



ORACLE

Oracle ACFS 21c



NAS Maximum Availability Extensions

2022年5月 | バージョン1.02

Copyright © 2022, Oracle and/or its affiliates

公開

免責事項

本文書には、ソフトウェアや印刷物など、いかなる形式のものも含め、オラクルの独占的な所有物である占有情報が含まれます。この機密文書へのアクセスと使用は、締結および遵守に同意したOracle Software License and Service Agreementの諸条件に従うものとします。本文書と本文書に含まれる情報は、オラクルの事前の書面による同意なしに、公開、複製、再作成、またはオラクルの外部に配布することはできません。本文書は、ライセンス契約の一部ではありません。また、オラクル、オラクルの子会社または関連会社との契約に組み込むことはできません。

本書は情報提供のみを目的としており、記載した製品機能の実装およびアップグレードの計画を支援することのみを意図しています。マテリアルやコード、機能の提供をコミットメント（確約）するものではなく、購買を決定する際の判断材料にするものでもありません。本書に記載されている機能の開発、リリースおよび時期については、オラクルの裁量により決定されます。

製品アーキテクチャの性質上、本書に記述されているすべての機能を安全に組み込むことができず、コードの不安定化という深刻なリスクを伴う場合があります。

目次

免責事項	1
はじめに	3
設計概要	4
Oracle ACFS NAS MAXの利点	4
サポート対象プラットフォーム	5
SMBのサポート	6
NFSのサポート	6
説明	7
Oracle ACFS NAS MAXのクライアント使用	7
12.2.0.1 SRVCTLコマンドの説明	8
サーバー側のエクスポート・オプション	9
クライアント側のマウント・オプション	9
NFSとサーバーの基本的セットアップ	10
エクスポートする初期ファイル・システムの作成	11
ファイル・システムの作成例	12
シンプルなOracle ACFS HA-NFSシナリオの構成（ロックなし）	12
ExportFSの登録	13
HA-SMBエクスポートの追加	15
HA-SMBのSamba構成	15
ノードの再配置	16
ロックありのHA-NFS V4のセットアップ	17
複雑なシナリオ	19
CRSのポリシー例 – 最適なノードの選択	20
どのような状況で、HAVIPとExportFSが他のノードに移動するのか	20
HAVIPのフェイルオーバー時間	22
計画的再配置	22
ノード障害	22
NFSのパフォーマンスに関する考慮事項	22
SMBのパフォーマンスに関する考慮事項	23
Oracle ACFS NAS MAXのスケラビリティ – ロックなしのNFSとSMB	23
ExportFSリソースの動作	23
クラスタ内のエクスポートの場所の制御	24
その他の検討事項	24
トラブルシューティング	24
結論	26

はじめに

クラウド。現在と未来はクラウドの世界へと向かっています。クラウド中心のアーキテクチャを基にビジネスを設計、計画し、正しく実装する企業は、大きな利益を上げて成功します。クラウドの世界では、複数のコンポーネントが考慮されますが、ストレージ・レイヤーを主要なコンポーネントに据えた基盤ほど重要なものではありません。このストレージ・レイヤーは、このクラウドという目的地に適した幅広い機能を備えている必要があります。

Oracle ACFSは、Oracle Clusterwareテクノロジー・セットの一部として提供されているクラスター・ファイル・システムであり、Oracle Gridクラスターの任意のノード上のファイルにアクセスできるようにします。このテクノロジーは両方のOracle Linux Supportライセンスで利用でき、Oracle Real Application Clusters (Oracle RAC) ライセンスの一部として利用することも可能です。

Oracle ACFSファイル・システムの上に構築されたOracle ACFS NAS Maximum Availability Extensions (Oracle ACFS NAS MAX) は、Oracle ACFSリソースとOracle Clusterwareリソースを利用するテクノロジー・セットであり、Oracle ACFSを介して特定のOSプロトコル (NFSやSMBなど) を実行できるようにします。これにより、Oracle ACFS NAS MAXは、Oracle RACクラスターの可用性の限界に達する高可用性サービス・アーキテクチャを提供します。Oracle ACFS NAS MAXは、Oracle Grid InfrastructureコンポーネントとOSプロトコル・テクノロジーによって提供される高可用性テクノロジーの上に構築されています。

12.2.0.1以降のOracle ACFS NAS MAXは、次の高可用性のフロントエンドを実現します。

- NFS v3, v4, およびv4 (ロックあり) (HA-NFS)
- WindowsおよびSamba用のSMBプロトコル (HA-SMB)

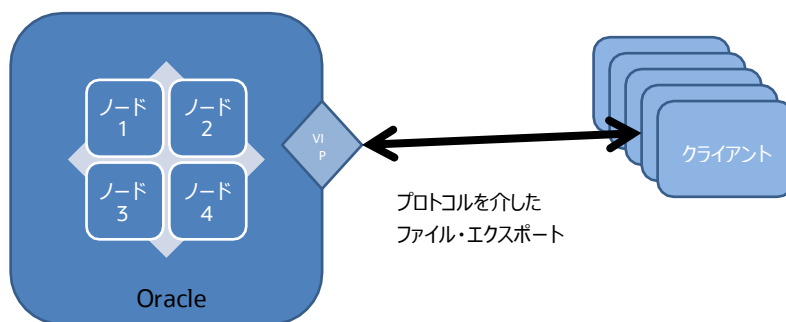
このテクノロジーにより、Oracle Gridクラスター内のいずれかのノードが損失しても存続する一連のNFSエクスポートまたはSMBエクスポートを作成して保持し、これらのエクスポートをクライアントで常に使用可能にできます。これは、Maximum Availability Architectureの重要な部分です。

この技術概要では、このテクノロジーの最新機能、基本構成と高度な構成、制限事項、およびシステム要件について説明します。

設計概要

おおまかには、Oracle ACFS NAS MAXはクライアントが仮想IP（VIP）に接続するためのアーキテクチャを提供します。現在のノードでサービスが利用できなくなった場合、このVIPはCRSリソースによって制御されて、クラスタ内を移動します。ほとんどの場合、クライアントはサーバーが別のサーバーに置き換わったことを認識せず、VIPがホストされている新しいサーバーとの間でトランザクションを続行します。

概念設計



Oracle RACクラスタ内のノードが使用可能である限り、VIP経由でエクスポートが引き続き論理的に使用可能。どのOracle RACノードがVIPへのアクセスを提供しているかを、クライアントは認識しない。

図1 - 概念設計

Oracle ACFS NAS MAXの利点

- Oracle ACFS NAS MAXはクラスタ・ファイル・システムOracle ACFSの上に構築されており、NFS共有およびSMB共有でOracle ACFSの管理インタフェースとメリットを活用できるようにします。
 - Oracle ACFSはクラスタ・ファイル・システムであるため、Oracle RACクラスタ内のすべてのノードのデータを利用できます。
 - Oracle ACFSタギングにより、属性でファイルをタグ付けできるため、レプリケーションやバックアップなどのために後で簡単にフィルタリングできます（タグ情報はホストでのみ使用可能です。クライアントでは使用できません）。
 - Oracle ACFSレプリケーションにより、複数のクラスタにわたってファイル・システム全体をレプリケートできるため、データセンター全体で高可用性が得られます。
 - Oracle ACFSスナップショットにより、管理者はデータのコピーを作成できるため、リカバリ、制御された変更、バックアップを迅速に実行できます。これらのスナップショットをOracle ACFS NAS MAXによってエクスポートできます。
 - Oracle ACFSセキュリティと暗号化により、柔軟なセキュリティ・ルールを使用してファイル・システムのファイルにアクセス権を定義でき、業界標準のディスク暗号化も使用できます。

- ファイル・システムをマウントして利用可能にしながら、ファイル・システムのサイズを自動的に変更できます。
 - Oracle ACFS圧縮により、Oracle ACFSファイル・システムに格納される情報のサイズを低減できます。
 - これらの機能およびOracle ACFSのその他の機能について、詳しくはOracle Advanced Cluster File Systemの管理者ガイドを参照してください。
- Oracle ACFS NAS MAXは、Grid Infrastructureスタックの起動と停止にSMBおよびNFSを統合し、従来使用されているSMBノードおよびNFSノードのinit起動スクリプト構成に代わる方法を提供します。Oracle ASMはブート・プロセスの後半まで使用できないため、ブート時にエクスポートのためにエクスポート対象のファイル・システムをエクスポート・ファイルまたはsamba設定ファイルに配置するという従来の方法は、Oracle ADVMボリュームにホストされているファイル・システムには使用できません。Oracle ACFS NAS MAXを使用してサーバーにエクスポートを提供すると、管理者は障害グループ、冗長性、管理インタフェースなどのOracle ASM機能を利用しながら、通常のNFS操作やSMB操作を引き続き実行できます。
 - スケーラビリティ – Oracle ACFS NAS MAXクラスタに新しいノードを追加し、バックエンドのネットワークとストレージ・ファブリックの限界にまでクラスタ・スループットのパフォーマンスを拡張できます。

サポート対象プラットフォーム

Oracle ACFS NAS MAXは、Oracle ACFSがサポートされている大部分のプラットフォームでサポートされています。最新の情報については、My Oracle Supportのナレッジベースの記事1369107.1¹を参照してください。

以下に示す特定のOSについては、固有の要件が存在します。

- AIX HA-SMB : Sambaバージョン3.6.24以降
- Solaris HA-SMB : Sambaバージョン3.6.24以降
- Linux HA-NFS : nfs-utils-1.0.9-60以降
- Linux HA-SMB : Sambaバージョン4.2以降
- Windows – SMBのみ
 - SMBはWindowsに組み込まれています。サーバーが新しいほど、新しいSMBバージョンがサポートされていますが、クライアントで古いバージョンを適用できます。サーバーでは最新のWindows OSを実行することを推奨します。
 - 20C以降、ACFSはWindowsでサポートされませんが、サーバー側のSMBエクスポート機能をLinuxから使用して、ACFSファイル・システムをWindowsにエクスポートすることができます。

¹ <https://support.oracle.com/oip/faces/secure/km/DocumentDisplay.jspx?id=2171587.1>

SMBのサポート

Windowsでは、SMBに複数のプロトコルがあります。プロトコル・ネゴシエーションは、次の2つの要因によって決まります。

1. 実行するサーバーOS – たとえば、Windows 2012R2は2008R2よりもはるかに新しいSMBプロトコルをサポートしています。
2. 実行するクライアントOS – サーバーは、クライアントとサーバーの両方が処理できるSMBバージョンに基づいて、合意されたバージョンをネゴシエートします。

詳しくは、Microsoftによるブログ記事²を参照してください。

エラーのリカバリ機能と再開機能（高可用性）が必要な場合は、サーバーとクライアントの両方でSMB 3.0以降のバージョンを使用することを強く推奨します。この場合、Windows 2012以降のOSバージョンが必要です。

他のバージョンのSMBも動作しますが、以下のことが生じる可能性があります。

- ネットワーク障害の後で、クライアントで共有の再マウントが必要になる可能性があります。
- 転送速度が低下する可能性があります。

Powershellコマンド`Get-SmbConnection`を使用すると、サーバーまたはクライアントで使用しているSMBのバージョンを確認できます。

Linux、Solaris、およびAIXでは、Sambaパッケージをインストールする必要があります。これらのパッケージは、www.samba.orgから入手できます。代替のSMB実装を提供しているOSベンダーもありますが、HA-SMBにはSamba.orgが必要です。Sambaの最新バージョンを使用することを推奨します。

NFSのサポート

AIX、Solaris、およびLinuxでは、HA-NFSはv3およびv4をロックなしでサポートしています。

また、Linuxのみで、HA-NFSはNFSv4をロックありでサポートしています。

NFSは通常、OSカーネルの一部です。対象OSの最新カーネルを使用することを推奨します。UEKカーネルとRHカーネルおよびNFSv4サポートについて、詳しくはMOS Note 2171587.1³を参照してください。

² <https://docs.microsoft.com/en-us/archive/blogs/josebda/windows-server-2012-r2-which-version-of-the-smb-protocol-smb-1-0-smb-2-0-smb-2-1-smb-3-0-or-smb-3-02-are-you-using>

³ <https://support.oracle.com/oip/faces/secure/km/DocumentDisplay.jspx?id=2171587.1>

説明

Oracle ACFS NAS MAXでは、Oracle ACFS、Oracle ADVM、およびOracle ASMだけでなく、12.1の新しいCRSリソースであるHAVIPおよびExportFSも使用します。ロックありのHA-NFSでは、12.2の新しいリソースであるNetStorageServiceリソースを使用します。

HAVIPリソースは、標準のOracleノードVIPリソースの特殊クラスです。このリソースは、クラスタ内の単一ノードに単一のIPアドレスを提供します。必要に応じてクラスタ内を移動して、エクスポート・ファイル・システム用のクライアント側インタフェースを提供します。HAVIPがノード上で適切に動作するためには、1つまたは複数の構成済みエクスポートが必要です。クラスタ内に他のHAVIPが構成されている場合、これらのHAVIPはクラスタ内の別のノードで動作するように試みます。12.2.0.1以降では、`-homenode`オプションを使用し、HAVIPを実行する優先ノードを指定できます。この場合、優先ノードが利用可能になると、HAVIPが優先ノードに戻るように試みます。

ExportFSリソースは、OSがNFSを介してファイル・システムをエクスポートするようにします。このリソースは、すべてのノードで利用できるように構成されたOracle ACFSファイル・システムで実行する必要があります。いずれかのノードでファイル・システムが利用できなくなった場合、このリソースは別のノードに移動し、ファイル・システムをエクスポートして、そのノードで引き続きエクスポート・サービスを提供します。ExportFSリソースがHAVIPリソースに接続され、HAVIPがクライアントにエクスポートを提供できるようになります。HAVIPの配置によって、エクスポートの提供元となるOracle RACノードが決まります。

新しいNetStorageServiceリソースは、OS NFSサーバーを管理します。このリソースは、ロックありのHA-NFSが有効になっている場合のみ利用可能です。ロック・モードのHA-NFSでは、Oracle RACクラスタ内のすべてのノードでNFSサービスが実行されるのではなく、クラスタ内の単一のNFSサーバーが実行されます。このNFSサーバーは、HAVIPリソースとともにクラスタ内を移動します（複数構成されている場合を含む）。この処理面の違いは意図的なもので、そのロック方法のためです。また、ロックのリカバリ猶予期間はOS NFSサーバーが処理します。

Oracle ACFS NAS MAXのクライアント使用

HA-SMBとHA-NFSを使用する場合、HAVIPリソースとExportFSリソースは単一グループとしてCRSによって処理されます。ロックありのHA-NFSを利用しない場合、高可用性のエクスポートを設定するときに、以下の簡単な経験則が役立ちます。HAVIPは、依存するファイル・システム（エクスポートを目的としたファイル・システム）の最大数が利用可能であり、かつ他のHAVIPが最少数であるノードで動作します。これにより、次の2つの経験則が成り立ちます。

1. HAVIPは、（接続されるExportFSリソースでの構成に応じて）エクスポートすることが想定されているファイル・システムの最大数が利用可能であることを確認する必要があります。HAVIPは、利用可能なファイル・システムがもっとも多いノードで動作するように試みます。
2. HAVIPは、自身を他のHAVIPサービスからクラスタ全体に分配するように試みます。これにより、最小レベルのロードバランシングが実行されるようにし、HAVIPが単一ノードに集中するのを防ぎます。優先のhomeノードはこのロードバランシングを無視するため、複数のHAVIPが単一のOracle RACノードで実行される可能性があります。

これらの2つの経験則を踏まえて、HANFSで最大限のスケーラビリティと可用性を実現するための一連のガイドラインを作成できます。

1. 最大限のスループットが必要なカテゴリと最大限の可用性が必要なカテゴリの2つのカテゴリに、エクスポートを分類する必要があります。
2. 接続されるすべてのExportFSリソースでの推定スループットが他のHAVIPとおおよそ同じになるように、HAVIPを構成する必要があります。

最大限の可用性が必要なエクスポートは、固有のHAVIPを割り当てるように構成する必要があります。

ロックありのHA-NFSを有効にしている場合、利用可能なHAVIPの数に関係なく、すべてのNFS共有が同じOSノードで動作します。

12.2.0.1 SRVCTLコマンドの説明

各種コマンドの全文は、『Oracle Clusterware管理およびデプロイメント・ガイド』の「付録G：SRVCTLのコマンド・リファレンス」に記載されています。または、`srvctl`コマンドを実行して`help`オプションを使用すると、最新の使用方法が表示されます。

システムに新しいHAVIPを追加するには、コマンド'`srvctl add HAVIP`'を使用します。このHAVIPをHA-NFSとHA-SMBの両方に使用できます。重要なオプションには、以下のようなものがあります。

-id	このHAVIPを識別するために使用する一意のIDです。このIDがリソース名になります。以降、このリソースのステータスを問い合わせる際にも使用されます。
-address	HAVIPによって使用されるIPアドレスまたはホスト名です。このアドレスは、非DHCP、非ラウンド・ロビンのDNSアドレスである必要があります。
-homenode	このHAVIPの推奨ノードです。ノードが稼働中で利用可能である場合、このHAVIPが優先ノードで動作します。このリソース移動により、他のHAVIPが優先ノードから移動するとこれらのHAVIPが自動的に再構成されるか、またはクラスタ内の単一ノードに複数のHAVIPが配置される場合があります。

ExportFSリソースを管理するには、`srvctl * exportfs`コマンド・ファミリーを使用します。ExportFSの作成時に、'`srvctl add exportfs`'を使用します。

-name	ExportFSの一意の識別子です。後の <code>status</code> コマンドで表示されます。
-path	エクスポートでエクスポートを実行するパス。特定のオペレーティング・システムでは、NFSの仕様の解釈が異なります。そのため、有効なパスはOSによって異なります。たとえば、Solarisでは、エクスポート済みのディレクトリのサブディレクトリのエクスポートは許可されません。この場合、ExportFSを開始すると、ユーザーにエラー・メッセージが表示されます。また、Solarisの場合は、ACFSスナップショットのエクスポートについても考慮する必要があります。すべてのスナップショットの親がエクスポートされます（リソース自体がこのエクスポートを処理します）。
-id	ExportFSの接続先のHAVIP IDです。
-options	NFSサーバーまたはSMBサーバーに渡されるオプションです。たとえば、RWまたはROの設定、さまざまなセキュリティ設定、ファイル・システムID、その他のOS固有の属性をここに指定できます。オプションが指定されていない場合は、'RO'が使用されます。詳しくは、「オプション」のセクションを参照してください。
-type	このタイプ（NFSまたはSMB）で、エクスポートのタイプが決まります。両方のタイプを同じHAVIPで使用できます。
-clients	(Linuxのみ) Linuxで、サブネット、IP、ホスト名などのさまざまなクライアント指定子をここに指定できます。これらのクライアントの検証は行われません。また、'*'が有効で、すべてのクライアントを意味します。

サーバー側のエクスポート・オプション

特定のエクスポートのオプション引数の扱いは、OSおよびプロトコル・タイプによって異なります。

- NFS
 - NFSでは、有効なNFSオプションをここに指定できます。
 - Linuxのみ – クライアントをここに指定しないでください。-clientオプションでクライアントを指定します。
 - SolarisおよびAIXの場合、これらのプラットフォームの標準のNFSオプションを使用してクライアントとセキュリティを指定してください。
 - オプションが指定されていない場合、NFSサーバー構成のデフォルト・オプションが使用されます。オプションが指定されていない場合は、'ro'がデフォルト・オプションです。
- SMB
 - SMBでは、クライアントをここに指定する必要があります。
 - サポートされるすべてのオプションをここに指定できます。また、カンマ区切りのキー/値ペアとして渡す必要があります。例：`"share=share1,group=accessgroup1,description="RW,HR,Accessgroup1"`
 - Windows - 任意のnet shareコマンド・オプションをここで使用できます。オプションが指定されていない場合は、/GRANT:everyone,READが使用されます。
 - Linux、Solaris、AIX - 使用するSambaバージョンでサポートされている任意のsmb.conf共有設定パラメータをここに指定できます。オプションが指定されていない場合は、`"read only = yes, browsable = yes"`が使用されます。

クライアント側のマウント・オプション

• HA-NFS

Oracle ACFS HA-NFSエクスポートをクライアントにマウントする場合、以下のオプションを推奨します。

- **hard** – NFSクライアントに対して、サーバーへの接続を試みて操作の再試行を続行するように指示します。これは、ソフト・マウントとは異なります。ソフト・マウントでは、ファイル・システムが利用できなくなるとすぐにエラーが返されます。ハード・マウントでは、ファイル・システムの状態を判断する前に、ファイル・システムが再度利用できるようになるまで待機します。
- **intr** – hardと併用すると、クライアントに対するNFS操作を中断できます（^Cや他のSIGINTなど）。これにより、ユーザーはハングしているように見える操作を終了できます。
- **noLOCK** – HANFSではNFSv3がサポートされています。このバージョンのNFSでは、すべてのプラットフォームでどのシナリオでも、ロックとロック・リカバリは適切に処理されません。そのため、安全のために、ロックが適切に機能しているとアプリケーションが認識しないように、ロック操作を禁止することを推奨します。これは、NFSサーバー固有です（注 - ロックありのHANFS v4を使用する場合は、必要ありません）。
- **Retrans=10000** – NFSクライアントに対して、障害発生時に操作を数回試行するように指示します。操作は通常、非常に高速であるため、これにより、ネットワークの異常とHAVIPのフェイルオーバーが解消されます。
- **Vers** – NFSv4ロックを利用する場合、vers=4を使用してクライアントでマウントするようにします（または、これをシステム・デフォルトとして設定します）。
- **その他のオプション** – wsizeやrsizeなどのその他のオプションは、NFSのパフォーマンス特性に大きく影響する可能性があります。個々の環境や使用方法に推奨されるNFSオプションについては、システム管理者に問い合わせるとともに、アプリケーションのドキュメントを参照してください。Oracle Applicationsについては、MOS Note 359515.1とMOS Note 384248.1、および使用するアプリケーション固有のドキュメントを参照してください。

クライアントにファイル・システムをマウントするための基本的な構文は以下のとおりです。

```
mount -t nfs <export HAVIP>:<export mount point> -o <options>
```

この構文は、使用するOS、アプリケーション、特定のネットワーク設定に応じて異なる場合があります。不確かな場合は、具体的なオプションについてシステム管理者に確認することを推奨します。

- **HA-SMB**

クライアントでOracle ACFS HA-SMBエクスポートをマウントする場合、指定するオプションはありません。オプションはサーバーで構成します。また、Oracle ACFSのHA-NFSと同様に、これらのオプションはパフォーマンスと信頼性に大きく影響します。必要なオプションについては、管理者に問い合わせるとともに、アプリケーションのドキュメントを参照してください。

NFSとサーバーの基本的セットアップ

この簡単なシナリオでは、単一サーバーにHA-NFSとHA-SMBの両方をセットアップします。使用するプラットフォームとインストールの構成手順については、システム管理者に問い合わせるとともに、ベンダーのOSドキュメントを参照してください。

このウォークスルーは、Oracle Linux 7で行っています。まず、このOSバージョンに必要なデーモンが実行されていることを確認します。以下を使用すると、各ノードでこれを確認できます。

- **# /bin/systemctl status rpcbind**

```
rpcbind.service - RPC bind service
```

```
Loaded: loaded (/usr/lib/systemd/system/rpcbind.service; static)
```

```
Active: active (running) since (sample date); 1 day 21h ago
```

```
Main PID:813 (rpcbind)
```

```
CGroup: /system.slice/rpcbind.service
```

```
??813 /sbin/rpcbind -w
```

```
Oct 12 18:30:12 host1 systemd[1]:Started RPC bind service.bash-3.2
```

- **# /bin/systemctl status nfs**

```
nfs-server.service - NFS server and services
```

```
Loaded: loaded (/usr/lib/systemd/system/nfs-server.service; enabled)
```

```
Active: active (exited) since (sample date); 1 day 21h ago
```

```
Main PID:1389 (code=exited, status=0/SUCCESS)
```

```
CGroup: /system.slice/nfs-server.service
```

- **[root@host1 bin]# systemctl status smb**

```
smb.service - Samba SMB Daemon
```

```
Loaded: loaded (/usr/lib/systemd/system/smb.service; disabled)
```

```
Active: active (running) since (sample date); 3s ago
```

```
Main PID:5206 (smbd)
```

```
Status: "smbd: ready to serve connections..."
```

```
CGroup: /system.slice/smb.service
```

```
??5206 /usr/sbin/smbd
```

```
??5215 /usr/sbin/smbd
```

```
Sample Date 16:21:22 host1.us.oracle.com systemd[1]:Starting Samba SMB Daemon...
Sample Date 16:21:22 host1.us.oracle.com smbd[5198]: (sample date)...
Sample Date 16:21:22 host1.us.oracle.com smbd[5198]: standard input is not a s...
Sample Date 16:21:22 host1.us.oracle.com systemd[1]: smb.service:Supervising ...
Sample Date 16:21:23 host1.us.oracle.com smbd[5206]: (sample date)...
Sample Date 16:21:23 host1.us.oracle.com systemd[1]:Started Samba SMB Daemon.
Hint:Some lines were ellipsized, use -l to show in full.
```

いずれかのデーモンが実行されていない場合は、以下を使用してデーモンを起動できます。

```
/sbin/systemctl start <service>
```

'systemctl'コマンドを使用すると、ブート時にこれらのサービスが起動されることを確認できます。

- `bash-3.2# /bin/systemctl enable nfs`
- `bash-3.2# /bin/systemctl enable rpcbind`
- `bash-3.2# /bin/systemctl enable smb`

SELinuxが構成されている場合、NFSファイル・システムおよびSMBファイル・システムで、システム自体にSELinuxのセットアップが正しく構成されていることを確認します。NFSおよびSMBのネットワーク・アクセスを許可できるモードで実行する必要があります。通常、これには'enforcing - targeted'または'permissive'を使用します。

エクスポートする初期ファイル・システムの作成

初期セットアップが終了したので、ファイル・システムを構成できます。Oracle ACFS NAS MAXには、単一のファイル・システム・リソースを使用してすべてのノードにマウントするように構成された、Oracle ACFSファイル・システムが必要です。これらのコマンドについて、詳しくは『Oracle Automatic Storage Management管理者ガイド』を参照してください。これを行うには、いくつかの方法があります。

- ASMCAの使用：
 - 「**Volumes**」ツリー・エントリをクリックします。
 - ペインの下部にある「**Create**」ボタンをクリックします。
 - プロンプトに従います。
 - 「**ACFS File Systems**」ツリー・エントリをクリックします。
 - ペインの下部にある「**Create**」ボタンをクリックします。
 - “Cluster File System”が“Type of ACFS”として指定されていることを確認します。

- “Auto Mount”が選択されていることを確認します。
- 「OK」をクリックしてスクリプトを実行するか、ASMCAでコマンドを実行します。

- コマンドライン :
 - asmcmdを使用してボリューム・デバイスを作成します。
 - 'mkfs'を使用してボリューム・デバイスをフォーマットします。
 - 'srvctl add filesystem -device <device> -path <mount path>'を使用してcrsでファイル・システムを登録します。
 - 'srvctl start filesystem -device <device>'を使用してパスをマウントします。

ファイル・システムの作成例

```
bash-3.2# /sbin/mkfs -t acfs /dev/asm/test1-194
mkfs.acfs: version                = 21.1.0.0.0
mkfs.acfs: on-disk version        = 39.0
mkfs.acfs: volume                 = /dev/asm/test1-194
mkfs.acfs: volume size            = 5368709120
mkfs.acfs:Format complete.
bash-3.2# mkdir /hr1
bash-3.2# /scratch/crs_home/bin/srvctl add filesystem -path /hr1 -device /dev/asm/test1-194
bash-3.2# /scratch/crs_home/bin/srvctl start filesystem -device /dev/asm/test1-194
bash-3.2# mount -t acfs
/dev/asm/test1-194 on /hr1 type acfs (rw)
bash-3.2# /scratch/crs_home/bin/srvctl status filesystem -device /dev/asm/hr1-194
ACFS file system /hr1 is mounted on nodes host1,host2
```

シンプルなOracle ACFS HA-NFSシナリオの構成（ロックなし）

このシナリオでは、2ノードOracle RACクラスタのシンプルなセットアップを前提とします。このクラスタから、以下のエクスポートが必要です。

- /hr1 – 想定平均スループット約500 KB/秒。おもに読取り。
- /hr2 – 想定平均スループット約500 KB/秒。おもに読取り。

各ファイル・システムは同じASMディスク・グループにホストされ、フェイルオーバーと外部冗長性は適用されません。管理者は、これらのファイル・システムで可用性の問題を想定していません。また、これらのファイル・システムは同じディスク・グループに配置されるため、ストレージの停止が両方のファイル・システムに同じように影響すると考えられます。両方のスループットを合わせても非常に小さいため、単一ノードで両方のエクスポートをホストする場合でも、ネットワーク帯域幅の問題はないと考えられます。

そのため、シンプルにするために、管理者はHAVIPを1つのみ作成して、両方のエクスポート・ファイル・システムをこのVIPに接続することを選択しました。これにより、すべてのクライアントが単一のHAVIPアドレスを使用して両方のマウント・ポイントにアクセスします。この場合、Oracle RAC HANFSクラスタに新しいノードを追加すると、クラスタ全体のスループットが自動的にスケールされないというマイナス面があります。

HAVIPの登録 :

```
bash-3.2# /scratch/crs_home/bin/srvctl add HAVIP -id HR1 -address HAVIP1.us.oracle.com - netnum 1 -description "HR
specific exports for the Omega Project"

bash-3.2# /scratch/crs_home/bin/srvctl status HAVIP -id HR1

HAVIP ora.hr1.HAVIP is enabled

HAVIP ora.hr1.HAVIP is not running

bash-3.2# /scratch/crs_home/bin/srvctl start HAVIP -id HR1

PRCR-1079 :Failed to start resource ora.hr1.HAVIP

CRS-2805:Unable to start 'ora.hr1.HAVIP' because it has a 'hard' dependency on resource type 'ora.HR1.export.type' and no
resource of that type can satisfy the dependency
```

障害が発生するのはなぜでしょうか。前述したように、HAVIPには1つまたは複数の構成済みのExportFSが必要です。ExportFSがないと、HAVIPは開始されません。クライアントがExportFSをマウントしており、使用可能なExportFSがない状態でHAVIPが開始された場合、クライアントはESTALEエラーを受け取ります。そのため、クライアントでNFSサービスが再開されません。

ExportFSの登録

```
bash-3.2# /scratch/crs_home/bin/srvctl add exportfs -path /hr1 -id HR1 -name HR1 -options "rw,no_root_squash" -
clients agraves-vm5,agraves-vm6

bash-3.2# /scratch/crs_home/bin/srvctl status exportfs -name HR1

export file system hr1 is enabled
export file system hr1 is not exported

bash-3.2# /scratch/crs_home/bin/srvctl add exportfs -path /hr2 -id HR1 -name HR2 -options "ro" -clients
10.149.236.0/22
```

この時点で、ExportFSまたはHAVIPを開始すると、このHAVIPの構成済みのExportFSすべてが開始されるか、または関連付けられたHAVIPが開始されます。

ここでは、2番目のExportFSであるHR2を、ネットワーク・リソースのサブネットにのみエクスポートすることを選択しました。

```
bash-3.2# /scratch/crs_home/bin/srvctl config exportfs -name HR2

export file system hr2 is configured

Exported path: /hr2

Export Options: ro

Configured Clients:10.149.236.0/22

bash-3.2# /scratch/crs_home/bin/srvctl config exportfs -name HR1

export file system hr1 is configured

Exported path: /hr1

Export Options: rw,no_root_squash

Configured Clients: host1,host2
```

2つのクライアント（host1とhost2）でのみ利用可能なHR1と比較します。

HAVIPからその他のリソースへの依存関係が構成されていることが分かります。

```
START_DEPENDENCIES=hard(ora.net1.network,uniform:type:ora.HR1.export.type)
attraction(ora.data.hr1.acfs,ora.data.hr2.acfs) dispersion:active(type:ora.HAVIP.type) pullup(ora.net1.network)
pullup:always(type:ora.HR1.export.type)
STOP_DEPENDENCIES=hard(intermediate:ora.net1.network,uniform:intermediate:type:ora.HR1.export
.type)
```

これらの依存関係により、HAVIPはExportFSリソースおよびOracle ACFSリソースの後に開始され、これらのリソースの前に停止されます。エクスポートを開始するには、いくつかの方法があります。

- `srvctl start exportfs -id <ID> - id <ID>`のHAVIPに接続されているすべてのエクスポートが開始されます。
- `srvctl start HAVIP -id <ID> - id <ID>`のHAVIPに接続されているすべてのエクスポートが開始されます。
- `srvctl start exportfs -name <NAME> - <NAME>` ExportFSおよびそのHAVIPのみ開始されます。

```
bash-3.2# /scratch/crs_home/bin/srvctl start exportfs -id HR1
bash-3.2# /scratch/crs_home/bin/srvctl status exportfs
export file system hr1 is enabled
export file system hr1 is exported on node host2
export file system hr2 is enabled
export file system hr2 is exported on node host2
bash-3.2# /usr/sbin/exportfs -v
/hr1
host1.us.oracle.com(rw,wdelay,no_root_squash,no_subtree_check,fsid=128850576,anonuid=65534,anongid=65534)
/hr1
host2.us.oracle.com(rw,wdelay,no_root_squash,no_subtree_check,fsid=128850576,anonuid=65534,anongid=65534)
/hr2
10.149.236.0/22(ro,wdelay,root_squash,no_subtree_check,fsid=1573414370,anonuid=65534,anongid=65534)
```

ここで、適切なオプションを使用してエクスポートが適切なクライアントにエクスポートされることを明確に確認できます。Linuxでは、同じファイル・システムまたは同じファイル・システムのディレクトリの複数エクスポートが可能です。他のOS NFS実装では実装が異なるため、親ディレクトリと子ディレクトリのエクスポートが制限される場合があります。Solarisでは、スナップショットのエクスポート時に、ファイル・システム・マウント・ポイントの親（ルート）がエクスポートされます。

HA-SMBエクスポートの追加

ここで、これにHA-SMBエクスポートを追加してみます。まず、同じファイル・システムに新しいマウントを追加します。
[root@host1 bin]# ./srvctl add exportfs -path /hr1/smb -id HR1 -name smbexport -type SMB
HA-NFSエクスポートと同じHAVIPにマウントを追加したことに注意してください。両方のエクスポートがクラスタ内を移動します。

```
[root@host1 bin]# ./srvctl status exportfs -id HR1
export file system hr1 is enabled
export file system hr1 is exported on node host2
export file system smbexport is enabled
```

HAVIPはすでに存在するため、新しいエクスポートを明示的に開始する必要があります。以降は、HAVIPが開始されると、HA-NFSエクスポートでエクスポートが開始されます。

```
[root@host1 bin]# ./srvctl start exportfs -name smbexport
[root@n bin]# ./srvctl status exportfs -id HR1
export file system hr1 is enabled
export file system hr1 is exported on node host2
export file system hr2 is enabled
export file system hr2 is exported on node host2
export file system smbexport is enabled
export file system smbexport is exported on node host2
```

HA-SMBのSamba構成

HA-SMBエクスポートを実行したら、Samba構成のバックエンド構造を確認できます。

```
[root@host2 ~]# testparm -s
Load smb config files from /etc/samba/smb.conf
rlimit_max: increasing rlimit_max (8192) to minimum Windows limit (16384)
Processing section "[smbexport]"
Processing section "[homes]"
Processing section "[printers]"
Loaded services file OK.
Server role:ROLE_STANDALONE
[global]
    workgroup = MYGROUP
    server string = Samba Server Version %v
    log file = /var/log/samba/log.%m
    max log size = 50
    idmap config *: backend = tdb
    cups options = raw
```



```
[smbexport]
    path = /hr1/smb
[homes]
    comment = Home Directories
    read only = No
    browseable = No
[printers]
    comment = All Printers
    path = /var/spool/samba
    printable = Yes
    print ok = Yes
    browseable = No
```

この情報はどこから得られるのでしょうか。新しいエクスポートごとに個別のインクルード・ファイルが生成されて、これらのファイルが /etc/samba/acfs/ に格納されます。この情報は、システムの新しい smb.conf にはリンクされない、単一の新しいインクルード・ファイル acfsinc.conf に格納されます。

ノードの再配置

ここで、ノードの再配置を行う必要があるとします。ノードの再配置に使用するコマンドは、'srvctl relocate HAVIP' です。

```
bash-3.2# /scratch/crs_home/bin/srvctl relocate HAVIP -id HR1 -node host1
bash-3.2# /scratch/crs_home/bin/srvctl status HAVIP -id HR1
HAVIP ora.hr1.HAVIP is enabled
HAVIP ora.hr1.HAVIP is running on node host1
```

リソース状態を特定できるコマンド（crsctl、srvctl）を使用して、リソースが新しいノードで実行されていることを確認できます。

```
bash-3.2# /scratch/crs_home/bin/crsctl stat res -w "TYPE = ora.HR1.export.type" -t
```

Name	Target	State	Server	State details
Cluster Resources				
ora.hr1.export 1	ONLINE	ONLINE	host1	STABLE
ora.hr2.export 1	ONLINE	ONLINE	host1	STABLE

ネットワーク障害、ストレージ障害、ノード全体の障害などの計画外停止の場合に、同じ原則が当てはまります。

ロックありのHA-NFS V4のセットアップ

ここでは、先に進めて、ロック・サポートありのHA-NFS v4について説明します。これは12.2.0.1で導入された新しいモードで、標準のHA-NFSモードやHA-SMBモードとは動作がやや異なります。このモードで動作している場合、Oracle ACFS NAS MAXのすべてのエクスポートが同じクラスタ・ノードで動作します。これは、基盤となるNFSサーバーによるロック処理のためです。これらのロックに対してフェイルオーバーとリカバリのセマンティクスを提供するために、NFSサーバーはロック・リカバリ時間の猶予時間を実装する必要があります。ロック状態は永続的な場所に保存され、サーバー起動時にこの状態の問合せが行われます。その間、新しいロックと操作は許可されません（これは猶予期間と呼ばれ、OSのパラメータで調整できます）。状態がメモリに読み取られると、クライアントがそれぞれのロック状態で応答できるようになり、通常の操作が再開されます。

クライアント・ノードで、NFSバージョンとしてNFS v4を指定してファイル・システムをマウントします。これにより、サーバーがデフォルトのNFS v3に設定されないようにし、NFS v4のロック機能のサポートを有効にします。ロックありのOracle ACFS HA-NFSを有効にするには、次の手順を実行します。

1. HANFS永続ストレージで使用する新しいボリュームを作成します。Oracle ADVMボリュームに関する制限事項は、以下のとおりです。
 1. この新しいOracle ADVMボリュームに既存のOracle ACFSリソースを配置できません。
 2. このOracle ADVMボリュームにOracle ACFSファイル・システムを配置できません。
 3. このOracle ADVMボリュームをクラスタ内で使用できません。
2. addnode操作を実行します。

```
bash-4.2# ./acfsnfs addnode -nfsv4lock -volume /dev/asm/hanfs-48
ACFS-9603:The script will do the following actions:
ACFS-9604: - Update the operating system startup procedure so that NFS does not automatically start.
ACFS-9605:      Management of the NFS daemons will be moved to Oracle Clusterware.
ACFS-9606: - Format the volume: /dev/asm/hanfs-48.
ACFS-9607: - Create an ACFS resource for the file system.
ACFS-9608: - Mount the ACFS file system on '/var/lib/nfs'.
ACFS-9609:Continue the installation? [1=yes,2=no]:

ACFS-9612:Stopping NFS Service.
Redirecting to /bin/systemctl stop nfs.service
ACFS-9179:Command executed: '/sbin/service nfs stop', output = '0'
ACFS-9179:Command executed: '/install/usb161118/bin/srvctl start filesystem - device /dev/asm/hanfs-48 -n host1.us.oracle.com', output = '2'
ACFS-9611:Starting NFS Service.
Redirecting to /bin/systemctl start nfs.service
ACFS-9179:Command executed: '/sbin/service nfs start', output = '0'
ACFS-9203: true
```

3. 新しくマウントされたNFSファイル・システムを確認します。OS NFSディレクトリにマウントされるため、クラスタ内のすべてのノードにこの情報が表示されます。この新しいACFSリソースは変更しないでください。変更すると、NFSロックが機能しなくなります。

```
bash-4.2# mount -t acfs
/dev/asm/hanfs-48 on /var/lib/nfs type acfs (rw,relatime,device,rootsuid,ordered)
```

4. ノードごとに、上記の手順を繰り返します。セットアップ後に追加されたノードには、この手順が自動的に実行されます。

これが完了すると、新しいリソースora.netstorageserviceを試すことができます。このリソースは、NFSのOSサービスを管理します。NFSエクスポートがクラスタ内を移動すると、各マシンがNFSサーバーを起動および停止します。起動時に、猶予期間がトリガーされて、OS NFSサーバーがフェイルオーバーとロック・リカバリを適切に処理できるようになります。

OS NFSサーバーはクラスタ内を移動して再起動されるため、Oracle ACFS HANFSエクスポートとOracle ACFS以外のエクスポートを混在することは推奨されません。OS NFSサーバーが再起動されると、Oracle ACFS以外のエクスポートへのアクセスが失われます。

新しいnetstorageserviceリソースがオンラインになっています。

```
NAME=ora.netstorageservice
TYPE=ora.netstorageservice.type
TARGET=ONLINE
STATE=ONLINE on host1
```

ここで、新しいエクスポートを追加してみます。ここでは、プロセスは前と同じで、srvctlを使用します。

```
bash-4.2# ./srvctl add HAVIP -address host1-HAVIP1 -id HAVIP1
bash-4.2# ./srvctl add exportfs -name export1 -id HAVIP1 -path /mnt/oracle/tfa
bash-4.2# ./srvctl status exportfs
export file system export1 is enabledexport file system
export1 is not exported
bash-4.2# ./srvctl start HAVIP -id HAVIP1
bash-4.2# ./srvctl status exportfs
export file system export1 is enabled
export file system export1 is exported on node host1
```

この新しいエクスポートで、依存関係を確認できます。ロックを有効化するようにOracle ACFS HA-NFS操作を変換するときに、HA-NFSリソースも変更されます。

```
START_DEPENDENCIES=hard(ora.mgmthost1.tfarepos.acfs,ora.netstorageservice)
attraction(ora.HAVIP1.HAVIP) dispersion(type:ora.export.type)
pullup(ora.mgmthost1.tfarepos.acfs,ora.netstorageservice)
```

依存関係チェーンが存在します。この依存関係チェーンによって、リソースがNFSサーバーと同じノードで実行されます。

複雑なシナリオ

ここでは、より複雑なシナリオについて説明します。管理者がHANFSに6つのマウント・ポイントを選択しました。

- /hr1 – 500 Kb/秒
- /hr – 10 Mb/秒 – 常に利用可能であることが必要
- /source – 100 Kb/秒 – 常に利用可能であることが必要
- /PDF – 10 Mb/秒
- /Games – 100 Mb/秒
- /Media – 1 Mb/秒

高可用性要件のために、この場合の最適な構成は次のようになります。

- /hr1、/PDF、および/Mediaを単一のHAVIPアドレスで構成
- /Gamesを単一のHAVIPアドレスで構成
- /sourceを単一のHAVIPアドレスで構成
- /hrを単一のHAVIPアドレスで構成

根拠

- /Gamesを固有のHAVIPアドレスに配置すると、/Gamesの高スループットが他のHAVIPから分離され、CRSがこのHAVIPおよび関連するExportFSを他のHAVIPとは別の固有のサーバーに配置できる可能性があります（クラスタ内に十分なサーバーがあることが前提）。
- /sourceを固有のHAVIPアドレスに配置すると、ストレージ接続の問題が発生した場合に、CRSがファイル・システムに対応できるクラスタ・メンバーに/sourceを移動できます。このHAVIPには1つのExportFSのみ存在するため、CRSはOracle ACFSファイル・システムが利用可能なノードを検出するだけでよく、ポリシーの決定やトレードオフは必要ありません。
- /hrを固有のHAVIPに配置すると、/sourceソースと同じロジックを/hrに適用できます。
- 残りをそれぞれ固有のHAVIPに配置すると、必要なIPアドレスの数が減ります。ごくまれですが、クラスタの各ノードで別のファイル・システムが利用可能になった場合には、CRSは最適と判断したノードにHAVIPを配置します。これにより、1つのファイル・システムが利用できなくなります。
- 同じことがHA-SMBエクスポートとHA-NFSエクスポートの両方に当てはまります。

CRSのポリシー例 – 最適なノードの選択

次のクラスタについて考えてみます。

ノード1 – 利用可能なファイル・システム : /fs1 /fs2 /fs3

ノード2 – 利用可能なファイル・システム : /fs1 /fs3

ノード3 – 利用可能なファイル・システム : /fs2 /fs3 /fs4

単一のHAVIPで4つのすべてのファイル・システムをエクスポートする場合、CRSはファイル・システムのエクスポート元とするのに最適なノードに関して方針を決定します。4つのすべてのファイル・システムが利用可能であるという点を完全に満たしているノードはありません。そのため、CRSはノード1またはノード3がHAVIPおよび関連するExportFSに最適な場所であると判断します。どちらのノードを選択した場合でも、ストレージ接続の問題のために、1つのファイル・システムが利用できなくなります。

どのような状況で、HAVIPとExportFSが他のノードに移動するのか

1. サーバー・クラスタ・メンバーシップ変更イベント（ノードのクラスタ離脱やクラスタ参加など）により、HAVIPが分配を再評価して、HAVIPとExportFSリソースをノード間で移動する可能性があります。その間、クライアントにはサービスの一時停止が表示されますが、エクスポートが再確立される（通常、3秒以下）とすぐに、サービスの中断がないかのようにクライアントが動作を続けます。
2. ごくまれですが、ストレージ障害が発生し、サーバーの特定ノードでファイル・システムにアクセスできなくなった場合、HAVIPは必須のすべてのファイル・システムに対して特定ノードがアクセスを改善できるかを評価します。改善できる場合は、HAVIPがそのノードに移動するため、利用可能なファイル・システムがもっとも多いノードが、HAVIPが配置されるノードになります。これにより、最大数のクライアントが最大数のファイル・システムに常にアクセスできるようになります。
3. 管理者が移動を必要とする場合、12.1の新しいコマンドを使用し、管理者はsrvctlコマンドで計画的な再配置を実行して、HAVIPおよび関連するExportFSリソースをクラスタ内の別のノードに強制的に移動できます。これはノードの計画停止の際に役立ちます。
4. クラスタ・メンバー固有のネットワーク接続に問題が発生した場合、クラスタ・メンバーはクラスタ名前空間から削除されて、HAVIPおよび関連するExportFSリソースは接続されているノードに移動します。
5. HAVIPに優先ノード・セットがある場合は、このノードがクラスタに再度参加すると、HAVIPがこのノードに戻ります。
6. ロックありのHA-NFSを使用する場合、現在ホスト中のノードでNFSサーバーに障害が発生すると、別のノードでサービスが再起動されます。この場合、一時停止状態が通常より長くなる可能性があります。他の場所でリソースを再起動するのにかかる時間に加えて、NFSサーバー自体でロック・リカバリのタイムアウト（通常、90秒間）が発生します。これにOSデフォルトを使用すると、かかる時間が短くなる（または長くなる）可能性があります。これを行う場合は注意が必要です。

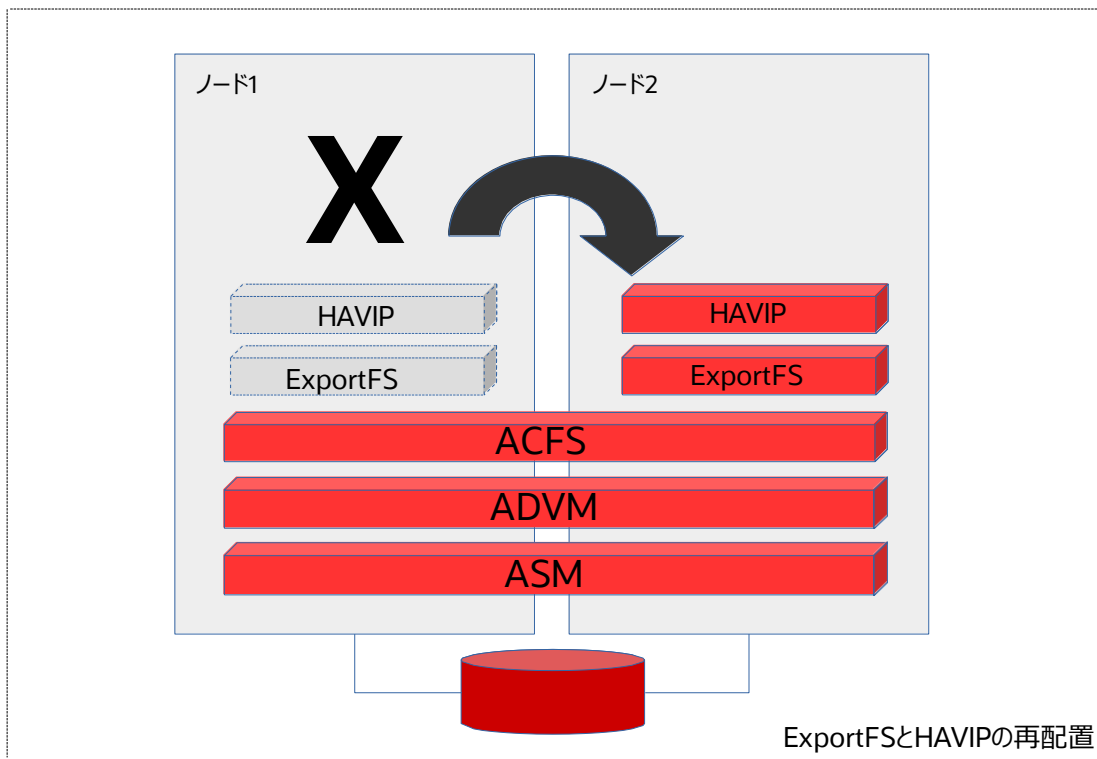
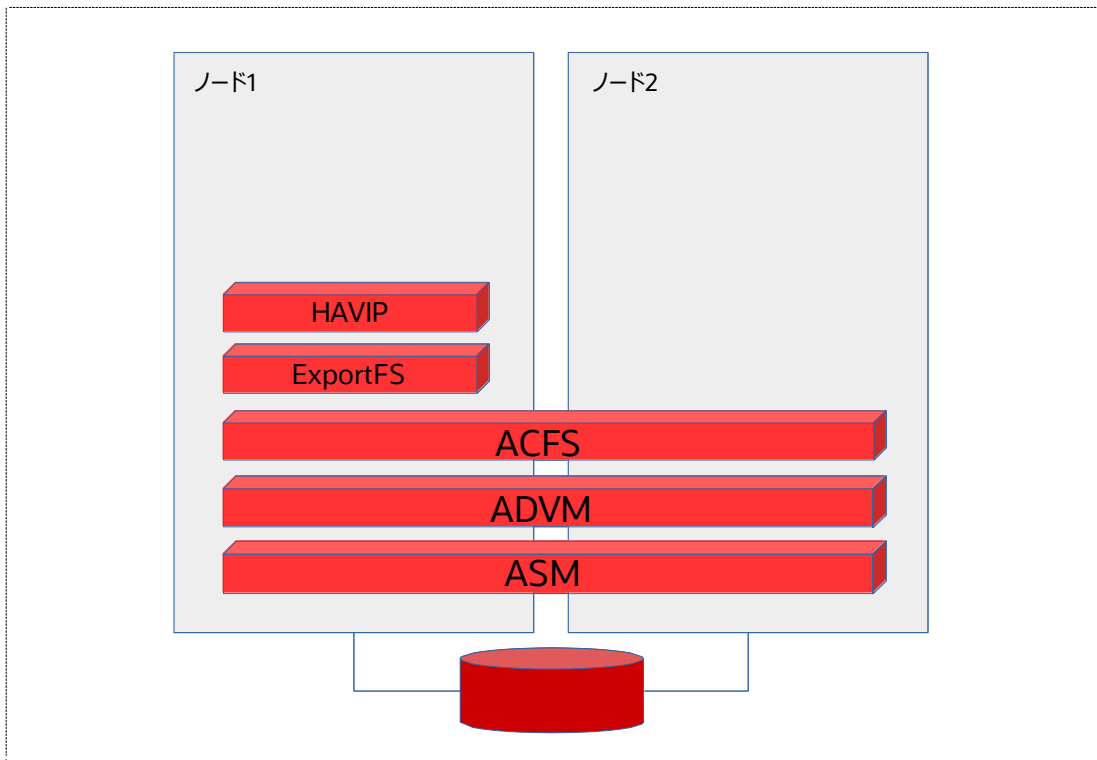


図2 - ExportFSとHAVIPの再配置

HAVIPのフェイルオーバー時間

ノードのフェイルオーバーを検討する際、次の2つのケースを考慮する必要があります。

- 計画的再配置
- ノード障害

計画的再配置

管理者がHAVIPおよび関連するExportFSリソースを別のノードに強制的に移動する場合、次の手順を実行します。

1. 最初のノードで、HAVIPを停止します。
2. 最初のノードで、関連するExportFSリソースを停止します。
3. 2番目のノードで、関連するExportFSリソースを起動します。
4. 最初のノードで、HAVIPを起動します。

HAVIPに多数の関連するExportFSリソースが構成されている場合、この計画フェイルオーバーにかかる時間が長くなる可能性があります。ノード1で個々のExportFSをそれぞれ停止するのにかかる時間は、最大で1~2秒です。すべてのExportFSを同時に停止できるだけの十分な処理能力がある場合は、停止がすばやく行われる可能性があります。ただし、CPU処理の制限のために、CRSが各ExportFSを順次停止する必要がある場合は、最悪のケースでは2秒×<ExportFSリソース数>になります。

同じシナリオが起動時間にも当てはまります。そのため、最悪のケースでは、再配置にかかる総時間 = (2秒×<ExportFSリソース数>) × 2 + 5秒 (5秒は、HAVIPの停止と起動にかかる時間) になります。

エクスポートされたファイル・システムは、HAVIPが起動されるまで、つまりこのプロセスが終了するまで、利用できません。

そのため、単一のHAVIPを使用し続ける場合は管理が簡単になりますが、ファイル・システムの処理が再開されてクライアントで利用可能になるまでの時間が長くなる可能性があります。

異なるバージョンのSMBでは、HAフェイルオーバーをサポートできる場合とできない場合があります。クライアントでファイル・システムの再マウントが必要になる可能性があります。

ノード障害

ノード障害も同様ですが、再配置よりも簡単です。CRSがノード障害を検出すると実行されるステップは、HAVIPの起動だけで、関連するすべてのExportFSリソースが起動された後で実行されます。そのため、かかる時間は再配置の時間の半分です。

NFSのパフォーマンスに関する考慮事項

HA-NFSはNFSに基づくため、標準のNFS構成とパフォーマンス・チューニングをOracle ACFS HA-NFS製品に適用できます。rsizeやwsizeなどのクライアント・オプションを使用すると、データ・アクセス速度に大きく差が出る可能性があります。また、クライアントに対するファイル・サーバーの位置も、NFSのパフォーマンスに影響します。通常、最適なパフォーマンスが得られるのは、NFSサーバーをクライアントと併置している場合です。

SMBのパフォーマンスに関する考慮事項

HA-SMBはSamba（Windowsの場合はSMB）に基づくため、標準の構成とパフォーマンス・チューニングを適用できます。Sambaのチューニングについては、<https://www.samba.org/samba/docs/man/Samba-HOWTO-Collection/speed.html>を参照してください。

Oracle ACFS NAS MAXのスケラビリティ – ロックなしのNFSとSMB

Oracle RAC HANFSクラスタのパフォーマンスについて、簡単に説明します。バッキングされたネットワーク・ファブリックが1 Gb/秒のスループットをサポートしているとして、各クラスタ・メンバーがこのネットワークに接続されている場合、単一マシンの論理的な最大スループットは、1 Gb/秒×（Oracle RAC HANFSクラスタ内のノード数）になります。ただし、CRSがHAVIPを移動させるため、単一ノードに複数のHAVIPがホストされている場合、最大スループットは、（1 Gb/秒×Oracle RAC HANFSクラスタ内のノード数） / （単一ノードのHAVIP数）になります。

これが高可用性ニーズ（前述）を満たしている場合、単一クラスタ内のHAVIP数をOracle RAC HANFSクラスタ内のノード数と同じにするとときのパフォーマンス上の利点について、簡単に説明します。各ノードが2つのHAVIPをホストする場合、クラスタ内のノード数を2倍にするだけで、Oracle RAC HANFSクラスタの合計スループットを2倍にできます。

また、単一ノードが2つのHAVIPをホストする場合を考えると、これらのHAVIPでエクスポートされるファイル・システムのパフォーマンスは50 %になると想定されます。バッキングされたストレージおよびネットワーク・ファブリックが想定スループットに妥当であるとすると、Oracle RAC HANFSクラスタに新しいノードを追加するだけで、この状況を改善できます。新しいノードを追加すると、各ノードが2つではなく1つのHAVIPのみをホストするように、CRSがHAVIPを移動させます。

ExportFSリソースの動作

- リソースの他の起動方法
 - 他のさまざまなリソースと同様に、透過的な高可用性によってエクスポート・リソースを起動できます。この場合、ExportFSリソースが、ファイル・システムのエクスポートをシステムがレポートするかどうかを監視します。そのため、'exportfs'（Linux）コマンドを使用してエクスポートをオンラインにできます。これを行った場合、ExportFSリソースが別のノードで実行される可能性があります。この状態は許容可能ですが、推奨されません。関連するHAVIPが起動されるとすぐに、ExportFSリソースは、そのリソースをホストしているノードに移動します。
- システム・ツールを使用したエクスポート・オプションの変更
 - 管理者がシステム・ツールでエクスポート・オプションを変更すると（rwからroへの変更、samba設定ファイルの変更など）、ExportFSリソースは'PARTIAL'状態によってこれを反映します。その際、ExportFSが実行中であることが管理者に通知されますが、構成されたオプションは、そのサーバーで現在エクスポートされているオプションとは異なります。
- THAを使用したExportFSの停止
 - 管理者がexportfsなどのコマンドライン・ツールでエクスポートを手動で削除すると、関連するExportFSリソースもオフラインになります。
- エクスポート対象のディレクトリの削除
 - エクスポート対象として構成されているディレクトリを'rm'で削除すると、関連するExportFSリソースもオフラインになります。

クラスタ内のエクスポートの場所の制御

エクスポートの場所をより詳細に制御することを希望する管理者は、'disable'コマンドと'enable'コマンドを使用し、HAVIPを特定のノードでのみ実行するように構成できます。

たとえば、管理者が2ノード・クラスタの1ノードでHR2とHR1（HAVIP HR1によってエクスポート）のみを実行する必要があるとします。リソースの追加後、管理者は次のように実行してリソースを制限できます。

```
bash-3.2# /scratch/crs_home/bin/srvctl disable havip -node host2 -id HR1
```

その他の検討事項

他のOracle ACFSテクノロジーおよびOracle ASMテクノロジーをOracle RAC HANFS製品に適用するかどうかは、読者にゆだねられます。次のような適用が考えられます。

- Oracle ACFS NAS MAXによってエクスポートされるACFSファイル・システムで、レプリケーションを利用する。このファイル・システムは、別のデータセンターにレプリケートされます。Oracle ACFSのレプリケーションにより、整合性のあるファイル・システムが別のデータセンターに提供され、HAVIPのアドレスを別のクラスタのHAVIPのアドレスに変更できます。これは高度にフォルト・トレラントなセットアップです。
- Oracle ACFS NAS MAXによってエクスポートされるOracle ACFSファイル・システムのACFSスナップショットを作成する。これにより、バックアップを作成できます。
- Oracle ACFS NAS MAXによってOracleホームをエクスポートし、Oracleホームを常に使用できるようにする。
- 不正アクセスを防ぐため、特定の期間は読み取り専用になるように、Oracle ACFS NAS MAX ExportFSをOracle ACFSレルムで構成する。
- 基盤ストレージが高度にフォルト・トレラントになるように、ASMディスク・グループにデフォルトの1つより多く障害グループを構成し、Oracle ACFS NAS MAXを使用する。これにより、単一のディスク・グループでストレージ障害が発生する可能性が事実上なくなります。また、Oracle RAC HANFSによってエクスポート自体が常に利用可能になり、高可用性に優れた単一のファイル・サーバーを実現できます。

トラブルシューティング

このホワイト・ペーパーで何度か述べたように、Oracle ACFS NAS MAXは基盤となるOSサーバー構成に大きく依存します。システム管理者と協力し、OSの基本機能を使用してレプリケートすれば、ほとんどのエラーを修正できます。

一般的な問題

- サーバーが応答しない
多くの場合、サーバーに正しいRPCデーモンがありません。この問題はクライアントで発生し、メッセージ“RPC – Program not registered”が表示されます。これはOS依存のメッセージで、OSによって異なる場合があります。これを修正するには、RPCデーモンがクライアントで実行されていることを確認します。

- 一般的なデバッグ・ツール
 - rpcinfo – 特定のサーバーのrpcデーモンに関する情報を表示します。
 - showmount – 特定のサーバーからのNFSエクスポートを表示します。
 - tcpdump – ノード間の正確なトラフィック・フローを表示します。ネットワークでパケットが破棄されている箇所など、問題が発生している箇所を特定するのに役立ちます。
 - traceroute – クライアントからサーバーへのルートが使用可能かどうかを確認します。
 - testparm – Samba設定ファイルを検証するために使用します。

- 複数のNIC – NFSサーバーでは、転送先はそれほど制御されません。NFSサーバーは、クライアント/サーバーへのルートがあると考えられるすべてのインタフェース経由でパケットを転送します。そのため、netstatルーティング・テーブル (netstat -rn) に、使用されているインタフェース経由でサーバー/クライアントへのルートが表示されることを確認することが非常に重要です。
 - また、route -rnおよびrouteコマンドを使用して、トラフィックの送信に正しいインタフェースが使用されていることを確認する必要があります。NFSは、各アドレス範囲にデフォルトのアダプタを使用します。クライアントとサーバーのIP関係にアダプタが定義されていない場合、正しくないネットワーク・アダプタが使用され、権限が拒否されて他の問題も発生する可能性があります。'ip route add'を使用して新しいルートを追加できます。常に、このようなコマンドは注意して使用してください。

- 権限の拒否 – もっとも一般的な問題でしょう。システム・ログで情報が見つかる可能性があります (/var/log/messagesなど)。ファイアウォールでパケットが破棄されている、セキュリティ構成 (SELinuxなど) のためにアクセスが制限されている、ディレクトリの権限が正しくない、ユーザー等価マッピング (root_squashなど) が正しくない、共有へのアクセス権限のないネットワーク・インタフェース経由でのアクセスをクライアントが要求しているなどの可能性があります。

- PowershellコマンドLet-Get-SmbConnectionを使用すると、サーバーまたはクライアントで使用しているSMBのバージョンを確認できます。詳しくは、<https://blogs.technet.microsoft.com/josebda/2012/06/06/windows-server-2012-which-version-of-the-smb-protocol-smb-1-0-smb-2-0-smb-2-1-or-smb-3-0-are-you-using-on-your-file-server/>を参照してください。

結論

このホワイト・ペーパーでは、関連するテクノロジーOracle ACFSおよびOracle Clusterwareの基礎知識を始めとし、Oracle ACFS NAS MAXの構成について説明しました。これらの構成要素の両方が、可用性に優れたOSプロトコルをセットアップするのに役立つMaximum Availability Architectureを実現します。これらは構成要素に過ぎませんが、Oracle ACFS NAS MAXはこれらの構成要素の高可用性を使用することで、基盤となるOSプロトコルをラップし、プロトコルの可用性を確保します。

システム管理者は、これらのテクノロジーをNFSおよびSMBで利用すると、クラスタ内のマシンを再起動する場合やこれらのマシンが利用できなくなった場合でも、単一のOracle Gridクラスタの使用期間全体を通じてクライアントにファイルを提供できる一連のエクスポートを作成できます。このセットアップにより、システム管理者が従来使用していた高コストのNASデバイスに代わる低コストの手段を、さらに高い可用性で実現できます。

構成セクションで示したように、これらのプロトコルのサポートはOS自体に大きく依存します。最新かつ最適なソフトウェアを実行することが常に推奨され、特定のパッケージや更新がシステム・インストールに必要となることが多いです。Oracle ACFS NAS MAXでは、基盤となるプロトコルが正しく構成されていない場合は、システムの動作が保証されません。また、Oracle ACFS NAS MAXインストールの構成時に、特定のNFS機能またはSMB機能が必要かどうかを考慮し、それに応じてシステムを構成する必要があります。

Oracle ACFS NAS MAXの可用性に加えて、Oracle ACFS NAS MAXの機能を拡張するOracle ACFSの他の機能も利用できます。これらの機能について、詳しくはOracle ASMの管理者ガイドおよび個別のホワイト・ペーパーを参照してください。これらの機能の併用と高可用性により、真に統合されたフル機能のソリューションを実現できます。

CONNECT WITH US

+1.800.ORACLE1までご連絡いただくか、oracle.comをご覧ください。
北米以外の地域では、oracle.com/contactで最寄りの営業所をご確認いただけます。

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

Copyright © 2022, Oracle and/or its affiliates. All rights reserved. 本文書は情報提供のみを目的として提供されており、ここに記載されている内容は予告なく変更されることがあります。本文書は、その内容に誤りがないことを保証するものではなく、また、口頭による明示的保証や法律による黙示的保証を含め、商品性ないし特定目的適合性に関する黙示的保証および条件などのいかなる保証および条件も提供するものではありません。オラクルは本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクルの書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

OracleおよびJavaはOracleおよびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

IntelおよびIntel XeonはIntel Corporationの商標または登録商標です。すべてのSPARC商標はライセンスに基づいて使用されるSPARC International, Inc.の商標または登録商標です。AMD、Opteron、AMDロゴおよびAMD Opteronロゴは、Advanced Micro Devicesの商標または登録商標です。UNIXは、The Open Groupの登録商標です。0120

Oracle ACFS 21c – NAS Maximum Availability Extensions 2022年5月

著者：Allan Graves

共著者：Oracle ACFS Product Management Team, Oracle ACFS Development team

