

Présentation Alfresco

Ludovic Plantin, Frédéric Sénèque, Xu Zhao

Polytech'Grenoble

Décembre 2008

Plan de la présentation

- 1 Introduction
- 2 Les CMS
- 3 La gestion des documents
- 4 Workflow
- 5 Répartition et réplication
- 6 Conclusion
- 7 Bibliographie

Introduction

Pourquoi une présentation sur Alfresco ?

- Un outil libre donnant l'occasion de parler des workflows en entreprise
- Un outil souvent utilisé par les sociétés de services (que vous aurez donc peut-être à utiliser un jour)
- Une approche différente des problèmes de répartition de charge, de la duplication de données...

Les CMS

1 Introduction

2 Les CMS

- Qu'est-ce qu'un CMS ?
- Les avantages des CMS
- Les différents CMS
- Les avantages des solutions Open Source
- Présentation Alfresco
- Le réseau de partenaire Alfresco

3 La gestion des documents

4 Workflow

5 Répartition et réplcation

6 Conclusion

7 Bibliographie

Qu'est-ce qu'un CMS ?

CMS/SGC : Une application informatique servant à créer, éditer, gérer et publier du contenu de manière organisée et automatisée.

Ils sont fréquemment utilisés pour le stockage de fichiers, le contrôle de version, et l'édition de documents tels que des articles de presse, etc.

Ils peuvent manipuler tous types de contenu :

- Documents bureautiques
- Des fichiers multimédia (image, audio, vidéo)
- Documents électroniques
- Contenu Web

Les avantages des CMS

- **Accessibilité** : L'interface d'administration, accessible à partir d'un simple navigateur internet, autorise la mise à jour du contenu pour tous les collaborateurs, qu'ils soient internes ou externes à la société ou l'organisation.
- **Autonomie, simplicité** : Les opérations élémentaires ne nécessitent aucune connaissance en programmation ni d'HTML
- **Multi-utilisateur et travail collaboratif** : Chaque collaborateur peut avoir un compte qui lui est propre et a des droits précis. Un espace de travail collaboratif est souvent présent.
- **Retour sur investissement rapide** :
 - Coût d'acquisition et d'intégration inexistant ou très faible pour les logiciels libres
 - Coût d'acquisition variable pour les logiciels commerciaux.
 - Gains sur les coûts de traitement : mise à jour simplifiée, automatisation des traitements (importation, exportation, mise en ligne, ...)

Les différents CMS

Plusieurs types de CMS :

- Publication d'articles sur le web : SPIP (Open-source)
- Gestion de documents d'entreprises : Alfresco (Open-source), Documentum (Propriétaire)
- " Assembleur de site web" : Drupal (Open-source)
- Publications de vidéos : CoreMedia CMS (Propriétaire)

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Les avantages des solutions Open Source

Les avantages des solutions Open Source :

- Universalité des développeurs(pas seulement une équipe)
- Correction des erreurs (bugs) rapides
- Meilleure sécurité
- Personnalisation possible
- Indépendance par rapport au fournisseur
- Support de la communauté
- Avantage économique

Présentation Alfresco

Alfresco :

- Un ECM(Enterprise Content Management) Open source en langage JAVA
- Créé en 2005 par John Newton et John Powell
- Dernière version stable (3.0) le 31 octobre 2008

Les avantages d'Alfresco :

- **Portabilité** : Compatible avec de nombreux OS, serveur d'applications et BD (application JEE)
- **Fonctionnalité** : Gestion de documents, de contenu web, ... Possibilité de créer des clients riches/lourds via l'API
- **Nombreuses extensions** : Facebook, MICROSOFT Office, Iphone, wiki...
- **Personnalisation** : Nombreuses solutions réalisées autour d'Alfresco (ex : Open Wide)

Le réseau de partenaire Alfresco

Alfresco est un logiciel Open-source qui peut donc être modifié pour fournir une application personnalisée. À ce titre Alfresco entretient un réseau de partenaires capables d'expertiser, configurer, personnaliser et mettre en place des solutions basées sur Alfresco pour des clients dans le monde entier.

Voici quelques unes des organisations utilisant Alfresco :



"Alfresco takes ECM from the Client-Server World of the 90's into the SOA, Web Services and Aspect Oriented World of Today"

Romain Sutton, PDG de Technical Architecture - Reed Managed Services plc

La gestion des documents

- 1 Introduction
- 2 Les CMS
- 3 La gestion des documents
 - Mode d'accès aux documents
 - Opérations possibles sur les documents
 - Mode de stockage des données
 - Démonstration
- 4 Workflow
- 5 Répartition et réplication
- 6 Conclusion
- 7 Bibliographie

Mode d'accès aux documents

- Via le navigateur web (HTTP)
- Via un disque dur monte en réseau :
 - Network File System(NFS)
 - Common Internet File System(CIFS)
 - Webdav (Web-based Distributed Authoring and Versioning)
- Via FTP (File Transfer Protocol)

Opérations possibles sur les documents

- Opérations standards (créer, copier, coller, déplacer, supprimer....)
- Prise et relâchement de verrou (Check in/check out)
- Gestion de version
- Indexage et recherches intelligentes
- Transformation de contenu (ex : génération automatique de pdfs)

Mode de stockage des données

- Informations sur les utilisateurs, les projets, les règles de workflows dans la base de données
- Méta-information des fichiers (chemin d'accès, nom, date de modification...) dans la base de données
- Contenu des fichiers dans un entrepôt de document (système de fichier)
- Fichiers de configuration du serveur au format XML

Démonstration

Démonstration

Workflow

- 1 Introduction
- 2 Les CMS
- 3 La gestion des documents
- 4 Workflow
 - Qu'est-ce qu'un workflow ?
 - Caractéristiques des Workflows
 - Les workflows dans Alfresco
 - Définitions de workflows
 - Implémentation
 - Démonstration
- 5 Répartition et réplication
- 6 Conclusion
- 7 Bibliographie

Qu'est-ce qu'un workflow ?

Workflow : gestion électronique des processus métiers ou "Flux de travail"

Modélisation et la gestion informatique de l'ensemble des tâches à accomplir et des différents acteurs impliqués dans la réalisation d'un processus métier

Un workflow comporte :

- Des acteurs
- Un circuit de validation
- Des tâches à accomplir entre les différents acteurs
- Des délais
- Des modes de validation

Caractéristiques des Workflows

Analogies dans le domaine ECM : ligne d'assemblage ou processus de fabrication

Exemples de Workflow dans la vie courante :

- Lorsqu'on fait une demande d'aide au logement, on démarre un workflow
- Quand on réalise le suivi d'un colis sur Internet, on demande des informations (méta-données) concernant ce workflow
- Lorsque quelqu'un écrit un commentaire sur blog, une étape du workflow est de valider ce commentaire pour l'afficher sur le blog.
- ...

Deux choses à retenir :

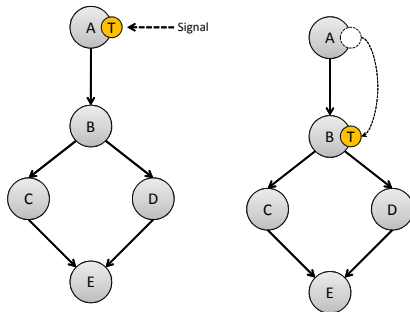
- Les deux types d'interactions qui existent sont : hommes à hommes et homme-machine
- Les workflow sont axés sur les contenus ou les documents

Les workflows dans Alfresco

Workflows basique	Workflows avancés
<ul style="list-style-type: none">● Configurable par des utilisateurs finaux non techniciens via l'interface web Alfresco● Utilise les règles, les espaces et les actions● Ne peut traiter qu'un processus avec des transitions avant et/ou arrière● Ne permet pas de faire des processus en parallèle, des coupures, des jointures et des décisions● Ne maintient pas l'état et les méta-données concernant le processus lui-même	<ul style="list-style-type: none">● Défini par les analystes et les développeurs via Eclipse● Utilise les fonctionnalités du "JBoss jBPM workflow Engine"● Modélise tous les processus métiers (décisions, split, jointure, flux parallèles, sous-processus, états d'attente et timer)● Ne permet pas de faire des processus en parallèle, des coupures, des jointures et des décisions● Maintient les états et les variables de processus

Moteur de Workflows

Type de Noeud	Description
Start-state	Noeud initiateur
End-state	Noeud de terminaison
State	Etat d'attente. Peut seulement être signalé en cas de transition activée
Node	Noeud où on peut executer une action
Fork	Crée des processus parallèles concurrents
Decision	Choix parmi plusieurs chemins d'executions



Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « ResourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « RessourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « RessourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « RessourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « ResourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « ResourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « RessourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Implémentation

Fichiers du workflow

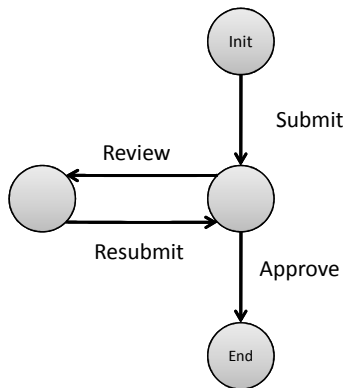
- Créer le fichier de définition (le squelette)
- Intégrer le comportement du workflow dans le squelette
- Créer le fichier de modèle des tâches
- Créer le « ResourceBundle »
- Déclarer les interfaces webs du modèle des tâches

Déploiement

- Définir le contexte d'utilisation du workflow
- Déployer le workflow via la console de workflow
- Redemarrer le serveur Alfresco

Démonstration

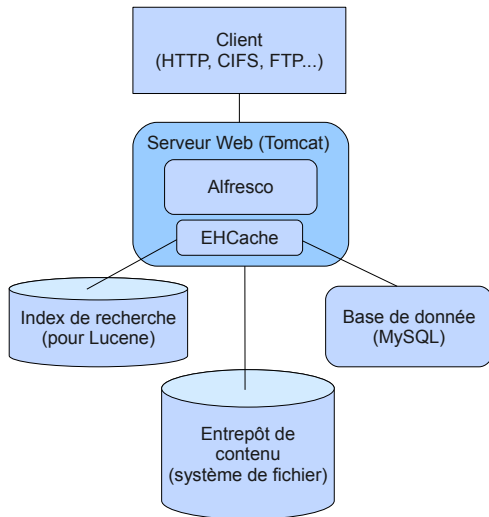
Démonstration



La répartition et la réplication

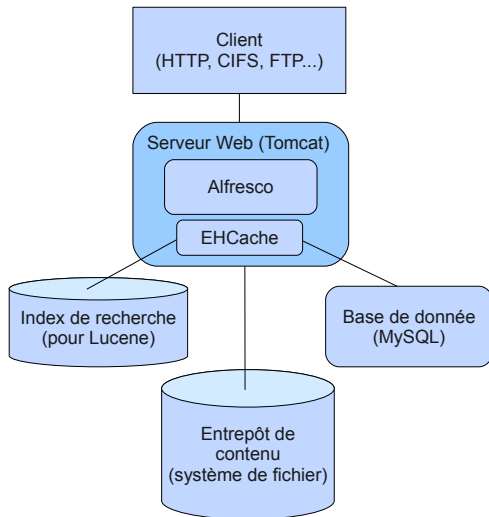
- 1 Introduction
- 2 Les CMS
- 3 La gestion des documents
- 4 Workflow
- 5 Répartition et réplication
 - Architecture de base
 - Répartition de la charge
 - Réplication de l'entrepôt de contenu
 - Réplication de la base de données
 - Démonstration
- 6 Conclusion
- 7 Bibliographie

Architecture de base



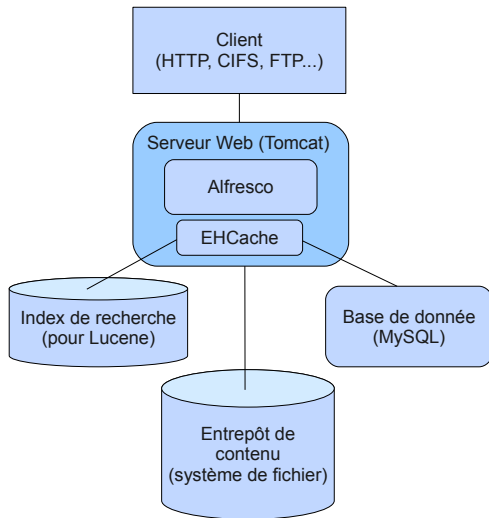
- Entrepôt de Contenu : sauvegarde le contenu de chaque fichier utilisateur dans un fichier avec un identifiant unique...
- Base de données : stocke les métas données, les workflows...
- Système d'indexation pour les recherches
- EHCACHE : cache applicatif (entre autres pour Hibernate)

Inconvénients de cette architecture



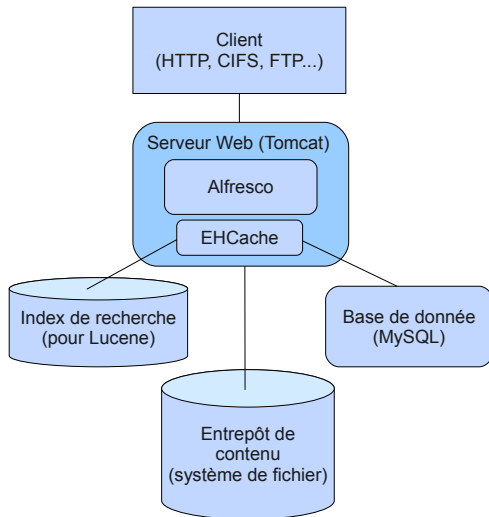
- Risque de pertes de données

Inconvénients de cette architecture



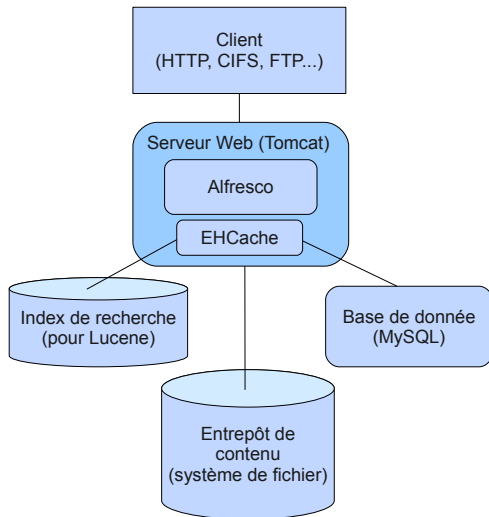
- Risque de pertes de données
- Multiples points de fragilités

Inconvénients de cette architecture



- Risque de pertes de données
- Multiples points de fragilités
- Passage à l'échelle uniquement vertical

Inconvénients de cette architecture



- Risque de pertes de données
- Multiples points de fragilités
- Passage à l'échelle uniquement vertical
- Faible disponibilité

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)
 - Entrées DNS multiples

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)
 - Entrées DNS multiples
 - Mix des solutions précédentes

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)
 - Entrées DNS multiples
 - Mix des solutions précédentes
- Comment offrir la même vision des données à travers les différents serveurs ?

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)
 - Entrées DNS multiples
 - Mix des solutions précédentes
- Comment offrir la même vision des données à travers les différents serveurs ?
 - Une solution est d'utiliser une base de données et un entrepôt de contenu commun

Répartition de charge

- Idée : Répartir les clients sur plusieurs serveurs
- Comment les répartir ?
 - Configuration statique
 - Routeurs répartiteurs de charge (conservation des associations serveur/ip)
 - Entrées DNS multiples
 - Mix des solutions précédentes
- Comment offrir la même vision des données à travers les différents serveurs ?
 - Une solution est d'utiliser une base de données et un entrepôt de contenu commun
 - Problème : deux points de fragilité limitant le passage à l'échelle

Réplication de l'entrepôt : Objectifs

Objectifs :

- Augmenter la fiabilité en dupliquant les données concernant le contenu

Réplication de l'entrepôt : Objectifs

Objectifs :

- Augmenter la fiabilité en dupliquant les données concernant le contenu
- Conserver l'indépendance des différents serveurs

Réplication de l'entrepôt : Objectifs

Objectifs :

- Augmenter la fiabilité en dupliquant les données concernant le contenu
- Conserver l'indépendance des différents serveurs
- Offrir une vision cohérente à travers les différents serveurs

Réplication de l'entrepôt : Objectifs

Objectifs :

- Augmenter la fiabilité en dupliquant les données concernant le contenu
- Conserver l'indépendance des différents serveurs
- Offrir une vision cohérente à travers les différents serveurs
- Pourvoir accéder "rapidement" aux données

Réplication de l'entrepôt : Objectifs

Objectifs :

- Augmenter la fiabilité en dupliquant les données concernant le contenu
- Conserver l'indépendance des différents serveurs
- Offrir une vision cohérente à travers les différents serveurs
- Pourvoir accéder "rapidement" aux données

Solution proposée :

- Un entrepôt primaire unique pour chaque serveur
- Un entrepôt secondaire partagé par tous

Réplication de l'entrepôt : Fonctionnement

Fonctionnement :

- En écriture :
 - Création d'un identifiant unique liant les informations métas du fichier et son contenu
 - Écriture du contenu avec l'identifiant dans l'entrepôt primaire
 - Duplication du contenu dans l'entrepôt secondaire
 - Écriture de la relation méta-infos/id dans la BD
- En lecture :
 - Récupération de l'id dans la BD
 - Recherche du contenu dans l'entrepôt primaire
 - S'il n'est pas présent, duplication de celui-ci depuis l'entrepôt secondaire vers l'entrepôt primaire

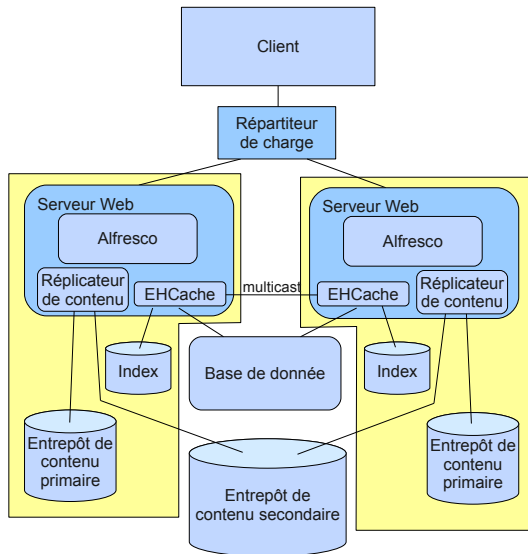
Réplication de l'entrepôt : Fonctionnement

Fonctionnement :

- En écriture :
 - Création d'un identifiant unique liant les informations métas du fichier et son contenu
 - Écriture du contenu avec l'identifiant dans l'entrepôt primaire
 - Duplication du contenu dans l'entrepôt secondaire
 - Écriture de la relation méta-infos/id dans la BD
- En lecture :
 - Récupération de l'id dans la BD
 - Recherche du contenu dans l'entrepôt primaire
 - S'il n'est pas présent, duplication de celui-ci depuis l'entrepôt secondaire vers l'entrepôt primaire

Problèmes : Validité du cache et de la base d'index pour les recherches

Réplication de l'entrepôt : nouvelle architecture



- Un entrepôt de contenu primaire par serveur
- Un entrepôt de contenu partagé entre les serveurs
- Un composant gérant la réplication du contenu dans les deux entrepôts par serveur
- Signaux multicasts entre les caches des serveurs pour maintenir leur cohérence

Avantages

- Cohérence des données entre les différents serveurs
- Pas de conflit écriture/écriture ou lecture/écriture possible sur l'entrepôt secondaire (grâce à l'identifiant unique)
- L'accès aux fichiers est plus rapide grâce à l'utilisation des entrepôts primaires
- Ajout possible d'un serveur sans avoir à redémarrer ou à reconfigurer les autres
- Les entrepôts primaires peuvent être configurés pour ne garder que les fichiers les plus utilisés, limitant ainsi l'espace disque nécessaire
- Possibilité d'utilisation d'un système de fichiers distribué (SAN, P2P ou autres) pour l'entrepôt secondaire.

Réplication de la base de données

- Différentes possibilités en fonction du choix de la base de données
- Configuration maître esclave, l'esclave sauvegarde et prend le relais si le primaire tombe
- Master/Master, les deux (ou plus) base de données traitent des requêtes en parallèle, et communiquent pour gérer les écritures concurrentes et les mises à jour. Efficace car plus de lectures que d'écritures et pas besoins de refaire les calculs avec les mises à jour.
- Autre exemple : MySQL cluster (haute disponibilité 99,999%) : utilise des nœuds de stockages et des nœuds de traitements

Démonstration

Démonstration

Conclusion

Pour résumer :

- Alfresco est un outil libre très employé pour gérer intelligemment les documents en entreprise
- Il offre la possibilité de créer des workflows adaptés à ses besoins
- Il supporte le passage l'échelle et peut être hautement disponible

Bibliographie

Site commercial d'Alfresco : <http://www.alfresco.com/fr/>

Wiki d'Alfresco : http://wiki.alfresco.com/wiki/High_Availability_Configuration_V1.3

Différentes présentation d'Alfresco : <http://www.slideshare.net/> (Anglais)

Présentation vidéo sur le clustering : <http://tinyurl.com/alfresco-cluster> (Anglais)

MySQL Cluster 6.2 : Livre blanc de Mikael Ronström, MySQL AB et Lars Thalmann, MySQL AB

Articles de Jeff Potts : <http://ecmarchitect.com/> (Anglais)

Ehcache : <http://ehcache.sourceforge.net/> (Anglais)

Des questions ?

Des questions ?