

Google Search Appliance

Planification et conception de la solution

Mars 2014



© Google 2014

Planification et conception de la solution

Ce document porte sur la planification et la conception d'une solution de mise en œuvre d'un système Google Search Appliance (GSA).

À propos de ce document

Les recommandations et informations rassemblées dans ce document sont le résultat de notre travail sur le terrain avec de nombreux clients et partenaires dans des environnements variés. Nous tenons à les remercier chaleureusement d'avoir partagé avec nous leurs expériences et leurs observations.

Thèmes abordés	Ce document couvre les principales activités de planification et les principes fondamentaux qui s'appliquent au déploiement de toute solution de recherche.
Lecteurs cibles	Responsables de projets et administrateurs GSA
Environnement informatique	GSA associé à divers mécanismes de sécurité et sources de contenu
Phases de déploiement	Planification du déploiement
Autres ressources	<ul style="list-style-type: none">● GSA en pratique : Architectures de déploiement● GSA en pratique : Sécurité● Le site Web Learngsa.com fournit des ressources pédagogiques sur le système GSA.● La documentation produit de GSA fournit des informations complètes sur le système.● Le Portail d'assistance Google for Work permet d'accéder à l'assistance Google.● Google for Work Connect

Sommaire

[À propos de ce document](#)

[Chapitre 1 : Présentation](#)

[Chapitre 2 : Recensement des exigences requises](#)

[Exigences utilisateur](#)

[Exigences en matière de sécurité et de contenu](#)

[Exigences en matière d'allocation de ressources](#)

[Exigences en matière de performances et d'évolutivité](#)

[Basculement/Haut niveau de disponibilité](#)

[Exigences en matière de création de rapports et d'administration](#)

[Chapitre 3 : Identification des phases](#)

[Phases de déploiement](#)

[Par où commencer ?](#)

[Échelonnement de vos activités](#)

[Combien de temps les phases doivent-elles durer ?](#)

[Chapitre 4 : Définition des critères de succès](#)

Chapitre 1 : Présentation

En théorie, la réussite d'une solution de recherche est très simple à évaluer puisque son objectif est d'aider les utilisateurs à trouver les informations dont ils ont besoin. La solution de recherche doit être rapide, simple et pertinente.

Le système Google Search Appliance est conçu pour être rapide, simple d'utilisation et pertinent. En revanche, vous devez planifier et exécuter le projet de façon à tirer pleinement parti de la puissance du système de recherche. La clé de cette approche consiste à maintenir des cycles de livraison courts et à structurer le travail autour de ces phases.

Le déploiement de chaque solution de recherche Google est unique. À titre d'exemple, vous pouvez avoir à fournir une solution de recherche sur des contenus SharePoint et élargir cette dernière avec des bons de commande provenant de SAP. Autre exemple : vous pouvez fournir une solution de recherche sur les millions de documents accumulés par votre entreprise au fil des ans, en les associant aux coordonnées des personnes qui les ont écrites et à des documents de réglementation.

Les sources de contenu, les exigences en matière de sécurité et les besoins des utilisateurs sont différents pour chaque déploiement. Cependant, certaines activités principales de planification et principes fondamentaux s'appliquent au déploiement de toutes les solutions de recherche. Ce document cible les activités principales de planification suivantes :

- [Recensement des exigences requises](#)
- [Identification des phases](#)
- [Définition des critères de succès](#)

Chapitre 2 : Recensement des exigences requises

Inventoriez les exigences requises et répartissez-les dans des groupes classés par ordre de priorité de façon à les intégrer dans les phases de travail. Voici les domaines vers lesquels orienter vos recherches :

- [Exigences utilisateur](#)
- [Exigences en matière de sécurité et de contenu](#)
- [Exigences en matière de performances et d'évolutivité](#)
- [Exigences en matière de création de rapports et d'administration](#)

Exigences utilisateur

Identifiez les éléments déterminants pour que les utilisateurs soient pleinement satisfaits du déploiement. En général, les exigences utilisateur reposent sur les critères suivants :

- [Facilité d'utilisation](#)
- [Étendue et profondeur](#)
- [Communication et commentaires](#)

Facilité d'utilisation

La recherche ne doit pas être une corvée pour l'utilisateur. La définition des exigences en matière de facilité d'utilisation peut contribuer à rendre votre solution de recherche intuitive et efficace pour les utilisateurs.

Lors de l'identification des exigences dans ce domaine, posez-vous les questions suivantes :

- Quelles sont les fonctionnalités relatives à l'utilisation de la solution de recherche qui ont vraiment de l'importance aux yeux des utilisateurs ?
 - Google Search Appliance propose de nombreuses fonctionnalités d'utilisabilité prêtes à l'emploi et simples à mettre en œuvre (suggestions de requêtes, correspondances, navigation dynamique, reconnaissance d'entités et recherche de spécialistes, pour ne citer que quelques exemples).
- Comment les utilisateurs accèdent-ils à la solution de recherche ?
 - Faites en sorte que les points d'accès à l'outil de recherche soient les plus pratiques possible pour les utilisateurs. La page de recherche peut être accessible directement sur le système GSA ou indirectement par l'intermédiaire d'une application Web qui analyse les résultats provenant du code XML du système GSA. Pour les portails indexés, il peut être intéressant de remplacer l'interface de recherche existante par la recherche Google (champ de recherche pour SharePoint, par exemple).
 - Veuillez noter que l'accès à la solution de recherche peut s'effectuer depuis plusieurs points d'entrée : l'Intranet, une application Web spécifique, une application mobile, ou même un client lourd. La mise en place de plusieurs points d'accès peut accroître la complexité de la solution et nécessiter des mécanismes d'authentification supplémentaires.

- Quelles sont les exigences de vos utilisateurs en matière de vitesse ?
 - Les technologies de style AJAX peuvent grandement améliorer la réactivité et les performances perçues, tout en enrichissant l'interface de recherche.

En général, répondez aux exigences en matière d'utilisation le plus tôt possible dans le cycle de mises à jour, car ces exigences ne sont pas reliées aux sources de contenu et peuvent créer un véritable engouement des utilisateurs pour la solution de recherche.

Étendue et profondeur

Une solution de recherche doit répondre aux souhaits de votre communauté d'utilisateurs. La définition des exigences en matière d'étendue et de profondeur permet de savoir si la solution de recherche répond aux besoins d'un groupe d'utilisateurs suffisamment large et fournit à ces derniers le contenu adéquat.

Lors de l'identification des exigences en matière d'étendue et de profondeur, répondez aux questions suivantes :

- Quelles sont les caractéristiques des groupes d'utilisateurs ?
 - Si possible, votre attention doit être portée en priorité sur les groupes les plus importants et les utilisateurs dont les besoins ne sont pas satisfaits.
 - En utilisant des frontaux, vous pouvez modifier l'interface et le confort d'utilisation du système de recherche, et afficher différents contenus en fonction des utilisateurs et de leurs besoins. Pour obtenir des informations sur les frontaux, consultez le document [Concevoir l'outil de recherche](#).
- Quels éléments les utilisateurs tentent-ils de trouver actuellement, sans succès ?
 - Ces contenus doivent être intégrés dès les premières phases.
 - L'intégration des utilisateurs doit être associée à celle du contenu qu'ils recherchent. De ce fait, essayez de ne pas fournir aux utilisateurs une fonctionnalité de recherche avant d'avoir inclus le contenu qui les intéresse.
- Certains utilisateurs ont-ils des besoins plus pointus en termes de recherche ?
 - La recherche avancée prête à l'emploi peut facilement être personnalisée avec des métadonnées et d'autres fonctionnalités principales.
 - Si cette personnalisation n'est pas requise, la recherche doit rester simple tout en apportant une richesse fonctionnelle.

Communication et commentaires

Les commentaires utilisateur représentent un outil efficace dans l'identification des problèmes d'utilisabilité. Lorsque vous sollicitez l'avis des utilisateurs sur le déploiement du système de recherche, vous leur accordez de l'intérêt. Le fait de définir les exigences en matière de communication et de commentaires permet à vos utilisateurs d'aider l'équipe responsable de la mise en œuvre à déployer le système de recherche en indiquant à cette dernière ce qui fonctionne pour eux et ce qui ne fonctionne pas.

Lors de l'identification des exigences en matières de communication et de commentaires, répondez aux questions suivantes :

- Quelles sont les informations à transmettre aux utilisateurs ?
 - L'ajout de nouveaux contenus et de nouvelles fonctionnalités ne suffit pas. Il est important d'en informer les utilisateurs pour maintenir leur intérêt dans le produit et faire en sorte que vos succès soient remarqués.
 - La plupart de vos utilisateurs savent déjà utiliser la technologie de recherche Google. Les besoins en formation sont donc réduits. Veillez simplement à ce que vos utilisateurs sachent que la recherche de contenu d'entreprise est désormais aussi simple que n'importe quelle recherche sur Internet.
- Comment recueillir les commentaires des utilisateurs ?
 - Une des manières de recueillir des commentaires sur l'outil de recherche et de peaufiner l'expérience utilisateur consiste à demander à l'administrateur du système de recherche de vérifier régulièrement les rapports GSA en rassemblant les données suivantes : requêtes les plus fréquentes, requêtes n'ayant renvoyé aucun résultat, etc. (Cette procédure peut être automatisée grâce à l'API Admin.)
 - Les commentaires des utilisateurs constituent l'un des meilleurs indicateurs de succès. Pensez à conduire régulièrement des enquêtes de satisfaction auprès des différents groupes d'utilisateurs.
 - Fournissez un lien par l'intermédiaire duquel vos utilisateurs peuvent transmettre leurs commentaires (intégration à votre système de suivi des demandes, par exemple).
 - Ajoutez un formulaire "Avez-vous trouvé ce que vous cherchiez" sur la page des résultats de recherche pour que les utilisateurs puissent immédiatement soumettre leurs commentaires.
 - Pour recueillir les commentaires utilisateur, vous pouvez également utiliser la boîte à outils gratuite [Search Quality Feedback](#).

Exigences en matière de sécurité et de contenu

Dans la plupart des entreprises, les deux éléments suivants vont de pair lors du déploiement d'un système de recherche :

- [Contenu](#)
- [Sécurité](#)

Le niveau de complexité des scénarios qui englobent le contenu et la sécurité peut varier et aller d'un site Web public non sécurisé de quelques pages à une intégration complexe d'un système ERP (Enterprise Resource Planning) tel que SAP ou PeopleSoft.

Prévoyez votre architecture finale dès les premières phases, mais pensez à intégrer à la fois le contenu et la sécurité. En d'autres mots, ne différez pas la livraison d'un outil de recherche de qualité sous prétexte de vouloir indexer le moindre élément de contenu ou mettre en œuvre un environnement de développement de la sécurité dont vos utilisateurs n'ont pour l'instant pas besoin.

Contenu

En général, vous devez analyser tous les référentiels organisationnels dès que possible dans le projet. Google Search Appliance excelle par la pertinence, la rapidité et la puissance avec laquelle il permet de rechercher du contenu non structuré. Vous ne devez néanmoins pas faire l'impasse sur vos contenus structurés (entrepôt de données, systèmes transactionnels, etc.).

Il est important de comprendre la manière dont les sources de contenu interagissent entre elles, car c'est cette interaction qui définit les étapes de déploiement du contenu. Par exemple, le contenu issu d'un système de gestion de dossiers peut être astucieusement complété par celui d'un catalogue produit. Les utilisateurs peuvent ainsi consulter les informations sur les produits, mais aussi les types de problèmes rencontrés lors de l'utilisation de ces produits.

Le tableau suivant répertorie différents types de sources de contenu, structurées et non structurées, ainsi que les éléments à prendre en compte pour programmer leur déploiement.

Source de contenu	Structurée/ Non structurée	Complexité (Faible/Moyenne/Élevée)	Éléments de réflexion
Sites Web publics	NS	F	<ul style="list-style-type: none"> • HTTP
Systèmes de fichiers	NS	M–E	<ul style="list-style-type: none"> • SMB (Connecteur) ou HTTP
Bases de données	S	F–E	<ul style="list-style-type: none"> • Utiliser un connecteur de base de données, un flux ou un élément compatible Web • Volume du contenu
Sites Web Intranet	NS	F	<ul style="list-style-type: none"> • La sécurité doit peut-être être prise en compte • Gestion de pages dynamiques

Portail interne	NS	M–E	<ul style="list-style-type: none"> • Les complexités relatives à la sécurité sont peut-être à étudier • La prise en compte des URL non uniques (une même URL contient différents contenus selon le rôle utilisateur) peut être nécessaire
Système de gestion de contenu	NS	F–M	<ul style="list-style-type: none"> • L'exploitation de métadonnées peut s'avérer nécessaire • La possibilité d'une exploration en mode natif doit être étudiée • La sécurité doit peut-être être prise en compte
Applications LOB (Lotus Notes, par exemple)	S	F–E	<ul style="list-style-type: none"> • Si ce sont des applications Web, il doit être possible d'explorer ces sources de contenu • Peut nécessiter un connecteur. • Sécurité
Applications d'entreprise (ERP, par exemple)	S	M–E	<ul style="list-style-type: none"> • Nécessite l'identification des données principales • Peut nécessiter des flux ou un connecteur • Sécurité
Autres systèmes transactionnels	S	F–E	<ul style="list-style-type: none"> • Peuvent être accessibles par un connecteur personnalisé ou un module OneBox • Sécurité

Sécurité

La sécurité peut revêtir une grande complexité dans le déploiement d'une solution de recherche. Lors de l'analyse du contenu, vérifiez si ce dernier est sécurisé et, si tel est le cas, par quel mécanisme (formulaire, système de sécurité au niveau de l'application, etc.).

Il est important, notamment pour les projets qui impliquent plusieurs sources de contenu, d'analyser en détail les options de sécurité avant de passer à la mise en œuvre. Cette étape est importante, car certaines informations sur la sécurité peuvent changer de façon significative le champ d'application d'un projet. Par conséquent, veillez au préalable à analyser la sécurité et à recenser les exigences en la matière.

Dans une solution de recherche, la sécurité a des répercussions dans deux principaux domaines :

- [Exploration et acquisition du contenu](#)
- [Traitement des résultats : authentification et autorisation des utilisateurs](#)

Pour obtenir des informations exhaustives sur le système de recherche et la sécurité, consultez la rubrique [Gestion de la recherche pour le contenu dont l'accès est contrôlé](#).

Exploration et acquisition du contenu

Google Search Appliance peut utiliser des protocoles de sécurité standards, tels que LDAP, HTTP Basic, NTLM, Kerberos, les certificats, SAML ou une sécurité basée sur des formulaires.

La connaissance de toutes les modifications apportées à la sécurité permet de planifier l'acquisition du contenu. Par exemple, la sécurité peut avoir un impact sur l'exploration du système de fichiers et du Web que vous devez planifier (exemple : configuration d'un proxy ou vérification de l'activation du protocole CIFS dans vos systèmes de fichiers Windows pour la compatibilité avec l'indexation SMB).

Un système de sécurité complexe peut nécessiter l'utilisation d'autres moyens d'acquisition du contenu, tels que les flux ou les connecteurs.

Le GSA est bien armé pour indexer la plupart des configurations standards. Cependant, dans certains cas, vous aurez besoin d'apporter de légers ajustements à votre environnement pour permettre au système de recherche d'explorer et d'acquérir le contenu sans utiliser de flux ou de connecteurs. Par exemple, il se peut que, pour élargir l'exploration à un nouveau sous-domaine, vous deviez modifier

un domaine de cookie. En effet, le système de recherche explore le contenu pour autoriser les cookies afin de se conformer à la requête pour les spécifications de commentaires (RFC). En principe, ce sont de légères modifications qui peuvent être mises en œuvre par plusieurs méthodes.

Les considérations en matière de sécurité peuvent nécessiter l'ajustement de la procédure d'indexation ou l'utilisation d'une autre approche d'acquisition de contenu. Dans ce cas, il peut y avoir des conséquences sur des éléments de la conception de la solution qui ne sont pas directement liés à l'architecture de la sécurité, tels que les flux ou les flux de travail de publication.

L'analyse de l'acquisition de contenu doit aller de pair avec l'autorisation des utilisateurs, car les méthodes d'autorisation peuvent varier selon que la liaison est précoce ou tardive. En général, la liaison précoce (avec les LCA) permet de renvoyer les résultats de recherche plus rapidement que la liaison tardive, car les autorisations de sécurité ne sont pas stockées sur le système GSA. La liaison précoce nécessite donc la mise en cache des autorisations sur le système GSA, ce qui implique que les informations doivent être mises à disposition du système GSA. Les informations LCA peuvent exister dans les flux, dans les informations d'en-tête HTML ou dans les en-têtes HTTP des serveurs. La liaison tardive correspond à la vérification effectuée lorsque les résultats sont comparés (en principe) au système source afin de vérifier immédiatement que l'utilisateur a accès. Il n'est donc pas nécessaire de personnaliser le contenu.

Pour en savoir plus, consultez le document [GSA en pratique : Sécurité](#).

Traitement des résultats : authentification et autorisation des utilisateurs

Lors du traitement du contenu sécurisé, Google Search Appliance vérifie dans un premier temps que l'utilisateur est autorisé à afficher les résultats en question. Si l'utilisateur n'est pas autorisé à afficher un document, ce dernier n'apparaît pas dans la liste des résultats.

Gardez à l'esprit que la transmission des résultats sécurisés est généralement gérée par deux processus : l'authentification et l'autorisation.

- L'authentification vérifie que l'utilisateur est vraiment qui il prétend être. Lors de ce processus, le système GSA rassemble également toutes les informations pertinentes relatives aux groupes de cet utilisateur.
- Le mécanisme d'autorisation se sert de l'identifiant de l'utilisateur pour vérifier que ce dernier a bien accès à un contenu donné.

Il est essentiel de comprendre ce principe pour indexer plusieurs sources de contenu sécurisées. Par exemple, il est possible d'utiliser un seul mécanisme d'authentification lorsque deux sources de contenu sécurisées se servent du même identifiant utilisateur. Cependant, si les deux sources de contenu requièrent des méthodes d'autorisation différentes, le système GSA devra être configuré pour utiliser deux méthodes d'autorisation (par exemple les LCA pour une source de contenu et l'autorisation par connecteur pour l'autre).

Bien entendu, vous avez toujours la possibilité de rendre ces résultats publics et de ne pas appliquer de sécurité au moment de la présentation des résultats. Dans de nombreuses situations, la recherche peut être déployée dans un premier temps de manière non sécurisée, le mécanisme de sécurité étant ajouté lorsque davantage de contenus sont acquis. La recherche publique (site Internet externe, par exemple) est généralement déployée de cette manière.

Google Search Appliance intègre par défaut de nombreux mécanismes de sécurité et peut répondre à diverses exigences hétérogènes en matière de sécurité, notamment l'authentification silencieuse (via des cookies ou Kerberos, par exemple) et l'intégration dans le système d'authentification unique (SSO). Dans le cas de scénarios d'intégration personnalisés, la SPI SAML est fournie, comme décrit dans le document [GSA en pratique : Sécurité](#). La SPI SAML sert à gérer les vérifications d'authentification et d'autorisation par l'intermédiaire de divers systèmes et protocoles. La fonctionnalité offre une grande flexibilité dans la manière de mettre en œuvre la sécurité. Vous pouvez :

- acheter des fournisseurs prédéfinis (voir des exemples sur Google for Work Solution Marketplace) ;
- concevoir une solution de sécurité personnalisée ;
- effectuer l'intégration d'un système SSO via SAML.

Google Search Appliance est également compatible avec la définition de listes de contrôle d'accès (LCA) pour que les vérifications d'autorisation puissent être effectuées sur les documents à l'aide de la liaison précoce. Les LCA accroissent les performances, mais offrent également davantage d'options de gestion de la sécurité. Cette fonctionnalité propose également des options pour établir les étapes du déploiement de votre solution de recherche. Pour obtenir des informations sur l'affichage des résultats de manière sécurisée et les règles LCA, consultez le document [GSA en pratique : Sécurité](#).

Une fois que le(s) identifiant(s) et les groupes de l'utilisateur ont été rassemblés, il peut être nécessaire de contrôler l'ordre d'utilisation des mécanismes d'autorisation. Par exemple, un CMS peut ne pas autoriser les résultats mis en cache, ce mécanisme étant en revanche hautement souhaitable pour une autre plate-forme. Ces règles, portant notamment le classement des méthodes d'autorisation pour un format donné, peuvent faire l'objet d'une règle d'[autorisation flexible](#) et doivent être étudiées avec soin dans des environnements complexes.

Exigences en matière d'allocation de ressources

Il est essentiel, pour la bonne livraison d'un projet, que les experts appropriés soient disponibles au bon moment. À défaut, un projet peut prendre énormément de retard. Il est donc important de veiller à ce que les bonnes personnes soient affectées à leurs tâches quand cela est nécessaire.

Par exemple, un administrateur SharePoint peut être sollicité pour configurer le connecteur GSA Connector for SharePoint ou participer au test du système.

Planifier les tâches en amont présente également l'avantage de tenir les intervenants informés des tâches spécifiques qu'ils auront à effectuer. Ces intervenants pourront alors les intégrer dans leurs emplois du temps respectifs.

En général, un projet mobilise les experts suivants pour le propriétaire du système GSA :

- Chef de projet
- Chef de projet technique
- Propriétaire de l'entreprise
- Analyste d'affaires
- Administrateur de référentiel
- Administrateur réseau
- Expert en sécurité
- Administrateur GSA
- Développeur

En général, un projet mobilise les experts suivants pour le partenaire Google for Work :

- Chef de projet
- Experts GSA
- Experts de référentiel
- Développeurs

Exigences en matière de performances et d'évolutivité

Les exigences non fonctionnelles (ENF) sont généralement purement techniques. Dans une solution de recherche, les ENF les plus fréquentes sont les suivantes :

- [Performances](#)
- [Évolutivité](#)

Performances

Les exigences en matière de performances tournent généralement autour de la rapidité avec laquelle la solution renvoie les résultats. Mais elles peuvent également inclure la vitesse d'acquisition des contenus. Les performances dépendent généralement d'un certain nombre de facteurs, notamment :

- Exigences en matière de sécurité
- Type de contenu
- Taille du corpus
- Type de requêtes exécutées
- Architecture et performances du réseau
- Autres fonctions de recherche utilisées (extension de requêtes, navigation dynamique, regroupement dynamique des résultats, par exemple).

Si des exigences spécifiques en matière de performances sont requises, il est admis de conduire un test des performances tôt dans le processus de déploiement afin de déterminer les modifications qui doivent être effectuées dans l'architecture de la solution.

Le système Google Search Appliance en lui-même ne peut pas être modifié, mais vous pouvez intégrer les modifications suivantes dans votre déploiement :

- Configuration des règles ou des LCA par URL pour améliorer le contrôle de la sécurité au moment de la présentation des résultats.
- Déploiement d'un proxy pour mettre en cache les éléments faisant l'objet de recherches récurrentes, le cas échéant. Cette modification n'est utile que pour les recherches effectuées sur du contenu public (non sécurisé).
- Réduction du trafic réseau entre le système Google Search Appliance et les sources de contenu. La diminution du temps de latence a surtout des répercussions sur l'exploration, mais elle permet également d'améliorer les performances de l'autorisation de liaison tardive.

- Amélioration des performances perçues par l'optimisation de la réactivité (AJAX, par exemple) et l'affichage d'un "sablier" de progression pour donner une impression de réactivité.
- Déploiement de systèmes de recherche supplémentaires afin de répartir la charge. Cette modification réduit la demande exercée sur un seul système de recherche et permet d'éviter que la capacité ne soit un facteur limitant.

Consultez le document [GSA en pratique : Architectures de déploiement](#) pour comprendre plus en détail l'architecture de recherche orientée performances.

Les exigences en matière de performances doivent également prendre en compte l'exploration et l'indexation. L'indexation du système de recherche alourdit vos systèmes de contenu. Vous devez savoir si les systèmes de contenu ne doivent pas être affectés à certaines heures de la journée afin de pouvoir configurer les charges d'hôtes du système de recherche en conséquence. En outre, si la charge exercée sur le système de contenu a atteint la limite fixée ou si ce dernier est particulièrement lent, vous devez envisager l'emploi de flux ou d'un connecteur.

Outre la méthode d'indexation (exploration ou actualisation) et la taille des documents, vous devez également noter que les fonctionnalités telles que la reconnaissance d'entités ou l'aperçu des documents peuvent également avoir un impact sur les performances, car elles traitent généralement d'autres données au cours de l'indexation.

Évolutivité

Les exigences en matière d'évolutivité tournent généralement autour du nombre de requêtes par seconde ou par minute. Comme pour les performances, le nombre de requêtes par seconde autorisé par la solution dépend des exigences en matière de sécurité, du type de contenu, du type de requête et des performances du réseau, entre autres facteurs.

Si des exigences en matière d'évolutivité ont été relevées, Google recommande de les réévaluer, de préférence en vous servant des données issues des recherches en cours. Les exigences en matière d'évolutivité peuvent généralement être revues à la baisse.

Les solutions de recherche peuvent être conçues pour être compatibles avec des centaines de requêtes par seconde, ce qui n'est souvent pas nécessaire en pratique. Les types d'exigences en matière d'évolutivité pour une solution de recherche sont considérablement différents de ceux d'un système transactionnel.

Pour en savoir plus sur la conception d'une solution de recherche dans le but d'accroître l'évolutivité, consultez le document [GSA en pratique : Architectures de déploiement](#). Pour obtenir des informations sur le nombre de connexions simultanées acceptées par Google Search Appliance, consultez la section [Concevoir une solution de recherche](#).

Basculement/Haut niveau de disponibilité

Pour de nombreuses entreprises, l'environnement de recherche se doit d'avoir un haut niveau de disponibilité. GSA fournit des outils qui facilitent la configuration et la maintenance d'une solution à haut niveau de disponibilité (mise en miroir de GSA à GSA, par exemple). Cependant, pour des environnements plus complexes, vous devez veiller à ce qu'une solution de basculement soit planifiée en conséquence.

En ce qui concerne l'environnement de recherche, une solution de basculement répond en général à deux exigences :

- [Haut niveau de disponibilité pour l'indexation](#)
- [Haut niveau de disponibilité pour le traitement](#)

Haut niveau de disponibilité pour l'indexation

L'indexation à haut niveau de disponibilité garantit que l'index du système GSA soit toujours actualisé, même si l'environnement principal ne peut pas effectuer le traitement. Le système GSA joue un rôle capital dans l'assurance d'un haut niveau de disponibilité, mais d'autres mécanismes y contribuent également (connecteurs ou outils d'actualisation, par exemple) et doivent être configurés pour le basculement. Le haut niveau de disponibilité lors de l'indexation figure souvent dans la liste des éléments souhaités, mais cette exigence reste peu satisfaite en pratique, car elle est particulièrement complexe et implique de nombreuses ressources.

Haut niveau de disponibilité pour le traitement

Un haut niveau de disponibilité pour le traitement garantit la réception des résultats par les utilisateurs, même si l'une des applications nécessaires au traitement est en panne. Pour le traitement public, dans le cas de scénarios de liaison précoce (règles LCA ou LCA par URL) et dans la plupart des scénarios de liaison tardive (requêtes Head), la fonctionnalité de mise en miroir des GSA est en principe suffisante. Cependant, il existe des scénarios qui ne sont pas aussi évidents :

- Utilisation de groupes renvoyés par un connecteur (Active Directory ou SharePoint, par exemple).
- Dans certains scénarios de liaison tardive (autorisation par connecteur, SAML Bridge ou déchiffrement de cookie, par exemple), vous devez également vous assurer d'avoir une solution de basculement, sans quoi les utilisateurs ne seront pas en mesure de recevoir les résultats de recherche si le connecteur principal est en panne.

Pour en savoir plus sur la conception d'une solution de recherche pour le basculement, consultez le document [GSA en pratique : Architectures de déploiement](#).

Exigences en matière de création de rapports et d'administration

La création de rapports tient une place importante dans toute application d'entreprise et la recherche ne fait pas exception. Outre les exigences en matière de création de rapports sur la recherche, identifiez les autres rapports requis. Portez une attention particulière aux exigences non fonctionnelles. Les exigences à prendre en compte incluent :

- la technologie d'analyse à utiliser (Google Analytics, rapports détaillés sur les recherches ou tout autre outil tiers, par exemple) ;
- la fréquence des rapports et le mode de distribution ;
- la nécessité de créer d'autres types de rapports (surveillance du système ou événements d'administration, par exemple).

Veillez à bien connaître les processus de l'entreprise qui utiliseront ces rapports. Par exemple, vous devez étudier les cas d'utilisation de vos exigences en matière de création de rapports et vous assurer que la stratégie de création de rapports permettra d'y répondre.

Chapitre 3 : Identification des phases

La plupart des déploiements sont classés dans l'une des catégories suivantes, répertoriées de la plus simple à la plus complexe :

- Déploiements spécialisés centrés sur la mise à disposition des clients d'une interface de recherche puissante et conforme à leurs habitudes leur permettant d'accéder aux informations externes publiques ou sécurisées d'une organisation.
- Déploiements de systèmes de recherche autonomes permettant aux entreprises de gagner en productivité et de mieux exploiter les informations qu'elles détiennent.
- Déploiements de systèmes de recherche exigés par un événement ou inclus dans un déploiement plus vaste (mise en œuvre d'un nouveau portail, mise en place d'un système de gestion de contenu ou lancement d'un nouveau projet d'architecture de l'information, par exemple).

Le déploiement d'un système de recherche exige en général des gains rapides, l'objectif étant d'offrir rapidement aux utilisateurs une interface de recherche riche et d'ajouter de la valeur de façon incrémentale et itérative tout au long du déploiement de la solution de recherche.

La valeur commerciale découle de l'étendue du contenu sur lequel la fonctionnalité de recherche repose ainsi que sur la facilité d'utilisation et l'efficacité de l'outil de recherche.

Phases de déploiement

Pour réussir le déploiement d'un système de recherche, il faut que ce dernier soit réalisé rapidement et fasse l'objet d'améliorations continues. N'essayez pas de tout traiter du premier coup. Donnez la possibilité à vos utilisateurs d'accéder au contenu qu'ils souhaitent le plus tôt possible. Les gains rapides générés par une livraison rapide permettent d'obtenir le soutien de vos actionnaires et de générer l'engouement de vos utilisateurs.

Le champ d'application des phases peut être défini en termes de :

- Sources de contenu
- Sécurité
- Groupes d'utilisateurs
- Fonctionnalités d'utilisabilité

Chaque phase doit inclure une tâche de test au cours de laquelle sont évaluées explicitement la satisfaction utilisateur ainsi que les fonctionnalités demandées par les utilisateurs, les unités commerciales et le département informatique, entre autres. Comme toujours, évaluez les demandes de fonctionnalités en étudiant notamment les risques associés à leur mise en œuvre et à la non-mise en œuvre.

En général, chaque phase s'étend sur une durée relativement courte. Vous pouvez donc utiliser la plupart des méthodologies de livraison (Agile ou Life Cycle, par exemple).

Si vous utilisez une méthode de développement classique, gardez à l'esprit que les phases de développement d'un projet de recherche Google sont relativement courtes. Vous devrez donc effectuer des ajustements pour fournir un outil de recherche de qualité, de manière flexible et efficace. Parmi les technologies que vous allez utiliser, nombreuses sont celles que vous pouvez mettre en œuvre et ajuster rapidement (feuilles de style XSLT, modules OneBox, par exemple). Votre approche doit être flexible afin de créer rapidement un prototype et enchaîner les livrables.

Cette section explique comment structurer vos livrables et planifier votre projet pour élargir l'empreinte de recherche et accroître l'utilisation de la solution de recherche. Chaque livraison fait avancer votre déploiement un peu plus loin sur la courbe de valeur.

Par où commencer ?

Google Search Appliance est conçu pour être déployé rapidement sur les principales sources de contenu. L'exploitation des normes et protocoles ouverts permet d'intégrer rapidement le contenu issu de différentes sources et de mettre en œuvre des fonctionnalités permettant d'enrichir l'utilisation (suggestions de requêtes, navigation dynamique, résultats ajoutés par l'utilisateur et clusters de résultats dynamiques, par exemple).

Les phases peuvent aller d'une semaine ou deux jusqu'à un mois. Google recommande de structurer le programme de travail en phases courtes, avec des livraisons rapides de fonctionnalités, contenus ou groupes d'utilisateurs itératifs.

Souvent, une seule phase de livraison rapide suffit. Même si votre déploiement fait partie d'un programme de travail complet, plus long à exécuter, visant à fournir une recherche universelle sur tous les actifs de votre entreprise, vous devez tout de même structurer vos phases de façon à obtenir des gains rapides.

Avant de commencer le déploiement de votre solution de recherche, réalisez les tâches principales suivantes pour que votre spécialiste en déploiement puisse lancer votre système le plus tôt possible.

Avant de commencer :

- Installez le système de recherche.
- Configurez les paramètres réseau.
- Faites l'inventaire de vos sources de contenu (y compris le nombre de documents).
- Faites l'inventaire de vos systèmes de sécurité, en vous concentrant sur l'authentification et l'autorisation.
- Configurez votre réseau afin de permettre au système de recherche d'accéder à toutes les sources de contenu et, le cas échéant, restreignez l'accès aux domaines sécurisés.
- Créez les identifiants utilisateurs requis par le système de recherche pour explorer le contenu.

En tant que partenaire Google for Work, veuillez à utiliser les outils fournis dans les [mises à jour produit](#) pour bénéficier de toutes les informations nécessaires à la réussite de votre projet.

Échelonnement de vos activités

Planifier vos activités comme décrit dans les sections suivantes :

- [Développement préliminaire](#)
- [Versions incrémentales](#)
- [Livraison avancée](#)

Développement préliminaire

Les éléments répertoriés dans le tableau ci-dessous peuvent généralement être livrés rapidement et simplement. Ils s'adaptent parfaitement à un développement préliminaire. Nombreux sont ceux qui doivent être considérés comme obligatoires (un frontal personnalisé, par exemple, doit toujours faire partie de la livraison principale, quel que soit son niveau de complexité).

Veuillez noter que la première phase implique l'indexation d'une source de contenu critique (priorité élevée). Cependant, surtout si vous ne connaissez pas GSA, il est fortement recommandé de vous entraîner en utilisant une source de contenu simple à indexer. Le fait de gérer dans un premier temps une source de contenu simple permet de mieux appréhender les phases ultérieures et d'acquérir ainsi une connaissance générale cruciale du système de recherche et des processus.

Le tableau ci-dessous répertorie des exemples à privilégier pour le développement préliminaire.

Livrable	Type	Complexité
Exploration Basic HTTP : <ul style="list-style-type: none"> ● Intranet ● Extranet ● Site Web ● Wiki ● Bases de connaissances Web (Lotus Notes, par exemple) 	Sources de contenu	Faible
Système de fichiers	Sources de contenu	Moyenne à élevée
Sites SharePoint	Sources de contenu	Moyenne à élevée
Modules OneBox basiques (PeopleSearch, par exemple)	Sources de contenu	Faible à moyenne
Recherche de spécialistes	Sources de contenu	Moyenne
Connecteur Active Directory Groups	Sécurité	Moyenne
Authentification LDAP	Sécurité	Faible
Intégration Kerberos	Sécurité	Moyenne
Suggestions de requête	Facilité d'utilisation	Faible à moyenne*
Résultats ajoutés par l'utilisateur	Facilité d'utilisation	Faible
Frontal personnalisé	Facilité d'utilisation	Faible à moyenne
Rapports détaillés sur les recherches	Facilité d'utilisation	Faible
Navigation dynamique	Facilité d'utilisation	Faible
Reconnaissance d'entités	Facilité d'utilisation	Faible à moyenne
Aperçu des documents	Facilité d'utilisation	Faible à moyenne
Recherche avec caractères génériques	Facilité d'utilisation	Faible
Utilisateurs du système principal	Groupes d'utilisateurs	Faible
Propriétaires de l'entreprise	Groupes d'utilisateurs	Faible

* selon que vous optez pour les fonctionnalités prêtes à l'emploi ou pour une mise en œuvre personnalisée.

La complexité peut varier en fonction de votre infrastructure et de la configuration de l'environnement.

Versions incrémentales

Les livrables répertoriés dans le tableau suivant sont adaptés à une version incrémentale. Étudiez ces éléments et planifiez leur déploiement en les classant par priorité (généralement selon le volume du contenu et l'importance de l'élément pour l'entreprise) et niveau de difficulté.

Il est souvent possible d'avancer la date de livraison en utilisant des outils tiers (connecteurs, par exemple) et en sollicitant l'assistance de partenaires Google for Work qualifiés qui connaissent les problèmes liés à l'intégration de Google Search Appliance. Certains de ces livrables (recherche avancée personnalisée, par exemple) nécessitent d'étudier les commentaires d'utilisateurs avant la mise en œuvre complète. Le tableau suivant répertorie les éléments adaptés à des versions incrémentales.

Livrable	Type	Complexité
Contenu de portail	Source de contenu	Moyenne à élevée
Bases de connaissances non Web (incluent un grand nombre de bases de données Lotus Notes)	Source de contenu	Moyenne
Systèmes de gestion de contenu	Source de contenu	Faible à élevée
Modules OneBox personnalisés (voire sécurisés)	Source de contenu	Faible à élevée
Contenu d'application personnalisée	Source de contenu	Faible à élevée
Connecteurs supplémentaires (FileNet, Livelink, Documentum)	Source de contenu	Faible à moyenne
Recherche avancée personnalisée	Facilité d'utilisation	Moyenne
Fonctionnalités avancées d'ergonomie (interface utilisateur AJAX, par exemple)	Facilité d'utilisation	Faible à moyenne
Utilisateurs supplémentaires associés au nouveau contenu	Groupes d'utilisateurs	Faible à moyenne
Gestion de plusieurs systèmes GSA (fédération/GSA^n)	Architecture	Élevée

Livraison avancée

Les livrables répertoriés dans le tableau ci-dessous sont adaptés à une livraison avancée. Les éléments adaptés à une livraison avancée requièrent davantage de temps de développement, sont plus difficiles à mettre en œuvre ou peuvent ne pas être requis. Si ces éléments font partie de votre déploiement de recherche, vous pouvez les mettre en œuvre en parallèle d'autres tâches de déploiement. Vous pouvez ainsi permettre aux utilisateurs d'accéder immédiatement au contenu principal.

Dans certains cas, les éléments sont des sources de données structurées qui nécessitent d'être analysées afin de pouvoir déterminer la meilleure façon de les intégrer dans l'outil de recherche (plates-formes Business Intelligence, par exemple). Le tableau suivant répertorie les éléments adaptés à une livraison avancée.

Livable	Type	Complexité
Sécurité avancée (LCA, déchiffrement de cookie, authentification silencieuse, par exemple)	Sécurité	Moyenne à élevée
Fournisseur SPI SAML personnalisé	Sécurité	Moyenne à élevée
Attribution de plusieurs noms d'utilisateur à un même utilisateur (en fonction de la source de contenu)	Sécurité	Élevée
Applications de confiance	Sécurité	Élevée
Système de gestion d'enregistrements	Sources de contenu	Moyenne à élevée
Systèmes ERP (SAP, Oracle, PeopleSoft)	Sources de contenu	Moyenne à élevée
Systèmes GRC (Siebel, par exemple)	Sources de contenu	Moyenne à élevée
Entrepôts de données/plates-formes BI	Sources de contenu	Moyenne à élevée
Autres systèmes LOB	Sources de contenu	Moyenne à élevée
Haut niveau de disponibilité pour l'indexation	Architecture	Élevée
Haut niveau de disponibilité pour le traitement	Architecture	Moyenne

Combien de temps les phases doivent-elles durer ?

En général, la durée des phases doit s'étendre de quelques jours à quelques semaines. La difficulté du travail varie et requiert des estimations spécifiques. En revanche, le temps requis pour effectuer les tâches doit être défini en fonction de la complexité, comme indiqué dans le tableau suivant.

Complexité	Durée
Faible	2–8 heures
Moyenne	1–5 jours
Élevée	1–4 semaines

Les durées inscrites dans ce tableau sont indiquées à titre informatif et varient en fonction de l'environnement et des exigences requises. Google recommande de déterminer le niveau de difficulté de votre déploiement.

Outre la difficulté, vous devez allouer suffisamment de temps à l'acquisition du contenu. Faites tout votre possible pour que l'index contienne un maximum de contenus provenant des sources de contenu ciblées. Cela ne veut évidemment pas dire que vous devez attendre d'avoir toutes les sources de contenu possibles dans votre solution de recherche. En revanche, l'index doit contenir l'ensemble du contenu provenant des systèmes à intégrer pour la version concernée.

Il est difficile de prédire la vitesse d'acquisition du contenu par Google Search Appliance, car le taux d'acquisition dépend de nombreux facteurs, notamment les suivants :

- Performances du réseau
- Performances du serveur
- Charges d'hôtes
- Type de contenu
- Taille du contenu
- Fonctionnalités (aperçu des documents ou reconnaissance d'entités, par exemple)

Google recommande d'effectuer des tests dès les premières étapes du cycle de vie du projet afin de déterminer la vitesse d'acquisition du contenu. Aidez-vous de ces informations pour effectuer votre planification.

Chapitre 4 : Définition des critères de succès

Avant d'effectuer la livraison, définissez les critères permettant de déterminer si le projet a atteint ses objectifs ou non. Le but est d'avoir un ensemble de critères d'acceptation clairement défini. Exemples de critères de succès pour le déploiement d'un système de recherche :

- Niveau d'appréciation de la nouvelle solution GSA par rapport à l'ancienne solution de recherche (test A/B, par exemple)
- Évaluations de pertinence réalisées par les utilisateurs (attribution de notes par les utilisateurs, par exemple)
- Tests de sécurité (vérifier que l'authentification et l'autorisation fonctionnent pour tous les systèmes sécurisés)
- Étendue du contenu (volume du contenu figurant dans l'index, 95 % de contenu dans un système, par exemple)
- Étendue du déploiement (pourcentage de comptes utilisateur activés)

Définissez les critères de succès en collaboration avec le propriétaire de l'entreprise et, pour être au plus près de la réalité, étudiez les commentaires portant sur les anciennes solutions de recherche. En outre, veillez à mener un processus d'analyse continu (après chaque phase, par exemple) afin d'effectuer les ajustements nécessaires le plus tôt possible.