



Haute disponibilité

Livre blanc

Par PaperCut Software – Juin 2018



Table des matières

Introduction	3
Les avantages de la haute disponibilité	4
Facteurs organisationnels	4
Valeur commerciale	4
Comportement de l'utilisateur	5
Infrastructure du système	5
Comment atteindre la haute disponibilité ?	5
Redondance	5
Reprise en cas d'incident	6
Quel niveau de haute disponibilité ?	6
Perte de données maximale admissible (PDMA ou en anglais Recovery Point Objective (RPO))	7
Durée d'interruption maximale admissible (DIMA ou en anglais Recovery Time Objective (RTO))	7
PaperCut et la haute disponibilité	8
Exemples d'utilisation appliqués à PaperCut	10
Architecture évolutive de PaperCut	10
Serveur d'applications PaperCut	11
Serveur de site PaperCut	11
Fournisseurs de services d'impression et serveurs d'impression	11
Base de données externe	12
Utilisation d'un NLB (équilibre de charge réseau)	12
Dans quel cas utiliser un NLB ?	12
Quels sont les avantages d'un NLB pour l'impression ?	13
L'équilibrage de charge fonctionne-t-il avec l'impression Find Me ?	15
Clustering de VM	18
Quel est le rôle des VM ?	18
Définition de l'environnement de clustering de la VM	19
Configuration du clustering	19
Utilisation du Microsoft Failover Cluster Manager	20
Qu'est-ce que le MFCM (Microsoft Failover Cluster Manager) ?	20
Comment configurer le basculement de cluster ?	20
Résilience du serveur de site	21
Considérations sur la résilience et la redondance	21

Introduction

Ce livre blanc présente la démarche pour mettre en œuvre la haute disponibilité au sein d'une organisation et explique notamment comment l'appliquer aux composants PaperCut et à l'infrastructure d'impression qui les accompagne. Notre guide s'adresse aux décideurs IT et à toute personne devant évaluer la technologie HA (High Availability ou Haute disponibilité en français) dans le cadre de l'implémentation de PaperCut au sein d'une organisation.

Le terme **High Availability (HA) ou Haute disponibilité** en français est généralement utilisé pour désigner le fait qu'un service informatique fonctionne sans interruption notable. La haute disponibilité procure une augmentation significative de la disponibilité bout à bout d'un système. Cependant, les opérations de déploiement et de maintenance des dispositifs recourant à cette technologie sont souvent plus compliquées. Il convient donc de se poser la question suivante : quels sont les systèmes concernés par une protection HA et à quel niveau?

Le concept de Haute disponibilité ne s'applique pas uniquement aux services informatiques. Prenons l'exemple des automobiles. Supposons que vous souhaitiez que vos voitures soient dotées de capacités à haute disponibilité. Vous allez faire en sorte de protéger les capacités propices aux pannes et défaillances, par exemple les pneus. Pour poursuivre avec cet exemple, vous savez tous que la plupart des voitures transportent une roue de secours pour parer aux risques de crevaison. Toutefois, compte-tenu du coût généré par cette mesure de protection, la plupart des voitures n'ont qu'une seule roue de secours. Dans ce cas comme en informatique, la haute disponibilité implique toujours un surcoût.

Ce serait formidable si les ordinateurs ne subissaient aucune défaillance ou si leurs disques durs ne tombaient jamais en panne. Mais ce n'est pas la réalité. Et même si c'était vrai, il y aurait des périodes d'indisponibilité dues à des facteurs externes (opérations de maintenance, erreurs humaines, attaques informatiques, problèmes de connectivité, catastrophes naturelles, etc.).



C'est pourquoi il est important d'identifier les systèmes à protéger et le niveau de protection. Pour reprendre l'exemple de la voiture, pourquoi transporter six roues de secours alors qu'une seule suffira ? La démarche est la même en informatique.

Les avantages de la haute disponibilité

La haute disponibilité préserve vos ressources informatiques des risques de pannes. À la clé, des opérations exécutées sans interruption ou sans baisse de rendement. La réduction du risque d'indisponibilité a pour principal avantage d'assurer ou de renforcer la continuité des opérations.

En éliminant les points de défaillance, la haute disponibilité contribue à réduire le risque d'indisponibilité. Si l'intégralité de votre système PaperCut est exécutée sur un serveur, ce dernier constitue un point de défaillance ainsi que ses composants (disque dur, alimentation, et carte réseau). Même des éléments logiciels tels que les mises à jour des pilotes ou les correctifs système sont susceptibles d'affecter le système s'ils comportent des défauts.

En conclusion, la haute disponibilité offre des méthodologies pour éliminer les points de défaillance, réduire les risques d'indisponibilité et assurer la continuité des opérations. Ses avantages sont indéniables!

Facteurs organisationnels

Les logiciels de gestion d'impression procurent les fonctionnalités de traçabilité et de sécurisation décrites dans ce livre blanc. Ainsi, la fonctionnalité de libération d'impression sécurisée (Secure print release) est essentielle pour tout système de gestion d'impression. PaperCut NG et PaperCut MF prennent en charge les dispositifs de sécurité présentés dans ce document.

Valeur commerciale

Premièrement, il est important d'identifier les opérations vitales. Si vous dirigez un hôpital, la panne du distributeur alimentaire n'aura pas un impact vital sur l'activité de l'hôpital. Toutefois, si c'est le distributeur alimentaire d'une université qui tombe en panne au moment des examens, la gêne occasionnée sera plus importante.

Comportement de l'utilisateur

Le comportement de l'utilisateur est un autre facteur à prendre en compte. Existe-t-il des périodes de pointe qui sollicitent fortement certains composants du système informatique ? Certaines applications ont-elle un effet important sur la charge système ? Quels utilisateurs doivent bénéficier d'un accès système, même si un incident est à craindre ?

Infrastructure du système

Différents types de dispositifs doivent-ils être protégés ? Tous les éléments du système ont-ils été identifiés ? Que se passe-t-il en cas de panne de courant dans le datacenter principal ?

Les techniques de protection des services informatiques sont généralement documentées. Quelle que soit l'origine d'une panne (un serveur, le système d'exploitation, les bases de données, ou l'électricité), il est nécessaire de comprendre les impératifs commerciaux et d'identifier les points de défaillance pouvant impacter le fonctionnement de l'entreprise.

Comment atteindre la haute disponibilité ?

La haute disponibilité peut être atteinte par la mise en oeuvre de la **redondance** et de la **reprise sur incident**. Toutes deux procurent la continuité de service nécessaire au système (c'est-à-dire la capacité à reprendre un fonctionnement normal).

Redondance

La redondance résout la possibilité d'une panne éventuelle par l'utilisation d'un deuxième dispositif. Elle fait appel à des technologies telles que le RAID, les [images de machine virtuelle \(VM\)](#), les clusters et [l'équilibrage de la charge réseau \(Network Load Balancers ou NLB\)](#).

Si vous utilisez la technologie RAID et qu'un disque dur tombe en panne, vous pouvez être tranquille. Grâce à la redondance, les données sont écrites sur d'autres disques. Si vous utilisez des machines virtuelles et qu'elles tombent en panne, pas de souci non plus. Vous pouvez simplement déplacer les dernières images virtuelles sur un nouveau serveur VM et reprendre le cours de vos activités.



Les clusters et les NLB sollicitent plusieurs serveurs ce qui fait que les requêtes peuvent être renvoyées d'un serveur en panne vers un serveur en fonctionnement. En théorie, le risque d'interruption est très faible si vous optez pour la redondance.

Toutefois, le coût d'un système redondant est souvent le double d'un système normal. De plus, les opérations de déploiement et de maintenance sont souvent complexes et nécessitent du personnel qualifié.

Pour déterminer le type et le niveau de redondance, il est nécessaire de prendre en compte certains critères. Un client de PaperCut utilisait un système Linux HA très sophistiqué avec des serveurs DNS en cluster et des serveurs de bases de données en cluster. Une protection à toute épreuve ! Toutefois, lorsque l'administrateur système a quitté l'entreprise, personne ne savait comment effectuer la maintenance du système. Les mises à jour système étaient laborieuses et chronophages. Pour finir, ce système trop complexe a été abandonné pour être remplacé par des serveurs VM plus flexibles. Le client disposait d'un niveau de redondance équivalent sans être contraint par la complexité du système antérieur.

Reprise en cas d'incident

La haute disponibilité peut aussi être atteinte grâce à un bon programme de reprise sur incident. Cette forme élémentaire de haute disponibilité ne requiert pas de techniques HA complexes. Un bon programme de reprise sur incident fait appel à des procédures minimisant l'interruption des systèmes critiques. Citons par exemple une procédure qui consisterait à sauvegarder une base de données chaque nuit et qui écrirait les transactions journalières sur un serveur hors site. Il serait ainsi possible de restaurer l'intégralité de la base de données en peu de temps, même si le serveur principal de la base de données tombait en panne et brûlait.

Quel niveau de haute disponibilité ?

L'un de nos clients avait mis en place une installation robuste incluant des serveurs d'applications en cluster, chacun pointant sur un SAN. D'un point de vue système, c'était un niveau élevé de haute disponibilité. Un jour, un incendie a détruit le datacenter, et il n'était plus question de haute disponibilité ni de disponibilité tout court. Le client a dû repenser toute son infrastructure et a porté son choix sur PaperCut. Il a ensuite opté pour des technologies de VM afin de bénéficier du même niveau de haute disponibilité.



Cela nous amène directement à la question du bon niveau de haute disponibilité. Le coût de propriété d'un temps de disponibilité de 99 % n'est pas comparable à celui d'un temps de disponibilité de 99,99 %. La différence justifie-t-elle un tel écart de coût ? Même si la haute disponibilité a un impact important sur l'activité de votre entreprise, et que vous savez comment l'atteindre, la question qui se pose porte sur le niveau de haute disponibilité. Selon les fonctions de vos interlocuteurs, vous obtiendrez des réponses différentes. Si le coût d'une panne de votre infrastructure est supérieur au coût de la haute disponibilité, il est évident que vos systèmes critiques doivent être protégés par la haute disponibilité.

Des systèmes critiques tels que systèmes de gestion de commandes, d'authentification utilisateur ou de bases de données doivent offrir un taux de disponibilité de 99,99 %. Diverses techniques telles que les images de VM, le clustering, la réplication synchrone, la reprise à chaud, les sauvegardes hors site permettent de relever le défi. Cependant, une interruption d'une heure est tout à fait acceptable pour un système d'impression.

En matière de haute disponibilité, il est important de prendre en compte les deux objectifs suivants :

Perte de données maximale admissible (PDMA ou en anglais Recovery Point Objective (RPO))

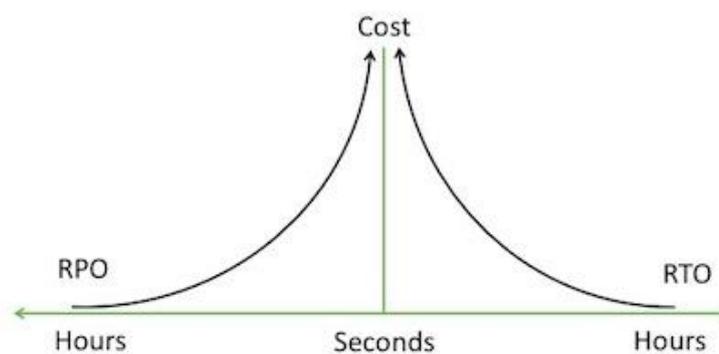
Le RPO est la durée écoulée entre les snapshots de données. Elle détermine jusqu'où vous devez aller pour obtenir un point de restauration. Elle correspond aussi à la quantité maximale acceptable de données perdues.

Durée d'interruption maximale admissible (DIMA ou en anglais Recovery Time Objective (RTO))

Le RTO (le temps moyen de restauration) est le délai maximal toléré pour revenir à un fonctionnement normal après un incident.



Les indicateurs RPO et RTO doivent être étudiés attentivement car ils déterminent le coût de la restauration. Plus les temps définis pour les objectifs RPO et RTO sont courts, plus le coût de restauration est élevé. Si les points de restauration sont mesurés en secondes et que le temps de restauration s'effectue en minutes, vous devez vous attendre à un coût de restauration très élevé.



PaperCut et la haute disponibilité

Appliquons maintenant ces principes au déploiement de PaperCut. Pour commencer, nous n'allons pas considérer que les systèmes d'impression doivent avoir le même niveau de protection que d'autres fonctions commerciales. Commençons par les objectifs commerciaux. Si vous perdez les tâches d'impression et leurs données comptables puis devez tout réimprimer, quelle est la quantité de données perdues acceptable (RPO) ? Quel est le délai acceptable entre la défaillance des systèmes



d'impression et le rétablissement complet (RTO) ? Et troisièmement, quel est le coût total de restauration estimé ?

Avec PaperCut, il est possible d'améliorer la continuité de service du système sur plusieurs points de l'infrastructure. Nous recommandons d'utiliser des technologies de haute disponibilité et des méthodes connues des utilisateurs. De plus, votre entreprise aura sans doute du personnel expert dans l'utilisation de ces technologies. Ainsi, dans l'éventualité d'une panne, vos équipes n'ont pas à se familiariser avec de nouveaux outils et procédures. C'est aussi la raison pour laquelle PaperCut n'impose pas la haute disponibilité ni ne crée ses propres produits de haute disponibilité.

Vous vous souvenez du client Linux haute disponibilité en cluster mentionné précédemment ? Il s'avère que leur point unique de défaillance (SPOF) était l'administrateur ayant mis en place l'environnement complexe avec plusieurs technologies de haute disponibilité. Lui seul était capable de comprendre le système qu'il avait configuré. PaperCut ne souhaite pas obliger ses clients à utiliser un outil supplémentaire qui viendrait s'ajouter à la longue liste des technologies de haute disponibilité.

Les technologies HA ont évolué et permettent maintenant d'atteindre des niveaux RPO et RTO favorables sans que cela ait un impact sur les coûts et la complexité des produits PaperCut. Le client n'a pas à apprendre comment nous atteignons la haute disponibilité, ses connaissances suffisent pour utiliser nos produits. Par ailleurs, il convient de rappeler que la source la plus fréquente d'interruption est externe : disques durs pleins, défaillance des cartes réseau, erreur humaine. L'intégralité du système d'impression, PaperCut y compris, peut être protégée contre les interruptions avec les technologies « clés en mains » de haute disponibilité.

Avant d'envisager de protéger PaperCut des interruptions, il est opportun de se poser les questions suivantes : « Comment le client peut-il se protéger des interruptions intervenant sur d'autres serveurs, et en particulier sur les systèmes d'impression ? Par exemple, il ne serait pas très utile de protéger le serveur PaperCut sans inclure les autres serveurs d'impression. Tout composant de l'infrastructure sur lequel PaperCut est exécuté doit être inclus dans le plan HA. En d'autres termes, il est essentiel d'inclure dans le plan HA le serveur d'applications PaperCut, [le serveur de site](#) PaperCut, le fournisseur d'impression PaperCut (PaperCut Print Providers, les serveurs d'impression), la base de données et les périphériques multifonctions (MFD). Si un client a déjà recours à des méthodes de HA pour protéger ses serveurs ou d'autres ressources,

il pourra inclure PaperCut à moindre frais et sans que cela nécessite des compétences supplémentaires.

Quelles que soient les méthodes utilisées pour viser la haute disponibilité avec PaperCut, « la règle générale est de commencer avec une infrastructure légère et de l'adapter progressivement » (Chris Dance, CEO de PaperCut).

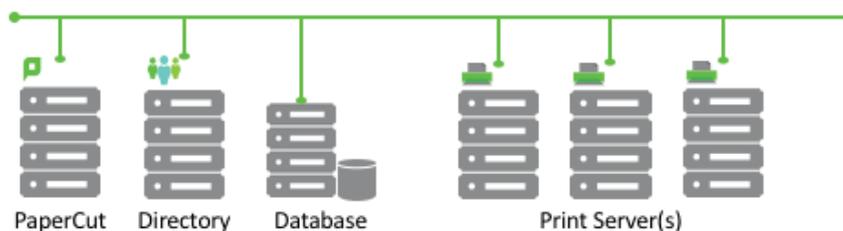
Exemples d'utilisation appliqués à PaperCut

En matière de haute disponibilité, diverses technologies sont disponibles, et chacune dispose de particularités qui impacteront votre choix. Des contraintes techniques peuvent aussi faire que certaines technologies ne seront pas adaptées (par exemple utiliser un NLB pour des bases de données SQL). Il convient de respecter les principes fondamentaux suivants :

- Utiliser des outils maîtrisés par vos équipes
- Fournir un niveau de haute disponibilité correspondant aux objectifs commerciaux de l'entreprise
- Commencer avec une infrastructure légère et l'adapter progressivement

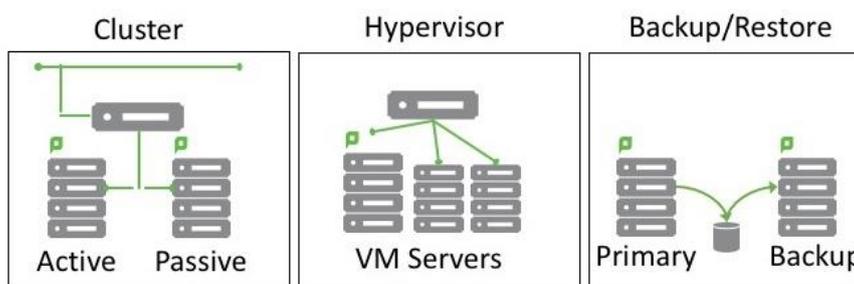
Architecture évolutive de PaperCut

PaperCut a été conçu pour supporter les montées en charge. Cela signifie que des composants critiques comme le serveur d'applications ou la base de données peuvent être situés sur des serveurs différents pour permettre des améliorations de performances. Cela a aussi pour avantage de protéger ces composants des interruptions. Le diagramme ci-dessous représente un déploiement typique avec un serveur d'applications PaperCut, un annuaire d'utilisateurs, un serveur de base de données, et plusieurs serveurs d'impression. Dans cet exemple, il n'y a pas beaucoup d'intérêt à protéger le serveur d'applications PaperCut car les données importantes se trouvent dans la base de données.



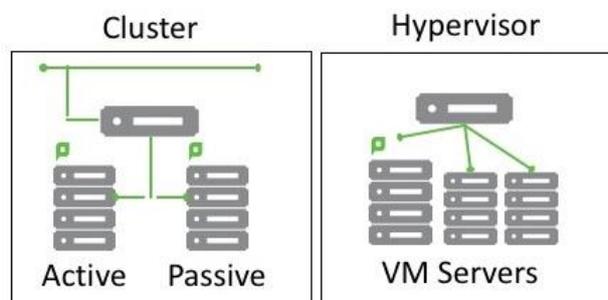
Serveur d'applications PaperCut

Le **serveur d'applications PaperCut (app server)** peut être protégé via le clustering, les VM ou une simple procédure de sauvegarde et restauration.



Serveur de site PaperCut

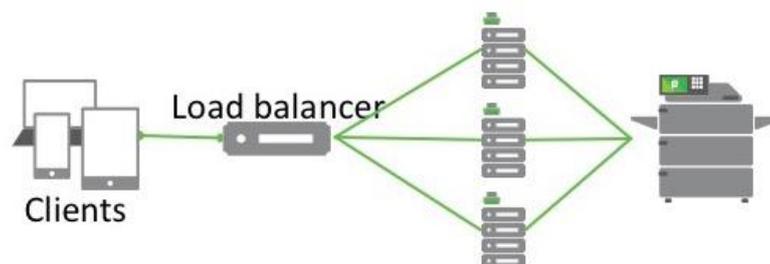
Le **serveur de site PaperCut (site server)** assure la continuité du système d'impression dans l'éventualité où le serveur d'applications deviendrait indisponible. Le serveur de site doit aussi être protégé par clustering ou en utilisant une VM.



Fournisseurs de services d'impression et serveurs d'impression

PaperCut Print Provider, le fournisseur de services d'impression, s'exécute sur des serveurs d'impression et peut utiliser toutes les

méthodes présentées ci-dessus ainsi que le **NLB (Network Load Balancing** ou **Équilibrage de la charge réseau**). Un NLB répartit le trafic des applications (les tâches d'impression dans notre cas) sur l'ensemble des serveurs. En plus d'effectuer l'équilibrage des tâches d'impression, un NLB peut contourner un serveur qui ne répond pas. Ce comportement procure une protection supplémentaire en cas de défaillance du serveur d'impression.



Base de données externe

Une **base de données externe** peut être protégée par clustering ou par l'utilisation de VM ainsi que par des techniques propres aux bases de données (réplication synchrone et journaux de transactions). Vérifiez avec votre client et l'éditeur de la base de données quel est le choix adapté à votre situation.

Utilisation d'un NLB (équilibrage de charge réseau)

Dans quel cas utiliser un NLB ?

Comme indiqué précédemment, un NLB répartit le trafic réseau et notamment les requêtes de service entre plusieurs serveurs capables de répondre aux demandes des clients. L'avantage étant que les clients envoient les requêtes au NLB et non aux serveurs individuels. Cette technique est largement répandue sur Internet.

Ainsi sur les sites, la façon la plus rapide de répondre aux augmentations de charge est d'utiliser un NLB. Si google.com était trop occupé, nous devrions tous naviguer sur google2.com et si les deux étaient trop occupés, il faudrait basculer sur google42.com. Ce qui serait particulièrement contraignant.



L'utilisation d'un NLB nous permet de tous naviguer sur un même site Internet (par exemple google.com). En effet, le NLB décide quel serveur est disponible pour répondre aux requêtes. Comme son nom l'indique, il équilibre et répartit la charge réseau entre les différents serveurs.

Le NLB s'avère particulièrement utile dans certaines situations. Outre sa capacité à répartir la charge réseau, il facilite les opérations de maintenance des serveurs individuels. Si votre système d'équilibrage de charge inclut 16 serveurs, et que vous devez changer un disque dur sur l'un d'eux, vous pouvez retirer le serveur du NLB, remplacer le disque, puis replacer le serveur dans le NLB.

La protection contre les pannes individuelles des serveurs compte parmi les autres fonctionnalités utiles des NLB. Si l'on reprend le scénario ci-dessus, que se passerait-il si ce disque tombait en panne avant que ne l'ayez changé ? Vous disposeriez toujours des 15 autres serveurs pour répartir la charge, ce qui éviterait une catastrophe.

Toutefois, le NLB peut aussi être à l'origine de problèmes. Les sessions Internet sont conçues pour être « stateless » (sans maintien de l'état). Cela signifie que le serveur ne fait pas le suivi de l'état de la session client. Ainsi, si vous naviguez sur google.com, vous risquez d'être renvoyé sur google7.com pour vous identifier, puis d'être renvoyé sur google166.com pour lancer votre recherche, puis sur google99.com lorsque vous cliquez sur un lien. Ces sessions Internet « sans maintien de l'état » fonctionnent bien avec un NLB.

En revanche, les connexions « stateful » (avec maintien de l'état) conservent des données importantes sur la session dans la mémoire du serveur. Si vous utilisez un NLB avec des serveurs « stateful », il y a de fortes chances pour qu'il les confonde. Les connexions des serveurs avec maintien de l'état sont plus efficaces que les connexions sans maintien de l'état car elles nécessitent moins d'échanges de données entre le client et le serveur.

Ainsi, le serveur d'applications de PaperCut est un serveur « stateful ». Il fait le suivi des clients connectés pour faciliter les communications. En conséquence, le serveur d'applications peut être protégé par clustering, VM ou processus de sauvegarde et réplication, ce qui n'est pas le cas du NLB.

Quels sont les avantages d'un NLB pour l'impression ?

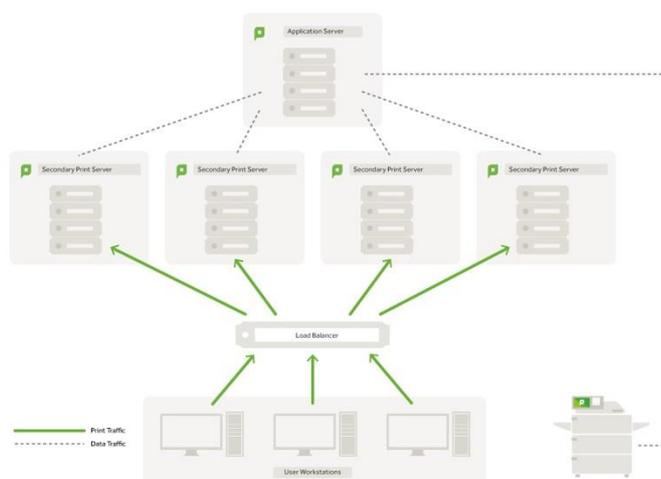
De la même façon qu'une ferme de serveurs web permet de répartir de nombreuses requêtes entre plusieurs serveurs web, un NLB permet de répartir les tâches d'impression. Il est en effet possible de configurer

un dispositif d'équilibrage de charge (comme F5, NetScaler) qui accepte les requêtes d'impression entrantes puis les répartit entre plusieurs serveurs d'impression.

Les NLB sont particulièrement utiles dans des environnements d'impression intensifs (par exemple un nombre important d'utilisateurs imprime constamment ou de grosses tâches d'impression sont régulièrement envoyées en impression). L'équilibrage de charge permet aux administrateurs système de mieux administrer l'environnement d'impression en adoptant une approche plus souple. PaperCut peut être utilisé conjointement avec ces dispositifs. Différentes méthodes permettent de combiner les différentes technologies.

Vous le savez peut-être déjà, PaperCut peut être installé en tant que serveur principal (aussi appelé serveur primaire), serveur secondaire et serveur de site. Le serveur principal correspond à l'application principale de PaperCut (la partie logicielle est gérée sur ce serveur).

Voici la configuration élémentaire d'une répartition de charge :



Les **serveurs d'impression secondaires de PaperCut** sont installés derrière le NLB.

Chaque serveur d'impression secondaire PaperCut est connecté au serveur d'applications et sera configuré avec toutes les files d'impression disponibles.

Par exemple, le réseau comporte cinq imprimantes (Printer1, Printer2, Printer3, Printer4, Printer5), et chaque serveur d'impression secondaire PaperCut doit avoir une file d'impression pour chacune des cinq imprimantes.

Le NLB peut être configuré de façon à accepter le trafic d'impression depuis les machines client puis à le rediriger vers un serveur d'impression secondaire PaperCut.

Dans la plupart des cas, vous configurerez un nom DNS qui fait référence à l'adresse IP du NLB.

Les clients utiliseront ce nom DNS pour référencer les imprimantes. Par exemple, si vous avez utilisé le nom DNS 'NLB-Uno' pour nommer le NLB, les clients se connecteront à NLB-Uno\Printer1 et à NLB-Uno\Printer2.

Lorsqu'un utilisateur envoie son impression sur l'une de ces files d'impression, le NLB redirige la tâche d'impression vers l'un des serveurs d'impression secondaires PaperCut. Cela signifie que la tâche d'impression peut s'effectuer sur l'un des quatre serveurs secondaires PaperCut (selon la façon dont le NLB a été configuré).

Le serveur d'applications voit chacun des serveurs secondaires PaperCut et les files d'impression associées. Il peut ensuite libérer chaque tâche sur le serveur d'impression approprié.

Si la [libération d'impression sécurisée](#) est utilisée, la tâche d'impression est redirigée par le serveur d'applications sur la bonne imprimante lors de l'authentification de l'utilisateur sur celle-ci. Sinon, la tâche d'impression n'est pas conservée sur le serveur et est envoyée directement sur l'imprimante.

L'équilibrage de charge fonctionne-t-il avec l'impression Find Me ?

Bien sûr que oui. Si vous utilisez l'impression [Find Me](#), vous devez configurer une file d'impression supplémentaire sur les serveurs d'impression secondaires PaperCut. Cette file sera votre file d'impression Find Me virtuelle. Assurez-vous de répéter l'opération pour chacun des serveurs secondaires PaperCut.

La file d'impression de l'utilisateur final serait NLB\Find Me. Le NLB peut être configuré pour rediriger la file d'impression Find Me entrante vers l'un des quatre serveurs d'impression.

Depuis le serveur d'applications, quatre files d'impression Find Me sont visibles et PaperCut les signale individuellement. Chaque file étant une instance séparée, notre solution permet de configurer des préférences de serveur pour les files d'impression. Cela garantit que le serveur qui reçoit la tâche d'impression est le même serveur que celui qui envoie la tâche d'impression à l'imprimante.

Les préférences de serveur permettent de réduire le trafic réseau. En effet, elles permettent d'éviter que les tâches d'impression soient reçues sur un serveur (par exemple Server1), puis transférées sur un autre serveur (par exemple Server4). Les résultats d'une configuration avec ou sans préférences sont visibles dans l'illustration ci-dessous.

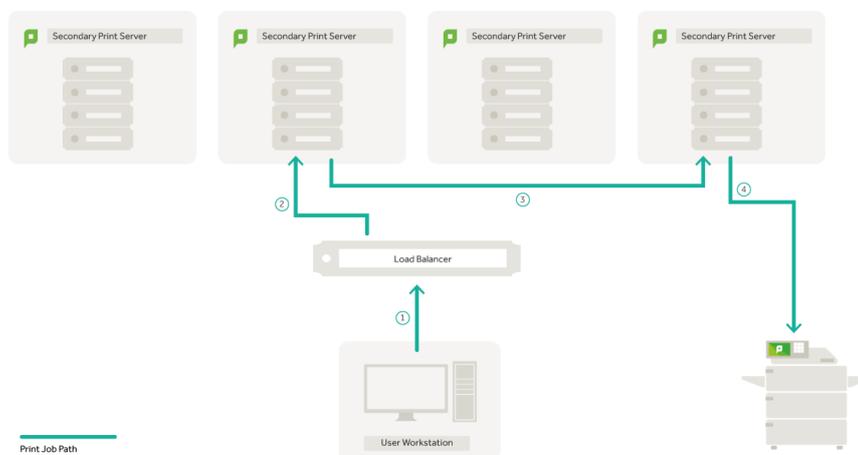


Figure 1 Sans préférence de serveur

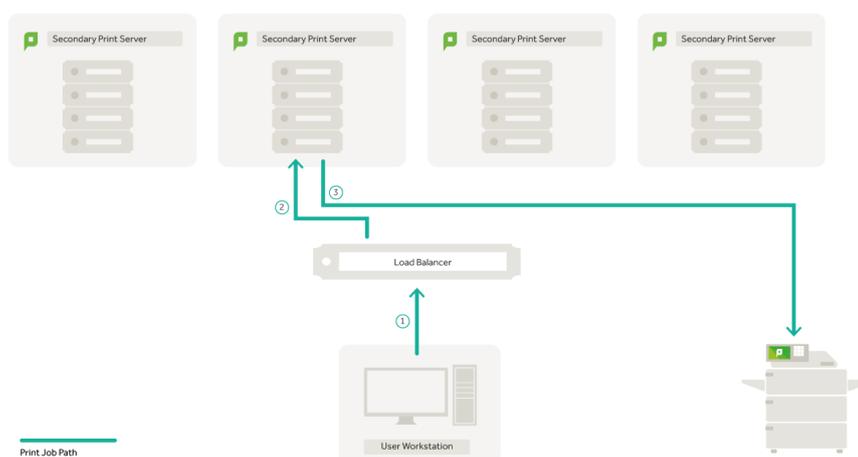


Figure 2 Avec des préférences de serveur

Pour configurer les préférences de serveur, il est nécessaire d'accomplir les étapes suivantes dans l'interface Admin de PaperCut :

- Connectez-vous à l'interface Admin de PaperCut.
- Accédez à 'Imprimantes'.
- Choisissez l'une des files d'impression Find Me (par ex. Server1\Find Me).

- Configurez-la comme une file d'impression virtuelle.
- Dans le champ relatif aux paramètres de redirection de tâche, choisissez les imprimantes sur lesquelles vous voulez libérer les tâches d'impression (par exemple la file d'impression Find Me).

Dans notre exemple, nous aurions choisi server1\printer1, server1\printer2, server1\printer3, server1\printer4 et server1\printer5.

- Enregistrez les paramètres.
- Désormais, les tâches d'impression envoyées dans la file Find Me sur Server1 peuvent uniquement être libérées dans une file d'impression sur ce serveur.
- Répétez ces étapes pour les files d'impression Find Me restantes, tout en vous assurant que vous sélectionnez uniquement les files sortantes pour chaque serveur approprié.

Désormais, PaperCut libérera uniquement les tâches d'impression sur le serveur qui les reçoit. Nous devons maintenant configurer chaque périphérique pour vérifier les quatre files d'impression Find Me :

- Accédez à 'Périphériques'.
- Choisissez un périphérique sur lequel vous voulez utiliser l'impression Find Me.
- Défilez jusqu'à 'Libération des impressions' et vérifiez que l'option est bien sélectionnée.
- Dans la liste des files d'impression, sélectionnez les quatre files d'impression Find Me.
- Enregistrez ces paramètres et répétez l'opération pour les périphériques restants.

Désormais, chaque périphérique vérifiera les files d'impression Find Me sur les quatre serveurs.

Du point de vue de l'utilisateur, une seule file d'impression est visible au moment de l'impression. Lorsque l'utilisateur libère des tâches d'impression, une seule file d'impression apparaît. Toutefois, les tâches d'impression peuvent aussi bien se trouver sur un seul serveur qu'être réparties sur plusieurs serveurs.

L'intérêt du dispositif est que vous disposez d'une alternative en cas de défaillance de l'un des serveurs d'impression. Cette solution facile à administrer supporte également très bien la montée en charge.

REMARQUE : L'équilibrage de charge ne s'applique pas au serveur d'applications PaperCut. Les tâches d'impression peuvent uniquement

être réparties entre les différents serveurs d'impression secondaires PaperCut.

Clustering de VM

Quel est le rôle des VM ?

Les technologies de machines virtuelles (VM ou Virtual Machine) telles que VMWare et Microsoft Hyper-V procurent une grande souplesse dans le déploiement des serveurs au sein d'une organisation. Les implémentations de VM sont aussi un moyen de fournir de la haute disponibilité par l'intermédiaire du clustering de VM. Lorsqu'une machine virtuelle est exécutée dans un cluster de VM à haute disponibilité, toute défaillance matérielle a un impact minimal sur la VM exécutée car elle est simplement transférée sur un autre nœud du cluster.

L'implémentation de la haute disponibilité par une infrastructure de VM est une bonne alternative aux pratiques classiques. Vous pouvez configurer PaperCut comme sur un serveur physique tout en permettant à l'infrastructure de VM de disposer de haute disponibilité.

Le clustering de VM offre de nombreux avantages par rapport aux configurations traditionnelles :

- L'installation et la configuration des logiciels, pilotes et paramètres s'effectuent en une seule fois sur une image de VM unique ;
- Selon votre infrastructure de VM, en cas de défaillance d'un nœud physique, la VM peut basculer sur un autre nœud avec un temps d'interruption minimal ;
- Des processus de sauvegarde simplifiés ;
- La prise en charge de la reprise sur incident (DR).

Les hyperviseurs de VM détectent si VM ne répond plus. Cette capacité peut être renforcée par un monitoring de niveau applicatif. Même si la VM semble fonctionner normalement, les applications sous-jacentes peuvent rencontrer des problèmes. Dans ce cas, un monitoring de niveau applicatif peut détecter et signaler le problème. Le monitoring de niveau applicatif peut s'effectuer de différentes façons :

- Chargement de l'URL du serveur d'applications pour tester le bon fonctionnement du serveur
- Pings de l'adresse IP
- Vérification de l'exécution des services PaperCut

Définition de l'environnement de clustering de la VM

Diverses stratégies de déploiement de VM sont disponibles selon la plateforme de VM utilisée. Les VM peuvent être hébergées sur différents emplacements physiques ou sur différents sites. Diverses options de reprise sur incident sont aussi disponibles.

Lors du choix d'un système de VM, il est particulièrement important de prendre en compte les critères suivants :

- Tolérance de pannes (FT) – la continuité de service des VM est assurée en créant et en maintenant une VM secondaire identique à la VM principale.
- Haute disponibilité – permet de créer une nouvelle VM pour réduire les temps d'interruption dans l'éventualité où une VM existante serait défaillante. Généralement la tolérance aux pannes s'ajoute à la haute disponibilité, car elle permet de basculer sur une autre VM s'il n'y a pas de perte de données.
- Haute disponibilité du niveau applicatif – permet de mettre à disposition une nouvelle VM pour réduire les temps d'interruption en cas de défaillance du niveau applicatif. Il est préférable d'avoir un système applicatif disposant de capacités de monitoring de niveau applicatif.
- Sauvegarde et réplication des données – cette capacité permet de sauvegarder et de restaurer les données en mémoire et les données applicatives. Dans le cas où des fonctionnalités de tolérance aux fautes ou de reprise sur incident sont utilisées, la réplication des données en mémoire joue un rôle capital. Les algorithmes propriétaires relatifs aux segments mémoire réduisent généralement la bande passante utilisée et sont très efficaces.

Configuration du clustering

Le déploiement de PaperCut sur une infrastructure VM peut s'effectuer via différentes méthodes. Prenez en compte les considérations suivantes pour implémenter une installation de PaperCut sur un cluster de VM.

Mode 1 : Configuration du clustering au niveau de la couche d'impression

Configurez les serveurs d'impression conformément aux exigences de votre environnement VM sans oublier de les paramétrer pour la haute disponibilité. Installez le serveur secondaire de PaperCut pour permettre le monitoring d'impression sur les serveurs d'impression.

Mode 2 : Configuration du clustering au niveau de la couche serveur d'applications

Configurez un nouveau serveur dans votre infrastructure VM pour héberger le serveur d'applications. Configurez cette VM pour la haute disponibilité. Installez le serveur d'applications PaperCut. Vous pouvez ensuite choisir de configurer vos serveurs d'impression pour la haute disponibilité.

Utilisation du Microsoft Failover Cluster Manager

Qu'est-ce que le MFCM (Microsoft Failover Cluster Manager) ?

Depuis Windows Server 2012, Microsoft a changé la façon de mettre en œuvre la haute disponibilité pour vos services d'impression. Il est désormais possible d'utiliser Hyper-V et le clustering de basculement pour que votre serveur d'impression devienne une VM à très haute disponibilité.

Cette solution procure des options permettant d'effectuer un basculement du serveur complet et peut être implémentée avec les serveurs PaperCut.

Comment configurer le basculement de cluster ?

Pour consulter un article détaillé sur la configuration des serveurs PaperCut sur Microsoft Failover Cluster Manager, reportez-vous à la section [Microsoft Failover Cluster Manager](#) du manuel de PaperCut. Sinon, vous trouverez ci-dessous des informations importantes résumant la marche à suivre.

Exigences système :

- Deux serveurs physiques ou plus
- MS Server 2012 ou 2016 avec Hyper-V
- SAN iSCSI avec 2 disques configurés
- Disque 1 : 5 Go (à utiliser pour le Quorum si seulement 2 nœuds sont utilisés)
- Disque 2 : stockage du serveur VM

Exigences supplémentaires :

- Droits d'administration pour ajouter des machines sur le domaine
- Adresse IP du SAN
- Adresses IP des composants suivants :
 - Les serveurs physiques
 - Le cluster
 - La VM

Effectuez la configuration en suivant les étapes ci-dessous :

Etape 1 – les rôles et fonctionnalités disponibles doivent être configurés pour la haute disponibilité

Etape 2 – connexion aux lecteurs réseau iSCSI

Etape 3 – création du cluster de basculement

Etape 4 – création d'une VM à haute disponibilité

Etape 5 – configuration du serveur d'impression

Résilience du serveur de site

Considérations sur la résilience et la redondance

Les clients utilisant des déploiements distribués doivent prendre sérieusement en compte des considérations telles que la redondance et la résilience.

En cas de panne réseau, une solution robuste doit être en mesure de protéger des points de défaillance critiques afin d'assurer la continuité des opérations au sein d'une organisation. En ce qui concerne PaperCut, cela signifie que les services d'impression doivent être assurés même si les liens réseau ne sont pas fiables.

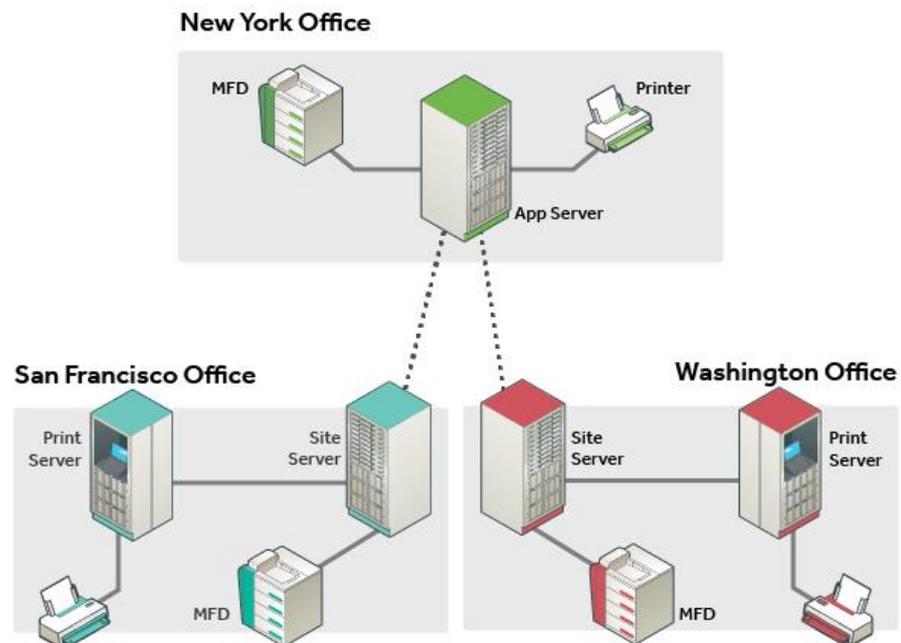
Grâce à l'installation des serveurs de site PaperCut, l'accès aux ressources d'impression n'est pas interrompu par les pannes réseau.

En cas de panne, le serveur de site duplique les fonctionnalités clés du serveur principal PaperCut sur un site local. Pour supprimer la dépendance au WAN, les MFD sont configurés de façon à se connecter au serveur de site comme s'il s'agissait du serveur principal. Les serveurs secondaires PaperCut (les fournisseurs de service d'impression) fournissent un serveur de basculement si le serveur principal ne peut pas être contacté.

Ce concept simple mais efficace a pour avantage de fournir de la haute disponibilité pour les MFD et prend en charge des fonctionnalités telles que la libération d'impression sécurisée et l'impression Find Me.

Le serveur de site s'installe en quelques minutes avec un minimum de configuration et ne requiert pas une administration permanente. Aucune expertise spécifique en gestion de base de données ou en réplication n'est nécessaire pour assurer la continuité des opérations. Le serveur de site garantit que ses informations sont à jour avec l'état en cours du serveur principal et accomplit son rôle de serveur principal en cas de panne. Lorsque la connexion au serveur principal peut être rétablie, la fusion des journaux et des transactions du serveur de site est effectuée vers le serveur principal.

La relation étroite entre le serveur de site et les serveurs principaux leur permet de prendre en charge les mêmes systèmes d'exploitation et bases de données. Ainsi, il est tout à fait possible qu'un serveur secondaire PaperCut existant soit promu en serveur de site PaperCut en vue d'améliorer la résilience du site.





Avec nos remerciements

Auteur

David O'Hara

Lead Solutions Architect, PaperCut Software

ACDI

sales@acd-inc.com

acd-inc.com

Support

support@acd-inc.com

acd-inc.com/support