight into the functionality of IDSx: How IDSx can be used to replicate content residing in the central Documentum repository, to multiple servers on a server farm. It describes how users can achieve replication in a secure, fast, and transactional manner. The paper also outlines how user-generated content on multiple servers (in a server farm) can be brought back and persisted into the central repository.

In several cases, large websites must be deployed on multiple servers to achieve a higher throughput and high availability in a load balanced environment.

At the same time, content served by these websites may be managed in a central content management repository using content management tools such as Documentum Web Publisher, Documentum Page Builder, Documentum Webtop, Documentum CenterStage, Documentum Digital Asset Manager, Documentum Media WorkSpace, and so on. This scenario identifies an implicit need for a tool that can publish content on demand or on schedule from a central repository to multiple servers.

In addition, with Web 2.0, user-generated content such as blogs, wikis, and feedback that is captured by websites for a specific content, must be sent to the content author associated with appropriate content and persisted in the central content management repository.

This paper provides an insight into how users can use IDSx to handle such use cases.

Audience

This white paper is intended for Content Management solution Architects, System Administrators, System Integrators, and others involved in evaluating, acquiring, managing, operating, and designing an EMC Documentum Content Management Solution.

Introduction

EMC Documentum Interactive Delivery Services Accelerated (IDSx) can publish content directly from a Documentum repository to a target machine, or to multiple server machines. IDSx can provide single-source publishing support using secure and efficient file transfer technology over any IP network.

This paper describes the following capabilities of IDSx:

- **Replication Manager:** This capability caters to replicating content residing in Documentum repository to multiple servers.
- **Ingestion Manager:** This capability caters to pulling user-generated content from multiple servers into the Documentum repository.

The IDSx Source and IDSx Target embed a UDP-based accelerated file transfer technology from a third-party. This technology helps to achieve high transfer rates

from the IDSx source to the staging target, and from the staging target to the replication target, with high latency and high packet loss networks.

IDSx exposes configurable parameters that can be used to change the various parameters of this technology based on customer environment. For example, transfer policy, minimum or maximum transfer rates, algorithm to check file checksum, and so on.

This technology is used to push data to the staging target or replication target, and also helps to pull data from these nodes back to the IDSx source.

Terminology

IDSx

An advanced version of IDS [formerly Site Caching Services (SCS)]. Its architecture is the same as the architecture of SCS. IDSx utilizes SCS concepts such as configuration management in Documentum Administrator, Source/Target combination, and so on, from the usability perspective.

IDSx Source

A web application hosted on a JBOSS instance and is co-located with Content Server. The IDSx source is primarily responsible for accepting requests related to publishing and replication, exporting data to a temporary location, and transferring exported data to the IDSx target using embedded UDP-based accelerated file transfer technology.

IDSx Staging Target/IDSx Replication Target

A stand-alone Java program running on a specified port. It accepts various requests from IDSx source such as login, file transfer commit, and so on. It also applies publish batch to the target directory and synchronized the target database.

Publish Batch

Output of a publish operation from an IDSx source to an IDSx target. It contains the new or modified content and its corresponding metadata.
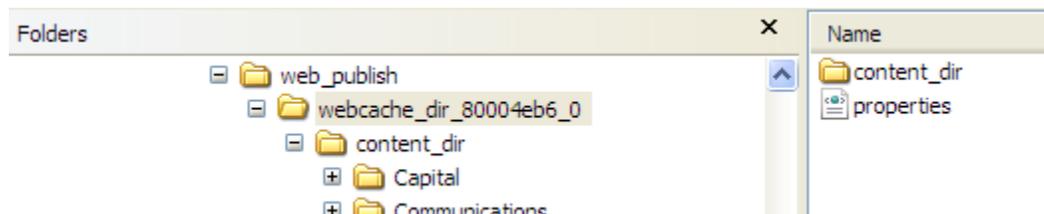

Figure 1. Publish Batch

Replication Batch

Output of combining multiple publish batch (generated during multiple publish operations) into one. **Figure 2** depicts how a typical Replication Batch looks.
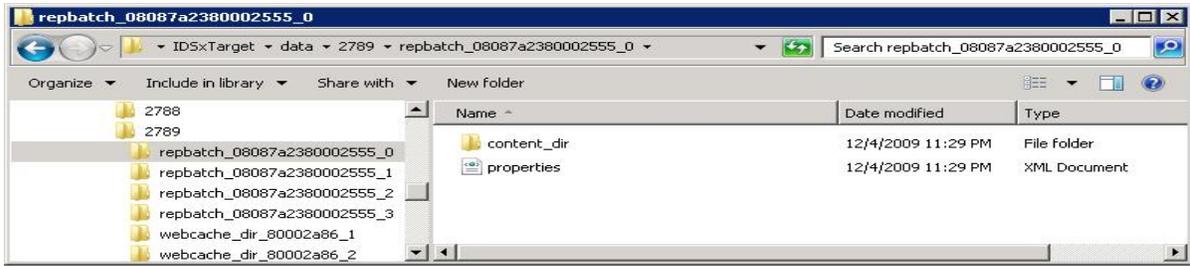
Figure 2. Replication Batch

Replication Manager

Replication Manager (RM) runs on the IDSx staging target, and is responsible for keeping target directories and metadata at IDSx replication target(s) in sync. The Replication Manager combines multiple publish batches into one replication batch on the IDSx staging target, and sends across the replication batch to IDSx replication targets to invoke synchronize operation. The Replication Manager is invoked by the IDSx Source when it receives a replicate request for an Interactive Delivery Services configuration.

Hot Folder or Target Ingest Directory

A configurable folder containing content that must be ingested based on IDS configuration.

Ingestion Batch

Any file(s) or folder(s) copied into the hot folder on a target machine is rolled up in the ingestion batch before it is pulled back to the IDSx source. A sample extract for an ingestion batch is given here:

```
ingest_batch_<configuration ObjectId>_<timestamp>     (directory)

- <Replication Target One Object Id> (directory)

- <Files and/or folders to be ingested>

- <Replication Target Two Object Id> (directory)

- <Files and/or folders to be ingested>

.

.

.

- <Replication Target Nth Object Id> (directory)

- <Files and/or folders to be ingested>
```

Ingestion Manager

The Ingestion Manager (IM) pulls data from multiple replication targets into a single ingestion batch on the staging target, which in turn is pulled in to the IDSx source. For example, if you have four replication targets, the data is pulled from all the four replication targets into a single ingestion batch on the staging target. This single

ingestion batch is then pushed to the IDSx source. The Ingestion Manager is a part of the IDSx target and is invoked by the IDSx source when it receives an Ingest request for an Interactive Delivery Services configuration.

Replication Manager

The Replication Manager helps to replicate content residing in the Documentum repository, to multiple servers.

The replication process is centrally managed using the Documentum Administrator client. Users can invoke replication on-demand or on schedule (using Content Server Jobs).

Although replication is a two-step process, the replication process is seamless:

1. Content is published to an IDSx staging target.

2. The Replication Manager replicates the content from the staging target to multiple IDSx replication target(s) in a transactional manner. To replicate content in a transactional manner, the size of content to replicate must be less than a configurable value. This rule is relevant whether data is applied to all replication targets or not. Based on the hardware deployed in your environment, it is recommended that you decide an appropriate content size value for transaction feature.
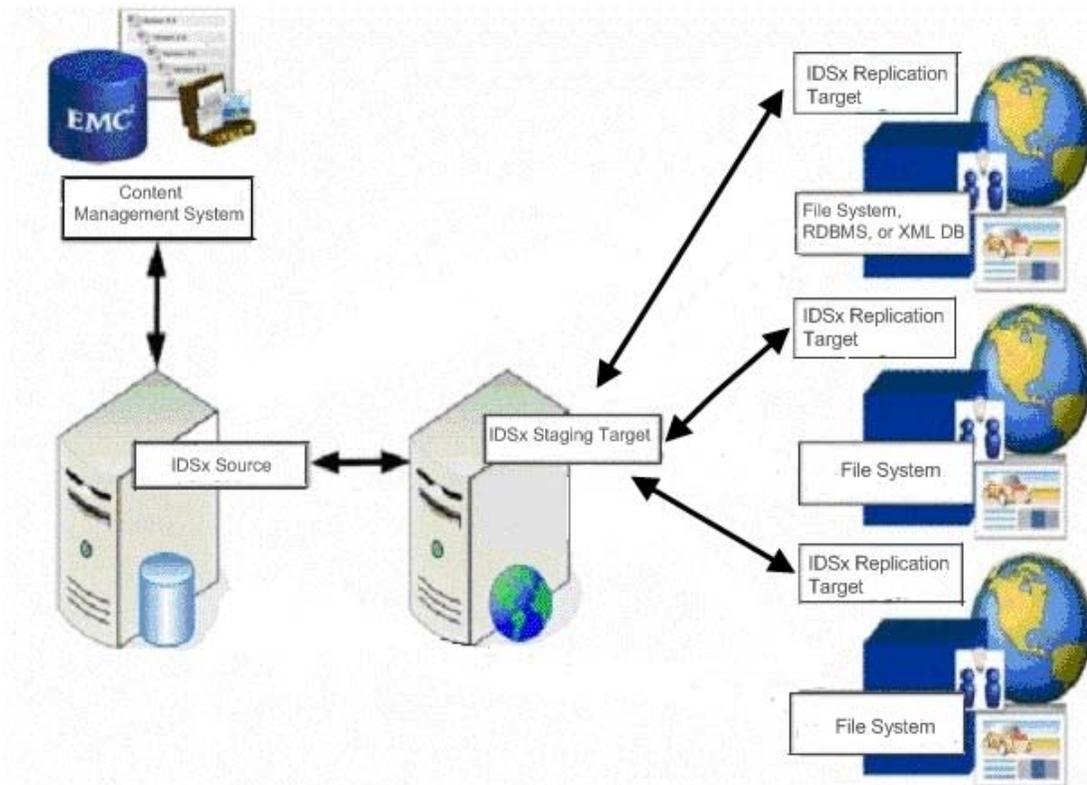

Figure 3. Replication setup in a production environment

Figure 3 depicts a typical replication configuration setup in a production environment. This configuration has been implemented by one of the largest financial companies in North America. The customer has deployed IDSx in the production environment with 14 replication targets and transactional capabilities enabled.

Ingestion Manager

The Ingestion Manager helps to ingest user-generated content residing on multiple servers, back to the Documentum repository.

The ingestion process can be managed centrally using the Documentum Administrator client. Users can invoke ingestion on demand or on schedule (using Content Server Jobs).

The process of ingestion involves the following steps:

1. The Ingestion Manager pulls user-generated content from IDSx replication targets to the IDSx staging target.

2. The user-generated content is then pulled from the staging target to the IDSx source and persisted into the repository using the following out-of-the box persistence handlers:

   - **Import Handler:** To create new content in the repository
   - **Check-in Handler:** To create a new version of the object
   - **Update Metadata Handler:** To update metadata for an existing object
   - **Rendition Handler:** To create a new rendition of the object
   - **Metric Handler:** To update metrics for a SysObject

Besides these out-of-the box handlers, the ingestion process offers an option to configure multiple custom persistence handlers that can be used to address the requirements of use cases that are specific to customer requirements.

Conclusion

This paper covers information about using the IDSx replication capability to replicate content from the central Documentum repository to multiple replication targets. This paper also describes how the IDSx ingestion capability can pull user-generated content such as blogs, wikis, and feedback, from the replication targets, to the Documentum repository. This forms a complete 360° round-trip of data using the new IDSx capabilities.

References

For more information about Documentum Interactive Delivery Services Accelerated, see the following documents on Powerlink:

- EMC Documentum Interactive Delivery Services Accelerated 6.5 SP1 User Guide
- EMC Documentum Interactive Delivery Services Accelerated 6.5 SP1 Installation Guide
- EMC Documentum Interactive Delivery Services Accelerated 6.5 SP2 Release Notes