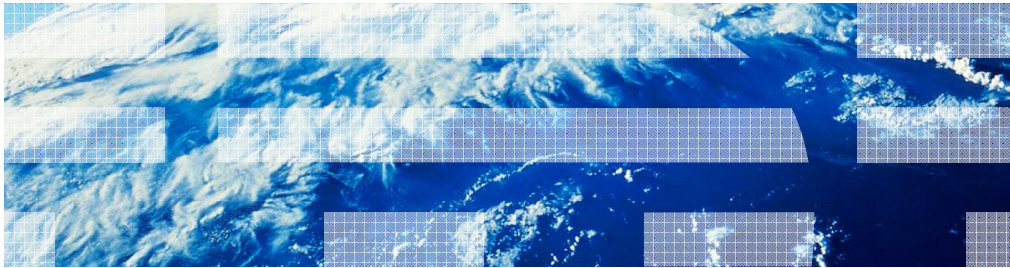


# IBM WebSphere Commerce V7 Feature Pack 2

## Web feed integration



This presentation provides an overview of the Web Feed Integration to the WebSphere Commerce V7.0 Feature Pack 2.

## Agenda

- Concepts
  - Content
  - Assets tool
- Problem areas before the Feature Pack 2 and the new solution provided in Feature Pack 2
- Solution architecture
- Configuration files
- Running the solution

This presentation first introduces two web feed integration related concepts: content and Assets tool. Then some issues regarding the content integration with WebSphere Commerce before the Feature Pack 2 are pointed out and the new content integration solution provided by Feature Pack 2 is introduced. The web feed integration solution architecture and the configuration files required by each integration scenario are covered. The presentation ends with an introduction on how to run the solution.

## Structured content versus unstructured content

- Structured content
  - Refer to information that has been classified using XML
  - The information stored in structured content is organized and stored in the WebSphere Commerce database
  
- Unstructured content
  - Refer to the static content like image files, URL and other multimedia files
  
- Web Feed Integration is about unstructured content

Web content is classified as structured content and unstructured content. Structured content refers to information that has been classified using XML. In contrast to the structured content, unstructured content refers to the static content like HTML, images and other multimedia files. The Web Feed Integration solution in Feature Pack 2 is about unstructured content integration.

## Content management in Assets tool

- Unstructured content is managed by the Assets tool as managed files and attachments
- Managed files
  - Written to the database and copied from the database to the WebSphere Commerce EAR file
- Attachments
  - A collection of one or more files, URLs that are associated to a catalog or marketing asset

The screenshot displays two windows from the IBM Assets tool. The top window, titled '\*Manual.doc', shows a 'References' tab with fields for File (Manual.doc), Name (Manual.doc), Size (1 KB), and Path (/Madisons/Attachment). The bottom window, titled '\*Attachment-demo', shows a 'Manage Attachment' tab with a 'Reference' sub-tab. It includes a 'General Attachment Information' section with fields for Code (Attachment-demo), Name (attachment-demo), Short Description, and Long Description. Below this is a 'Files or URLs' section with a table:

Files or URLs	Type	Path	Language (United St...)	Language (b
		/Madisons/Attachment/Manual.doc		

At the bottom of the attachment window, it indicates '0 of 1 selected'. The footer of the interface includes 'Web feed integration' and '© 2010 IBM Corporation'.

In WebSphere Commerce, unstructured content is managed by Management Center Assets tool as managed files and attachments.

After you upload the managed files into the Assets tool, the files are initially written to the WebSphere Commerce database. Once certain criteria is met, managed files are copied from the WebSphere Commerce database to the WebSphere Commerce EAR file.

The attachment in WebSphere Commerce is a collection of one or more files and URLs that are associated to a catalog or marketing asset.

## Problem areas

- WebSphere Commerce consumes the unstructured content in 3 ways:
  - Managed files are served by Commerce server
    - Access managed files through attachments
  - Contents are served by external system
    - Access external URLs through attachment
  - Contents are served by web server
    - Access the contents through Commerce web server
- Integration is needed if unstructured content is managed by a vendor's Content Management System (CMS)
  - Synchronization of files and URLs is needed between the CMS and WebSphere Commerce
- Before Feature Pack 2, WebSphere Commerce did not provide integration code

WebSphere Commerce consumes the unstructured content in 3 ways. The first way, the managed files are served by WebSphere Commerce server. You create attachments with the managed file asset and you associate the attachments with Commerce objects. The second way, the content is served by an external system. You create attachments with external content URL assets and you associate the attachment with Commerce objects. The third way, the content is served by WebSphere Commerce web server. You can also create an URL attachment and associate with other WebSphere Commerce objects.

Integration is needed if unstructured content is managed by a vendor's content management system (CMS), the contents in the CMS must be synchronized with the ones in WebSphere Commerce environment.

Before Feature Pack 2, WebSphere Commerce did not provide integration code. If you use the vendor's Content Management product to manage the static content, you need to write your own integration code to integrate CMS with WebSphere Commerce.

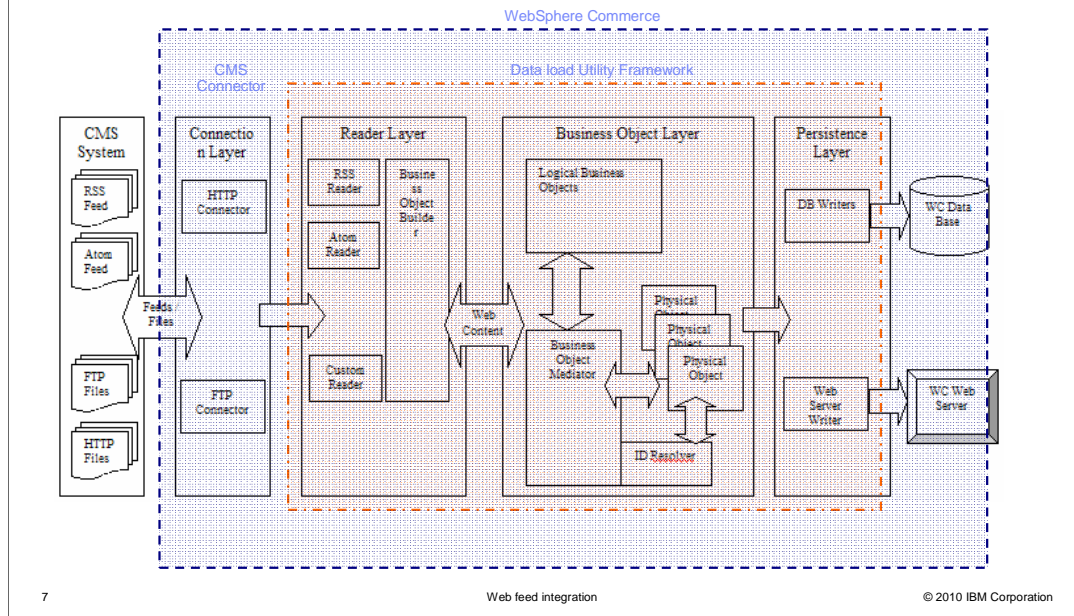
## Solution provided in Feature Pack 2

- Utility is provided to interact with any vendor's CMS with RSS/ATOM feed support
  - Solution provides loose coupling that allows any CMS system to be integrated
  - Retrieve URL contents from feed and create URL attachments in WebSphere Commerce
  - Retrieve file contents from feed, save the files into WebSphere Commerce database as managed files and create file attachments for these files
  - Retrieve file from CMS system and copy or FTP file to WebSphere Commerce web server or a specified location

The Web Feed Integration solution interacts with vendor's CMS in a loosely coupled mode. It integrates with the CMS through the configurations, you do not need to write any specific vendor's platform interface code. This solution supports any vendor's CMS that can provide the contents as ATOM/RSS feeds.

A utility is provided to allow you to integrate with CMS in three ways. You can receive content as an URL to an external location where your content is managed. An attachment with the URL is then created for business users to reference. You can receive the content as a file. The file is saved into WebSphere Commerce as a managed file. An attachment is created for the file. Finally, you can receive a file from a CMS feed and then copy or FTP the file to the WebSphere Commerce web server. An URL attachment for the file is created for you during the process.

## Interaction architecture



The Web Feed Integration solution is based on the existing Data Load framework which was developed in Version 7. A connector layer is added on top of the Data Load framework to handle the connection between the external CMS system and the Commerce Data Load Utility Framework.

The Connector connects to the CMS system to read the RSS/ATOM feeds. It filters the already updated **entries** from the **feed** and passes on the updated or newly available **entries** to reader in Data Load. The Data Load Reader parses the feed and retrieves the content from the URL location specified in the feed. The content is then loaded into the WebSphere Commerce database using the Database Writer or the content is saved into a folder that the web server can access using the **Web Server Writer**. The writers also create the attachment which references the content received.

## Connector

- Actions performed by Connector
  - Read the web feed configuration file
  - Generate Data Load configuration files required by the Data Load framework
    - For example: wc-dataload-config.xml and wc-loader-managedFile.xml
  - Connect to the content location using the appropriate protocol and authentication
- Supported connection
  - HTTP Connector (HTTP and HTTPs)
  - FTP Connector (secure FTP is not supported)
- Authentication
  - Basic authentication (ID and password)
    - Support both HTTP and HTTPs
    - FTP
  - Client certificate authentication
    - Support HTTPs only

The Connector performs these actions when you run the web feed utility. It first reads the web feed configuration file where you specify the feed end points, authentication information and the kind of attachments that need to be created. In the CMS feed configuration file, you can specify whether data load configuration files need to be generated. These Data Load configuration files are required by the Data Load framework. If Data Load configuration generation is enabled, the data load configuration files are generated on the fly. The Connector connects to the content location using appropriate protocol and authentication.

The connection supports FTP and HTTP to retrieve files from CMS. If the entries in the feed point to a FTP URL from where content needs to be fetched, the FTP connector is used. Otherwise the HTTP connector is used.

The connection supports basic authentication over HTTP, HTTPs and FTP. To use the basic authentication, you need to provide the user ID and password in the web feed configuration file. The connection also supports client certificate authentication. To use this authentication, in the web feed configuration file, you need to provide the client's certificate location and client certificate's password.



## Data Load Readers

- Actions performed by Data Load Reader
  - Filter the already updated entries from the feed and pass on the updated or newly available entries to reader in Data Load layer
  - Add log into database table **cmfeedlog**

### Supported Reader types

- RSS Reader
  - Parse and process feeds that comply with the **RSS version 2.0** standard
- ATOM Reader
  - Parse and process feeds that comply with the **Atom 1.0** standard
- Write your own Data Load Reader if your web feed is not supported

The Data Load Reader will perform actions when you run the web feed utility. The Reader reads a feed entry from the feed and checks for updates. It compares the feed entries with the feed log table **cmfeedlog**. If an update or new entry is found, the Reader logs that entry in the table **cmfeedlog**. The Reader reads data from the feed. The returned data is passed to the business object builder to build business objects.

The Feature Pack 2 provides two web feed readers. The RSS Reader helps parse and process feeds that comply with the RSS 2.0 standard. The ATOM Reader helps parse and process feeds that comply with the ATOM 1.0 standard. If your web feed is not supported, you need to write your own Data Load Reader.

## Data Load Writer

- Actions performed by Data Load Writer
  - Write the physical objects generated by the Business Object Mediators to the final destination location
  
- Supported Writer types
  - DB Writer – For Managed Files or URL scenario
    - Write the physical objects to the database
  
  - Web Server Writer – For Static Content (web server) scenario
    - Copy content to local file system
    - FTP to a web server on a remote machine

There are two writers, namely DB Writer and web Server Writer. The writers write the physical objects generated by the mediators to the final destination location. If you want to load the web feed contents as Managed Files or URLs, the DB writer is used. The contents are loaded into the WebSphere Commerce database. If you want to load the web feed contents to WebSphere Commerce web server, the web Server Writer is called to either copy files to a local directory or FTP to the web server on a remote machine.

## Web feed sample

- A sample ATOM web feed with only one entry in the feed

```

- <feed>
  <title>Announcements List</title>
  <link href="http://aimcp211.austin.ibm.com:10040/wps/wcm/connect/Intranet+Web+Content+6.1.0/IntranetSite/Home/Announcements/AnnouncementsList/">
  <updated>2008-04-06T01:06:16Z</updated>
  - <author>
    <name>uid=xyzadmin,o=defaultWIMFileBasedRealm</name>
  - <author>
    http://aimcp211.austin.ibm.com:10040/wps/wcm/connect/Intranet+Web+Content+6.1.0/IntranetSite/Home/Announcements/AnnouncementsList
  </id>
  <generator uri="http://www.ibm.com/software/lotus/products/webcontentmanagement">Lotus Web Content Management</generator>
- <entry>
  <title type="text">TestContent</title>
  <link rel="alternate" type="text/html" href="http://aimcp211.austin.ibm.com:10040/wps/wcm/connect/Intranet+Web+Content+6.1.0/IntranetSite/Home/Announcements/TestContent/">
  <id>75b9028043af19828522cf0fd5b6ed91</id>
  <published>2010-23-08T01:10:56Z</published>
  <updated>2010-05-09T01:04:29Z</updated>
  <summary>
  <content type="image/jpeg" src="http://aimcp211.austin.ibm.com:10040/wps/wcm/connect/dc68198043d87f61a67cae2c1f94b5af/community.jpg?MOD=AJPERES&amp;CONVERT_TO=url&amp;CACHEID=dc68198043d87f61a67cae2c1f94b5af"/>
  <category term="">
  - <author>
    <name>uid=xyzadmin,o=defaultWIMFileBasedRealm</name>
  - <author>
  </author>
  </entry>
</feed>

```

A sample ATOM web feed with one entry in the feed is shown here. An image file is provided as content in the feed.

## Managed file scenario - content received as managed file

- Utility retrieves the content files from the CMS system and stores them into the database as managed files
- Two assets are created in the Assets tool:
  - Managed file
  - File attachment

Sample feed configuration file:

```

- <config:CMSConfiguration xsi:schemaLocation="http://www.ibm.com/xmlns/prod/commerce/foundation/CMS/config CMSIntegration_config.xsd">
- <config:feedEndPoint>
- <config:content>
- <config:managedAsset>
  <config:name>ProductImages1</config:name>
- <config:feedURL>
  http://aunmcp211.austin.ibm.com:10040/wps/wcm/connect/Intranet+Web+Content+6.1.0/IntranetSite/Home/Announcements
  /AnnouncementsList?srv=cmprnt&source=library&cmprntname=Intranet+Web+Content+6.1.0/ATOM+Feed+Menu
</config:feedURL>
  <config:managedDirectory>/mytest/images</config:managedDirectory>
  <config:allowedTypes>*.jpg</config:allowedTypes>
  <config:tempFileLocation>C:/temp/tempFile/</config:tempFileLocation>
</config:managedAsset>
</config:content>
</config:feedEndPoint>
  <config:dataLoadConfigFile generate="true" envFileLocation="C:/IBM/WCDE_ENT70/samples/DataLoad/Catalog/wc-dataload-env.xml">C:/temp/wc-
  dataload-config.xml</config:dataLoadConfigFile>
</config:CMSConfiguration>

```

12

Web feed integration

© 2010 IBM Corporation

For the Managed file scenario, the feed received from the CMS system contains the URL to the content file. The web feed utility parses the feed and retrieves the content file from the CMS system and stores it in the WebSphere Commerce system database as a managed file. It also creates a file attachment in WebSphere Commerce.

To run the Managed file scenario, you need to provide a web feed configuration file to the utility. In the configuration file, you can specify into which managed directory your managed files are saved, content types you want to fetch and if the utility generates the Data Load configuration files for you or not. A sample of web feed configuration file is shown here.

## URL attachment scenario - content received as URL

- The utility retrieves the URLs from the feed
- The utility creates an URL attachment for each URL in the Assets tool
- Content life cycle is managed by CMS, WebSphere Commerce retrieves the content during the runtime

Sample feed configuration file:

```
-<config:CMSConfiguration xsi:schemaLocation="http://www.ibm.com/xmlns/prod/commerce/foundation/CMS/config CMSIntegration_config.xsd">
  -<config:feedEndPoint>
    -<config:content>
      -<config:URLAttachment>
        <config:name>ProductImages</config:name>
      -<config:feedURL>
        http://aimcp211.austin.ibm.com:10040/wps/wcm/connect/Intranet+Web+Content+6.1.0/IntranetSite/Home/Announcements
        /AnnouncementsList?srv=cmpnt&source=library&cmpntname=Intranet+Web+Content+6.1.0/ATOM+Feed+Menu
      </config:feedURL>
    </config:URLAttachment>
  </config:content>
</config:feedEndPoint>
<config:dataLoadConfigFile generate="true" envFileLocation="C:/IBM/WCDE_ENT70/samples/DataLoad/Catalog/wc-dataload-env.xml">C:/temp/wc-
dataload-config.xml</config:dataLoadConfigFile>
</config:CMSConfiguration>
```

For the URL attachment scenario, the utility creates an attachment in WebSphere Commerce for each URL received from the feed. The actual contents stay at locations external to the WebSphere Commerce and the life cycle of content is managed by CMS.

For this scenario, the content life cycle is managed by CMS, WebSphere Commerce only retrieves the content files during the runtime.

## Web server scenario - contents received as files and stored in WebSphere Commerce web server

For the web server scenario, the utility retrieves the content files from the CMS system and stores the files in the WebSphere Commerce web server. The Web Feed Integration Utility also creates an URL attachment for each of the content in WebSphere Commerce. The URL in the attachment points to the file in the web server. The content files are served by the web server during the runtime.

The web server scenario supports an update operation, it will override the same file if it already exists in the web server location. But the delete file operation is not supported.

A sample web feed configuration file is provided here.

For web feed integration, if there is a lot of content to be created for the site to be displayed in production, then the content should not be received into WebSphere Commerce as managed files. Storing the content as managed files into WebSphere Commerce will impact the size of the application directly. It can impact receiving additional managed files, fix pack updates, and application deployment.

## Executing the Solution from scheduler

- Scheduler command – Create a scheduled job from Administration Console

### **FeedDataLoadSchedulerCmd**

- Input parameters

**configFile:** The path of the web feed configuration file

**XmlValidation:** The flag that can enable or disable XML validations

**errorLogPath:** The path of the generated error log file

**ConsoleHandler.Level:** Sets the logging level for the console handler

**FileHandler.Level:** Sets the logging level for the file handler

**LoadOrder:** Defines the load order that needs to be used

The screenshot shows the 'New Scheduled Job' configuration page in the WebSphere Commerce Administration Console. The 'Job command' is set to 'FeedDataLoadSchedulerCmd'. The 'Job parameters' field contains 'configFile=C:\temp\CMS-Feed-config-Manag'. The 'Start Date' is set to 2010-11-18 and the 'Start time' is 2:00. The 'Associated user' is 'wcsadmin'. The 'Schedule interval' is 0, 'Job attempts' is 0, and 'Seconds to retry' is 0. The 'Scheduler policy' is 'Run only once' and the 'Job priority' is 5.

You can schedule the command **FeedDataLoadSchedulerCmd** to run through the WebSphere Commerce Administration Console with the interval and retry behavior.

This slide lists all the parameters that can be specified as input parameters to the scheduler command. The parameter **configFile** is mandatory and an error is thrown if it is not specified. Other parameters are optional.

## Executing the solution from command line

- Toolkit:  
`webFeedLoad.bat /full_path/wc-feed-config.xml [-Dname=value]`
- WebSphere Commerce Server:  
`webFeedLoad /full_path/webfeedconfig.xml -Dwaspath=PathToWC.ear [-Dname=value]*`
- Parameter `/full_path/wc-feed-config.xml` is mandatory, parameters `[-Dname=value]` are optional
- Parameter options :
  - DXmlValidation=<true|false>
  - DGenerateDataLoadConfigOnly=<true|false>
  - DLoadOrder=<loadItemName1, loadItemName2>
  - DConsoleHandler.level=<level name>
  - DFileHandler.level=<level name>
  - D<package\_name>.level=<level name> where <level name>=<OFF|SEVERE|WARNING|INFO|CONFIG|FINE|FINER|FINEST|ALL>
  - any other variable names defined in the dataload configuration files

A command-line utility is provided to run a one time fetch of the feed data. The syntax of the command line in different execution environments is shown here.



## Problem determination

- Web Feed Integration solution log goes to Data Load utility's log file
  - \logs\wc-dataload.log
- The database table **CMFEEDLOG** logs information about each feed that has been processed
  - The Web Feed Integration utility checks this table to only load updated or new content

The Web Feed Integration solution relies on the Data Load utility for the error handling mechanism. So the Web Feed Integration utility and Data Load utility use the same log file.

The database table **CMFEEDLOG** is used to log information about each record of a feed which is processed. If a content is already in the Commerce environment, the content is not reloaded to WebSphere Commerce environment unless that content is updated.

## Summary

- Content Management Integration solution supports integration with any CMS which can provide ATOM/RSS feeds
- Three scenarios are supported
  - Managed file scenario - content received as files
  - URL attachment scenario - content received as URLs
  - Web server scenario - content received as files and stored in WebSphere Commerce web server

Content Management integration solution lets you obtain unstructured data such as images and text from a vendor's Content Management System through ATOM and RSS feeds, and store them on the WebSphere Commerce Server. Three scenarios are supported by the solution: Managed file scenario, URL attachment scenario and web server scenario.

---

## Reference

- RSS Feed  
<http://en.wikipedia.org/wiki/RSS>
- Atom Feed  
[http://en.wikipedia.org/wiki/Atom\\_\(standard\)](http://en.wikipedia.org/wiki/Atom_(standard))
- Scheduling a site-level job  
<http://publib.boulder.ibm.com/infocenter/wchelp/v7r0m0/topic/com.ibm.commerce.admin.doc/tasks/tjsregjob.htm>

Some useful references are listed here for further reading.

## Feedback

Your feedback is valuable

You can help improve the quality of IBM Education Assistant content to better meet your needs by providing feedback.

- Did you find this module useful?
- Did it help you solve a problem or answer a question?
- Do you have suggestions for improvements?

Click to send email feedback:

[mailto:iea@us.ibm.com?subject=Feedback\\_about\\_WebFeedIntegration.ppt](mailto:iea@us.ibm.com?subject=Feedback_about_WebFeedIntegration.ppt)

This module is also available in PDF format at: <WebFeedIntegration.pdf>

You can help improve the quality of IBM Education Assistant content by providing feedback.



## Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, DB, and WebSphere are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at <http://www.ibm.com/legal/copytrade.shtml>

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2011. All rights reserved.