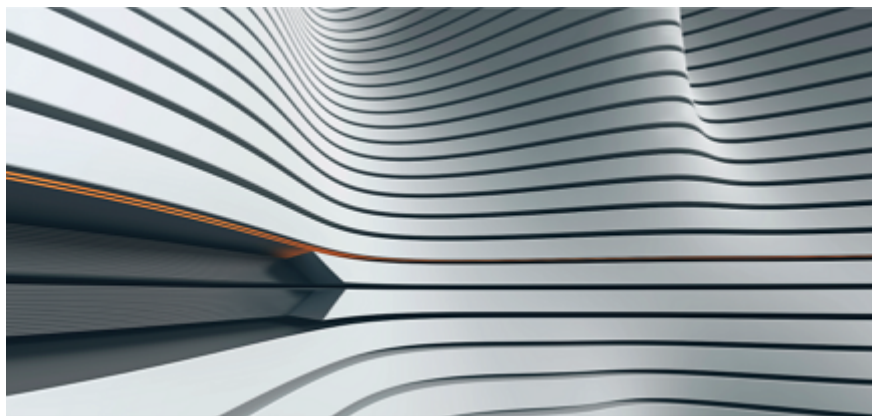


ABBYY FlexiCapture : apprendre la mise en place d'installations performantes



Par ABBYY

Date de publication : 30 janvier 2020

TOUT PUBLIC

Ce tutoriel propose les meilleures pratiques pour mettre en place des installations performantes d'ABBYY© FlexiCapture© Édition distribuée (ci-après dénommée FlexiCapture).

Vous y trouverez également :

- des approches pour améliorer la performance de FlexiCapture, identifier les goulots d'étranglement et optimiser les performances du système ;
- les indicateurs de performance et la méthodologie de test de FlexiCapture ;
- les limites d'applicabilité des résultats de test.

La redondance de FlexiCapture n'est pas couverte par ce tutoriel.

Un espace de discussion vous est proposé sur le forum. N'hésitez pas à y apporter vos contributions. **Commentez**

I - Préambule.....	4
II - Aperçu des performances de FlexiCapture.....	4
III - Architecture.....	4
III-A - Côté serveur.....	5
III-B - Côté client.....	6
IV - Interaction des composants.....	6
V - Indicateurs de performance.....	10
V-A - Paramètres impactant la charge de travail du système.....	10
V-B - Taille moyenne des lots.....	11
V-C - Mode couleur de l'image.....	11
V-D - Pages par jour et pages par heure.....	11
V-E - Nombre d'utilisateurs.....	12
V-F - Durée de stockage des documents.....	12
VI - Mise à l'échelle.....	13
VI-A - Estimation du nombre de cœurs de processeurs requis par le serveur d'application.....	19
VI-A-1 - Stations automatisées.....	19
VI-A-2 - Opérateurs de vérification.....	20
VI-A-3 - Scanner les opérateurs.....	20
VI-B - Stations de traitement.....	21
VI-B-1 - Comment régler les performances d'une station de traitement ?.....	21
VI-B-2 - Comment calculer le nombre de postes de traitement ?.....	23
VI-C - Stations de scannage.....	25
VI-C-1 - Support des stations de scannage.....	26
VI-D - Stations de vérification.....	27
VI-E - Flux de travail.....	27
VII - Valeurs optimales et limites des performances du système.....	28
VIII - Surveillance du système et détection des goulets d'étranglement.....	31
IX - Tests de performance.....	37
IX-A - Dispositif de tests minimal.....	37
IX-B - Test de FlexiCapture 12 contre FlexiCapture 11 au banc d'essai ABBYY (Small).....	38
IX-B-1 - Spécifications des installations de test.....	39
IX-C - Résultats des tests au banc ABBYY (Moyen).....	40
IX-C-1 - Spécifications des installations de test.....	41
IX-D - Test des stations de traitement avec différents types d'unités centrales au banc ABBYY (Petit).....	43
IX-D-1 - Spécifications des installations de test.....	43
IX-E - Résultats des tests pour Microsoft Azure (Large).....	45
IX-E-1 - Spécifications des installations de test.....	45
IX-F - Résultats des tests pour le Huawei FusionCube 6000 (Grand).....	47
IX-F-1 - Spécification de l'installation de test.....	47
X - Remerciements Developpez.com.....	50

I - Préambule

Ce tutoriel contient des informations et des recommandations qui décrivent l'utilisation et les performances d'ABBYY FlexiCapture 12. Il fournit les meilleures pratiques d'ABBYY pour l'architecture et les performances du système FlexiCapture, ainsi que les résultats des tests. Un utilisateur peut appliquer ces informations pour décider si les informations fournies correspondent aux besoins des processus.

Tous les tests ont été effectués sur l'infrastructure d'ABBYY et ne peuvent être considérés comme une expérience ultime.

Les performances du système dépendent toujours des caractéristiques de l'utilisateur ou de facteurs environnementaux.

Ces documents ne peuvent pas être copiés sans référence aux droits d'auteur d'ABBYY.

Si vous avez des questions supplémentaires, contactez votre représentant ABBYY local dont la liste figure à l'adresse www.abbyy.com/contacts.

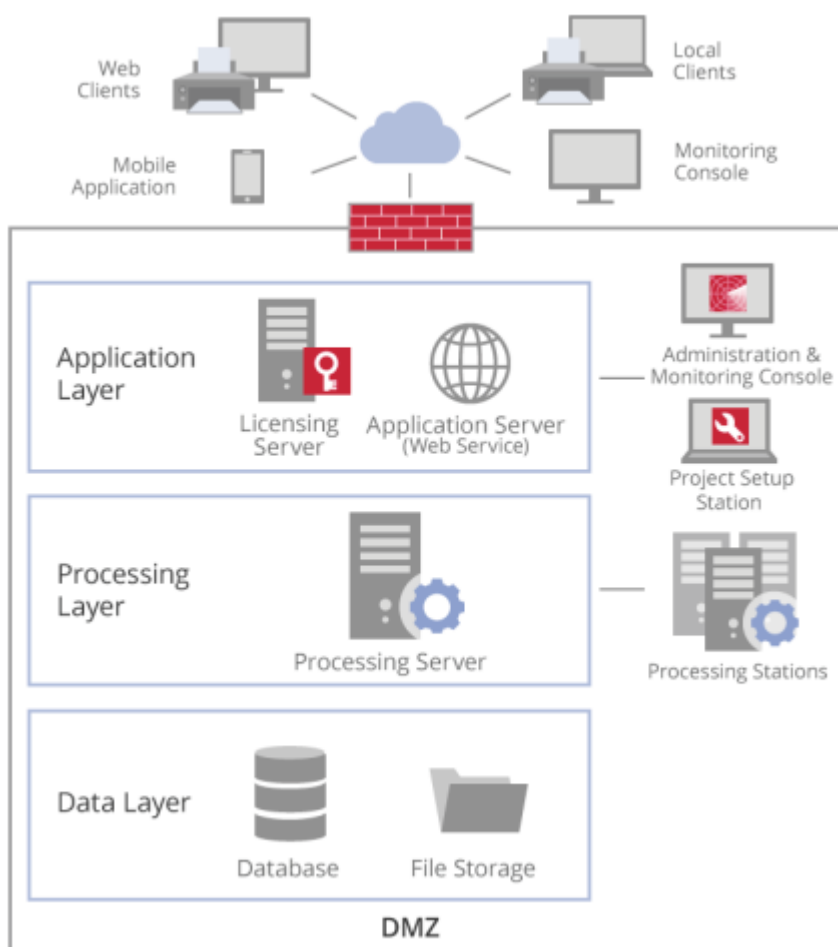
II - Aperçu des performances de FlexiCapture

ABBYY FlexiCapture est une plate-forme de capture de données prête à l'emploi, capable de :

- s'adapter à toutes les tâches en utilisant des installations autonomes ou en location sur le cloud ;
- s'optimiser pour une allocation efficace des ressources de l'entreprise (matériel et argent) ;
- traiter jusqu'à 1 million de pages en couleur ou 3 millions de pages en noir et blanc en 24 heures ;
- réduire la consommation de papiers, même dans les plus grandes entreprises à forte consommation de papier ;
- procurer une tolérance aux pannes, une disponibilité et une multilocation élevées ;
- gérer plusieurs comptes de façon séparée (locataires).

III - Architecture

ABBYY FlexiCapture est doté d'une architecture client-serveur multicouche.



III-A - Côté serveur

Le côté serveur contient trois couches de composants.

- **Couche Application**
 - Serveur d'application : c'est un web-service dans IIS (Information Internet Services de Microsoft), qui est la principale passerelle pour le trafic HTTP/HTTPS. Il vérifie l'authentification et l'autorisation des utilisateurs, et effectue la logique commerciale de FlexiCapture.
 - Serveur de licence : est un service qui contrôle les informations relatives à la licence actuelle et aux performances du système juridique.
- **Couche Traitement**
 - Le serveur de traitement est un service qui gère un ensemble de stations de traitement en informatique distribuée.
- **Couche Stockage de données**
 - La base de données est un répertoire des paramètres de traitement, des données personnelles des utilisateurs et des statistiques sur les documents traités et les documents en cours.
 - Le composant FileStorage est un stock d'images et de données de documents.

Chacun de ces composants peut être installé sur un ordinateur séparé – pour une sécurité et une fiabilité personnalisables, et une mise à niveau indépendante des autres composants.

III-B - Côté client

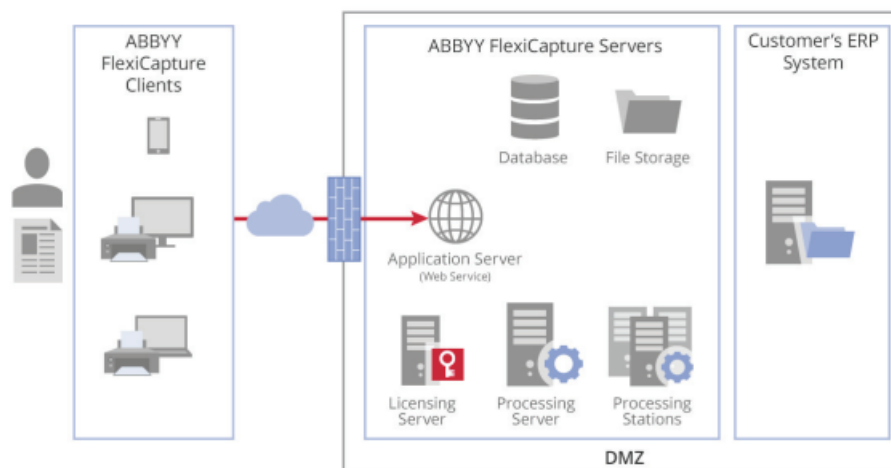
Le côté client comprend :

- les stations de numérisation et les stations de vérification, qui sont des applications permettant de créer des images de documents, de les transmettre à FlexiCapture et de vérifier les données extraites. Ces applications sont disponibles sous le nom de :
 - clients locaux,
 - des clients web s'exécutant dans un navigateur,
 - les applications mobiles ;
- les stations de traitement, qui importent et traitent les images, effectuent la reconnaissance optique de caractères sur les documents, exécutent des scripts personnalisés, exportent les données saisies dans le système ERP du client et effectuent un certain nombre d'opérations de service ;
- console web d'administration et de surveillance ;
- station de configuration du projet utilisée pour configurer les paramètres de traitement des documents et des données.

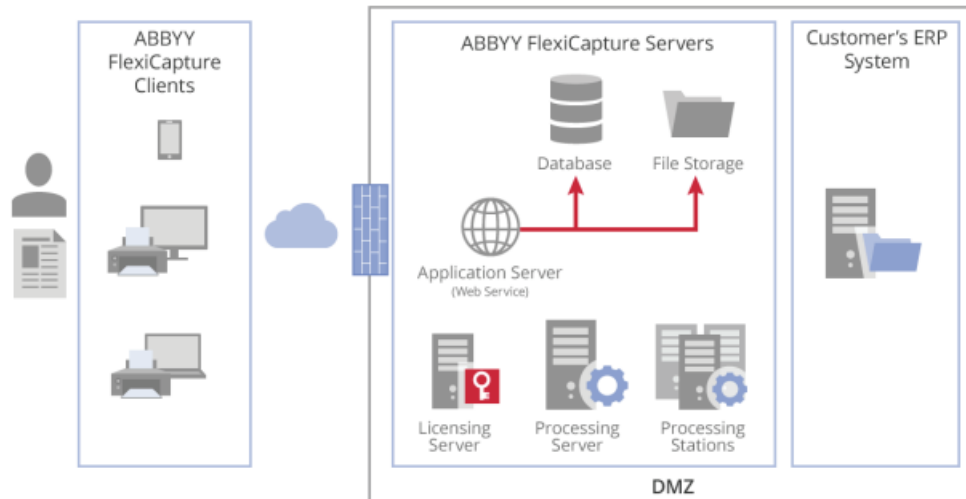
IV - Interaction des composants

On part d'un exemple pratique : un utilisateur alimente des images de documents vers FlexiCapture via une application mobile ou tout autre client (web ou local).

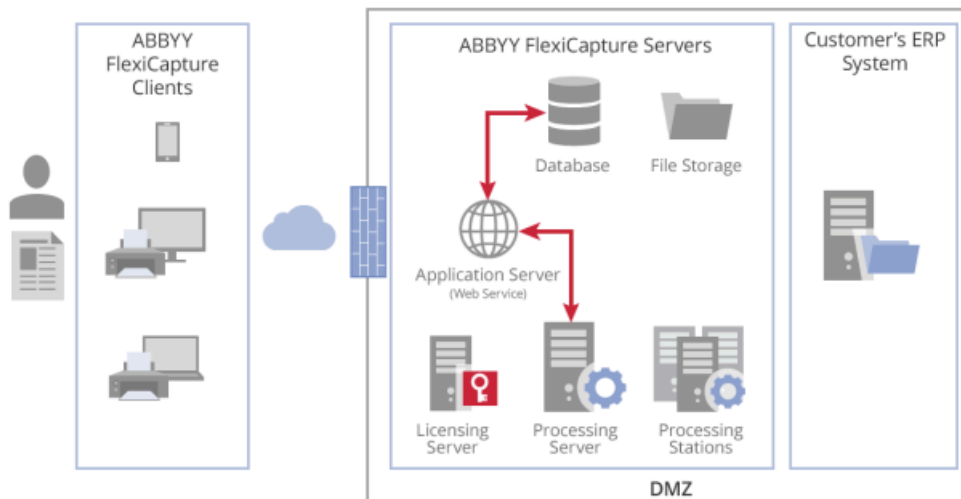
- 1 Son application client se connecte via HTTP/HTTPS au serveur d'application en demandant l'authentification de l'utilisateur.
- 2 De la même manière, elle envoie des images de documents – et parfois des informations supplémentaires pour que le serveur d'application puisse déterminer les paramètres de traitement à appliquer.



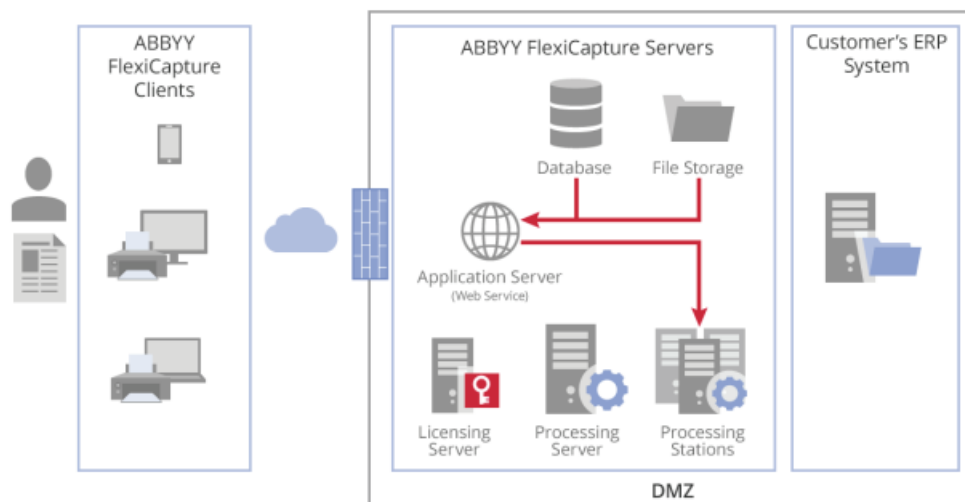
- 3 Le serveur d'application enregistre ces images dans le **FileStorage**. Dans la base de données, il crée un ensemble d'enregistrements :
 - un nouveau document est arrivé pour être traité ;
 - l'étape de traitement en cours de ce document ;
 - les paramètres de traitement à appliquer ;
 - un chemin d'accès aux images du document stockées dans le **FileStorage**.



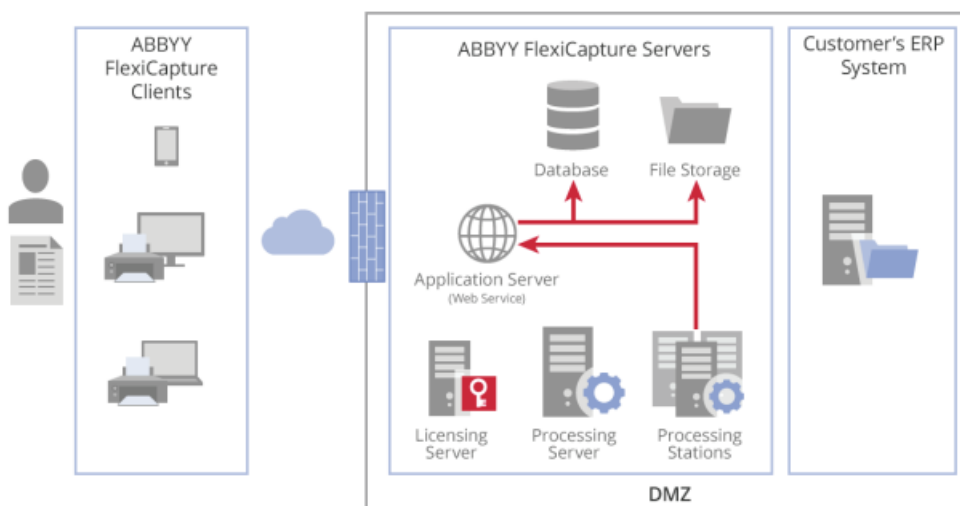
- 4 Le serveur de traitement contacte régulièrement le serveur d'application pour de nouvelles tâches de traitement. Lorsqu'il prend connaissance d'un nouveau document et des paramètres à appliquer, il attribue une tâche à une station de traitement libre.



- 5 La station de traitement obtient plus de détails sur les tâches du serveur d'application, notamment :
- les images de documents ;
 - un ensemble d'opérations de traitement à effectuer ;
 - les paramètres de traitement à appliquer.

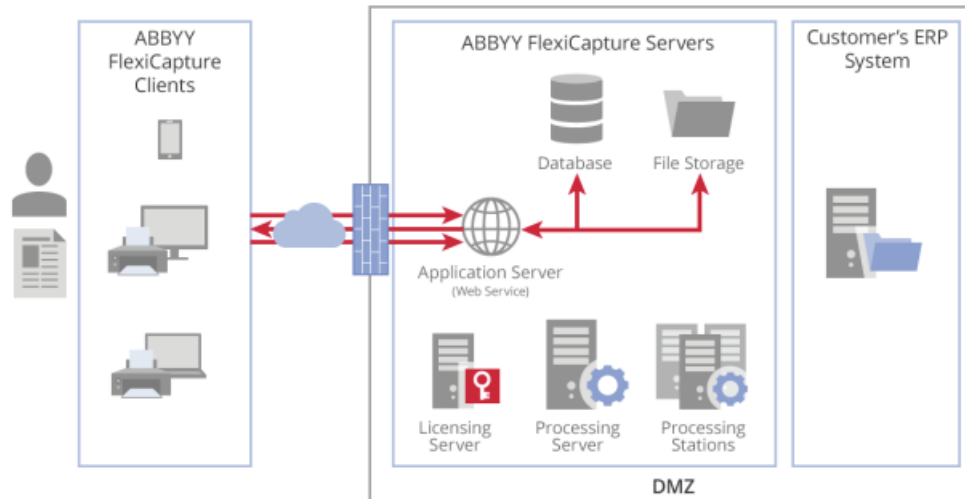


- 6 Une fois le traitement terminé, les résultats arrivent au serveur d'application, où les données correspondantes sont modifiées dans le FileStorage et le statut du document dans la base de données est mis à jour.

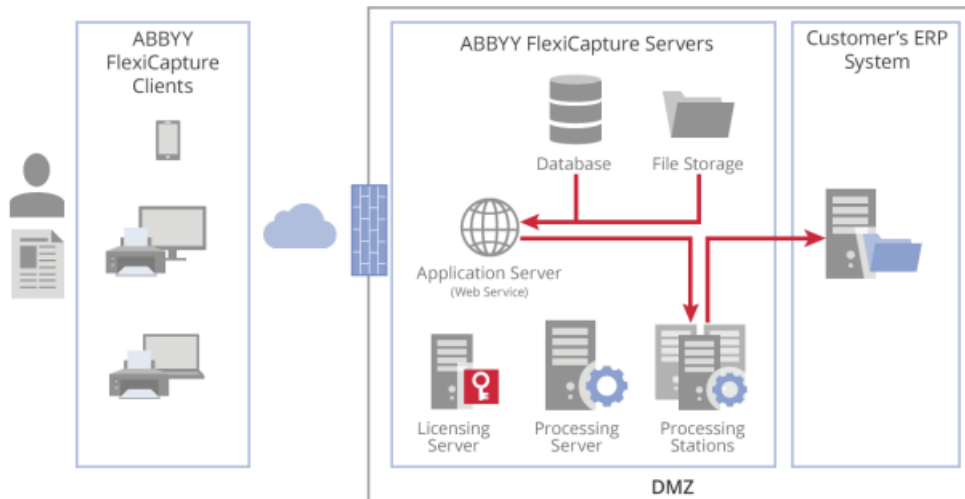


- 7 Le document traité peut être examiné « manuellement » par des vérificateurs humains si :
- les paramètres de traitement et les contrôles automatisés le permettent ;
 - ces personnes ont certains droits d'accès ;
 - elles peuvent utiliser un client de vérification – local, web ou mobile – spécifiquement installé pour cette tâche.

Ce client se connecte au serveur d'application et reçoit les images des documents et les données extraites pour vérification. Les données vérifiées arrivent au serveur d'application : il modifie les données correspondantes dans le FileStorage, et met à jour le statut du document dans la base de données.



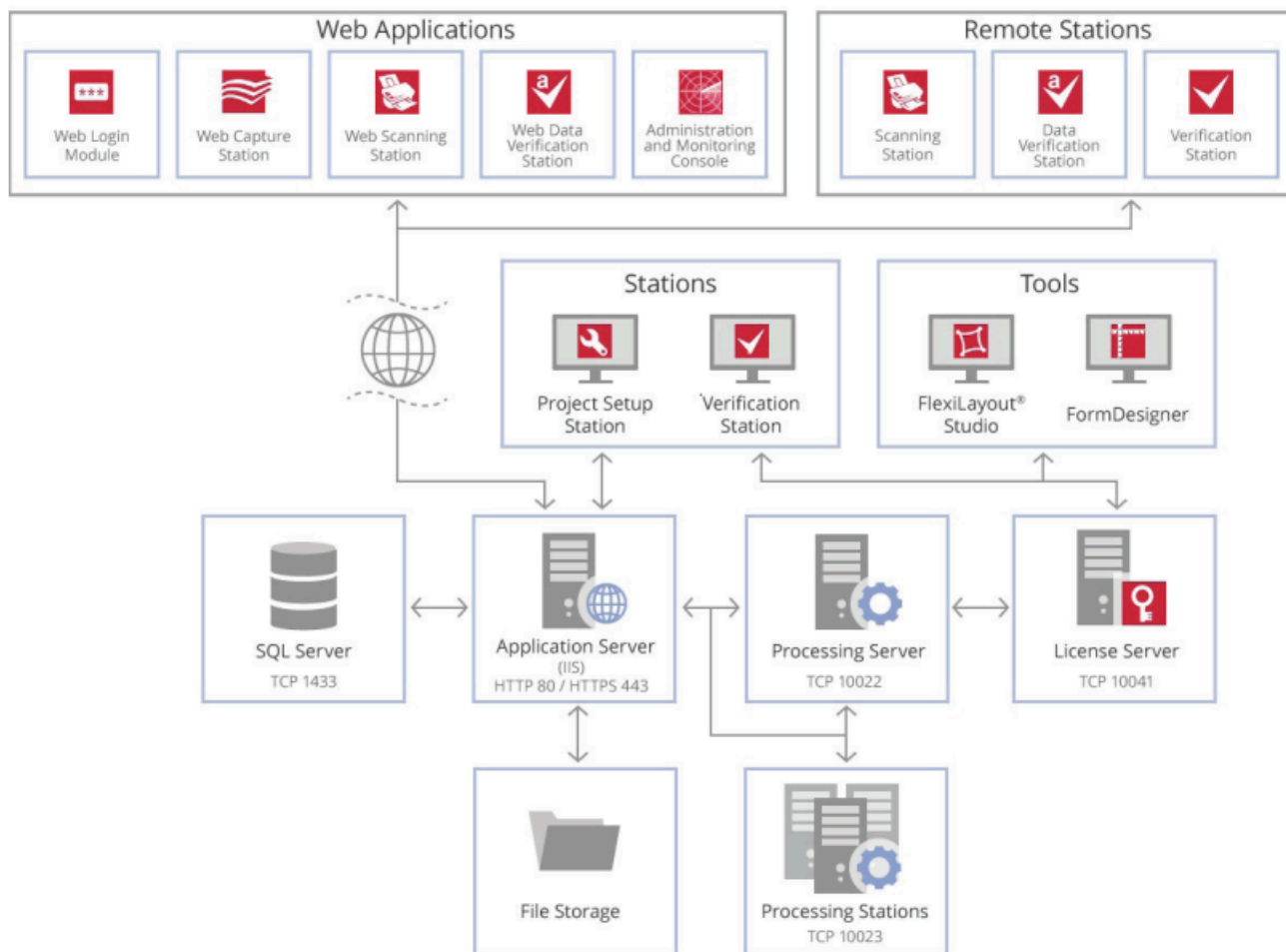
- 8 Un document entièrement traité retourne à sa station de traitement, où les images et les données sont converties dans les formats requis et exportées vers le système ERP du client, et le serveur d'application est informé que le travail est fait.



- 9 Le serveur d'application marque le document comme étant traité :
- il collecte des statistiques de traitement pour ce document – pour chaque étape qu'il a franchie ;
 - et les enregistre dans des tableaux pour générer des rapports de traitement.
- 10 Le document traité va au FileStorage et y reste jusqu'à la fin du délai de conservation fixé par le client. Le serveur d'application supprime alors ses images du FileStorage et efface tous les enregistrements de la base de données.

D'une manière générale, les composants de FlexiCapture interagissent aussi bien pour le traitement des documents que pour les tâches de service

- par exemple, les vérifications de permis.



V - Indicateurs de performance

ABBYY FlexiCapture (ou le système) extrait les données des documents arrivant en flux et c'est pourquoi nous mesurons la performance en volumes traités par période de temps.

Pour concevoir le système, définissez la **performance cible** à l'aide de mesures de performance.

Le temps de traitement requis est défini par les procédures internes, les accords de niveau de service et les exigences des processus commerciaux d'une entreprise cliente.

Les volumes de traitement sont basés sur les données antérieures et les tendances de développement des entreprises, ou sur le plan d'affaires d'une entreprise. Des sauts de volume occasionnels ou saisonniers peuvent se produire en raison de campagnes publicitaires réussies ou de la fin de l'exercice financier, etc.

V-A - Paramètres impactant la charge de travail du système

Ces paramètres façonnent la charge de travail du système :

Taille moyenne des lots en pages	Taille moyenne des documents en pages
Mode couleur de l'image : couleur, niveaux de gris, noir et blanc	Nombre d'opérateurs de scannage
Pages par jour (c'est-à-dire 24 heures), moyenne/crête	Nombre d'opérateurs de vérification
Pages par heure, moyenne/crête	Durée de conservation des documents

V-B - Taille moyenne des lots

Un lot représente un ensemble de documents apparentés traités ensemble.

Par exemple : un client soumet une douzaine de documents pour traitement – tous sous la même demande, car les recoupements et la logique commerciale interdisent leur traitement indépendant.

V-C - Mode couleur de l'image

Les images de documents sont de toutes formes et de toutes tailles comme :

- des copies scannées en couleur, en niveaux de gris ou en noir et blanc ;
- des photos en différentes résolutions ;
- des pièces jointes à des courriels – fichiers PDF vectoriels, etc.

Le niveau de couleur des images de documents dépend de :

- 1 La capacité à contrôler et à modifier les données d'entrée.
Par exemple : si les clients FlexiCapture sont choisis pour la numérisation, une entreprise peut définir le même mode de numérisation (couleur) pour tous les documents entrants ;
- 2 Besoins de stockage à long terme.
Par exemple : selon la réglementation de l'entreprise, tous les documents doivent être conservés pendant cinq ans en niveaux de gris des images seulement. Dans ce cas, les clients de FlexiCapture peuvent convertir les images en couleur en images en niveaux de gris à l'étape de balayage.

Bien que les entreprises soient souvent obligées de stocker les documents entrants dans leur format d'origine, elles sont en mesure d'estimer les formats auxquels elles peuvent s'attendre – et de fournir quelques exemples d'images. Le scénario le plus coûteux est celui où toutes les images des documents sont en couleur (coûts de transmission sur le réseau et de stockage des fichiers).

V-D - Pages par jour et pages par heure

Les **performances moyennes et maximales** sont définies comme **le nombre moyen et maximal de pages en couleur, en niveaux de gris ou en noir et blanc traitées dans une période de temps** qu'une entreprise juge préférable (1 heure, 24 heures, etc.).

- Spécifiez des intervalles de temps précis : « 24 heures » est mieux que « 1 jour », ce qui peut être mal interprété comme 1 jour de travail, c'est-à-dire 8 à 12 heures seulement.
- Rendez-les significatives pour vous – et voyez facilement si le système fonctionne selon vos besoins et les attentes.
Par exemple : un meilleur point de contrôle pour un client est le devis « 1000 pages en 24 heures », et non « 0,01 page par seconde ».

Nous utilisons des pages plutôt que des documents pour estimer le volume de traitement, car la taille des documents varie considérablement. Dans le même temps, il est généralement facile de deviner la taille moyenne des documents d'un type donné en pages. Par exemple, une facture peut contenir une page ou jusqu'à plus de 100 pages, mais elle comporte généralement trois pages en moyenne.

Enfin, nous devons trouver des chiffres en octets et en bits par seconde qui sont couramment utilisés pour calculer les performances du matériel. Pour ce faire, nous utilisons des formats typiques de page A4 de différents modes de couleur :

- A4 noir et blanc – 100 KB ;
- niveau de gris A4 – 3 MB ;

- A4 couleur – 10 MB.

Pour une estimation plus précise, un échantillon de documents réels est nécessaire.

Avoir des tailles typiques pour une page de différents modes de couleurs, et le nombre moyen et maximum de pages par jour ou heure, vous pouvez estimer le débit d'entrée moyen et de pointe en octets par seconde.

V-E - Nombre d'utilisateurs

Il peut arriver qu'un **certain nombre d'utilisateurs** accèdent simultanément au système lorsque le traitement des documents est en cours.

Il existe deux types d'utilisateurs :

- **les opérateurs de numérisation** : ils numérisent, vérifient et modifient les images des documents, puis les envoient au serveur d'application ;
- **les opérateurs de vérification** vérifient et révisent les données extraites, en téléchargeant les images et en envoyant les données corrigées au serveur d'application.

V-F - Durée de stockage des documents

La durée de stockage des documents a un grand impact sur la configuration du système et les coûts du matériel, car des temps de stockage plus longs nécessitent un plus grand FileStorage.

Le **temps de stockage des documents dans le système** est un paramètre important ; il ne doit pas être confondu avec le **temps de stockage des documents dans l'organisation**.

La **durée moyenne de stockage des documents** dans le système correspond souvent au **temps de traitement moyen**.

Parfois, lorsque **plusieurs étapes de traitement avec des opérations manuelles** sont impliquées, cela peut prendre des semaines.

Cependant, il y a des cas où le **temps moyen de stockage des documents dans le système** correspond en fait à leur **temps de traitement moyen plus le temps de stockage des images et des données** au stade du traitement. Cela se produit parce que FlexiCapture traite un document comme traité après son exportation vers le système ERP de l'entreprise, même si son traitement au sein de l'organisation est encore en cours, ce qui signifie que ce document peut être renvoyé à l'une des étapes initiales de traitement au sein du système.

Pour cette raison, les documents ayant le statut « **Traité** » (c'est-à-dire les images de document et les données saisies) sont stockés à l'intérieur de FlexiCapture jusqu'à :

- ce qu'ils soient passés en revue par tous les processus opérationnels ;
- ce qu'ils soient placés dans les archives de l'entreprise.



Note : FlexiCapture n'est pas un système d'archivage en tant que tel. Un délai de conservation typique pour un document dans le système est de deux semaines.

VI - Mise à l'échelle

FlexiCapture peut traiter de plusieurs centaines à plusieurs millions de pages par jour, et prendre en charge jusqu'à des milliers d'opérateurs. Avec les directives de ce document, il est facile d'estimer la charge du système à l'avance et choisir l'architecture et le matériel appropriés pour les serveurs.

Le système s'étend :

- en augmentant le nombre de clients de balayage, de clients de vérification et de stations de traitement ;
- en augmentant la puissance des machines pour les serveurs d'applications, de traitement, de licences et de bases de données, et le FileStorage, en utilisant plusieurs machines pour ces rôles.

Les numéros ci-dessous permettent d'évaluer ou de sélectionner une configuration préliminaire du composant du serveur FlexiCapture.

Nombre de pages traitées en 24 heures			Nombre de cœurs de traitement	Nombre d'opérateur de vérification	Nombre d'opérateur de scannage	Configuration
en noir et blanc uniquement	en niveaux de gris uniquement	couleur uniquement				
20 000	5000	1000	8	3	3	Démo
1 million	500 000	300 000	80	100	300	Média
3 millions	2 millions	1 million	120	500	1000	Grand (Média 10 Gb/s)
Beaucoup plus						xLarge (combinaison d'installations ABBYY FlexiCapture)

La surveillance des goulots garantit que le matériel utilisé n'est pas suffisant pour obtenir les performances souhaitées et qu'il est grand temps de passer à l'échelle supérieure.

La démonstration est une configuration typique pour les démonstrations ou les projets pilotes, **non recommandée** pour les projets à l'échelle de la production. Tous les composants du système sont installés sur une machine virtuelle ou déployés sur un PC.

Rôle de la machine	Conditions requises
ABBY FlexiCapture	1 processeur 4 cœurs, 2,4Ghz 8 GO DE RAM Disque dur : <ul style="list-style-type: none"> • 100 Go pour le système d'exploitation et les fichiers temporaires ; • 100 Go pour la base de données et le stockage de fichiers. Système d'exploitation : Windows© 2012 ou ultérieur.

MS SQL Express® peut être utilisé comme serveur de base de données et installé sur la même machine que les serveurs FlexiCapture. Au lieu d'utiliser un FileStorage séparé, les fichiers peuvent être stockés directement dans la base de données. Les opérateurs et les postes de traitement peuvent être installés sur la même machine.



Remarque : dans les projets commerciaux, le poste de traitement ne doit jamais être installé sur un ordinateur hébergeant des serveurs FlexiCapture ou un serveur de base de données, car il monopolise toutes les ressources et les performances du serveur se détériorent.

Medium est une configuration typique pour les projets commerciaux, car elle est évolutive : chaque composant du serveur est installé sur une machine dédiée.

Le serveur d'application doit être installé sur une machine dédiée, car il utilise une approche de montée en charge qui est différente de celle des serveurs de base de données, de traitement et de licences.



Remarque: techniquement, le serveur d'application, le serveur de traitement et le serveur de licence peuvent être installés sur le même ordinateur. La redondance du serveur sera assurée, mais l'évolutivité du serveur d'application ne le sera pas.

- Le serveur d'application est un service web dans IIS ; son évolutivité et sa fiabilité sont obtenues par un clustering qui utilise la technologie d'équilibrage de charge réseau Microsoft®. Tous les nœuds du cluster sont des pairs fonctionnant en mode actif-actif et peuvent être désactivés à tout moment.
- Le serveur de traitement et le serveur de licence sont des services Windows® ; leur fiabilité est obtenue par la création d'un cluster actif-passif basé sur la technologie Microsoft® Failover Cluster.

Microsoft® interdit clairement l'utilisation de l'ensemble de ces technologies sur un même ordinateur.

Si la fiabilité est tout ce dont vous avez besoin, intégrez le serveur d'applications dans IIS, qui prend également en charge le clustering par Microsoft® Failover Cluster.

Les serveurs de licence et de traitement peuvent être installés sur la même machine.

Nous recommandons d'installer le serveur de base de données sur une machine dédiée. Il est très gourmand en ressources et si vous le combinez avec certains autres serveurs FlexiCapture, limitez son utilisation du processeur et de la mémoire vive et localisez les fichiers de la base de données sur un disque dur physiquement séparé, afin de ne pas affecter les performances du serveur voisin.

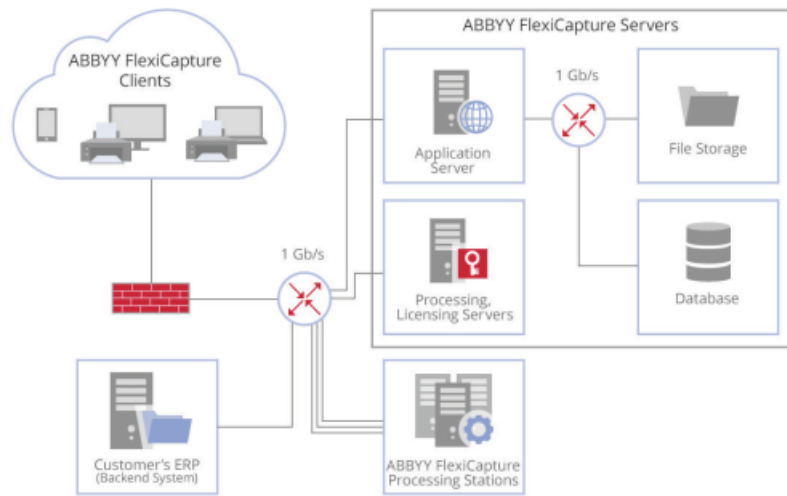
Pour de petites charges et de meilleures performances, vous pouvez utiliser des disques durs rapides sur la machine Application Server comme un FileStorage : par exemple des disques SATA2 à 15 000 tr/min ou plus, disposés au moins en RAID1 pour la redondance, ou en RAID10 pour une meilleure performance également.

Toutefois, à des stades ultérieurs du projet, si le volume de pages à traiter augmente, cette configuration résulte probablement en un goulot d'étranglement, en particulier pour le traitement des images en niveaux de gris ou en couleur, et le problème est qu'il ne peut pas être mis à l'échelle à la volée – il faudra que le système tombe en panne et que d'autres disques durs soient connectés.

Utiliser des stockages externes comme le NAS ou le SAN, auxquels le serveur d'application a accès en lecture-écriture à 1 Gb/s sur LAN, SCSI, Fibre Channel, etc. Cela permettra une mise à l'échelle en douceur du FileStorage.

Le texte suivant contient une explication sur la façon de calculer la performance requise du FileStorage du matériel.

Une configuration réseau FlexiCapture typique dans un environnement d'entreprise :



Rôle de la machine **Serveur d'application**

Conditions requises

CPU : 8 cœurs physiques, 2,4 GHz ou plus
16 GO DE RAM
DISQUE DUR : 100 GB
2 NICs, 1 Gb/s :

- un pour se connecter au réseau local et
- un pour se connecter au serveur de la base de données

FileStorage : Si un SAN est utilisé, connectez-le à l'aide de SCSI, Fibre Channel ou InfiniBand.

Système d'exploitation : Windows© 2012 ou ultérieur

Le serveur d'application est à la fois un service web et la plaque tournante de toutes les communications FlexiCapture :

- le transfert de grands corps binaires ;
- des réponses rapides aux petites demandes de service SOAP/JSON.

Les ressources essentielles sont :

- 1 Une interface réseau rapide pour la connexion aux clients ;
 - 2 Connexion rapide et stable au serveur de stockage de fichiers et de bases de données ;
 - 3 CPU multicœur à haute vitesse ;
 - 4 Mémoire vive suffisante, au moins 2 Go par cœur physique.
Si l'une de ces ressources provoque un goulot d'étranglement, augmentez la capacité du serveur d'application :
 - 5 Dans tous les cas, toutes les machines ayant le rôle de serveur d'application doivent être connectées de la même manière à la même base de données et au même stockage de fichiers.
- plus la vitesse est élevée, plus chaque demande est traitée rapidement,

- plus il y a de cœurs physiques, plus il y a de demandes traitées en même temps. Pour tirer le meilleur parti de l'unité centrale, pour le pool d'applications de services web FlexiCapture, il faut deux fois plus de processus de travail IIS, comme le nombre de cœurs physiques. Par exemple, 16 IIS Worker Processes pour un processeur à 8 cœurs ;
- via la technologie d'équilibrage de la charge du réseau Microsoft® – il regroupe plusieurs ordinateurs avec le rôle de serveur d'applications. Voir les instructions détaillées dans le guide de l'administrateur du système FlexiCapture,
- au niveau matériel en connectant une série de clients différents à des machines différentes avec le rôle de serveur d'application. Par exemple, vous pouvez utiliser une machine pour servir tous les traitements automatiques, et une autre pour les exposer à des clients externes.

Serveur de traitement, Serveur de licence	CPU à 4 cœurs, 2,4 GHz ou plus rapide 8 GO DE RAM DISQUE DUR : 100 GB NIC 1 GB/s pour la connexion au LAN Système d'exploitation : Windows Server © 2012 ou version ultérieure
--	---

Une connexion réseau stable est essentielle pour les serveurs, car, dans le cas contraire, le traitement des documents s'arrêtera.

Pour assurer la redondance, utilisez le cluster de basculement Microsoft®.

Voir les instructions détaillées dans le guide de l'administrateur du système FlexiCapture.

Le serveur de licence gère les copies des licences pour tous les clients concurrents dans sa mémoire. Concentrez-vous sur ce point si vous allez utiliser simultanément un grand nombre d'opérateurs de balayage et de vérification. Le serveur de licence est un service Windows® 32 bits, il ne peut donc pas occuper plus de 2 Go de mémoire vive (Remarque : dans FlexiCapture 12, il est déjà un processus de 64 bits). Selon les tests, 2 Go de RAM suffisent pour gérer les licences de 1000 clients. Considérez l'utilisation de plusieurs serveurs de licence pour desservir plusieurs clients simultanément.

Serveur de traitement, Serveur de licence	CPU à 4 cœurs, 2,4 GHz ou plus rapide 8 GO DE RAM DISQUE DUR : 100 GB NIC 1 GB/s pour la connexion au LAN Système d'exploitation : Windows® 2012 ou version ultérieure
--	---

Serveur de base de données	Pour MS SQL Server® : Base de données : MS SQL Server® 2014 ou supérieur, édition standard ou entreprise Le matériel : PROCESSEUR : 8 cœurs physiques, 3,4 GHz ou plus 16 Go de RAM ou plus DISQUE DUR : 400 GB Système d'exploitation : Windows® 2012 ou version ultérieure
Pour Oracle® :	Base de données : Oracle 12c Enterprise Edition Le matériel : Machine de base de données Oracle

Exadata X2-2, Quarter Rack

ABBYY FlexiCapture prend en charge MS SQL Server® et Oracle installés sur n'importe quelle plate-forme. Les deux serveurs de base de données conservent leurs propres enregistrements sur les paramètres optimaux, la mise à l'échelle et la tolérance aux pannes.

Recommandé pour le serveur MS SQL Server® :

- plus de RAM si possible sur la machine du serveur de base de données pour héberger la plus grande partie des fichiers de base de données dans RAM et d'y accéder plus rapidement ;
- un disque dur rapide pour un accès rapide à la partie de la base de données hébergée sur le disque (nous recommandons l'utilisation de SSD à cette fin) ;
- éviter les modes de la base de données avec des retards de transaction (Mirroring, etc.) ;
- choisir un modèle de récupération de base de données simple ;
- la base de données et son journal sont stockés sur des disques séparés ;
- mises à jour régulières de l'index pour les tables qui changent fréquemment (Document, Page, Batch, Task et EventLog). En cas d'échec, la taille d'un indice peut devenir supérieure à la taille des données du tableau.

FileStorage	NAS ou SAN Connecté via un réseau local, SCSI, Fibre Channel ou InfiniBand Vitesse de lecture-écriture : 100 Mo/s*. Capacité : 5 TB*.
-------------	--

- Les exigences en matière de lecture et d'écriture et de capacité dépendent fortement de ces deux facteurs : **nombre moyen et maximum de pages traitées par jour** (c'est-à-dire 24 heures) et par heure, et leur mode couleur. Comme mentionné dans la section sur les mesures de performance, nous pouvons estimer le flux d'entrée en octets par seconde si nous prenons quelques tailles de fichiers typiques pour les pages scannées en couleur, en niveaux de gris et en noir et blanc.

Les images constituent la majorité des données transférées dans le système. En analysant le traitement workflow, définissons les deux valeurs :

- le nombre R d'étapes où les images des pages **sont téléchargées à partir** du serveur d'application ;
- le nombre W d'étapes où les images des pages **sont téléchargées** vers le serveur d'application.

- 1 Les exigences en matière de vitesse de lecture-écriture peuvent être calculées comme suit :
- 2 **Exemple** : un client doit traiter 10 000 pages en niveaux de gris par heure. Le flux de traitement comprend trois étapes.
- 3 Une station de traitement télécharge les images d'un hot folder, les préreconnait et les télécharge vers le Serveur d'application (W=1, R=0).
- 4 Une autre station de traitement récupère ces images du serveur d'application, effectue la reconnaissance et les résultats de l'OCR arrivent au serveur d'application (W=1, R=1).
- 5 Un opérateur de vérification télécharge les images et les données reconnues pour la vérification et envoie les résultats vérifiés au serveur d'application (W=1, R=1). (W=1, R=2) .
- 6 Enfin, une station de traitement télécharge les images et les données vérifiées, pour les envoyer au système d'arrière-plan (W=1, R=3).

En supposant que la taille de fichier d'une numérisation moyenne en niveaux de gris A4 est de 3 Mo, nous avons les calculs suivants :

Flux d'entrée = 10 000 images de page en niveaux de gris/heure = 2,8 images en niveaux de gris/s = 8,4 Mo/s.

Vitesse d'écriture requise = 1 x 8,4 Mo/s = 8,4 Mo/s.

Vitesse de lecture requise = 3 x 8,4 Mo/s = 25,2 Mo/s.

Pour évaluer les performances du disque dur, vous pouvez utiliser un outil [CrystalDiskMark 2.2](#), distribué sous licence du MIT.

1 La durée de stockage des documents dans le système.

Exemple : un client doit traiter 100 000 images en niveaux de gris en 24 heures. Sous le niveau de service Accord, le délai de traitement est de deux jours par document. Les documents traités sont conservés pendant deux semaines en raison des contrôles supplémentaires dans le système ERP du client ; en cas de divergences, les documents sont édités dans FlexiCapture et téléchargés à nouveau dans le système ERP.

- Vitesse d'écriture requise = $W \times \text{débit d'entrée en octets par seconde}$.
- Vitesse de lecture requise = $R \times \text{flux d'entrée en octets par seconde}$.

Ainsi, les images doivent être stockées pendant $2+14 = 16$ jours, et le système accumulera $16 \times 100\,000$ niveaux de gris images $\times 3$ MB (taille moyenne du fichier pour une image en niveaux de gris A4) = 4,8 TB de données.

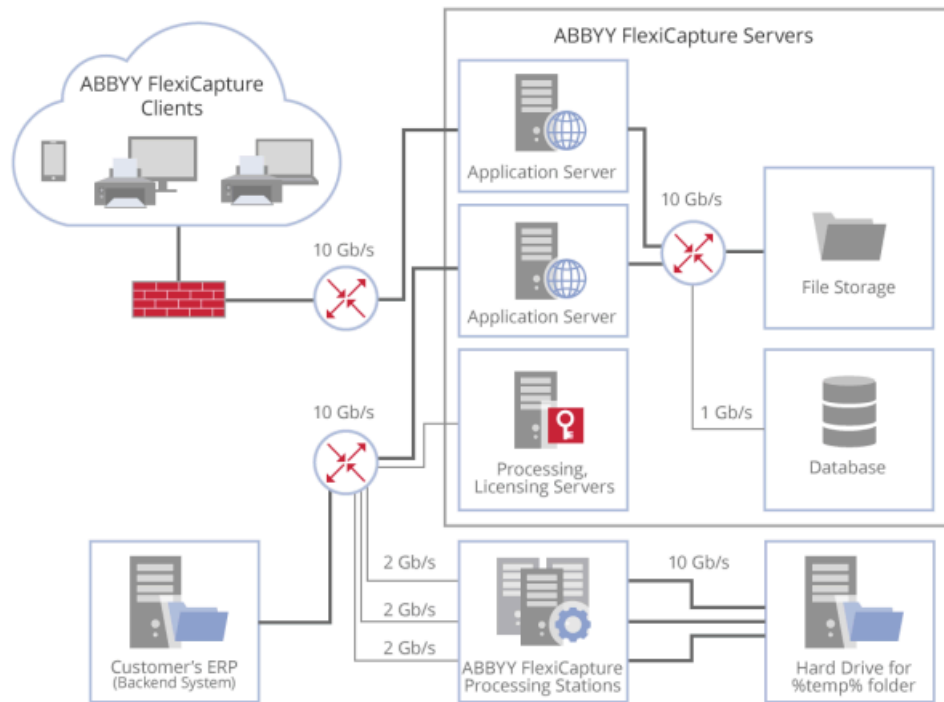


Remarque : nous recommandons vivement l'utilisation d'une technologie de stockage tolérante aux pannes, par exemple le RAID 10. L'indexation des recherches et l'analyse antivirus du contenu de FileStorage peuvent entraîner une diminution des performances ou bloquer l'accès aux fichiers, qui sont traités dans le système lui-même.

Une configuration importante est nécessaire lorsque vous traitez un volume important (plus de 300 000) de pages en couleur. Nous déclarons qu'elle peut atteindre jusqu'à 3 millions de pages en noir et blanc ou jusqu'à 1 million de pages en couleur en 24 heures.

Tout ce qui est mentionné ci-dessus à propos de la configuration moyenne reste valable pour la configuration large. La différence ici est que vous devez suivre toutes les recommandations d'optimisation et prêter une attention particulière à chaque partie du système – pour calculer la charge et choisir un matériel suffisamment puissant, mais pas trop cher. Entre autres choses, testez la connexion Internet et le connecteur dorsal pour vous assurer qu'ils peuvent fonctionner au niveau de performance souhaité.

Dès le début, envisagez d'utiliser un réseau de 10 Gb/s et un puissant FileStorage. L'architecture réseau possible pour cette configuration est présentée ci-dessous.



Au lieu de fournir des exigences système typiques pour les grandes configurations, nous recommandons d'examiner les configurations qui ont été testées, et leurs performances, comme indiqué dans ce tutoriel.

Pour obtenir des performances encore meilleures, il convient de combiner plusieurs installations FlexiCapture indépendantes sous un seul point d'administration et de contrôle – appelé configuration xLarge – ce qui dépasse le cadre du présent document.

VI-A - Estimation du nombre de cœurs de processeurs requis par le serveur d'application

Les demandes des clients sont traitées par le serveur d'application.

Plus les performances d'un cœur de processeur sont élevées, plus le traitement de chaque demande est rapide.

Toutefois, en règle générale, un grand nombre de clients différents enverront des demandes simultanément, de sorte que le serveur d'application devra les mettre en attente. Pour pouvoir traiter simultanément les demandes entrantes, vous devez déterminer le nombre optimal de cœurs de processeur pour votre nombre de clients.

Il existe trois types de clients dans le système :

- les stations automatisées ;
- les scanners des opérateurs ;
- les opérateurs de vérification.

Nous examinons ci-dessous chaque type de client en détail.

VI-A-1 - Stations automatisées

Des tests ont été effectués pour estimer le nombre de cœurs de processeurs requis par le serveur d'application pour desservir les stations automatisées.

Deux installations ont été testées, avec différents types de cœurs CPU utilisés pour le serveur d'application et les stations de traitement.

Les demandes des clients sont traitées par le serveur d'application.

	Serveur d'application	Stations de traitement
Installation 1	Intel Xeon® Platinum 8168 (SkyLake) 2.7 GHz	Intel Xeon® Platinum 8168 (SkyLake) 2.7 GHz
Installation 2	Intel Xeon® E5-2680 v4 2.4 GHz	Intel Xeon® E5-2680 v4 2.4 GHz

Les tests ont montré que 100 cœurs de traitement chargent complètement 6 cœurs du serveur d'application s'il n'y a pas de goulets d'étranglement.

Avec une utilisation du processeur du serveur d'application de 80 %, le serveur d'application devient lui-même un goulot d'étranglement. C'est pourquoi nous recommandons d'avoir huit cœurs pour le serveur d'application dans les installations comportant cent cœurs de traitement.

VI-A-2 - Opérateurs de vérification

Du point de vue du serveur d'application, les demandes provenant d'un opérateur de vérification ne sont pas très différentes des demandes envoyées par un cœur de traitement.

Toutefois, un opérateur humain est plus lent qu'une unité centrale et, dans le même laps de temps, il enverra moins de demandes qu'un cœur de traitement.

Pour des raisons pratiques, on peut supposer qu'en termes de charge générée, un cœur de traitement équivaut approximativement à 5 opérateurs de vérification.

Exemple.

Estimons le nombre de cœurs requis par le serveur d'application accédé par 100 vérifications opérateurs.

Un cœur de station de traitement génère autant de charge que 5 opérateurs de vérification.

Par conséquent, 100 opérateurs de vérification équivalent à 20 cœurs de traitement.

En tenant compte des résultats des tests décrits ci-dessus, nous supposons qu'un cœur de serveur d'application traite les demandes de 10 cœurs de stations de traitement.

Il s'ensuit donc que 2 cœurs de serveur d'application sont nécessaires pour traiter les demandes provenant de 100 vérifications opérateurs.

Remarque : pour les projets à forte charge comportant plus de 100 opérateurs de vérification et plus de 50 opérateurs de balayage, nous recommandons d'utiliser plusieurs serveurs d'application ou des installations en grappe. Cela permettra de rendre le système plus réactif et d'améliorer l'expérience des opérateurs.

VI-A-3 - Scanner les opérateurs

Du point de vue du serveur d'application, les demandes provenant d'un opérateur de balayage ne sont pas non plus très différentes des demandes envoyées par un cœur de traitement.

Toutefois, les demandes des opérateurs de balayage sont unidirectionnelles, car ils ne font que télécharger des données sur le serveur.

Tout comme un opérateur de vérification, un opérateur de scannage génère moins de charge qu'un cœur de traitement.

Pour des raisons pratiques, on peut supposer qu'en termes de charge générée, un cœur de traitement équivaut approximativement à 10 opérateurs de balayage.

VI-B - Stations de traitement

Les stations de traitement s'occupent des tâches de traitement automatisées, de la conversion des images en d'autres formats, de l'OCR (reconnaissance), de l'exécution des scripts utilisateur, etc. Elles sont toutes connectées au serveur de traitement qui surveille leur disponibilité pour attribuer des tâches de traitement parallèle ; une station par ordinateur.

Pour faire évoluer le système, vous devez :

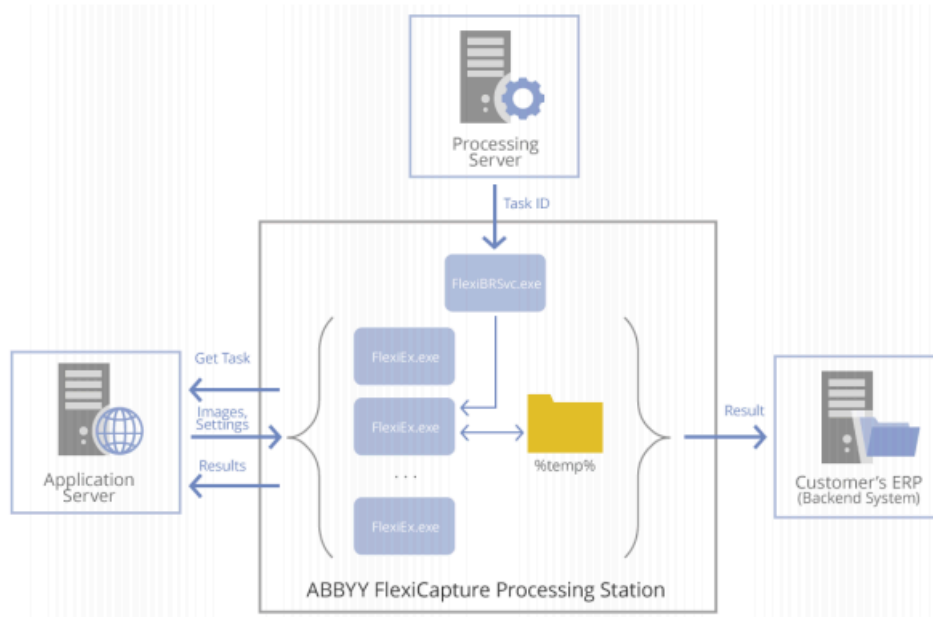
- régler les performances de chaque station de traitement afin d'accélérer le traitement ;
- ajouter d'autres stations de traitement jusqu'à ce que vous atteigniez le niveau de performance souhaité.

VI-B-1 - Comment régler les performances d'une station de traitement ?

Une station de traitement est une application de service Windows®. Pour traiter une tâche, une station :

- se connecte au serveur de traitement, pour obtenir les identificateurs des tâches à traiter ;
- se connecte au serveur d'application via HTTP/HTTPS et télécharge des images, des données de documents, et les paramètres de traitement ;
- lance plusieurs processus exécutifs pour effectuer des tâches de traitement ;
- télécharge les résultats sur le serveur d'application ou sur un système de gestion (par exemple, un système ERP ou DMS).

Ces processus utilisent intensivement le disque dur pour sauvegarder les données de traitement intermédiaires dans un dossier temporaire.



Le matériel utilisé pour les stations de traitement a un impact considérable sur les performances de FlexiCapture.

Rôle de la machine

Station de traitement

Conditions requises

CPU : 8 cœurs physiques avec Hyper-Threading, 2,4 GHz ou plus

16 Go de mémoire vive (RAM)

DISQUE DUR : 150 GB

NIC : 1 GB/s

Système d'exploitation : Windows© 2012
ou version ultérieure

Une station lance un processus exécutif pour chaque cœur de processeur, ce qui permet de traiter plusieurs tâches simultanément. Pour améliorer les performances du CPU, utilisez l'Hyper-Threading lorsque c'est techniquement possible.



Remarque : l'utilisation de plus de 16 cœurs de processeur logiques est un mauvais choix : plusieurs processus exécutifs se feront concurrence pour le temps du disque dur et la mémoire cache du CPU.

Au moins 1 Go de RAM par cœur logique est suffisant pour le traitement.

La vitesse de traitement dépend fortement de la fréquence du processeur et de la vitesse de lecture-écriture du disque dur. Il est recommandé de configurer un disque dur rapide pour une station de traitement, ou de combiner plusieurs disques durs en RAID0 pour obtenir une plus grande vitesse d'accès dans les processus exécutifs aux dossiers temporaires.

Si la quantité de mémoire vive disponible est supérieure à la quantité recommandée de 1 Go par cœur logique, il est recommandé de créer un disque dur virtuel dans la mémoire vive et y placer un dossier temporaire pour les processus exécutifs – cela peut entraîner jusqu'à 30 % de gain de vitesse de traitement.



Note : comment estimer la taille d'un dossier temporaire pour les processus exécutifs.

L'**espace maximum requis sur le disque dur pour un dossier temporaire** est en fait la taille totale du document dans un lot typique, en Mo, multiplié par le nombre de processus d'exécution, qui par défaut est le nombre de cœurs de processeurs logiques.

Exemple : calculons la taille maximale d'un dossier temporaire dans une configuration où les images en niveaux de gris sont traitées par lots de 100 pages sur une station à 8 cœurs avec Hyper-Threading activé.

La taille d'un lot en MB = 100 pages x 3 MB, dont la taille typique d'une page en niveaux de gris en MB = 300 MB.

Un ordinateur à 8 cœurs avec Hyper-Threading activé fournit 16 cœurs logiques, la station de traitement exécutera donc 16 processus exécutifs simultanés. Ainsi, l'espace requis pour le dossier temporaire est de 300 Mo x 16 processus exécutifs = 4,8 Go.

Si le dossier temporaire est hébergé en RAM, la taille de la RAM requise est de 1 Go pour chaque cœur logique, comme requis pour le traitement x 16 processus exécutifs + 4,8 Go pour le dossier temporaire = environ 21 Go de RAM.

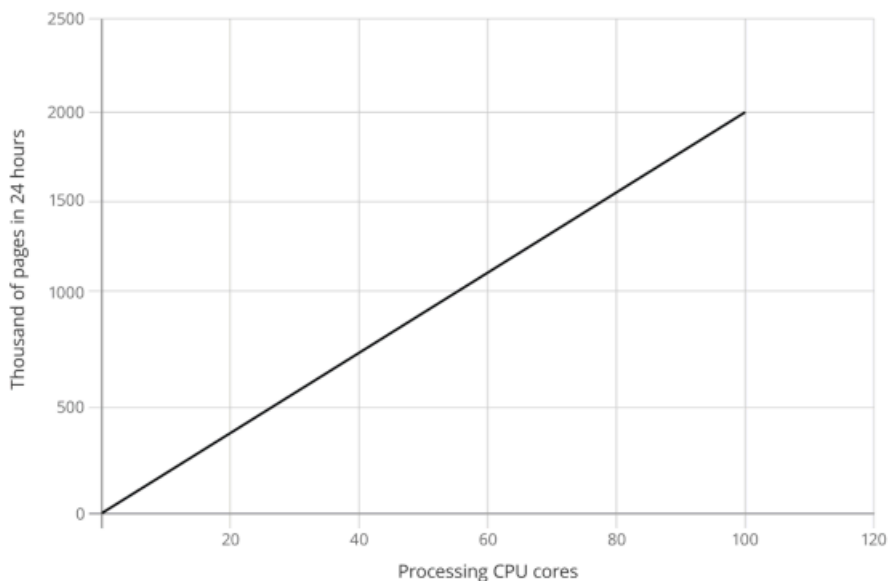
Il n'est pas nécessaire de prévoir une redondance pour les disques durs de la station de traitement. En cas de panne, seuls les résultats du traitement en cours seront perdus, les images seront transmises à une autre station de traitement et traitées dans celle-ci – pour cela, il faut certainement disposer d'au moins deux stations de traitement dans le système.

VI-B-2 - Comment calculer le nombre de postes de traitement ?

Pour tirer le meilleur parti des ressources informatiques, chaque station exécute plusieurs threads de traitement en même temps ; plus il y a de cœurs de processeur disponibles, plus il y a de threads parallèles traités. Comme le nombre de cœurs de processeur varie d'un ordinateur à l'autre, il est logique de compter le nombre total de cœurs de processeur de traitement dans le système FlexiCapture.

S'il n'y a pas de goulets d'étranglement dans le système, chaque nouveau cœur de traitement contribue de la même manière à la performance de l'ensemble du système. Vous devez donc estimer la contribution d'un cœur, puis estimer le nombre de cœurs nécessaires pour atteindre les performances visées.

Le nombre de pages qu'un cœur de traitement est capable de traiter pendant une période donnée dépend largement du flux de travail (par exemple, le nombre d'étapes), des paramètres de traitement (opération d'amélioration de l'image, mode de reconnaissance, paramètres d'exportation), de la mise en œuvre d'étapes personnalisées (moteurs et règles de script personnalisés, accès à des ressources externes) et du matériel. Lorsque vous n'avez aucune idée de ces détails, mais que vous avez déjà besoin d'une estimation, vous pouvez utiliser le graphique suivant comme référence. Cependant, il est fort probable que vous obteniez d'autres résultats dans le cadre de votre projet.



La dépendance de la performance sur le nombre des cœurs de traitement

Configuration

La démo « SingleEntryPoint » (point d'entrée unique)

projet : traitement sans surveillance, l'exportation vers des fichiers PDF.

Pour les pages en noir et blanc :

10 stations de traitement de base,

2,4GHz, 16Go de RAM, SSD,

1 Gb/s NIC

Pour estimer le nombre de cœurs de traitement nécessaires, vous pouvez procéder comme suit :

- 1 Configurez le déroulement de votre projet, prenez la station de traitement la plus proche en termes des paramètres matériels de ce qui va être utilisé dans la production, et créez un lot type d'images ;
- 2 Nous allons mesurer le temps nécessaire au traitement d'un lot pour un cœur.
Il ne suffit pas de traiter un lot une seule fois, car FlexiCapture peut distribuer les traitements entre tous les cœurs disponibles, et il faudra moins de temps pour traiter un lot pendant des tests tandis que pendant la production réelle, d'autres cœurs seront occupés à traiter d'autres lots.

Pour obtenir une estimation fiable, nous recommandons de créer plusieurs copies du lot type – au moins le même nombre que le nombre de cœurs, mais il vaut mieux multiplier par N (qui est au moins 3) pour minimiser l'erreur de mesure, et les placer tous dans le traitement simultané. Le temps nécessaire pour traiter un lot par cœur est alors le total temps de traitement divisé par N. Cette estimation tient compte de la concurrence éventuelle entre les cœurs de traitement sur les ressources partagées de la station de traitement.

Exemple : Nous avons une station de traitement à 8 cœurs avec hypertreading activé, ce qui nous donne 16 cœurs logiques et processus exécutifs à cette station. Nous devons créer au moins 16 copies d'un lot typique, mais nous ferions mieux de créer $16 \times 3 = 48$ copies pour minimiser les erreurs de mesure. Nous mettons tous les lots dans le classeur à chaud FlexiCapture, démarrons le minuteur à la première tâche d'importation créée et l'arrêtons après que le dernier résultat a été exporté vers le backend – il affichera 15 minutes. Cette fois-ci, chaque cœur doit traiter 3 lots, d'où le temps Le traitement d'un lot est d'environ 5 minutes. Notre lot compte 69 pages, et nous pouvons dire qu'il faut 4,35 secondes pour traiter une page ;

- 3 Une fois que nous connaissons le rendement souhaité en pages par heure ou par jour, nous pouvons établir une estimation du nombre de cœurs souhaité.

Supposons que vous deviez traiter P pages en T temps. Nous savons déjà, d'après ce qui précède, qu'un cœur a besoin de t temps pour traiter 1 page. Il faut donc $N = (P \times t) / T$ cœurs.

Exemple : un client a besoin de traiter 200 000 pages en 8 heures, soit 28 800 secondes. Comme nous le savons d'après ce qui précède, il faut 4,35 secondes à un cœur pour traiter une page. Par conséquent,

nous avons besoin de $(200\ 000 \times 4,35) / 28\ 800 = 31$ cœurs. Ainsi, 2 stations de traitement avec 8 cœurs et hyperthreading activé (32 cœurs logiques au total) seront suffisantes pour le traitement automatique.

Il existe deux facteurs limitant le nombre de cœurs de traitement dans le système :

1 La charge totale sur l'infrastructure qui peut entraîner des goulets d'étranglement :

- sur le matériel du serveur FlexiCapture,
- sur le réseau ou
- sur les ressources partagées externes (comme les bases de données, les services externes, etc.) qui sont demandées à partir de scripts de traitement personnalisés.

Un goulot d'étranglement entraînera une saturation des performances – l'ajout d'un nouveau cœur de traitement aura un effet négatif ou simplement aucun effet sur les performances totales. Ce document décrit comment concevoir le système pour éviter les goulets d'étranglement (voir ci-dessus) et comment surveiller le matériel et l'infrastructure pour détecter les goulets d'étranglement.

Cependant, même s'il n'y a pas de goulets d'étranglement clairement détectés, la concurrence entre les cœurs de traitement pour les ressources partagées s'accroît lorsque de nouveaux cœurs sont ajoutés au système. Si vous devez utiliser plus de 50 % de la capacité de lecture/écriture du réseau ou de FileStorage (selon les calculs effectués dans ce document), ajoutez 20 % au temps de traitement de chaque page dans les exemples ci-dessus – cela nécessitera en fait 20 % de cœurs de traitement supplémentaires dans le système.

Utilisez la mise en cache des cœurs de traitement pour accéder plus rapidement aux ressources externes – par exemple, au lieu de vous connecter directement à la base de données, la connecter à l'ensemble de données FlexiCapture et demander ensuite l'ensemble de données à partir des scripts ;

2 Le nombre de cœurs de traitement pouvant être desservis par le serveur de traitement. Ce nombre dépend du temps moyen dont un cœur a besoin pour effectuer une tâche. Le temps moyen dépend largement de la taille du lot dans les pages et de la personnalisation mise en œuvre. Habituellement, si vous avez environ 10 pages dans un lot, le serveur de traitement est capable de servir 120 cœurs de traitement. Cependant, si vous créez un grand nombre d'étapes personnalisées avec des scripts très rapides, ou si vous allez traiter une page par lot, vous diminuerez considérablement le temps moyen de la tâche, ce qui peut entraîner une légère diminution du nombre maximum de cœurs de traitement.

Note : *pour détecter ce problème, vous devez surveiller le compteur de cœurs de traitement libres sur le serveur de traitement. Si vous constatez que, malgré cela, vous avez une file d'attente de documents à traiter, que le nombre de cœurs occupés a atteint un certain point de saturation et ne dépasse presque jamais, vous avez obtenu l'effet décrit. Pour y remédier :*

- *traitez l'ensemble du lot sans le diviser en petites tâches, si possible (voir les propriétés des étapes dans la boîte de dialogue des paramètres du flux de travail) ;*
- *traitez les pages par portions plus importantes : augmentez le nombre moyen de pages par lot ;*
- *fusionnez plusieurs étapes de personnalisation en une seule, ou amenez la personnalisation à un stade standard, par exemple en l'ajoutant à un événement de routage dans le scénario de cette étape.*

VI-C - Stations de scannage

ABBYY FlexiCapture permet d'importer des images à partir d'un scanner personnel (via un client léger ou un client riche local), ou d'un scanner réseau (les images vont dans un dossier ou une boîte de réception de courrier électronique), ou encore à partir d'une application mobile. Les performances de chaque client de numérisation sont limitées par la **vitesse du scanner** et la **bande passante de transfert des données**.

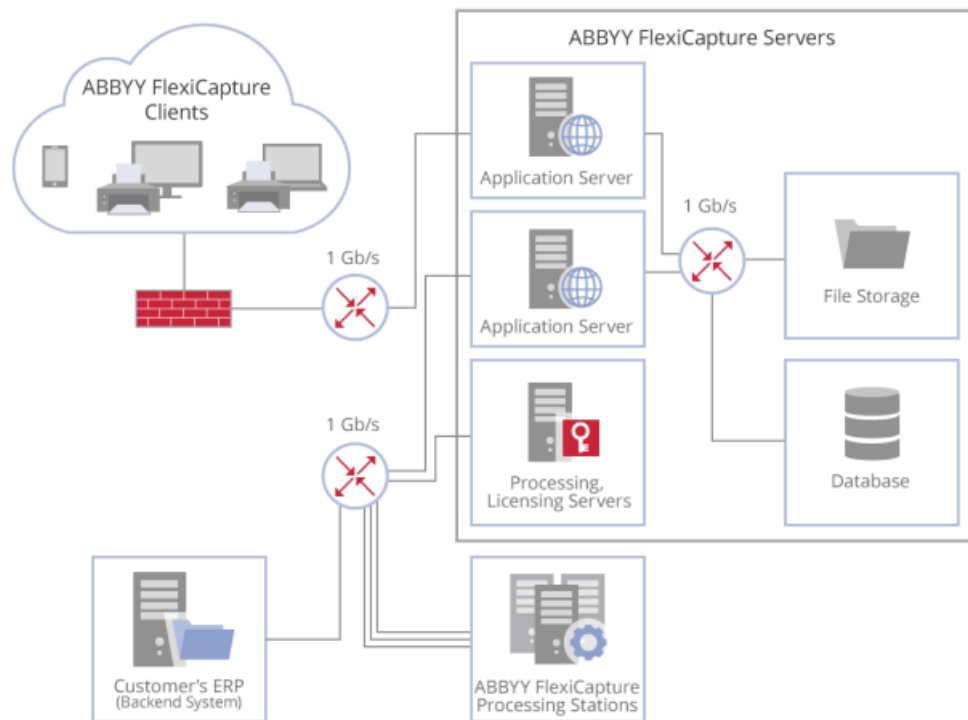
Le nombre total de clients de balayage n'est pas aussi important pour la performance que le **flux d'entrée** total – le **nombre moyen et maximal de pages traitées par heure ou par 24 heures**, et la **taille de chaque page**, en fonction de son **mode de couleur**. Le débit d'entrée de pointe ne doit pas dépasser les capacités du système.

Le trafic de toutes les stations de balayage, de vérification et de traitement passe par le même canal à la passerelle du serveur d'application.

Lorsque le trafic des stations de balayage occupe la moitié de la bande passante du canal ou plus, ou présente des pics importants, allouez une interface réseau séparée sur le serveur d'application aux clients de balayage. Cela permet d'éviter les situations où les pics de trafic entraînent des retards sur les stations de vérification et les stations de traitement.

Si le serveur d'application est déployé sur un groupe de plusieurs ordinateurs, le trafic peut être réparti entre eux par l'un ou l'autre :

- en utilisant les paramètres d'affinité du NLB pour le cluster (*le niveau du logiciel*) ;
- en acheminant les connexions réseau vers des nœuds de grappe spécifiques (*le niveau matériel*).



Numérisation et vérification

Les clients utilisent une connexion indépendante du serveur d'application.

VI-C-1 - Support des stations de scannage

- Reprise automatique des téléchargements d'images vers le serveur d'application après l'échec de la connexion réseau. Cela permet d'atténuer les pics de trafic provenant des stations de balayage.
- Paramétrage centralisé des options de numérisation, d'amélioration des images et d'exportation vers le serveur d'applications – par exemple, vous pouvez définir le mode couleur des images numérisées, détecter et supprimer toutes les pages vides inutiles, produites par la numérisation recto-verso, pour réduire le flux d'entrée vers le serveur d'applications.

- Programmation des téléchargements d'images vers le serveur d'application pour équilibrer les charges du réseau (par exemple en attribuant des temps de téléchargement différents aux différents bureaux régionaux).

VI-D - Stations de vérification

Les documents traités automatiquement peuvent être vérifiés manuellement, si nécessaire. Pour cette raison, ABBYY FlexiCapture fournit des clients de vérification riches et légers.

La vérification est un processus lent et coûteux. ABBYY FlexiCapture offre une fonctionnalité de vérification automatique des règles de validation qui permettent de valider automatiquement les documents, ce qui signifie que la vérification manuelle peut être évitée.

Un autre moyen de réduire le travail de vérification consiste à préciser avec le client quels sont les champs du document qui doivent être extraits avec une qualité de 100 % – parfois, il ne s'agit pas de tous les champs du document, ce qui permet également de concentrer la vérification uniquement sur les documents présentant des problèmes dans ces champs.

Pour calculer le nombre d'opérateurs de vérification, vous devez comprendre le nombre de documents à vérifier traités, combien d'entre eux nécessitent une vérification, le délai de traitement du document selon le contrat de niveau de service et le temps moyen nécessaire pour vérifier un document.

Les vérificateurs génèrent également une charge de travail sur le système. Une station de vérification interagit avec le serveur d'application de la même manière qu'une station de traitement : il demande des tâches et télécharge des images et des données de documents à partir du serveur d'application, et renvoie les données modifiées.

- La vitesse de traitement des stations de vérification est beaucoup plus lente, car la vérification manuelle prend généralement beaucoup plus de temps que le traitement automatique sur une station de traitement.
- Les opérateurs de vérification n'ont pas toujours besoin de voir les images des documents dans leur qualité d'origine. Les paramètres de FlexiCapture permettent de modifier la compression (qui est de 60 % par défaut) des images téléchargées par les opérateurs à partir du serveur d'application.

Ainsi, nous supposons qu'un vérificateur travaillant au maximum de sa capacité génère jusqu'à 1/5 de la charge créée par un cœur de traitement d'une station de traitement.

Vous pouvez utiliser cette hypothèse pour interpréter les résultats des tests, effectués sans vérificateurs, en utilisant uniquement le traitement sans surveillance : si vous constatez un fonctionnement stable du système avec, par exemple, 100 cœurs de traitement, cela signifie que vous pouvez remplacer en toute sécurité un certain nombre d'entre eux par un certain nombre d'opérateurs de vérification travaillant simultanément, multiplié par 5.

Exemple : *un client doit traiter 100 000 documents en 8 heures de travail.*

Comme prévu initialement, seuls 30 % des documents devront être vérifiés manuellement, et la vérification de chaque document prend jusqu'à 2 minutes. Par conséquent, jusqu'à 125 vérificateurs seront nécessaires.

Chaque document compte environ 3 pages en moyenne. Vous pouvez créer un lot de test à partir de documents types et tester le système avant de le mettre en service en mode de traitement non surveillé. Disons que le système est stable et que vous ne voyez pas de congestions en utilisant 100 cœurs de traitement, alors que 60 cœurs de traitement suffisent déjà pour traiter la quantité souhaitée de 300 000 pages en 8 heures. Ainsi, le système pourra facilement traiter 125 vérificateurs sur 60 cœurs de traitement (puisque la limite supérieure estimée pour 125 vérificateurs est de 25 cœurs).

VI-E - Flux de travail

La configuration du flux de travail a un impact important sur les performances du système et la charge du matériel.

Les considérations ci-dessus portent sur la charge, produite par le flux de travail par défaut qui contient les étapes de prétraitement, de reconnaissance, de vérification et d'exportation.

Pour répondre aux exigences de projets spécifiques, vous pouvez ajouter des étapes de traitement supplémentaires, les réorganiser et configurer des règles de routage sophistiquées. Vous devez garder à l'esprit ce qui suit.

- 1 Évitez de multiplier les étapes
 Chaque étape augmente le volume des ressources nécessaires – télécharger les données à traiter, obtenir quelqu'un pour effectuer le traitement et renvoyer les résultats du traitement au serveur – et, par conséquent, le coût total du projet.
 Par exemple, si vous allez ajouter une nouvelle étape personnalisée pour un script automatique, envisagez la possibilité d'exécuter ce scénario en utilisant des règles, ou des événements prédéfinis, ou de le combiner avec une autre étape existante.
- 2 L'étape la plus lente limite la performance
 En général, les étapes les plus lentes sont celles qui nécessitent un travail manuel. Il est moins évident que, même en cas de traitement non surveillé, des goulets d'étranglement peuvent apparaître, causés par des scripts personnalisés non optimaux ou un accès lent à des ressources externes non mises en cache.
 Observez les files d'attente aux différentes étapes du flux de travail en utilisant la console d'administration et de surveillance pour identifier l'étape la plus lente. Envisagez la possibilité d'accélérer l'étape ou au moins de mettre en parallèle le traitement en utilisant l'option « Documents par tâche » dans les propriétés des étapes.
- 3 Ne produisez pas de tâches trop petites en parallélisant le traitement à une étape
 Lorsque vous mettez le traitement en parallèle à une étape, évitez de diviser le traitement en trop de morceaux ; la manipulation de chaque morceau nécessitera un travail supplémentaire de la part du système.
 En particulier, un grand nombre de très petites tâches automatiques peuvent ralentir le serveur de traitement qui répartit chaque tâche entre les exécuteurs.
 Si vous devez accélérer une étape d'un facteur de deux seulement et que vous avez généralement 10 documents dans un lot, il suffit déjà de créer une tâche pour 2 séries de 5 documents chacune au lieu d'une tâche pour l'ensemble du lot comme par défaut. Toutefois, essayez d'éviter de créer une tâche par document, alors que vous n'en avez pas réellement besoin.
 N'oubliez pas non plus que la création d'une tâche plus petite qu'un lot limite l'agilité de l'exécuteur : si, dans certains scénarios, un vérificateur peut travailler avec chaque document indépendamment, alors pour l'assemblage automatique des documents, il est essentiel d'avoir toutes les pages d'un lot comme une seule tâche.

VII - Valeurs optimales et limites des performances du système

Facteurs d'influence	Valeurs optimales et limites	Commentaires
Performance du système en pages par 24 heures	FlexiCapture est capable de traiter : - jusqu'à 20 000 pages en noir et blanc ou jusqu'à 1000 pages en couleur par 24 heures sur un ordinateur personnel ; - jusqu'à 1 million de pages en noir et blanc ou jusqu'à 300 000 pages par 24 heures, en utilisant une ferme d'ordinateurs ordinaires (Medium) configurés selon les exigences du présent document ; - jusqu'à 3 millions de pages en noir et blanc ou jusqu'à 1 million de pages en couleur par	

24 heures, en utilisant le matériel de niveau production de l'entreprise (Grand), comme indiqué dans les résultats des tests de performance ci-dessous.

Nombre d'opérateurs de scannage FlexiCapture est capable de accueillir 1000 opérateurs de scannage. Cette valeur est principalement limitée par la quantité de trafic produite aux heures de pointe et en moyenne – voir la section spéciale (page 21) de ce document.

Nombre d'opérateurs de vérification FlexiCapture est capable de accueillir 500 opérateurs de vérification. Nous supposons que la charge du système produite par 5 opérateurs de vérification est égale à la charge d'un cœur sur une station de traitement. Cette connaissance peut être utilisée pour prévoir le nombre de vérificateurs autorisés sur la base des tests de traitement non surveillés. Voir la section spéciale (page 22) dans le document.

Nombre de stations de traitement Nous avons utilisé jusqu'à 200 cœurs au total pour l'ensemble des stations de serveurs d'application et traitement. Cela dépend en grande partie de la puissance des serveurs d'application et du temps nécessaire pour traiter une tâche dans une station de traitement. Voir la section spéciale (page 18) dans le document.

Nombre de cœurs par station de traitement Avec un lecteur de disque normal (SATA2 7 500 tr/min) : jusqu'à 8 cœurs logiques.
Avec un lecteur de disque rapide (SAS 15 000 tr/min) : jusqu'à 16 cœurs logiques.
Avec lecteur RAM : jusqu'à 32 cœurs logiques.

Nombre de pages dans un lot La valeur optimale est de 10 à 1000 pages par lot. Les petits lots (3 pages ou moins) entraînent une surcharge de traitement par page, de sorte que la performance totale en pages en 24 heures diminue. En particulier, le nombre maximum de cœurs du serveur de traitement peut diminuer en raison de la taille trop réduite des tâches.

		De très grands lots (2000 pages et plus) imposent une charge trop importante au serveur d'application et à la base de données lors du routage d'une étape à l'autre. Ils peuvent également se heurter à des limites de temps et de taille maximale des requêtes dans le réseau et les paramètres des logiciels sous-jacents.
Nombre de pages d'un document	La valeur optimale est de 100 pages maximum dans un document ad hoc	Les documents volumineux peuvent entraîner des lenteurs dans le travail des opérateurs : il faut beaucoup de temps pour charger toutes les images des pages et calculer les règles qui utilisent un grand nombre de champs, par exemple, de grands tableaux à plusieurs pages.
Nombre de pages, de documents et de lots dans le système	Cela dépend fortement du matériel utilisé. Pour une grande configuration, jusqu'à 100 000 lots, ou 1 million de documents, ou 10 millions de données de pages est normal.	Un très grand nombre de pages, de documents et de lots dans le système peut conduire à ce que le serveur de base de données agisse plus lentement sur les requêtes. Nous recommandons alors d'utiliser un matériel plus puissant pour le serveur de base de données et de reconstruire périodiquement les index des tables.
Durée de stockage des données	Les magasins FlexiCapture : <ul style="list-style-type: none"> pages, documents et lots, qui sont ou ont été traités ; le journal des événements enregistre le traitement ; des statistiques sur le traitement en vue de l'établissement de rapports. <p>En règle générale, les pages, documents, lots et journaux d'événements sont stockés dans le</p>	Nous recommandons de supprimer les pages, les documents, les lots et les enregistrements du journal des événements associés peu après le traitement - le plus tôt sera le mieux. Le stockage de lots traités pendant une longue période dans le système entraînera leur accumulation dans celui-ci, ce qui pourrait ralentir la base de données.

système pendant deux semaines au maximum. Les statistiques pour l'établissement de rapports peuvent être stockées pendant des années sans incidence sur les performances.

VIII - Surveillance du système et détection des goulets d'étranglement

La surveillance du système comprend :

- la surveillance du traitement des documents via la console d'administration et de surveillance ;
- la surveillance du matériel pour chaque composant du serveur FlexiCapture en utilisant les différentes performances de Windows®. Surveiller les compteurs.

Vous pouvez effectuer une surveillance du hardware de l'ensemble du système en utilisant l'application gratuite Web Performance Monitor (<http://www.iis.net/downloads/community/2007/01/web-performance-monitor>) , le plus puissant gestionnaire des opérations du centre de système Microsoft®, ou des outils similaires.

Ce sont des paramètres clés à suivre sur chaque ordinateur, quel que soit son rôle dans FlexiCapture.

REMARQUE : la pénurie d'une seule ressource peut entraîner une surcharge de n'importe quel composant, par exemple une pénurie de mémoire vive peut entraîner une utilisation très intensive du disque dur. C'est pourquoi l'ordre dans lequel vous enquêtez sur les paramètres pour trouver un goulot d'étranglement est important. Veuillez suivre le modèle comme dans le tableau.

Mémoire

Lorsque le compteur **Memory\Available Bytes** (la mémoire non occupée par les processus en cours d'exécution et un cache sur le disque dur) est constamment faible, tandis que le compteur **Memory\Pages/sec** (le nombre de pages de mémoire demandées sur le disque dur ou téléchargées sur le disque dur pour en libérer davantage RAM) est en constante évolution, il est probable que l'ordinateur ne dispose pas d'une mémoire vive suffisante.

Les compteurs **Process(<Toutes les instances>)\Working Set** indiquent le nombre de pages de mémoire allouées à chaque processus.

- Chaque processus peut retenir une grande quantité de mémoire et la mémoire totale disponible peut être faible – mais cela ne signifie pas qu'il y a un manque de mémoire dans le système.
- Cependant, si vous constatez que plusieurs processus agrandissent

leurs ensembles de travail et que certains autres les réduisent légèrement, que la quantité de mémoire disponible est faible et que le compteur de pages/seconde augmente constamment, il est probable qu'il s'agisse d'un goulot d'étranglement de la RAM.



REMARQUE : les processus 32 bits ne peuvent pas allouer plus de 2 Go de RAM, même s'il y a beaucoup de la mémoire vive disponible dans le système.

Unité centrale

Pour plus de détails, voir <https://technet.microsoft.com/en-us/library/cc938577.aspx>.

Lorsque le **Process(<Toutes les instances>)\% Processor Time** du processeur (le pourcentage de temps que le processeur est occupé) affiche plus de 80 % pendant des périodes importantes, et le compteur **System \Processor Queue Length** (le nombre de threads dans la file d'attente du processeur) dépasse le double du nombre de CPU, alors le CPU est très probablement à l'origine d'un goulot d'étranglement.

Pour plus de détails, voir <https://technet.microsoft.com/en-us/library/cc938609.aspx>.

Le compteur **Process(<Toutes les instances>)\% Processor Time** permet de déterminer quels sont les processus qui « consomment » le temps CPU.

Disque dur

Lors de la vérification du disque dur, assurez-vous que le système dispose de suffisamment de mémoire vive (voir la section Mémoire colonne ci-dessus).

Le compteur **LogicalDisk(<All instances>)\Free Megabytes** indique l'espace libre sur un disque logique. Si l'espace libre n'est pas suffisant, les performances du système diminueront considérablement.

Lorsque le compteur de temps de disque **PhysicalDisk(<Toutes les instances>)\%DiskTime** (le pourcentage de temps le disque consacré au traitement des demandes de lecture et d'écriture) montre plus de 90 %, et le **PhysicalDisk(<Toutes les instances>)\Avg. Compteur de longueur de file d'attente de disque** (le nombre

Réseau

moyen de demandes en attente et en cours pendant la période de suivi) montre constamment plus que 2 par fil de disque dur, alors le disque dur est probablement à l'origine d'un goulot d'étranglement.

Pour plus de détails, voir <https://technet.microsoft.com/en-us/library/cc940380.aspx>.

Lorsque le compteur de la longueur de la file d'attente de sortie **Network Interface(<Toutes les instances>)\Output Queue Length** (le nombre de paquets réseau sortants dans une file d'attente) affiche constamment plus de 2, l'adaptateur réseau est très probablement en attente de la connexion, ce qui retarde les demandes du serveur. Lorsque le compteur de l'interface réseau **Network Interface(<Toutes les instances>)\Packets Outbound Discarded** (compteur de paquets sortants rejetés) est en constante augmentation, le canal est tellement surchargé que la mémoire tampon de l'adaptateur réseau ne peut pas traiter toutes les demandes sortantes.

Lorsque le compteur de l'interface réseau **Network Interface(<Toutes les instances>)\Bytes Total/sec** (la quantité de données transitant par le NIC) fait 65 % (ou plus) du compteur **Network Interface(<All instances>)\Current Bandwidth** (la bande passante disponible), utiliser un canal avec une bande passante plus élevée ou segmenter le réseau pour minimiser les collisions dans le canal.

Rôle de la machine Serveur d'application

Compteur à surveiller

Le serveur d'application est un service web sur IIS servi par les processus w3wp.exe.

Le temps CPU et la quantité de mémoire utilisés par les instances de ces processus doivent être supervisés. La charge du réseau de connexion de données (DCN) sur un ordinateur avec le serveur d'application mérite une attention particulière.

Utilisez ces compteurs pour suivre les chargements de l'IIS :

- **Service web (site web par défaut)**
\Octets reçus/sec ;
- **Service web (site web par défaut)**
\Octets envoyés/sec.

Lorsqu'ils affichent plus de 65 % de la bande passante DCN disponible, utilisez l'une de ces solutions :

- augmenter la largeur de bande ;
- segmenter le réseau ;
- ajouter plus d'ordinateurs exécutant IIS, et équilibrer les charges.

Lorsque la valeur du compteur **W3SVC_W3WP(_Total)\Active Threads Count** (le nombre de requêtes de maintenance des threads) atteint la valeur du compteur **W3SVC_W3WP(_Total)\Maximum Threads Count** (le nombre maximum de threads disponibles pour l'entretien), IIS est fortement surchargé.

Utilisez ces compteurs pour suivre les pics d'activité actuels par rapport à activités antérieures :

- **W3SVC_W3WP(_Total)\Active Requests** (le nombre de demandes actives) ;
- **W3SVC_W3WP(_Total)\Requests/sec** (le taux de traitement des demandes) ;
- **Web Service(Default Web Site)\Current Connections** (le nombre de connexions actives entre les clients et le service web).

Utilisez les compteurs du serveur d'application et des stations de traitement pour estimer le temps de réponse du serveur de traitement :

FlexiCapture(All instances)\ASCT Latency – Latence du fil de communication du serveur d'application (en millisecondes).

La valeur idéale est proche de 0 ; 1 seconde est la norme. À des valeurs supérieures à 30 secondes, le serveur de traitement cesse de prendre en charge de nouvelles tâches de traitement. Cela signifie que le serveur d'application ou le serveur de base de données est surchargé.

Serveur de traitement

Le serveur de traitement est un service Windows® – le processus FlexiBrSvc.exe.

FlexiCapture(Processing Server)\Primary Thread Latency indique le temps nécessaire au serveur de traitement pour répondre à ses clients – le moniteur du serveur de

traitement et les stations de traitement (en millisecondes).

La valeur idéale est proche de 0. Les valeurs supérieures à 10 secondes indiquent de sérieux problèmes soit des charges excessives sur l'ordinateur du serveur de traitement, soit une connexion défectueuse avec un ou plusieurs postes de traitement.

FlexiCapture(Serveur de traitement)\Cores Count indique le nombre total de traitements des cœurs de processeurs sur toutes les stations de traitement dans l'état « Started ».

FlexiCapture(Processing Server)\Free Cores indique le nombre total des cœurs de processeurs de traitement disponibles qui ne traitent rien. Ce compteur permet de déterminer s'il y a suffisamment de cœurs de processeur de traitement dans ABBYY FlexiCapture.

Certains cœurs sont toujours libres alors qu'il y a une file d'attente aux étapes de traitement automatique.

Cela peut se produire parce que :

- certaines stations de traitement sont configurées pour traiter un type de tâches spécifiques et lorsqu'ils sont gratuits, la puissance de traitement est insuffisante pour les autres types de tâches ;
- le serveur de traitement connaît un goulot d'étranglement :
 - vérifiez s'il y a un manque de mémoire vive, de processeur, de disque dur ou de ressources réseau sur le machine serveur de traitement,
 - augmenter la taille de la tâche, car les cœurs de traitement traitent les tâches trop rapidement et le serveur de traitement ne répartit pas les tâches entre eux.

FlexiCapture(Processing Server)\Pending Tasks : indique le nombre de tâches demandées au serveur d'application, mais non attribuées à l'une des stations de traitement. Ces tâches apparaissent dans le moniteur du serveur de traitement avec le statut « Pending » (en attente).



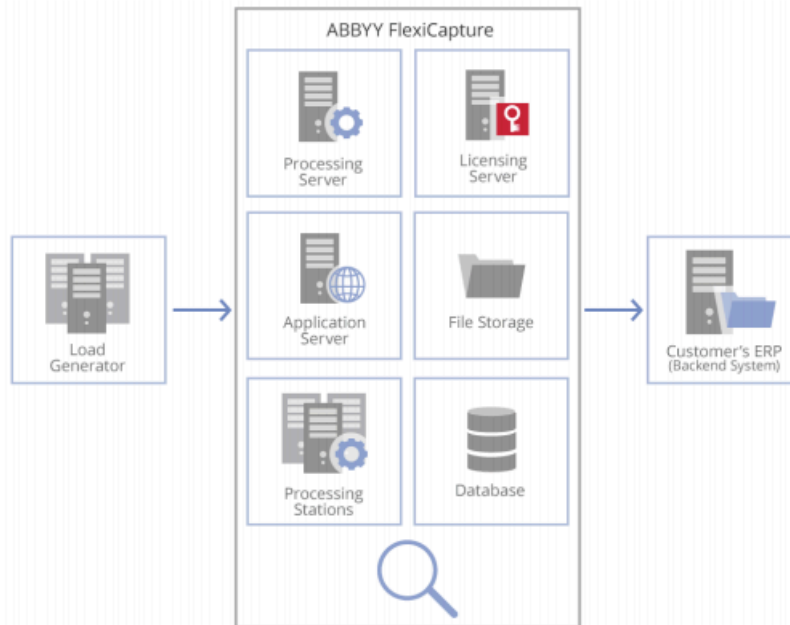
Note : l'ensemble des tâches en attente de traitement ne peut être consulté qu'à l'adresse Console de surveillance et d'administration.

	<p>Sa valeur ne doit pas dépasser le double du nombre de cœurs disponibles (Cœurs Compter). Si le nombre de tâches en attente est en constante augmentation, cela signifie certaines options de traitement sont désactivées sur certains postes, ou il y a les erreurs de communication entre la station et le serveur : ce dernier considère certaines stations doivent être en direct, mais ne sont pas en mesure de leur fournir une tâche.</p>
Serveur de licence	<p>Le serveur de licence est un service Windows® – le processus LicensingService.exe.</p> <p>Vous pouvez utiliser l'objet COM du serveur de licence pour surveiller l'état des licences, mais dans la plupart des cas, il n'y a pas de données utiles pour les performances du système.</p> <p>Gardez un œil sur la consommation de mémoire – elle peut augmenter avec un certain nombre de clients dans le système. Le processus LicensingService.exe est un processus 32 bits (64 bits dans le cas de l'application FlexiCapture 12) qui ne peut pas allouer plus de 2 Go de mémoire vive. Considérez l'ajout de serveurs de licence supplémentaires au système lorsque la consommation de mémoire est élevée et vous avez besoin de plus de clients.</p>
Serveur de base de données	<p>En plus des compteurs standard de surveillance du système (voir ci-dessus), vous pouvez également utiliser des compteurs qui affichent des données sur une base de données spécifique. Pour plus de détails, consultez la documentation du serveur de base de données.</p>
Stockage de fichiers	<p>Pour le disque dur en tant que FileStorage, utilisez les compteurs standard de surveillance du système (voir ci-dessus). Si vous utilisez un SAN ou un NAS, consultez la documentation de votre matériel pour plus de détails.</p>
Stations de traitement	<p>Utilisez les compteurs standard du système pour surveiller les postes de traitement.</p>

IX - Tests de performance

IX-A - Dispositif de tests minimal

En envoyant progressivement des lots d'images identiques au serveur d'application, les stations de numérisation automatisées contribuent à générer une charge de travail spécifique.



Les images d'entrée sont traitées automatiquement par le système :

Importation à partir d'une station de balayage -> Prétraitement -> Reconnaissance -> Exportation
-> Étape de traitement.

Le SingleEntryPoint livré avec ABBYY FlexiCapture est utilisé en standard pour les tests de performance.

Il contient cinq définitions de documents : quatre flexibles (Lettres, Contrats, Prix, Factures) et une fixe (Banque), qui comprend trois sections et des pages annexes.

Les images d'entrée arrivent pour être traitées sous forme d'ensembles de pages séparés. Le lot utilisé dans le test en tant que norme contient 69 pages (27 documents). En fonction des objectifs du test, le mode couleur et le nombre d'images du lot peuvent varier. Au stade de la reconnaissance, elles sont assemblées en documents après l'application des définitions des documents.

Les documents traités sont reçus dans le dossier partagé SMB :

- les données extraites sont exportées sous forme de fichiers CSV ;
- les images enregistrées sous forme de fichiers TIFF couleur.

Les anciens fichiers et données sont automatiquement nettoyés par une procédure standard sur le serveur d'application : les lots datant de plus d'une heure, ainsi que les journaux et les statistiques datant de plus de 24 heures sont supprimés au cours du nettoyage, qui commence toutes les heures, si le système dispose de ressources suffisantes.

La configuration du serveur et le nombre de cœurs de processeurs de traitement varient en fonction des objectifs du test.

Le système est constamment surveillé pendant les tests. En plus des compteurs, (voir les sections relatives à la Surveillance du système et au Système de détection des goulots d'étranglement), ces paramètres sont également contrôlés :

- 1 Le débit d'entrée en pages/seconde et en pages/24 heures ;
- 2 Les performances du système en pages/seconde et en pages/24 heures ;
- 3 Le temps moyen de traitement d'un lot ;
- 4 La performance du nettoyage automatisé en pages/24 heures.

Chaque installation de test devrait fonctionner pendant au moins 8 heures à charge de travail constante. Les périodes de test typiques sont les suivantes :

- Tests express : 24 heures ;
- Tests standard : 72 heures ;
- Tests de stabilité : 1 à 2 semaines.

Avec l'étape de vérification, le déroulement des opérations est le suivant :

```
Importation à partir d'une station de balayage -> Prétraitement -> Reconnaissance -> Vérification  
-> Exportation -> Étape de traitement.
```

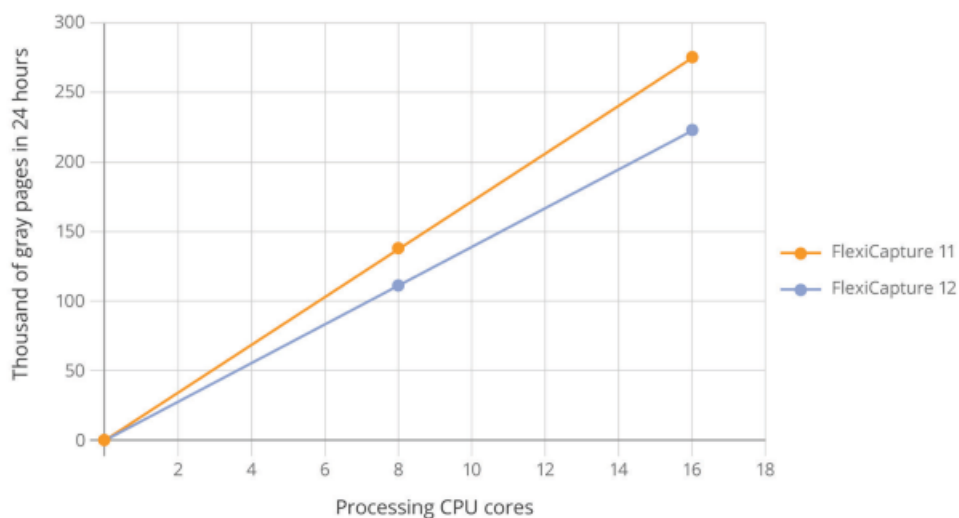
Au stade de la vérification, les clients automatisés imitent le travail des opérateurs de vérification en recevant les tâches, en enregistrant les documents modifiés et en transmettant les tâches traitées à l'étape suivante. Dans ce cas, les paramètres suivants sont également contrôlés :

- 1 Le nombre de vérificateurs ;
- 2 Le temps nécessaire pour obtenir une tâche et le temps nécessaire pour enregistrer les résultats de la vérification en quelques secondes – indicateurs clés de la façon dont il est facile pour les opérateurs de vérification d'utiliser le système.

IX-B - Test de FlexiCapture 12 contre FlexiCapture 11 au banc d'essai ABBYY (Small)

Nos tests ont révélé une différence de performance et d'évolutivité entre FlexiCapture 12 et FlexiCapture 11 lors du traitement des images en niveaux de gris.

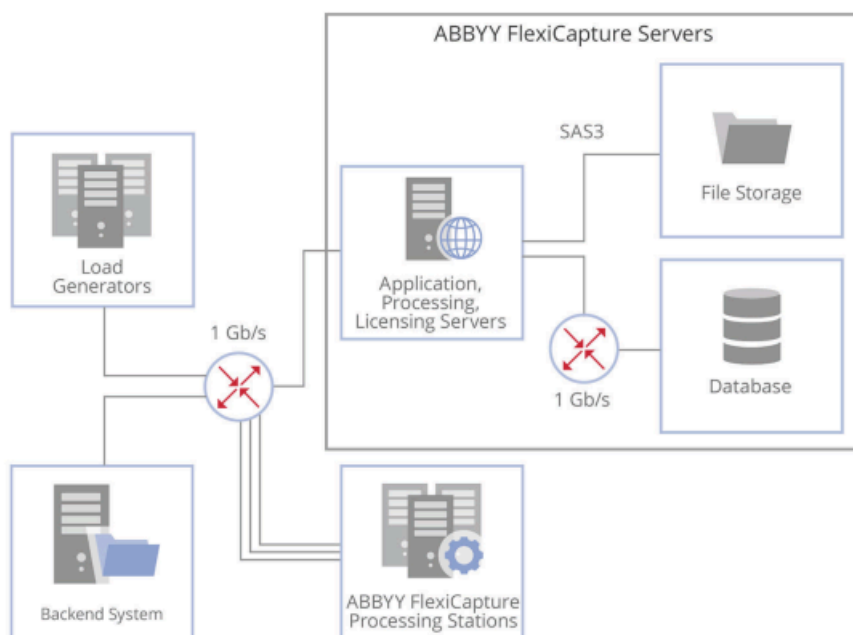
FlexiCapture 12 offre des performances supérieures de 20 % à celles de FlexiCapture 11.



IX-B-1 - Spécifications des installations de test

Les tests ont été réalisés dans l'environnement d'entreprise d'ABBYY en utilisant la méthodologie de test décrite ci-dessus.

Nous avons utilisé un lot standard de 69 pages en niveaux de gris provenant du projet SingleEntryPoint.

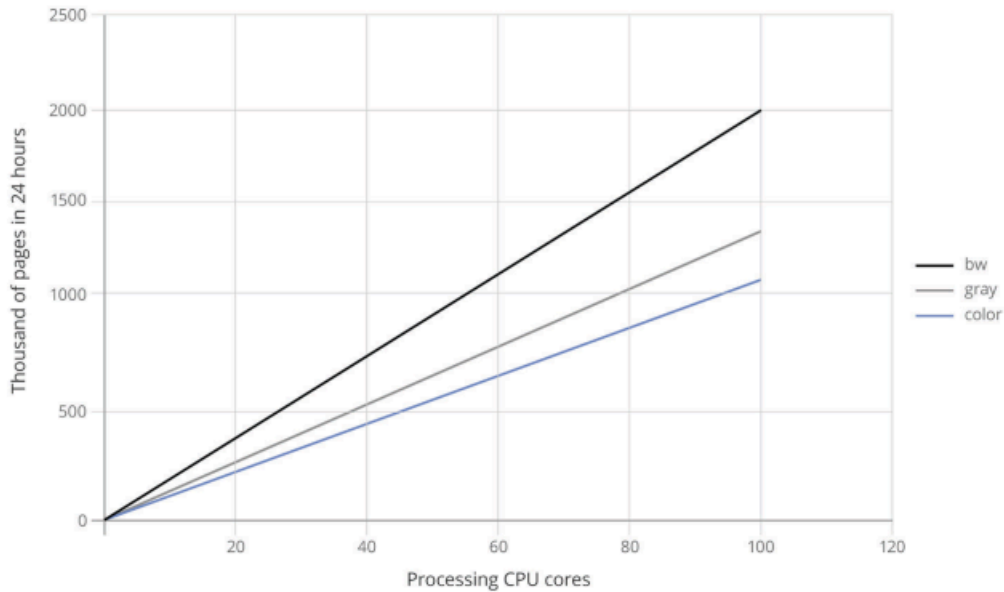


Rôle de la machine	Conditions requises
Serveurs ABBYY FlexiCapture : <ul style="list-style-type: none"> • Serveur d'application • Serveur de traitement • Serveur de licence 	Intel Xeon® E5-2680 v4 2,4 GHz 8 cœurs de processeurs virtuels, 32 GO DE RAM 2 IAS, 1 Gb/sec chacune

Serveur de base de données	Windows 2016 Microsoft SQL Server Developer 2016, ou Oracle 12c Enterprise Edition Intel Xeon® E5-2680 v4 2,4 GHz 8 cœurs de processeurs virtuels, 128 GO DE RAM 2 NIC, 1 Gb/sec chacun
Stockage de fichiers	Windows 2016 Disques SSD fonctionnant sur SAS3 avec : - Lecture continue à 3000 Mo/sec, écriture à 2660 Mo/sec - Accès aléatoire à un seul flux avec un bloc de 4 Ko : lecture à 19,4 Mo/sec, écriture à 26,1 Mo/sec. Les données acquises à l'aide de CrystalDiskMark 6.0.0
Station de traitement type et générateur de charge de travail	Intel Xeon® E5-2680 v4 2,4 GHz 8 cœurs virtuels Cœurs de CPU, 32 GO DE RAM 1 Gb/sec IAS
Backend	Dossier partagé du réseau SMB où les résultats du traitement sont exportés Intel Core i5-2400, 3,1 GHz, 4 cœurs physiques de l'unité centrale, Pas d'hyper-filtrage 4 GO DE RAM 1 Gb/sec NIC Windows 2008 R2 SP1

IX-C - Résultats des tests au banc ABBYY (Moyen)

Les tests ont révélé la dépendance des performances du système au nombre de cœurs de traitement, les images de traitement étant scannées dans différents modes de couleur.



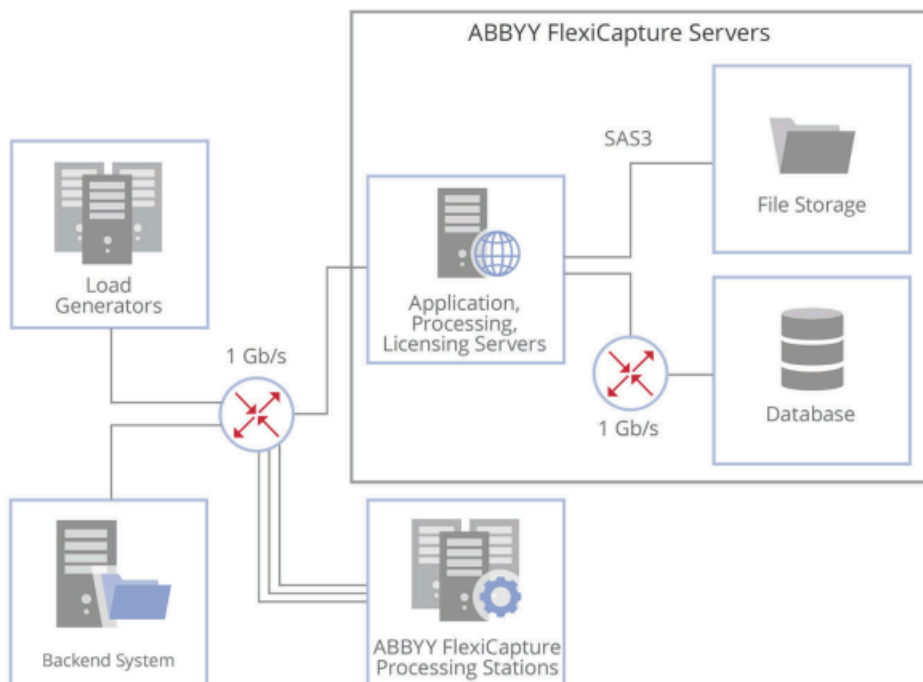
IX-C-1 - Spécifications des installations de test

Les serveurs FlexiCapture et le serveur de base de données ont été installés sur des machines virtuelles.

Tous les serveurs ABBYY FlexiCapture sont installés sur le même ordinateur, auquel ils sont connectés :

- le serveur de base de données via une interface réseau distincte avec une bande passante de 1 Go/seconde ;
- le FileStorage via une interface SCSI ;
- au réseau local externe via une interface réseau séparée avec une bande passante de 1 Gb/seconde.

Tous les générateurs de charge de travail, les stations de traitement et le système dorsal se trouvent dans le réseau local.



Rôle de la machine	Conditions requises
<p>Serveurs ABBYY FlexiCapture :</p> <ul style="list-style-type: none"> • Serveur d'application • Serveur de traitement • Serveur de licence <p>Serveur de base de données</p>	<p>Intel Xeon® E5-2680 v4 2,4 GHz 8 cœurs de processeurs virtuels, 32 GO DE RAM 2 IAS, 1 Gb/sec chacune Windows 2016</p> <p>Microsoft SQL Server Developer 2016 , ou Oracle 12c Enterprise Edition Intel Xeon® E5-2680 v4 2,4 GHz 8 cœurs de processeurs virtuels, 128 GO DE RAM 2 NIC, 1 Gb/sec chacun Windows 2016</p>
FileStorage	<p>Disques SSD fonctionnant sur SAS3 avec :</p> <ul style="list-style-type: none"> - Lecture continue à 3000 Mo/sec, écriture à 2660 Mo/sec - Accès aléatoire à un seul flux avec un bloc de 4 Ko : lecture à 19,4 Mo/sec, écriture à 26,1 Mo/sec. <p>Les données acquises à l'aide de CrystalDiskMark 6.0.0</p>
Station de traitement typique et générateur de charge de travail	<p>Intel Xeon® E5-2680 v4 2,4 GHz 10 cœurs virtuels Cœurs de CPU, 32 GO DE RAM 1 Gb/sec IAS et Intel Core i7-2600, 3,4 GHz 4 cœurs physiques de l'unité centrale, 8 GO DE RAM 1 Gb/sec IAS</p>

Backend

 Windows 2016
 Dossier partagé du réseau SMB où les résultats du traitement sont exportés
 Intel Core i5-2400, 3,1 GHz,
 4 cœurs physiques de l'unité centrale,
 Pas d'hyper-filtrage
 4 GO DE RAM
 1 Gb/sec NIC
 Windows 2008 R2 SP1

IX-D - Test des stations de traitement avec différents types d'unités centrales au banc ABBYY (Petit)

Ce test a comparé les performances des stations de traitement utilisant différents types d'unités centrales et a estimé le nombre de cœurs de processeurs nécessaires pour atteindre les charges souhaitées.

Il convient de noter que la performance a été mesurée en utilisant une seule station de traitement, ce qui signifie que le résultat final n'a pas été affecté de manière significative par la charge du réseau, du stockage des fichiers, de la base de données et du serveur FlexiCapture. Si l'on ajoute d'autres postes de traitement, l'image changera, montrant une augmentation non linéaire des performances.

Milliers de pages en niveaux de gris traitées en 24 heures					
Nombre de cœurs de CPU	Intel Xeon E5-2680 v4 2,4 GHz	Intel Xeon E5-2660 V2 2,6 GHz	Intel Xeon E5-2697A v4 2,6 GHz	Intel Xeon E5520 2,27 GHz	Intel Xeon E5-2640 v4 2,4 GHz
4	75	83.5	90	50	95
8	142	124	150	120	170
12	205	140	211	n/a	n/a

C'est pourquoi les données du tableau ci-dessus ne doivent pas être utilisées directement pour estimer le nombre de cœurs de processeurs nécessaires pour obtenir des performances spécifiques.

Veuillez tenir compte des recommandations suivantes.

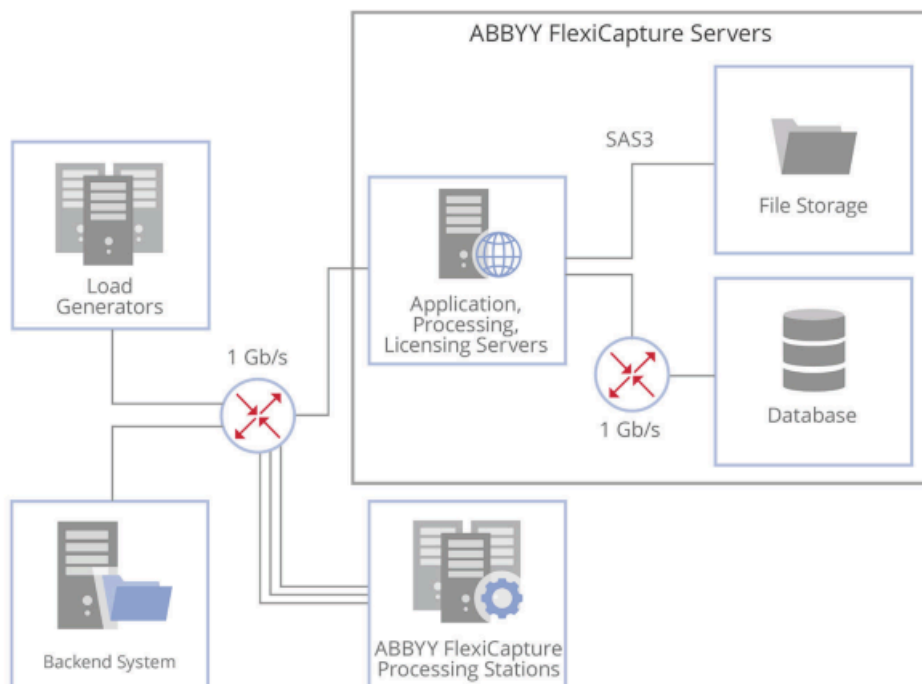
- Si vos estimations montrent qu'il faut plus de 50 cœurs d'unité centrale, testez le système entièrement assemblé pour vous assurer qu'il n'y a pas de goulots d'étranglement.
- Si votre estimation du nombre de cœurs d'unité centrale se situe entre 50 et 100, augmentez ce nombre de 25 %.
- Si vos estimations montrent qu'il faut plus de 100 cœurs de processeurs, un test de charge est obligatoire.

Notez également que les performances de vos stations de traitement seront affectées par les performances de votre sous-système de disque.

IX-D-1 - Spécifications des installations de test

Les tests ont été réalisés dans l'environnement d'entreprise d'ABBYY en utilisant la méthodologie de test décrite ci-dessus.

Nous avons utilisé un lot standard de 69 pages en niveaux de gris provenant du projet SingleEntryPoint.

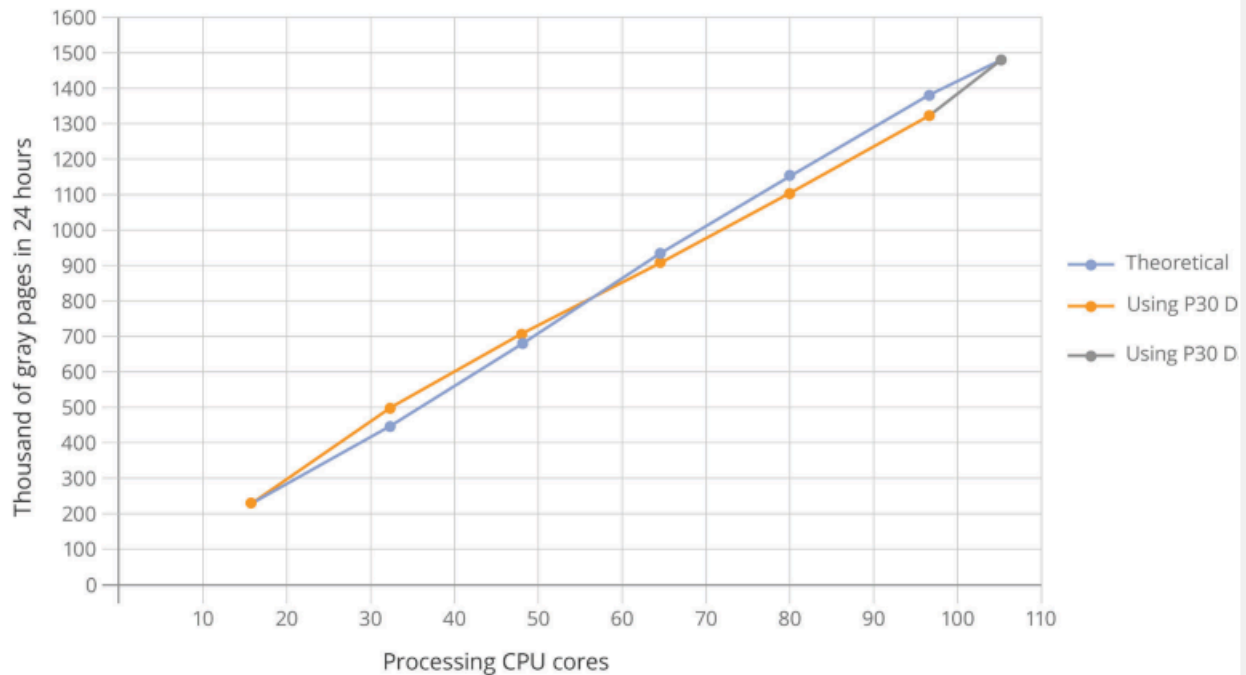


Rôle de la Machine	Conditions requises
Serveurs ABBYY FlexiCapture : <ul style="list-style-type: none"> • Serveur d'application • Serveur de traitement • Serveur de licence 	Intel Xeon® E5-2680 v4 2.4 GHz 8 virtual CPU cores, 32 GB RAM 2 IASs, 1 Gb/sec each Windows 2016
Serveur de base de données	Microsoft SQL Server Developer 2016, or Oracle 12c Enterprise Edition Intel Xeon® E5-2680 v4 2.4 GHz 8 virtual CPU cores, 128 GB RAM 2 IASs, 1 Gb/sec each Windows 2016
FileStorage	Disques SSD fonctionnant sur SAS3 avec : <ul style="list-style-type: none"> - Lecture continue à 3000 Mo/sec, écriture à 2660 Mo/sec - Accès aléatoire à un seul flux avec un bloc de 4 Ko : lecture à 19,4 Mo/sec, avec une vitesse de 26,1 Mo/sec. Les données acquises à l'aide de CrystalDiskMark 6.0.0
Backend	Dossier partagé du réseau SMB où les résultats du traitement sont exportés Intel Core i5-2400, 3,1 GHz, 4 cœurs physiques de l'unité centrale, Pas d'hyper-filtrage 4 GO DE RAM 1 Gb/sec NIC Windows 2008 R2 SP1 Dossier partagé du réseau SMB où les résultats du traitement sont exportés Intel Core i5-2400, 3,1 GHz, 4 cœurs physiques de l'unité centrale,

Pas d'hyper-filtrage
4 GO DE RAM
1 Gb/sec NIC
Windows 2008 R2 SP1

IX-E - Résultats des tests pour Microsoft Azure (Large)

Ce test a permis d'évaluer les performances et l'évolutivité du système sur la puissance de calcul louée à Microsoft Azure.



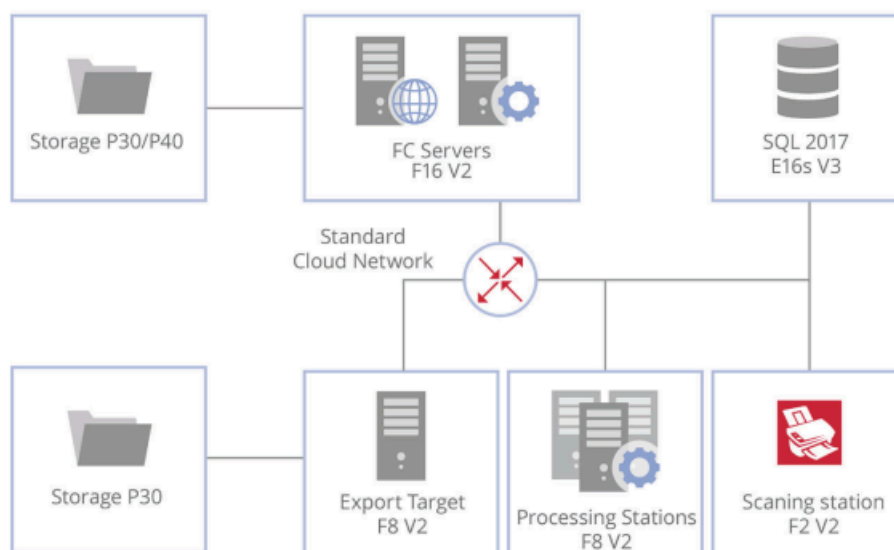
IX-E-1 - Spécifications des installations de test

Les serveurs FlexiCapture et le serveur de base de données ont été installés sur des machines virtuelles louées à Microsoft Azure.

La configuration était celle d'une installation typique sur site.

À un moment du test, les performances ont atteint leur maximum en raison des limitations du sous-système de disques.

Le disque peu performant a été remplacé par le disque le plus rapide disponible, mais cela n'a pas entraîné d'amélioration significative des performances, car le potentiel du disque a été rapidement épuisé.

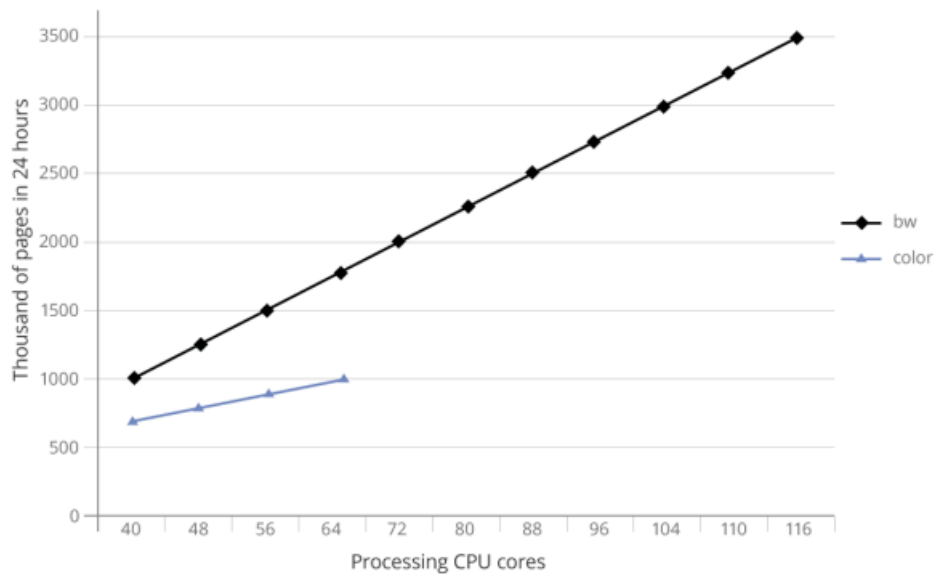


Rôle de la machine	Conditions requises
Serveurs ABBYY FlexiCapture : <ul style="list-style-type: none"> • Serveur d'application • Serveur de traitement • Serveur de licence 	Instance d'Azure : F16 v2 Intel Xeon® Platinum 8168 (SkyLake), 2,7 GHz 16 cœurs de processeurs virtuels 32.00 GiB Windows 2016
Serveur de base de données	Microsoft SQL Server Developer 2017 Instance d'azur : E16 v2 Intel XEON® E5-2673 v4 (Broadwell), 2,3 GHz 16 cœurs de processeurs virtuels 128.00 GiB Windows 2016
FileStorage	Disque géré SSD Premium, attaché à la machine virtuelle Instance d'azur : P30 Débit par disque : 200 Mo/seconde Instance d'azur : P40 Débit par disque : 250 Mo/seconde
Station de traitement	Instance d'azur : F8 v2 Intel Xeon® Platinum 8168 (SkyLake), 2,7 GHz 8 cœurs de processeurs virtuels 16.00 GiB Windows 2016
Générateur de charge de travail	Instance d'azur : F2 v2 Intel Xeon® Platinum 8168 (SkyLake), 2,7 GHz 2 cœurs d'unité centrale virtuels 4.00 GiB Windows 2016
Backend	Dossier partagé du réseau SMB où les résultats du traitement sont exportés Instance d'azur : F8 v2 Intel Xeon® Platinum 8168 (SkyLake), 2,7 GHz 8 cœurs de processeurs virtuels

16.00 GiB
Disque géré SSD Premium, attaché à la machine virtuelle
Instance d'azur : P30
Débit par disque : 200 Mo/seconde
Windows 2016

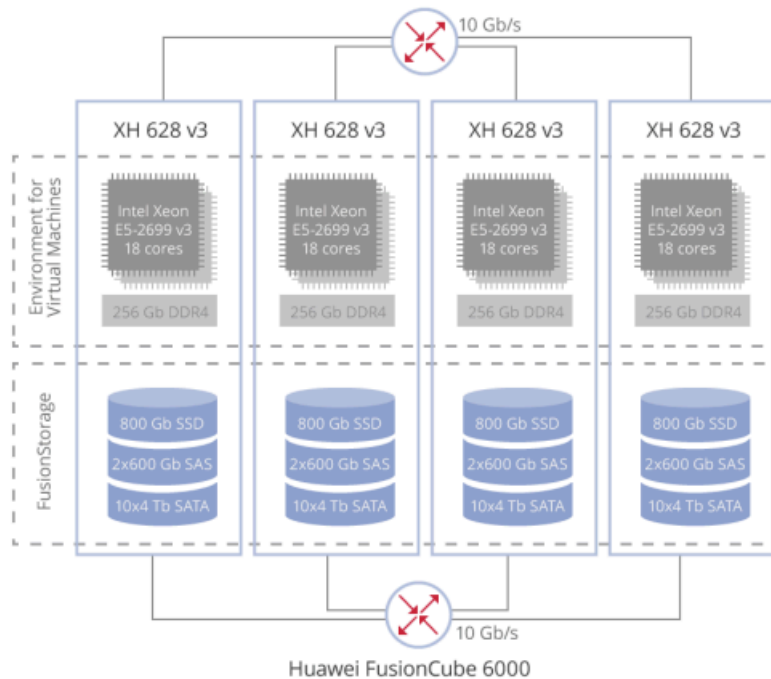
IX-F - Résultats des tests pour le Huawei FusionCube 6000 (Grand)

Lors des tests, la performance maximale du système à une charge de travail maximale a été mesurée, lorsque les images sont traitées dans différents modes de couleur, y compris en noir et blanc.



IX-F-1 - Spécification de l'installation de test

ABBYY FlexiCapture et le serveur de base de données sont déployés sur un serveur Huawei FusionCube 6000. Le Huawei FusionCube 6000 contient deux cartes réseau de 10 Gb/s et 4 nœuds XH628 v3.



Chaque nœud de la configuration supérieure est constitué de :

- 2 x processeurs Intel Xeon E5-2699 v3, 18 cœurs (36 threads) chacun ;
- 256 GB DDR4 RAM ;
- Des matrices de disques durs de 800 Go de cache de cartes PCIe SSD, 2 x 600 Go de SAS et 10 x 4 To de disques SATA.

Ces nœuds hébergent des machines virtuelles qui sont les composants de FlexiCapture et remplissent d'autres rôles selon les besoins des tests :

- 1 machine virtuelle avec le rôle de serveur d'application ;
- 1 machine virtuelle pour le serveur de base de données ;
- 1 machine virtuelle pour les serveurs de traitement et de licence ;
- 10 machines virtuelles pour la station de traitement ;
- 8 machines virtuelles pour les générateurs de charge afin d'émuler les entrées des utilisateurs ;
- 1 machine virtuelle pour l'émulation du système dorsal, où tous les résultats de traitement doivent être exportés.

Une carte réseau est utilisée pour créer un VLAN qui assure la communication entre toutes les machines virtuelles à l'intérieur du FusionCube. Une autre carte réseau est utilisée pour assurer la connexion entre chaque machine virtuelle et FusionStorage qui combine toutes les matrices de disques durs en deux stockages séparés : un stockage est utilisé à titre privé pour FlexiCapture FileStorage (500/600 Mo/s en lecture/écriture), tandis que l'autre héberge les disques durs de tous des machines virtuelles (900/700 Mo/s en lecture/écriture).

Ainsi, chaque machine virtuelle dispose de deux cartes réseau de 2 Gb/s : l'une à connecter au VLAN et l'autre au système FileStorage.

Chaque machine virtuelle possède un certain nombre de cœurs de processeur virtuels qui sont en fait représentés par des threads sur Intel Xeon Processeurs E5-2699 v3.

Rôle de la machine	Conditions requises
ABBY FlexiCaptureServers :	1 machine virtuelle sur 12 cœurs de processeurs Huawei FusionCube 6000: virtuels Intel Xeon E5-2699 v3
• Serveur d'application	24 GO DE RAM Disque dur de 150 Go chez FusionStorage (900/700 Mo/s en lecture/écriture) NIC 2Gb/s pour se connecter au VLAN NIC 2Gb/s pour se connecter à FusionStorage Windows© 2012 R2
ABBY FlexiCaptureServers :	1 machine virtuelle sur 4 cœurs de processeurs Huawei FusionCube 6000: logiques Intel Xeon E5-2699 v3
• Serveur de traitement	8 GO DE RAM
• Serveur de licence	Disque dur de 100 GB chez FusionStorage (900/700 MB/s en lecture/écriture) NIC 2Gb/s pour se connecter au VLAN NIC 2Gb/s pour se connecter à FusionStorage Windows© 2012 R2
Serveur de base de données	de Développeur MS SQL 12 logical CPU cores Intel Server© 2012 Xeon E5-2699 v3 SP1 sur une machine virtuelle sur Huawei FusionCube 6000: 24 GB RAM 400 GB hard drive at FusionStorage (900/700 MB/s read/write) 2Gb/s NIC to connect to VLAN 2Gb/s NIC to connect to FusionStorage Windows© 2012 R2
FileStorage	Réseau de disques 5 TB FusionStorage à l'intérieur du Huawei FusionCube 6000: Vitesse de lecture de 500 Mo/s Vitesse d'écriture de 600 Mo/s Les données acquises à l'aide de CrystalDiskMark 2.2.
Station de traitement	10 machines virtuelles sur 12 cœurs de processeurs Huawei FusionCube 6000 : logiques Intel Xeon E5-2699 v3 36 GO DE RAM Disque dur de 100 Go chez FusionStorage (900/700 MB/s en lecture/écriture) NIC 2Gb/s pour se connecter au VLAN NIC 2Gb/s pour se connecter à FusionStorage

		Windows Server © 2012 R2
Générateur de charge de travail typique	de 8 machines virtuelles sur 12 cœurs de processeurs Huawei FusionCube 6000 :	logiques Intel Xeon E5-2699 v3 36 GO DE RAM Disque dur de 100 Go chez FusionStorage (900/700 MB/s en lecture/écriture) NIC 2Gb/s pour se connecter au VLAN NIC 2Gb/s pour se connecter à FusionStorage Windows Server© 2012 R2
Backend	1 machine virtuelle sur 4 logical CPU cores Huawei FusionCube 6000:	Xeon E5-2699 v3 4 GB RAM 1 TB hard drive at FusionStorage (900/700 MB/s read/write) 2Gb/s NIC to connect to VLAN 2Gb/s NIC to connect to FusionStorage Windows Server© 2012 R2

X - Remerciements Developpez.com

Developpez.com remercie ABBYY pour l'autorisation de publication de ce tutoriel. Tous les remerciements aussi à **Guillaume SIGUI** pour la mise au gabarit et **Claude Leloup** pour la relecture orthographique.