

Google Search Appliance

Présentation de l'intégration de contenus

Août 2014



Intégration de contenus

Cet article présente les fonctionnalités que vous pouvez utiliser pour intégrer les contenus de différents référentiels dans le système Google Search Appliance (GSA).

À propos de ce document

Les recommandations et les informations contenues dans ce document sont issues de notre travail sur le terrain avec plusieurs clients et dans différents environnements. Nous remercions nos clients et partenaires d'avoir partagé leurs expériences et impressions.

Sujets abordés	Cet article concerne l'utilisation des flux, des connecteurs et de l'accès au cloud.
Public principal	Chefs de projets, administrateurs du système GSA et développeurs de connecteurs.
Environnement informatique	GSA et différents référentiels de données externes, tels que systèmes de gestion de contenu d'entreprise, serveurs LDAP et Google Apps.
Phases de déploiement	Configuration initiale du système GSA.
Autres sources d'information	• Learngsa.com fournit des ressources de formation sur le système GSA. • La documentation produit de GSA fournit des informations complètes sur le système GSA. • Le portail d'assistance Google for Work permet d'accéder à l'assistance Google.

Sommaire

[À propos de ce document](#)

[Introduction](#)

[Chapitre 1 Flux personnalisés](#)

[Présentation](#)

[Types de flux](#)

[Mise à jour des flux](#)

[Intégrité des flux](#)

[Bonnes pratiques](#)

[Sécurité](#)

[Déploiements à grande échelle](#)

[Chapitre 2 Considérations générales sur les connecteurs](#)

[Présentation](#)

[Taille du référentiel de contenu](#)

[Sécurité](#)

[Emplacement physique](#)

[Personnalisation du connecteur](#)

[Performances](#)

[Bonnes pratiques de création d'un connecteur](#)

[Chapitre 3 Flux, robot d'exploration et proxy, structure d'adaptateurs ou structure de connecteurs ?](#)

[Présentation](#)

[Utilisation de la structure de connecteurs](#)

[Utilisation de la structure d'adaptateurs \(Plexi\)](#)

[Utilisation du robot d'exploration via un proxy](#)

[Utilisation d'un client de flux](#)

[Chapitre 4 Connecteur SharePoint de Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Liaison précoce de groupes](#)

[Où est mon fichier ?](#)

[Description](#)

[Recherche sécurisée](#)

[Performances de traitement](#)

[Connecteur interne ou externe ?](#)

[Chapitre 5 Connecteur de systèmes de fichiers Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Où est mon fichier ?](#)

[Description](#)

[Recherche sécurisée](#)

[Connecteur interne ou externe ?](#)

[Chapitre 6 Connecteur de bases de données de Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Où est mon enregistrement ?](#)

[Description](#)

[Recherche sécurisée](#)

[Chapitre 7 Connecteur Lotus Notes de Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Où est mon document ?](#)

[Description](#)

[Recherche sécurisée](#)

[Conseils et astuces](#)

[Chapitre 8 Connecteur Livelink de Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Où est mon enregistrement ?](#)

[Description](#)

[Recherche sécurisée](#)

[Chapitre 9 Connecteur EMC Documentum de Google Search Appliance](#)

[Présentation](#)

[Où est mon enregistrement ?](#)

[Description](#)

[Recherche sécurisée](#)

[Chapitre 10 Connecteur IBM FileNet de Google Search Appliance](#)

[Présentation](#)

[Balayage](#)

[Où est mon enregistrement ?](#)

[Description](#)

[Recherche sécurisée](#)

[Chapitre 11 Connecteur LDAP de Google Search Appliance](#)

[Présentation](#)

[Où est mon enregistrement ?](#)

[Description](#)

[Recherche sécurisée](#)

[Limites](#)

[Chapitre 12 Architecture de déploiement des connecteurs](#)

[Présentation](#)

[Chapitre 13 Accès au cloud](#)

[Présentation](#)

[Dépannage](#)

[Vérification des identifiants OAuth](#)

[Vérification du principal identifiant validé](#)

[Vérification des paramètres de pare-feu](#)

Introduction

Pour pouvoir effectuer une recherche, nous devons indexer l'ensemble du contenu au préalable. Il existe pour cela différentes méthodes d'indexation :

- **Exploration** : processus qui consiste à suivre les liens à partir d'un point de départ, à télécharger l'ensemble du contenu et à l'ajouter à l'index.
- **Exploration via un proxy** : processus similaire au précédent, mais qui passe par un proxy, par exemple pour ajouter des métadonnées (c'est-à-dire microdonnées, paramètres de sécurité) ou modifier du contenu.
- **Exploration de bases de données** : indexation de bases de données à l'aide de requêtes SQL. Une requête JDBC-SQL permet de mapper le modèle de données sur le format du flux.
- **Utilisation de flux** : un flux est un document XML qui décrit les documents/contenus, et permet de contrôler les métadonnées et la sécurité de chaque document.
- **Connecteur** : logiciel qui balaye les documents d'un référentiel, généralement un système DMS ou CMS, et les transmet au système GSA pour indexation.

Dans les chapitres qui suivent, nous examinerons en détail les différentes méthodes d'indexation, en précisant les avantages et les inconvénients de chacune d'elles.

Chapitre 1 Flux personnalisés

Présentation

Un flux est un document XML indiquant à Google Search Appliance quels contenus vous souhaitez indexer. Cela permet de transmettre des données au système de recherche. L'utilisation de flux peut être indiquée pour différentes raisons :

- Accès au contenu
- Liens JavaScript
- Métadonnées externes
- Listes de contrôle d'accès (LCA) par URL
- Code d'état incorrect
- Indexations urgentes

Accès au contenu

Vous pouvez envisager d'utiliser des flux lorsque vous n'avez pas d'interface Web ni d'autre moyen d'accéder au contenu via l'exploration standard. L'envoi d'un flux de contenu est alors la seule manière de procéder.

Liens JavaScript

Certaines pages Web contiennent des données ou des liens JavaScript générés automatiquement que GSA n'est pas en mesure d'explorer.

GSA ne peut analyser ce type de liens que dans certaines portions de code JavaScript intégré. Ainsi, les flux sont parfois le seul moyen d'indexer tout le contenu d'un site Web, en particulier lorsque celui-ci est en grande partie généré de façon dynamique.

Métadonnées externes

Les flux peuvent vous permettre d'associer des métadonnées externes à vos documents. Les propriétés incluses dans les documents et les balises META des en-têtes HTML sont automatiquement associées aux documents indexés. Cependant, vous pouvez inclure des métadonnées supplémentaires (provenant d'un système de gestion de contenu, par exemple) en les envoyant au système GSA à l'aide de flux.

Listes de contrôle d'accès (LCA) par URL

Pour mettre en œuvre les fonctionnalités avancées (espaces de noms et sensibilité à la casse, par exemple) des LCA par URL dans le cadre des autorisations, vous devez utiliser un flux.

Utilisation incorrecte du code d'état HTTP 200

Dans certains cas, les pages supprimées ne renvoient pas le code d'état 404. Lorsqu'une URL renvoie une page d'erreur associée au code 200 (OK), GSA la traite comme une page existante, puis l'indexe. Le meilleur moyen de remédier à ce problème consiste à envoyer des flux de "suppression".

Indexations urgentes

Les flux peuvent vous permettre d'indexer rapidement des documents nouveaux ou mis à jour. Pour les communiqués de presse et les autres types de contenus ayant un intérêt immédiat et devant être indexés rapidement, il est préférable d'utiliser des flux indiquant à GSA de placer ces éléments en tête de la file d'attente d'exploration, même s'il s'agit de contenus modifiés ou d'URL figurant déjà dans l'index.

Types de flux

Le système Google Search Appliance accepte les deux types de flux suivants :

- Flux Web/métadonnées-et-URL
- Flux de contenu

Flux Web/métadonnées-et-URL

Un flux Web/métadonnées-et-URL fournit au système de recherche une liste d'URL associées à des métadonnées. GSA se charge ensuite d'explorer ces URL. L'envoi d'un flux métadonnées-et-URL permet non seulement d'associer des métadonnées supplémentaires aux documents, mais aussi de déclencher une nouvelle exploration du contenu. Les URL intégrées dans les documents seront également explorées si elles correspondent aux **formats à suivre** que vous avez définis.

Flux de contenu

Le flux de contenu fournit au système de recherche à la fois les URL et leur contenu. Il existe deux types de flux de contenu :

- Flux complet
- Flux incrémentiel

Avec un flux complet, les contenus obsolètes associés à des URL existantes sont supprimés avant l'indexation des nouveaux contenus. Avec un flux incrémentiel, si un contenu est envoyé vide avec des métadonnées associées, seules les métadonnées sont mises à jour, et inversement. Les URL intégrées dans ces documents seront également explorées si elles correspondent aux **formats à suivre et à explorer** que vous avez définis.

Mise à jour des flux

Les flux complets ne peuvent pas être modifiés (si le nom de leur source de données ne change pas). Si vous remplacez un flux complet par un flux incrémentiel, les documents associés au flux initial seront supprimés de l'index. Cela signifie qu'il est préférable de toujours utiliser des flux incrémentiels lorsque le contenu a de fortes chances d'être modifié.

Si vous envoyez un flux incrémentiel incluant des modifications de métadonnées, les nouvelles métadonnées remplacent les métadonnées existantes.

Remarque : Il est impossible de mettre à jour des éléments de métadonnées individuellement, vous devez mettre à jour tous les champs simultanément.

Intégrité des flux

Un bon client de flux ne se contente pas d'envoyer les flux, il s'assure également que les documents ajoutés sont indexés correctement. Les flux sont d'abord prétraités pour vérifier leur format, par rapport aux **formats d'URL à suivre et à explorer** et aux **formats à ne pas suivre**. S'il s'agit d'un flux de type métadonnées-et-URL, GSA tente de l'explorer. Dans le cas contraire, le contenu est indexé directement. Il est possible que des erreurs surviennent pendant cette phase du processus. Les clients de flux peuvent utiliser l'[API Administrative](#) de GSA pour contrôler l'état du processus durant les deux phases suivantes :

- [Récupération des données du flux fourni par la source](#)
- [Récupération de l'état des documents](#)

Bonnes pratiques

Une solution de gestion des flux bien conçue doit posséder les caractéristiques suivantes :

- Noms de source de données peu nombreux.
- GSA effectue le suivi de l'état des flux dans des fichiers. Lorsque les sources de données sont trop nombreuses, les fichiers associés le sont également et cela réduit les performances de la console d'administration. En règle générale, une source de données correspond à une source de contenu.
- Un seul fichier de flux pour de nombreux documents.
- La taille de votre fichier de flux peut atteindre 1 Go. GSA réalisera un traitement largement plus efficace sur un fichier volumineux comprenant de nombreux documents que sur de nombreux petits fichiers contenant seulement quelques documents. Il faut évidemment tenir compte du délai de récupération des erreurs : si le traitement échoue, le nombre de documents à renvoyer sera d'autant plus grand que le flux est volumineux.
- Compression des flux volumineux pour améliorer les performances.
- Codage en base64 pour les métadonnées contenant des caractères non-ASCII.
- Propriété "last-modified" permettant à GSA de traiter plus rapidement les fichiers de flux.
- Si vous envoyez des flux de contenu composés exclusivement de texte, les performances seront largement plus élevées que si vous envoyez des fichiers complets, codés en base64, qui requièrent un décodage et parfois une conversion au format HTML (comme pour des documents Word, PDF, etc.).
- Vous pouvez consulter le nombre de flux en retard à l'adresse <http://<GSA>:19900/getbacklogcount>. Un client de flux efficace doit permettre de spécifier le nombre maximal de flux en retard au-delà duquel l'envoi de flux au système GSA est interrompu.
- Pour améliorer encore la vitesse de traitement des flux, vous pouvez répartir ces derniers équitablement sur différents noms de source de données (trois, par exemple) qui seront traités en parallèle.
- Si vous incluez des LCA dans vos flux, exploitez la fonctionnalité d'héritage de ces listes pour réduire leur incidence sur le système de contenu et l'index de GSA.
- Vous ne pouvez pas modifier la sensibilité à la casse des LCA après leur inclusion dans l'index. La sensibilité à la casse fonctionne différemment d'un système de contenu à l'autre et doit être prise en compte dès la conception de votre client de flux.

Sécurité

Si vous utilisez des flux métadonnées-et-URL, l'attribut `authmethod` associé aide le système GSA à identifier le protocole d'authentification à mettre en œuvre lors de l'exploration des URL. Cet attribut indique également si les URL doivent être exclues de la recherche publique (toute valeur différente de `authmethod="none"`). À partir de la version 7.2 de GSA, il est recommandé de toujours spécifier l'attribut `authmethod`, même lorsque sa valeur est `"none"`.

L'utilisation de l'attribut `authmethod` présente toutefois certains inconvénients, car il implique un nouvel envoi des flux en cas de modification. Par conséquent, il est préférable d'opter pour la méthode suivante :

1. Définissez l'attribut sur une valeur fixe telle que `httpbasic`.
2. Configurez la fonctionnalité "Accès du robot d'exploration" ou "Authentification par formulaire" sous **Sources de contenu > Exploration du Web > Exploration sécurisée** dans la console d'administration.

Pour marquer les contenus comme publics ou sécurisés à la volée, cochez ou décochez **"Rendre public"** sur la page **Sources de contenu > Exploration du Web > Exploration sécurisée > Accès du robot d'exploration**.

Déploiements à grande échelle

Tenez compte des points suivants lorsque vous créez des flux destinés à des déploiements avancés :

- Dans une configuration en miroir, vous ne pouvez envoyer les flux qu'au nœud maître.
- Dans une configuration d'exploration et de traitement distribuée, vous ne pouvez envoyer les flux qu'aux nœuds maîtres.
- Dans une configuration unifiée, vous pouvez envoyer les flux à n'importe quel nœud.

Chapitre 2 Considérations générales sur les connecteurs

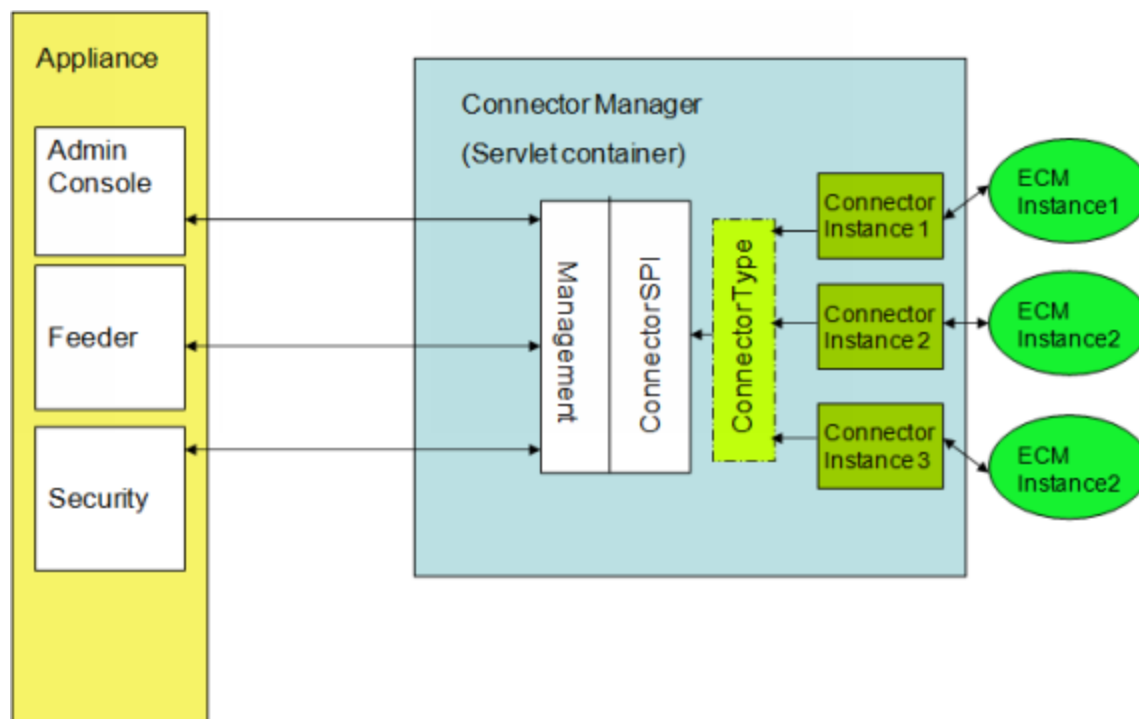
Présentation

Les [connecteurs](#) permettent d'étendre la portée du système Google Search Appliance à des référentiels non Web tels que des systèmes de gestion de contenu d'entreprise, afin de rendre davantage de contenus accessibles via la recherche. Google Search Appliance fournit des connecteurs pour les référentiels de documents suivants :

- Microsoft SharePoint
- Microsoft Active Directory
- Systèmes de fichiers
- Bases de données
- Lotus Notes
- Open Text Livelink
- EMC Documentum
- IBM FileNet
- LDAP

GSA 7.2 comprend des versions bêta du connecteur 4.0. Ce document concerne la version 3.2 stable des connecteurs.

Pour découvrir les documents d'un référentiel, Google Search Appliance n'explore pas son contenu. Il fait appel à un processus appelé "balayage" au cours duquel le connecteur envoie des requêtes au référentiel pour récupérer des données relatives aux documents, qui sont ensuite transmises au système de recherche pour indexation. Le diagramme suivant donne un aperçu des différents composants d'une solution faisant intervenir un connecteur.



Les connecteurs permettent à Google Search Appliance non seulement de traiter et d'effectuer des recherches dans des documents stockés dans des référentiels non Web, mais aussi d'explorer des interfaces Web dont l'accès est restreint :

- Sites Web extrêmement complexes, en raison de l'utilisation intensive de JavaScript par exemple, sur lesquels le système de recherche ne parvient pas à atteindre tous les documents
- Interfaces dans lesquelles les documents sont dissociés de leurs métadonnées (métadonnées supplémentaires stockées dans une base de données distincte, par exemple)

De nombreux éléments sont à prendre en considération lorsque vous planifiez le déploiement d'un connecteur, y compris :

- Taille du référentiel de contenu
- Sécurité
- Emplacement physique
- Personnalisation du connecteur

Taille du référentiel de contenu

Le premier élément à examiner est la taille du référentiel de contenu. Les référentiels volumineux sont plus longs à balayer. Avec certains connecteurs, il est possible de réduire la durée du balayage en exécutant plusieurs instances du connecteur, en veillant à ne pas dupliquer le contenu.

La taille du contenu peut également avoir une incidence sur la topologie du déploiement. Les connecteurs sont habituellement déployés sur des serveurs distincts du système GSA. Cependant, le système de recherche comprend plusieurs connecteurs intégrés qui peuvent être exécutés sans matériel supplémentaire. Ces connecteurs "internes" incluent SharePoint, File Share et LDAP.

Google recommande l'utilisation des connecteurs internes uniquement si le référentiel à traiter comprend moins de 500 000 documents. Avec des référentiels plus volumineux, les connecteurs internes occupent davantage de ressources sur le système GSA et réduisent les performances globales de ce dernier. En revanche, les connecteurs "externes", qui ne sont pas inclus dans le système de recherche, permettent d'utiliser des configurations matérielles personnalisables et de mieux contrôler le fonctionnement du connecteur, le dépannage et les mises à jour.

Sécurité

Autre élément essentiel de la planification du déploiement d'un connecteur : la sécurité. Les connecteurs peuvent gérer à la fois l'authentification (AuthN) d'un utilisateur, la résolution de groupes et les autorisations (AuthZ) associées aux résultats de recherche. En règle générale, les connecteurs sont utilisés pour gérer uniquement les autorisations. Ils peuvent s'appuyer sur les API natives des référentiels de contenu.

Si vous devez sécuriser la fonction de recherche, veillez à planifier soigneusement l'authentification via un ou deux mécanismes compatibles avec le système de recherche tels que Kerberos, les formulaires ou SAML, l'objectif étant d'établir l'identité de l'utilisateur afin que les connecteurs puissent fournir les autorisations correspondantes.

Emplacement physique

L'emplacement physique des connecteurs est un point important à prendre en compte si vous souhaitez que Google Search Appliance indexe des sources de contenu situées dans différents centres de données. Vous devez prévoir au moins trois éléments d'architecture dans votre déploiement pour permettre aux connecteurs de gérer les autorisations lors de la présentation des résultats :

- Google Search Appliance
- Sources de contenu
- Connecteur

Pour minimiser la latence, au moins deux de ces éléments doivent être situés au même endroit.

Personnalisation du connecteur

La modification du fonctionnement d'un connecteur existant peut être nécessaire dans certains cas, par exemple si vous souhaitez ajouter une fonctionnalité d'authentification au connecteur de bases de données fourni par Google. Le gestionnaire de connecteurs est conçu pour faciliter ces personnalisations. Il utilise la structure Spring qui vous permet de remplacer des mises en œuvre existantes de la SPI du gestionnaire de connecteurs par les vôtres en toute simplicité.

Le filtrage des documents est une autre fonctionnalité importante et relativement nouvelle du gestionnaire de connecteurs. Vous pouvez appliquer des filtres aux documents pour compléter leurs propriétés et en supprimer ou en modifier certaines. Il est également possible d'associer plusieurs filtres pour former un processus de traitement des documents. Par exemple, vous pouvez définir un filtre personnalisé pour déchiffrer les documents chiffrés de votre système de contenu avant leur inclusion dans GSA.

Performances

Si vous utilisez des machines virtuelles exécutées sur un matériel partagé proche de sa capacité maximale, le balayage sera beaucoup plus lent que sur des serveurs réels. Le cas échéant, assurez-vous que le serveur hôte de vos serveurs virtuels offre une capacité suffisante et que les serveurs virtuels disposent d'un espace libre supérieur au minimum requis par les connecteurs :

- 2 Go de mémoire vive par gestionnaire de connecteurs (vous pouvez sélectionner une taille pendant l'installation)
- 20 Go d'espace sur disque dur
- Processeur à 2,4 GHz

Veillez également à inclure le paramètre `filter=p` ou `filter=0` dans les requêtes de recherche envoyées au système GSA. Cela réduit le nombre de requêtes d'autorisation nécessaire à chaque requête de recherche. Pour plus d'informations sur les paramètres des requêtes, consultez la [référence du protocole de recherche](#).

Nous vous recommandons de contrôler les ressources du processeur, de la mémoire et des disques pour corriger les problèmes de capacité liés au connecteur ou au serveur sous-jacent.

Bonnes pratiques de création d'un connecteur

Les connecteurs utilisant la bibliothèque Connector Framework Library mettent en œuvre un ensemble d'interfaces de programmation. Connector Framework Library est une bibliothèque Open Source. Cela

signifie qu'au-delà des connecteurs fournis par Google et ses partenaires, vous pouvez utiliser des connecteurs personnalisés pour étendre la portée de votre déploiement à toutes les sources de contenu dont vous avez besoin.

En tant que structure de création de connecteurs personnalisés, le gestionnaire de connecteurs Java offre les avantages suivants :

- Interfaces bien conçues pour définir des protocoles de balayage et de sécurité
- Exécution intégrée de plusieurs fils d'exécution simultanés pour optimiser le balayage, la planification et l'équilibrage de la charge sur l'hôte
- Classes utilitaires pour gérer les tâches courantes
- Structure entièrement compatible et régulièrement mise à jour pour tirer parti des dernières fonctionnalités de Google Search Appliance

Voici quelques-unes des bonnes pratiques à adopter pour créer un connecteur :

- Commencer par le connecteur "Hello World" présenté dans le [guide du développeur de connecteurs](#)
- Toujours répondre aux trois questions suivantes :
 - Comment (et en combien de temps) puis-je réaliser l'intégration initiale des contenus dans GSA ?
 - Comment effectuer le suivi des documents modifiés et ajoutés et dans quels délais seront-ils découverts ?
 - Comment retirer de l'index les documents supprimés ?
- Le préfixe `googleconnector://` permet au système GSA d'identifier l'instance du connecteur ayant émis l'autorisation d'accès à certaines URL de l'index. Vous avez la possibilité de modifier le gestionnaire de connecteurs de manière à envoyer de véritables URL, à la place de ces URL artificielles. Cela pourrait améliorer la pertinence des résultats.
- Le contrôle des autorisations d'accès par le connecteur n'est pas obligatoire. Il est parfaitement possible de confier cette tâche à un fournisseur SAML, par exemple. Cela peut permettre de simplifier l'architecture et d'améliorer les performances. Ainsi, vous pouvez contrôler les autorisations d'accès à l'aide d'un service SAML ou d'un service Web personnalisé fourni par un serveur d'applications existant.
- Le cas échéant, le système utilisera les LCA par URL définies pour ces URL pour contrôler les autorisations d'accès.
- En règle générale, l'authentification par le connecteur n'est pas nécessaire, car la plupart des systèmes de contenu réalisent une forme d'authentification silencieuse. Il ne reste plus alors au connecteur qu'à contrôler les autorisations d'accès aux documents.

Chapitre 3 Flux, robot d'exploration et proxy, structure d'adaptateurs ou structure de connecteurs ?

Présentation

Vous aurez peut-être à déterminer la solution la mieux adaptée à vos besoins parmi les suivantes : [structure de connecteurs Google](#), structure d'adaptateurs Google (Plexi), exploration via un proxy ou client de gestion des flux. Voici quelques pistes pour vous aider à arrêter votre choix :

Utilisation de la structure de connecteurs

La structure de connecteurs Google est basée sur les flux et offre les principales fonctionnalités suivantes :

- Configuration, planification, intégration dans l'interface utilisateur, et plus encore
- Construction spécifique des URL permettant au système GSA de diriger les requêtes d'autorisation vers les connecteurs (à savoir, `googleconnector://`)

La structure de connecteurs offre de nombreuses fonctionnalités utiles, mais la mise en œuvre d'un connecteur personnalisé pour la première fois implique un apprentissage poussé. Ainsi, le développement d'un connecteur personnalisé est indiqué dans les cas suivants :

- Nécessité de configurer des autorisations à liaison tardive avec un système de contenu offrant une API pour contrôler les autorisations d'accès des utilisateurs aux documents
- Réutilisation du connecteur dans d'autres déploiements du système GSA

Si vous n'exploitez pas la structure de connecteurs dans son ensemble, vous pouvez tout de même mettre en œuvre certains de ses utilitaires dans votre solution basée sur des flux, notamment le gestionnaire de flux de GSA décrit dans la section suivante. La structure de connecteurs est disponible à l'adresse <http://code.google.com/p/google-enterprise-connector-manager/>

Utilisation de la structure d'adaptateurs (Plexi)

La structure d'adaptateurs, que nous appelons Plexi en interne, est la nouvelle version de la structure de connecteurs. Cette dernière a été entièrement réécrite pour créer une version nouvelle qui se distingue principalement de la précédente par son architecture désormais basée sur le modèle [Répertoire/Récupérer](#).

L'architecture **répertoire/récupérer** des adaptateurs offre de nombreux avantages par rapport aux flux lorsque vous avez la possibilité d'accéder au contenu de façon aléatoire : simplicité, évolutivité, autoréparation, fonctionnement sans état et transparent (surveillance et débogage faciles) et utilisation de l'infrastructure existante.

Cette [FAQ destinée aux développeurs](#) répond à différentes questions relatives au développement basé sur la structure d'adaptateurs.

Pour consulter le référentiel de la structure d'adaptateurs, cliquez sur ce lien : <https://code.google.com/p/plexi/>

Utilisation du robot d'exploration via un proxy

Il est possible d'associer un proxy au robot d'exploration dans les deux situations suivantes :

- Comme prérequis permettant aux systèmes de contrôler le robot d'exploration GSA et d'utiliser la mémoire cache du proxy pour éviter les surcharges réseau ou simplement pour contrôler le robot d'exploration GSA.
- Comme outil permettant de réécrire du code HTML à la volée et d'associer aux métadonnées ou au contenu de nouvelles informations extraites du contenu lui-même ou d'autres sources telles que des bases de données.

Le proxy est particulièrement utile au processus d'intégration des contenus dans cette dernière situation, en tant que proxy de réécriture, notamment lorsqu'il n'y a pas de métadonnées à traiter. Vous pouvez extraire les métadonnées du contenu et les ajouter sous forme de métadonnées normales à partir de l'en-tête du document, les [microformats](#) sont généralement traités de cette manière.

Utilisation d'un client de flux

Les flux peuvent s'avérer utiles comme alternative à la création de connecteurs personnalisés. Étant plus faciles à créer, ils peuvent être adaptés dans les situations suivantes :

- Vous pouvez facilement identifier les contenus nouveaux ou modifiés par le biais de requêtes.
- Vous utilisez un déclencheur d'événements ou un système de surveillance pour les ajouts ou les modifications de contenu.
- Le système de contenu n'est pas compatible avec les API clientes Java.
- Le langage de développement choisi n'est pas Java.
- La mise en œuvre consiste en une intégration unique (autrement dit, l'indexation régulière n'est pas requise).
- Les documents sont considérés comme publics et non sécurisés ou les autorisations d'accès sont basées sur des LCA par URL.

Pour transmettre un flux au système de recherche, vous devez utiliser un client de flux. Le gestionnaire de flux fait partie des outils permettant d'envoyer des données au système GSA par le biais de flux. Pour plus d'informations sur le gestionnaire de flux, consultez la page <http://code.google.com/p/gsafeedmanager/>. Le gestionnaire de flux GSA n'est pas inclus dans l'assistance Google for Work.

Pour plus d'informations sur les flux, consultez le [guide du protocole des flux destiné aux développeurs](#).

Chapitre 4 Connecteur SharePoint de Google Search Appliance

Présentation

Le connecteur Microsoft SharePoint permet au système Google Search Appliance d'indexer des fichiers de contenu et des métadonnées stockés dans Microsoft SharePoint afin d'y effectuer des recherches. Pour utiliser le connecteur Microsoft SharePoint, vous avez besoin des composants suivants :

- Connecteur exécuté en interne ou sur un hôte distinct
- Ensemble de services Web personnalisés appelé Services Google pour SharePoint
- [Facultatif] Champ de recherche Google pour SharePoint, qui remplace le champ de recherche intégré dans le frontal Web de SharePoint

Si vous n'avez pas la possibilité d'activer Kerberos sur le système Google Search Appliance, vous aurez également besoin de SAML Bridge pour Windows.

1. En règle générale, le processus de déploiement se déroule comme suit :
2. Installez les Services Google pour SharePoint sur les frontaux Web de SharePoint.
3. Installez et configurez le connecteur SharePoint, puis vérifiez que le contenu est bien indexé par le système GSA.
4. Configurez la sécurité lors de la recherche (authentification et autorisation d'accès), puis vérifiez le bon fonctionnement de l'authentification silencieuse, si nécessaire.
5. (Facultatif) Si vous souhaitez que les utilisateurs puissent se servir de Google Search Appliance sans quitter l'interface SharePoint, installez le champ de recherche Google pour SharePoint, puis vérifiez que la recherche sécurisée à partir des frontaux Web de SharePoint fonctionne correctement.

Balayage

Les performances du balayage, qui correspondent au temps nécessaire au connecteur pour réaliser un balayage lors du déploiement initial, peuvent varier en fonction de différents facteurs, y compris :

- la vitesse de balayage définie ;
- les performances du serveur qui héberge le connecteur SharePoint ;
- les performances du frontal Web SharePoint ;
- le nombre de documents, le nombre de collections de sites et la profondeur des sites.

En moyenne, le connecteur parvient à balayer près d'un million de documents par jour. Pour accroître la vitesse de balayage, vous pouvez :

- définir une vitesse de balayage supérieure ;
- configurer plusieurs instances du connecteur.

Augmentation de la vitesse de balayage

Assurez-vous que la mémoire disponible sur le serveur hôte est suffisante en surveillant la consommation de celui-ci. Par défaut, la mémoire allouée s'élève à 2 Go. Vous pouvez augmenter cette valeur si ce n'est pas suffisant. Vérifiez également que le serveur SharePoint sera en mesure de traiter la charge que représentent les requêtes.

Configuration de plusieurs instances du connecteur

Utilisez les formats d'exclusion disponibles sur la page de configuration pour définir plusieurs instances du connecteur traitant chacune une partie des sites. Par exemple, si vous avez des centaines ou des milliers de collections de sites, répartissez les collections principales sur les différentes instances du connecteur ; l'instance A traitant les sites commençant par la lettre a, b ou c, et l'instance B traitant les sites commençant par la lettre d, e ou f.

Liaison précoce de groupes

À partir de la version 3.0, Google Search Appliance inclut un nouveau connecteur de groupes Active Directory. Ce connecteur n'est pas une solution de récupération de contenus, mais une solution de sécurité. Il remplace la méthode de liaison tardive utilisée dans les versions précédentes pour la résolution de groupes par une liaison précoce qui consiste à balayer les groupes Active Directory et à les stocker dans sa propre base de données.

Le connecteur SharePoint 3.0 intégré comprend donc un connecteur de groupes Active Directory. Cependant, le processus d'autorisation d'accès par LCA requiert les groupes locaux SharePoint, en plus des groupes Active Directory. Le connecteur SharePoint se charge de récupérer et de stocker ces deux types de groupes dans une base de données. Pour un domaine Windows ou une batterie de serveurs SharePoint de grande envergure, le temps nécessaire à cette opération vient s'ajouter à celui du balayage.

Où est mon fichier ?

En parcourant l'ensemble des collections de sites ou les sous-sites, le connecteur SharePoint effectue un balayage vertical. Si l'arborescence de la collection de sites est vraiment très profonde, cela retarde l'intégration des contenus appartenant à la collection suivante.

Avant de tenter de contrôler qu'un document particulier a été indexé, accédez au répertoire d'une instance de connecteur, puis vérifiez la valeur du paramètre `FullRecrawlFlag` d'un site Web dans le fichier `sharepoint_state.xml`. Un fichier XSL est à votre disposition pour rendre le fichier d'état plus lisible. Vous pouvez télécharger le fichier (`SharePointStateReport.xslt`) à partir de l'URL suivante : <http://code.google.com/p/google-enterprise-connector-sharepoint/downloads/list>

Lorsque vous définissez des formats d'exclusion dans le connecteur SharePoint, le temps nécessaire à l'indexation d'un petit nombre de documents peut s'avérer très long si la plupart des documents découverts sont ignorés.

Si vous utilisez des flux métadonnées-et-URL, vous pouvez aisément retrouver un fichier dans la section **Indexer > Diagnostic > Diagnostics d'index** de la console d'administration GSA. Si vous utilisez des flux de contenu, les URL de SharePoint ne sont pas utilisées comme URL d'enregistrement dans les flux. Les URL du serveur SharePoint apparaissent toutefois comme URL d'affichage. Par défaut, les éléments des listes SharePoint sont indexés avec la syntaxe

googleconnector://...?docid=http://...AllItems.aspx?<ItemID> et le système GSA les explore comme des documents distincts. Pour rechercher des documents dans la section **"Diagnostics d'index"**, vous devez trouver l'URL d'enregistrement par le biais de l'URL d'affichage. À partir de GSA 6.8, vous pouvez utiliser l'URL d'affichage avec les opérateurs "site:" et "inurl:". Tout d'abord, effectuez une recherche avec la syntaxe suivante : `inurl:<fragment de l'URL SharePoint>`. Assurez-vous que vous êtes autorisé à accéder au fichier concerné. Localisez le fichier dans les résultats de recherche, puis cliquez sur le lien "En cache". En haut de la page en cache, vous trouverez l'URL d'enregistrement qui commence par "googleconnector://". Vous pouvez alors utiliser cette URL pour localiser le document dans la section **"Diagnostics d'index"**.

Pour savoir si un document a été découvert, mais exclu, vous pouvez consulter le fichier `excluded_url.txt` dans le répertoire de l'instance du connecteur afin de vérifier que le connecteur n'exclut aucun élément que vous souhaiteriez inclure dans l'index.

Le connecteur SharePoint et GSA utilisent des bibliothèques différentes pour analyser les formats à base d'expressions rationnelles. Veillez à vous reporter à la documentation pour écrire vos formats d'inclusion ou d'exclusion.

Description

Le contenu SharePoint peut être traité directement depuis Google Search Appliance, ou depuis le frontal Web de SharePoint si vous y installez le champ de recherche Google pour SharePoint.

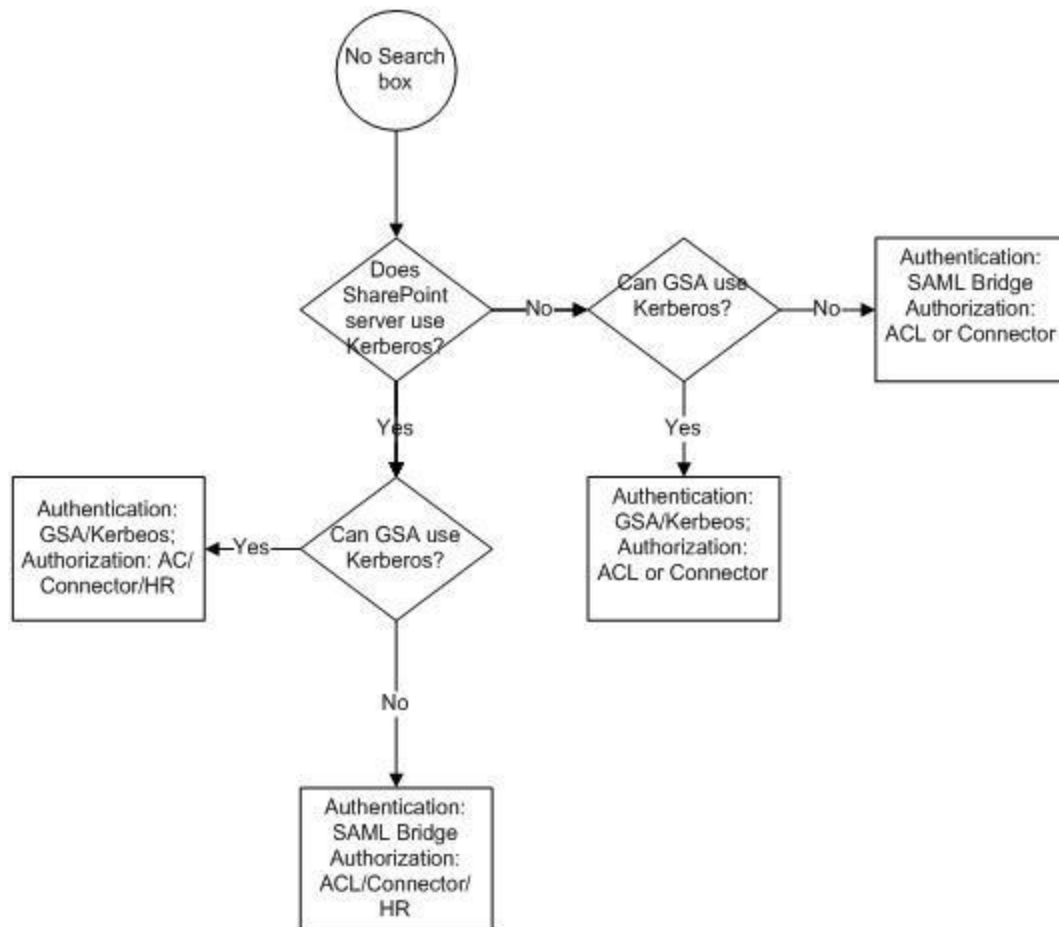
Champ de recherche Google pour SharePoint

Le champ de recherche Google pour SharePoint offre une expérience intégrée et une apparence similaire à la fonction de recherche native de SharePoint. Cependant, il ne transmet pas les requêtes au moteur de recherche de SharePoint, mais au système de recherche. Il réalise l'authentification de manière silencieuse pour les utilisateurs authentifiés lorsque le protocole de sécurité mis en œuvre est l'authentification intégrée Windows, et inclut également les recherches publiques (documents non sécurisés).

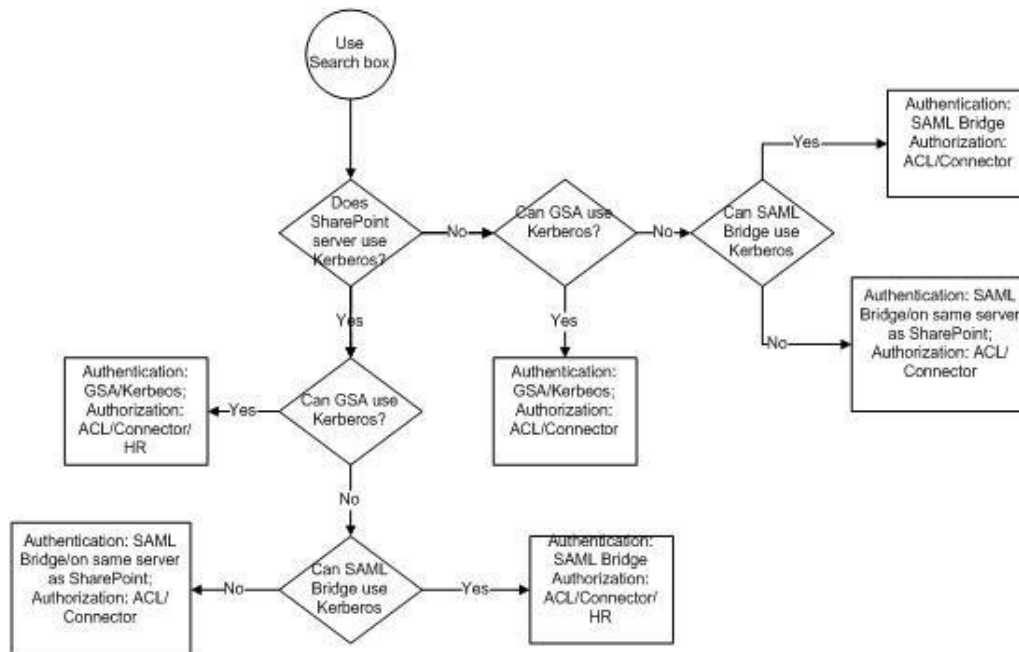
Recherche sécurisée

L'authentification silencieuse est une exigence très courante en matière de recherche sécurisée. Elle peut être mise en œuvre de différentes manières. Les deux diagrammes ci-dessous indiquent les composants à utiliser dans ce cadre.

Le diagramme ci-dessous représente l'authentification silencieuse sans champ de recherche pour SharePoint (l'abréviation HR fait référence à une requête HTTP HEAD).



Le diagramme suivant représente l'authentification silencieuse avec le champ de recherche pour SharePoint. La mention "No Kerberos" signifie que le protocole NTLM est utilisé.



Pour plus d'informations sur le champ de recherche Google pour SharePoint, consultez la page de wiki [google-enterprise-connector-sharepoint](#). Pour plus d'informations sur SAML Bridge pour Windows, reportez-vous à l'article relatif à l'[activation de l'authentification intégrée Windows](#).

Performances de traitement

Pour bénéficier de performances de traitement optimales, veuillez à observer les consignes suivantes (par ordre de préférence) :

- Utilisez des LCA pour les autorisations d'accès. Cette fonctionnalité est disponible depuis la version 2.8 et elle peut offrir les meilleures performances. À partir de sa version 3.0, le connecteur réalise la résolution des groupes (Active Directory et SharePoint) par liaison précoce grâce au connecteur de groupes Active Directory de GSA. Ces nouvelles fonctionnalités améliorent largement les performances en matière d'autorisation.
- Utilisez le connecteur pour traiter les autorisations plutôt que des requêtes HEAD, si possible (qu'il soit configuré avec Kerberos ou SAML Bridge, le système GSA peut envoyer des requêtes HEAD).
- Si vous utilisez SAML Bridge pour traiter les autorisations, veuillez à activer l'option "**Utiliser des requêtes d'autorisation SAML par lots**" sur la page **Rechercher > Recherche sécurisée > Contrôle d'accès**.

Connecteur interne ou externe ?

Nous vous recommandons de toujours utiliser le connecteur SharePoint 3.0 de façon externe avec une base de données externe.

Nous invitons également les partenaires GSA officiels à consulter les [points essentiels concernant GSA](#) sur Google for Work Connect (dans la section relative au déploiement).

Chapitre 5 Connecteur de systèmes de fichiers de Google Search Appliance

Présentation

Google recommande l'utilisation du connecteur de systèmes de fichiers pour indexer des partages de fichiers. La passerelle intégrée pour systèmes de fichiers n'est plus disponible depuis la version 7.2.

Le connecteur de systèmes de fichiers a fait l'objet de modifications importantes dans la version 3.2 et il fonctionne assez différemment des autres connecteurs.

Balayage

La version 3.0 du connecteur est basée sur le modèle "répertoire/récupérer" : il découvre les documents, puis envoie les propriétés et les LCA des fichiers au système GSA à l'aide de flux métadonnées-et-URL. GSA explore les fichiers via le connecteur de systèmes de fichiers. La fréquence des explorations et le fonctionnement global sont les mêmes que pour l'exploration du Web.

Où est mon fichier ?

Dans la version 3.0, le connecteur de systèmes de fichiers alimente GSA en fichiers à l'aide d'une URL dont le format est le suivant :

`http://<hôte du connecteur:port>/connector-manager/getDocumentContent?ConnectorName=<nom du connecteur>&docid=smb://<chemin d'accès au fichier>`. Notez que l'URL ne commence plus par "googleconnector://".

Sachant que le lien se termine par le véritable chemin d'accès au fichier, le terme de requête `filetype` fonctionne. Cela signifie également que vous pouvez utiliser des "**Formats à ne pas suivre**" pour exclure certaines extensions de fichier.

Le meilleur moyen de vérifier qu'un fichier a bien été indexé est d'accéder à la page **Indexer > Diagnostic > Diagnostics d'index** ou d'effectuer une recherche de type `inurl:<chemin d'accès au fichier>`.

Le connecteur de systèmes de fichiers filtre également les fichiers en fonction de leur type MIME, en comparant ce dernier avec les types MIME exclus dans les paramètres du connecteur.

Description

Le lien d'accès aux fichiers figurant dans les résultats de recherche commence par "`file://`". Ce lien ne permet pas toujours d'ouvrir le fichier. Si vous souhaitez qu'un clic sur le lien déclenche systématiquement l'ouverture du fichier, vous devez personnaliser le frontal de GSA. Veillez à tenir compte des différences entre les navigateurs.

Recherche sécurisée

Le connecteur de systèmes de fichiers transmet les entrées de LCA associées à chaque fichier qu'il découvre et des LCA par URL sont stockées dans l'index du système GSA. Le connecteur de systèmes de fichiers peut ainsi traiter des millions de documents.

À partir de la version 7.0, Google Search Appliance propose des fonctionnalités de refus et d'héritage de LCA.

Le connecteur de groupes Active Directory est nécessaire à la résolution des groupes et à l'utilisation des LCA du connecteur de systèmes de fichiers.

Connecteur interne ou externe ?

À partir de sa version 6.8, GSA inclut un connecteur de systèmes de fichiers interne que vous pouvez utiliser pour des déploiements légers. Cependant, l'utilisation d'un connecteur de systèmes de fichiers externe présente de nombreux avantages :

- Évolutivité pour traiter des partages de fichiers très volumineux
- Compatibilité avec la mise en miroir GSA
- Résolution des problèmes facilitée, le cas échéant
- Configurations avancées non disponibles via la console d'administration
- Nous invitons également les partenaires GSA officiels à consulter les [points essentiels concernant GSA](#) sur Google for Work Connect (dans la section relative au déploiement).

Chapitre 6 Connecteur de bases de données de Google Search Appliance

Présentation

Google Search Appliance propose deux méthodes d'exploration des bases de données :

- L'explorateur de bases de données intégré (abandonné à partir de la version 7.2)
- Le connecteur de bases de données

Il est préférable d'utiliser le connecteur de bases de données, plutôt que l'explorateur de bases de données intégré dans le système de recherche. Le connecteur fournit deux fonctionnalités importantes qui ne sont pas disponibles dans l'explorateur de bases de données intégré.

- Recherche sécurisée
- Exploration planifiée

Balayage

Le connecteur de bases de données peut fonctionner dans différents scénarios tels que :

- Chaque ligne est traitée comme un document distinct.
- La colonne des résultats de la requête contient un champ BLOB/CLOB.
- Une URL est associée à un enregistrement de base de données.

Chaque ligne est traitée comme un document distinct.

Ce scénario implique que chaque ligne soit traitée comme un document distinct lors de l'indexation des enregistrements de la base de données. Notez que vous pouvez utiliser la requête du connecteur pour regrouper de nombreux enregistrements de base de données dans un enregistrement "unique" du point de vue du connecteur.

Ce mode de traitement des bases de données est généralement choisi, plutôt que le module OneBox, essentiellement pour les deux raisons suivantes :

- Les enregistrements de base de données indexés offrent une meilleure pertinence.
- Sachant que le traitement du contenu balayé par le connecteur de bases de données n'implique aucun envoi de requête à la base de données, cette méthode offre généralement des performances bien meilleures, car les résultats sont extraits directement de l'index.

Voici comment les enregistrements sont indexés dans ce scénario :

- Chaque ligne est convertie en fichier XML, chaque nom de colonne correspondant à un élément XML.
- Les flux de contenus sont générés et l'URL principale est au format suivant :
`googleconnector://localhost/<nom de l'hôte>/DocID`
Le paramètre DocID est un codage base64 des clés primaires de la requête de balayage.
- Les URL d'affichage sont au format suivant :
`dbconnector://<nom de l'hôte>/<nom de la base de données>/<docId>`

Avant que le document soit envoyé au système de recherche, une feuille de style XSL est appliquée pour transformer le fichier XML en fichier HTML. Le connecteur comprend une feuille de style par défaut que vous pouvez personnaliser en fonction de vos besoins, par exemple :

- Pour indexer certaines colonnes sous forme de métadonnées par transformation des éléments XML associés en en-têtes de métadonnées HTTP.
- En respectant les bonnes pratiques d'optimisation du référencement, pour créer des fichiers HTML permettant d'attribuer une pertinence supérieure à certaines colonnes de la base de données (par exemple, en-têtes HTML, texte en gras, tailles de police).

La colonne des résultats de la requête contient un champ BLOB/CLOB.

Lorsque la colonne des résultats de la requête comprend un champ BLOB/CLOB, les flux de contenu sont générés en traitant les données du champ BLOB/CLOB comme un document et les autres colonnes comme des métadonnées. Le type MIME du document est ignoré si les paramètres du gestionnaire de contenu indiquent de l'exclure.

Une URL est associée à un enregistrement de base de données.

Lorsqu'une URL est associée à un enregistrement de base de données, le connecteur fournit des flux métadonnées-et-URL.

Vitesse de balayage

Le connecteur de bases de données effectue le suivi des modifications des enregistrements en prenant un instantané de chaque fichier. Lorsqu'il exécute des requêtes sur de grands ensembles de données, le connecteur de bases de données renvoie des résultats par lots. Si la vitesse de balayage est trop faible, l'opération risque de durer trop longtemps. Si la vitesse de balayage est trop élevée, la requête peut se bloquer. Il est donc important de trouver la vitesse de balayage adéquate. Google vous recommande de définir une vitesse de balayage comprise entre 1 000 et 1 500 documents par minute.

Où est mon enregistrement ?

L'identifiant des documents indexés correspond à la somme de contrôle des valeurs de la clé primaire. Par conséquent, il n'y a pas de moyen direct de retrouver un enregistrement dans les diagnostics de l'exploration. Le meilleur moyen d'y parvenir consiste à rechercher les valeurs directement.

Description

Lorsque chaque enregistrement est traité comme un document distinct et que des flux de contenus sont envoyés, l'URL d'affichage commence par "dbconnector://". Dans le frontal par défaut, ils sont transformés en liens vers les résultats en cache. Aucune requête n'est envoyée au système GSA lors de la présentation des résultats.

Recherche sécurisée

Lorsque le connecteur de bases de données est configuré pour envoyer des flux métadonnées-et-URL, l'autorisation d'accès aux documents ne passe pas par le connecteur.

Le connecteur de bases de données traite l'autorisation à l'aide d'une requête de flux de contenus à la base de données, ce qui requiert une identité validée. Vous pouvez procéder de différentes manières pour obtenir une identité validée. Pour plus d'informations sur les identités validées, reportez-vous à l'article concernant la [gestion de la recherche pour des contenus dont l'accès est contrôlé](#). L'administrateur doit spécifier la requête d'autorisation avec des paramètres substituables pour le nom d'utilisateur et l'identifiant du document.

Lorsqu'une instruction SQL d'autorisation est configurée, tous les enregistrements du flux sont associés à l'attribut `authmethod="httpbasic"` indiquant qu'il s'agit d'un contenu sécurisé.

Le connecteur de bases de données ne fournit pas d'interface pour son authentification dans le cadre de la recherche sécurisée. L'authentification doit être effectuée par d'autres composants du système GSA tels que Kerberos ou l'altération des cookies. Sinon, vous devez personnaliser le connecteur de manière à fournir une implémentation de la SPI `AuthenticationManager`.

Pour plus d'informations sur la configuration de la recherche sécurisée pour le connecteur, reportez-vous à la page de wiki [google-enterprise-connector-database](#).

Chapitre 7 Connecteur Lotus Notes de Google Search Appliance

Présentation

Le connecteur Lotus Notes version 3.0 constitue une mise à jour importante par rapport aux versions précédentes. Ce nouveau connecteur :

- s'exécute dans le gestionnaire de connecteurs et dans Tomcat ;
- utilise le client Notes pour accéder au serveur Domino ;
- n'utilise pas de gestionnaire d'agents ni de serveur HTTP sur le serveur Domino ;
- utilise les LCA hiérarchiques de GSA pour les LCA de la base de données de l'application Notes et les noms de lecteur du document ;
- utilise une seule base de données de configuration dans Domino ;
- optimise l'évolutivité pour traiter des millions de documents ;
- ne stocke plus de métadonnées ni de rapports dans la base de données de configuration.

Balayage

Le balayage est effectué par le gestionnaire de connecteurs. Cependant, les documents sont placés temporairement en file d'attente dans la base de données de configuration et sur le système de fichiers avant d'être envoyés au système GSA :

1. Le connecteur interroge les bases de données configurées pour obtenir la liste des documents modifiés/mis à jour, puis crée un stub contenant l'identifiant du document dans la file d'attente d'exploration de la base de données du connecteur.
2. Les fils d'exécution du robot d'exploration récupèrent les documents et les pièces jointes dans la base de données source et procèdent à la transformation des métadonnées, puis placent les documents dans la file d'attente d'envoi.
3. Le gestionnaire de connecteurs crée un flux à partir des documents de la file d'attente d'envoi, qu'il transmet au système GSA. Ensuite, il ne conserve dans sa base de données H2 locale que quelques informations relatives au document envoyé.

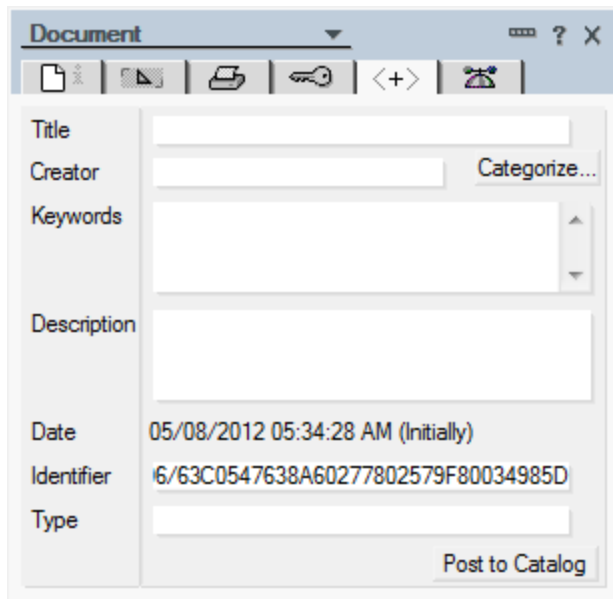
Où est mon document ?

Le connecteur interroge les bases de données selon la planification établie dans la console d'administration GSA. Vous pouvez également indiquer dans la base de données de configuration la fréquence à laquelle une base de données particulière doit être interrogée.

Procédez comme suit pour vérifier qu'un document a bien été traité :

- Réglez le niveau de journalisation sur TOUT afin de consigner toutes les informations de diagnostic dans les journaux.
- Redémarrez le balayage de la base de données en procédant comme suit :
 - a. Ouvrez la base de données de configuration du connecteur.
 - b. Ouvrez le document de configuration de la base de données.
 - c. Cliquez sur "**Modifier**".
 - d. Cliquez sur "**Redémarrer le balayage**".
 - e. Cliquez sur "**Enregistrer**", puis sur "**Quitter**".

- Pour trouver l'UNID du document, ouvrez-le dans le client Notes, ouvrez la boîte de dialogue des propriétés du document, puis accédez à l'onglet d'information <+>. Dans l'identifiant, l'UNID est la chaîne qui suit le dernier caractère /.



- Recherchez l'UNID du document dans le journal afin de vérifier qu'il a bien été envoyé au système GSA.
- Accédez à la page **Sources de contenu > Flux** de la console d'administration pour vous assurer que les flux ont fonctionné correctement.
- Consultez la page **Indexer > Diagnostic > Diagnostics d'index** pour rechercher les documents qui auraient dû être explorés.

Enfin, si une base de données a été traitée et que certains documents n'ont pas été indexés, procédez comme suit :

1. Dans le modèle configuré pour cette base de données, vérifiez que la formule de sélection inclut le document en question.
2. Assurez-vous que l'identifiant utilisateur Notes que vous utilisez est autorisé à accéder au document (ouvrez le document dans le client Notes en utilisant cet identifiant utilisateur Notes).
Le connecteur Notes n'indexe pas les documents marqués comme "Conflits de réplication ou d'enregistrement".

Description

Par défaut, les résultats de recherche sont présentés selon une feuille de style XSLT interne qui établit le lien vers les documents à l'aide du protocole HTTP. Sachant que le lien utilisé par le protocole Notes est inclus dans un champ de métadonnées des résultats de recherche, il est possible de personnaliser la feuille de style XSLT de manière à utiliser ce lien dans les résultats de recherche. Cette méthode permet d'ouvrir les résultats de recherche dans le client Notes, à condition que ce dernier soit installé sur l'ordinateur de l'utilisateur (et enregistré dans le navigateur). Voici une illustration de cette méthode.

```
<xsl:when test="starts-with(U, $db_url_protocol)">
  <xsl:value-of disable-output-escaping='yes'
    select="concat('db/', $temp_url)"/>
</xsl:when>

<!-- *** DÉBUT : ajout pour le remplacement des URL Notes *** -->
<xsl:when test="starts-with(lower-case($stripped_url), 'newyork')">
  <xsl:value-of select="MT[@N='dom_noteslink']/@V"/>
</xsl:when>
<!-- *** FIN : ajout pour le remplacement des URL Notes *** -->

<!-- *** Échappement des URI pour smb ou NFS nécessaire, car elles apparaissent
dans la requête d'URI *** -->
<xsl:when test="$protocol='nfs' or $protocol='smb'">
  <xsl:value-of disable-output-escaping='yes'
    select="concat($protocol, '/', $temp_url)"/>
</xsl:when>
```

Recherche sécurisée

Le mécanisme de recherche sécurisée par défaut utilise les autorisations fournies par le connecteur pour autoriser les résultats de recherche au moment de leur présentation.

Il est également possible d'envoyer des LCA au système GSA pour mettre en œuvre une liaison précoce permettant des recherches sécurisées beaucoup plus rapides. Cette méthode présente toutefois les limites suivantes :

- Avec les versions 6.14 et antérieures de GSA, seule la LCA de la base de données est envoyée. La sécurité des documents liée aux champs du lecteur et de l'auteur n'est pas respectée.
- Les entrées "No Access" de la LCA de la base de données ne sont pas compatibles.

Conseils et astuces

Ne pas utiliser d'entrées "No Access" dans la liste de contrôle d'accès de la base de données

Pour le moment, les entrées "No Access" des LCA ne sont pas compatibles avec la liaison précoce (LCA par URL). Pour remédier à ce problème, définissez les accès comme suit :

- Utilisez "No Access" comme paramètre par défaut.
- Ajoutez des groupes à la LCA et attribuez-leur des autorisations d'accès appropriées.

Ne pas indexer les champs à visibilité limitée

Pour des questions de cohérence, n'incluez pas de champs dont les paramètres de sécurité sont masqués dans l'interface utilisateur de Notes. Pour ce faire, modifiez le modèle utilisé pour l'exploration de la base de données concernée.

Désactiver le filtrage

Pour afficher tous les documents d'une base de données et améliorer les performances de traitement, définissez le filtrage comme suit : `filter=0` ou `filter=p`.

Chapitre 8 Connecteur Livelink de Google Search Appliance

Présentation

Le connecteur OpenText Livelink permet au système GSA d'indexer les fichiers de contenu et les métadonnées stockés dans un ECM OpenText Livelink et d'y effectuer des recherches. Le connecteur formate le contenu et les métadonnées du référentiel pour les transmettre au système GSA sous forme de flux de contenus.

Le connecteur indexe les mêmes éléments que Livelink, à savoir les documents, les dossiers, les projets et les discussions. Vous pouvez configurer avec précision les types de fichiers et les emplacements à indexer. Pour les documents, seules les métadonnées et la version actuelle du fichier d'origine sont indexées. Les précédentes versions et interprétations ne sont pas indexées.

Pour plus d'informations sur ce connecteur, consultez les ressources suivantes :

- [Configuration du connecteur Livelink](#)
- [Connecteur Livelink de Google Search Appliance](#)

Balayage

Le connecteur balaye le contenu Livelink via l'API Livelink en utilisant le nom d'utilisateur et le mot de passe de balayage spécifiés dans la configuration du connecteur. Vous pouvez contrôler la portée de récupération des contenus Livelink de deux manières :

- Avec les identifiants d'objet spécifiés dans les configurations du connecteur
- En limitant l'accès au contenu pour l'utilisateur associé au balayage

Si vous optez pour cette dernière méthode, ne renseignez aucun identifiant d'objet de sorte que le connecteur récupère tous les éléments auxquels l'utilisateur associé au balayage a accès.

Vous pouvez configurer le balayage effectué par le connecteur Livelink sur le référentiel Livelink de différentes manières. Par défaut, dans la version 2.6.10, le connecteur Livelink fait appel à une méthode de balayage basée sur un tableau du système Livelink appelé DtreeAncestors. Cette méthode offre des performances de balayage optimales, mais elle est dépendante d'un tableau (DtreeAncestors) qui peut ne pas être à jour ou contenir des informations incorrectes.

En raison de ces limitations, Google vous recommande de configurer le connecteur de sorte qu'il utilise une méthode de balayage "généalogique" moins rapide, mais plus fiable, en particulier la méthode BatchGenealogist qui est la plus performante des méthodes de balayage généalogique. Pour plus d'informations, reportez-vous à la [configuration avancée](#).

Où est mon enregistrement ?

Si des contenus ne semblent pas apparaître dans le système GSA, procédez comme suit pour corriger le problème :

- Accédez à la page "Flux" de la console d'administration GSA pour vérifier que les flux fonctionnent correctement.
- Dans la console d'administration GSA, consultez les diagnostics d'exploration des pages que vous recherchez. Pour ce faire, vous aurez besoin de l'identifiant d'objet Livelink des documents en question, car ils sont inclus dans les URL des diagnostics d'exploration.
- Vérifiez que les journaux du connecteur n'indiquent pas d'erreur connexe. Si nécessaire, augmentez le niveau de journalisation du connecteur, comme indiqué sur la page relative à la [configuration du connecteur Livelink](#).

Description

Vous pouvez utiliser la feuille de style XSLT par défaut pour formater les résultats de recherche et la configurer pour afficher les métadonnées associées au contenu (à condition que le connecteur soit configuré pour les récupérer).

Lorsque vous configurez le connecteur Livelink, vous spécifiez l'URL Livelink des résultats de recherche. Cette URL se présente généralement comme suit : `http://host::port/Livelink/livelink.exe`

Il s'agit de l'URL standard qui permet aux utilisateurs d'accéder à Livelink. Notez que l'URL Livelink doit comprendre un nom d'hôte complet pour que les liens "En cache" et "Version textuelle" des résultats de recherche fonctionnent.

Recherche sécurisée

Le connecteur permet l'authentification sur la base de données Livelink et sur un service d'annuaire externe. Il obtient l'autorisation du serveur Livelink lorsqu'un utilisateur demande l'affichage d'un élément donné et du contenu associé. Le système GSA ne requiert pas de configuration particulière pour utiliser les mécanismes d'authentification utilisateur et d'autorisation de Livelink.

Pour le moment, ce connecteur ne vous permet pas de transmettre des LCA au système GSA et n'est donc pas compatible avec le mécanisme d'autorisation par liaison précoce.

Chapitre 9 Connecteur EMC Documentum de Google Search Appliance

Présentation

Le connecteur EMC Documentum formate le contenu et les métadonnées d'un référentiel EMC Documentum pour les transmettre au système Google Search Appliance.

Par défaut, le connecteur explore et indexe les fichiers de contenu associés au type d'objet `dm_document`. Vous avez la possibilité d'indexer les autres sous-types des objets `dm_sysobject` et `dm_document`, ainsi que des sous-types personnalisés, et d'y effectuer des recherches, mais vous devrez pour cela ajouter manuellement les types voulus à la page "**Configuration avancée**" lorsque vous configurez un connecteur.

Pour plus d'informations, reportez-vous à l'article relatif à la [configuration du connecteur pour Documentum](#).

Où est mon enregistrement ?

Si des contenus ne semblent pas apparaître dans le système GSA, posez-vous les questions suivantes :

- Le contenu a-t-il été ignoré par la requête ? Vérifiez que le type d'objet en question est configuré pour l'exploration et que votre contenu n'est pas exclu par la clause `WHERE`.
- Le contenu a-t-il été ignoré par le connecteur ? Assurez-vous que le contenu n'a pas été exclu par le connecteur. Cela peut se produire pour différentes raisons :
 - Type MIME incompatible
 - Fichier trop volumineux
 - Erreur de récupération (par exemple, anciens types d'objet DFC)
- Le flux a-t-il été rejeté par GSA ? Vérifiez quels sont les hôtes de confiance dont GSA accepte les flux, ainsi que les formats d'URL à **suivre et explorer**, et assurez-vous que les flux sont bien traités par le système GSA.

Description

Vous pouvez utiliser la feuille de style XSLT par défaut pour formater les résultats de recherche et la configurer pour afficher les métadonnées associées au contenu (à condition que le connecteur soit configuré pour les récupérer).

Vous pouvez également définir le mode d'ouverture des résultats de recherche en modifiant l'élément URL de Webtop" sur la page de configuration du connecteur. Cela vous permet d'ouvrir les documents de différentes manières, par exemple :

- Ouverture pour consultation ou modification
- Consultation
- Téléchargement direct
- Accès aux propriétés du document

Pour plus d'informations, reportez-vous à l'article relatif à la [définition du format de l'URL Webtop](#).

Vous avez également la possibilité de créer une personnalisation WDK ou une application DFC légère afin de présenter le contenu hors contexte. Vous devez simplement faire en sorte que le connecteur accole l'identifiant de l'objet à l'URL de Webtop que vous fournissez et qui peut être totalement étrangère à Webtop. Vous devez également authentifier l'utilisateur et vérifier qu'il dispose des autorisations requises avant de renvoyer des documents, à moins que vous souhaitiez fournir un accès totalement ouvert aux documents.

Recherche sécurisée

Pour traiter le contenu sécurisé, le connecteur peut faire appel à n'importe quel mécanisme d'authentification des utilisateurs de Documentum.

Lors de la présentation des résultats, le connecteur demande les identifiants de l'utilisateur ayant envoyé la requête de recherche. Ces identifiants sont transmis aux serveurs de contenu qui authentifient l'utilisateur et déterminent quels résultats l'utilisateur est autorisé à consulter. Le système GSA ne requiert pas de configuration spécifique pour gérer les mécanismes d'autorisation et d'authentification des utilisateurs Documentum.

Lorsque vous configurez le connecteur dans la console d'administration, vous pouvez définir des contenus comme public en cochant "**Rendre public**" sur la page **Sources de contenu > Exploration du Web > Exploration sécurisée > Accès du robot d'exploration**. Le contenu en question sera ainsi accessible via la recherche publique, quelles que soient les autorisations d'accès définies dans le référentiel.

Pour le moment, ce connecteur ne vous permet pas de transmettre des LCA au système GSA et n'est donc pas compatible avec le mécanisme d'autorisation par liaison précoce.

Chapitre 10 Connecteur IBM FileNet de Google Search Appliance

Présentation

Le connecteur IBM FileNet permet au système GSA de balayer, d'indexer et de traiter le contenu d'un référentiel IBM FileNet. Pour plus d'informations, reportez-vous à l'article relatif à la [configuration du connecteur FileNet](#).

Pour le moment, le connecteur indexe uniquement les entités associées à la classe Document (et à ses sous-classes) ou à un processus Workflow.

Balayage

Le balayage commence par le document associé à la première date de modification. Une fois le balayage initial effectué, le connecteur fonctionne de façon incrémentielle pour indexer les documents ajoutés ou modifiés.

Pour savoir quels documents ont été explorés, vous pouvez consulter l'état du connecteur au point de contrôle. Pour ce faire, ouvrez le fichier

```
WEB-INF\...\<nom-instance-du-connecteur>_state.txt
```

Le point de contrôle comprend les éléments suivants :

- uuid : dernier document ajouté à l'index
- lastModified : date du dernier ajout de document
- uuidToDelete : dernier document supprimé de l'index
- lastRemoveDate : date de la dernière suppression de document

Où est mon enregistrement ?

L'acquisition des contenus étant effectuée à l'aide d'un flux, vous pouvez consulter l'état du connecteur et la page **Sources de contenu > Flux** dans la console d'administration GSA pour vous assurer que les flux sont traités correctement. Vous pouvez également contrôler chaque document sur la page **Indexer > Diagnostic > Diagnostics d'index** de la console d'administration. Pour localiser un document particulier, vous devez connaître son identifiant FileNet et le rechercher sur la page "**Diagnostics d'index**".

Description

Vous pouvez configurer les résultats de recherche fournis par le connecteur FileNet de sorte qu'ils pointent vers le document lui-même ou vers ses métadonnées, via le client FileNet Workplace. Pour plus d'informations sur le paramètre "**URL du bureau électronique**", reportez-vous aux [paramètres des instances de connecteur](#).

Recherche sécurisée

Le connecteur comprend un paramètre de configuration permettant de définir l'affichage des résultats publics et sécurisés. Pour plus d'informations sur le paramètre "**Rendre public**", reportez-vous aux [paramètres des instances de connecteur](#).

La recherche sécurisée, l'authentification et les autorisations d'accès sont gérées par le connecteur via les API FileNet Java. Le connecteur traite les authentifications et les autorisations d'accès en temps réel, par la méthode de liaison tardive. La liaison précoce n'est pas compatible avec le connecteur FileNet pour le moment.

Chapitre 11 Connecteur LDAP de Google Search Appliance

Présentation

Le connecteur LDAP existe depuis la version 6.8 du système GSA. Il met en œuvre un flux pour transmettre le contenu et les métadonnées LDAP à GSA.

Pour plus d'informations sur ce connecteur, reportez-vous à l'article relatif à la [configuration du connecteur LDAP](#).

Où est mon enregistrement ?

Pour vous assurer que le contenu est bien envoyé au système GSA, utilisez la page **Sources de contenu > Flux** de la console d'administration afin de vérifier que le contenu a été reçu et traité par le connecteur LDAP.

Description

La principale utilité du connecteur LDAP est de permettre l'utilisation de la fonctionnalité "Recherche de spécialistes" qui fournit des coordonnées d'utilisateurs dont vous pouvez personnaliser l'affichage dans les résultats de recherche. Pour plus d'information sur la recherche de spécialistes, consultez cette [page du centre d'aide](#).

Vous pouvez également présenter le contenu fourni par le connecteur LDAP à côté des résultats de recherche traditionnels et des autres documents correspondant aux critères de recherche de l'utilisateur. Dans ce cas, il peut être souhaitable de personnaliser la feuille de style XSLT de sorte que les résultats du connecteur LDAP soient présentés de façon similaire à ceux de l'élément de barre latérale appelé "Recherche de contacts".

Recherche sécurisée

Le connecteur LDAP effectue uniquement des recherches publiques. Il n'est pas compatible avec la recherche sécurisée.

Limites

Le connecteur LDAP utilise une fonctionnalité de pagination pour passer en revue les résultats LDAP plutôt que de réunir tous les objets correspondant à une recherche dans une seule requête. Certains serveurs LDAP ne sont pas compatibles avec cette fonctionnalité. Dans ce cas, il convient de définir la propriété LDAP appelée "MaxPageSize" (ou équivalente) sur une valeur suffisamment élevée pour pouvoir renvoyer tous les objets dans un résultat de recherche unique.

Chapitre 12 Architecture de déploiement des connecteurs

Présentation

Ce chapitre traite des configurations de connecteur à haute disponibilité. Ces configurations sont basées sur les prérequis suivants :

- La mise en miroir est utilisée pour permettre la haute disponibilité du système GSA.
- La haute disponibilité n'est pas indispensable, et donc pas requise, pendant l'exploration.
- La recherche sécurisée est requise, d'où l'importance de la haute disponibilité au moment de la présentation des résultats.
- L'authentification n'est pas effectuée par un connecteur, mais par un autre mécanisme tel que GSA avec Kerberos, altération des cookies ou fournisseur SAML.

Pour la recherche sécurisée, nous examinerons deux cas d'utilisation impliquant un connecteur devant être configuré pour la haute disponibilité :

- GSA met en œuvre la liaison tardive par le biais d'une API d'autorisation passant par le connecteur.
- LCA par URL : sachant que les LCA sont stockées dans l'index de GSA, le connecteur ne sera pas utilisé pour les autorisations, mais il sera peut-être nécessaire pour la résolution des groupes. La configuration exacte dépendra de la manière dont cela est mis en œuvre par chaque connecteur. À partir de là, deux scénarios sont possibles :
 - Le système GSA ou d'autres composants externes réalisent la résolution des groupes, par exemple LDAP ou SAML. Dans ce cas, le connecteur n'est pas du tout utilisé lors de la présentation des résultats.
 - Les connecteurs sont utilisés pour la résolution des groupes. Au moment de l'écriture, seuls le connecteur SharePoint et le connecteur de groupes Active Directory sont compatibles avec la résolution des groupes.

Dans ce chapitre, nous vous présenterons une architecture permettant de mettre en œuvre ces deux cas d'utilisation, mais nous ne décrirons en détail qu'une seule configuration. Sachez toutefois que vous pouvez procéder de manière différente et que le niveau de haute disponibilité de votre système relève d'un choix à la fois technique et stratégique. La configuration décrite ci-dessous offre un fonctionnement sans interruption. Si vous estimez possible d'interrompre le service pendant quelques heures en cas de panne d'un composant, vous aurez peut-être intérêt à renoncer à la méthode d'équilibrage des charges et à sauvegarder régulièrement la configuration et les fichiers d'état ou la base de données de vos connecteurs. Nous vous recommandons d'analyser les coûts et les avantages associés à chaque méthode pour choisir celle qui répondra le mieux à vos besoins.

Les configurations décrites ici incluent les quatre composants suivants :

- Systèmes Google Search Appliance
- Conteneur (Tomcat) de l'application Web de gestion des connecteurs
- Gestionnaires de connecteurs
- Instances de connecteurs dans les gestionnaires de connecteurs

Il est important de bien différencier le gestionnaire de connecteurs et les instances de connecteur qu'il héberge (lors de l'attribution des noms), car cela évite les confusions durant la phase de conception. Il est également important de savoir que les connecteurs internes ne sont pas compatibles avec la mise en miroir GSA. Vous devez utiliser des connecteurs externes.

Ce scénario comprend deux systèmes GSA, un "Maître" (Master GSA) et sa "Réplique" (Replica GSA), et s'applique aux versions suivantes :

- GSA 7.0 et versions ultérieures
- Connecteur 3.0 et versions ultérieures

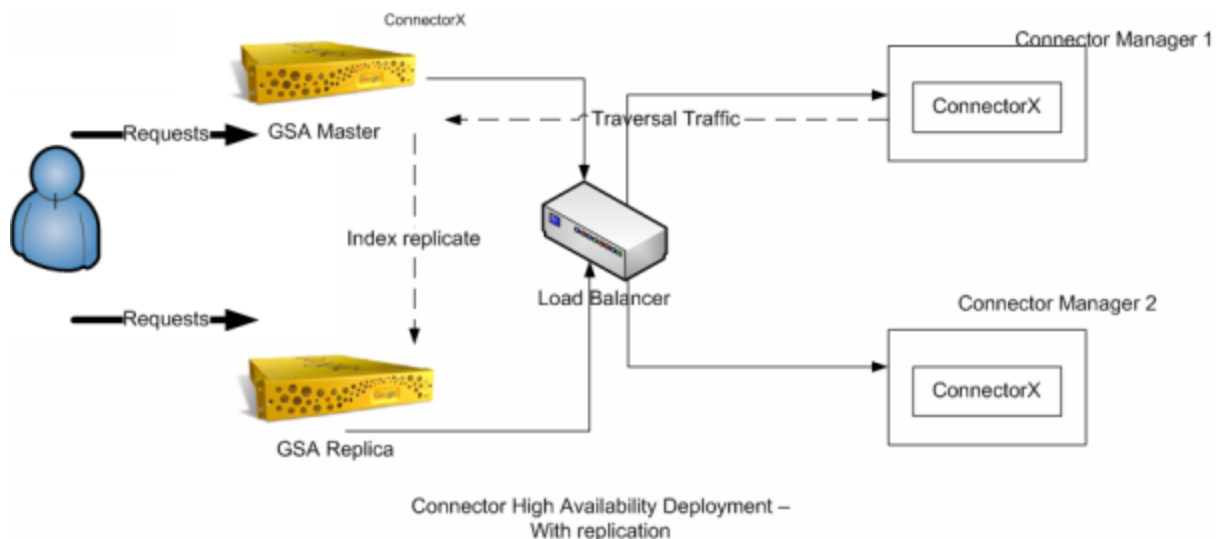
Les conditions requises pour le déploiement peuvent changer avec les nouvelles versions du connecteur.

Le contenu est transmis uniquement au système GSA maître contenant l'index de recherche complet. Il est ensuite copié sur le système GSA réplique à l'aide des mécanismes de réplication GSA. La réplique est créée à l'aide de la fonctionnalité "**Administration de la mise en miroir GSA**" disponible sur la page **GSA** > **Configuration**. La configuration de la réplication n'est pas décrite dans ce document.

La réplique récupère les index de recherche du maître, donc elle **ne doit pas** recevoir directement les contenus fournis par le connecteur. Cette configuration comprend le gestionnaire de connecteurs 1 (Connector Manager 1) alimentant le système GSA maître, et le gestionnaire de connecteurs 2 (Connector Manager 2) qui ne fournit aucun contenu.

Lors de la présentation des résultats, l'un ou l'autre des connecteurs reçoit les requêtes du système GSA. Si les connecteurs fournissent directement les autorisations, ils font appel aux API associées au système de contenu. Si les connecteurs traitent la résolution des groupes, ils font appel à une base de données partagée pour déterminer les groupes de l'utilisateur actuel. Évidemment, le serveur de base de données et le système de contenu devront être configurés pour la haute disponibilité, mais nous n'aborderons pas ce point ici.

Le diagramme ci-dessous décrit une architecture de connecteurs haute disponibilité dans une configuration en miroir.



La mise en œuvre d'une configuration en miroir implique les grandes étapes suivantes :

1. Configuration du système GSA à l'aide de la console d'administration
2. Configuration manuelle du serveur Tomcat
3. Ajout de l'équilibreur de charge au système

Chacune de ces étapes est décrite en détail dans les sections suivantes.

Étape 1 : Configuration du système GSA à l'aide de la console d'administration

Sur le GSA maître :

1. Ajoutez la véritable adresse IP du gestionnaire de connecteurs 1 à la "**Liste des adresses IP de confiance**" sur la page **Sources de contenu > Flux**.

N'ACTIVEZ PAS la "Fédération de GSA". Cela bloquerait le processus d'autorisation des recherches via l'équilibreur de charge.

2. Ajoutez le gestionnaire de connecteurs 1 à l'aide de l'option "**Enregistrer un nouveau gestionnaire de connecteurs**" disponible sur la page **Sources de contenu > Gestionnaires de connecteurs**.

Le nom du gestionnaire doit être propre à ce gestionnaire de connecteurs (par exemple, "GestionnaireConnecteurs1"). Dans le champ "**Lieu**" ou "**URL du service**", renseignez la véritable adresse IP ou le nom d'hôte de la machine (ou machine virtuelle) exécutant GestionnaireConnecteurs1, et non l'adresse de proxy de l'équilibreur de charge.

3. Ajoutez une instance de connecteur à GestionnaireConnecteurs1 à l'aide de la page **Sources de contenu > Connecteurs**.

Dans le champ "**Nom du connecteur**", saisissez un nom non unique qui désignera les instances du connecteur dans les deux gestionnaires de connecteurs. *Les connecteurs doivent porter ce même nom pour que les authentifications et/ou les autorisations de recherche avec équilibrage de charge puissent fonctionner.*

4. Vérifiez que les balayages sont activés et qu'un planning des balayages est établi.
5. Arrêtez GestionnaireConnecteurs1.
6. Revenez à l'étape 3 pour enregistrer GestionnaireConnecteurs2.
7. Ajoutez une instance de connecteur à GestionnaireConnecteurs2 à l'aide de la page **Sources de contenu > Connecteurs**.

Dans le champ "**Nom du connecteur**", saisissez un nom **non unique** qui désignera les instances du connecteur dans les deux gestionnaires de connecteurs.

8. Vérifiez que les balayages sont **désactivés** en cochant "**Désactiver le balayage**".

Étape 2 : Configuration manuelle du serveur Tomcat

Ajoutez l'adresse IP de l'équilibreur de charge à la liste des adresses IP autorisées. Il est également recommandé d'ajouter l'adresse IP de certaines machines servant à l'administration du système, car cela pourrait vous être utile pour dépanner ce dernier.

Le gestionnaire de connecteurs inclut un sélecteur d'adresse Tomcat distante activé. Celui-ci restreint le trafic en direction du gestionnaire de connecteurs en autorisant uniquement l'accès depuis l'hôte local et le système GSA configuré lors de l'installation. Cette étape vous permettra de modifier le paramètre du sélecteur d'adresse distante (RemoteAddressValve) de manière à autoriser le trafic provenant de l'équilibreur de charge. Vous trouverez des instructions détaillées pour modifier le sélecteur d'adresse distante sur la page de wiki relative au paramètre [RemoteAddrValve](#).

Pour chaque gestionnaire de connecteurs :

1. Ajoutez l'adresse IP de l'équilibreur de charge à la liste des adresses IP autorisées.
2. Recommandé : Ajoutez l'adresse IP de certaines machines servant à l'administration du système, car cela pourrait vous être utile pour dépanner ce dernier.

Étape 3 : Ajout de l'équilibreur de charge au système

Cette étape vous permettra de rediriger le trafic des systèmes GSA vers l'équilibreur de charge, et non plus vers les gestionnaires de connecteurs.

1. Arrêtez les gestionnaires de connecteurs.

Cela vous permettra de supprimer le lien entre le gestionnaire de connecteurs et les systèmes GSA, tout en conservant la configuration des connecteurs. GSA ne vous autorisera pas à supprimer un gestionnaire de connecteurs lorsque des connecteurs sont en cours d'exécution.

Il ne vous autorisera pas non plus à ajouter un gestionnaire de connecteurs comprenant des instances de connecteur dont le nom existe déjà.

2. Sur le système GSA maître, effectuez les opérations suivantes :
 - a. Sur la page **Sources de contenu > Gestionnaires de connecteurs**, effacez l'enregistrement de GestionnaireConnecteur1 et de GestionnaireConnecteurs2.
 - b. Redémarrez uniquement GestionnaireConnecteurs1.
 - c. Sur la page **Sources de contenu > Gestionnaires de connecteurs**, enregistrez l'équilibreur de charge comme un nouveau gestionnaire de connecteurs. Son état devrait apparaître sous la forme d'un rond vert. La page **Sources de contenu > Connecteurs** devrait comprendre une instance du connecteur associée au gestionnaire de connecteurs de l'équilibreur de charge.
 - d. Redémarrez GestionnaireConnecteur2.

Remarque : À partir de cette étape, vous ne devez plus modifier la configuration ni la planification du connecteur via la console d'administration d'un de vos systèmes GSA, car cela pourrait générer des incohérences au niveau de l'état, voire des blocages.

Remplacer un connecteur défectueux

En cas de défaillance d'un serveur hébergeant un connecteur, l'équilibreur de charge redirige automatiquement le trafic vers le connecteur actif. Vous devez alors mettre le connecteur défaillant hors service, puis enregistrer un nouveau connecteur. Le service ne devrait pas être interrompu entre-temps. Nous utiliserons la méthode de migration de la configuration et de l'état, plutôt que celle de l'arrêt/redémarrage lors de la configuration initiale. Cette méthode peut également être utilisée lors de la configuration initiale de HA. La procédure décrite ci-dessous est générique et peut varier selon le type de connecteur concerné. Par exemple, avec un connecteur SharePoint, vous devrez définir une méthode de migration sécurisée des fichiers d'état, si nécessaire.

1. Installez un connecteur du même type sur un nouveau serveur (le serveur défaillant étant probablement inutilisable). Assurez-vous que le service associé au connecteur est bien arrêté.
2. Copiez le dossier contenant l'instance du connecteur stocké sur le serveur du connecteur actif pour le coller sur le serveur du nouveau connecteur.
3. Sur le système GSA, modifiez le gestionnaire du connecteur défaillant de manière à le rediriger vers le nouveau serveur.

Chapitre 13 Accès au cloud

Présentation

La fonctionnalité Accès au cloud est un composant AJAX écrit en JavaScript qui appelle un service sur le système GSA qui communique ensuite avec l'API Google Apps Search. Les réponses de Google Apps sont traitées par le système GSA, puis affichées. Les résultats sont présentés sous forme d'éléments de barre latérale comprenant 10 résultats maximum.

Pour configurer l'intégration de contenu personnel, utilisez la page **Sources de contenu > Google Apps** (avant la version 7.2 **Accès au cloud > Google Apps**) de la console d'administration. Dans la configuration, saisissez le domaine Google Apps dans le champ "**Domaine**", sans préfixe d'aucune sorte tel que "www". Le champ "**Clé client OAuth**" représente un client (également un domaine) enregistré auprès de Google.

Par exemple, recherche.votreentreprise.com est une **clé client OAuth** sur le domaine Google Apps votreentreprise.com. Les deux domaines peuvent également ne rien avoir en commun. Par exemple, votreentreprise1.com est une **clé client OAuth** sur le domaine Google Apps votreentreprise.com. La clé client OAuth doit être enregistrée en tant que client tiers autorisé à accéder aux API Domain.

Pour plus d'informations sur l'enregistrement d'un client pour un domaine Google Apps, reportez-vous à l'article relatif à l'[utilisation d'OAuth 2.0 avec des applications de serveur Web](#).

Dépannage

Le principal problème rencontré avec la fonctionnalité Accès au cloud est l'absence de résultats. Pour résoudre les problèmes de connectivité avec Google Apps, procédez comme suit :

1. Vérification des identifiants OAuth
2. Vérification du principal identifiant validé
3. Vérification des paramètres du pare-feu

Chacune de ces étapes est décrite en détail dans les sections suivantes.

Vérification des identifiants OAuth

Sachant que les appels au service Google Apps Search sont envoyés depuis un navigateur, il est difficile de déterminer la cause du problème. Nous avons donc créé le site OAuth Playground pour vous permettre d'effectuer des tests : <https://developers.google.com/oauthplayground/>

Pour vérifier les identifiants OAuth à l'aide du site OAuth Playground, procédez comme suit :

1. Dans la section 1 permettant de **sélectionner des champs d'application**, saisissez :
`https://www.googleapis.com/auth/appssearch`
2. Dans la section 2 permettant de **modifier les paramètres OAuth**, saisissez les informations suivantes :
oauth_signature_method: HMAC_SHA1
oauth_consumer_key: <clé client de l'utilisateur provenant de la console d'administration Google Apps>
consumer secret: <code secret client provenant de la console d'administration Google Apps>

3. Dans la section 3, cliquez sur le bouton **"Request Token"** (Demander un jeton) pour obtenir un jeton de requête. `oauth_token` sera renseigné automatiquement.
4. Dans la section 4, cliquez sur le bouton **"Authorize"** (Autoriser) pour autoriser le jeton de requête. `oauth_token` sera renseigné automatiquement.
5. Dans la section 6, cliquez sur le bouton **"Access token"** (Jeton d'accès) pour passer à un jeton accès. `oauth_token` sera transformé en jeton d'accès.
6. Dans la section 6, collez la chaîne suivante après avoir remplacé les paramètres indiqués entre crochets "<>" par leur véritable valeur : `https://www.googleapis.com/apps/search/v1r1?q=<terme de la requête>&xoauth_requestor_id=<adresse e-mail Google Apps complète>`
7. Cliquez sur le bouton d'**exécution**.

Vous devriez obtenir une réponse HTTP 200 accompagnée de quelques résultats. Dans le cas contraire, cela signifie que les identifiants OAuth ne sont pas configurés correctement.

Vérification du principal identifiant validé

Le principal identifiant validé (identifiant utilisateur associé au groupe d'informations d'identification par défaut) est utilisé comme identifiant pour la fonctionnalité Accès au cloud. Pour vérifier le principal identifiant validé :

1. Accédez à la page **Rechercher > Recherche sécurisée > Connexion universelle** de la console d'administration.
2. Téléchargez le journal SegMgr.
3. Vérifiez que le groupe d'informations d'identification par défaut (**Default** credential group) est correctement résolu.

Alias

Si les noms d'utilisateur obtenus comme principal identifiant validé par le système Google Search Appliance sont différents des identifiants Gaia du domaine Google Apps, ils doivent être [ajoutés](#) comme alias par l'administrateur du domaine Google Apps.

Vérification des paramètres du pare-feu

Pour que la fonctionnalité Accès au cloud soit opérationnelle, le système GSA doit pouvoir accéder à l'adresse `https://www.googleapis.com/apps/search/v1r1/` via le port 443. Elle ne peut pas fonctionner via un proxy. Il est recommandé d'ajouter également "google.com" au pare-feu pour permettre aux systèmes GSA de communiquer ouvertement avec les API Google.