HyperContent 2.0b2

Slides for the Introduction to HyperContent seminar at the JA-SIG conference in Baltimore June 12, 20045

by

av317

last updated Jul 13, 2005 12:27:37 PM EDT

HyperContent 2.0 (b2)

by Alex Vigdor

June 12, 2005

Agenda:

- Introductory slides
- Hands-on demonstration

Table of Contents

What is HyperContent 2?

HyperContent 2.0 is a web content management server.

- Author, store, and serve files of any type
- Maintain history of file revisions
- Enforce access controls to projects and files
- Separate design and delivery from data
- Use workflow scripts to orchestrate and augment the flow of information
- Distribute batch rendering and publishing across servers
- Integrate with external authentication, web and file servers

Compare Versions 1 & 2

Version 1.x

- uPortal channel
- static content constructed and pushed via asynchronous batch requests
- fixed three-step workflow (edit, build, publish)
- fixed management GUI

Version 2

- standalone web server
- render dynamic and static content in real-time, or static content in asynchronous batches
- configurable event-driven workflow
- customizable management screens

HyperContent 2 Design

Architecture

HyperContent 2.0 offers a pre-integrated set of components that have been designed for flexibility and tuned for performance and stability. This section details some aspects of

Pipelines

Pipelines:

Pipelines form the backbone of the rendering engine in HyperContent. Pipelines feed both the web server and the batch processing engine. Pipelines are configured with any number of stages, which can be wired in by classname for extensibility. Every project can have its own independent pipeline configuration.

Some built-in Pipeline Stages

- XML includes
- XSLT transformations
- Velocity templates
- Form binding
- Trim whitespace
- Resize image
- HTTP redirect

Static Pipeline

> Used for for static views of content which could be pushed to a separate web server or cached for re-use. Primary caching is done on disk, with soft references used to keep some cached data in memory.

Modal Pipeline

> Used for dynamic and interactive views which cannot be cached or pushed to another server. Identified by a "mode" parameter.

Web Server

HTTP 1.1 Server

- Forces browsers to check cache validity on every request
- Uses 304 response whenever possible
- Applies ETags to views in order to avoid browser confusion
- Offers GZIP compression
- Maps every incoming request to a specific project, whose pipelines are then used to generate the output
- Uses request/response API whose default implementations enable it to run as a servlet.

Workflow

Scripts

Workflows are configured in each project with XML scripts that define the kickoff events for a workflow, its state queues, and the events available for each queue. Event handling is configured by chaining a series of commands.

Events

- System or workflow-specific user events
- Attribute descriptors provide auto-form generation and validation
- Restrict handlers with groups and permissions
- Attributes can be stored as workflow instance variables and referenced by further commands

Queues

- represents a particular state
- has a unique set of event handlers

Commands

- Can be configured with pre-defined tags for built in commands like queue, dequeue, email, render, copy, zip, delete or redirect.
- Can be configured with an "exec" tag to run any WorkCommand implementation by classname

Batch Processing

Distributed Asynchronous Processing

HyperContent 1.x offered an asynchronous processing engine which started a single thread per request, which ran on the machine that received the request until it was complete. While this has proved very stable in production, there are concerns about the scalability of such an approach. HyperContent 2.0 offers a redesigned processing engine:

- Large batch executables are broken down into a set of smaller executables
- All asynchronous executables are serialized to disk and enqueued for processing
- Any number of servers can feed from the same executable queues
- A dedicated Batch Monitor thread recombines results from broken down batches and cleans up completed executables
- Out-of-the-box executables use the ContentServer.throttle() method to effectively yield processor time to real time web requests.
- Sensitive data that is required to process an executable (e.g. an SFTP password) is encrypted to disk using 2 key encryption, where the private key is automatically generated at startup and never written to disk.

AuthN & AuthZ

Authentication

HyperContent 2 uses JAAS for authentication. The built-in LoginModule implementation stores an MD5 Hash of the user's password in an XML file. CAS support for JAAS is coming soon.

Authorization

HyperContent 1.x was able to leverage uPortal Groups and Permissions management for authorization. In 2.0, groups and permissions are stored in XML files in a project's repository. uPortal Groups and Permissions is currently being migrated to a standalone component, which will be an optional plugin in a future HyperContent 2.x release.

I18n & L10n

Internationalization

HyperContent is UTF-8 transparent, supporting the full range of Unicode characters

New in 2.0: even properties files are UTF-8 transparent - no more \uXXXX encoding

Localization

HyperContent provides support for mapping localized resource bundles into either XSL or Velocity templates.

Authoring Tools

HyperContent versions 1.4 and 2.0 provide the same core set of 7 authoring tools:

- WYSIWYG xhtml + xml
- Dublin Core metadata
- Upload/download
- Spell check
- Image crop/resize
- VCard (contact info)
- Navigation / site map

Unfinished for 2.0

Scheduling

A scheduling component will allow configuration of timed and repeating workflow events.

Testing

More testing is always good!

Administrative scripts

Rebuild search index, clear locks, maybe more.

Beyond 2.0

And then?

- AJAX & DHTML redesign of XML/WYSIWYG editor
- Enhanced configuration file format
- WSRP/UDDI (serve up portlet content)
- JA-SIG Groups & Permissions integration
- WebDAV (seamless desktop authoring)
- Site template library
- Project Configuration wizard