Guide des notions fondamentales de Sun Cluster pour SE Solaris

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054
U.S.A.

Référence : 819–2060
Août 2005, Révision A

050812@12762

Table des matières

Préface

Le *guide des notions fondamentales de Sun™ Cluster pour Solaris* contient des notions conceptuelles et des références pour le système SunPlex™ sur des systèmes SPARC® et x86.

Remarque – dans ce document, le terme “x86” fait référence à la gamme de puces microprocesseurs de la gamme 32 bits d’Intel et aux puces microprocesseurs conçues par AMD.

Le système SunPlex comprend tous les composants matériels et logiciels de la solution cluster de Sun.

Ce document est destiné aux administrateurs système expérimentés, formés pour l’utilisation du logiciel Sun Cluster. Ne l’utilisez pas comme guide de planification ou de pré-vente. Vous devez déjà avoir déterminé vos besoins système et acheté l’équipement et les logiciels appropriés avant de lire ce document.

Pour comprendre les notions fondamentales décrites dans ce document, vous devez connaître le système d’exploitation Solaris™ et maîtriser le gestionnaire de volume utilisé avec le système Sun Cluster.

Remarque – le logiciel Sun Cluster fonctionne sur deux plates-formes, SPARC et x86. Les informations contenues dans ce document s’appliquent aux deux, sauf indication contraire dans un chapitre, une rubrique, une remarque, une liste à puces, une figure, un tableau ou un exemple spécifique.

Conventions typographiques

Le tableau suivant présente les modifications typographiques utilisées dans ce manuel.

TABLEAU P–1 Conventions typographiques

Type de caractère ou symbole	Signification	Exemple
`AaBbCc123`	Noms de commandes, fichiers, répertoires et messages système s'affichant à l'écran.	Modifiez votre fichier `.login`. Utilisez `ls -a` pour afficher la liste de tous les fichiers. `nom_machine% Vous avez reçu du courrier.`
`AaBbCc123`	Ce que vous entrez, par opposition à ce qui s'affiche à l'écran.	`nom_machine%` **`su`** `Mot de passe :`
AaBbCc123	Paramètre substituable de ligne de commande à remplacer par un nom ou une valeur	La commande permettant de supprimer un fichier est `rm>` *nom_fichier*.
AaBbCc123	Titres de manuels, termes nouveaux et mis en évidence.	Reportez-vous au chapitre 6 du *Guide de l'utilisateur*. Effectuez une *analyse de patchs*. *N'*enregistrez *pas* le fichier. [Notez que certains éléments mis en évidence s'affichent en gras sur le site.]

Invites de shell dans les exemples de commandes

Le tableau suivant présente les invites système et les invites de superutilisateur par défaut des shells
C, Bourne et Korn.

TABLEAU P–2 Invites du shell

Shell	Invite
Invite en C shell	`nom_machine%`
Invite du superutilisateur en C shell	`nom_machine#`
Invites en Bourne et Korn shells	`$`
Invite de superutilisateur en Bourne et Korn shells	`#`

Documentation connexe

Vous trouverez dans le tableau suivant les manuels contenant des informations sur des sujets connexes associés à Sun Cluster. Toute la documentation relative à Sun Cluster est disponible à l'adresse http://docs.sun.com.

Rubrique	Documentation
Présentation	*Présentation de Sun Cluster pour SE Solaris*
Concepts	*Guide des notions fondamentales de Sun Cluster pour SE Solaris*
Installation et administration matérielle	*Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS*
	Guides d'administration matérielle individuelle
Installation du logiciel	*Guide d'installation du logiciel Sun Cluster pour SE Solaris*
Installation et administration de services de données	*Sun Cluster Data Services Planning and Administration Guide for Solaris OS*
	Guides des services de données individuels
Développement de services de données	*Guide du développeur de services de données Sun Cluster pour SE Solaris*
Administration du système	*Guide d'administration système de Sun Cluster pour SE Solaris*
Messages d'erreur	*Sun Cluster Error Messages Guide for Solaris OS*
Références sur les commandes et les fonctions	*Sun Cluster Reference Manual for Solaris OS*

Pour obtenir une liste complète de la documentation Sun Cluster, reportez-vous aux notes de version relatives à votre version du logiciel Sun Cluster à l'adresse http://docs.sun.com/app/docs.

Accès à la documentation Sun en ligne

Le site Web docs.sun.com[SM] vous permet d'accéder à la documentation technique Sun en ligne. Vous pouvez le parcourir ou y rechercher un titre de manuel ou un sujet particulier. L'URL est `http://docs.sun.com`.

Commande de documents Sun

Sun Microsystems offre une sélection de documentation produit imprimée. Pour consulter la liste des documents disponibles et savoir comment passer une commande, reportez-vous à la rubrique "Acheter de la documentation papier" à l'adresse `http://docs.sun.com`.

Accès à l'aide

Si vous rencontrez des difficultés lors de l'installation ou de l'utilisation du système Sun Cluster, contactez votre fournisseur de services et donnez-lui les informations suivantes :

- votre nom et votre adresse de courrier électronique (le cas échéant) ;
- le nom, l'adresse et le numéro de téléphone de votre société ;
- les numéros de modèle et de série de vos systèmes ;
- le numéro de version de l'environnement d'exploitation (par exemple Solaris 9) ;
- le numéro de version du logiciel Sun Cluster (par exemple, 3.1 8/05).

Pour obtenir des informations sur chaque nœud de votre système, exécutez les commandes suivantes :

Commande	Fonction
`prtconf -v`	Indique la taille de la mémoire système et affiche des informations sur les périphériques.
`psrinfo -v`	Affiche des informations sur les processeurs.

Commande	Fonction
`showrev -p`	Indique les patchs installés.
SPARC : `prtdiag -v`	Affiche des informations diagnostiques sur le système.
`scinstall -pv`	Affiche des informations sur la version du package et du logiciel Sun Cluster.
`scstat`	Fournit un aperçu ponctuel de l'état du cluster.
`scconf -p`	Présente les informations de configuration du cluster.
`scrgadm -p`	Affiche des informations sur les ressources installées, les groupes de ressources et les types de ressources.

Gardez également à disposition le contenu du fichier `/var/adm/messages`.

Formation aux produits

Sun Microsystems propose des formations sur de nombreuses technologies Sun par le biais d'une gamme variée de cours dirigés par un formateur ou de cours auto-dirigés. Pour plus d'informations sur les cours de formation proposés par Sun et pour vous inscrire à un cours, consultez le site des Services de formation à l'adresse `http://training.sun.com/`.

CHAPITRE 1

Introduction et présentation

Le système Sun Cluster est une solution matérielle et logicielle Sun Cluster intégrée, destinée à la création de services évolutifs et hautement disponibles.

Le Guide des notions fondamentales de Sun Cluster pour SE Solaris présente les notions fondamentales nécessaires aux principaux utilisateurs de la documentation Sun Cluster. Ces utilisateurs comprennent :

- les fournisseurs de services assurant l'installation et la maintenance du matériel de cluster ;
- les administrateurs système chargés d'installer, de configurer et d'administrer le logiciel Sun Cluster ;
- les développeurs d'applications qui développent des services de basculement et évolutifs pour des applications qui ne sont pas actuellement incluses dans le produit Sun Cluster.

Lisez ce document et l'ensemble de la documentation Sun Cluster pour obtenir une vue complète du système Sun Cluster.

Ce chapitre

- fournit une présentation générale du système Sun Cluster ;
- décrit plusieurs points de vue d'utilisateurs de Sun Cluster ;
- identifie les principales notions fondamentales dont vous avez besoin pour utiliser le système Sun Cluster ;
- établit une correspondance entre les principales notions fondamentales, les procédures et les informations connexes qui figurent dans la documentation Sun Cluster ;
- Répertorie les tâches liées au cluster de la documentation contenant les procédures d'exécution de ces tâches.

Présentation du système Sun Cluster

Le système Sun Cluster étend le système d'exploitation Solaris en un système d'exploitation de cluster. Un cluster, ou plex, est un ensemble de noeuds en configuration dispersée fournissant au client une vue unique des services et applications réseau, notamment des bases de données, des services Web et des services de fichiers.

Chaque noeud du cluster est un serveur autonome fonctionnant avec ses propres processus. Ceux-ci communiquent entre eux pour former ce qui ressemble (pour un client réseau) à un système unique fournissant les applications, les ressources système et les données aux utilisateurs.

Un cluster présente plusieurs avantages par rapport aux systèmes à serveur unique traditionnels. Il prend notamment en charge les services évolutifs et de basculement, se prête à une évolution modulaire et son coût est relativement peu élevé par rapport aux systèmes de tolérance aux pannes traditionnels.

Les objectifs du système Sun Cluster sont les suivants :

- réduire ou éliminer les temps d'arrêt système liés aux pannes logicielles ou matérielles ;
- assurer la disponibilité des données et applications aux utilisateurs finaux, même en présence d'une panne entraînant normalement l'arrêt d'un système à serveur unique ;
- augmenter le rendement des applications en permettant aux services de se mettre en corrélation avec d'autres processeurs en ajoutant des noeuds au cluster ;
- accroître la disponibilité du système en vous permettant de réaliser des opérations de maintenance sans arrêter la totalité du cluster.

Pour plus d'informations sur la tolérance de pannes et la haute disponibilité, voir "Haute disponibilité des applications avec Sun Cluster" du *Présentation de Sun Cluster pour SE Solaris*.

Pour consulter les questions et les réponses sur la haute disponibilité, "FAQ sur la haute disponibilité " à la page 95.

Trois points de vue sur le système Sun Cluster

Cette section décrit trois points de vue différents sur le système Sun Cluster ainsi que les principales notions fondamentales et les documents appropriés. Ces points de vue nous sont donnés par des professionnels :

- du personnel chargé de l'installation et des prestations matérielles ;
- des administrateurs système ;
- des développeurs d'applications.

Point de vue du personnel chargé de l'installation et des prestations matérielles

Pour le personnel chargé des prestations matérielles, le système Sun Cluster ressemble à une collection de matériel en stock, notamment des serveurs, des réseaux et des périphériques de stockage. Ces composants sont tous reliés les uns aux autres à l'aide d'un câble, de sorte que chacun possède une sauvegarde et qu'il n'existe aucun point de panne unique.

Notions-clés : matériel

Le personnel chargé des prestations matérielles doit comprendre les notions fondamentales suivantes :

- configurations et câblage du matériel de cluster ;
- installation et maintenance (ajout, suppression, remplacement) :
  - des composants de l'interface réseau (adaptateurs, jonctions, câbles) ;
  - des cartes d'interface de disques ;
  - des tableaux de disques ;
  - Unités de disque
  - de la console administrative et du périphérique d'accès à la console ;
- paramétrage de la console administrative et du périphérique d'accès à la console.

Plus de notions fondamentales sur le matériel

Les rubriques suivantes contiennent le matériel correspondant aux notions-clés précédentes.

- "Noeuds de cluster " à la page 22

Documentation Sun Cluster destinée au personnel chargé des prestations matérielles

Le document Sun Cluster suivant contient les procédures et les informations associées aux notions fondamentales relatives aux prestations matérielles :

Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS

Point de vue des administrateurs système

Pour les administrateurs système, le système Sun Cluster est un ensemble de serveurs (nœuds) câblés qui partagent des périphériques de stockage et le logiciel est perçu comme :

- Un logiciel de cluster spécialisé intégré au logiciel Solaris pour contrôler la connectivité entre les noeuds de cluster.
- Un logiciel spécialisé contrôlant l’état des programmes d’application utilisateur exécutés sur les noeuds de cluster.
- Un logiciel de gestion des volumes définissant et administrant les disques.
- Un logiciel de cluster spécialisé permettant à tous les noeuds d’accéder à tous les dispositifs de stockage, même à ceux n’étant pas directement connectés aux disques.
- Un logiciel de cluster spécialisé, permettant aux fichiers d’apparaître sur chaque noeud comme s’ils étaient localement reliés à ce noeud.

Notions-clés : administration système

Les administrateurs système doivent maîtriser les notions et processus suivants :

- l’interaction entre les composants matériels et logiciels ;
- la procédure générale d’installation et de configuration du cluster, notamment :
  - l’installation du système d’exploitation Solaris ;

- installation et configuration du logiciel Sun Cluster ;
- l'installation et la configuration d'un gestionnaire de volume ;
- l'installation et la configuration des logiciels prêts à être clusterisés ;
- l'installation et la configuration du logiciel de services de données Sun Cluster ;

- les procédures d'administration de cluster pour l'ajout, la suppression, le remplacement et la maintenance des composants matériels et logiciels de cluster ;
- les modifications de configuration pour l'amélioration des performances.

Plus de notions fondamentales sur l'administration système

Les rubriques suivantes contiennent le matériel correspondant aux notions-clés précédentes.

- "Interfaces administratives " à la page 38
- "Heure du cluster " à la page 38
- "Structure à haute disponibilité " à la page 39
- "Périphériques globaux " à la page 42
- "Groupes de périphériques de disques " à la page 43
- "Espace de noms global " à la page 47
- "Systèmes de fichiers Cluster " à la page 48
- "Contrôle de chemin de disque " à la page 51
- "À propos de la séparation en cas d'échec " à la page 57
- "Services de données " à la page 65

Documentation Sun Cluster destinée aux administrateurs système

Les documents Sun Cluster suivants contiennent les procédures et informations associées aux notions fondamentales relatives à l'administration système :

- *Guide d'installation du logiciel Sun Cluster pour SE Solaris*
- *Guide d'administration système de Sun Cluster pour SE Solaris*
- *Sun Cluster Error Messages Guide for Solaris OS*
- *Notes de version de Sun Cluster 3.1 8/05 pour SE Solaris*
- *Sun Cluster 3.0-3.1 Release Notes Supplement*

Point de vue des développeurs d'applications

Le système Sun Cluster fournit des *services de données* pour des applications comme Oracle, NFS, DNS, Sun™ Java System Web Server, Apache Web Server (sur des systèmes SPARC) et Sun Java System Directory Server. Vous créez des services de données en configurant des applications prêtes à l'emploi de façon qu'elles s'exécutent

sous le contrôle du logiciel Sun Cluster. Le logiciel Sun Cluster fournit des fichiers de configuration et des méthodes de gestion qui démarrent, arrêtent et surveillent les applications. Si vous devez créer un nouveau service de basculement ou évolutif, l'API Sun Cluster et l'API DSET (Data Service Enabling Technologies) vous permettent de développer les fichiers de configuration et les méthodes de gestion nécessaires à l'exécution de son application en tant que service de données sur le cluster.

Notions-clés : développement d'applications

Les développeurs d'applications doivent comprendre les éléments décrits ci-dessous.

- Les caractéristiques de leur application afin de déterminer si elle doit s'exécuter en tant que service évolutif ou de basculement.
- L'API Sun Cluster, l'API DSET et le service de données « générique ». Les développeurs doivent identifier l'outil le mieux approprié pour développer des programmes ou écrire des scripts permettant de configurer leur application pour l'environnement du cluster.

Plus de notions fondamentales sur le développement d'applications

Les rubriques suivantes contiennent le matériel correspondant aux notions-clés précédentes.

- "Services de données " à la page 65
- "Ressources, groupes de ressources et types de ressource " à la page 75
- Chapitre 4

Documentation Sun Cluster destinée aux développeurs d'applications

Les documents Sun Cluster suivants contiennent les procédures et informations associées aux notions fondamentales relatives au développement d'applications :

- *Guide du développeur de services de données Sun Cluster pour SE Solaris*
- *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*

Tâches du système Sun Cluster

L'exécution de toutes les tâches du système Sun Cluster nécessitent l'acquisition de certaines notions fondamentales. Le tableau suivant présente une vue d'ensemble détaillée des tâches et de la documentation illustrant les étapes d'exécution de ces tâches. Les rubriques du présent ouvrage décrivent le lien entre les notions fondamentales et ces tâches.

TABLEAU 1–1 Plan des tâches : correspondance entre la documentation et les tâches utilisateur

Tâche	Instructions
Installation du matériel de cluster	*Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS*
Installation du logiciel Solaris sur le cluster	*Guide d'installation du logiciel Sun Cluster pour SE Solaris*
SPARC : installation du logiciel Sun™ Management Center	*Guide d'installation du logiciel Sun Cluster pour SE Solaris*
Installation et configuration du logiciel Sun Cluster	*Guide d'installation du logiciel Sun Cluster pour SE Solaris*
Installation et configuration du logiciel de gestion de volumes	*Guide d'installation du logiciel Sun Cluster pour SE Solaris* Votre documentation relative à la gestion de volumes
Installation et configuration des services de données Sun Cluster	*Sun Cluster Data Services Planning and Administration Guide for Solaris OS*
Maintenance du matériel de cluster	*Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS*
Administration du logiciel Sun Cluster	*Guide d'administration système de Sun Cluster pour SE Solaris*
Administration du logiciel de gestion de volumes	Document *Guide d'administration système de Sun Cluster pour SE Solaris* et documentation de gestion de votre volume
Administration du logiciel d'application	La documentation de l'application
Identification des problèmes et suggestions d'actions aux utilisateurs	*Sun Cluster Error Messages Guide for Solaris OS*
Création d'un nouveau service de données	Document *Guide du développeur de services de données Sun Cluster pour SE Solaris*

CHAPITRE 2

Principales notions fondamentales destinées aux fournisseurs de services matériels

Ce chapitre décrit les notions fondamentales liées aux composants matériel d'une configuration système Sun Cluster. Les sujets traités sont les suivants :

Matériel système et composants logiciels Sun Cluster

Ces informations s'adressent en priorité aux fournisseurs de services matériel. Ces notions peuvent aider les fournisseurs de services à mieux comprendre les relations entre les divers composants matériels avant d'installer, de configurer ou d'entretenir le matériel de cluster. Elles peuvent aussi s'avérer utiles pour les administrateurs système qui y trouveront des informations de référence pour l'installation, la configuration et l'administration du logiciel de cluster.

Un cluster comprend plusieurs composants matériels, notamment :

- des noeuds de cluster avec des disques locaux (non partagés) ;
- un stockage multihôte (disques partagés entre les noeuds) ;

- des médias amovibles (bandes et CD-ROM) ;
- interconnexion de cluster ;
- des interfaces de réseau public ;
- des systèmes client ;
- une console administrative ;
- des dispositifs d'accès à la console.

Le système Sun Cluster vous permet de combiner ces composants et d'obtenir de nombreuses configurations. Les sections suivantes décrivent ces configurations :

- "SPARC : Topologies Sun Cluster pour SPARC " à la page 30
- "x86 : Topologies Sun Cluster pour x86 " à la page 34

Pour obtenir un exemple de configuration d'un cluster à deux nœuds, voir "Environnement matériel de Sun Cluster" du *Présentation de Sun Cluster pour SE Solaris*.

Noeuds de cluster

Un nœud de cluster est une machine exécutant à la fois le système d'exploitation Solaris et le logiciel Sun Cluster. Un nœud de cluster est également un membre actuel du cluster (un *membre de cluster*) ou un membre potentiel.

- le logiciel SPARC : Sun Cluster prend en charge un à seize nœuds par cluster. Pour plus d'informations sur les configurations de nœuds prises en charge, voir "SPARC : Topologies Sun Cluster pour SPARC " à la page 30.
- le logiciel x86: Sun Cluster prend en charge deux nœuds par cluster. Pour plus d'informations sur les configurations de nœuds prises en charge, voir "x86 : Topologies Sun Cluster pour x86 " à la page 34.

Les nœuds de cluster sont généralement connectés à un ou plusieurs périphériques multihôtes. Ceux qui ne le sont pas utilisent le système de fichiers de cluster pour accéder aux périphériques multihôtes. Par exemple, dans une configuration de services évolutifs, les nœuds peuvent traiter les requêtes sans être directement connectés aux périphériques multihôtes.

De plus, les nœuds des configurations de bases de données parallèles partagent un accès simultané à tous les disques.

- Pour plus d'informations sur l'accès simultané aux disques, voir "Périphériques multihôtes " à la page 24.
- Pour plus d'informations sur les configurations de bases de données parallèles, voir "SPARC : Topologie de paire clusterisée pour SPARC " à la page 30 et "x86 : Topologie de paire clusterisée pour x86 " à la page 34.

Tous les nœuds du cluster sont regroupés sous un nom commun, « le nom du cluster », utilisé pour accéder au cluster et le gérer.

Les adaptateurs de réseau public relient les nœuds aux réseaux publics, permettant ainsi au client d'accéder au cluster.

Les membres du cluster communiquent avec les autres nœuds du cluster par le biais d'un ou de plusieurs réseaux physiquement indépendants. Cet ensemble de réseaux physiquement indépendants est désigné sous le nom d'*interconnexion de cluster*.

Chaque noeud du cluster sait lorsqu'un autre noeud rejoint ou quitte le cluster. De même, il sait quelles ressources fonctionnent localement et quelles ressources fonctionnent sur les autres nœuds du cluster.

Les noeuds d'un même cluster doivent avoir des capacités de traitement, de mémoire et d'E/S similaires afin que les basculements éventuels s'effectuent sans dégradation significative des performances. En raison de la possibilité de basculements, tous les nœuds doivent disposer d'une capacité excédentaire leur permettant de gérer la charge de travail de tous les nœuds dont ils constituent le noeud secondaire ou de sauvegarde.

Chaque noeud initialise son propre système de fichiers racine (/).

Composants logiciels des membres matériels du cluster

Pour fonctionner comme un membre du cluster, un noeud doit être équipé des logiciels suivants :

- Système d'exploitation Solaris
- Logiciel Sun Cluster
- application de service de données ;
- Gestion des volumes (Solaris Volume Manager™ ou VERITAS Volume Manager)

  Une configuration utilisant un RAID (réseau redondant de disques indépendants) matériel constitue une exception. Cette configuration ne nécessite pas de gestionnaire de volume logiciel comme Solaris Volume Manager ou VERITAS Volume Manager.
- Pour plus d'informations sur l'installation du système d'exploitation Solaris, de Sun Cluster et du logiciel de gestion de volume, voir *Guide d'installation du logiciel Sun Cluster pour SE Solaris*.
- Pour plus d'informations sur l'installation et la configuration des services de données, voir *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.
- Pour plus d'informations conceptuelles sur les composants logiciels précédents, voir Chapitre 3.

La figure ci-après affiche une vue générale des composants logiciels qui fonctionnent ensemble pour créer l'environnement logiciel Sun Cluster.

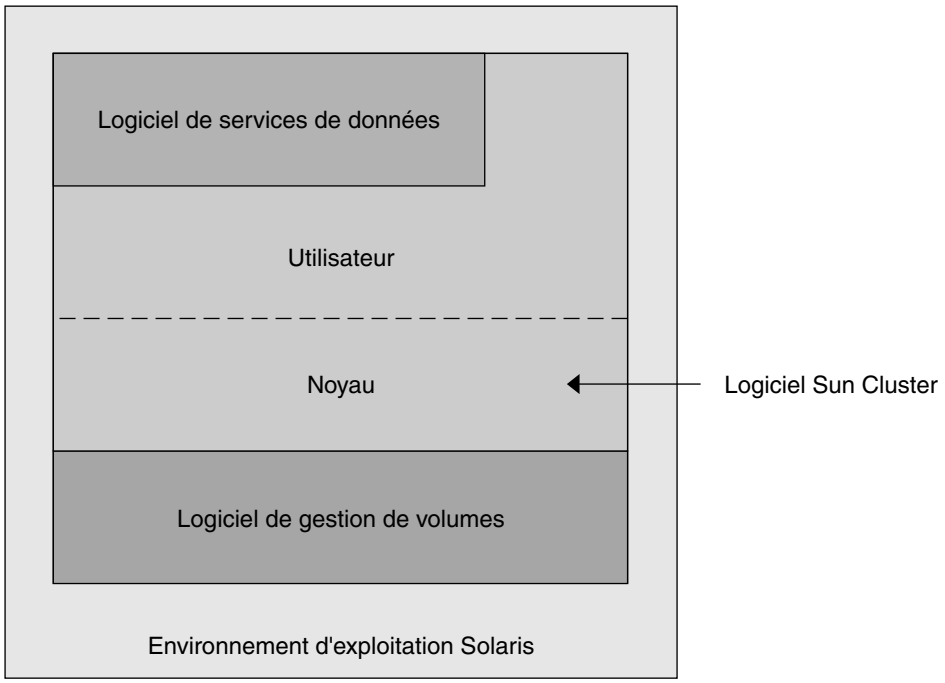


FIGURE 2–1 Relations générales des composants logiciels Sun Cluster

Pour consulter les questions et les réponses sur les membres de cluster, voir Chapitre 4.

Périphériques multihôtes

Les disques pouvant être connectés à plus d'un nœud à la fois sont appelés périphériques multihôtes. Dans l'environnement Sun Cluster, le stockage multihôte rend les disques hautement disponibles. Le logiciel Sun Cluster requiert le stockage multihôte pour les clusters à deux nœuds, afin d'établir un quorum. Les clusters supérieurs à deux nœuds ne nécessitent pas de périphériques de quorum. Pour plus d'informations sur le quorum, voir "Quorum et périphériques de quorum " à la page 55.

Les périphériques multihôtes possèdent les caractéristiques suivantes :

- la tolérance aux pannes de nœuds uniques.
- la capacité à stocker des données d'application, des binaires d'applications et des fichiers de configuration.

- la protection contre les pannes de nœuds. Si les requêtes client accèdent aux données via un nœud défaillant, elles sont basculées vers un autre nœud ayant une connexion directe aux mêmes disques.
- accès soit global par un nœud principal « gérant » les disques, soit simultané direct via les chemins locaux. Actuellement, Oracle Real Application Clusters Guard est la seule application à utiliser l'accès simultané direct.

Un gestionnaire de volume fournit aux configurations miroirs ou RAID-5 une redondance de données des périphériques multihôtes. Actuellement, Sun Cluster prend en charge les gestionnaires de volume Solaris Volume Manager et VERITAS Volume Manager, qui peut être utilisé uniquement sur les clusters SPARC, ainsi que le contrôleur de matériel RAID-5 RDAC sur plusieurs plates-formes RAID matérielles.

La combinaison de périphériques multihôtes avec la mise en miroir et l'entrelacement (striping) protège à la fois contre les pannes de nœuds et les pannes de disques individuels.

Pour consulter les questions et les réponses sur le stockage multihôte, voir Chapitre 4.

SCSI multi-initiateur

Cette rubrique s'applique uniquement aux périphériques de stockage SCSI et non au stockage fibre Channel utilisé pour les périphériques multihôtes.

Sur un serveur autonome, le noeud contrôle les activités du bus SCSI par le biais du circuit de l'adaptateur d'hôte SCSI connectant ce serveur à un bus SCSI particulier. Ce circuit est appelé *SCSI initiateur*. il initie toutes les activités de ce bus SCSI. L'adresse SCSI par défaut des cartes de contrôleur SCSI dans les systèmes Sun est 7.

Dans les configurations en cluster, le stockage est partagé entre plusieurs nœuds du serveur à l'aide de périphériques multihôtes. Si le stockage de cluster est constitué de périphériques SCSI en mode asymétrique ou différentiel, la configuration est appelée « SCSI multi-initiateur ». Comme son nom l'indique, plusieurs initiateurs SCSI figurent sur le bus SCSI.

Cette spécification SCSI nécessite que chaque périphérique d'un bus SCSI ait une adresse SCSI unique. (la carte de contrôleur est également un dispositif du bus SCSI). La configuration matérielle par défaut dans un environnement multi-initiateur entraîne un conflit, car toutes les cartes de contrôleur sont définies sur 7 par défaut.

Pour résoudre ce conflit, sur chaque bus SCSI, laissez l'adresse d'une des cartes de contrôleur sur 7, et définissez tous les autres sur des adresses SCSI non utilisées. Afin de bien répartir ces adresses SCSI « non utilisées », choisissez des adresses non utilisées à l'heure actuelle et non utilisables dans le futur. Par exemple, l'ajout de capacité de stockage par l'installation de nouveaux lecteurs à des emplacements vides crée des adresses non utilisables dans le futur.

Dans la plupart des configurations, l'adresse SCSI disponible pour un second adaptateur d'hôte est 6.

Vous pouvez modifier les adresses SCSI sélectionnées pour ces cartes de contrôleur à l'aide de l'un des outils suivants permettant de définir la propriété `scsi-initiator-id` :

- `eeprom`(1M)
- PROM OpenBoot sur un système SPARC ;
- utilitaire SCSI exécuté de façon optionnelle après initialisation du BIOS sur un système x86.

Elle peut être définie globalement pour un nœud ou individuellement pour chaque carte de contrôleur. Les instructions de définition d'une propriété `scsi-initiator-id` unique pour chaque adaptateur hôte SCSI sont incluses dans *Sun Cluster 3.0-3.1 With SCSI JBOD Storage Device Manual for Solaris OS*.

Disques locaux

Les disques locaux ne sont connectés qu'à un seul noeud. Par conséquent, les disques locaux ne sont pas protégés contre les pannes de nœuds (ils ne sont donc pas hautement disponibles). Toutefois, tous les disques, y compris les disques locaux, sont inclus dans l'espace de noms global et sont configurés comme *périphériques globaux*. Ainsi, les disques eux-mêmes sont visibles depuis tous les nœuds de cluster.

Vous pouvez mettre les systèmes de fichiers des disques locaux à la disposition d'autres nœuds en les plaçant sous un point de montage global. Si un des nœuds sur lequel est monté l'un de ces systèmes de fichiers globaux échoue, tous les nœuds perdent l'accès à ce système de fichiers. L'utilisation d'un gestionnaire de volume vous permet de mettre ces disques en miroir, évitant ainsi qu'une panne ne rende ces systèmes de fichiers inaccessibles ; les gestionnaires de volumes ne protègent cependant pas des pannes de nœuds.

Pour plus d'informations sur les périphériques globaux, voir "Périphériques globaux " à la page 42.

Support amovible

Les supports amovibles tels que les lecteurs de bandes et lecteurs de CD sont pris en charge par le cluster. En général, vous installez, configurez et utilisez ces périphériques de la même façon que dans un environnement non clusterisé. Ces périphériques sont configurés en tant que périphériques globaux dans Sun Cluster, de façon que chaque nœud du cluster puisse accéder à chaque périphérique. Pour plus d'informations sur l'installation et la configuration des supports amovibles, voir *Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS*.

Pour plus d'informations sur les périphériques globaux, reportez-vous à "Périphériques globaux " à la page 42.

Interconnexion de cluster

L'*interconnexion de cluster* est la configuration physique des périphériques utilisés pour le transfert de communications privées du cluster et de celles des services de données entre les noeuds du cluster. L'interconnexion étant fortement sollicitée pour les communications privées du cluster, les performances peuvent être amoindries.

Seuls les nœuds du cluster peuvent être y connectés. Le modèle de sécurité de Sun Cluster suppose que seuls les nœuds de cluster disposent d'un accès physique à l'interconnexion de cluster.

Tous les nœuds doivent être connectés par l'interconnexion de cluster via au minimum deux réseaux ou chemins d'accès redondants physiquement indépendants, afin d'éviter un point de panne unique. Il peut y avoir plusieurs réseaux physiquement indépendants (de deux à six) entre deux noeuds, quels qu'ils soient.

L'interconnexion de cluster se compose de trois éléments matériels : les adaptateurs, les jonctions et les câbles. La liste présentée ci-dessous décrit chacun de ces éléments matériels.

- Adaptateurs : cartes d'interface réseau résidant dans chaque nœud du cluster. Leur nom se compose d'un nom de périphérique immédiatement suivi d'un numéro d'unité physique, par exemple, qfe2. Certains adaptateurs n'ont qu'une connexion physique au réseau, tandis que d'autres, comme la carte qfe, en possèdent plusieurs. Certains contiennent à la fois des interfaces réseau et des interfaces de stockage.

  Un adaptateur réseau à plusieurs interfaces peut entraîner un point de panne unique s'il tombe en panne. Pour assurer une disponibilité maximale, planifiez votre cluster de sorte que l'unique chemin d'accès entre deux nœuds ne dépende pas d'un seul adaptateur réseau.
- Jonctions : commutateurs résidant à l'extérieur des noeuds du cluster. Ils remplissent des fonctions d'intercommunication et de commutation vous permettant de connecter plus de deux nœuds entre eux. Dans un cluster à deux noeuds, vous n'avez pas besoin de jonctions car les noeuds peuvent être directement connectés l'un à l'autre via des câbles physiques redondants connectés à des adaptateurs redondants sur chaque noeud. Les configurations à plus de deux noeuds requièrent généralement des jonctions.
- Câbles : connexions physiques installées entre deux adaptateurs réseau ou entre un adaptateur et une jonction.

Pour consulter les questions et les réponses sur l'interconnexion de cluster, voir Chapitre 4.

Interfaces de réseau public

Les clients se connectent au cluster via des interfaces de réseau public. Chaque carte d'adaptateur réseau peut être connectée à un ou plusieurs réseaux publics, selon si elle possède plusieurs interfaces matérielles ou non. Vous pouvez définir des noeuds pour

qu'ils prennent en charge plusieurs cartes d'interfaces de réseau public configurées de façon à ce qu'elles soient toutes actives, et puissent ainsi se servir mutuellement de sauvegarde en cas de basculement. Si l'un des adaptateurs tombe en panne, le logiciel multi-acheminement sur réseau IP est appelé et bascule l'interface défectueuse sur un autre adaptateur du groupe.

Aucune remarque d'ordre matériel particulière ne s'applique au clustering pour les interfaces de réseau public.

Pour consulter les questions et les réponses sur les réseaux publics, voir Chapitre 4.

Systèmes client

Les systèmes client comprennent des stations de travail et autres serveurs accédant au cluster via le réseau public. Les programmes côté client utilisent les données ou d'autres services des applications côté serveur qui s'exécutent sur le cluster.

Les systèmes client ne sont pas hautement disponibles. Par contre, les données et applications du cluster le sont.

Pour consulter les questions et les réponses sur les systèmes client, voir Chapitre 4.

Dispositifs d'accès à la console

Vous devez disposer d'un accès par console à l'ensemble des noeuds de cluster. Pour accéder à la console, utilisez l'un des périphériques suivants :

- le concentrateur de terminaux acquis avec votre matériel de cluster ;
- le processeur SSP sur les serveurs Sun Enterprise E10000 (pour les clusters SPARC) ;
- le contrôleur de système sur les serveurs Sun Fire™ (également pour les clusters SPARC) ;
- un autre périphérique pouvant accéder à `ttya` sur chaque nœud.

Un seul concentrateur de terminaux pris en charge est disponible auprès de Sun et son utilisation n'est pas obligatoire. Il permet d'accéder à `/dev/console` sur chaque noeud, via un réseau TCP/IP. Vous disposez ainsi d'un accès par console pour chaque nœud à partir d'une station de travail distante située n'importe où sur le réseau.

Le SSP propose un accès à la console aux serveurs Sun Enterprise E1000. Le SSP est une machine installée sur un réseau Ethernet qui est configurée pour prendre en charge le serveur Sun Enterprise E1000. Pour le serveur Sun Enterprise E1000, il s'agit de la console administrative. En utilisant les fonctions de la console de réseau Sun Enterprise E10000, toutes les stations de travail du réseau peuvent ouvrir une session de console hôte.

Les autres méthodes d'accès par la console comprennent des concentrateurs de terminaux, un accès au port série `tip(1)` à partir d'un autre nœud et des terminaux non intelligents. Vous pouvez utiliser les claviers et moniteurs Sun™ ou d'autres périphériques port série si votre fournisseur de services matériels les prend en charge.

Console administrative

Vous pouvez utiliser une station de travail UltraSPARC® dédiée ou un serveur Sun Fire V65x, appelés *console administrative*, pour administrer le cluster actif. Habituellement, vous pouvez installer et exécuter un logiciel outil administratif, par exemple le CCP et le module Sun Cluster pour le produit Sun Management Center (à utiliser uniquement avec des clusters SPARC) sur la console administrative. L'option `cconsole` du CCP vous permet de vous connecter à plus d'un noeud en même temps. Pour plus d'informations sur l'utilisation du CCP, voir Chapitre 1, "Introduction à l'administration de Sun Cluster" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

La console administrative n'est pas un nœud de cluster. Elle permet d'accéder à distance aux noeuds du cluster, soit via le réseau public, soit éventuellement à travers un concentrateur de terminaux basé sur le réseau. Si votre cluster est constitué de la plate-forme Sun Enterprise E10000, vous devez vous connecter au SSP à partir de la console administrative, à l'aide de la commande `netcon(1M)`.

Les noeuds sont généralement configurés sans moniteur. Vous accédez à leur console par le biais d'une session `telnet` à partir de la console administrative. Cette console est connectée à un concentrateur de terminaux, lequel est connecté au port série du nœud. S'il s'agit d'un serveur Sun Enterprise E1000, vous vous connectez à partir du SSP. Pour plus d'informations, voir "Dispositifs d'accès à la console " à la page 28.

Avec Sun Cluster, il n'est pas nécessaire d'avoir une console administrative dédiée, mais elle présente les avantages suivants :

- Elle permet une gestion centralisée des clusters en regroupant les outils de gestion et de console sur la même machine.
- Elle peut permettre une résolution plus rapide des problèmes par votre fournisseur de services matériels.

Pour consulter les questions et les réponses sur la console administrative, voir Chapitre 4.

SPARC : Topologies Sun Cluster pour SPARC

Une topologie est un plan de connexion reliant les nœuds de cluster aux plates-formes de stockage d'un environnement Sun Cluster. Le logiciel Sun Cluster prend en charge toutes les topologies respectant les directives indiquées ci-dessous.

- Un environnement Sun Cluster composé de systèmes SPARC prend en charge un maximum de seize nœuds d'un cluster, quelles que soient les configurations de stockage implémentées.
- Un périphérique de stockage partagé peut être connecté à autant de noeuds qu'il peut en prendre en charge.
- Les périphériques de stockage partagés n'ont pas besoin d'être connectés à tous les nœuds du cluster. Toutefois, ils doivent être connectés à au moins deux noeuds.

Le logiciel Sun Cluster ne nécessite pas que vous configuriez un cluster en utilisant des topologies particulières. Les topologies décrites ci-dessous ont pour but de fournir le vocabulaire nécessaire à une discussion relative à un plan de connexion de cluster, et correspondent à des plans de connexions classiques :

- Paire clusterisée
- paire+N ;
- N+1 (étoile) ;
- N*N (évolutive).

Les rubriques suivantes présentent des exemples schématiques de chaque topologie.

SPARC : Topologie de paire clusterisée pour SPARC

Une topologie de paire clusterisée correspond à deux nœuds fonctionnant dans le cadre administratif d'un même cluster. Dans cette configuration, les basculements ne se produisent qu'entre deux éléments d'une paire. Toutefois, tous les nœuds sont connectés par l'interconnexion de cluster et sont contrôlés par le logiciel Sun Cluster. Cette topologie peut être utilisée pour exécuter une application de base de données parallèle sur une paire, et une application évolutive ou de basculement sur une autre paire.

En utilisant le système de fichiers de cluster, vous pouvez également disposer d'une configuration à deux paires. Plus de deux nœuds peuvent exécuter un service évolutif ou une base de données parallèle même si tous les nœuds ne sont pas directement connectés aux disques qui stockent les données d'application.

La figure présentée ci-après illustre une configuration de paires clusterisées.

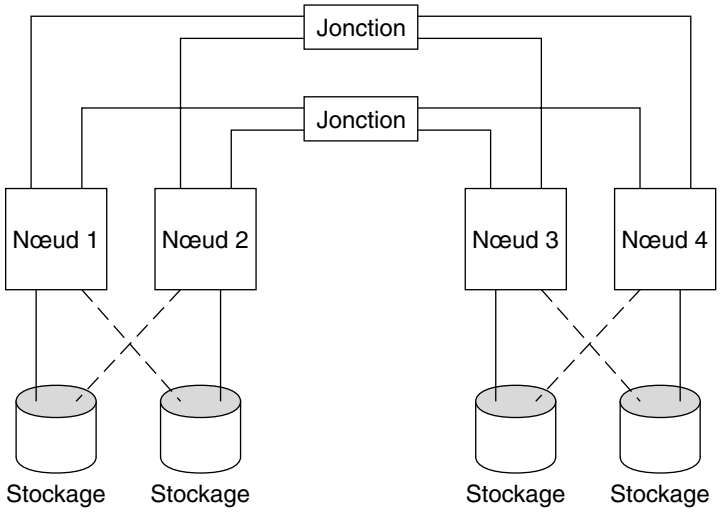


FIGURE 2–2 SPARC : topologie de paire clusterisée

SPARC : Topologie paire+N pour SPARC

La topologie paire+N comprend une paire de nœuds directement connectés au système de stockage partagé et un ensemble de nœuds utilisant l'interconnexion de cluster pour accéder au système de stockage partagé ; ils n'ont eux-mêmes pas de connexion directe.

La figure présentée ci-après illustre une topologie paire+N où deux des quatre noeuds (le noeud 3 et le noeud 4) utilisent l'interconnexion de cluster pour accéder au système de stockage. Cette configuration peut être étendue et inclure des noeuds supplémentaires n'ayant pas d'accès direct au système de stockage partagé.

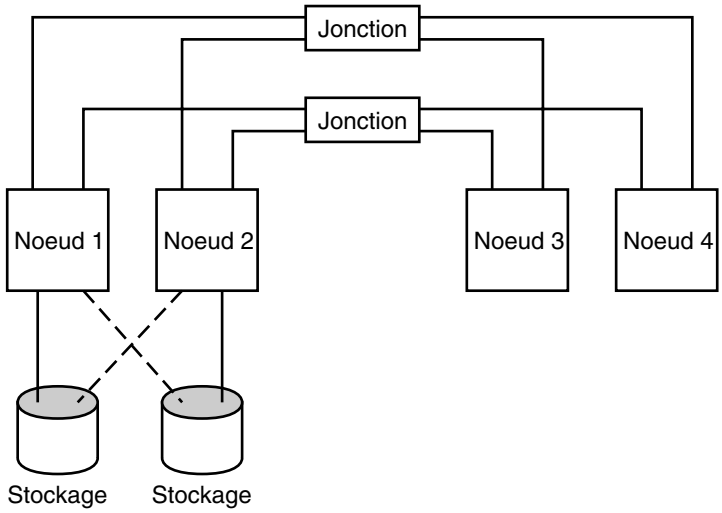


FIGURE 2–3 topologie paire+N

SPARC : Topologie N+1 (étoile) pour SPARC

Une topologie N+1 comprend un certain nombre de noeuds principaux et un noeud secondaire. Vous n'avez pas besoin de configurer les noeuds principaux et le noeud secondaire de la même manière. Les noeuds principaux sont activement impliqués dans la fourniture de services d'application. Le nœud secondaire n'a pas besoin d'être inactif en attendant la panne d'un nœud principal.

Le noeud secondaire est le seul noeud de la configuration à être physiquement connecté à tout le système de stockage multihôte.

Si une erreur se produit sur un nœud principal, Sun Cluster bascule les ressources sur le nœud secondaire, sur lequel elles fonctionnent jusqu'à ce qu'elles basculent de nouveau (automatiquement ou manuellement) sur le nœud principal.

Le nœud secondaire doit toujours avoir une capacité de CPU suffisamment importante pour assumer la charge supplémentaire d'un nœud principal défectueux.

La figure présentée ci-après illustre une configuration N+1.

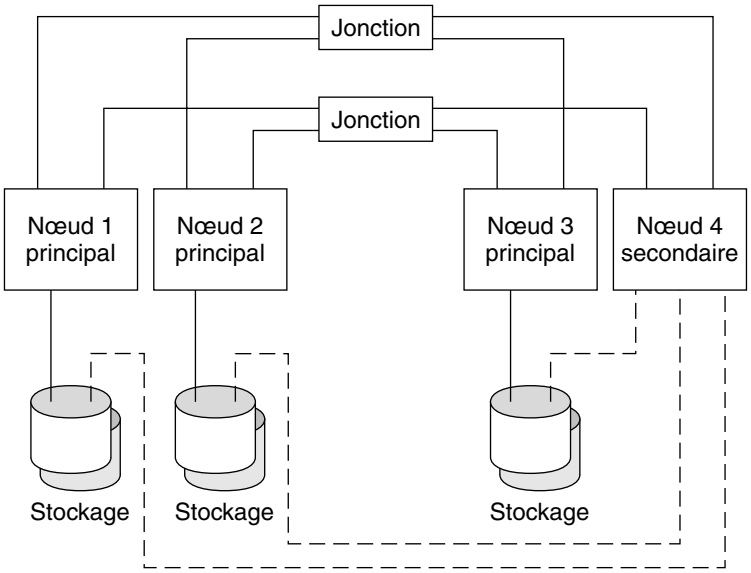


FIGURE 2–4 SPARC : topologie N+1

SPARC : Topologie N*N (évolutive) pour SPARC

Une topologie N*N permet à tous les périphériques de stockage partagés du cluster de se connecter à tous les nœuds de ce dernier. Cette topologie permet aux applications à haut niveau de disponibilité de basculer d'un nœud sur un autre sans occasionner de dégradation de service. En cas de basculement, le nouveau nœud peut accéder au périphérique de stockage en utilisant un chemin d'accès local et non l'interconnexion privée.

La figure suivante illustre une configuration N*N.

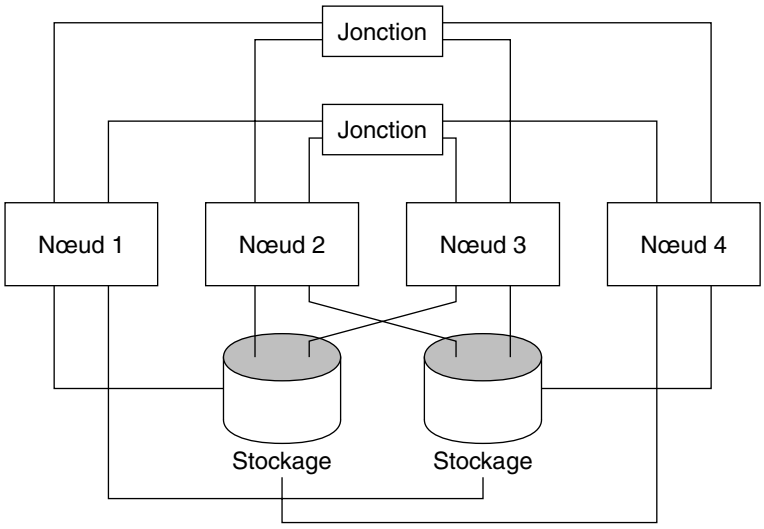


FIGURE 2–5 SPARC : topologie N*N

x86 : Topologies Sun Cluster pour x86

Une topologie est un plan de connexion reliant les nœuds de cluster aux plates-formes de stockage du cluster. Sun Cluster prend en charge toutes les topologies respectant les directives indiquées ci-dessous.

- Sun Cluster constitué de systèmes x86 prend en charge des clusters à deux nœuds.
- Les périphériques de stockage partagés doivent être connectés aux deux noeuds.

Sun Cluster ne requiert pas de topologies spécifiques pour la configuration d'un cluster. La topologie de paire clusterisée suivante, seule topologie pour clusters composés de nœuds x86, est présentée pour introduire la terminologie du schéma de connexion d'un cluster. Elle correspond à un schéma de connexion type.

La rubrique suivante présente un exemple de diagramme de cette topologie.

x86 : Topologie de paire clusterisée pour x86

Une topologie de paire clusterisée correspond à deux nœuds fonctionnant dans le cadre administratif d'un même cluster. Dans cette configuration, les basculements ne se produisent qu'entre deux éléments d'une paire. Toutefois, tous les nœuds sont

connectés par l'interconnexion de cluster et fonctionnent sous le contrôle du logiciel Sun Cluster. Cette topologie peut être utilisée pour exécuter une application de base de données parallèle, évolutive ou de basculement sur la paire.

La figure présentée ci-après illustre une configuration de paires clusterisées.

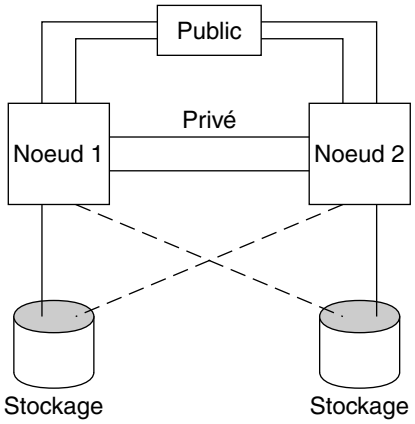


FIGURE 2–6 x86 : topologie de paire clusterisée

CHAPITRE 3

Notions-clés destinées aux administrateurs système et aux développeurs d'applications

Ce chapitre décrit les notions-clés associées aux composants logiciels d'un système Sun Cluster. Il aborde les sujets suivants :

Ces informations sont destinées principalement aux administrateurs système et aux développeurs d'applications utilisant l'API et le kit de développement logiciel (SDK) Sun Cluster. Les administrateurs système y trouveront des informations de fond pour l'installation, la configuration et l'administration du logiciel de cluster. Les développeurs d'applications y trouveront les informations nécessaires à la compréhension de l'environnement de cluster dans lequel ils travailleront.

Interfaces administratives

Plusieurs interfaces utilisateur vous permettent d'installer, de configurer et d'administrer le système Sun Cluster. Les tâches d'administration système peuvent être effectuées par l'intermédiaire de l'interface utilisateur graphique de SunPlex Manager ou via l'interface de ligne de commande. Outre l'interface de ligne de commande, il existe des utilitaires, comme `scinstall` et `scsetup` qui permettent de simplifier les tâches d'installation et de configuration sélectionnées. Le système Sun Cluster intègre également un module fonctionnant avec Sun Management Center qui fournit une interface graphique utilisateur pour l'exécution de certaines tâches du cluster. Ce module est disponible uniquement pour les clusters SPARC. Pour une description complète des interfaces administratives, voir "Outils d'administration" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

Heure du cluster

L'heure entre tous les nœuds du cluster doit être synchronisée. La synchronisation éventuelle des nœuds du cluster avec une heure extérieure n'a aucune incidence sur son fonctionnement. Le système Sun Cluster utilise le protocole NTP pour synchroniser les horloges entre les nœuds.

En général, si l'horloge système est modifiée d'une fraction de seconde, cela ne pose aucun problème. Cependant, si vous exécutez `date(1)`, `rdate(1M)` ou `xntpdate(1M)` (de façon interactive ou dans des scripts `cron`) sur un cluster actif, vous pouvez modifier une heure de plus d'une fraction de seconde pour synchroniser l'horloge système avec la source de synchronisation. Cette modification forcée peut alors entraîner des problèmes avec l'horodateur des modifications de fichiers ou troubler le service NTP.

Si vous installez le système d'exploitation Solaris sur chaque nœud de cluster, vous pouvez modifier l'heure et la date par défaut de chaque nœud. Vous pouvez en général accepter les paramètres par défaut.

Si vous installez le logiciel Sun Cluster à l'aide de `scinstall(1M)`, vous devez configurer le protocole NTP pour le cluster. Le logiciel Sun Cluster contient un fichier modèle, `ntp.cluster` (voir `/etc/inet/ntp.cluster` sur un nœud de cluster installé) qui crée une relation d'homologue entre tous les nœuds de cluster. Un nœud est désigné comme étant le nœud "préféré". Les nœuds sont identifiés par leurs noms d'hôte privés et la synchronisation de l'heure se produit à travers l'interconnexion du cluster. Pour obtenir des instructions sur la configuration du cluster pour le protocole NTP, reportez-vous au Chapitre 2, "Installation et configuration du logiciel Sun Cluster" du *Guide d'installation du logiciel Sun Cluster pour SE Solaris*.

Vous pouvez aussi installer un ou plusieurs serveurs NTP à l'extérieur du cluster et modifier le fichier `ntp.conf` pour qu'il reflète cette configuration.

Lors d'un fonctionnement normal, vous ne devriez jamais avoir à modifier l'heure du cluster. Cependant si l'heure n'était pas réglée correctement lorsque vous avez installé le système d'exploitation Solaris et si vous souhaitez la modifier, la procédure adéquate est incluse dans le Chapitre 7, "Administration du cluster" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

Structure à haute disponibilité

Grâce au système Sun Cluster, tous les composants qui se trouvent sur le « chemin » reliant les utilisateurs aux données, y compris les interfaces réseau, les applications elles-mêmes, le système de fichiers et les disques multihôtes, sont hautement disponibles. Un composant du cluster est dit hautement disponible s'il est capable de survivre à toute panne (matérielle ou logicielle) unique du système.

Le tableau suivant affiche les types de panne des composants Sun Cluster (tant matériels que logiciels) et les types de récupération intégrés à la structure à haute disponibilité :

TABLEAU 3–1 Niveaux de détection de panne et de récupération de Sun Cluster

Composant du cluster en panne	Récupération logicielle	Récupération matérielle
Service de données	API HD, structure HD	Non applicable
Adaptateur de réseau public	multi-acheminement sur réseau IP	Plusieurs cartes d'adaptateur de réseau public
Systèmes de fichiers de cluster	Répliques principales et secondaires	Périphériques multihôtes
Périphérique multihôte mis en miroir	Gestion des volumes (Solaris Volume Manager et VERITAS Volume Manager, qui est disponible uniquement sur les clusters SPARC)	RAID-5 matériel (par exemple, Sun StorEdge™ A3x00)
Périphérique global	Répliques principales et secondaires	Plusieurs chemins d'accès au périphérique, jonctions de transport intracluster
Réseau privé	Logiciel de transport HD	Plusieurs réseaux matériels privés indépendants

TABLEAU 3–1 Niveaux de détection de panne et de récupération de Sun Cluster *(Suite)*

Composant du cluster en panne	Récupération logicielle	Récupération matérielle
Nœud	Moniteur d'appartenance au cluster, pilote failfast	Plusieurs nœuds

La structure à haute disponibilité du logiciel Sun Cluster détecte rapidement une panne de nœud et crée un serveur équivalent pour les ressources de la structure sur un nœud restant du cluster. Les ressources de la structure ne sont jamais toutes indisponibles en même temps. Celles qui ne sont pas affectées par une panne de nœud restent totalement disponibles pendant la récupération. En outre, celles du nœud défectueux redeviennent disponibles dès que la récupération est terminée. Une ressource de la structure récupérée n'a pas à attendre la récupération de toutes les autres.

La plupart des ressources de la structure à haute disponibilité sont récupérées de façon transparente pour les applications (services de données) les utilisant. La sémantique d'accès aux ressources de la structure est totalement préservée après l'échec d'un nœud. Les applications ne peuvent pas détecter que le serveur des ressources de la structure a été déplacé sur un autre nœud. La panne d'un nœud unique est complètement transparente pour les programmes des nœuds restants utilisant les fichiers, périphériques et volumes de disques connectés à ce nœud. Cette transparence est possible s'il existe un autre chemin matériel vers les disques d'un autre nœud. On peut par exemple utiliser des périphériques multihôtes ayant des ports connectés à plusieurs nœuds.

Moniteur d'appartenance au cluster

Pour assurer que les données ne s'altèrent pas, tous les nœuds doivent arriver à un accord cohérent sur l'appartenance au cluster. Si nécessaire, le moniteur d'appartenance au cluster coordonne la reconfiguration des services (applications) du cluster en réponse à une panne.

Le MAC reçoit des informations sur la connectivité aux autres nœuds depuis la couche de transport intracluster. Il utilise l'interconnexion du cluster pour échanger des informations d'état au cours d'une reconfiguration.

Après avoir détecté une modification d'appartenance au cluster, le CMM effectue une configuration synchronisée du cluster. Dans une configuration synchronisée, les ressources du cluster peuvent être redistribuées, en fonction de la nouvelle appartenance au cluster.

Contrairement aux versions précédentes du logiciel Sun Cluster, le CMM s'exécute entièrement dans le noyau.

Pour plus d'informations sur la protection du cluster contre le partitionnement en plusieurs clusters distincts, voir "À propos de la séparation en cas d'échec " à la page 57.

Mécanisme failfast

Si le CMM détecte un problème crucial sur un nœud, il demande à la structure du cluster de l'arrêter de force et de le supprimer de l'appartenance au cluster. Ce mécanisme d'arrêt et de suppression est appelé *failfast*. Il entraîne l'arrêt d'un nœud de deux façons.

- Si un nœud quitte le cluster puis tente de démarrer un nouveau cluster sans avoir de quorum, il est « séparé » pour être empêché d'accéder aux disques partagés. Pour plus d'informations sur l'utilisation du mécanisme failfast, voir "À propos de la séparation en cas d'échec " à la page 57.
- Si un ou plusieurs démons propres au cluster meurent (`clexecd`, `rpc.pmfd`, `rgmd` ou `rpc.ed`), la panne est détectée par le CMM et le nœud panique.

Lorsque la mort d'un démon de cluster provoque la panique d'un nœud, un message semblable au suivant s'affiche sur la console correspondant à ce nœud.

```
panic[cpu0]/thread=40e60: Failfast: Aborting because "pmfd" died 35 seconds ago.
409b8 cl_runtime:__0FZsc_syslog_msg_log_no_argsPviTCPCcTB+48 (70f900, 30, 70df54, 407acc, 0)
%l0-7: 1006c80 000000a 000000a 10093bc 406d3c80 7110340 0000000 4001 fbf0
```

Après la panique, le nœud peut redémarrer et tenter de rejoindre le cluster. Si le cluster est composé de systèmes SPARC, le nœud peut rester à l'invite de la PROM OpenBoot™. L'action suivante du nœud est déterminée par la définition du paramètre `auto-boot?`. Vous pouvez le définir avec `eeprom`(1M) à l'invite `ok` de la PROM OpenBoot.

Référentiel de configuration du cluster (CCR)

Le CCR utilise un algorithme de validation à deux phases pour les mises à jour : une mise à jour doit être effectuée sur tous les membres de cluster, faute de quoi elle est annulée. Le CCR utilise l'interconnexion du cluster pour appliquer les mises à jour distribuées.

Caution – Le CCR est constitué de fichiers texte, mais vous ne devez jamais les modifier manuellement. Chaque fichier contient une somme de contrôle enregistrée pour assurer la cohérence entre les nœuds. Une mise à jour manuelle de ces fichiers peut provoquer l'arrêt d'un nœud ou de l'ensemble du cluster.

Le CCR s'appuie sur le moniteur d'appartenance pour garantir qu'un cluster ne fonctionne que si un quorum a été atteint. Il est chargé de vérifier la cohérence des données au sein du cluster, d'effectuer des récupérations lorsque cela s'aèvre nécessaire et de faciliter les mises à jour des données.

Périphériques globaux

Le système Sun Cluster utilise des *périphériques globaux* pour fournir à tout périphérique d'un cluster un accès hautement disponible dans l'ensemble du cluster, à partir de n'importe quel nœud, quelle que soit la connexion physique de ce périphérique. En général, si un nœud tombe en panne alors qu'un périphérique global y a accès, le logiciel Sun Cluster change automatiquement le chemin vers ce périphérique et redirige l'accès en utilisant ce nouveau chemin. Les périphériques globaux de Sun Cluster comprennent les disques, les CD et les bandes. Cependant les disques sont les seuls périphériques globaux multiports pris en charge par le logiciel Sun Cluster. Actuellement, les lecteurs de CD-ROM et de bande ne sont pas des périphériques hautement disponibles. Les disques locaux installés sur chaque serveur ne disposent pas non plus d'accès multiples et ne sont donc pas hautement disponibles.

Le cluster assigne automatiquement un ID unique à chaque disque, CD-ROM et périphérique de bande du cluster. Cela permet un accès cohérent à chaque périphérique à partir de n'importe quel nœud du cluster. L'espace de noms du périphérique global se trouve dans le répertoire `/dev/global`. Pour plus d'informations, voir "Espace de noms global " à la page 47.

Les périphériques globaux à accès multiples fournissent plus d'un chemin d'accès au périphérique. Comme les disques multihôtes font partie d'un groupe de périphériques de disques hébergé par plusieurs nœuds, ils sont hautement disponibles.

ID de périphérique et pilote de pseudo IDP

Le logiciel Sun Cluster gère les périphériques globaux à travers une structure appelée pilote de pseudo IDP (ID de périphérique). Ce pilote est utilisé pour assigner automatiquement un ID unique à chaque périphérique du cluster, notamment aux disques multihôtes, aux lecteurs de bandes et aux CD.

Ce pilote fait partie intégrante de la fonction d'accès aux périphériques globaux du cluster. Il teste tous les nœuds du cluster et établit la liste des périphériques de disques uniques, assignant à chacun un numéro unique majeur et mineur d'une façon cohérente. L'accès aux périphériques globaux est effectué via l'ID de périphérique unique et non les ID de périphérique standard Solaris, par exemple `c0t0d0` pour un disque.

Cette méthode garantit que toute application accédant aux disques (comme un gestionnaire de volume ou des applications utilisant des périphériques bruts) utilise un chemin d'accès cohérent à travers le cluster. Cette cohérence est particulièrement importante pour les disques multihôtes, car les numéros majeurs et mineurs de chaque périphérique peuvent varier d'un nœud à l'autre, modifiant ainsi les conventions

d'attribution de nom des périphériques Solaris. Le nœud 1 (`Node1`) peut par exemple considérer un disque multihôte comme `c1t2d0`, tandis que le nœud 2 (`Node2`) peut considérer ce même disque de manière totalement différente, comme `c3t2d0` , par exemple. Le pilote d'IDP assigne un nom global, tel que `d10`, que les nœuds utilisent. Ainsi, chaque nœud obtient un mappage cohérent vers les disques multihôtes.

Les commandes `scdidadm(1M)` et `scgdevs(1M)` vous permettent de mettre à jour et de gérer les ID de périphérique. Pour plus d'informations, reportez-vous aux pages man suivantes :

- `scdidadm`(1M)
- `scgdevs`(1M)

Groupes de périphériques de disques

Dans le système Sun Cluster, tous les périphériques multihôtes doivent être sous le contrôle du logiciel Sun Cluster. Créez d'abord des groupes de disques du gestionnaire de volume—des ensembles de disques Solaris Volume Manager ou des groupes de disques VERITAS Volume Manager (qui peuvent être utilisés uniquement sur des clusters SPARC)—sur les disques multihôtes. Vous enregistrez ensuite les groupes de disques du gestionnaire de volume comme *groupes de périphériques de disques*. Un groupe de périphériques de disques est un type de périphérique global. En outre, Sun Cluster crée automatiquement un groupe de périphériques de disques bruts pour chaque disque et bande du cluster. Toutefois, ces groupes de périphériques du cluster restent à l'état hors ligne tant que vous ne les utilisez pas comme périphériques globaux.

L'enregistrement fournit au système Sun Cluster des informations permettant de savoir quels nœuds possèdent un chemin d'accès à quels groupes de disques du gestionnaire de volume. Ceux-ci deviennent alors globalement accessibles au sein du cluster. Si plus d'un nœud peut écrire sur (contrôler) un groupe de périphériques de disques, les données stockées sur ce groupe deviennent hautement disponibles. Ce groupe hautement disponible permet de contenir les systèmes de fichiers du cluster.

Remarque – les groupes de périphériques de disques sont indépendants des groupes de ressources. Un nœud peut contrôler un groupe de ressources (représentant un groupe de services de données) tandis qu'un autre peut contrôler les groupes de disques auxquels les services de données accèdent. Toutefois, il est conseillé de conserver sur un même nœud le groupe de périphériques de disques qui stocke les données d'une application particulière et le groupe de ressources qui contient les ressources de cette application (démon de l'application). Pour plus d'informations sur l'association entre les groupes de périphériques de disques et les groupes de ressources, voir "Relationship Between Resource Groups and Disk Device Groups" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

Lorsqu'un nœud utilise un groupe de périphériques de disques, le groupe de disques du gestionnaire de volume devient « global », car il prend en charge plusieurs chemins vers les disques sous-jacents. Chaque nœud du cluster physiquement relié aux disques multihôtes fournit un chemin d'accès au groupe de périphériques de disques.

Basculement de groupes de périphériques d'un disque

Un boîtier de disque étant connecté à plusieurs nœuds, tous les groupes de périphériques de ce boîtier sont accessibles via un autre chemin en cas d'échec du nœud contrôlant le groupe de périphériques. Cette panne n'affecte pas l'accès au groupe de périphériques sauf pendant le laps de temps nécessaire à la récupération et aux contrôles de cohérence. Durant ce laps de temps, toutes les requêtes sont bloquées (de manière transparente pour l'application) jusqu'à ce que le système rende disponible le groupe de périphériques.

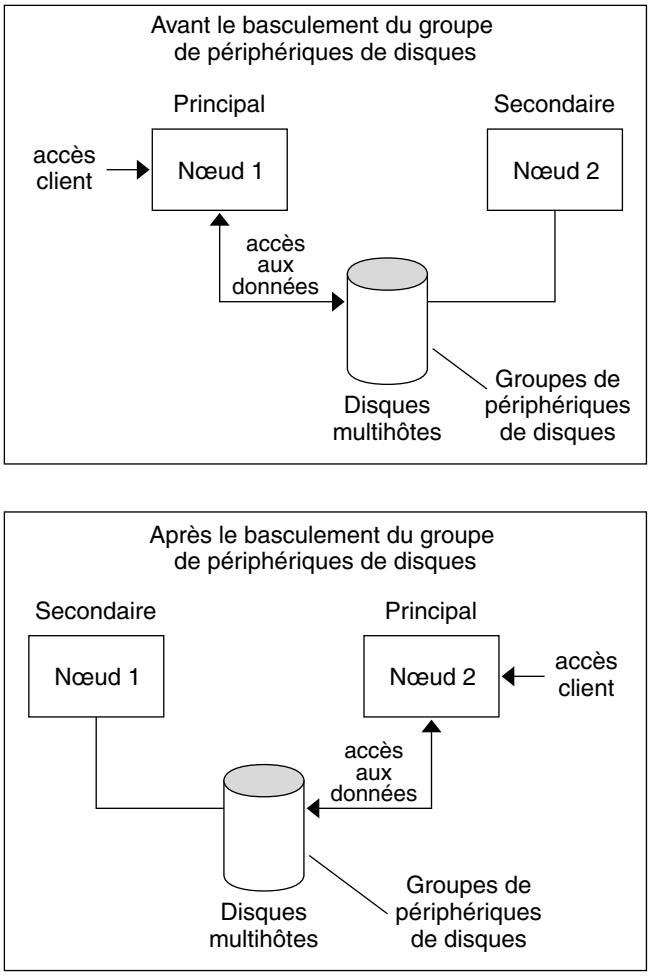


FIGURE 3–1 Groupe de périphériques de disques avant et après le basculement

Groupes de périphériques de disques multiports

Cette section décrit les propriétés des groupes de périphériques de disques qui vous permettent d'équilibrer les performances et la disponibilité dans une configuration de disques multiports. Le logiciel Sun Cluster propose deux propriétés qui permettent de configurer des disques multiports : `preferenced` et `numsecondaries`. Vous pouvez contrôler l'ordre dans lequel les nœuds essaient de prendre le contrôle en cas de basculement à l'aide de la propriété `preferenced`. La propriété `nombre_nœuds_secondaires` permet de définir le nombre souhaité de nœuds secondaires d'un groupe de périphériques.

Un service hautement disponible est dit arrêté si le nœud principal tombe en panne et si aucun nœud secondaire ne peut être promu au rang de nœud principal. Si le service bascule, la propriété `preferenced` étant définie sur `true`, alors les nœuds suivent l'ordre de la liste de nœuds pour en sélectionner un secondaire. La liste de nœuds définit l'ordre dans lequel les nœuds essaient de prendre le contrôle du nœud principal ou de passer de l'état de remplacement à celui de secondaire. La préférence d'un service de périphérique peut être modifiée de manière dynamique à l'aide de l'utilitaire `scsetup`(1M). La préférence associée aux fournisseurs de services dépendants, par exemple un système de fichiers global, est identique à celle du service de périphérique.

Les nœuds secondaires sont contrôlés par le nœud principal au cours du fonctionnement normal. Dans une configuration de disques à accès multiples, le contrôle de chaque nœud secondaire entraîne une dégradation des performances et une surconsommation de mémoire. La prise en charge des nœuds de remplacement a été implémentée pour réduire au minimum la dégradation des performances et la surconsommation de la mémoire provoquées par le contrôle de chaque nœud. Par défaut, votre groupe de périphériques de disques dispose d'un nœud principal et d'un nœud secondaire. Les nœuds disponibles restants deviennent des nœuds de remplacement. En cas de basculement, le nœud secondaire devient principal et le nœud dont la priorité est la plus élevée dans la liste de nœuds devient secondaire.

Le nombre de nœuds secondaires souhaité peut être défini sur n'importe quel entier compris entre un et le nombre de nœuds de fournisseur non principaux opérationnels dans le groupe de périphériques.

Remarque – Si vous utilisez Solaris Volume Manager, vous devez créer le groupe de périphériques de disques avant de définir la propriété `numsecondaries` sur un nombre autre que la valeur par défaut.

Le nombre souhaité de nœuds secondaires par défaut pour les services de périphériques est un. Le nombre réel de fournisseurs secondaires géré par la structure des répliques correspond au nombre souhaité à moins que le nombre de fournisseurs non principaux opérationnels soit inférieur à celui qu'on attend. Vous devez modifier la propriété `numsecondaries` et vérifier la liste de nœuds si vous ajoutez ou supprimez des nœuds dans votre configuration. La gestion de la liste de nœuds et du nombre souhaité de nœuds secondaires empêche tout conflit entre le nombre configuré de nœuds secondaires et le nombre réel autorisé par la structure.

- (Solaris Volume Manager) Utilisez la commande `metaset`(1M) pour les groupes de périphériques Solaris Volume Manager avec les propriétés `preferenced` et `numsecondaries` pour gérer l'ajout et la suppression de nœuds dans votre configuration.
- (Veritas Volume Manager) Utilisez la commande `scconf`(1M) pour les groupes de périphériques de disques VxVM avec les propriétés `preferenced` et `numsecondaries` pour gérer l'ajout et la suppression de nœuds dans votre

configuration.

- Pour plus d'informations sur les procédures à suivre pour modifier les propriétés d'un groupe de périphériques de disques, voir "Administration de systèmes de fichiers de cluster : présentation" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

Espace de noms global

Le mécanisme du logiciel Sun Cluster qui active les périphériques globaux est l'*espace de noms global*. Cet espace comprend la hiérarchie `/dev/global/` ainsi que les espaces de noms du gestionnaire de volume. L'espace de noms global reflète les disques multihôtes et les disques locaux (et tout autre périphérique du cluster, tel que CD et bandes) et fournit aux disques multihôtes plusieurs chemins de basculement. Chaque nœud physiquement connecté aux disques multihôtes fournit un chemin d'accès au stockage pour tout nœud du cluster.

Normalement, dans le cas de Solaris Volume Manager, les espaces de noms du gestionnaire de volume sont situés dans les répertoires `/dev/md/`*diskset*`/dsk` (et `rdsk`). Dans le cas de Veritas VxVM, les espaces de noms du gestionnaire de volume sont situés dans les répertoires `/dev/vx/dsk/`*disk-group* et `/dev/vx/rdsk/`*disk-group*. Ces espaces de noms sont constitués de répertoires pour chaque ensemble de disques Solaris Volume Manager et chaque groupe de disques VxVM importés dans le cluster. Chacun de ces répertoires contient un nœud de périphérique pour chaque métapériphérique ou volume de cet ensemble de disques ou groupe de disques.

Dans le système Sun Cluster, chaque nœud de périphérique de l'espace de noms du gestionnaire de volume local est remplacé par un lien symbolique vers un nœud de périphérique du système de fichiers `/global/.devices/node@` *nodeID*, où *nodeID* est un nombre entier représentant les nœuds du cluster. Le logiciel Sun Cluster continue de présenter les périphériques du gestionnaire de volume, sous forme de liens symboliques, à leur emplacement standard. L'espace de noms global et l'espace de noms du gestionnaire de volume standard sont tous deux disponibles à partir de n'importe quel nœud du cluster.

Les avantages de l'espace de noms global sont décrits ci-dessous.

- Chaque nœud demeure pratiquement indépendant et le modèle d'administration du périphérique est légèrement modifié.
- Les périphériques peuvent devenir globaux de manière sélective.
- Les générateurs de liens de tiers demeurent opérationnels.
- À partir d'un nom de périphérique local donné, on peut facilement établir un mappage pour obtenir son nom global.

Exemples d'espaces de noms locaux et globaux

Le tableau suivant montre les mappages entre les espaces de noms locaux et globaux d'un disque multihôte, `c0t0d0s0`.

TABLEAU 3–2 Mappages des espaces de noms locaux et globaux

Composant ou chemin d'accès	Espace de noms du nœud local	Espace de noms global
Nom logique Solaris	`/dev/dsk/c0t0d0s0`	`/global/.devices/node@`*nodeID*`/dev/dsk/c0t0d0s0`
Nom de l'ID de périphérique	`/dev/did/dsk/d0s0`	`/global/.devices/node@`*nodeID*`/dev/did/dsk/d0s0`
Solaris Volume Manager	`/dev/md/`*diskset*`/dsk/d0`	`/global/.devices/node@`*nodeID*`/dev/md/`*diskset*`/dsk/d0`
SPARC : VERITAS Volume Manager	`/dev/vx/dsk/`*disk-group*`/v0`	`/global/.devices/node@`*nodeID*`/dev/vx/dsk/`*disk-group*`/v0`

L'espace de noms global est automatiquement généré au moment de l'installation et mis à jour à chaque réinitialisation de configuration. Vous pouvez aussi le générer en exécutant la commande `scgdevs(1M)`.

Systèmes de fichiers Cluster

Le système de fichiers de cluster possède les caractéristiques suivantes :

- Les emplacements d'accès aux fichiers sont transparents. Un processus peut ouvrir un fichier situé n'importe où sur le système. Les processus sur tous les nœuds peuvent utiliser le même nom de chemin pour localiser un fichier.

  Remarque – lorsqu'un système de fichiers de cluster lit des fichiers, il ne met pas à jour l'horaire d'accès sur ces fichiers.

- Des protocoles de cohérence sont utilisés pour préserver la sémantique d'accès aux fichiers UNIX même lorsqu'on accède au fichier simultanément à partir de plusieurs nœuds.
- La mise en mémoire cache extensive est utilisée avec des mouvements d'entrée/sortie de masse sans copie pour déplacer les données des fichiers de manière efficace.
- Le système de fichiers d'un cluster fournit des fonctionnalités de verrouillage de fichiers informatif hautement disponible par le biais des interfaces `fcntl(2)`. Les applications exécutées sur plusieurs nœuds peuvent synchroniser l'accès aux

données en utilisant le verrouillage de fichiers informatif sur le fichier d'un système de fichiers du cluster. Les verrous de fichiers sont immédiatement récupérés à partir de nœuds quittant le cluster ou d'applications échouant au verrouillage.

- L'accès aux données est assuré, même en cas de pannes. Les applications ne sont pas affectées par les pannes tant qu'un chemin d'accès aux disques demeure opérationnel. Cette garantie est aussi valable pour l'accès aux disques bruts et pour toutes les opérations du système de fichiers.
- Les systèmes de fichiers de cluster sont indépendants du système de fichiers sous-jacent et du logiciel de gestion de volumes. Les systèmes de fichiers de cluster rendent globaux tous les systèmes de fichiers sur les disques pris en charge.

Vous pouvez monter globalement un système de fichiers sur un périphérique global avec `mount -g` ou localement avec `mount`.

Les programmes peuvent accéder à un fichier d'un système de fichiers de cluster à partir de tout nœud du cluster par le biais du même nom de fichier (par exemple, `/global/foo`).

Un système de fichiers de cluster est monté sur tous les membres du cluster. Il est impossible de le monter sur un sous-ensemble des membres du cluster.

Un système de fichiers de cluster n'est pas un type de système de fichiers à part. Les clients vérifient le système de fichiers sous-jacent (par exemple UFS).

Utilisation des systèmes de fichiers de cluster

Dans le système Sun Cluster, tous les disques multihôtes sont placés dans des groupes de périphériques de disques, tels que des ensembles de disques Solaris Volume Manager, des groupes de disques VxVM ou des disques individuels n'étant pas sous le contrôle d'un gestionnaire de volume de logiciel.

Pour qu'un système de fichiers de cluster soit hautement disponible, le périphérique de stockage de disques sous-jacent doit être connecté à plusieurs nœuds. Ainsi, un système de fichiers local (stocké sur le disque local d'un nœud) créé à l'intérieur d'un système de fichiers de cluster n'est pas hautement disponible.

Vous pouvez monter des systèmes de fichiers de cluster comme vous monteriez tout autre système de fichiers :

- **Manuellement** – Utilisez la commande `mount` et les options de montage `-g` ou `-o global` pour monter le système de fichiers de cluster à partir de la ligne de commande, par exemple :

  ```
  SPARC : # mount -g /dev/global/dsk/d0s0 /global/oracle/data
  ```

- **Automatiquement**– Créez une entrée dans le fichier `/etc/vfstab` avec une option de montage `global` pour monter le système de fichiers de cluster à l'initialisation. Créez ensuite un point de montage sous le répertoire `/global` sur

tous les nœuds. Le répertoire `/global` est conseillé : il n'est nullement impératif. Exemple de ligne pour un système de fichiers de cluster dans un fichier `/etc/vfstab` :

```
SPARC : /dev/md/oracle/dsk/d1 /dev/md/oracle/rdsk/d1 /global/oracle/data ufs 2 yes global,logging
```

Remarque – Le logiciel Sun Cluster n'impose pas de stratégie de dénomination pour les systèmes de fichiers de cluster, mais vous pouvez en faciliter l'administration en créant un point de montage pour tous les systèmes de fichiers de cluster sous le même répertoire, par exemple `/global/`*disk-device-group*. Pour plus d'informations, voir *Sun Cluster 3.1 9/04 Software Collection for Solaris OS (SPARC Platform Edition)* and *Guide d'administration système de Sun Cluster pour SE Solaris*.

Type de ressource `HAStoragePlus`

Le type de ressource `HAStoragePlus` est conçu pour optimiser la disponibilité des configurations des systèmes de fichiers locaux comme UFS (système de fichiers UNIX) et VxFS. Utilisez `HAStoragePlus` pour intégrer votre système de fichiers local à l'environnement Sun Cluster et en optimiser la disponibilité. `HAStoragePlus` intègre des fonctions supplémentaires en matière de système de fichiers telles que les contrôles, les montages et les démontages forcés qui permettent à Sun Cluster de basculer sur les systèmes de fichiers locaux en cas de panne. Pour permettre le basculement, le système de fichiers local doit résider sur des groupes de disques globaux dont les bascules correspondantes sont activées.

Pour plus d'informations sur l'utilisation du type de ressource `HAStoragePlus`, voir "Enabling Highly Available Local File Systems" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

`HAStoragePlus` vous permet également de synchroniser le démarrage des ressources et des groupes de périphériques de disques dont les ressources dépendent. Pour plus d'informations, voir "Ressources, groupes de ressources et types de ressource " à la page 75.

Option de montage `Syncdir`

Vous pouvez utiliser l'option de montage `syncdir` pour les systèmes de fichiers de cluster utilisant UFS en tant que système de fichiers sous-jacent. Toutefois, les performances seront sensiblement améliorées si vous ne la spécifiez pas. Si vous la spécifiez, les écritures sont compatibles POSIX. Si vous ne la spécifiez pas, le comportement est identique à celui des systèmes de fichiers NFS. Par exemple, sans `syncdir`, vous pourriez ne pas détecter de condition d'espace saturé avant de fermer un fichier. Avec `syncdir` (et le système POSIX), l'insuffisance d'espace disponible est découverte au moment de l'opération d'écriture. Les cas dans lesquels vous risquez d'avoir des problèmes si vous n'indiquez pas `syncdir` sont rares.

Si vous utilisez un cluster SPARC, VxFS ne dispose pas d'une option de montage équivalente à l'option de montage `syncdir` du système de fichiers UNIX. VxFS se comporte comme le système de fichiers UNIX lorsque l'option `syncdir` n'est pas spécifiée.

Pour consulter les questions fréquemment posées sur les périphériques globaux et les systèmes de fichiers de cluster, voir "FAQ sur les systèmes de fichiers " à la page 96.

Contrôle de chemin de disque

La version actuelle du logiciel Sun Cluster prend en charge le contrôle de chemin de disque (CCD). Cette rubrique donne des informations théoriques sur le CCD, le démon CCD et les outils d'administration utilisés pour contrôler les chemins d'accès aux disques. Pour plus d'informations sur les procédures de contrôle, de désactivation du contrôle et de vérification du statut des chemins de disques, voir *Guide d'administration système de Sun Cluster pour SE Solaris*.

Remarque – Le CCD n'est pas pris en charge sur les nœuds qui exécutent des versions commercialisées avant le logiciel Sun Cluster 3.1 10/03. N'utilisez pas les commandes de CCD au cours d'une mise à niveau progressive. Lorsque tous les nœuds ont été mis à niveau, ils doivent être en ligne pour permettre l'utilisation des commandes de CCD.

Présentation du CCD

Le CCD améliore la fiabilité globale des basculements et commutations en contrôlant la disponibilité du chemin d'accès au disque secondaire. La commande `scdpm` permet de vérifier la disponibilité du chemin d'accès au disque utilisé par une ressource avant sa commutation. Les options de cette commande vous permettent de contrôler les chemins de disques sur un nœud ou sur tous les nœuds du cluster. Pour plus d'informations sur les options de ligne de commande, voir `scdpm`(1M).

Les composants CCD sont installés à partir du package `SUNWscu`. Ce package est installé par la procédure d'installation de Sun Cluster standard. Pour plus d'informations sur l'interface d'installation, voir `scinstall`(1M). Le tableau suivant décrit l'emplacement d'installation par défaut des composants CCD :

Lieu	Composant
Démon	`/usr/cluster/lib/sc/scdpmd`

Lieu	Composant
Interface de ligne de commande	`/usr/cluster/bin/scdpm`
Bibliothèques partagées	`/user/cluster/lib/libscdpm.so`
Fichier de statut du démon (créé au moment de l'exécution)	`/var/run/cluster/scdpm.status`

Un démon CCD à multifile tourne sur chaque nœud. Le démon CCD (`scdpmd`) est lancé par un script `rc.d` lorsqu'un nœud s'initialise. Si un problème se produit, il est géré par `pmfd` et relancé automatiquement. La liste présentée ci-dessous décrit le fonctionnement de `scdpmd` au moment du démarrage initial.

Remarque – au démarrage, le statut de chaque chemin d'accès au disque est initialisé sur `UNKNOWN`.

---

1. Le démon CCD collecte des informations sur les chemins de disques et les noms des nœuds dans le fichier de statut précédent ou dans la base de données CCR. Pour plus d'informations sur le CCR, voir "Référentiel de configuration du cluster (CCR) " à la page 41. Une fois le démon CCD lancé, vous pouvez le forcer à lire la liste des disques contrôlés à partir d'un nom de fichier spécifié.
2. Le démon CCD initialise l'interface de communication pour répondre aux requêtes de composants extérieurs au démon, tels que l'interface de ligne de commande.
3. Le démon CCD pingue, toutes les dix minutes, l'état de chaque chemin d'accès aux disques inclus dans la liste contrôlée à l'aide des commandes `scsi_inquiry`. Chaque entrée est verrouillée pour empêcher l'interface de communication d'accéder au contenu d'une entrée en cours de modification.
4. Le démon CCD envoie une notification à Sun Cluster Event Framework et consigne le nouveau statut du chemin par le biais du mécanisme UNIX `syslogd`(1M).

Remarque – Toutes les erreurs associées au démon sont rapportées par `pmfd` (1M). Toutes les fonctions de l'API renvoient la valeur `0` en cas de succès et `-1` en cas d'échec.

Le démon CCD contrôle la disponibilité du chemin d'accès logique visible via plusieurs pilotes à chemins multiples, notamment Sun StorEdge Traffic Manager, HDLM et PowerPath. Les chemins d'accès physiques individuels gérés par ces pilotes ne sont pas contrôlés parce que le pilote multivoie masque les pannes individuelles du démon CCD.

Contrôle de chemins de disques

Cette rubrique présente deux méthodes de contrôle des chemins d'accès aux disques dans le cluster. La première méthode consiste à utiliser la commande `scdpm`. Cette commande permet de contrôler, de désactiver le contrôle ou d'afficher le statut des chemins d'accès aux disques dans le cluster. Elle permet également d'imprimer la liste des disques défectueux et de contrôler les chemins d'accès aux disques à partir d'un fichier.

La seconde méthode consiste à utiliser l'interface utilisateur graphique de SunPlex Manager. SunPlex Manager propose une vue topologique des chemins d'accès aux disques contrôlés dans le cluster. Cette vue est mise à jour toutes les 10 minutes afin de donner des informations sur le nombre de pings ayant échoué. Les informations fournies par l'interface utilisateur graphique de SunPlex doivent être utilisées en conjonction avec la commande `scdpm`(1M) pour administrer les chemins d'accès aux disques. Pour plus d'informations sur SunPlex Manager, voir Chapitre 10, "Administration de Sun Cluster avec les interfaces graphiques" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

Utilisation de la commande `scdpm` pour contrôler les chemins d'accès aux disques

La commande `scdpm`(1M) intègre des commandes d'administration CCD vous permettant d'effectuer les tâches suivantes :

- contrôler un nouveau chemin de disque ;
- désactiver le contrôle d'un chemin de disque ;
- relire les données de configuration à partir de la base de données du CCR ;
- lire les disques pour les contrôler ou désactiver le contrôle depuis un fichier spécifié ;
- rapporter le statut d'un chemin ou de tous les chemins d'accès aux disques dans le cluster ;
- imprimer tous les chemins d'accès aux disques accessibles depuis un nœud.

Pour effectuer des tâches d'administration CCD sur le cluster, exécutez la commande `scdpm`(1M) avec l'argument chemin de disque à partir de n'importe quel nœud actif. Celui-ci est toujours constitué d'un nom de nœud et d'un nom de disque. Le nom de nœud n'est pas obligatoire et la valeur par défaut est `all` si aucun nom n'est spécifié. Le tableau présenté ci-après décrit les conventions d'attribution de noms applicables aux chemins d'accès aux disques.

Remarque – l'utilisation du nom de chemin de disque global est fortement recommandé, car il est cohérent dans l'intégralité du cluster. Le nom de chemin de disque UNIX ne l'est pas, le chemin de disque UNIX d'un disque peut varier d'un nœud de cluster à l'autre. Il peut par exemple être `c1t0d0` sur un nœud, et `c2t0d0` sur un autre. Si vous utilisez des noms de chemins de disques UNIX, utilisez la commande `scdidadm -L` pour les mapper sur les noms de chemins de disques globaux avant d'utiliser des commandes de CCD. Pour plus d'informations, voir `scdidadm`(1M).

TABLEAU 3–3 Exemples de noms de chemins de disques

Type de nom	Exemple de nom de chemin de disque	Description
Chemin de disque global	`schost-1:/dev/did/dsk/d1`	Chemin de disque `d1` sur le nœud `schost-1`
`all:d1`	Chemin de disque `d1` sur tous les nœuds du cluster	
Chemin de disque UNIX	`schost-1:/dev/rdsk/c0t0d0s0`	Chemin de disque `c0t0d0s0` sur le nœud `schost-1`
`schost-1:all`	Tous les chemins de disques sur le nœud `schost-1`	
Tous les chemins de disques	`all:all`	Tous les chemins de disques sur tous les nœuds du cluster

Utilisation de SunPlex Manager pour contrôler les chemins d'accès aux disques

SunPlex Manager vous permet de réaliser les tâches d'administration CCD de base suivantes :

- contrôler un chemin de disque ;
- désactiver le contrôle d'un chemin de disque ;
- afficher le statut de tous les chemins de disques dans le cluster.

Consultez l'aide en ligne de SunPlex Manager pour de plus amples informations sur la procédure d'administration des chemins de disques avec SunPlex Manager.

Quorum et périphériques de quorum

Cette section comporte les rubriques suivantes :

Remarque – Pour connaître la liste des périphériques pris en charge comme périphériques de quorum par le logiciel Sun Cluster, contactez votre fournisseur de services Sun.

Comme les nœuds de cluster partagent des données et des ressources, un cluster ne doit jamais être divisé entre des partitions séparées actives simultanément, au risque d'altérer les données. Le moniteur d'appartenance au cluster (MAC) et l'algorithme de quorum garantissent l'exécution d'une instance au plus du même cluster à un moment donné, même si l'interconnexion de cluster est partitionnée.

Pour plus d'informations sur le quorum et le moniteur d'appartenance au cluster, reportez-vous à "Appartenance au cluster" du *Présentation de Sun Cluster pour SE Solaris*.

Deux types de problèmes proviennent des partitions de cluster :

- Split brain
- Amnésie

Le Split brain se produit lorsque l'interconnexion de cluster entre les nœuds est perdue et que le cluster se retrouve partitionné en sous-clusters. Chaque partition « croit » être la seule partition, car les nœuds d'une partition ne peuvent pas communiquer avec ceux de l'autre partition.

Une amnésie survient lorsque le cluster redémarre après un arrêt avec des données de configuration antérieures au moment de l'arrêt. Ce problème peut se produire si vous démarrez le cluster sur un nœud qui n'était pas sur la partition de cluster qui fonctionnait en dernier.

Le logiciel Sun Cluster évite les split brain et amnésies en :

- affectant un vote à chaque nœud ;

- mandatant une majorité de votes en faveur d'un cluster opérationnel.

Une partition dotée d'une majorité de votes obtient un *quorum* et est autorisée à fonctionner. Ce mécanisme de vote majoritaire évite les phénomènes de split brain et d'amnésie lorsque plus de deux nœuds sont configurés au sein d'un cluster. Toutefois, le décompte des votes des nœuds seul n'est pas suffisant lorsque plus de deux nœuds sont configurés dans le cluster. Dans un cluster à deux noeuds, la majorité est deux. Si un tel cluster à deux nœuds est partitionné, un vote externe est nécessaire pour que l'une des partitions obtienne le quorum. Ce vote externe est alors fourni par un *périphérique de quorum*.

À propos des décomptes de votes de quorum

Utilisez la commande `scstat -q` pour déterminer les informations suivantes :

- Nombre total de votes configurés
- Votes présents actuels
- Votes requis pour le quorum

Pour plus d'informations sur cette commande, reportez-vous à `scstat`(1M).

Les nœuds comme les périphériques de quorum apportent leurs votes au cluster afin de constituer le quorum.

Un nœud apporte ses votes en fonction de son état :

- Un nœud apporte *un* vote lorsqu'il est initialisé et devient membre du cluster.
- Un nœud apporte *zéro* vote lors de son installation.
- Un nœud apporte *zéro* vote lorsqu'un administrateur système place le nœud à l'état de maintenance.

Les périphériques de quorum apportent des votes en fonction du nombre de votes qui leur sont connectés. Lorsque vous configurez un périphérique de quorum, le logiciel Sun Cluster lui affecte un décompte de votes de *N*-1, où *N* correspond au nombre de votes connectés au périphérique de quorum. Par exemple, un périphérique de quorum connecté à deux noeuds avec un nombre de votes différent de zéro possède un nombre de quorums égal à un (deux moins un).

Un périphérique de quorum apporte son vote si *l'une* des deux conditions suivantes est vérifiée :

- L'un des nœuds au moins auquel le périphérique de quorum est connecté est membre du cluster.
- L'un des nœuds au moins auquel le périphérique de quorum est connecté est en cours d'initialisation et ce nœud était membre de la dernière partition de cluster ayant possédé le périphérique de quorum.

Vous pouvez configurer les périphériques de quorum pendant l'installation du cluster ou ultérieurement, à l'aide des procédures décrites dans le Chapitre 5, "Administration du quorum" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

À propos de la séparation en cas d'échec

Une panne entraînant la partition du cluster (appelée *split brain*) est un des problèmes majeurs que peut rencontrer un cluster. Lorsque ce phénomène se produit, les nœuds ne peuvent pas tous communiquer, ainsi, des nœuds individuels ou des sous-ensembles de nœuds risquent de tenter de former des clusters individuels ou des sous-ensembles. Chaque partition ou sous-ensemble peut « croire » être seul propriétaire des disques multihôtes et à en posséder l'accès. Si plusieurs nœuds tentent d'écrire sur les disques, les données peuvent être endommagées.

La séparation en cas d'échec limite l'accès des nœuds aux périphériques multihôtes en les empêchant physiquement d'accéder aux disques. Lorsqu'un noeud quitte le cluster (parce qu'il a échoué ou a été partitionné), la séparation en cas d'échec assure qu'il ne peut plus accéder aux disques. Seuls les membres actuels des nœuds ont accès aux disques, garantissant ainsi l'intégrité des données.

Les services des périphériques de disques intègrent des capacités de basculement pour les services utilisant des périphériques multihôtes. Si un membre de cluster, actuellement membre principal (propriétaire) du groupe de périphériques de disques échoue ou est inaccessible, un nouveau nœud principal est choisi. Ce nouveau nœud permet d'accéder au groupe de périphériques de disques pour poursuivre en n'étant soumis qu'à des interruptions mineures. Pendant ce processus, l'ancien nœud principal doit perdre l'accès aux périphériques pour que le nouveau nœud principal puisse être démarré. Toutefois, lorsqu'un membre se détache du cluster et n'est plus joignable, le cluster ne peut pas informer ce nœud de libérer les périphériques dont il était le nœud principal. Il faut donc trouver un moyen de permettre aux membres survivants d'accéder aux périphériques globaux et d'en prendre le contrôle à la place des membres défectueux.

Le système Sun Cluster utilise le mode de réservation des disques SCSI pour implémenter la séparation en cas d'échec. Grâce au système de réservation SCSI, les nœuds défectueux sont « isolés » à l'extérieur des périphériques multihôtes, pour les empêcher d'accéder à ces disques.

Les réservations de disques SCSI-2 prennent en charge une forme de réservations qui octroie l'accès à tous les nœuds liés au disque (en cas d'absence de réservation) ou à un seul nœud (celui qui détient la réservation).

Lorsqu'un membre détecte qu'un autre noeud ne communique plus sur l'interconnexion du cluster, il lance une procédure de séparation en cas d'échec pour empêcher le noeud d'accéder aux disques partagés. Lors de la séparation en cas d'échec, le nœud séparé panique et un message de « conflit de réservation » s'affiche sur la console.

Si un nœud est détecté comme n'étant plus membre de cluster, une réservation SCSI est déclenchée sur tous les disques partagés entre ce nœud et les autres. Le nœud séparé risque de ne pas « savoir » qu'il est en cours de séparation et s'il essaie d'accéder à l'un des disques partagés, il détectera la réservation et s'emballera.

Mécanisme failfast pour la séparation en cas d'échec

On appelle *failfast* le mécanisme par le biais duquel la structure du cluster s'assure qu'un nœud défectueux ne peut pas redémarrer et commencer à écrire dans le stockage partagé.

Les nœuds membres du cluster activent en permanence un ioctl spécifique, MHIOCENFAILFAST, pour les disques auxquels ils ont accès, notamment les disques de quorum. Cet ioctl est une directive adressée au pilote de disque. Il permet à un nœud de paniquer s'il ne peut pas accéder à un disque réservé par un autre nœud.

L'ioctl MHIOCENFAILFAST provoque la vérification par le pilote de l'erreur renvoyée par toutes les lectures et écritures d'un nœud sur le disque s'il s'agit du code d'erreur `Reservation_Conflict` . À l'arrière-plan, l'ioctl teste régulièrement le disque pour rechercher le code d'erreur `Reservation_Conflict`. Les chemins de flux de contrôle au premier plan et à l'arrière-plan paniquent si le code `Reservation_Conflict` est renvoyé.

Pour les disques SCSI-2, les réservations ne sont pas persistantes ; elles ne survivent pas aux réinitialisations des nœuds. Pour les disques SCSI-3 dotés de la fonction PGR (réservation de groupe persistante), les informations de réservation sont stockées sur le disque et persistent après réinitialisation des noeuds. Le mécanisme failfast fonctionne de la même façon, que vous ayez des disques SCSI-2 ou SCSI-3.

Si un nœud perd la connectivité aux autres nœuds du cluster et qu'il ne fait pas partie d'une partition pouvant atteindre un quorum, il est supprimé de force du cluster par un autre nœud. Un autre noeud faisant partie de la partition pouvant atteindre le quorum effectue des réservations sur les disques partagés. Si le nœud qui ne possède pas de quorum essaie d'accéder aux disques partagés, il reçoit un conflit de réservation et panique du fait du mécanisme failfast.

Après la panique, le noeud peut soit se réinitialiser et tenter de rejoindre le cluster, soit rester sur l'invite de la PROM OpenBoot™ (OBP) si le cluster est constitué de systèmes SPARC. L'action entreprise est déterminée par la définition du paramètre `auto-boot?`. Vous pouvez définir `auto-boot?` sur `eeprom`(1M), à l'invite `ok` de la PROM OpenBoot dans un cluster SPARC. Vous pouvez également configurer ce paramètre avec l'utilitaire SCSI que vous pouvez exécuter après le démarrage du BIOS dans un cluster x86.

À propos des configurations de quorum

La liste suivante présente des faits concernant les configurations de quorum :

- Les périphériques de quorum peuvent contenir des données utilisateur.
- Dans une configuration N+1, où les périphériques de quorum *N* sont chacun connectés à l'un des `1` par le biais des nœuds *N* et du nœud `N+1`, le cluster survit à la mort de tous les `1` par le biais des nœuds *N* ou de l'un des nœuds *N*/**2**. Cette disponibilité suppose que le périphérique de quorum fonctionne correctement.
- Dans une configuration *N*-nœud où un seul périphérique de quorum se connecte à tous les nœuds, le cluster peut survivre à la mort d'un des nœuds *N*-**1**. Cette disponibilité suppose que le périphérique de quorum fonctionne correctement.
- Dans une configuration à *N* nœuds où un périphérique de quorum unique se connecte à tous les nœuds, le cluster peut survivre à la défaillance du périphérique de quorum si tous les nœuds du cluster sont disponibles.

Pour consulter les exemples de configurations de quorum à éviter, voir "Configurations de quorum déconseillées " à la page 63. Pour consulter les exemples de configurations de quorum conseillées, voir "Configurations de quorum conseillées " à la page 61.

Respect des contraintes des périphériques de quorum

Vous devez respecter les conditions requises suivantes. Si vous ignorez ces conditions, vous risquez de compromettre la disponibilité du cluster.

- Vérifiez que le logiciel Sun Cluster prend en charge votre périphérique comme périphérique de quorum.

Remarque – Pour connaître la liste des périphériques pris en charge comme périphériques de quorum par le logiciel Sun Cluster, contactez votre fournisseur de services Sun.

  Le logiciel Sun Cluster prend en charge deux types de périphérique de quorum :

  - Les disques partagés multihôtes prenant en charge les réservations PGR SCSI-3.
  - Les disques partagés à deux hôtes prenant en charge les réservations SCSI-2.
- Dans une configuration à deux nœuds, vous devez configurer au moins un périphérique de quorum pour garantir qu'un nœud unique puisse continuer en cas de défaillance de l'autre nœud. Reportez-vous à la Figure 3–2.

Pour consulter les exemples de configurations de quorum à éviter, reportez-vous à "Configurations de quorum déconseillées " à la page 63. Pour consulter les exemples de configurations de quorum conseillées, voir "Configurations de quorum conseillées " à la page 61.

Respect des recommandations portant sur les périphériques de quorum

Utilisez les informations suivantes pour évaluer la configuration de quorum la mieux adaptée à votre topologie :

- Disposez-vous d'un périphérique pouvant être connecté à tous les nœuds du cluster ?
  - Si oui, configurez ce périphérique comme périphérique de quorum unique. Vous n'êtes *pas* tenu de configurer un autre périphérique de quorum, car votre configuration est optimale.

Attention – Si vous ne tenez pas compte de ce conseil et ajoutez un autre périphérique de quorum, ce dernier va réduire la disponibilité de votre cluster.

  - Sinon, configurez votre ou vos périphériques sur deux ports.
- Assurez-vous que le nombre total de votes apportés par les périphériques de quorum est strictement inférieur au nombre de votes apportés par les nœuds. Sinon, vos nœuds ne pourraient pas constituer un cluster en cas d'indisponibilité de tous les disques, et ce même si tous les nœuds fonctionnent.

Remarque – Dans certains environnements, vous pouvez réduire la disponibilité globale du cluster en fonction de vos besoins. Dans ces cas-là, vous pouvez ignorer les techniques habituelles préconisées. Soyez conscient toutefois que le non-respect de cette technique conseillée réduit la disponibilité globale. Par exemple, dans la configuration indiquée dans "Configurations de quorum atypiques " à la page 63, le cluster est moins disponible : Les votes du quorum excèdent les votes des nœuds. Si l'accès au stockage partagé entre `Nodes A` et `Node B` est perdu, tout le cluster échoue.

Pour consulter l'exception à cette recommandation, voir "Configurations de quorum atypiques " à la page 63.

- Spécifiez un périphérique de quorum entre chaque paire de nœuds partageant l'accès au périphérique de stockage. Cette configuration de quorum accélère le processus de séparation en cas d'échec. Voir "Quorum dans les configurations supérieures à deux nœuds " à la page 61.

- En général, si l'ajout d'un périphérique de quorum permet de rendre pair le nombre de votes, la disponibilité globale du cluster diminue.
- Les périphériques de quorum ralentissement légèrement les reconfigurations suite à l'arrivée ou à la mort d'un nœud. Limitez donc l'ajout de périphériques de quorum au strict minimum.

Pour consulter les exemples de configurations de quorum à éviter, reportez-vous à "Configurations de quorum déconseillées " à la page 63. Pour consulter les exemples de configurations de quorum conseillées, reportez-vous à "Configurations de quorum conseillées " à la page 61.

Configurations de quorum conseillées

Cette section indique des exemples de configurations de quorum conseillées. Pour consulter les exemples de configurations de quorum à éviter, reportez-vous à "Configurations de quorum déconseillées " à la page 63.

Quorum dans les configurations à deux nœuds

Deux votes de quorum sont nécessaires pour la formation d'un cluster à deux nœuds. Ces deux votes peuvent dériver des deux nœuds de cluster ou d'un seul nœud et d'un périphérique de quorum.

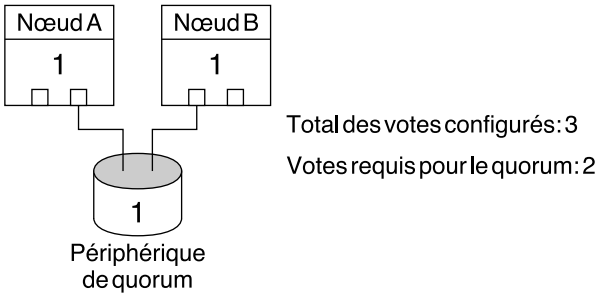
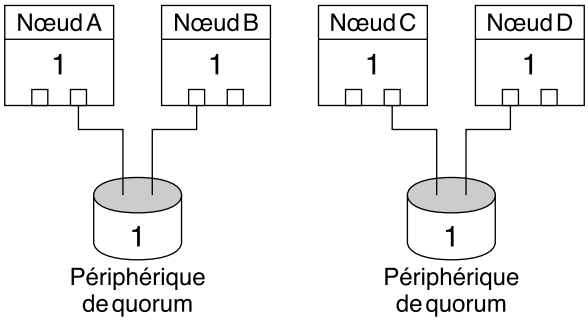


FIGURE 3–2 Configuration à deux nœuds

Quorum dans les configurations supérieures à deux nœuds

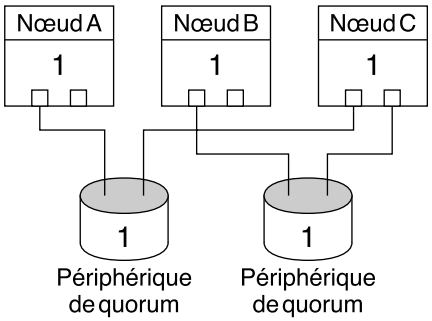
Vous pouvez configurer un cluster supérieur à deux nœuds sans périphérique de quorum. Cependant, si vous procédez ainsi, vous ne pouvez pas démarrer le cluster sans une majorité de nœuds dans le cluster.



Total des votes configurés : 6

Votes requis pour le quorum : 4

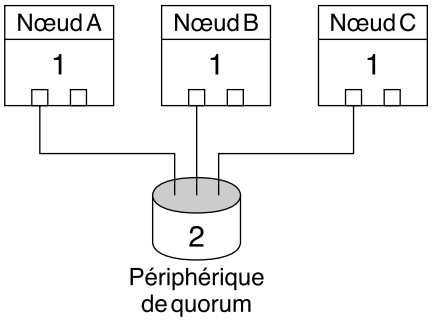
Dans cette configuration, chaque paire doit être disponible afin que n'importe laquelle d'entre elles puisse survivre.



Total des votes configurés : 5

Votes requis pour le quorum : 3

Dans cette configuration, les applications sont généralement paramétrées afin d'être exécutées sur le nœud A et le nœud B et d'utiliser le nœud C comme hot spare.



Total des votes configurés : 5

Votes requis pour le quorum : 3

Dans cette configuration, l'association d'un ou de plusieurs nœuds et du périphérique de quorum peut constituer un cluster.

Configurations de quorum atypiques

La Figure 3–3 suppose que vous exécutez des applications stratégiques (une base de données Oracle, par exemple) sur `Node A` et `Node B`. Si le `nœud A` et le `nœud B` sont indisponibles et ne peuvent accéder aux données partagées, vous pouvez souhaiter la mise hors service totale du cluster. Sinon, cette configuration n'est pas véritablement optimale car elle n'offre pas la meilleure disponibilité.

Pour plus d'informations sur la recommandation à laquelle cette exception est liée, voir "Respect des recommandations portant sur les périphériques de quorum " à la page 60.

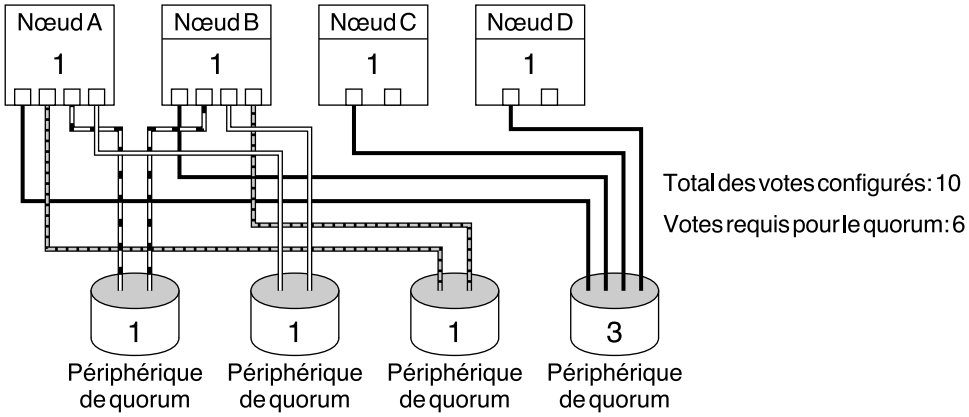
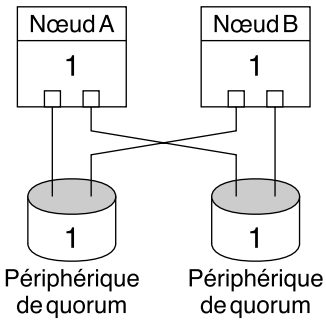


FIGURE 3–3 Configuration atypique

Configurations de quorum déconseillées

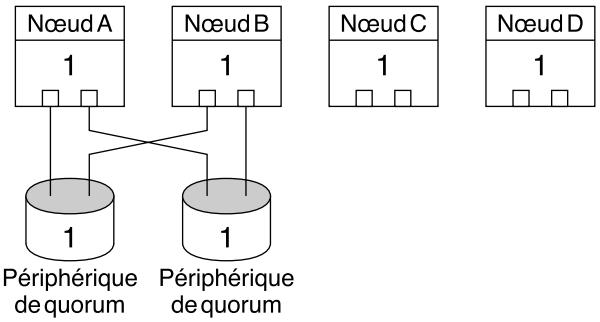
Cette section donne des exemples de configurations de quorum à éviter. Pour consulter les exemples de configurations de quorum conseillées, reportez-vous à "Configurations de quorum conseillées " à la page 61.



Total des votes configurés : 4

Votes requis pour le quorum : 3

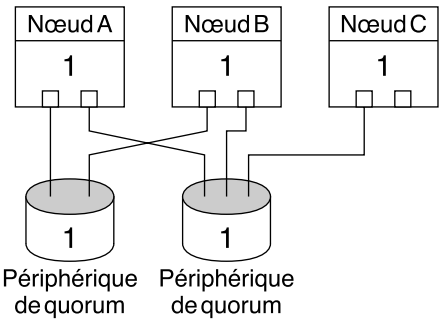
Cette configuration est contraire aux recommandations selon lesquelles le nombre de votes de périphérique de quorum doit être strictement inférieur au nombre de votes des nœuds.



Total des votes configurés : 6

Votes requis pour le quorum : 4

Cette configuration est contraire aux recommandations selon lesquelles vous ne devez pas ajouter de périphériques de quorum de sorte que le nombre total de votes soit un nombre pair.



Total des votes configurés : 5

Votes requis pour le quorum : 3

Cette configuration est contraire aux recommandations selon lesquelles le nombre de votes de périphérique de quorum doit être strictement inférieur au nombre de votes des nœuds.

Services de données

Le terme *service de données* décrit une application telle qu'Oracle ou Sun Java System Web Server, configurée pour fonctionner sur un cluster plutôt que sur un seul serveur. Un service de données se compose d'une application, de fichiers de configuration Sun Cluster spécialisés et de méthodes de gestion Sun Cluster contrôlant les actions suivantes de l'application :

- Début
- Arrêter
- le contrôle et la prise de mesures correctives.

Pour plus d'informations sur les types de services de données, voir "Services de données" du *Présentation de Sun Cluster pour SE Solaris*.

La Figure 3–4 compare une application s'exécutant sur un seul serveur d'applications (modèle de serveur unique) à la même application s'exécutant sur un cluster (modèle de serveur clusterisé). La seule différence entre les deux configurations réside dans le fait que l'application clusterisée peut s'exécuter plus rapidement et avec un niveau supérieur de disponibilité.

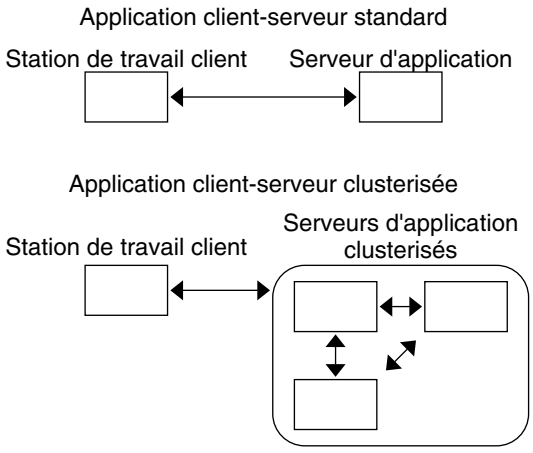


FIGURE 3–4 Configuration client-serveur standard et configuration client-serveur clusterisée

Dans le modèle de serveur unique, vous configurez l'application pour accéder au serveur par le biais d'une interface de réseau public particulier (un nom d'hôte). Le nom d'hôte est associé à ce serveur physique.

Dans le modèle de serveur clusterisé, l'interface de réseau public est un *nom d'hôte logique* ou une *adresse partagée*. Le terme *ressources réseau* se rapporte à la fois aux noms d'hôtes logiques et aux adresses partagées.

Certains services de données nécessitent que vous indiquiez des noms d'hôtes logiques ou des adresses partagées comme interfaces réseau. Les noms d'hôtes logiques et les adresses partagées ne sont pas interchangeables. D'autres services de données vous permettent de spécifier les noms d'hôtes logiques comme les adresses partagées. Pour plus d'informations sur le type d'interface que vous devez spécifier, reportez-vous à l'installation et à la configuration de chaque service de données.

Une ressource réseau n'est pas associée à un serveur physique particulier. Elle peut migrer entre des serveurs physiques.

Une ressource réseau est initialement associée à un nœud, le *nœud principal*. En cas de panne du nœud principal, la ressource réseau et la ressource d'application basculent sur un autre nœud du cluster (un nœud secondaire). Lorsque la ressource réseau bascule, la ressource d'application continue de fonctionner sur le nœud secondaire après un bref laps de temps.

La Figure 3–5 compare le modèle de serveur unique au modèle de serveur clusterisé. Notez que dans le modèle de serveur clusterisé, une ressource réseau (nom d'hôte logique dans cet exemple) peut se déplacer entre plusieurs nœuds du cluster. L'application est configurée pour utiliser ce nom d'hôte logique à la place d'un nom d'hôte associé à un serveur particulier.

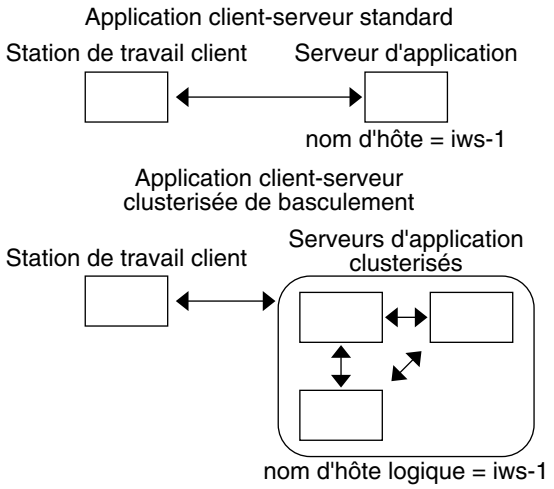


FIGURE 3–5 Nom d'hôte fixe et nom d'hôte logique

Une adresse partagée est également associée initialement à un nœud. Ce nœud s'appelle le nœud d'interface globale. Une adresse partagée (également appelée *interface globale*) est utilisée comme interface de réseau unique sur le cluster.

Le modèle de nom d'hôte logique et le modèle de service évolutif diffèrent, car dans ce dernier, chaque nœud dispose également de l'adresse partagée, configurée de manière active dans son interface de loopback. Cette configuration permet à plusieurs instances

d'un service de données de s'activer simultanément sur plusieurs nœuds. Le terme « service évolutif » signifie que vous pouvez augmenter la puissance de calcul de l'application en ajoutant des nœuds de cluster supplémentaires afin d'améliorer les performances.

En cas d'échec du nœud d'interface globale, l'adresse partagée peut être démarrée sur un autre nœud où tourne aussi une instance de l'application (faisant ainsi de cet autre nœud le nouveau nœud d'interface globale). L'adresse partagée peut également basculer sur un autre nœud du cluster n'ayant pas exécuté l'application auparavant.

La Figure 3–6 compare la configuration à serveur unique à la configuration de service évolutif clusterisé. Notez que dans la configuration service évolutif, l'adresse partagée est présente sur tous les nœuds. De la même façon qu'un nom d'hôte logique est utilisé dans un service de données à basculement, l'application est configurée pour utiliser cette adresse partagée à la place d'un nom d'hôte associé à un serveur particulier.

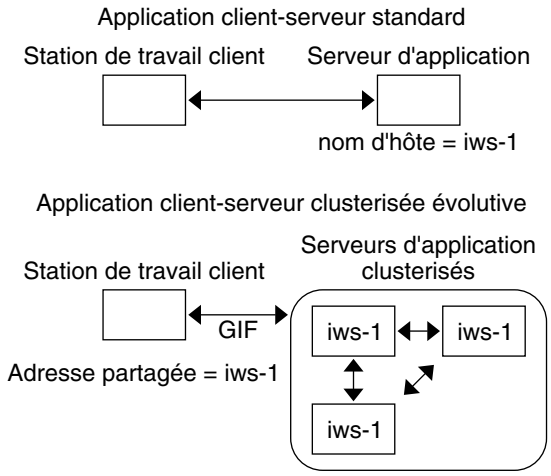


FIGURE 3–6 Nom d'hôte fixe et adresse partagée

Méthodes des services de données

Le logiciel Sun Cluster intègre un ensemble de méthodes de gestion de services. Ces méthodes s'exécutent sous le contrôle du produit Gestionnaire du groupe de ressources (RGM), qui les utilise pour démarrer, arrêter et contrôler l'application sur les nœuds du cluster. Ces méthodes, avec les logiciels de cluster et les périphériques multihôtes, permettent aux applications de devenir des services de données évolutifs ou de basculement.

Le gestionnaire du groupe de ressources (RGM) gère aussi les ressources du cluster, notamment les instances d'une application et les ressources réseau (noms d'hôtes logiques et adresses partagées).

Outre les méthodes du logiciel Sun Cluster, le système Sun Cluster fournit également une API et plusieurs outils de développement de services de données. Ces outils permettent aux développeurs d'applications de développer les méthodes de services de données nécessaires à l'exécution d'autres applications en tant que services de données hautement disponibles avec le logiciel Sun Cluster.

Services de données de basculement

Si le nœud sur lequel le service de données fonctionne (nœud principal) échoue, le service est déplacé vers un autre nœud opérationnel sans intervention de l'utilisateur. Les services de basculement utilisent un *groupe de ressources de basculement* contenant des ressources d'instances d'application et des ressources réseau (*noms d'hôtes logiques*). Les noms d'hôtes logiques sont des adresses IP pouvant être configurées sur un nœud, puis automatiquement retirées pour être configurées sur un autre nœud.

Avec les services de données, les instances d'application ne tournent que sur un seul nœud. Si le détecteur de pannes détecte une panne, il essaie de redémarrer l'instance sur un même nœud ou de la lancer sur un autre nœud (basculement). Le résultat dépend de la configuration du service de données.

Services de données évolutifs

Le service de données évolutif permet d'avoir des instances fonctionnant sur plusieurs nœuds. Les services évolutifs utilisent les deux groupes de ressources suivants :

- un *groupe de ressources évolutif*, qui contient les ressources d'applications ;
- un groupe de ressources de basculement, qui contient les ressources réseau (*adresses partagées*) dont dépend le service évolutif.

Le groupe de ressources évolutif peut être connecté à plusieurs nœuds, permettant ainsi à plusieurs instances du service de fonctionner en même temps. Le groupe de ressources de basculement hébergeant les adresses partagées ne peut être connecté qu'à un seul nœud à la fois. Tous les nœuds hébergeant un service évolutif utilisent la même adresse partagée pour héberger le service.

Les demandes de service entrent dans le cluster par le biais de l'interface de réseau unique (interface globale). Ces demandes sont ensuite distribuées aux nœuds, en fonction d'un des algorithmes prédéfinis définis par la *règle d'équilibrage de charge*. Le cluster peut utiliser cette règle pour équilibrer la charge de service entre plusieurs nœuds. Plusieurs interfaces globales peuvent exister sur différents nœuds qui hébergent d'autres adresses partagées.

Avec les services évolutifs, les instances d'application tournent sur plusieurs nœuds en même temps. Si le nœud hébergeant l'interface globale échoue, celle-ci bascule sur un autre nœud. Si une instance d'application en cours d'exécution échoue, elle essaie de redémarrer sur le même nœud.

S'il lui est impossible de redémarrer sur le même nœud et qu'un autre nœud non utilisé est configuré pour exécuter le service, celui-ci bascule sur le nœud non utilisé. Faute de quoi, le service continue de s'exécuter sur les nœuds restants, ce qui peut provoquer une dégradation de ses capacités de traitement.

Remarque – l'état TCP de chaque instance d'application est conservé sur le nœud contenant l'instance et non sur le nœud d'interface globale. Ainsi, l'échec du nœud d'interface globale n'a pas d'incidence sur la connexion.

La Figure 3–7 affiche un exemple de groupe de ressources de basculement et de groupes de ressources évolutifs ainsi que les dépendances qui existent entre eux pour les services évolutifs. Cet exemple présente trois groupes de ressources. Le groupe de ressources de basculement contient les ressources d'applications des serveurs DNS à haute disponibilité et les ressources réseau utilisées à la fois par les serveurs DNS à haute disponibilité et les serveurs Web Apache à haute disponibilité (disponibles uniquement sur les clusters SPARC). Les groupes de ressources évolutifs ne contiennent que les instances d'application du serveur Web Apache. Notez que des dépendances de groupes de ressources existent entre les groupes de ressources évolutifs et de basculement (lignes pleines). Par ailleurs, toutes les ressources de l'application Apache dépendent de la ressource réseau `schost-2`, qui est une adresse partagée (pointillés).

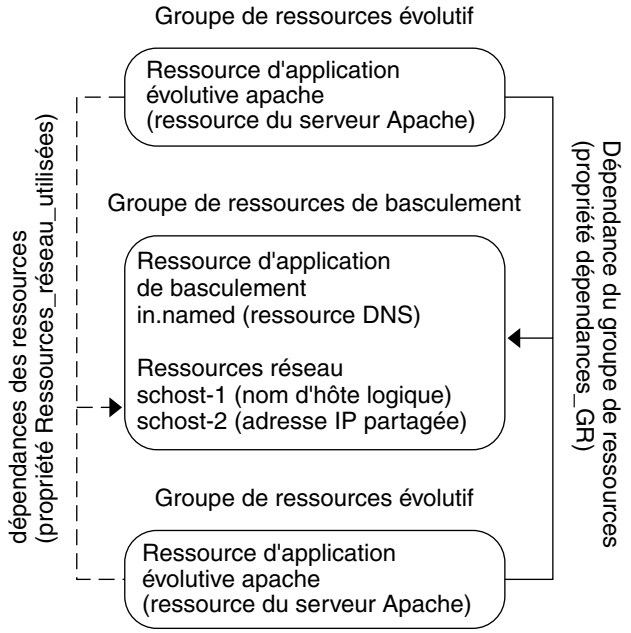


FIGURE 3–7 SPARC : exemple de groupe de ressources évolutif et de basculement

Règles d'équilibrage de la charge

L'équilibrage de la charge permet d'améliorer les performances du service évolutif, tant en matière de temps de réponse que de rendement. Il existe deux classes de services de données évolutifs.

- Pur
- Sticky

Un service *pur* est à même de faire répondre l'une de ses instances aux demandes d'un client. Un service *sticky* est à même de demander à un client d'envoyer des demandes à la même instance. Ces dernières ne sont pas redirigées vers d'autres instances.

Un service pur utilise une règle d'équilibrage de la charge pondérée. Sous cette règle, les requêtes du client sont réparties de manière uniforme entre les instances de serveur du cluster. Supposons par exemple que chaque nœud d'un cluster à trois nœuds ait un poids de 1. Chaque nœud traitera 1/3 des demandes d'un client pour le compte de ce service. L'administrateur peut à tout moment changer les poids par le biais de l'interface de contrôle `scrgadm(1M)` ou de l'interface utilisateur graphique de Gestionnaire SunPlex.

Il existe deux types de services sticky : *sticky ordinaire* et *sticky joker*. Les services sticky permettent à plusieurs sessions d'application simultanées utilisant plusieurs connexions TCP de partager la mémoire d'état (état de la session d'application).

Les services sticky ordinaires permettent à un client de partager l'état entre plusieurs connexions TCP simultanées. Le client est dit « sticky » envers l'instance du serveur écoutant sur un port unique. Le client a la garantie que toutes ses requêtes vont vers la même instance de serveur, sous réserve que cette instance soit active et accessible et que la règle d'équilibrage de la charge ne soit pas modifiée alors que le service est en ligne.

Par exemple, un navigateur Web du client se connecte à une adresse IP partagée sur le port 80 à l'aide de trois connexions TCP différentes. Toutefois, les connexions échangent, au niveau du service, les informations de session mises en cache.

Une généralisation d'une règle sticky s'étend à plusieurs services évolutifs qui échangent des informations de session à l'arrière-plan et au niveau de la même instance. Le cas échéant, le client est dit « sticky » envers plusieurs instances de serveur sur le même nœud écoutant des ports différents.

Prenons l'exemple d'un client sur un site d'e-commerce, qui remplit son panier d'achat en utilisant le protocole HTTP sur le port 80. Le client bascule ensuite sur le protocole SSL sur le port 443 pour envoyer des données sécurisées afin de payer par carte de crédit les articles contenus dans le panier.

Les services sticky *joker* utilisent des numéros de port assignés de manière dynamique, mais attendent toujours des demandes du client qu'elles se dirigent vers le même nœud. Le client est dit « sticky joker » sur les ports qui ont la même adresse IP.

Un bon exemple de cette règle est le protocole FTP en mode passif. Par exemple, un client se connecte à un serveur FTP sur le port 21. Le serveur demande ensuite au client de se reconnecter à un serveur de ports d'écoute dans la plage des ports dynamiques. Toutes les demandes relatives à cette adresse IP sont transférées sur le même nœud par le biais duquel le serveur a informé le client des informations de contrôle.

La règle d'équilibrage de charge pondérée est appliquée par défaut pour chacune de ces règles sticky. Ainsi la requête initiale d'un client est dirigée vers l'instance imposée par l'équilibreur de charge. Une fois que le client a établi une affinité pour le nœud sur lequel l'instance s'exécute, les demandes futures sont dirigées sous conditions vers cette instance. Le nœud doit être accessible et la règle d'équilibrage de charge ne doit pas avoir été modifiée.

D'autres informations sur les règles d'équilibrage de charge spécifiques sont indiquées ci-dessous.

- Pondérée. La charge est répartie entre les différents nœuds en fonction de valeurs spécifiées. Cette règle est définie via l'utilisation de la valeur `LB_WEIGHTED` pour la propriété `Load_balancing_weights`. Si la charge d'un nœud n'est pas explicitement définie, elle sera par défaut égale à 1.

  La règle pondérée redirige un certain pourcentage du trafic provenant des clients vers un nœud particulier. Soit X=le poids et A=le total des poids de tous les nœuds actifs : un nœud actif peut s'attendre qu'environ X/A de toutes les nouvelles connexions soient dirigées vers lui. Toutefois, le nombre total de connexions doit être suffisamment important. Cette règle ne s'applique pas à des requêtes individuelles.

  Notez qu'elle ne fonctionne pas à tour de rôle. Si on utilisait une règle à tour de rôle (circulaire), chaque demande d'un client irait toujours vers un nœud différent. Par exemple, la première demande irait vers le nœud 1, la seconde, vers le nœud 2, etc.
- Sticky. Avec cette règle, le jeu de ports est connu au moment de la configuration des ressources d'application. Cette règle est définie via l'utilisation de la valeur `LB_STICKY` pour la propriété de ressource `Load_balancing_policy`.
- Sticky joker. Cette règle est un sur-ensemble de la règle « sticky » ordinaire. S'il s'agit d'un service évolutif identifié par l'adresse IP, les ports sont assignés par le serveur (et ne sont pas connus à l'avance). Les ports peuvent varier. Cette règle est définie via l'utilisation de la valeur `LB_STICKY_WILD` pour la propriété de ressource `Load_balancing_policy`.

Paramètres de rétablissement

Les groupes de ressources basculent d'un nœud sur un autre. Le cas échéant, le nœud secondaire d'origine devient le nouveau nœud principal. Les paramètres de rétablissement spécifient les actions qui se produisent lorsque le nœud principal d'origine est de nouveau en ligne. Vous pouvez soit autoriser le nœud principal d'origine à reprendre son rôle (rétablissement), soit permettre au nœud principal actuel de rester. Spécifiez l'option souhaitée en utilisant le paramètre de propriété du groupe de ressources `Failback`.

Si le nœud d'origine qui héberge le groupe de ressources est défectueux et redémarre sans arrêt, ce rétablissement peut réduire la disponibilité du groupe de ressources.

Détecteurs de pannes des services de données

Chaque service de données Sun Cluster intègre un détecteur de pannes le sondant périodiquement pour vérifier son état. Un détecteur de pannes vérifie que le(s) démon(s) de l'application fonctionnent et que les clients sont servis. En fonction des informations renvoyées par les sondes, des actions prédéfinies telles que le redémarrage des démons ou le déclenchement d'un basculement peuvent être initiées.

Développement de nouveaux services de données

Sun propose des fichiers de configuration et des modèles de méthodes de gestion vous permettant de faire fonctionner différentes applications comme service évolutif ou de basculement au sein d'un cluster. Si Sun ne propose pas l'application que vous voulez exécuter en tant que service de basculement ou évolutif, vous avez une autre solution. Utilisez une API Sun Cluster ou l'API DSET pour configurer l'application à exécuter en tant que service de basculement ou service évolutif. Cependant, les applications ne peuvent pas toutes devenir un service évolutif.

Caractéristiques des services évolutifs

Un ensemble de critères détermine si une application peut devenir un service évolutif. Pour déterminer si c'est le cas de la vôtre, reportez-vous à la section "Analyse du caractère approprié de l'application" du *Guide du développeur de services de données Sun Cluster pour SE Solaris*. Cet ensemble de critères est résumé ci-dessous.

- Tout d'abord, un tel service est composé d'une ou de plusieurs *instances* de serveur. Chaque instance tourne sur un nœud différent du cluster et plusieurs instances du même service ne peuvent pas fonctionner sur le même nœud.

- Si le service fournit un magasin de données logique externe, vous devez être prudent. Vous devez synchroniser les accès simultanés à ce magasin depuis plusieurs instances de serveur pour éviter de perdre des mises à jour ou de lire des données en cours de modification. Notez l'utilisation du terme « externe » pour que ce magasin ne soit pas considéré comme en mémoire. Le terme « logique » indique que ce magasin s'affiche en tant qu'entité unique, bien qu'il puisse être répliqué. En outre, à chaque fois qu'une instance de serveur met à jour ce magasin de données logique, les autres instances peuvent immédiatement « voir » cette mise à jour.

  Le système Sun Cluster propose un stockage externe de ce type par le biais de son système de fichiers de cluster et de ses partitions brutes globales. Supposons par exemple qu'un service écrive des données sur un fichier journal externe ou qu'il modifie les données en place. Si plusieurs instances de ce service s'exécutent, chacune a accès à ce journal externe, auquel elles peuvent accéder simultanément. Les instances doivent alors synchroniser leur accès pour éviter toute interférence entre elles. Le service peut utiliser le verrouillage de fichier ordinaire Solaris via `fcntl(2)` et `lockf(3C)` pour obtenir la synchronisation souhaitée.

  Un autre exemple de ce type de magasin est une base de données d'arrière-plan, notamment la base de données Oracle Real Application Clusters Guard à haute disponibilité pour les clusters SPARC ou Oracle. Ce type de serveur de base de données d'arrière-plan propose une synchronisation intégrée en utilisant des requêtes de base de données ou des transactions de mise à jour. Ainsi, plusieurs instances de serveur n'ont pas besoin d'implémenter leur propre synchronisation.

  Le serveur Sun IMAP est un exemple de service non évolutif. Le service met à jour une mémoire, mais cette mémoire est privée et lorsque les instances IMAP écrivent dans cette mémoire, elles s'écrasent mutuellement parce que les mises à jour ne sont pas synchronisées. Le serveur IMAP doit être récrit pour synchroniser les accès simultanés.
- Pour finir, notez que les instances peuvent disposer de données privées disjointes de celles des autres instances. Dans un tel cas, le service n'a pas besoin d'accès simultané synchronisé, car les données sont privées : seule l'instance en question peut les manipuler. Dans ce cas, vous devez veiller à ne pas stocker ces données privées dans le système de fichiers de cluster, car ces données seront alors globalement accessibles.

API de services de données et API de bibliothèque de développement de service de données

Pour rendre les applications hautement disponibles, le système Sun Cluster fournit les éléments suivants :

- des services de données fournis dans le cadre du système Sun Cluster ;
- une API (interface de programme d'application) de services de données ;
- une API de bibliothèque de développement pour les services de données ;

- un service de données « générique ».

Le guide *Sun Cluster Data Services Planning and Administration Guide for Solaris OS* explique comment installer et configurer les services de données du système Sun Cluster. La *collection de logiciels Sun Cluster 3.1 9/04 pour Solaris (Édition pour plate-forme SPARC)* explique comment instrumenter d'autres applications pour qu'elles aient un haut niveau de disponibilité dans la structure Sun Cluster.

Les API Sun Cluster permettent aux développeurs d'applications de développer des détecteurs de pannes et des scripts qui démarrent et arrêtent les instances de services de données. Grâce à ces outils, une application peut être implémentée comme service de basculement ou service de données évolutif. Le système Sun Cluster propose un service de données « générique ». Utilisez-le pour générer rapidement les méthodes de démarrage et d'arrêt d'application requises et pour implémenter le service de données comme service de basculement ou service évolutif.

Utilisation de l'interconnexion de cluster pour le trafic de services de données

Un cluster doit avoir de multiples connexions réseau entre les nœuds pour former une interconnexion de cluster. Le logiciel Sun Cluster utilise plusieurs interconnexions pour atteindre les objectifs suivants :

- haute disponiblité ;
- performances améliorées.

Pour le trafic interne, notamment les données des systèmes de fichiers ou celles des services évolutifs, les messages sont agrégés sur toutes les interconnexions disponibles à tour de rôle. L'interconnexion de cluster est également mise à la disposition des applications pour garantir une communication hautement disponible entre les nœuds. Par exemple, une application répartie peut avoir des composants exécutés sur différents nœuds et ayant besoin de communiquer entre eux. En utilisant l'interconnexion de cluster plutôt que le transport public, ces connexions peuvent résister à l'échec d'un lien individuel.

Pour utiliser l'interconnexion de cluster dans le cadre des communications entre les nœuds, l'application doit adopter les noms d'hôtes privés que vous avez configurés lors de l'installation de Sun Cluster. Par exemple, si le nom d'hôte privé du nœud 1 est `clusternode1-priv` , utilisez ce nom pour communiquer sur l'interconnexion de cluster vers le nœud 1 (`node 1`). Les sockets TCP ouverts à l'aide de ce nom sont dirigés vers l'interconnexion de cluster et peuvent être redirigés de manière transparente en cas de panne du réseau.

Comme vous pouvez configurer les noms d'hôtes privés pendant l'installation de Sun Cluster, l'interconnexion de cluster utilise le nom que vous avez choisi à ce moment-là. Pour identifier le nom réel, utilisez la commande `scha_cluster_get`(3HA) avec l'argument `scha_privatelink_hostname_node`.

Les communications d'applications et les communications de cluster interne sont agrégées sur toutes les interconnexions. Comme les applications partagent l'interconnexion de cluster avec le trafic de cluster interne, la bande passante disponible pour les applications dépend de celle qui est utilisée par le reste du trafic de cluster. En cas de panne, le trafic interne et le trafic d'applications s'agrègent sur toutes les interconnexions disponibles.

Une adresse fixe est également assignée à chaque nœud. Cette adresse est transférée sur le pilote `clprivnet`. L'adresse IP effectue la correspondance avec le nom d'hôte privé du nœud : `clusternode1-priv`. Pour plus d'informations sur le pilote de réseau privé de Sun Cluster, consultez la page de manuel se référant à `clprivnet`(7).

Si votre application nécessite des adresses IP cohérentes sur tous les points, configurez l'application à lier à l'adresse par nœud, tant sur le client que sur le serveur. Toutes les connexions semblent alors provenir de cette adresse et y retourner.

Ressources, groupes de ressources et types de ressource

Les services de données utilisent plusieurs types de *ressources* : Les applications telles que Sun Java System Web Server ou le serveur Web Apache utilisent des adresses réseau (noms d'hôtes logiques et adresses partagées) dont les applications dépendent. Les ressources de l'application et du réseau constituent une unité de base que gère le RGM.

Les services de données sont des types de ressource. Par exemple, Sun Cluster HA pour Oracle est le type de ressource `SUNW.oracle-server` et Sun Cluster HA pour Apache le type de ressource `SUNW.apache`.

Une ressource est une instanciation d'un *type de ressource* défini au niveau du cluster. Plusieurs types de ressource sont définis.

Les ressources réseau sont des types de ressource `SUNW.LogicalHostname` ou `SUNW.SharedAddress`. Ces deux types de ressource sont pré-enregistrés dans le logiciel Sun Cluster.

Les types de ressource `HAStorage` et `HAStoragePlus` servent à synchroniser le démarrage des ressources et des groupes de périphériques de disques dont les ressources dépendent. Avant le démarrage d'un service de données, ils garantissent la disponibilité des chemins d'accès aux points de montage d'un système de fichiers de

cluster, aux périphériques globaux et aux noms des groupes de périphériques. Pour plus d'informations, reportez-vous à « Synchronisation des démarrages entre les groupes de ressources et les groupes de périphériques de disques », du *Guide d'installation et de configuration des services de données*. Le type de ressource `HAStoragePlus` est désormais disponible dans Sun Cluster 3.0 5/02 et lui confère une autre fonctionnalité en optimisant la disponibilité des systèmes de fichiers locaux. Pour plus d'informations sur cette fonctionnalité, voir "Type de ressource `HAStoragePlus` " à la page 50.

Les ressources gérées par le RGM sont placées dans des groupes, appelés *groupes de ressources*, afin de pouvoir être gérées en tant qu'unité. Si une commutation ou un basculement est initié sur un groupe de ressources, ce dernier se transforme en unité.

Remarque – Si vous activez un groupe de ressources qui contient des ressources d'applications en ligne, l'application est lancée. La méthode de démarrage des services de données attend l'exécution de l'application avant de se fermer. La méthode permettant de définir à quel moment l'application est opérationnelle est identique à la méthode utilisée par le contrôleur de panne pour déterminer si un service de données est en train de servir des clients. Pour plus d'informations sur ce processus, voir *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

Gestionnaire du groupe de ressources (RGM)

Le gestionnaire du groupe de ressources (RGM) contrôle les services de données (applications) comme des ressources gérées par des implémentations de *type de ressource*. Ces implémentations peuvent être fournies par Sun ou créées par un développeur à l'aide d'un modèle de service de données générique, l'API de la bibliothèque de développement d'un service de données (API BDSD) ou l'API de la gestion de ressources (API GR). L'administrateur du cluster crée et gère les ressources dans des conteneurs appelés *groupes de ressources*. Le RGM arrête et démarre les groupes de ressources des nœuds sélectionnés en réponse aux modifications des membres du cluster.

Le RGM agit sur les *ressources* et les *groupes de ressources.* Les actions du RGM peuvent faire passer les ressources et les groupes de ressources de l'état en ligne à l'état hors ligne et inversement. La section "États et paramètres des ressources et des groupes de ressources " à la page 77 décrit en détail les états et les paramètres qui peuvent être appliqués aux ressources et aux groupes de ressources.

Pour plus d'informations sur le lancement de projets Solaris sous le contrôle du RGM, voir "Configuration d'un projet de services de données " à la page 78.

États et paramètres des ressources et des groupes de ressources

Un administrateur applique des paramètres statiques aux ressources et groupes de ressources. Ces paramètres ne peuvent être modifiés qu'à travers des actions d'administration. Le RGM déplace les groupes de ressources entre les « états » dynamiques décrits ci-après.

- Gérés ou non gérés : ces paramètres s'appliquent uniquement aux groupes de ressources d'un cluster. Les groupes de ressources sont gérés par le RGM. La commande `scrgadm(1M)` permet de gérer un groupe de ressources(ou de désactiver cette gestion) via le RGM. Ces paramètres ne changent pas après une reconfiguration du cluster.

  Au moment de sa création, un groupe de ressources n'est pas géré. Il doit l'être pour que ses ressources puissent devenir actives.

  Dans certains services de données, par exemple un serveur Web évolutif, des tâches doivent être effectuées avant le démarrage des ressources réseau et après leur arrêt. Ces tâches s'accomplissent par le biais des méthodes initialisation (INIT) et fin (FINI) des services de données. Les méthodes INIT ne fonctionnent que si le groupe de ressources dans lequel réside les ressources est à l'état « géré ».

  Lorsqu'un groupe de ressources passe de l'état « non géré » à « géré », toutes les méthodes INIT enregistrées pour le groupe sont exécutées sur les ressources du groupe.

  Lorsqu'un groupe de ressources passe de l'état « géré » à « non géré », toutes les méthodes FINI enregistrées sont appelées pour procéder à un nettoyage.

  Les méthodes INIT et FINI sont utilisées le plus couramment pour les ressources réseau des services évolutifs. Toutefois, vous pouvez les utiliser pour des tâches d'initialisation et de nettoyage qui ne sont pas effectuées par l'application.

- Activé ou désactivé : ces paramètres s'appliquent aux ressources d'un cluster. La commande `scrgadm(1M)` permet d'activer ou de désactiver une ressource. Ces paramètres ne changent pas après une reconfiguration du cluster.

  Une ressource est normalement activée et fonctionne dans le système.

  Si vous souhaitez supprimer la disponibilité de la ressource sur tous les nœuds de cluster, désactivez la ressource. Un ressource désactivée devient inutilisable.

- En ligne ou hors ligne : états dynamiques s'appliquant tant aux ressources qu'aux groupes de ressources.

  Ces états changent lorsque le cluster est reconfiguré lors d'un basculement ou d'une commutation. Vous pouvez également les modifier par le biais d'actions administratives. Utilisez la commande `scswitch(1M)` pour changer l'état en ligne ou hors ligne d'une ressource ou d'un groupe de ressources.

  Une ressource de basculement ou un groupe de ressources ne peuvent être connectés qu'à un seul nœud à la fois. Une ressource évolutive ou un groupe de ressources peuvent être en ligne sur certains nœuds et hors ligne sur d'autres. Durant une commutation ou un basculement, les groupes de ressources et les

ressources qu'ils contiennent sont déconnectés d'un nœud puis reconnectés à un autre nœud.

Si un groupe de ressources est hors ligne, alors toutes ses ressources le sont. Si un groupe de ressources est en ligne, alors toutes ses ressources activées le sont.

Les groupes de ressources peuvent contenir plusieurs ressources ayant des dépendances entre elles. Ces dépendances nécessitent que les ressources soient mises en ligne et hors ligne dans un ordre particulier. Le temps nécessaire aux méthodes utilisées pour connecter et déconnecter les ressources peut varier pour chaque ressource. Du fait de l'existence de dépendances entre les ressources et du caractère variable des temps de démarrage et d'arrêt, les ressources d'un même groupe peuvent avoir différents états « en ligne » et « hors ligne » durant une reconfiguration du cluster.

Propriétés des ressources et des groupes de ressources

Vous pouvez configurer les valeurs des propriétés des ressources et des groupes de ressources de vos services de données Sun Cluster. Il existe des propriétés standard communes à tous les services de données et des propriétés d'extension propres à chaque service de données. Certaines propriétés standard et propriétés d'extension sont définies par des paramètres par défaut et vous n'avez pas à les modifier. D'autres propriétés doivent être définies au moment de la création et de la configuration des ressources. La documentation de chaque service de données indique quelles propriétés de ressources peuvent être définies et comment les définir.

Les propriétés standard permettent de configurer les propriétés des ressources et groupes de ressources habituellement indépendantes de tout service de données. Pour consulter cet ensemble de propriétés standard, voir Annexe A, "Standard Properties" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

Les propriétés d'extension du RGM (gestionnaire du groupe de ressources) donne des informations sur l'emplacement des binaires d'application et des fichiers de configuration. Les propriétés d'extension peuvent être modifiées lorsque vous configurez vos services de données. Elles sont décrites dans le manuel consacré au service de données.

Configuration d'un projet de services de données

Les services de données peuvent être configurés pour être lancés sous un nom de projet Solaris lorsqu'ils sont mis en ligne à l'aide du RGM. La configuration associe une ressource ou un groupe de ressources géré par le RGM à une ID de projet Solaris.

La mise en correspondance de votre ressource ou groupe de ressources vers un ID projet vous permet d'utiliser des contrôles avancés, disponibles dans le système d'exploitation Solaris, pour gérer les charges de travail et la consommation du cluster.

Remarque – Vous pouvez effectuer cette configuration uniquement si vous exécutez la version actuelle du logiciel Sun Cluster avec au minimum Solaris 9.

L'utilisation des fonctionnalités de gestion de Solaris dans un environnement Sun Cluster vous permet de vous assurer que vos applications les plus importantes sont prioritaires lors du partage d'un nœud avec d'autres applications. Il arrive que les applications partagent un nœud si vous avez des services consolidés ou si les applications ont basculé. L'utilisation des fonctionnalités de gestion décrites dans ce document peut améliorer la disponibilité d'une application stratégique en empêchant les applications à faible priorité de consommer des ressources système, par exemple le temps CPU.

Remarque – La documentation Solaris relative à cette fonctionnalité décrit les « ressources » que sont le temps CPU, les processus, les tâches et composants semblables. La documentation Sun Cluster utilise le terme « ressources » pour décrire les entités qui sont sous le contrôle du RGM. La section suivante utilise le terme « ressource » pour désigner des entités Sun Cluster sous le contrôle du RGM. Elle utilise le terme « ressources » pour désigner le temps CPU, les processus et les tâches.

Cette section décrit la procédure de configuration des services de données pour le lancement de processus dans un `project`(4) Solaris 9 spécifié. Elle décrit également plusieurs scénarios de basculement et donne des suggestions de planification pour l'utilisation des fonctionnalités de gestion du système d'exploitation Solaris.

Pour plus d'informations sur les notions fondamentales et les procédures relatives à la fonctionnalité de gestion, voir Chapitre 1, "Network Service (Overview)" du *System Administration Guide: Network Services*.

Lorsque vous configurez des ressources et des groupes de ressources pour qu'ils utilisent les fonctionnalités de gestion de Solaris dans un cluster, suivez le processus général suivant :

1. Configurez les applications comme partie intégrante de la ressource.
2. Configurez les ressources comme partie intégrante d'un groupe de ressources.
3. Activez les ressources du groupe de ressources.
4. Activez la gestion du groupe de ressources.
5. Créez un projet Solaris pour votre groupe de ressources.
6. Configurez des propriétés standard pour associer le nom du groupe de ressources au projet que vous avez créé à l'étape 5.

7. Mettez le groupe de ressources en ligne.

Pour configurer les propriétés standard `Resource_project_name` ou `RG_project_name` afin d'associer l'ID projet Solaris à la ressource ou au groupe de ressources, utilisez l'option `-y` avec la commande `scrgadm`(1M). Définissez les valeurs des propriétés de la ressource ou du groupe de ressources. Pour consulter les définitions des propriétés, voir Annexe A, "Standard Properties" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*. Reportez-vous aux descriptions des propriétés `r_properties`(5) et `rg_properties`(5).

Le nom de projet spécifié doit figurer dans la base de données de projets (`/etc/project`) et le superutilisateur doit être configuré en tant que membre du projet nommé. Pour plus d'informations sur les notions fondamentales relatives à la base de données de projets, voir Chapitre 2, "Projects and Tasks (Overview)" du *System Administration Guide: Solaris Containers-Resource Management and Solaris Zones*. Pour consulter la description de la syntaxe d'un fichier projet, voir `project`(4).

Lorsque le RGM met les ressources ou les groupes de ressources en ligne, il lance les processus connexes sous le nom du projet.

Remarque – les utilisateurs peuvent à tout moment associer la ressource ou le groupe de ressources à un projet. Toutefois, le nom du projet n'est pas effectif tant que la ressource ou le groupe de ressources n'ont pas été mis hors ligne, puis remis en ligne à l'aide du RGM.

Le lancement des ressources ou des groupes de ressources sous le nom de projet vous permet de configurer les fonctions indiquées ci-après pour gérer les ressources système au sein du cluster.

- Comptabilité étendue : elle constitue un moyen souple d'enregistrer la consommation d'une tâche ou d'un procédé. La comptabilité étendue vous permet d'examiner l'historique de l'utilisation et d'évaluer les besoins en capacité des charges de travail futures.
- Contrôles : ce mécanisme permet d'appliquer des contraintes aux ressources système. On peut empêcher les processus, tâches et projets de surconsommer certaines ressources système.
- Fair Share Scheduling (FSS) : cette fonction offre la possibilité de contrôler l'allocation de temps CPU aux charges de travail, en fonction de leur importance. L'importance des charges de travail est définie par le nombre de parts de temps CPU attribué à chaque charge. Pour plus d'informations, reportez-vous aux pages suivantes.
  - `dispadmin`(1M)
  - `priocntl`(1)
  - `ps`(1)
  - `FSS`(7).

- Pools : cette fonction offre la possibilité d'utiliser des partitions pour les applications interactives en fonction des besoins de l'application. Les pools permettent de segmenter un serveur qui prend en charge un certain nombre d'applications différentes. Grâce à l'utilisation de pools les réponses des applications deviennent plus prévisibles.

Détermination des besoins pour la configuration de projets

Avant de configurer des services de données pour l'utilisation des contrôles de Solaris dans un environnement Sun Cluster, vous devez choisir la méthode de contrôle et de suivi des ressources au travers des commutations et des basculements. Identifiez les dépendances dans le cluster avant de configurer un nouveau projet. Par exemple, les ressources et groupes de ressources dépendent des groupes de périphériques de disques.

Utilisez les propriétés de groupe de ressources `nodelist`, `failback`, `maximum_primaries` et `desired_primaries` qui sont configurées avec `scrgadm`(1M) afin d'identifier les propriétés de la liste de nœuds de votre groupe de ressources.

- Pour consulter une brève présentation des dépendances de la liste de nœuds entre les groupes de ressources et les groupes de périphériques de disques, voir "Relationship Between Resource Groups and Disk Device Groups" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.
- Pour une description détaillée des propriétés, voir `rg_properties`(5).

Utilisez les propriétés `preferenced` et `failback` configurées avec `scrgadm`(1M) et `scsetup`(1M) pour identifier les priorités de la liste de nœuds du groupe de périphériques de disques.

- Pour plus d'informations sur les notions fondamentales relatives à la propriété `preferenced`, voir "Groupes de périphériques de disques multiports " à la page 45.
- Pour plus d'informations sur les procédures, reportez-vous au paragraphe "Comment changer les propriétés des périphériques de disques" dans "Administration des groupes de périphériques de disques" du *Guide d'administration système de Sun Cluster pour SE Solaris*.
- Pour plus d'informations sur les notions fondamentales relatives à la configuration des nœuds et au comportement des services de données de basculement et évolutifs, voir "Matériel système et composants logiciels Sun Cluster " à la page 21.

Si vous configurez tous les nœuds du cluster de la même façon, les limites d'utilisation sont appliquées de manière identique aux nœuds principaux et aux nœuds secondaires. Les paramètres de configuration des projets ne sont pas tenus d'être identiques pour toutes les applications dans les fichiers de configuration. Tous les

projets associés à l'application doivent au moins être accessibles par le biais de la base de données de projets sur tous les maîtres potentiels de l'application. Supposons que l'application 1 soit contrôlée par *phys-schost-1*, mais puisse être basculée sur *phys-schost-2* ou *phys-schost-3*. Le projet associé à l'application 1 doit être accessible sur les trois nœuds (*phys-schost-1*, *phys-schost-2* et *phys-schost-3*).

Remarque – Les informations de la base de données de projets peuvent être contenues dans un fichier de base de données local `/etc/project` ou stockées dans la carte NIS ou dans le service d'annuaire LDAP.

Le système d'exploitation Solaris permet de configurer des paramètres d'utilisation de façon souple ; peu de limites sont imposées par Sun Cluster. Les choix de configuration sont tributaires des besoins du site. Tenez compte des indications précisées aux rubriques suivantes pour la configuration de vos systèmes.

Définition des limites de mémoire virtuelle par processus

Pour limiter la mémoire virtuelle par processus, définissez le paramètre `process.max-address-space`. Pour plus d'informations sur la définition de la valeur `process.max-address-space`, voir `rctladm`(1M).

Si vous utilisez des contrôles de gestion avec le logiciel Sun Cluster, configurez des limites de mémoire de façon appropriée, pour empêcher tout basculement superflu et tout effet « ping-pong » des applications. En règle générale, respectez les consignes indiquées ci-dessous.

- Ne définissez pas de limites de mémoire trop basses.

  Lorsqu'une application atteint sa limite de mémoire, elle peut basculer. Cet aspect est particulièrement important pour les applications de base de données, où les conséquences liées à l'atteinte de la limite de mémoire virtuelle sont imprévisibles.

- Les limites de mémoire des nœuds principaux et des nœuds secondaires ne doivent pas être identiques.

  Si elles l'étaient, un effet ping-pong pourrait se produire au moment où l'application atteint sa limite de mémoire et bascule sur un nœud secondaire ayant une limite de mémoire identique. Définissez une limite de mémoire légèrement supérieure sur le nœud secondaire. Vous préviendrez ainsi des effets ping-pong et l'administrateur système aura un laps de temps lui permettant de régler les paramètres si nécessaire.

- Utilisez les limites de la mémoire de gestion de ressources pour l'équilibrage de la charge.

  Vous pouvez par exemple utiliser les limites de la mémoire pour empêcher une application errante de consommer trop d'espace de swap.

Scénarios de basculement

Vous pouvez configurer les paramètres de gestion de sorte que l'allocation de la configuration de projet (`/etc/project`) soit opérationnelle durant le fonctionnement normal du cluster et dans les situations de commutation et de basculement.

Les rubriques suivantes présentent des exemples de scénarios.

- Les deux premières rubriques, "Cluster à deux nœuds avec deux applications" et "Cluster à deux nœuds avec trois applications", présentent des scénarios de basculement de nœuds complets.
- La rubrique "Basculement de groupes de ressources uniquement" illustre les opérations de basculement d'une application seulement.

Dans un environnement Sun Cluster, vous configurez une application comme partie d'une ressource. Vous configurez ensuite une ressource comme partie d'un groupe de ressources. En cas de panne, le groupe de ressources ainsi que les applications qui lui sont associées basculent sur un autre nœud. Dans les exemples suivants les ressources n'apparaissent pas de manière explicite. Supposons que chaque ressource n'a qu'une seule application.

Remarque – un basculement intervient en fonction de l'ordre de préférence de la liste de nœuds définie par le RGM.

Les exemples présentés ci-après comportent les contraintes suivantes :

- Application 1 (App-1) est configurée dans le groupe de ressources GR-1.
- Application 2 (App-2) est configurée dans le groupe de ressources GR-2.
- Application 3 (App-3) est configurée dans le groupe de ressources GR-3.

Bien que le nombre de parts attribuées reste le même, le pourcentage de temps CPU alloué à chaque application change après un basculement. Ce pourcentage dépend du nombre d'applications tournant sur le nœud et du nombre de parts attribuées à chaque application active.

Pour ces scénarios, supposons que nous avons les configurations suivantes :

- Toutes les applications sont configurées sous un même projet.
- Chaque ressource n'a qu'une seule application.
- Les applications ne sont pas les seuls processus actifs sur les nœuds.
- Les bases de données de projets sont configurées de la même manière sur tous les nœuds du cluster.

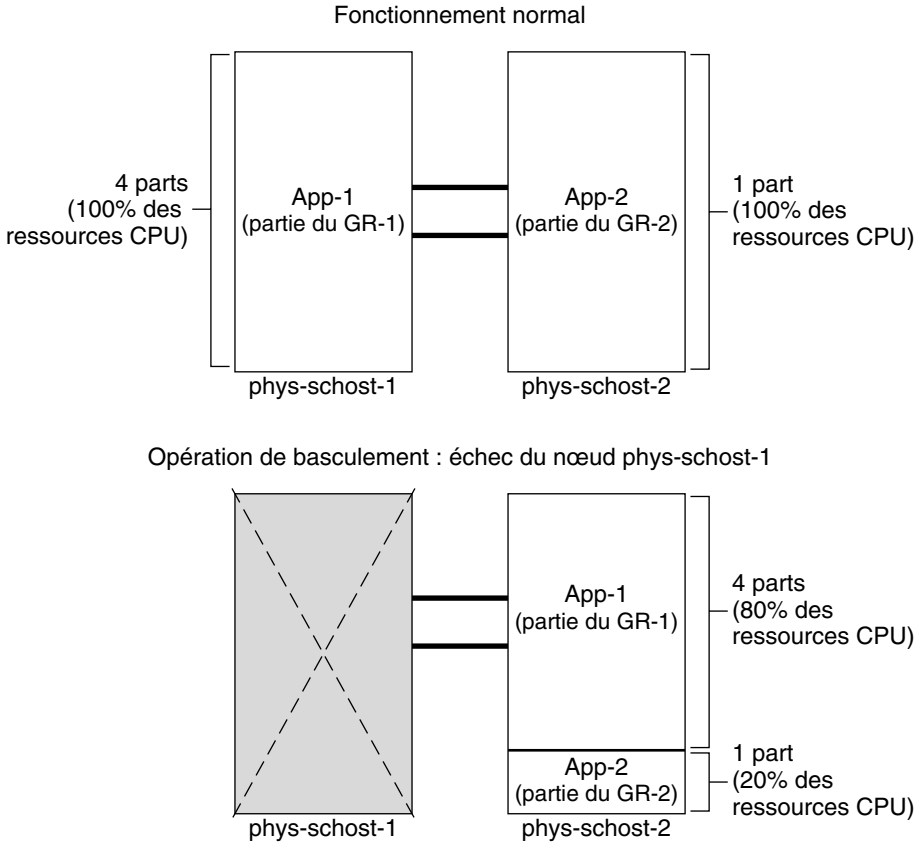
Cluster à deux nœuds avec deux applications

Vous pouvez configurer deux applications sur un cluster à deux nœuds pour vous assurer que chaque hôte physique (*phys-schost-1*, *phys-schost-2*) sert de maître par défaut pour une application. Chaque hôte physique sert de nœud secondaire à l'autre hôte physique. Tous les projets associés à l'application 1 et à l'application 2 doivent être représentés dans les fichiers de bases de données de projets sur les deux nœuds. Lorsque le cluster fonctionne normalement, chaque application tourne sur son maître par défaut, où le temps UC lui est attribué par la fonction de gestion.

Lorsqu'un basculement ou une commutation a lieu, les deux applications tournent sur un seul nœud, où les parts précisées dans le fichier de configuration leur sont attribuées. Par exemple, cette entrée du fichier `/etc/project` indique que 4 parts sont allouées à l'application 1 et qu'une part est allouée à l'application 2.

```
Prj_1:100:project for App-1:root::project.cpu-shares=(privileged,4,none)
Prj_2:101:project for App-2:root::project.cpu-shares=(privileged,1,none)
```

Le schéma indiqué ci-après présente le fonctionnement de cette configuration en situation normale et en cas de basculement. Le nombre de parts assignées ne change pas. Cependant, le pourcentage de temps CPU disponible pour chaque application peut changer. Il dépend du nombre de parts assignées à chaque processus qui exige du temps CPU.



Cluster à deux nœuds avec trois applications

Sur un cluster à deux nœuds avec trois applications, vous pouvez configurer un hôte physique (*phys-schost-1*) comme maître par défaut d'une application. Vous pouvez configurer le second hôte physique (*phys-schost-2*) comme maître par défaut pour les deux applications restantes. Dans l'exemple suivant, supposons que le fichier de base de données des projets se trouve sur chaque nœud. Le fichier de base de données des projets ne change pas lorsqu'un basculement ou une commutation a lieu.

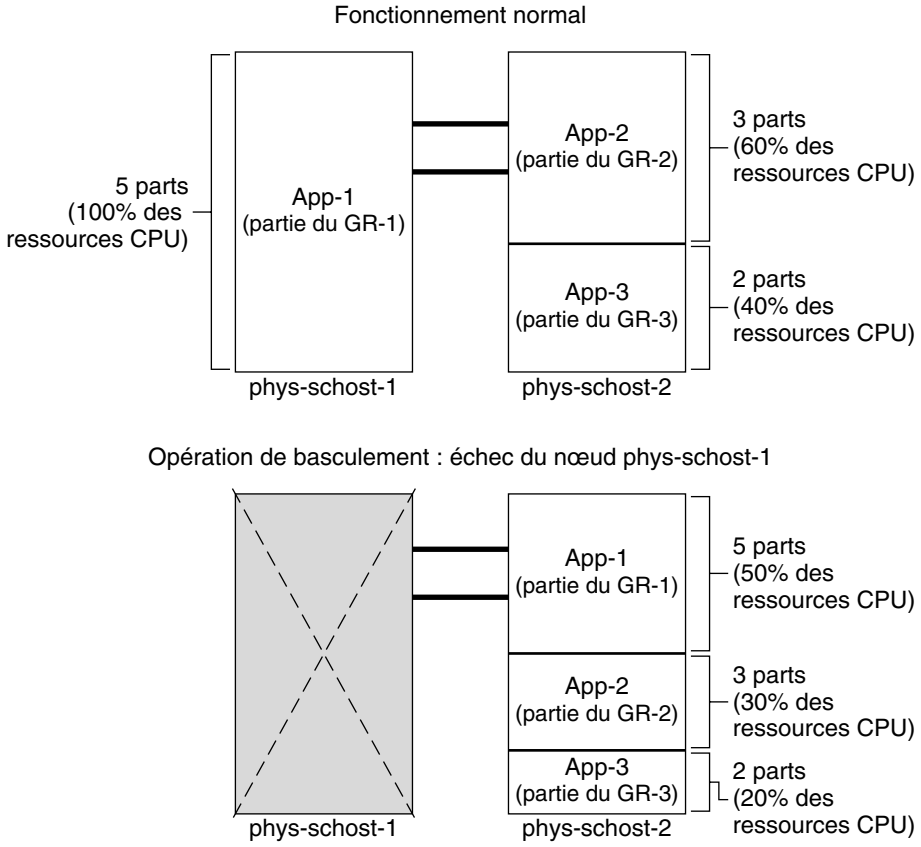
```
Prj_1:103:project for App-1:root::project.cpu-shares=(privileged,5,none)
Prj_2:104:project for App_2:root::project.cpu-shares=(privileged,3,none)
Prj_3:105:project for App_3:root::project.cpu-shares=(privileged,2,none)
```

Lorsque le cluster fonctionne normalement, 5 parts sont allouées à l'application 1 sur son maître par défaut, *phys-schost-1*. Ce nombre équivaut à 100 pour cent du temps UC parce que c'est la seule application qui demande du temps UC sur ce nœud. Trois parts et deux parts sont allouées respectivement aux applications 2 et 3 sur leur maître par défaut, *phys-schost-2*. L'application 2 reçoit 60 pour cent du temps CPU et l'application 3 reçoit 40 pour cent du temps CPU au cours du fonctionnement normal.

Si un basculement ou une commutation se produisent et que l'application 1 est basculée sur *phys-schost-2*, les parts des trois applications restent les mêmes. Toutefois, les pourcentages de ressources CPU sont allouées en fonction du fichier de la base de données des projets.

- L'application 1, avec 5 parts, reçoit 50 pour cent de l'UC.
- L'application 2, avec 3 parts, reçoit 30 pour cent de CPU.
- L'application 3, avec 2 parts, reçoit 20 pour cent de l'UC.

Le schéma indiqué ci-après présente le fonctionnement de cette configuration en situation normale et en cas de basculement.



Basculement du groupe de ressources uniquement

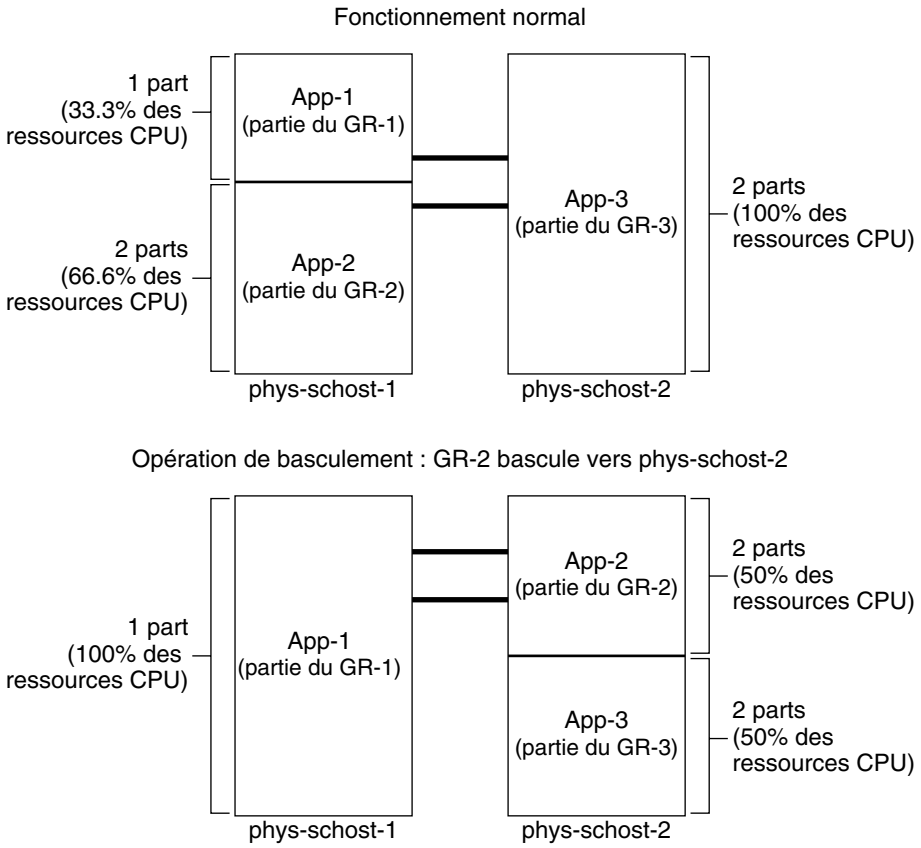
Dans une configuration où plusieurs groupes de ressources ont le même maître par défaut, un groupe de ressources (et ses applications associées) peuvent basculer ou commuter sur un nœud secondaire. Pendant ce temps, le maître par défaut continue de fonctionner dans le cluster.

Remarque – durant un basculement, l'application basculant se voit attribuer des ressources tel que spécifié dans le fichier de configuration du nœud secondaire. Dans cet exemple, les fichiers de base de données de projets sur les nœuds principal et secondaire ont les mêmes configurations.

Par exemple, cet échantillon de fichier de configuration spécifie que 1 part est allouée à l'application, 2 parts sont allouées à l'application 2 et 2 parts à l'application 3.

```
Prj_1:106:project for App_1:root::project.cpu-shares=(privileged,1,none)
Prj_2:107:project for App_2:root::project.cpu-shares=(privileged,2,none)
Prj_3:108:project for App_3:root::project.cpu-shares=(privileged,2,none)
```

Le diagramme suivant illustre les opérations normales et de basculement de cette configuration, où RG-2, qui contient l'application 2, bascule sur *phys-schost-2*. Notez que le nombre de parts assignées ne change pas. Toutefois, le pourcentage de temps CPU disponible pour chaque application peut changer, en fonction du nombre de parts assignées à chaque application qui exige du temps CPU.



Adaptateurs de réseau public et multi-acheminement sur réseau IP

Les clients adressent des requêtes de données au cluster à travers le réseau public. Chaque nœud du cluster est connecté à au moins un réseau public à travers une paire d'adaptateurs de réseau public.

Le logiciel de multi-acheminement sur réseau IP Solaris de Sun Cluster propose un mécanisme de base pour le contrôle des adaptateurs de réseau public et le basculement sur des adresses IP d'un adaptateur à un autre en cas de détection de panne. Chaque nœud du cluster a sa propre configuration de multi-acheminement sur réseau IP, qui peut être différente de celle des autres nœuds.

Les adaptateurs de réseau public sont organisés en *groupes de multi-acheminement sur IP* (groupes de multi-acheminement). Chaque groupe de multi-acheminement possède un ou plusieurs adaptateurs de réseau public. Chaque adaptateur d'un groupe peut être actif. Vous pouvez également configurer des interfaces de réserve qui sont inactives excepté en cas de basculement.

Le démon de multi-acheminement `in.mpathd` utilise une adresse IP de test pour détecter les pannes et les réparations. S'il détecte une panne sur l'un des adaptateurs, un basculement a lieu. Tout l'accès réseau bascule de l'adaptateur défectueux vers un autre adaptateur opérationnel du groupe de multi-acheminement. Ainsi, le démon conserve la connectivité du réseau public pour le nœud. Si vous avez configuré une interface de réserve, le démon la choisit. Sinon, le démon choisit l'interface ayant le plus petit nombre d'adresses IP. En cas de basculement au niveau de l'interface de l'adaptateur, les connexions de niveau supérieur, notamment TCP, ne sont pas affectées, excepté pendant un bref instant lors du basculement. En cas de succès du basculement des adresses IP, des diffusions ARP sont envoyées. Ainsi, le démon conserve les connexions aux clients distants.

Remarque – En raison des caractéristiques de récupération de surcharge du protocole TCP, les points finaux TCP peuvent connaître un retard supplémentaire après un basculement réussi. Certains segments peuvent avoir été perdus pendant le basculement, activant le mécanisme de contrôle de la surcharge du protocole TCP.

Les groupes de multi-acheminement procurent aux blocs assemblés des ressources de nom d'hôte logique et d'adresse partagée. Vous pouvez aussi créer des groupes de multi-acheminement indépendamment des ressources de nom d'hôte logique et d'adresse partagée pour contrôler la connexion des nœuds du cluster au réseau public. Sur un nœud, le même groupe de multiacheminement peut héberger un nombre indéfini de ressources de nom d'hôte logique ou d'adresse partagée. Pour plus d'informations sur les ressources de nom d'hôte logique et d'adresse partagée, voir *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

Remarque – Le mécanisme multi-acheminement sur réseau IP est conçu pour détecter et masquer les pannes des adaptateurs. Il n'est pas destiné à effectuer une récupération à partir de l'utilisation d'un administrateur de `ifconfig(1M)` pour la suppression d'une des adresses IP logiques (ou partagées). Le logiciel Sun Cluster affiche les adresses IP logiques et partagées comme ressources gérées par le RGM. Pour ajouter ou supprimer une adresse IP, un administrateur doit utiliser `scrgadm`(1M) pour modifier le groupe de ressources qui contient la ressource.

Pour plus d'informations sur l'implémentation du multi-acheminement sur réseau IP Solaris, reportez-vous à la documentation appropriée du système d'exploitation Solaris installé sur le cluster.

Version du système d'exploitation	Instructions
Système d'exploitation Solaris 8	*IP Network Multipathing Administration Guide*
Système d'exploitation Solaris 9	Chapitre 1, "IP Network Multipathing (Overview)" du *IP Network Multipathing Administration Guide*
Système d'exploitation Solaris 10	Partie VI, "IPMP" du *System Administration Guide: IP Services*

SPARC : Prise en charge de la reconfiguration dynamique

La prise en charge de la fonction de reconfiguration dynamique (DR) par Sun Cluster 3.1 8/05 est développée par phases successives. Cette section propose des explications et des observations concernant la prise en charge de la fonction DR par Sun Cluster 3.1 8/05.

Toutes les exigences de configuration, procédures et restrictions applicables à la reconfiguration dynamique (DR) de Solaris s'appliquent également à la DR de Sun Cluster (à l'exception de l'opération d'arrêt progressif de l'environnement d'exploitation). Reportez-vous donc à la documentation relative à la DR de Solaris *avant* d'utiliser la fonction DR du logiciel Sun Cluster. Vous devez consulter en particulier les problèmes qui affectent les périphériques d'E/S non réseau pendant une séparation DR.

Les documents *Sun Enterprise 10000 Dynamic Reconfiguration User Guide* et *Sun Enterprise 10000 Dynamic Reconfiguration Reference Manual* (des collections *Solaris 8 on Sun Hardware* ou *Solaris 9 on Sun Hardware*) peuvent tous deux être téléchargés à l'adresse http://docs.sun.com.

SPARC : Description générale d'une reconfiguration dynamique

La fonction DR permet des opérations, comme la suppression de matériel système, sur des systèmes en cours d'exécution. Les processus de la reconfiguration dynamique ont pour but d'assurer la continuité du fonctionnement du système. Il n'est pas nécessaire d'arrêter le système et d'interrompre la disponibilité du cluster.

La reconfiguration dynamique (DR) opère au niveau de la carte. Ainsi, une DR affecte tous les composants d'une carte. Chaque carte peut contenir plusieurs composants, notamment la CPU, la mémoire et les jonctions de périphériques des lecteurs de disques, lecteurs de bandes et connexions réseau.

Le retrait d'une carte qui contient des composants actifs entraîne des erreurs système. Avant la suppression d'une carte, le sous-système de la DR interroge d'autres systèmes, tels que Sun Cluster, pour déterminer si les composants de la carte sont en cours d'utilisation. Si le sous-système de la DR découvre qu'une carte est en cours d'utilisation, l'opération de suppression de la carte n'est pas effectuée. Ainsi, il est toujours conseillé de retirer une carte DR, car le sous-système de la DR rejette les opérations sur des cartes qui contiennent des composants actifs.

L'ajout de carte DR est également toujours conseillé. Les unités centrales et la mémoire d'une nouvelle carte sont automatiquement mises en service par le système. Cependant, l'administrateur système doit configurer manuellement le cluster pour l'utilisation de façon active des composants qui sont contenus sur cette nouvelle carte.

Remarque – le sous-système de la DR comporte plusieurs niveaux. Si un niveau inférieur rapporte une erreur, le niveau supérieur fait de même. Cependant, si le niveau inférieur rapporte une erreur spécifique, le niveau supérieur indique `Erreur inconnue`. Vous pouvez ignorer sans risque cette erreur.

Les rubriques suivantes présentent quelques observations sur les conséquences de l'utilisation de la reconfiguration dynamique avec différents types de périphériques.

SPARC : Clustering DR : éléments à prendre en compte pour les périphériques CPU

Le logiciel Sun Cluster ne rejette pas le retrait de carte DR en raison de la présence de périphériques CPU.

Lorsqu'une opération d'ajout de carte par reconfiguration dynamique a réussi, les CPU de la carte ajoutée sont automatiquement incorporées au système.

SPARC : Clustering DR : éléments à prendre en compte pour la mémoire

Pour les besoins de la DR, tenez compte de deux types de mémoires.

- mémoire noyau
- mémoire non noyau

Ces deux types ne diffèrent qu'au niveau de l'utilisation. Le matériel utilisé est le même. La mémoire noyau est la mémoire utilisée par le système d'exploitation Solaris. Le logiciel Sun Cluster ne prend pas en charge les retraits d'une carte qui contient la mémoire noyau et rejette toute opération de ce type. Si le retrait de la carte DR affecte

la mémoire autre que la mémoire noyau, le logiciel Sun Cluster ne rejette pas cette opération. Lorsqu'une carte qui contient de la mémoire est ajoutée par reconfiguration dynamique, la mémoire de cette carte est automatiquement incorporée au système.

SPARC : Clustering DR : éléments à prendre en compte pour les lecteurs de disques et de bandes

Sun Cluster rejette les opérations de suppression de cartes par reconfiguration dynamique sur les lecteurs actifs dans le nœud principal. Ces opérations peuvent être effectuées sur des lecteurs inactifs dans le nœud principal ou sur tous les lecteurs dans le nœud secondaire. Après l'opération DR, l'accès aux données de cluster se poursuit comme auparavant.

Remarque – Sun Cluster rejette les opérations DR ayant un impact sur la disponibilité des périphériques de quorum. Pour consulter les éléments à prendre en compte sur les périphériques de quorum et la procédure à suivre pour effectuer des opérations DR sur ceux-ci, voir "SPARC : Clustering DR : éléments à prendre en compte pour les périphériques de quorum " à la page 92.

Pour consulter les instructions détaillées sur l'exécution de ces actions, voir "Reconfiguration dynamique avec périphériques de quorum" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

SPARC : Clustering DR : éléments à prendre en compte pour les périphériques de quorum

Si le retrait de carte DR affecte une carte qui contient l'interface d'un périphérique de quorum, le logiciel Sun Cluster rejette cette opération. Le logiciel Sun Cluster identifie également le périphérique de quorum affecté par cette opération. Vous devez désactiver la fonction de quorum du périphérique avant d'effectuer une opération de suppression de carte par reconfiguration dynamique.

Pour consulter les instructions détaillées sur l'administration du quorum, voir Chapitre 5, "Administration du quorum" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

SPARC : Clustering DR : éléments à prendre en compte pour les interfaces d'interconnexion de cluster

Si le retrait de carte DR affecte une carte qui contient l'interface d'une interconnexion de cluster active, le logiciel Sun Cluster rejette cette opération. Le logiciel Sun Cluster identifie également l'interface affectée par cette opération. Vous devez utiliser un outil d'administration de Sun Cluster pour désactiver cette interface afin que cette opération puisse aboutir.

Caution – Le logiciel Sun Cluster nécessite que chaque nœud du cluster dispose au moins d'un chemin d'accès opérationnel à tous les nœuds du cluster. Ne désactivez pas une interface d'interconnexion privée prenant en charge le dernier chemin d'accès à un nœud du cluster.

Pour consulter les instructions détaillées sur l'exécution de ces actions, voir "Administration des interconnexions de cluster" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

SPARC : Clustering DR : éléments à prendre en compte pour les interfaces de réseau public

Si le retrait de carte DR affecte une carte qui contient une interface de réseau public active, le logiciel Sun Cluster rejette cette opération. Le logiciel Sun Cluster identifie également l'interface affectée par cette opération. Avant de retirer une carte qui contient une interface de réseau active, basculez tout le trafic de cette interface sur une autre interface opérationnelle dans le groupe de multi-acheminement, à l'aide de la commande `if_mpadm`(1M).

Caution – si l'adaptateur de réseau restant échoue au moment où vous supprimez l'adaptateur de réseau désactivé par reconfiguration dynamique, la disponibilité du cluster peut être menacée. L'adaptateur restant ne peut pas effectuer de basculement pendant toute la durée de l'opération DR.

Pour consulter les instructions détaillées sur l'exécution d'un retrait DR sur une interface de réseau public, voir "Administration du réseau public" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

CHAPITRE 4

Questions fréquemment posées

Ce chapitre donne les réponses aux questions les plus fréquemment posées sur le système Sun Cluster. Ces questions sont classées par thème.

FAQ sur la haute disponibilité

Question : Qu'est-ce qu'un système hautement disponible ?

Réponse : Le système Sun Cluster définit la haute disponibilité comme étant la capacité d'un cluster à assister l'exécution d'une application. L'application s'exécute même s'il se produit une panne qui rendrait un système serveur indisponible a priori.

Question : Quel est le processus grâce auquel le cluster peut fournir la haute disponibilité ?

Réponse : Le cluster fournit un environnement hautement disponible à travers un processus appelé basculement. Le basculement est une série d'actions exécutées par le cluster pour déplacer les ressources d'un service de données d'un noeud défectueux vers un autre noeud actif du cluster.

Question : Quelle est la différence entre un service de données évolutif et un service de données de basculement ?

Réponse : Il existe deux types de services de données hautement disponibles :

- Basculement
- Évolutif

Un service de données de basculement exécute une application sur un seul noeud principal du cluster à la fois. Les autres noeuds peuvent exécuter d'autres applications, mais chaque application ne tourne que sur un seul noeud. En cas d'échec d'un nœud principal, les applications qui s'exécutent sur ce nœud basculent sur un autre et continuent de s'exécuter.

Un service évolutif exécute une application sur plusieurs nœuds pour créer un service logique unique. Les services évolutifs utilisent tous les nœuds et processeurs du cluster sur lequel ils tournent.

Pour chaque application, un nœud héberge l'interface physique pour le cluster. Ce nœud s'appelle nœud d'interface globale (ou nœud GIF). Il peut exister plusieurs nœuds GIF dans le cluster. Chacun héberge une ou plusieurs interfaces logiques pouvant être utilisées par les services évolutifs. Ces interfaces logiques sont appelées *interfaces globales*. Un nœud GIF héberge une interface globale pour toutes les requêtes d'une application particulière et les distribue aux nœuds sur lesquels tourne le serveur d'application. Si le noeud GIF tombe en panne, l'interface globale bascule sur un noeud encore actif.

Si l'un des nœuds sur lequel l'application s'exécute tombe en panne, celle-ci continue de s'exécuter sur les autres nœuds, ses performances étant quelque peu réduites. Ce processus se poursuit jusqu'à ce que le nœud défectueux revienne dans le cluster.

FAQ sur les systèmes de fichiers

Question : Puis-je exécuter un ou plusieurs des nœuds de cluster en tant que serveurs NFS hautement disponibles avec d'autres nœuds de cluster clients ?

Réponse : Non, il ne faut pas faire de montage en boucle.

Question : Puis-je utiliser un système de fichiers de cluster pour les applications qui ne sont pas sous le contrôle de Gestionnaire du groupe de ressources ?

Réponse : Oui. Toutefois, sans contrôle de RGM vous devez redémarrer manuellement les applications après la panne du nœud sur lequel elles s'exécutent.

Question : Tous les systèmes de fichiers de cluster doivent-ils posséder un point de montage dans le répertoire `/global` ?

Réponse : Non. Toutefois, en plaçant tous les systèmes de fichiers de cluster sous le même point de montage, par exemple `/global`, vous en faciliterez l'organisation et la gestion.

Question : Quelle différence y-a-t-il entre l'utilisation des systèmes de fichiers de cluster et l'exportation de systèmes de fichiers NFS ?

Réponse : Il existe plusieurs différences :

1. Le système de fichiers de cluster prend en charge les périphériques globaux. NFS ne prend pas en charge l'accès distant aux périphériques.
2. Le système de fichiers de cluster a un espace de noms global. Seulement une commande de montage est requise. Avec NFS, vous devez monter le système de fichiers sur chaque noeud.

3. Le système de fichiers de cluster met en cache des fichiers plus souvent que ne le fait NFS. Par exemple, il met en cache un fichier lorsque celui-ci est utilisé par plusieurs nœuds pour des opérations de lecture, d'écriture, de verrouillage de fichiers et d'E/S asynchrones.
4. Le système de fichiers de cluster a été conçu pour exploiter les futures capacités d'interconnexion rapide du cluster apportant des fonctions de DMA distant et de zéro copie.
5. Si vous modifiez les attributs d'un fichier (par exemple, en utilisant `chmod(1M)`) dans un système de fichiers de cluster, la modification est immédiatement appliquée à tous les nœuds. S'il s'agit d'un système de fichiers NFS exporté, l'application de cette modification peut prendre plus de temps.

Question : Le système de fichiers `/global/.devices/node@`*nodeID* s'affiche sur mes nœuds de cluster. Puis-je l'utiliser pour stocker des données destinées à être hautement disponibles et globales ?

Réponse : Ces systèmes de fichiers stockent l'espace de noms des périphériques globaux. Ils ne sont pas destinés à une utilisation générale. Ils sont globaux, mais sans être accessibles d'une manière globale—chaque nœud accède seulement à son propre espace de noms des périphériques globaux. Si un nœud a échoué, les autres ne peuvent pas accéder à son espace de noms. Ces systèmes de fichiers ne sont pas hautement disponibles. Ils ne doivent pas être utilisés pour stocker des données accessibles globalement ou hautement disponibles.

FAQ sur la gestion des volumes

Question : Dois-je mettre en miroir tous les périphériques de disque ?

Réponse : Pour qu'un périphérique de disque soit considéré comme hautement disponible, il doit être mis en miroir ou utiliser le matériel RAID 5. Tous les services de données doivent utiliser des périphériques de disque hautement disponibles ou des systèmes de fichiers de cluster montés sur ce type de périphériques. Ces configurations peuvent supporter des pannes de disque uniques.

Question : Puis-je utiliser deux gestionnaires de volumes différents pour les disques locaux (disque d'initialisation) et pour les disques multihôtes ?

Réponse : SPARC : cette configuration est prise en charge, le logiciel Solaris Volume Manager gérant les disques locaux et VERITAS Volume Manager gérant les disques multihôtes. Aucune autre combinaison n'est prise en charge.

x86: non, cette configuration n'est pas prise en charge, car seul le logiciel Solaris Volume Manager est pris en charge sur les clusters x86.

FAQ sur les services de données

Question : Quels sont les services de données Sun Cluster disponibles ?

Réponse : La liste des services de données pris en charge figure dans "Produits pris en charge" du *Notes de version de Sun Cluster 3.1 8/05 pour SE Solaris*.

Question : Quelles sont les versions d'application prises en charge par les services de données Sun Cluster ?

Réponse : La liste des versions d'application prises en charge figure dans "Produits pris en charge" du *Notes de version de Sun Cluster 3.1 8/05 pour SE Solaris*.

Question : Puis-je écrire mon propre service de données ?

Réponse : Oui. Pour plus d'informations, voir Chapitre 11, "Fonctions de l'API de la bibliothèque DSDL" du *Guide du développeur de services de données Sun Cluster pour SE Solaris*.

Question : Lors de la création des ressources réseau, dois-je spécifier les adresses IP numériques ou les noms d'hôtes ?

Réponse : Le meilleur moyen de spécifier les ressources réseau est d'utiliser le nom d'hôte UNIX plutôt que l'adresse IP numérique.

Question : Lors de la création de ressources réseau, quelle est la différence entre l'utilisation d'un nom d'hôte logique (ressource LogicalHostname) et l'utilisation d'une adresse partagée (ressource SharedAddress) ?

Réponse : Excepté dans le cas de Sun Cluster HA pour NFS, où la documentation conseille d'utiliser une ressource `LogicalHostname` dans un groupe de ressources en mode `Failover`, vous pouvez utiliser de façon interchangeable la ressource `SharedAddress` ou la ressource `LogicalHostname`. L'utilisation d'une ressource `SharedAddress` implique un temps système supplémentaire car le logiciel de mise en réseau de cluster est configuré pour une ressource `SharedAddress`, mais pas pour une ressource `LogicalHostname`.

L'avantage lié à l'utilisation d'une ressource `SharedAddress` est manifeste si vous configurez à la fois des services de données évolutifs et de basculement et souhaitez que les clients puissent accéder à ces deux services en utilisant un même nom d'hôte. Dans ce cas, les ressources `SharedAddress` ainsi que la ressource d'application de basculement sont contenues dans un seul groupe de ressources. La ressource de service évolutif est contenue dans un groupe de ressources distinct et configurée pour utiliser la ressource `SharedAddress`. Les services évolutif et de basculement peuvent alors utiliser le même ensemble de noms d'hôte/adresses configurés dans la ressource `SharedAddress`.

FAQ sur les réseaux publics

Question : Quels sont les adaptateurs de réseau public pris en charge par le système Sun Cluster ?

Réponse : Actuellement, le système Sun Cluster prend en charge les adaptateurs de réseau public Ethernet (10/100BASE-T et 1000BASE-SX Gb). De nouvelles interfaces pouvant être prises en charge dans le futur, informez-vous auprès de votre représentant Sun.

Question : Quel est le rôle de l'adresse MAC au cours d'un basculement ?

Réponse : Lorsqu'un basculement se produit, de nouveaux paquets du protocole ARP (Address Resolution Protocol) sont générés et diffusés. Ils contiennent la nouvelle adresse MAC (du nouvel adaptateur physique vers lequel le nœud a basculé) et l'ancienne adresse IP. Lorsqu'une autre machine sur le réseau reçoit un de ces paquets, elle vide son cache ARP de l'ancien mappage MAC-IP et utilise le nouveau.

Question : Le système Sun Cluster prend-il en charge la définition `local-mac-address?=true` *?*

Réponse : Oui. En fait, le multiacheminement sur réseau IP *requiert* la définition de `local-mac-address?` sur `true`.

Vous pouvez définir `local-mac-address?` sur `eeprom`(1M), à l'invite `ok` de la PROM OpenBoot dans un cluster SPARC. Vous pouvez également définir l'adresse MAC avec l'utilitaire SCSI que vous pouvez exécuter après le démarrage du BIOS sur un cluster x86.

Question : Quel est le délai d'attente lorsque multi-acheminement sur réseau IP effectue un basculement entre les adaptateurs ?

Réponse : Ce délai d'attente peut être de quelques minutes. En effet, lors d'un basculement multi-acheminement sur réseau IP, un ARP gratuit est envoyé. Toutefois, il n'est pas certain que le routeur situé entre le client et le cluster utilisera l'ARP gratuit. Ainsi, tant que le délai d'attente de l'entrée de cache ARP correspondant à cette adresse IP n'a pas expiré, l'entrée peut utiliser l'adresse MAC périmée.

Question : Combien de temps faut-il pour détecter la panne d'un adaptateur réseau ?

Réponse : Le temps de détection de la panne par défaut est de 10 secondes. L'algorithme essaie de se conformer à ce temps de détection, mais le temps réel dépend de la charge du réseau.

FAQ sur les membres de cluster

Question : Tous les membres du cluster doivent-ils avoir le même mot de passe racine ?

Réponse : Non, ce n'est pas obligatoire. Toutefois, en adoptant le même mot de passe, vous simplifierez l'administration du cluster.

Question : L'ordre dans lequel les noeuds sont initialisés est-il important ?

Réponse : Dans la plupart des cas, non. Cependant l'ordre d'initialisation est important pour empêcher l'amnésie. Par exemple, si le noeud 2 est propriétaire du périphérique de quorum et que le noeud 1 est arrêté, si vous arrêtez ensuite le noeud 2, vous devrez l'activer de nouveau avant d'activer le noeud 1. Cet ordre vous empêche d'activer par inadvertance un nœud avec des informations de configuration de cluster périmées. Pour plus d'informations sur l'amnésie, voir "À propos de la séparation en cas d'échec " à la page 57.

Question : Dois-je mettre en miroir les disques locaux d'un noeud du cluster ?

Réponse : Oui. Bien que cette mise en miroir ne soit pas indispensable, elle empêche l'arrêt du nœud du cluster en cas de panne d'un disque non mis en miroir. Cette mise en miroir a cependant pour conséquence d'augmenter la charge d'administration du système.

Question : Quels sont les problèmes liés à la sauvegarde des membres du cluster ?

Réponse : Il existe plusieurs méthodes de sauvegarde d'un cluster. Vous pouvez disposer d'un nœud configuré comme nœud de sauvegarde, accompagné d'un lecteur de bande ou d'une bibliothèque. Le système de fichiers de cluster est ensuite utilisé pour sauvegarder les données. Ne connectez pas ce noeud aux disques partagés.

Pour plus d'informations sur la sauvegarde et la restauration des données, voir Chapitre 9, "Sauvegarde et restauration d'un cluster" du *Guide d'administration système de Sun Cluster pour SE Solaris*.

Question : Quand un noeud est-il suffisamment sain pour être utilisé comme noeud secondaire ?

Réponse : Solaris 8 et Solaris 9 :

Après une réinitialisation, un nœud est suffisamment sain pour devenir nœud secondaire lorsqu'il affiche l'invite de connexion.

Solaris:

Un nœud est suffisamment sain pour devenir un nœud secondaire lorsque le jalon `multi-user-server` est en cours d'exécution.

```
# svcs -a | grep multi-user-server:default
```

FAQ sur le stockage de cluster

Question : Qu'est-ce qui rend le stockage multihôte hautement disponible ?

Réponse : Il est hautement disponible, car il peut survivre à la perte d'un disque en raison de la mise en miroir (ou des contrôleurs RAID-5 basés sur le matériel). Un périphérique de stockage multihôte ayant plus d'une connexion hôte, il peut aussi résister à la perte d'un noeud auquel il est connecté. En outre, chaque noeud possède des chemins d'accès redondants vers le périphérique de stockage auquel il est relié afin de supporter les pannes des adaptateurs de bus hôtes, des câbles ou des contrôleurs de disques.

FAQ sur l'interconnexion de cluster

Question : Quelles sont les interconnexions de cluster prises en charge par le système Sun Cluster ?

Réponse : Actuellement, le système Sun Cluster prend en charge les interconnexions de cluster suivantes :

- Ethernet (100BASE-T Fast Ethernet et 1000BASE-SX Gb) sur les clusters SPARC et x86 ;
- Infiniband sur les clusters SPARC et x86 ;
- SCI uniquement sur les clusters SPARC.

Question : Quelle est la différence entre un « câble » et un « chemin » d'accès ?

Réponse : Les câbles de transport intracluster sont configurés à l'aide des adaptateurs et des commutateurs de transport. Les câbles relient les adaptateurs et les commutateurs d'un composant à l'autre. Le gestionnaire de la topologie du cluster utilise les câbles disponibles pour établir des chemins d'accès de bout en bout entre les noeuds. Un câble ne correspond pas directement à un chemin d'accès.

Les câbles sont « activés » et « désactivés » de manière statique par un administrateur. On parle « d'état » des câbles (activé ou désactivé), mais pas de « statut ». Si un câble est désactivé, c'est comme s'il n'était pas configuré. Les câbles désactivés ne peuvent pas être utilisés comme chemins d'accès. Leur état n'est pas connu puisqu'ils n'ont pas été testés. La commande `scconf -p` vous permet d'obtenir l'état d'un câble.

Les chemins d'accès sont définis de manière dynamique par le gestionnaire de topologie du cluster, déterminant également leur « statut ». Un chemin d'accès peut avoir le statut « en ligne » ou « hors ligne ». La commande `scstat(1M)` vous permet d'obtenir le statut d'un chemin d'accès.

Prenons l'exemple suivant d'un cluster à deux noeuds avec quatre câbles.

```
node1:adapter0      to switch1, port0
node1:adapter1      to switch2, port0
node2:adapter0      to switch1, port1
node2:adapter1      to switch2, port1
```

Deux chemins d'accès sont possibles à partir de ces quatre câbles.

```
node1:adapter0    to node2:adapter0
node2:adapter1    to node2:adapter1
```

FAQ sur les systèmes client

Question : Dois-je prendre en compte certains besoins ou restrictions spécifiques au client dans le cadre de l'utilisation d'un cluster ?

Réponse : Les systèmes client se connectent au cluster comme ils le feraient à tout autre serveur. En fonction de l'application du service de données, vous pouvez dans certains cas avoir à installer un logiciel client ou procéder à d'autres modifications de configuration pour que le client puisse se connecter à l'application du service de données. Pour plus d'informations sur la configuration requise côté client, voir Chapitre 1, "Planning for Sun Cluster Data Services" du *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

FAQ sur la console administrative

Question : Le système Sun Cluster nécessite-t-il une console administrative ?

Réponse : Oui.

Question : La console administrative doit-elle être dédiée au cluster ou peut-elle être utilisée pour d'autres tâches ?

Réponse : Le système Sun Cluster ne requiert pas de console administrative dédiée, mais si vous n'en utilisez qu'une, vous bénéficierez des avantages décrits ci-dessous.

- Elle permet une gestion centralisée des clusters en regroupant les outils de gestion et de console sur la même machine.
- Elle peut permettre une résolution plus rapide des problèmes par votre fournisseur de services matériels.

Question : La console administrative doit-elle être située « à proximité » du cluster, par exemple, dans la même pièce ?

Réponse : Consultez votre fournisseur de services. Ce fournisseur peut exiger que la console soit située à proximité du cluster. Il n'existe cependant aucune raison technique pour que la console soit placée dans la même pièce.

Question : Une console administrative peut-elle servir plusieurs clusters si des conditions de proximité sont également respectées ?

Réponse : Oui. Il est possible de contrôler plusieurs clusters à partir d'une seule console administrative. Vous pouvez aussi partager un seul concentrateur de terminaux entre les clusters.

FAQ sur le concentrateur de terminaux ou le processeur de services système (SSP)

Question : Le système Sun Cluster nécessite-t-il un concentrateur de terminaux ?

Réponse : Depuis Sun Cluster version 3.0, aucune version du logiciel ne nécessite de concentrateur de terminaux pour s'exécuter. Contrairement à Sun Cluster version 2.2, qui nécessitait un concentrateur de terminaux pour la séparation en cas d'échec, les versions ultérieures ne dépendent pas du concentrateur de terminaux.

Question : La plupart des serveurs Sun Cluster utilisent un concentrateur de terminaux, contrairement au serveur Sun Enterprise E1000. Pourquoi ?

Réponse : Le concentrateur de terminaux constitue en effet un convertisseur série/Ethernet pour la plupart des serveurs. La console du concentrateur de terminaux dispose d'un port série. Le serveur Sun Enterprise E1000 ne possède pas de console série. C'est le SSP (System Service Processor) qui sert de console, à travers un port Ethernet ou un port `jtag`. Vous utilisez toujours SSP pour les consoles des serveurs Sun Enterprise E1000.

Question : Quels sont les avantages de l'utilisation d'un concentrateur de terminaux ?

Réponse : Un concentrateur de terminaux permet à chaque nœud d'une station de travail distante, quel que soit son emplacement sur le réseau, d'accéder à la console. Cet accès est fourni même lorsque le nœud est situé à l'invite de la PROM OpenBoot s'il s'agit d'un nœud SPARC ou sur un sous-système d'initialisation s'il s'agit d'un nœud x86.

Question : Si j'utilise un concentrateur de terminaux que Sun ne prend pas en charge, que dois-je savoir pour l'habiliter ?

Réponse : La principale différence entre le concentrateur de terminaux pris en charge par Sun et les autres périphériques de console réside dans le fait que le premier possède un microprogramme spécial. Ce microprogramme empêche le concentrateur de terminaux d'envoyer une valeur d'interruption à la console lors de son initialisation. Si vous possédez un périphérique de console qui peut envoyer une valeur de ce type ou un signal pouvant être interprété comme tel, la valeur d'interruption désactive le nœud.

Question : Puis-je libérer un port verrouillé du concentrateur de terminaux pris en charge par Sun sans le réinitialiser ?

Réponse : Oui. Notez le numéro de port à réinitialiser et entrez les commandes suivantes :

```
telnet tc
Entrez le nom ou le numéro de port annexe : cli
annexe : su -
annex# admin
admin : reset numéro_port
admin : quit
annex# hangup
#
```

Pour plus d'informations sur la configuration et l'administration du concentrateur de terminaux pris en charge par Sun, reportez-vous aux manuels suivants :

- "Présentation de l'administration de Sun Cluster" du *Guide d'administration système de Sun Cluster pour SE Solaris*
- Chapitre 2, "Installing and Configuring the Terminal Concentrator" du *Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS*

Question : Que se passe-t-il si le concentrateur de terminaux tombe en panne ? Dois-je en avoir un autre à disposition ?

Réponse : Non. Le cluster garde toute sa disponibilité en cas de panne du concentrateur de terminaux. Vous perdez la capacité de vous connecter aux consoles du nœud jusqu'à ce que le concentrateur soit de nouveau en service.

Question : Qu'en est-il de la sécurité lorsque j'utilise un concentrateur de terminaux ?

Réponse : En général, le concentrateur de terminaux est connecté à un petit réseau utilisé par les administrateurs système, et non à un réseau utilisé pour les autres accès client. Vous pouvez contrôler la sécurité en limitant l'accès à ce réseau particulier.

Question : SPARC : comment utiliser la reconfiguration dynamique avec un lecteur de disques ou de bandes ?

Réponse : Procédez comme suit :

- Déterminez si le lecteur de disques ou de bandes fait partie d'un groupe de périphériques actif. Si ce n'est pas le cas, des opérations de suppression par reconfiguration dynamique peuvent être effectuées sur ce lecteur.
- Si l'opération de suppression par reconfiguration dynamique touche un lecteur de disques ou de bandes actif, le système rejette l'opération et identifie les lecteurs susceptibles d'être affectés par cette opération. Si le lecteur fait partie d'un groupe de périphériques actif, accédez à "SPARC : Clustering DR : éléments à prendre en compte pour les lecteurs de disques et de bandes " à la page 92.
- Déterminez si le lecteur est un composant du noeud principal ou du noeud secondaire. Si le lecteur est un composant du noeud secondaire, vous pouvez y effectuer l'opération de suppression par reconfiguration dynamique.

- Si le lecteur est un composant du nœud principal, vous devez commuter les nœuds principal et secondaire avant d'effectuer l'opération de suppression par reconfiguration dynamique sur ce périphérique.

Caution – tout échec sur le noeud principal, alors que vous effectuez une opération DR sur un noeud secondaire, a une incidence sur la disponibilité du cluster. Le noeud principal n'a pas d'endroit pour basculer, jusqu'à ce qu'un nouveau noeud secondaire soit défini.

Index

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V