NetBackup™ for Hadoop 管理者ガイド

UNIX、Windows および Linux

リリース 10.4

VERITAS™

NetBackup™ for Hadoop 管理者ガイド

最終更新日: 2024-05-14

法的通知と登録商標

Veritas、Veritas ロゴ、Veritas Alta、NetBackup は、Veritas Technologies LLC または関連会社の米国およびその他の国における商標または登録商標です。その他の会社名、製品名は各社の登録商標または商標です。

この製品には、Veritas 社がサードパーティへの帰属を示す必要があるサードパーティ製ソフトウェア(「サードパーティ製プログラム」)が含まれる場合があります。サードパーティプログラムの一部は、オープンソースまたはフリーソフトウェアライセンスで提供されます。本ソフトウェアに含まれる本使用許諾契約は、オープンソースまたはフリーソフトウェアライセンスでお客様が有する権利または義務を変更しないものとします。このVeritas製品に付属するサードパーティの法的通知文書は次の場所で入手できます。

https://www.veritas.com/about/legal/license-agreements

本書に記載されている製品は、その使用、コピー、頒布、逆コンパイルおよびリバースエンジニアリングを制限するライセンスに基づいて頒布されます。Veritas Technologies LLC からの書面による許可なく本書を複製することはできません。

本書は、現状のままで提供されるものであり、その商品性、特定目的への適合性、または不侵害の暗黙的な保証を含む、明示的あるいは暗黙的な条件、表明、および保証はすべて免責されるものとします。ただし、これらの免責が法的に無効であるとされる場合を除きます。Veritas Technologies LLC およびその関連会社は、本書の提供、パフォーマンスまたは使用に関連する付随的または間接的損害に対して、一切責任を負わないものとします。本書に記載の情報は、予告なく変更される場合があります。

ライセンスソフトウェアおよび文書は、FAR 12.212 に定義される商用コンピュータソフトウェアと見なされ、Veritasがオンプレミスまたはホスト型サービスとして提供するかを問わず、必要に応じて FAR 52.227-19「商用コンピュータソフトウェア - 制限される権利 (Commercial Computer Software - Restricted Rights)」、DFARS 227.7202「商用コンピュータソフトウェアおよび商用コンピュータソフトウェア文書 (Commercial Computer Software and Commercial Computer Software Documentation)」、およびそれらの後継の規制に定める制限される権利の対象となります。米国政府によるライセンス対象ソフトウェアおよび資料の使用、修正、複製のリリース、実演、表示または開示は、本使用許諾契約の条項に従ってのみ行われるものとします。

Veritas Technologies LLC
2625 Augustine Drive
Santa Clara, CA 95054

http://www.veritas.com

テクニカルサポート

テクニカルサポートはグローバルにサポートセンターを管理しています。すべてのサポートサービスは、サポート契約と現在のエンタープライズテクニカルサポートポリシーに応じて提供されます。サ

ポート内容およびテクニカルサポートの利用方法に関する情報については、次の Web サイトにアクセスしてください。

https://www.veritas.com/support

次の URL で Veritas Account の情報を管理できます。

https://my.veritas.com

現在のサポート契約についてご不明な点がある場合は、次に示すお住まいの地域のサポート契約管理チームに電子メールでお問い合わせください。

世界共通 (日本を除く)	CustomerCare@veritas.com
日本	CustomerCare_Japan@veritas.com

マニュアル

マニュアルの最新バージョンがあることを確認してください。各マニュアルには、2 ページ目に最終更新日が記載されています。最新のマニュアルは、Veritasの Web サイトで入手できます。

https://sort.veritas.com/documents

マニュアルに対するご意見

お客様のご意見は弊社の財産です。改善点のご指摘やマニュアルの誤謬脱漏などの報告をお願いします。その際には、マニュアルのタイトル、バージョン、章タイトル、セクションタイトルも合わせてご報告ください。ご意見は次のアドレスに送信してください。

NB.docs@veritas.com

次の Veritas コミュニティサイトでマニュアルの情報を参照したり、質問したりすることもできます。

http://www.veritas.com/community/

Veritas Services and Operations Readiness Tools (SORT)

Veritas SORT (Service and Operations Readiness Tools) は、特定の時間がかかる管理タスクを自動化および簡素化するための情報とツールを提供する Web サイトです。製品によって異なりますが、SORT はインストールとアップグレードの準備、データセンターにおけるリスクの識別、および運用効率の向上を支援します。SORT がお客様の製品に提供できるサービスとツールについては、次のデータシートを参照してください。

https://sort.veritas.com/data/support/SORT_Data_Sheet.pdf

目次

1

概要

この章では以下の項目について説明しています。

- NetBackup を使用した NetBackup for Hadoop データの保護
- NetBackup for Hadoop データのバックアップ
- NetBackup for Hadoop データのリストア
- NetBackup for NetBackup for Hadoop の用語
- 制限事項

NetBackup を使用した NetBackup for Hadoop データの保護

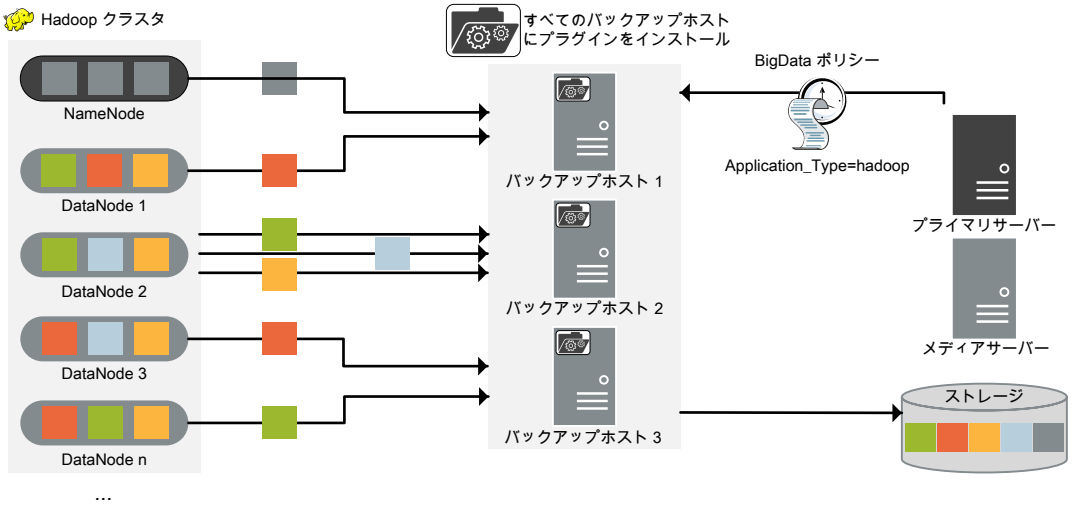
NetBackup の並列ストリームフレームワーク (PSF) を使用した場合、NetBackup を使用して NetBackup for Hadoop データを保護できるようになりました。

次の図は、NetBackup によって NetBackup for Hadoop データを保護する方法の概要を示しています。

また、Hadoop の関連する用語も確認します。

p.11 の「NetBackup for NetBackup for Hadoop の用語」を参照してください。

図 1-1　　アーキテクチャの概要



図では次の内容を説明しています。

- データは並列ストリームでバックアップされ、バックアップ時に DataNodes はデータブロックを同時に複数のバックアップホストに対してストリームします。ジョブの処理速度が、複数のバックアップホストと並列ストリームによって向上します。
- NetBackup for Hadoop クラスタと NetBackup 間の通信は、NetBackup for Hadoop の NetBackup プラグインを使用して有効になります。
  プラグインは NetBackup のインストール時にインストールされます。
- NetBackup 通信の場合、BigData ポリシーを構成し、関連するバックアップホストを追加する必要があります。
- NetBackup のメディアサーバー、クライアント、またはプライマリサーバーをバックアップホストとして構成できます。また、DataNodes の数によっては、バックアップホストを追加または削除できます。バックアップホストをさらに追加することで使用環境の規模を簡単に拡大できます。
- NetBackup 並列ストリームフレームワークにより、エージェントレスのバックアップが可能で、バックアップとリストア操作はバックアップホストで実行します。クラスタノードには、エージェントの占有域がありません。また、NetBackup は NetBackup for Hadoop クラスタのアップグレードやメンテナンスの影響を受けません。

詳細情報:

- p.9 の「NetBackup for Hadoop データのバックアップ」を参照してください。
- p.10 の「NetBackup for Hadoop データのリストア」を参照してください。

- p.13 の「制限事項」を参照してください。
- NetBackup 並列ストリームフレームワーク (PSF) については、『NetBackup 管理者ガイド Vol. 1』を参照してください。

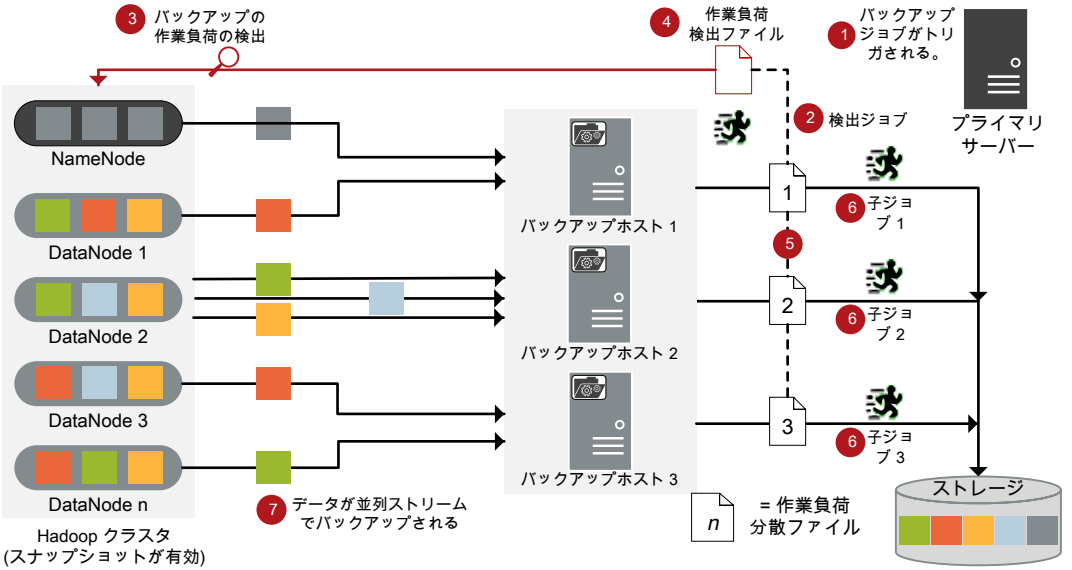
NetBackup for Hadoop データのバックアップ

NetBackup for Hadoop データは並列ストリームでバックアップされ、バックアップ時に NetBackup for Hadoop DataNodes はデータブロックを同時に複数のバックアップホストに対してストリームします。

メモ: NetBackup for Hadoop バックアップ対象で指定されたすべてのディレクトリは、バックアップ前にスナップショット対応に設定する必要があります。

次の図は、バックアップフローの概要を示しています。

図 1-2　　バックアップフロー



次の図に示されているようになります。

1. スケジュールされたバックアップジョブはプライマリサーバーからトリガされます。
2. NetBackup for Hadoop データのバックアップジョブは複合ジョブです。バックアップジョブがトリガされると、最初に検出ジョブが実行されます。

3. 検出中に、最初のバックアップホストは NameNode と接続し、検出を実行して、バックアップする必要があるデータの詳細を取得します。
4. 作業負荷検出ファイルは、バックアップホストに作成されます。作業負荷検出ファイルには、さまざまな DataNodes からバックアップする必要があるデータの詳細が含まれています。
5. バックアップホストは作業負荷検出ファイルを使用し、作業負荷が複数のバックアップホスト間でどのように分散されるかを決定します。作業負荷分散ファイルは、バックアップホストごとに作成されます。
6. バックアップホストごとに個別の子ジョブが実行されます。作業負荷分散ファイルで指定されたデータがバックアップされます。
7. データブロックは、異なる DataNodes から複数のバックアップホストに同時にストリームします。

すべての子ジョブが完了するまで、複合バックアップジョブは完了しません。子ジョブが完了すると、NetBackup は NameNode からすべてのスナップショットをクリーンアップします。クリーンアップ活動が完了した後にのみ、複合バックアップジョブは完了します。

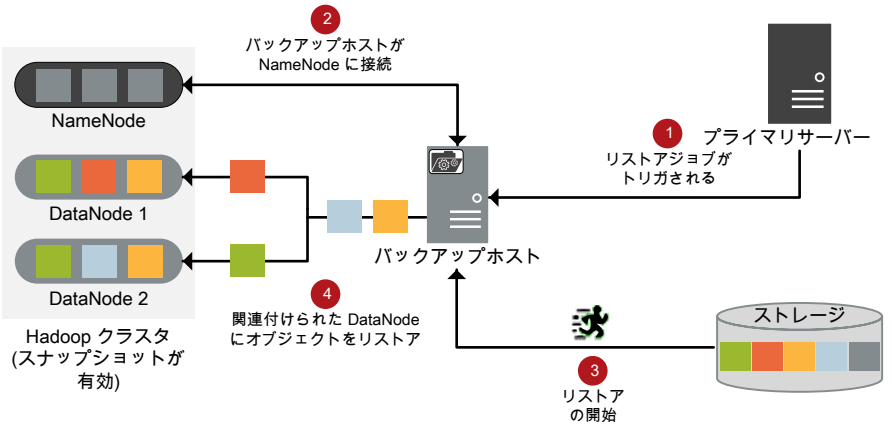
p.43 の「NetBackup for Hadoop クラスタのバックアップについて」を参照してください。

NetBackup for Hadoop データのリストア

リストアに使用されるのは、1 つのバックアップホストのみです。

次の図は、リストアフローの概要を示しています。

図 1-3　リストアフロー



図では次の内容を説明しています。

1. プライマリサーバーからのリストアジョブがトリガされます。
2. バックアップホストが NameNode と接続します。バックアップホストは宛先クライアントでもあります。
3. ストレージメディアからの実際のデータリストアが開始されます。
4. データブロックは DataNodes にリストアされます。

p.45 の「NetBackup for Hadoop クラスタのリストアについて」を参照してください。

NetBackup for NetBackup for Hadoop の用語

次の表では、NetBackup for Hadoop クラスタの保護に NetBackup を使用するときに使われる用語を定義しています。

表 1-1 NetBackup の用語

用語	定義
複合ジョブ	NetBackup for Hadoop データのバックアップジョブは複合ジョブです。 ■ バックアップジョブは、バックアップするデータの情報を取得するための検出ジョブを実行します。 ■ 子ジョブは、実際のデータ転送を実行する各バックアップホストに対して作成されます。 ■ バックアップが完了すると、ジョブは NameNode 上のスナップショットをクリーンアップし、その後ジョブ自体に完了したというマークが付けられます。
検出ジョブ	バックアップジョブを実行すると、最初に検出ジョブが作成されます。検出ジョブは NameNode と通信し、バックアップする必要があるブロックの情報と、関連する DataNodes の情報を収集します。検出の最後に、ジョブは作業負荷検出ファイルにデータを入力します。ファイルはその後 NetBackup によってバックアップホスト間で作業負荷を分散させるために使用されます。
子ジョブ	バックアップの場合、ストレージメディアにデータを転送するバックアップホストごとに個別の子ジョブが作成されます。子ジョブは、複数の DataNodes からデータブロックを転送できます。
作業負荷検出ファイル	検出時のバックアップホストが NameNode と通信するときに、作業負荷検出ファイルが作成されます。このファイルには、バックアップするデータブロックと、関連付けられている DataNodes についての情報が含まれています。
作業負荷分散ファイル	検出が完了すると、NetBackup はバックアップホストごとに作業負荷配布ファイルを作成します。これらのファイルには、それぞれのバックアップホストで転送されるデータの情報が含まれます。

用語	定義
並列ストリーム	NetBackup 並列ストリームフレームワークにより、複数の DataNodes からのデータブロックを、複数のバックアップホストを同時に使用してバックアップできます。
バックアップホスト	バックアップホストは、プロキシクライアントとして機能します。すべてのバックアップとリストア操作は、バックアップホストで実行されます。 メディアサーバー、クライアント、またはプライマリサーバーを、バックアップホストとして構成できます。 バックアップホストは、リストア中に宛先クライアントとしても使用されます。
BigData ポリシー	BigData ポリシーは以下を実行するために導入されました。 ■ アプリケーションの種類を指定します。 ■ 分散マルチノード環境のバックアップを可能にします。 ■ バックアップホストを関連付けます。 ■ 作業負荷分散を実行します。
アプリケーションサーバー	Namenode は、NetBackup ではアプリケーションサーバーと呼ばれます。
プライマリ NameNode	高可用性シナリオでは、1 つの NameNode を BigData ポリシーと tpconfig コマンドで指定する必要があります。この NameNode はプライマリ NameNode と呼ばれます。
フェールオーバー NameNode	高可用性シナリオでは、hadoop.conf ファイル内で更新されるプライマリ NameNode 以外の NameNode は、フェールオーバー NameNode と呼ばれます。

表 1-2 NetBackup for Hadoop の用語

用語	定義
NameNode	NameNode は、リストア中にソースクライアントとしても使用されます。
DataNode	DataNode は、NetBackup for Hadoop で実際のデータを格納する役割を果たします。

用語	定義
スナップショット対応ディレクトリ (スナップショット可能)	スナップショットは、ディレクトリがスナップショット対応になれば、どのディレクトリでも実行できます。 ■ 各スナップショット対応ディレクトリは、65,536 の同時スナップショットに対応できます。スナップショット対応ディレクトリの数に制限はありません。 ■ 管理者は、どのディレクトリでもスナップショット対応に設定できます。 ■ スナップショット対応ディレクトリは、その中にスナップショットがあると、それらすべてのスナップショットが削除されるまでは削除したり名前を変更したりできません。 ■ 親または子のいずれかがスナップショット対応ディレクトリである場合、ディレクトリはスナップショット対応にできません。

制限事項

NetBackup for Hadoop プラグインを配備する前に、次の制限事項を確認します。

- RHEL および SUSE プラットフォームのみが、バックアップホストのサポート対象です。Hadoop クラスタでサポートされるプラットフォームについては、『NetBackup Database and Application Agent Compatibility List』を参照してください。
- 委任トークン認証方法は、NetBackup for Hadoop クラスタではサポートされていません。
- Hadoop プラグインはバックアップ中にはオブジェクトの拡張属性 (xattrs) またはアクセス制御リスト (ACL) をキャプチャしないため、それらはリストアされたファイルまたはフォルダに対しては設定されません。
- 高可用性 NetBackup for Hadoop クラスタでは、バックアップまたはリストア操作中にフェールオーバーが発生すると、ジョブは失敗します。
- バックアップ操作の検出ジョブが進行中のときにバックアップジョブを手動でキャンセルしても、スナップショットエントリは Hadoop Web グラフィカルユーザーインターフェース (GUI) から削除されません。
- HTTPS ベースの Hadoop クラスタのバックアップ中に CRL の期限が切れた場合、バックアップは部分的に実行されます。
- 複数の CRL ベースの Hadoop クラスタがある場合は、クラスタごとに異なるバックアップホストを追加していることを確認します。
- `bp.conf` で `NB_FIPS_MODE` が有効になっている場合、Kerberos 認証では、バックアップおよびリストア操作はサポートされません。

メモ: Kerberos 認証でバックアップを実行するには、NB_FIPS_MODE=0 を指定するか、無効にして、新しいバックアップホストを配備します。

2

NetBackup 用 NetBackup for Hadoop プラグインの前提条件およびベストプラクティス

この章では以下の項目について説明しています。

Hadoop プラグインの配備について

Hadoop プラグインは NetBackup と共にインストールされます。配備を完了するには次のトピックを確認してください。

表 2-1 Hadoop プラグインの配備

タスク	参照先
前提条件と要件	p.16 の「NetBackup for Hadoop プラグインの前提条件」を参照してください。
Hadoop クラスタの準備	p.16 の「NetBackup for Hadoop クラスタの準備」を参照してください。

タスク	参照先
ベストプラクティス	p.17 の「NetBackup for Hadoop プラグインを配備するためのベストプラクティス」を参照してください。
構成	p.20 の「NetBackup for NetBackup for Hadoop の構成について」を参照してください。

NetBackup for Hadoop プラグインの前提条件

NetBackup for Hadoop プラグインを使用する前に、次の前提条件が満たされていることを確認します。

- p.16 の「オペレーティングシステムとプラットフォームの互換性」を参照してください。
- p.16 の「NetBackup の NetBackup for Hadoop プラグインのライセンス」を参照してください。

オペレーティングシステムとプラットフォームの互換性

このリリースでは、RHEL および SUSE プラットフォームが NetBackup for Hadoop クラスタと NetBackup バックアップホストのサポート対象です。

詳しくは、NetBackup プライマリ互換性リストを参照してください。

NetBackup の NetBackup for Hadoop プラグインのライセンス

NetBackup 用 Hadoop プラグインを使用するバックアップおよびリストア操作では、アプリケーションとデータベースパックライセンスが必要です。

ライセンスを追加する方法に関する詳細情報を参照できます。

『NetBackup 管理者ガイド Vol. 1』を参照してください。

NetBackup for Hadoop クラスタの準備

NetBackup の NetBackup for Hadoop クラスタを準備するには、次のタスクを実行します。

- NetBackup for Hadoop ディレクトリがスナップショット対応であることを確認します。ディレクトリをスナップショット可能にするには、NameNodes で次のコマンドを実行します。

  ```
  hdfs dfsadmin -allowSnapshot directory_name
  ```

メモ: 親または子のいずれかがスナップショット対応ディレクトリである場合、ディレクトリはスナップショット対応にできません。

詳しくは、NetBackup for Hadoop のマニュアルを参照してください。

- バックアップホストが NetBackup for Hadoop クラスタと通信できるように、ファイアウォールの設定を更新 (適切なポートが Hadoop クレデンシャルとともに追加されていることを確認) します。
- すべての NameNodes と DataNodes のエントリを、すべてのバックアップホスト上の `/etc/hosts` ファイルに追加します。ホスト名は FQDN 形式で追加する必要があります。
  または
  `/etc/resolv.conf` ファイルに適切な DNS エントリを追加します。
- NetBackup for Hadoop クラスタで `webhdfs` サービスが有効になっていることを確認します。

NetBackup for Hadoop プラグインを配備するためのベストプラクティス

NetBackup for Hadoop プラグインを配備して、NetBackup for NetBackup for Hadoop を構成するには、次のようにします。

- バックアップホスト、メディアサーバー、プライマリサーバーのホスト名に、一貫性がある表記規則を使用します。たとえば、hadoop.veritas.com というホスト名 (FQDN 形式) を使用している場合は、他のすべての場所で同じ形式を使用します。
- すべての NameNodes と DataNodes のエントリを、すべてのバックアップホスト上の `/etc/hosts` ファイルに追加します。ホスト名は FQDN 形式で追加する必要があります。
  または
  `/etc/resolv.conf` ファイルに適切な DNS エントリを追加します。
- NameNode と DataNodes を必ず FQDN 形式で指定します。
- バックアップホストから (FQDN を使用して) すべてのノードに ping を実行します。
- NameNode のホスト名とポートは、NetBackup for Hadoop クラスタの `core-site.xml` 内の http アドレスパラメータで指定したものと同じでなければなりません。
- 複合リストアジョブの親ジョブをキャンセルしても、子リストアジョブはキャンセルされません。子リストアジョブは手動でキャンセルする必要があります。
- SSL (HTTPS) が有効になっている Hadoop クラスタに対して次のことを確認します。

- Hadoop クラスタのすべてのノードからの公開鍵を含むバックアップホスト上に有効な証明書が存在する。
- CRL を使用する Hadoop クラスタの場合は、CRL が有効で期限が切れていないことを確認する。

- メディアサーバーに十分な空きポートがあることを確認します。
- HDFS (Hadoop 分散ファイルシステム) に特殊文字 % または ^ を含むファイルまたはディレクトリ名を作成しないでください。

3

NetBackup for Hadoop の構成

この章では以下の項目について説明しています。

NetBackup for NetBackup for Hadoop の構成について

表 3-1 NetBackup for NetBackup for Hadoop の構成

タスク	参照先
バックアップホストの追加	p.20 の「バックアップホストの管理」を参照してください。 NetBackup クライアントをバックアップホストとして使用する場合、プライマリサーバーの許可リストに NetBackup クライアントを含める必要があります。 p.22 の「NetBackup プライマリサーバーの許可リストに NetBackup クライアントを含める」を参照してください。
NetBackup での NetBackup for Hadoop クレデンシャルの追加	p.24 の「NetBackup での NetBackup for Hadoop クレデンシャルの追加」を参照してください。
NetBackup for Hadoop 構成ファイルを使用した NetBackup for Hadoop プラグインの構成	p.25 の「NetBackup for Hadoop 構成ファイルを使用した NetBackup for Hadoop プラグインの構成」を参照してください。 p.26 の「高可用性 NetBackup for Hadoop クラスタ用の NetBackup の構成」を参照してください。 p.30 の「バックアップホストのスレッド数の設定」を参照してください。 p.31 の「バックアップホストの分散アルゴリズムとゴールデン比率の構成」を参照してください。 p.31 の「バックアップホストのストリーム数の構成」を参照してください。
Kerberos を使用する NetBackup for Hadoop クラスタ用のバックアップホストの構成	p.39 の「Kerberos を使用する NetBackup for Hadoop クラスタの設定」を参照してください。
NetBackup for Hadoop プラグイン用の NetBackup ポリシーの構成	p.40 の「Hadoop クラスタ用の BigData ポリシーの作成」を参照してください。

バックアップホストの管理

バックアップホストは、Hadoop クラスタのすべてのバックアップとリストア操作をホストするプロキシクライアントとして機能します。NetBackup の Hadoop プラグインの場合、バック

アップホストはすべてのバックアップとリストア操作を実行します。別途エージェントを Hadoop クラスタにインストールする必要はありません。

バックアップホストは、Linux コンピュータである必要があります。NetBackup 10.4 のリリースでは、バックアップホストとして RHEL および SUSE プラットフォームのみをサポートします。

バックアップホストとして、NetBackup クライアント、メディアサーバー、またはプライマリサーバーを使用できます。NetBackup ではバックアップホストとしてメディアサーバーを設定することをお勧めします。

バックアップホストを追加する前に、次の点を考慮します。

- バックアップ操作用に、1 つ以上のバックアップホストを追加できます。
- リストア操作用に、バックアップホストを 1 つだけ追加できます。
- プライマリサーバー、メディアサーバー、またはクライアントが、バックアップホストの役割を実行できます。
- NetBackup の Hadoop プラグインは、すべてのバックアップホストにインストールされます。

バックアップホストの追加

バックアップホストを追加するには

1. NetBackup Web UI を開きます。
2. BigData ポリシーを作成します。

   p.40 の「Hadoop クラスタ用の BigData ポリシーの作成」を参照してください。

3 [バックアップ対象 (Backup selections)]タブで[追加 (Add)]をクリックし、次の形式でバックアップホストを追加します。

Backup_Host=IP_address or hostname

また、次のコマンドを使用して、バックアップホストを追加することもできます。

Windows の場合:

```
<install_path>¥NetBackup¥bin¥admincmd¥bpplinclude PolicyName -add
Backup_Host=IP_address or hostname
```

UNIX の場合:

```
/usr/openv/var/global/bin/admincmd/bpplinclude PolicyName -add
Backup_Host=IP_address or hostname
```

4 ベストプラクティスとして、すべてのバックアップホスト上の `/etc/hosts` ファイルにすべての NameNodes と DataNodes のエントリを追加します。FQDN 形式でホスト名を追加する必要があります。

または

`/etc/resolv.conf` ファイルに適切な DNS エントリを追加します。

バックアップホストの削除

バックアップホストを削除するには

1 [バックアップ対象]タブで、削除するバックアップホストを選択します。

2 選択したバックアップホストを右クリックし、[削除]をクリックします。

また、次のコマンドを使用して、バックアップホストを削除することもできます。

Windows の場合:

```
Install_Path¥NetBackup¥bin¥admincmd¥bpplinclude PolicyName -delete
Backup_Host=IP_address or hostname
```

UNIX の場合:

```
/usr/openv/var/global/bin/admincmd/bpplinclude PolicyName -delete
'Backup_Host=IP_address or hostname'
```

NetBackup プライマリサーバーの許可リストに NetBackup クライアントを含める

バックアップホストとして NetBackup クライアントを使用するには、許可リストに含める必要があります。NetBackup プライマリサーバーで許可リストへの追加手順を実行します。

許可リストへの追加は、ソフトウェアまたはアプリケーションが安全な実行を承認されていないかぎり、それらを実行しないようにシステムを制限するセキュリティ手法です。

NetBackup プライマリサーバーの許可リストに NetBackup クライアントを追加するには

- NetBackup プライマリサーバーで次のコマンドを実行します。
  - UNIX の場合
    コマンドへのディレクトリパスは
    `/usr/openv/var/global/bin/admincmd/bpsetconfig` です。

```
bpsetconfig -h primaryserver
bpsetconfig> APP_PROXY_SERVER = clientname.domain.org
bpsetconfig>
UNIX systems: <ctl-D>
```

  - Windows の場合
    コマンドへのディレクトリパスは
    `<Install_Path>¥NetBackup¥bin¥admincmd¥bpsetconfig` です。

```
bpsetconfig -h primaryserver
bpsetconfig> APP_PROXY_SERVER = clientname1.domain.org
bpsetconfig> APP_PROXY_SERVER = clientname2.domain.org
bpsetconfig>
Windows systems: <ctl-Z>
```

このコマンドは *APP_PROXY_SERVER = clientname* エントリをバックアップ構成 (bp.conf) ファイルに設定します。

APP_PROXY_SERVER = clientname について詳しくは、『NetBackup 管理者ガイド Vol. 1』の「NetBackup クライアントの構成オプション」のセクションを参照してください。

Veritas NetBackup のドキュメント

バックアップホストとしての NetBackup アプライアンスの設定

NetBackup アプライアンスをバックアップホストとして使用する場合、次の記事を確認してください。

- NetBackup アプライアンスを Kerberos 認証を使用する NetBackup for Hadoop のバックアップホストとして使用する
  詳しくは、ベリタスのテクニカルサポートに連絡し、担当者に 100039992 の記事を参照するよう伝えてください。
- 高可用性 NetBackup for Hadoop クラスタによって、NetBackup アプライアンスをバックアップホストとして使用する
  詳しくは、ベリタスのテクニカルサポートに連絡し、担当者に 100039990 の記事を参照するよう伝えてください。

NetBackup での NetBackup for Hadoop クレデンシャルの追加

正常なバックアップとリストア操作のために NetBackup for Hadoop クラスタと NetBackup との間でシームレスな通信を確立するには、NetBackup for Hadoop クレデンシャルを NetBackup プライマリサーバーに追加して更新する必要があります。

tpconfig コマンドを使用して、NetBackup プライマリサーバーに NetBackup for Hadoop クレデンシャルを追加します。

tpconfig コマンドを使用してクレデンシャルの削除と更新を行うパラメータについて詳しくは https://www.veritas.com/content/support/ja_JP/DocumentBrowsing.html?product=NetBackup『NetBackup コマンドリファレンスガイド』を参照してください。

NetBackup for Hadoop クレデンシャルを追加する場合は、次の点を考慮します。

- 高可用性 NetBackup for Hadoop クラスタの場合、プライマリとフェールオーバーの NameNode のユーザーが同じであることを確認します。
- BigData ポリシーを構成するときに使用するアプリケーションサーバーのクレデンシャルを使用します。
- Kerberos を使用する NetBackup for Hadoop クラスタの場合、「kerberos」を application_server_user_id 値として指定します。
- NameNode のホスト名とポートは、NetBackup for Hadoop クラスタの core-site.xml 内の http アドレスパラメータで指定したものと同じでなければなりません。
- パスワードについては、ランダムな値を指定します。たとえば、Hadoop です。

NetBackup で Hadoop クレデンシャルを追加するには

1. 次のディレクトリパスから tpconfig コマンドを実行します。

   UNIX システムでは、/usr/openv/volmgr/bin/

   Windows システムでは、install_path¥Volmgr¥bin¥

2. tpconfig --help コマンドを実行します。Hadoop クレデンシャルを追加、更新、および削除するのに必要なオプションのリストが表示されます。

3 `tpconfig -add -application_server` *`application_server_name`* `-application_server_user_id` *`user_ID`* `-application_type` *`application_type`* `-requiredport` *`IP_port_number`* `[-password` *`password`* `[-key` *`encryption_key`*`]]` コマンドを各パラメータに適切な値を入力して実行し、Hadoop クレデンシャルを追加します。

たとえば、*application_server_name* が *hadoop1* である Hadoop サーバーのクレデンシャルを追加する場合、適切な *<user_ID>* と *<password>* の詳細を使用して次のコマンドを実行します。

```
tpconfig -add -application_server hadoop1 -application_type hadoop
-application_server_user_id Hadoop -requiredport 50070 -password
Hadoop
```

ここで、`-application_type` パラメータに指定された値 `hadoop` は Hadoop に対応します。

4 `tpconfig -dappservers` コマンドを実行し、NetBackup プライマリサーバーに追加された Hadoop クレデンシャルがあることを確認します。

NetBackup for Hadoop 構成ファイルを使用した NetBackup for Hadoop プラグインの構成

バックアップホストは、NetBackup for Hadoop プラグインの設定を保存するために `hadoop.conf` ファイルを使用します。各バックアップホストに対して個別のファイルを作成し、`/usr/openv/var/global/` にコピーする必要があります。`hadoop.conf`ファイルは、JSON 形式で手動で作成する必要があります。デフォルトでは、インストーラはこのファイルを使用できません。

メモ: どのパラメータにも空白値は指定できません。指定するとバックアップジョブは失敗します。

バックアップとリストア操作を正常に実行するために必要なすべてのパラメータを設定していることを確認します。

このリリースでは、次のプラグインを設定できます。

- p.26 の「高可用性 NetBackup for Hadoop クラスタ用の NetBackup の構成」を参照してください。
- p.29 の「NetBackup for Hadoop クラスタのカスタムポートの設定」を参照してください。
- p.30 の「バックアップホストのスレッド数の設定」を参照してください。

- p.32 の「NetBackup および Hadoop クラスタ間での SSL 対応 (HTTPS) 通信の設定」を参照してください。

hadoop.conf ファイルの例を次に示します。

メモ: HA 以外の環境では、フェールオーバーパラメータは必要ありません。

```
{
 "application_servers":
 {
  "hostname_of_the_primary_namenode":
  {
   "failover_namenodes":
   [
    {
     "hostname":"hostname_of_failover_namenode",
     "port":port_of_the_failover_namenode
    }
   ],
   "port":port_of_the_primary_namenode
   "distro_algo": distribution_algorithm,
   "num_streams": number_of_streams
   }
  }
 },
  "number_of_threads":number_of_threads
}
```

高可用性 NetBackup for Hadoop クラスタ用の NetBackup の構成

NetBackup for NetBackup for Hadoop クラスタの構成時に高可用性 NetBackup for Hadoop クラスタを保護するには、次のようにします。

- BigData ポリシーでクライアントとしていずれかの NameNodes (プライマリ) を指定します。
- tpconfig コマンドを実行するときに、アプリケーションサーバーと同じ NameNode (プライマリとフェールオーバー) を指定します。
- hadoop.conf ファイルを作成して、NameNode (プライマリとフェールオーバー) の詳細で更新し、すべてのバックアップホストに複製します。hadoop.conf ファイルは JSON 形式です。

- NameNode のホスト名とポートは、NetBackup for Hadoop クラスタの `core-site.xml` 内の http アドレスパラメータで指定したものと同じでなければなりません。
- プライマリとフェールオーバーの NameNode のユーザー名は同じでなければなりません。
- どのパラメータにも空白値は指定できません。指定するとバックアップジョブは失敗します。

高可用性 NetBackup for Hadoop クラスタの hadoop.conf ファイルを更新するには

1 次のパラメータで hadoop.conf ファイルを更新します。

```
{
 "application_servers":
 {
  "hostname_of_primary_namenode1":
  {
   "failover_namenodes":
   [
    {
     "hostname": "hostname_of_failover_namenode1",
     "port": port_of_failover_namenode1
    }
   ],
   "port":port_of_primary_namenode1
  }
 }
}
```

2 複数の NetBackup for Hadoop クラスタがある場合、同じ hadoop.conf ファイルを使用して詳細を更新します。次に例を示します。

```
{
 "application_servers":
 {
  "hostname_of_primary_namenode1":
  {
   "failover_namenodes":
   [
    {
     "hostname": "hostname_of_failover_namenode1",
     "port": port_of_failover_namenode1
    }
   ],
   "port"::port_of_primary_namenode1
  },
  "hostname_of_primary_namenode2":
  {
   "failover_namenodes":
   [
    {
     "hostname": "hostname_of_failover_namenode2",
     "port": port_of_failover_namenode2
    }
   ],
   "port":port_of_primary_namenode2
  }
 }
}
```

3 このファイルをすべてのバックアップホストの次の場所に複製します。

```
/usr/openv/var/global/
```

NetBackup for Hadoop クラスタのカスタムポートの設定

NetBackup for Hadoop 設定ファイルを使用すると、カスタムポートを設定できます。デフォルトでは、NetBackup は 50070 番ポートを使用します。

NetBackup for Hadoop クラスタのカスタムポートを設定するには

1 次のパラメータで hadoop.conf ファイルを更新します。

```
{
  "application_servers": {
   "hostname_of_namenode1":{

     "port":port_of_namenode1
  }
}
```

2 このファイルをすべてのバックアップホストの次の場所に複製します。

```
/usr/openv/var/global/
```

バックアップホストのスレッド数の設定

バックアップのパフォーマンスを向上させるため、各バックアップホストが許容するスレッド数 (ストリーム) を設定できます。さらにバックアップホストを追加するか、バックアップホストあたりのスレッド数を増やすことで、バックアップのパフォーマンスを改善できます。

スレッド数を決定する場合、次の内容を考慮してください。

- デフォルトの値は 4 です。
- 各バックアップホストに対して最小で 1 個、最大で 32 個のスレッドを設定できます。
- 各バックアップホストに異なるスレッド数を設定できます。
- スレッド数を設定するときは、利用可能なコア数と使用するコア数を考慮してください。ベストプラクティスとして、1 コアに 1 スレッド設定することをお勧めします。たとえば、8 つのコアを利用可能で、4 つのコアを使用する場合、4 個のスレッドを設定します。

/usr/openv/var/global/スレッド数の設定のために hadoop.conf ファイルを更新するには

1 次のパラメータで hadoop.conf ファイルを更新します。

```
{
 "number_of_threads": number_of_threads
}
```

2 このファイルをバックアップホストの次の場所にコピーします。

```
/usr/openv/var/global/
```

バックアップホストのストリーム数の構成

バックアップのパフォーマンスを向上させるため、各バックアップホストが許容するストリーム数を構成できます。さらにバックアップホストを追加するか、バックアップホストあたりのストリーム数を増やすことで、バックアップのパフォーマンスを改善できます。

ストリーム数を決定する場合、次の内容を考慮してください。

- デフォルト値は 1 です。
- 並列ストリーム数はチューニングパラメータに基づいています。

ストリーム数の構成のために hadoop.conf ファイルを更新するには

1. 次のパラメータで hadoop.conf ファイルを更新します。

```
{
 "num_of_streams": number_of_streams
}
```

2. このファイルをバックアップホストの次の場所にコピーします。

```
/usr/openv/var/global/
```

メモ: ストリーム数を増やした場合は、クライアントあたりのジョブの最大数を更新し、複数のスレッドの stu 設定とクライアントタイムアウトを更新して、エラーが突然発生しないようにします。

バックアップホストの分散アルゴリズムとゴールデン比率の構成

バックアップパフォーマンスを向上させるために、チューニングパラメータに基づいて分散アルゴリズムとゴールデン比率を構成できます。配布アルゴリズムとゴールデン比率の組み合わせによって、これらのアルゴリズムのパフォーマンスの微調整が可能になり、バックアップパフォーマンスを向上させることができます。

配布アルゴリズムとゴールデン比率を決定するには、次の点を考慮してください。

- データセットに**少数の大きいサイズのファイル**がある場合: 配布アルゴリズム 1 を使用します。ゴールデン比率の変更は反映されません。
- データセットに**多数の小さいサイズのファイル**がある場合: 配布アルゴリズム 2 を使用します。ゴールデン比率の変更は反映されません。
- データセットに、**少数のサイズが非常に大きいファイルと多数のサイズが小さいファイル**がある場合: 配布アルゴリズム 4 または 5 と、配備に適したゴールデン比率を使用します。ゴールデン比率でサポートされる範囲は 1 から 100 です。指定しない場合、デフォルトで 75 と見なされます。

メモ: この値を調整すると、パフォーマンスが大幅に変わる可能性があります。

/usr/openv/var/global/ アルゴリズムとゴールデン比率の構成のために hadoop.conf ファイルを更新するには

1 次のパラメータで hadoop.conf ファイルを更新します。

```
{
"distro_algo": distribution_algorithm and
"golden_ratio":godlen_ratio
}
```

2 このファイルをバックアップホストの次の場所にコピーします。

/usr/openv/var/global/

NetBackup および Hadoop クラスタ間での SSL 対応 (HTTPS) 通信の設定

NetBackup と Hadoop クラスタ間で SSL 対応 (HTTPS) 通信を有効にするには、次の手順を実行します。

- use_ssl パラメータを次の形式で使用して、バックアップホストの /usr/openv/var/global/ ディレクトリにある hadoop.conf ファイルを更新します。

```
{
  "application_servers":
  {
   "hostname_of_namenode1":
    {
       "use_ssl":true
    }
  }
}
```

SSL と HA の構成ファイルの形式:

```
{
   "application_servers":
   {
    "primary.host.com":
    {
      "use_ssl":true,
     "failover_namenodes":
      [
```

```
        {
         "hostname":"secondary.host.com",
         "use_ssl":true,
         "port":11111
        }
       ]
      }
     }
}
```

デフォルトでは、この値は `false` に設定されています。
複数のバックアップホストを使用する場合、`hadoop.conf` ファイルで `use_ssl` パラメータを定義したバックアップホストが通信に使用されます。
すべての Hadoop クラスタについて、`hadoop.conf` ファイルで `use_ssl` パラメータを定義する必要があります。

- `nbsetconfig` コマンドを使用して、アクセスホストで次の NetBackup 構成オプションを構成します。
  構成オプションについて詳しくは、『NetBackup 管理者ガイド』を参照してください。

`ECA_TRUST_STORE_PATH`

信頼できるすべてのルート CA 証明書を含む証明書バンドルファイルのファイルパスを指定します。

この外部 CA のオプションをすでに構成してある場合は、Hadoop の CA 証明書を既存の外部証明書トラストストアに追加します。

このオプションを構成していない場合は、必要な Hadoop サーバーの CA 証明書をすべてトラストストアに追加して、このオプションを設定します。

p.34 の「NetBackup サーバーとクライアントの ECA_TRUST_STORE_PATH」を参照してください。

`ECA_CRL_PATH`

外部 CA の証明書失効リスト (CRL) が保存されているディレクトリのパスを指定します。

この外部 CA のオプションをすでに構成してある場合は、Hadoop サーバーの CRL を CRL キャッシュに追加します。

このオプションを構成していない場合は、必要なすべての CRL を CRL キャッシュに追加してオプションを設定します。

p.35 の「NetBackup サーバーとクライアントの ECA_CRL_PATH」を参照してください。

HADOOP_SECURE_CONNECT_ENABLED	このオプションは、Hadoop の安全な通信に影響します。 hadoop.conf ファイルで use_ssl を true に設定した場合は、この値を YES に設定します。use_ssl を true に設定すると、すべての Hadoop クラスタに単一の値を適用できます。 Hadoop では、デフォルトで安全な通信が有効です。 このオプションを使用すると、セキュリティ証明書検証をスキップできます。 p.37 の「サーバーとクライアントの HADOOP_SECURE_CONNECT_ENABLED」を参照してください。
HADOOP_CRL_CHECK	CRL で Hadoop サーバー証明書の失効状態を検証できます。 use_ssl を true に設定すると、すべての Hadoop クラスタに単一の値を適用できます。 デフォルトでは、このオプションは無効になっています。 p.38 の「NetBackup サーバーとクライアントの HADOOP_CRL_CHECK」を参照してください。

NetBackup サーバーとクライアントの ECA_TRUST_STORE_PATH

ECA_TRUST_STORE_PATH オプションでは、信頼できるすべてのルート CA 証明書を含む証明書バンドルファイルへのファイルパスを指定します。

この証明書ファイルには、PEM 形式の 1 つ以上の証明書が必要です。

Windows 証明書ストアを使用する場合、ECA_TRUST_STORE_PATH オプションを指定しないでください。

トラストストアは次の形式の証明書をサポートします。

- 信頼できるルート認証局の、バンドルされている証明書を持つ PKCS #7 または P7B ファイル。このファイルは、PEM または DER でエンコードされている場合があります。
- 信頼できるルート認証局の PEM エンコードされた証明書が連結されて含まれるファイル。

このオプションは、ファイルベースの証明書で必須です。

Cloudera ディストリビューションのルート CA 証明書は、Cloudera 管理者から取得できます。Hadoop クラスタで手動 TLS 構成または自動 TLS が有効になっている場合があります。いずれの場合も、NetBackup では管理者からのルート CA 証明書が必要になります。

セキュア (SSL) クラスタの場合、Hadoop クラスタのルート CA 証明書を使用してすべてのノードの証明書を検証し、NetBackup でバックアップおよびリストアプロセスを実行できます。このルート CA 証明書は、このようなすべてのノードに対して発行された証明書のバンドルです。

自己署名 CA 環境、サードパーティ CA 環境、ローカル/中間 CA 環境の場合、ECA_TRUST_STORE_PATH でルート CA の証明書を構成する必要があります。たとえば、自動 TLS が有効な Cloudera 環境では、通常、cm-auto-global_cacerts.pem という名前のルート CA ファイルが /var/lib/cloudera-scm-agent/agent-cert のパスに置かれています。詳しくは、Cloudera のマニュアルを参照してください。

表 3-2 ECA_TRUST_STORE_PATH の情報

使用方法	説明
使用する場所	NetBackup サーバーまたはクライアント上。 VMware、Red Hat Virtualization サーバー、Nutanix AHV に対して証明書の検証が必要な場合、NetBackup がホストの通信に使用する認証局 (NetBackup CA または外部 CA) に関係なく、NetBackup プライマリサーバーとそれぞれのアクセスホストでこのオプションを設定する必要があります。
使用方法	オプションを表示、追加、変更するには、nbgetconfig コマンドと nbsetconfig コマンドを使用します。 これらのコマンドについて詳しくは、『NetBackup コマンドリファレンスガイド』を参照してください。 次の形式を使用します。 ECA_TRUST_STORE_PATH = *Path to the external CA certificate* 例: c:¥rootCA.pem Flex Appliance アプリケーションインスタンスでこのオプションを使用する場合、パスは /mnt/nbdata/hostcert/ である必要があります。
同等の UI プロパティ	相当するエントリは存在しません。

NetBackup サーバーとクライアントの ECA_CRL_PATH

ECA_CRL_PATH オプションは、ECA (外部認証局) の CRL (証明書失効リスト) が保存されているディレクトリのパスを指定します。

これらの CRL は、NetBackup CRL キャッシュにコピーされます。CRL キャッシュの CRL で外部証明書の失効状態が検証されます。

CRL キャッシュ内の CRL は、ECA_CRL_PATH に指定された場所にある CRL で、ECA_CRL_PATH_SYNC_HOURS オプションに基づいて定期的に更新されます。

ECA_CRL_CHECK または HADOOP_CRL_CHECK オプションが DISABLE (または 0) に設定されておらず、ECA_CRL_PATH オプションが指定されていない場合、NetBackup は CRL 配布ポイント (CDP) で指定された URL から CRL をダウンロードし、それらを使用してピアホストの証明書の失効状態を検証します。

メモ: 仮想化サーバー証明書の失効状態の検証には、VIRTUALIZATION_CRL_CHECK オプションを使用します。

Hadoop サーバー証明書の失効状態の検証には、HADOOP_CRL_CHECK オプションを使用します。

表 3-3 ECA_CRL_PATH の情報

使用方法	説明
使用する場所	NetBackup サーバーまたはクライアント上。 VMware、Red Hat Virtualization サーバー、Nutanix AHV、または Hadoop に対して証明書の検証が必要な場合、NetBackup がホストの通信に使用する認証局 (NetBackup CA または外部 CA) に関係なく、NetBackup プライマリサーバーとそれぞれのアクセスホストまたはバックアップホストでこのオプションを設定する必要があります。 VMware、Red Hat Virtualization サーバー、または Hadoop に対して証明書の検証が必要な場合、NetBackup がホストの通信に使用する認証局 (NetBackup CA または外部 CA) に関係なく、NetBackup プライマリサーバーとそれぞれのアクセスホストまたはバックアップホストでこのオプションを設定する必要があります。

使用方法	説明
使用方法	オプションを表示、追加、変更するには、nbgetconfig コマンドと nbsetconfig コマンドを使用します。 これらのコマンドについて詳しくは、『NetBackup コマンドリファレンスガイド』を参照してください。 次の形式を使用して、CRL ディレクトリのパスを指定します。 ECA_CRL_PATH = *Path to the CRL directory* 次に例を示します。 ECA_CRL_PATH = /usr/eca/crl/eca_crl_file.crl Flex Appliance アプリケーションインスタンスでこのオプションを使用する場合、パスは /mnt/nbdata/hostcert/crl である必要があります。
同等の UI プロパティ	相当するエントリは存在しません。

サーバーとクライアントの HADOOP_SECURE_CONNECT_ENABLED

HADOOP_SECURE_CONNECT_ENABLED オプションを指定すると、Hadoop サーバー証明書をルートまたは中間の認証局 (CA) 証明書を使用して検証できます。

表 3-4 HADOOP_SECURE_CONNECT_ENABLED の情報

使用方法	説明
使用する場所	すべてのバックアップホスト
使用方法	オプションを表示、追加、変更するには、nbgetconfig コマンドと nbsetconfig コマンドを使用します。 これらのコマンドについて詳しくは、『NetBackup コマンドリファレンスガイド』を参照してください。 デフォルトでは、HADOOP_SECURE_CONNECT_ENABLED は YES に設定されています。 Hadoop の証明書検証を有効にするには、次の形式を使用します。 HADOOP_SECURE_CONNECT_ENABLED = YES
同等の UI プロパティ	相当するエントリは存在しません。

NetBackup サーバーとクライアントの HADOOP_CRL_CHECK

HADOOP_CRL_CHECK オプションを使用すると、Hadoop サーバーの外部証明書の失効の確認レベルを指定できます。確認に基づいて、ホストとの通信時に、証明書失効リスト (CRL) に対して Hadoop サーバー証明書の失効状態が検証されます。

デフォルトでは、HADOOP_CRL_CHECK は無効になっています。証明書失効リスト (CRL) に対して Hadoop サーバー証明書の失効状態を検証する場合は、オプションを別の値に設定します。

ECA_CRL_PATH 構成オプションまたは CRL 配布ポイント (CDP) で指定されているディレクトリの CRL を使用できます。

p.35 の「NetBackup サーバーとクライアントの ECA_CRL_PATH」を参照してください。

表 3-5 HADOOP_CRL_CHECK の情報

使用方法	説明
使用する場所	すべてのバックアップホスト
使用方法	オプションを表示、追加、変更するには、nbgetconfig コマンドと nbsetconfig コマンドを使用します。 これらのコマンドについて詳しくは、『NetBackup コマンドリファレンスガイド』を参照してください。 次の形式を使用します。 `HADOOP_CRL_CHECK = CRL check` 次のいずれかを指定できます。 ■ DISABLE (または 0) - 失効の確認を無効にします。ホストとの通信時に、CRL で証明書の失効状態は検証されません。これはデフォルト値です。 ■ LEAF (または 1) - CRL でリーフ証明書の失効状態が検証されます。 ■ CHAIN (または 2) - CRL で証明書チェーンの証明書すべての失効状態が検証されます。
同等の UI プロパティ	相当するエントリは存在しません。

bp.conf ファイルのパラメータ値の例

SSL 対応 (HTTPS) CRL ベースの Hadoop クラスタの bp.conf ファイルに追加された値の例を次に示します。

```
ECA_TRUST_STORE_PATH=/tmp/cacert.pem
ECA_CRL_PATH=/tmp/backuphostdirectory
```

```
HADOOP_SECURE_CONNECT_ENABLED=YES/NO
HADOOP_CRL_CHECK=DISABLE / LEAF / CHAIN
```

Kerberos を使用する NetBackup for Hadoop クラスタの設定

Kerberos を使用する NetBackup for Hadoop クラスタについては、すべてのバックアップホストで次のタスクを実行します。

- すべてのバックアップホストに Kerberos パッケージが配布されていることを確認します。
  - RHEL の場合は krb5-workstation パッケージ
  - SUSE の場合は krb5-client
- `keytab` ファイルを取得して、バックアップホストの安全な場所にコピーします。
- `keytab` に必要なプリンシパルがあることを確認します。
- 適切な KDC サーバーとレルムの詳細で `krb5.conf` ファイルを手動で更新します。

メモ: `default_ccache_name` パラメータの値が KEYRING:persistent:%{uid} に設定されていないことを確認してください。パラメータをコメントアウトしてデフォルトを使用することもできますし、FILE:/tmp/krb_file_name:%{uid} などのファイル名を指定することもできます。

- NetBackup for Hadoop のクレデンシャルを NetBackup に追加するときに、`application_server_user_id` の値として「kerberos」を指定します。p.24 の「NetBackup での NetBackup for Hadoop クレデンシャルの追加」を参照してください。
- Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップとリストア操作については、NetBackup for Hadoop クラスタを認証するため、NetBackup for Hadoop に有効な Kerberos チケット認可チケット (TGT) が必要となります。p.44 の「Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップおよびリストア操作実行の前提条件」を参照してください。
- Kerberos を使用するには、ユーザーは、HDFS のフルアクセスと所有権を持つスーパーユーザーである必要があります。バックアップホストのユーザーに有効なトークンが必要です。

並列リストアの hadoop.conf の構成

TBD

```
"application_servers": {
"punnbuucsm5b29-v14.vxindia.veritas.com": {
"port": 9000,
"distro_algo": 4,
"num_streams": 2,
"golden_ratio": 80,
"additionalBackupHosts": ["bh1.vxindia.veritas.com",
"bh2.vxindia.veritas.com"]
  }
 },
"number_of_threads": 10
}
------------
```

num_stream: リストアのパフォーマンスを向上させるため、各バックアップホストが許容するストリーム数を構成できます。デフォルト値は 1 です。

additionalBackupHosts: リストアのパフォーマンスを向上させるために、追加のバックアップホストの詳細を構成できます。追加のバックアップホストのホスト名を指定できます。

注意:

- 利用可能な追加のバックアップホストがない場合は、additionalBackupHosts を空にしておく必要があります。
- hadoop.conf の構成は、すべてのバックアップホストで同じである必要があります。
- num_stream の構成は、バックアップ処理とリストア処理で同じである必要があります。
- Hadoop の設定と NetBackup の設定は、同じタイムゾーンにする必要があります。
- ストリーム数を増やした場合は、クライアントあたりのジョブの最大数を調整し、複数スレッドの stu 設定とクライアントタイムアウトを更新して、エラーが突然発生しないようにします。

Hadoop クラスタ用の BigData ポリシーの作成

バックアップポリシーは、NetBackup がクライアントをバックアップするときに従う指示を提供します。NetBackup の Hadoop プラグインのバックアップポリシーを構成するには、[ポリシー形式 (Policy type)]として[BigData]、[Hadoop]形式を使用します。

メモ: NameNode のホスト名とポートは、NetBackup for Hadoop クラスタの `core-site.xml` 内の HTTP アドレスパラメータで指定した値と同じでなければなりません。

Hadoop クラスタ用の BigData ポリシーを作成するには

1. NetBackup Web UI を開きます。
2. 左側で[保護 (Protection)]、[ポリシー (Policies)]の順に選択します。
3. [ポリシー (Policies)]タブで、[追加 (Add)]をクリックします。
4. [属性 (Attributes)]タブで、[ポリシー形式 (Policy type)]に[BigData]を選択します。
5. [スケジュール (Schedules)]タブで[追加 (Add)]をクリックして、新しいスケジュールを作成します。

   BigData ポリシーの完全バックアップ、差分増分バックアップ、または累積増分バックアップのスケジュールを作成できます。スケジュールを設定すると、Hadoop データは、ユーザーがそれ以上介入しなくても、設定されたスケジュールに従って自動的にバックアップされます。
6. [クライアント (Clients)]タブには、`NameNode` の IP アドレスまたはホスト名を入力します。
7. [バックアップ対象 (Backup selections)]タブで、次のようにパラメータとその値を入力します。

   - *Application_Type=hadoop*
     これらのパラメータ値では、大文字と小文字が区別されます。
   - *Backup_Host=IP_address or hostname*
     バックアップホストは、Linux コンピュータである必要があります。バックアップホストには、NetBackup クライアントまたはメディアサーバーを指定できます。
     複数のバックアップホストを指定できます。
   - バックアップを作成するファイルのパスまたはディレクトリです。
     複数のファイルパスを指定できます。

   メモ: BigData ポリシーを Application_Type=hadoop で定義するときにバックアップ対象に対して指定されるディレクトリまたはフォルダには、名前にスペースまたはカンマを含めることはできません。
8. [作成 (Create)]をクリックします。

BigData アプリケーションの場合の NetBackup の使用について詳しくは、Veritas NetBackup のドキュメントのページを参照してください。

NetBackup for Hadoop クラスタのディザスタリカバリ

NetBackup for Hadoop クラスタをディザスタリカバリする場合、次のタスクを実行します。

表 3-6　　ディザスタリカバリの実行

タスク	説明
NetBackup for Hadoop クラスタとノードが起動した後、クラスタで NetBackup による操作の準備をします。	次のタスクを実行します。 ファイアウォールの設定を更新して、バックアップホストが NetBackup for Hadoop クラスタと通信できるようにします。 NetBackup for Hadoop クラスタで webhdfs サービスが有効になっていることを確認します。 p.16 の「NetBackup for Hadoop クラスタの準備」を参照してください。
正常なバックアップとリストア操作のために NetBackup for Hadoop クラスタと NetBackup の間のシームレスな通信を確立するには、NetBackup for Hadoop のクレデンシャルを NetBackup プライマリサーバーに追加して更新する必要があります。	`tpconfig` コマンドを使用して、NetBackup プライマリサーバーに NetBackup for Hadoop クレデンシャルを追加します。 p.24 の「NetBackup での NetBackup for Hadoop クレデンシャルの追加」を参照してください。
バックアップホストは、NetBackup for Hadoop プラグインの設定を保存するために `hadoop.conf` ファイルを使用します。各バックアップホストに個別のファイルを作成して、`/usr/openv/var/global/` にコピーする必要があります。`hadoop.conf` ファイルは JSON 形式で作成する必要があります。	このリリースでは、次のプラグインを設定できます。 ■ p.26 の「高可用性 NetBackup for Hadoop クラスタ用の NetBackup の構成」を参照してください。 ■ p.30 の「バックアップホストのスレッド数の設定」を参照してください。
元の NameNode 名で BigData ポリシーを更新します。	p.40 の「Hadoop クラスタ用の BigData ポリシーの作成」を参照してください。

4

Hadoop のバックアップとリストアの実行

この章では以下の項目について説明しています。

- NetBackup for Hadoop クラスタのバックアップについて
- NetBackup for Hadoop クラスタのリストアについて
- バックアップおよびリストア時のパフォーマンスを向上するためのベストプラクティス

NetBackup for Hadoop クラスタのバックアップについて

NetBackup Web UI を使用してバックアップ操作を管理します。

表 4-1 NetBackup for Hadoop データのバックアップ

タスク	参照先
プロセスの理解	p.9 の「NetBackup for Hadoop データのバックアップ」を参照してください。
(オプション) Kerberos の前提条件をすべて満たす	p.44 の「Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップおよびリストア操作実行の前提条件」を参照してください。
NetBackup for Hadoop クラスタのバックアップ	p.45 の「NetBackup for Hadoop クラスタのバックアップ」を参照してください。

タスク	参照先
ベストプラクティス	p.44 の「NetBackup for Hadoop クラスタのバックアップを作成するためのベストプラクティス」を参照してください。
トラブルシューティングのヒント	検出とクリーンアップの関連ログについては、検出をトリガした最初のバックアップホスト上の次のログファイルを確認します。 `/usr/openv/var/global/logs/nbaapidiscv` データ転送関連ログについては、プライマリサーバー上のログファイルから、対応するバックアップホストを (ホスト名を使用して) 検索します。 p.56 の「NetBackup for Hadoop データのバックアップ問題のトラブルシューティング」を参照してください。

Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップおよびリストア操作実行の前提条件

Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップとリストア操作については、NetBackup for Hadoop クラスタを認証するため、NetBackup for Hadoop に有効な Kerberos チケット認可チケット (TGT) が必要となります。

メモ: バックアップ操作中とリストア操作中は、TGT を有効にする必要があります。このため、適切な形で TGT の有効期間を指定するか、操作中必要なときに更新する必要があります。

次のコマンドを実行して TGT を生成します。

```
kinit -k -t /keytab_file_location/keytab_filename principal_name
```

次に例を示します。

```
kinit -k -t /usr/openv/var/global/nbusers/hdfs_mykeytabfile.keytab
hdfs@MYCOMPANY.COM
```

設定に関連する情報も確認してください。p.39 の「Kerberos を使用する NetBackup for Hadoop クラスタの設定」を参照してください。

NetBackup for Hadoop クラスタのバックアップを作成するためのベストプラクティス

NetBackup for Hadoop クラスタのバックアップを作成する前に、次の点を考慮します。

- NetBackup for Hadoop ファイルシステム全体のバックアップを作成する前に、バックアップ対象として「/」を指定し、「/」でスナップショットが有効になっていることを確認します。

- バックアップジョブを実行する前に、すべてのノードでバックアップホストからホスト名 (FQDN) への正常な ping のレスポンスが返ることを確認します。
- ファイアウォールの設定を更新して、バックアップホストが NetBackup for Hadoop クラスタと通信できるようにします。
- HDFS ノードとバックアップホストのローカル時刻が NTP サーバーと同期していることを確認します。
- SSL (HTTPS) が有効になっている Hadoop クラスタに、有効な証明書があることを確認します。

NetBackup for Hadoop クラスタのバックアップ

バックアップ用のポリシーを作成するか、バックアップを手動で実行できます。

p.40 の「Hadoop クラスタ用の BigData ポリシーの作成」を参照してください。

バックアッププロセスの概要が利用可能です。

p.9 の「NetBackup for Hadoop データのバックアップ」を参照してください。

バックアッププロセスは、次のステージで構成されます。

1. 事前処理: 事前処理のステージでは、BigData ポリシーで構成した最初のバックアップホストが検出をトリガします。このステージでは、バックアップ対象全体のスナップショットが生成されます。スナップショットの詳細は、NameNode Web インターフェースに表示されます。
2. データ転送: データ転送処理中には、バックアップホストごとに 1 つの子ジョブが作成されます。
3. 事後処理: 事後処理の一部として、NetBackup は NameNode 上のスナップショットをクリーンアップします。

NetBackup for Hadoop クラスタのリストアについて

NetBackup Web UI を使用してリストア操作を管理します。

表 4-2 NetBackup for Hadoop データのリストア

作業	参照先
プロセスの理解	p.10 の「NetBackup for Hadoop データのリストア」を参照してください。
Kerberos の前提条件をすべて満たす	p.44 の「Kerberos 認証を使用する NetBackup for Hadoop クラスタのバックアップおよびリストア操作実行の前提条件」を参照してください。

作業	参照先
同じ NameNode または NetBackup for Hadoop クラスタへの NetBackup for Hadoop データのリストア	p.47 の「同じ Hadoop クラスタ上での Hadoop データのリストア」を参照してください。
代替 NameNode または NetBackup for Hadoop クラスタへの NetBackup for Hadoop データのリストア このタスクは bprestore コマンドを使用してのみ実行できます。	p.48 の「代替の Hadoop クラスタ上での Hadoop データのリストア」を参照してください。
ベストプラクティス	p.46 の「Hadoop クラスタをリストアするためのベストプラクティス」を参照してください。
トラブルシューティングのヒント	p.61 の「NetBackup for Hadoop データのリストア問題のトラブルシューティング」を参照してください。

Hadoop クラスタをリストアするためのベストプラクティス

Hadoop クラスタをリストアするときは、次の点を考慮してください。

- リストアジョブを実行する前に、クラスタにリストアジョブを完了するための十分な領域があることを確認します。
- ファイアウォールの設定を更新して、バックアップホストが NetBackup for Hadoop クラスタと通信できるようにします。
- SSL (HTTPS) が有効になっている Hadoop クラスタのすべてのクラスタノードに、有効な証明書があることを確認します。
- バックアップホストに有効な PEM 証明書ファイルがあることを確認します。
- HTTP または HTTPS ベースのクラスタ用に、適切なパラメータが `hadoop.conf` ファイルに追加されていることを確認します。
- バックアップホストに期限が切れていない有効な CRL が含まれていることを確認します。

- アプリケーションレベルまたはファイルシステムレベルの暗号化は Hadoop ではサポートされません。リストアが正しく動作するには、Hadoop のスーパーユーザーである必要があります。

同じ Hadoop クラスタ上での Hadoop データのリストア

同じ Hadoop クラスタ上で Hadoop データをリストアするには、次の点を考慮してください。

- NetBackup Web UI を使用して、Hadoop データのリストア操作を開始します。このインターフェースでは、リストアするオブジェクトが存在する NetBackup サーバー、およびバックアップイメージを表示するクライアントを選択できます。これらの選択に基づいて、バックアップイメージの履歴の表示、個々の項目の選択およびリストアの開始を行うことができます。
- リストアブラウザを使用すると、Hadoop ディレクトリオブジェクトを表示することができます。オブジェクトは階層表示され、リストアに使用するオブジェクトを選択できます。Hadoop クラスタを構成するオブジェクト (Hadoop ディレクトリまたはファイル) は、個々のディレクトリを展開すると表示されます。
- 管理者は、Hadoop ディレクトリおよび個々の項目を参照し、リストアすることができます。ユーザーがリストアできるオブジェクトには、Hadoop ファイルとフォルダが含まれます。

同じ Hadoop クラスタ上での Hadoop データのリストア

このトピックでは、同じ Hadoop クラスタ上の Hadoop データをリストアする方法について説明します。

同じ Hadoop クラスタ上の Hadoop データをリストアするには

1. NetBackup Web UI を開きます。
2. 左側の[リカバリ (Recovery)]を選択します。
3. [標準リカバリ (Regular recovery)]で[リカバリの開始 (Start recovery)]をクリックします。
4. [基本プロパティ (Basic properties)]タブで、以下を入力します。
   - [ポリシー形式 (Policy type)]で、[BigData]、[Hadoop]の順に選択します。
   - リストア操作を実行するソースとして NetBackup for Hadoop アプリケーションサーバーを指定します。
     [ソースクライアント (Source client)]リストから、必要なアプリケーションサーバーを選択します。
   - バックアップホストを宛先クライアントとして指定します。

［宛先クライアント (Destination client)］リストから、必要なバックアップホストを選択します。バックアップホストがノードをバックアップしたメディアサーバーの場合、リストアはより短時間になります。

- ［次へ (Next)］をクリックします。

5 ［リカバリの詳細 (Recovery details)］タブで、次の操作を実行します。

- データセット全体をリストアする適切な日付範囲を選択するか、［バックアップ履歴の使用 (Use backup history)］に移動して、リストアするバックアップイメージを選択します。
- 左側のディレクトリ階層から、リストアするファイルとフォルダを選択します。

メモ: そのディレクトリの下にある後続のすべてのファイルとフォルダが、右側のペインに表示されます。

- ［次へ (Next)］をクリックします。

6 ［リカバリオプション (Recovery Options)］タブで、次の操作を実行します。

- バックアップを実行したのと同じ場所にファイルをリストアする場合は、［元の位置にすべてをリストア (Restore everything to its original location)］を選択します。
- バックアップの場所とは異なる場所にファイルをリストアする場合は、［すべてを異なる位置にリストア］を選択します。
  パスを指定します。
- ファイルとディレクトリを別の場所にリストアするには、［個々のディレクトリやファイルを異なる位置にリストア (Restore individual directories and files to different locations)］を選択します。
  ファイルパスを編集および追加します。
- ［リカバリオプション (Recovery options)］で、適切なオプションを選択します。
- ［次へ (Next)］をクリックします。

7 ［レビュー (Review)］タブで、詳細を確認して［リカバリの開始 (Start Recovery)］をクリックします。

代替の Hadoop クラスタ上での Hadoop データのリストア

NetBackup では、別の NameNode または Hadoop クラスタに Hadoop データをリストアできます。この種類のリストア方法は、リダイレクトリストアと呼ばれます。

メモ: NetBackup プライマリサーバーに代替 NameNode または Hadoop クラスタのクレデンシャルが追加されており、NetBackup プライマリサーバーで許可リストへの追加タスクが完了していることを確認してください。NetBackup で Hadoop クレデンシャルを追加する方法と許可リストに登録する手順について詳しくは、p.24 の「NetBackup での NetBackup for Hadoop クレデンシャルの追加」を参照してください。 p.22 の「NetBackup プライマリサーバーの許可リストに NetBackup クライアントを含める」を参照してください。

Hadoop のリダイレクトリストアを実行するには

1 *rename_file* および *listfile* の値を次のように変更します。

パラメータ	値
rename_file	/<source_folder_path> を /<destination_folder_path> ALT_APPLICATION_SERVER=<alternate name node> に変更します。
listfile	リストアするすべての Hadoop ファイルのリスト

2 手順 1 で説明したパラメータに、変更した値を使用して、NetBackup プライマリサーバーで `bprestore -S primary_server -D backup_host -C client -R rename_file -t 44 -L progress log -f listfile` コマンドを実行します。

手順の詳細:

`-S primary_server`

NetBackup プライマリサーバーの名前を指定します。

`-D backup host`

バックアップホストの名前を指定します。

`-C client`

このオプションでは、ファイルのリストア元のバックアップまたはアーカイブの検索に使用するソースとして NameNode を指定します。この名前は、NetBackup カタログに表示される名前と一致している必要があります。

`-f listfile`

このオプションでは、リストアするファイルのリストを含むファイル (listfile) を指定します。このオプションは、ファイル名オプションの代わりに使用できます。listfile では、各ファイルパスを個別の行に指定する必要があります。

`-L progress_log`

このオプションでは、進捗情報を書き込む許可リストファイルパスの名前を指定します。

`-t 44`

ポリシー形式として BigData を指定します。

`-R rename_file`

このオプションでは、代替パスへのリストアのために名前を変更するファイル名を指定します。

ファイル名の変更を記述するファイルのエントリには、次の形式を使用します。

```
change backup_filepath to restore_filepath
ALT_APPLICATION_SERVER=<Application Server Name>
```

ファイルパスは / (スラッシュ) で始まる必要があります。

メモ: NetBackup インストールパスの一部としてまだ組み込まれていない、<rename_file_path>、<progress_log_path> などのすべてのファイルパスを許可リストに載せたことを確認します。

バックアップおよびリストア時のパフォーマンスを向上するためのベストプラクティス

SSL 環境 (HTTPS) を使用した Hadoop のバックアップとリカバリ中に、スループットが低下したり、CPU 使用率が高くなるなどのパフォーマンスの問題が発生します。この問題は、Hadoop の内部通信が暗号化されていない場合に発生します。Hadoop の内部通信とパフォーマンスを改善するため、HDFS 構成を HDFS クラスタで正しく調整する必要があります。また、これにより、バックアップとリカバリのパフォーマンスも向上させることができます。

- バックアップとリストアのパフォーマンスを向上させるために、NetBackup では、使用中の Apache または Hadoop 分散からの Hadoop 構成の推奨事項に従うことをお勧めします。
- クラスタ内で Hadoop 暗号化を有効にしている場合は、使用中の Apache または Hadoop 分散の推奨事項に従って、Hadoop クラスタ内のデータ転送に使用する正しい暗号とビット長を選択します。
- データブロック転送中に AES 128 をデータ暗号化に使用すると、バックアップおよびリストア時の NetBackup のパフォーマンスが向上します。
- また、バックアップのパフォーマンスを向上させるために、Hadoop クラスタで複数のフォルダをバックアップする場合、バックアップホストの数を増やすこともできます。最大のメリットを得るには、Hadoop クラスタ内のフォルダごとに最大 1 つのバックアップホストを設定できます。
- また、バックアップ操作中に NetBackup が Hadoop クラスタからデータをフェッチするために使用されるバックアップホストごとのスレッド数を増やすこともできます。数十 GB のサイズ範囲のファイルがある場合は、パフォーマンスを向上するためにスレッドの数を増やすことができます。スレッドのデフォルト数は 4 です。
- 並列ストリームに使用されるバックアップホストごとのストリーム数を増やすこともできます。
- 配置に最適なデータ配布アルゴリズムのいずれかを選択できます。
  - データセットに含まれる少数の大きいファイルで、配布アルゴリズム 1 を使用します。
  - データセットに含まれる多数の小さいファイルで、配布アルゴリズム 2 を使用します。
  - サイズが非常に大きい少数のファイルとサイズが小さい多数のファイルがデータセットに混在する場合は、配布アルゴリズムとゴールデン比率の適切な組み合わせを使用します。次の例を参照してください。

表 4-3 多数の小さいファイルと少数の大きいファイルの例

データサイズ	バックアップホストの数	スレッド数	ストリーム数	配布アルゴリズム	ゴールデン比率
最大 1 TB	4	16	5	4	80
最大 50 TB	5	32	5	4	80
> 50 TB	6	32	5	4	80

詳しくは、『Apache Hadoop のマニュアル』のセキュアモードを参照してください。

さらに、パフォーマンスを最適化するために、次のことを確認します。

- プライマリサーバーはバックアップホストとして使用されていません。
- 複数のポリシーが並行してトリガされるようにスケジュールされている場合:
  - すべてのポリシーで同じ検出ホストを使用しないようにします。
- これらのポリシーでは、最後の Backup_Host エントリが異なります。

メモ: 検出ホストは、Backup_Host リストの最後のエントリです。

5

トラブルシューティング

この章では以下の項目について説明しています。

- NetBackup for NetBackup for Hadoop の問題のトラブルシューティングについて
- NetBackup for Hadoop のデバッグログについて
- NetBackup for Hadoop データのバックアップ問題のトラブルシューティング
- NetBackup for Hadoop データのリストア問題のトラブルシューティング

NetBackup for NetBackup for Hadoop の問題のトラブルシューティングについて

表 5-1 NetBackup for NetBackup for Hadoop の問題のトラブルシューティング

領域	参照
一般的なログとデバッグ	p.55 の「NetBackup for Hadoop のデバッグログについて」を参照してください。
バックアップの問題	p.56 の「NetBackup for Hadoop データのバックアップ問題のトラブルシューティング」を参照してください。
リストアの問題	p.61 の「NetBackup for Hadoop データのリストア問題のトラブルシューティング」を参照してください。

領域	参照
問題を避けるために、ベストプラクティスも確認	p.17 の「NetBackup for Hadoop プラグインを配備するためのベストプラクティス」を参照してください。 p.44 の「NetBackup for Hadoop クラスタのバックアップを作成するためのベストプラクティス」を参照してください。 p.46 の「Hadoop クラスタをリストアするためのベストプラクティス」を参照してください。

NetBackup for Hadoop のデバッグログについて

NetBackup は、バックアップ操作とリストア操作に関連するさまざまなプロセスのプロセス固有のログを保持します。これらのログを調べて、問題の根本原因を見つけることができます。

これらのログフォルダは、ログの記録用にあらかじめ存在している必要があります。これらのフォルダが存在しない場合は作成する必要があります。

次のディレクトリにあるログフォルダ

- Windows の場合: `install_path¥NetBackup¥logs`
- UNIX または Linux の場合: `/usr/openv/var/global/logs`

表 5-2 Hadoop に関連する NetBackup ログ

ログフォルダ	メッセージの内容	ログの場所
`install_path/NetBackup/logs/bpVMutil`	ポリシーの構成	プライマリサーバー
`install_path/NetBackup/logs/nbaapidiscv`	BigData フレームワーク、検出、および NetBackup for Hadoop 構成ファイルのログ	バックアップホスト
`install_path/NetBackup/logs/bpbrm`	ポリシー検証、バックアップ、およびリストア操作	メディアサーバー
`install_path/NetBackup/logs/bpbkar`	バックアップ	バックアップホスト
`install_path/NetBackup/logs/tar`	リストアおよび NetBackup for Hadoop 構成ファイル	バックアップホスト

詳しくは、『NetBackup ログリファレンスガイド』を参照してください。

NetBackup for Hadoop データのバックアップ問題のトラブルシューティング

次の項を参照してください。

- p.55 の「NetBackup for Hadoop のデバッグログについて」を参照してください。
- p.56 の「バックアップ操作がエラー 6609 で失敗する」を参照してください。
- p.56 の「バックアップ操作がエラー 6618 で失敗した」を参照してください。
- p.57 の「バックアップ操作がエラー 6647 で失敗する」を参照してください。
- p.57 の「Hadoop で拡張属性 (xattrs) とアクセス制御リスト (ACL) がバックアップまたはリストアされない」を参照してください。
- p.58 の「バックアップ操作がエラー 6654 で失敗する」を参照してください。
- p.58 の「バックアップ操作が bpbrm エラー 8857 で失敗する」を参照してください。
- p.59 の「バックアップ操作がエラー 6617 で失敗する」を参照してください。
- p.59 の「バックアップ操作がエラー 6616 で失敗する」を参照してください。

バックアップ操作がエラー 6609 で失敗する

このエラーは、次のシナリオ中に発生します。

1. NetBackup for Hadoop プラグインファイルが、バックアップホスト (1 つまたは複数) から削除されているか失われています。

   回避方法:

   NetBackup for Hadoop プラグインをダウンロードしてインストールします。

2. `Application_Type` の詳細が不正です。

   回避方法:

   `Hadoop` の代わりに `hadoop` を使用して、`Application_Type` を指定します。

バックアップ操作がエラー 6618 で失敗した

バックアップ操作がエラー 6618 で失敗し、次のエラーが表示されます。

```
NetBackup cannot find the file to complete the operation.(6618)
```

このエラーは、バックアップ選択として無効なディレクトリを指定した場合に発生します。

回避方法:

BigData ポリシーのバックアップ選択として有効なディレクトリを指定します。

バックアップ操作がエラー 6647 で失敗する

バックアップ操作がエラー 6647 で失敗し、次のエラーが表示されます。

```
Unable to create or access a directory or a path. (6647)
```

このエラーは、次の状況のいずれかで発生します。

- ディレクトリでスナップショットが有効になっていない
- ポリシーではバックアップ選択としてルートフォルダのスナップショットを作成するように設定されているが、子フォルダの 1 つですでにスナップショットが有効になっている
- ポリシーではバックアップ選択として子フォルダのスナップショットを作成するように設定されているが、親フォルダの 1 つですでにスナップショットが有効になっている
- ポリシーでバックアップ選択としてファイルのスナップショットを作成するように設定されている

回避方法:

NetBackup for Hadoop では、入れ子になったディレクトリでスナップショットを有効にすることは許可されていません。親ディレクトリですでにスナップショットが有効である場合、親ディレクトリの下の他の子ディレクトリではスナップショットを有効にできません。Bigdata ポリシー形式でのバックアップ選択については、バックアップに対してスナップショットが有効なディレクトリのみを選択する必要があります。その他の子ディレクトリを選択してはいけません。

Hadoop で拡張属性 (xattrs) とアクセス制御リスト (ACL) がバックアップまたはリストアされない

拡張属性によって、ユーザーアプリケーションは、追加のメタデータを Hadoop のファイルまたはディレクトリに関連付けることができます。これは、デフォルトでは、Hadoop 分散ファイルシステム (HDFS) で有効になっています。

アクセス制御リストによって、特定の名前付きユーザーや名前付きグループに対して、標準のアクセス許可も含めて、アクセス許可を個別に設定できます。これは、デフォルトでは、HDFS で無効になっています。

Hadoop プラグインは、バックアップ中に、オブジェクトの拡張属性またはアクセス制御リスト (ACL) をキャプチャしないため、リストアされるファイルやフォルダにはこれらは設定されません。

回避方法:

`Application_Type = hadoop` の BigData ポリシーを使用してバックアップが作成されるファイルまたはディレクトリのいずれかに拡張属性が設定されている場合、リストアされるデータに拡張属性を明示的に設定する必要があります。

拡張属性は、fs -getfattr や hadoop fs -setfattr など、Hadoop シェルコマンドを使用して設定できます。

Application_Type = hadoop の BigData ポリシーを使用してバックアップが作成されるファイルまたはディレクトリのいずれかでアクセス制御リスト (ACL) が有効で、設定されている場合、リストアされるデータに ACL を明示的に設定する必要があります。

ACL は、hadoop fs -getfacl や hadoop fs -setfacl など、Hadoop シェルコマンドを使用して設定できます。

バックアップ操作がエラー 6654 で失敗する

このエラーは、次のシナリオ中に発生します。

- NetBackup for Hadoop のクレデンシャルが NetBackup プライマリサーバーに追加されていない場合。
  回避方法:
  NetBackup for Hadoop のクレデンシャルが NetBackup プライマリサーバーに追加されていることを確認します。tpconfig コマンドを使用します。詳しくは、p.24 の「NetBackup での NetBackup for Hadoop クレデンシャルの追加」を参照してください。を参照してください。
- バックアップホストに NetBackup for Hadoop プラグインがインストールされていない場合
  回避方法:
  バックアップ操作を開始する前に、すべてのバックアップホストに確実に NetBackup for Hadoop プラグインファイルをインストールします。
- バックアップホストとして使用されている NetBackup クライアントが許可リストに載っていない場合
  回避方法:
  バックアップ操作を開始する前に、バックアップホストとして使用されている NetBackup クライアントが許可リストに載っていることを確認します。
  p.22 の「NetBackup プライマリサーバーの許可リストに NetBackup クライアントを含める」を参照してください。

バックアップ操作が bpbrm エラー 8857 で失敗する

このエラーは、NetBackup プライマリサーバーで NetBackup クライアントが許可リストに追加されていない場合に発生します。

回避方法:

バックアップホストとして NetBackup クライアントを使用する場合は、NetBackup プライマリサーバーで許可リストに追加する手順を実行する必要があります。詳しくは、p.22 の

「NetBackup プライマリサーバーの許可リストに NetBackup クライアントを含める」を参照してください。

バックアップ操作がエラー 6617 で失敗する

バックアップ操作がエラー 6617 で失敗し、次のエラーが表示されます。

システムコールに失敗しました。

Kerberos 対応 NetBackup for Hadoop クラスタの場合には、バックアップホストに有効なチケット認可チケット (TGT) があることを確認します。

回避方法:

TGT を更新します。

バックアップ操作がエラー 6616 で失敗する

バックアップ操作がエラー 6616 で失敗し、次のエラーがログに記録されます。

`hadoopOpenConfig: 構成ファイルから JSON オブジェクトを作成するのに失敗しました。`

回避方法:

`hadoop.conf` ファイルを検証して、パラメータ値に空白の値または不正な構文が使用されていないことを確認します。

バックアップ操作がエラー 84 で失敗する

バックアップ操作がエラー 84 メディア書き込みエラーで失敗しました。

回避方法:

- 有効なメディアサーバーを使用してバックアップを実行します。
- メディアサーバーストレージの 1 つを停止します。
- 完全バックアップを再度実行します。

コンテナベースの NetBackup Appliance を再起動した後、NetBackup 構成ファイルおよび証明書ファイルが保持されない

コンテナベースの NetBackup Appliance を何らかの理由で再起動した後、`hadoop.conf` または `hbase.conf` などの NetBackup 構成ファイル、または SSL 証明書や CRL パスが保持されません。この問題は、バックアップホストとしてコンテナベースの NetBackup Appliance を使用して Hadoop または HBase の作業負荷を保護する場合に該当します。

理由:

NetBackup Appliance 環境では、Docker ホストの永続的な場所で利用可能なファイルは再起動操作後も保持されます。hadoop.conf と hbase.conf ファイルはカスタム構成ファイルであり、永続的な場所に一覧表示されません。

構成ファイルは、フェールオーバー中の HA (高可用性) ノードやバックアップのスレッド数などの値を定義するために使用されます。これらのファイルが削除された場合、バックアップでは、HA とスレッド数の両方にデフォルト値 (それぞれ、プライマリ名ノードと 4) が使用されます。このようなケースでは、プライマリノードが停止した場合のみ、プラグインがセカンダリサーバーの検出に失敗するためバックアップは失敗します。

SSL 証明書と CRL パスのファイルが永続的ではない場所に格納されている場合、アプライアンスを再起動するとバックアップとリストア操作は失敗します。

回避方法:

Hadoop と HBase のカスタム構成ファイルが再起動後に削除された場合は、次の場所にファイルを手動で作成できます。

- Hadoop: /usr/openv/var/global/hadoop.conf
- HBase: /usr/openv/var/global/hbase.conf

Hadoop または HBase の SSL 証明書と CRL に署名した CA 証明書は、次の場所に格納できます。

/usr/openv/var/global/

バックアップイメージの選択でイメージが表示されているにもかかわらず、リストア時に増分バックアップイメージが表示されない

この問題は、増分バックアップイメージをリストアしようとしたときに、バックアップポリシーのバックアップ対象リストに、/ のサブフォルダ内のバックアップ対象が含まれている場合に発生します。

次に例を示します。

```
/data/1
/data/2
```

回避方法

増分バックアップイメージからリストアできる利用可能なデータを表示するには、増分バックアップイメージとともに関連する完全バックアップイメージを選択します。

子バックアップジョブの 1 つがキューに投入された状態になる

複数のバックアップホストがあるシナリオで、子バックアップジョブの 1 つがキューに投入された状態になり、メディアサーバーを待機し続けます。

理由:

この問題は、複数のバックアップホストが使用されていて、メディアサーバーが非アクティブな状態になっている NetBackup Appliance 環境で確認できます。

回避方法:

NetBackup Web UI を開きます。左側で［ストレージ (Storage)］、［メディアサーバー (Media servers)］の順に選択します。メディアサーバーを特定して選択します。次に、［有効化 (Activate)］をクリックします。

NetBackup for Hadoop データのリストア問題のトラブルシューティング

- p.61 の「リストアが 2850 エラーコードで失敗する」を参照してください。
- p.62 の「NetBackup の NetBackup for Hadoop のリストアジョブが部分的に完了する」を参照してください。
- p.57 の「Hadoop で拡張属性 (xattrs) とアクセス制御リスト (ACL) がバックアップまたはリストアされない」を参照してください。
- p.62 の「Hadoop プラグインファイルがバックアップホスト上にない場合、リストア操作が失敗する」を参照してください。
- p.62 の「リストアが bpbrm エラー 54932 で失敗する」を参照してください。
- p.62 の「リストア操作が bpbrm エラー 21296 で失敗する」を参照してください。

リストアが 2850 エラーコードで失敗する

このエラーは、次のシナリオで発生する場合があります。

- Error:`2850 "errno = 62 - Timer expired"`
  回避方法:
  ファイアウォールの設定を更新して、バックアップホストが NetBackup for Hadoop クラスタと通信できるようにします。
- 要求されたファイルはリカバリされません。
  回避方法:
  Kerberos 対応 NetBackup for Hadoop クラスタの場合には、バックアップホストに有効なチケット認可チケット (TGT) があることを確認します。
  TGT を更新します。
- アプリケーションサーバーの値が不正で、クレデンシャルが無効です。
  回避方法:
  リストア中に宛先 NetBackup for Hadoop クラスタのホスト名を正しく入力したことを確認してください。これは `tpconfig` コマンドで提供されたものと同じである必要があります。

NetBackup の NetBackup for Hadoop のリストアジョブが部分的に完了する

リストアデータの容量が NetBackup for Hadoop クラスタで使用できる領域を超えている場合、リストアジョブが部分的に完了します。

回避方法:

NetBackup for Hadoop クラスタの領域をクリーンアップします。

Hadoop で拡張属性 (xattrs) とアクセス制御リスト (ACL) がバックアップまたはリストアされない

この問題について詳しくは、p.57 の「Hadoop で拡張属性 (xattrs) とアクセス制御リスト (ACL) がバックアップまたはリストアされない」を参照してください。を参照してください。

Hadoop プラグインファイルがバックアップホスト上にない場合、リストア操作が失敗する

Hadoop プラグインファイルがインストールされていないバックアップホストでリストアジョブがトリガされると、リストア操作が次のエラーで失敗します。

```
client restore EXIT STATUS 50: client process aborted
```

回避策: NetBackup for Hadoop プラグインをダウンロードしてインストールします。

リストアが bpbrm エラー 54932 で失敗する

このエラーは、リストアするファイルのバックアップが正常に作成されていない場合に発生します。

回避方法:

リストア操作を開始する前に、バックアップ作成が正常に完了していることを確認します。

あるいは、アクティビティモニターメニューで[ジョブ状態]タブをクリックして、特定のジョブ ID を探し、エラーメッセージの詳細を確認します。

リストア操作が bpbrm エラー 21296 で失敗する

`<application_server_name>` に不正な値を指定し、Hadoop クレデンシャルを NetBackup プライマリサーバーに追加した場合に、このエラーが発生します。

回避方法:

`<application_server_name>` で指定される詳細が正しいかどうかを確認します。

Kerberos を使用した Hadoop のリストアジョブがエラー 2850 で失敗する

Kerberos を使用した Hadoop のリストアジョブがエラー 2850 で失敗します。この問題は、HDFS 所有者がファイルとディレクトリの所有権を設定していない場合、または Kerberos の構成に問題がある場合に発生します。

回避方法: リストアする前に、次のことを確認してください。

- Kerberos のバックアップに HDFS 所有者ユーザーが使用されていることを確認します。
- 現在の Kerberos ユーザーで、`chown` や `setfacl` などの HDFS コマンドを使用して、所有者/ACLS を手動で設定できることを確認します。

詳しくは、『NetBackup for Hadoop 管理者ガイド』を参照してください。

ディザスタリカバリ後に構成ファイルがリカバリされない

SSL 対応 (HTTPS) の NetBackup for Hadoop クラスタまたは NetBackup for Hadoop クラスタで高可用性のために NetBackup プライマリサーバーをバックアップホストとして使用する場合、完全なカタログリカバリを実行すると、`hadoop.conf` 構成ファイルがリカバリされません。

構成ファイルは手動で作成してください。構成ファイルには、次の形式を使用してください。

```
{
   "application_servers":
   {
    "primary.host.com":
    {
      "use_ssl":true
     "failover_namenodes":
     [
      {
       "hostname":"secondary.host.com",
       "use_ssl":true
       "port":11111
      }
     ],
     "port":11111
    }
   },
   "number_of_threads":5
}
```