

IBM SPSS Analytic Server
バージョン 2.1

管理者ガイド

IBM

注記

本書および本書で紹介する製品をご使用になる前に、13 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Analytic Server バージョン 2.1、リリース 1、モディフィケーション 0、および新しい版で明記されていない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM SPSS Analytic Server
Version 2.1
Administrator's Guide

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 概要	1	バージョン情報	11
アーキテクチャー	2	ログ・コレクター	11
バージョン 2.1 の管理者向けの新機能	2	一般的な問題	12
第 2 章 テナント管理	3	特記事項	13
命名規則	4	商標	15
第 3 章 ユーザーの作業開始	7		
第 4 章 Analytic Server ジョブ名	9		
第 5 章 トラブルシューティング	11		
ロギング	11		

第 1 章 概要

IBM® SPSS® Analytic Server は、IBM SPSS テクノロジーをビッグデータ・システムと組み合わせたビッグデータ分析ソリューションであり、使いやすい IBM SPSS ユーザー・インターフェースを使用して、以前は達成不可能であった規模で問題を解決できるようにします。

ビッグデータ分析が重要な理由

組織で収集されるデータ・ボリュームは飛躍的に増加しています。例えば、金融業や小売業では 1 年間 (または 2 年間や 10 年間) のすべての顧客取引があり、通信プロバイダーではコール・データ・レコード (CDR) やデバイス・センサーのデータがあり、インターネット会社では Web クロールの結果があります。

ビッグデータ分析は、以下が存在する場合に必要です。

- 特に構造化データと非構造化データの混合である場合の大量 (テラバイト単位、ペタバイト単位、エクサバイト単位) のデータ。
- 急速に変化/蓄積するデータ

また、ビッグデータ分析は、以下の場合に役立ちます。

- 多数 (数千) のモデルが構築されている
- モデルが頻繁に構築/更新される

課題

次のように多様な理由から、大量のデータを収集しているのと同じ組織で、そのデータを実際に利用するのが困難であることがよくあります。

- 従来型の分析製品のアーキテクチャーが分散型計算に適していない
- 既存の統計アルゴリズムが、ビッグデータを処理するように設計されていない (このようなアルゴリズムはデータがアルゴリズムに移動されることを予期しますが、ビッグデータの移動には高額なコストがかかります)。そのため、次のような問題が起こります。
- ビッグデータに対して最先端の分析を実行するには、新たなスキルとビッグデータ・システムについての徹底した知識が必要である。このようなスキルを持つ分析者は非常にわずかです。
- インメモリ・ソリューションは中規模の問題に対しては機能するが、本当のビッグデータに対応して十分に拡張することはできない。

解決策

Analytic Server は以下を提供します。

- Hadoop Map/Reduce などのビッグデータ・システムを HDFS 内のデータと共に活用するデータ中心型アーキテクチャー
- データに向かうよう設計された新しい統計アルゴリズムを取り込むための定義済みインターフェース
- 分析者がデータの分析に集中できるようにビッグデータ環境の詳細を非表示にする、使いやすい IBM SPSS ユーザー・インターフェース
- どの規模の問題にも対応して拡張できるソリューション

アーキテクチャー

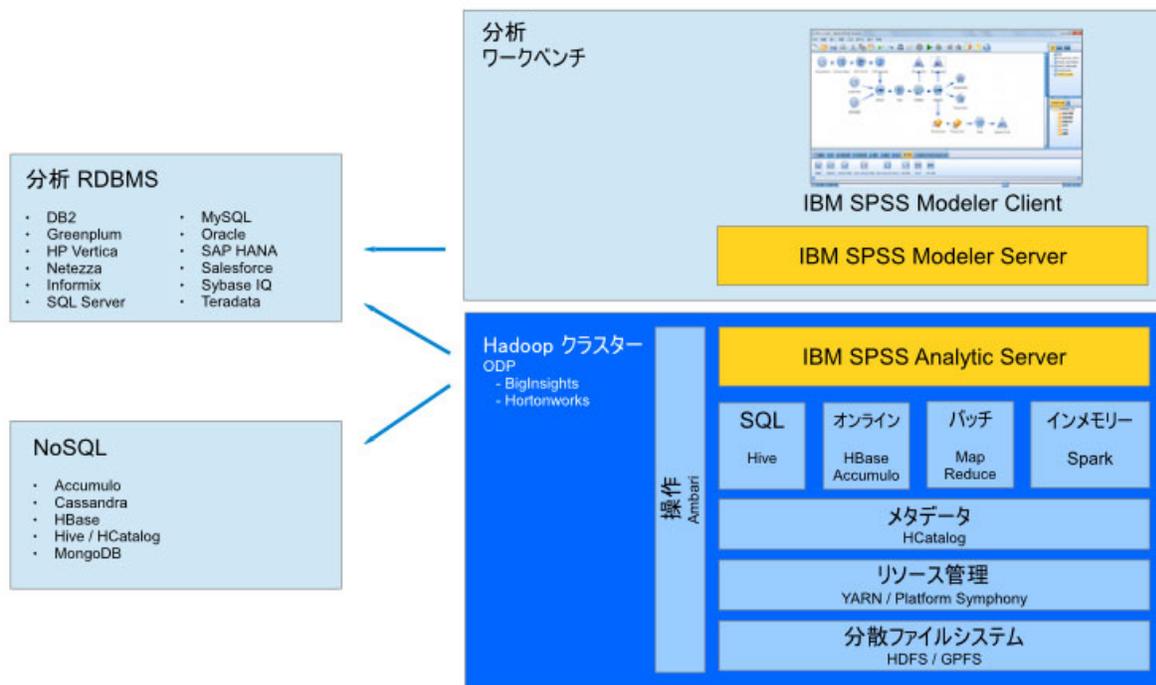


図 1. アーキテクチャー

Analytic Server はクライアント・アプリケーションと Hadoop クラウドの間に位置しています。データがクラウドにあることを想定して、Analytic Server での作業に関する概要は次のとおりです。

1. クラウドのデータに対して Analytic Server データ・ソースを定義します。
2. クライアント・アプリケーションで実行する分析を定義します。現行リリースでは、クライアント・アプリケーションは IBM SPSS Modeler です。
3. 分析の実行時に、クライアント・アプリケーションは Analytic Server 実行要求を実行依頼します。
4. Analytic Server は、Hadoop クラウドで実行されるジョブを調整して、結果をクライアント・アプリケーションに報告します。
5. 結果を使用して、さらなる分析を定義することができます。このサイクルが繰り返されます。

バージョン 2.1 の管理者向けの新機能

Analytic Server コンソール

読者役割

テナント内で、ユーザーやグループを読者役割に割り当てることができます。この役割は、Analytic Server にはログインできませんが、Modeler の Analytic Server ソース・ノードを介して Analytic Server のデータ・ソースを読み取ることができます。

第 2 章 テナント管理

テナントにより、オブジェクトをテナント間で共有できないように、ユーザー、プロジェクト、およびデータ・ソースの大まかな分類が提供されます。各ユーザーは、割り当てられているテナントに照らしてシステムにアクセスします。

Analytic Server コンソールでテナントを管理して、ユーザーをテナントに割り当てます。「テナント」ページのビューは、コンソールにログオンしているユーザーの役割によって異なります。

- インストール時にセットアップされる「スーパーユーザー」管理者は、テナント管理者です。このユーザーのみが新規テナントを作成して、すべてのテナントのプロパティを編集することができます。
- 管理者役割のユーザーは、ログインしているテナントのプロパティを編集できます。
- ユーザー役割のユーザーは、テナントのプロパティを編集できません。「テナント」ページは、これらのユーザーに対して非表示になります。
- 読者役割のユーザーは、データ・ソースを編集できません。また、Analytic Server コンソールにログインすることもできません。

管理者は、「プロジェクト」ページと「データ・ソース」ページにアクセスして、クリーンアップと管理のためにプロジェクトやデータ・ソースを管理できます。詳しくは、「*IBM SPSS Analytic Server ユーザーズ・ガイド*」を参照してください。

テナントのリスト表示

メインの「テナント」ページには、既存のテナントが表に表示されます。「スーパーユーザー」管理者のみがこのページで編集を行うことができます。

- テナントの名前をクリックして、その詳細を表示し、プロパティを編集します。
- テナントの URL をクリックして、そのテナントのコンテキストでコンソールを開きます。

注: コンソールからログアウトされ、そのテナントに有効な資格情報を使用してログインする必要があります。

- 名前に検索文字列が含まれるテナントのみを表示するようにリスト表示をフィルタリングするには、検索域に入力します。
- 「新規」をクリックして、「**新規テナントの追加 (Add new tenant)**」ダイアログで指定した名前でも新規テナントを作成します。テナントに付ける名前に関する制約事項については、4 ページの『命名規則』を参照してください。
- 選択したテナントを削除するには、「**削除**」をクリックします。
- リスト表示を更新するには、「**更新**」をクリックします。

個々のテナントの詳細

コンテンツ領域は、いくつかの縮小可能なセクションに分かれています。

詳細

名前 テナントの名前を表示する編集可能なテキスト・フィールド。

説明 テナントに関する説明テキストを指定できる編集可能なテキスト・フィールド。

URL これは、Analytic Server コンソールを使用してテナントにログインし、SPSS Modeler サー

パーの構成に使用するためにユーザーに指定する URL です。SPSS Modeler の構成について詳しくは、「*IBM SPSS Analytic Server* インストールと構成のガイド」を参照してください。

状態 「**アクティブ (Active)**」なテナントは現在使用されています。テナントを「**非アクティブ (Inactive)**」にすると、ユーザーがそのテナントにログインすることを防ぎますが、基本情報は削除されません。

プリンシパル

プリンシパルは、インストール時にセットアップされるセキュリティー・プロバイダーから引き出されるユーザーとグループです。管理者またはユーザーとしてプリンシパルをテナントに追加できます。

- テキスト・ボックスに入力すると、検索文字列が名前に含まれるユーザーとグループがフィルタリングされます。ドロップダウン・リストから「**管理者**」、「**ユーザー**」、または「**読者 (Reader)**」を選択して、テナント内の役割を割り当てます。「**参加者の追加 (Add participant)**」をクリックして、作成者のリストに追加します。
- 参加者を削除するには、メンバー・リストからユーザーまたはグループを選択して、「**参加者を削除 (Remove participant)**」をクリックします。

メトリック

テナントのリソースの制限を構成できます。テナントによって現在使用されているディスク・スペースを報告します。

- テナントの最大ディスク・スペース割り当て量を設定できます。この制限に達すると、テナントのディスク・スペース使用量が割り当て量を下回るように十分なディスク・スペースが消去されるまで、このテナントでディスクにこれ以上のデータを書き込むことはできません。
- テナントのディスク・スペース警告レベルを設定できます。この割り当て量を超えると、テナントのディスク・スペース使用量が割り当て量を下回るように十分なディスク・スペースが消去されるまで、このテナントでプリンシパルが分析ジョブを実行依頼することはできません。
- このテナントで一度に実行できる並行ジョブの最大数を設定できます。この割り当て量を超えると、現在実行中のジョブが完了するまで、このテナントでプリンシパルが分析ジョブを実行依頼することはできません。
- 1 つのデータ・ソースが保持できるフィールドの最大数を設定できます。データ・ソースが作成または更新されるたびに、この制限が確認されます。
- ファイルの最大サイズをメガバイトで設定できます。ファイルがアップロードされる際にこの制限が確認されます。

セキュリティー・プロバイダー構成

ユーザー認証プロバイダーを指定できます。「**デフォルト**」では、インストールおよび構成時にセットアップされたデフォルトのテナントのプロバイダーが使用されます。「**LDAP**」では、Active Directory や OpenLDAP などの外部 LDAP サーバーを使用してユーザーを認証できます。プロバイダーの設定を指定して、オプションで「**プリンシパル**」セクションで選択可能なユーザーとグループを制御するためのフィルター設定を指定します。

命名規則

データ・ソースやプロジェクトなど、Analytic Server で固有の名前を付けることができるすべてのものの名前には、以下の規則が適用されます。

- 1 つのテナント内では、同じタイプのオブジェクト内で名前が固有でなければなりません。例えば、2 つのデータ・ソースの両方に `insuranceClaims` という名前を付けることはできませんが、データ・ソースとプロジェクトのそれぞれに `insuranceClaims` という名前を付けることはできます。
- 名前では大文字と小文字が区別されます。例えば、`insuranceClaims` と `InsuranceClaims` は固有の名前と見なされます。
- 名前では、先頭と末尾の空白文字は無視されます。
- 以下の文字は、名前では無効です。
~, #, %, &, *, {, }, ¥¥, :, <, >, ?, /, |, ", ¥t, ¥r, ¥n

第 3 章 ユーザーの作業開始

http://<host>:<port>/<context-root>/admin/<tenant> にナビゲートし、ユーザー名およびパスワードを入力して Analytic Server コンソールにログオンするようにユーザーに通知します。

<host> Analytic Server ホストのアドレス。

<port> Analytic Server が listen するポート。デフォルトは 8080 です。

<context-root>

Analytic Server のコンテキスト・ルート。デフォルトは `analyticserver` です。

<tenant>

複数テナント環境では、ユーザーが所属するテナント。単一テナント環境の場合、デフォルトのテナントは `ibm` です。

例えば、ホスト・マシンの IP アドレスが 9.86.44.232 であり、"mycompany" テナントを作成してユーザーを追加済みで、その他の設定はデフォルトのままである場合、ユーザーは `http://9.86.44.232:8080/analyticserver/admin/mycompany` にナビゲートして Analytic Server コンソールにアクセスします。

第 4 章 Analytic Server ジョブ名

Analytic Server は、Hadoop クラスターのジョブ・トラッカーを使用してモニターできる map-reduce ジョブを作成します。

map-reduce ジョブ名は以下の構造になっています。

AS/{tenant name}/{user name}/{algorithm name}

{tenant name}

これは、ジョブが実行されるテナントの名前です。

{user name}

これは、ジョブを要求したユーザーです。

{algorithm name}

これは、ジョブの 1 次アルゴリズムです。単一のストリームが複数の map-reduce ジョブを生成する可能性があることに注意してください。同様に、1 つのストリーム内の複数の操作が単一の map-reduce ジョブに含まれる可能性があります。

第 5 章 トラブルシューティング

Analytic Server では、問題の判別に役立つ複数のツールを提供しています。

ロギング

Analytic Server は、問題の診断に役立つカスタマー・ログ・ファイルおよびトレース・ファイルを作成します。デフォルトの Liberty インストール済み環境では、これらのログ・ファイルは {AS_ROOT}/ae_wlpserver/usr/servers/aeserver/logs ディレクトリーにあります。

デフォルトのロギング構成では、2 つのログ・ファイルが作成されます。これらのログ・ファイルは、毎日ロールオーバーします。

as.log このファイルには、通知目的の警告およびエラー・メッセージの大まかな要約が含まれます。ユーザー・インターフェースに表示されるエラー・メッセージでは解決できないサーバー・エラーが発生した場合は、最初にこのファイルを確認してください。

as_trace.log

このファイルには、ae.log 内のすべてのエントリーが含まれますが、これに加え、主に IBM サポートおよび開発を対象としたデバッグ用の情報が追加されます。

Analytic Server は、基礎となるロギング機構として Apache LOG4J を使用します。LOG4J を使用することで、{AS_SERVER_ROOT}/configuration/log4j.xml 構成ファイルを編集してロギングを動的に調整できます。ロギングの変更は、問題の診断に役立つ目的でサポートから要請される場合も、保持されるログ・ファイルの数を制限するためにユーザーが行う場合もあります。このファイルに対する変更は数秒以内に自動的に検出されるため、Analytic Server を再始動する必要はありません。

log4j およびこの構成ファイルについて詳しくは、Apache 公式 Web サイトの資料 (<http://logging.apache.org/log4j/>) を参照してください。

バージョン情報

{AS_ROOT}/properties/version フォルダを調べると、どのバージョンの Analytic Server がインストールされているかが分かります。バージョン情報は、以下のファイルに含まれています。

IBM_SPSS_Analytic_Server-*.swtag

詳細な製品情報が記載されています。

version.txt

インストールされている製品のバージョンおよびビルド番号。

ログ・コレクター

ログ・ファイルを直接検討しても問題を解決できない場合は、すべてのログをバンドルして IBM サポートに送信することができます。必要なすべてのデータを容易に収集するためのユーティリティーが提供されています。

コマンド・シェルを使用して、以下のコマンドを実行します。

```
cd {AS_ROOT}/bin
run >sh ./logcollector.sh
```

これらのコマンドにより、{AS_ROOT}/bin に圧縮ファイルが作成されます。圧縮ファイルには、すべてのログ・ファイルと製品のバージョン情報が含まれています。

一般的な問題

このセクションでは、いくつかの一般的な管理上の問題と、それらの修正方法について説明します。

セキュリティー

HCatalog データ・ソースにアクセスしようとするとき Kerberos 認証が失敗する

以下のようなエラーがログに表示される場合:

```
cause:javax.security.sasl.SaslException: Failure to initialize security context
```

```
com.spss.analyticframework.api.exceptions.ComponentException: Cannot access HCatalog
```

HDFS ユーザーの Kerberos TGT がキャッシュされ、Analytic Server サーバー・ホスト上にあるようにしてください。これを行うには、以下の手順を実行します。

1. Analytic Server プロセスを停止します。
2. Analytic Server ホストから `kinit -f $hdfs.user` を実行します。ここで、`$hdfs.user` は `config.properties` ファイルに定義されたとおりであり、Analytic Server のルート・ディレクトリーに対する書き込み許可を持っています。
3. Analytic Server を開始します。

Analytic Server コンソール

iOS 上の Safari からの Analytic Server コンソールへのアクセス

テナント状況のドロップダウン・リストが機能せず、ファイル・ベースのデータ・ソースのデータ・モデルを更新できません。これらのアクションを実行する際は別のブラウザを使用してください。

ストリームの実行

PySpark ジョブを実行できない

1. Ambari コンソールで、YARN サービスの `yarn-env` パラメーターに `export SPARK_HOME=/usr/iop/current/spark-client` を追加します。
2. YARN サービスを再始動します。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

以下の保証は、国または地域の法律に沿わない場合は、適用されません。IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性があります。その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者をお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

表示されている IBM の価格は IBM が小売り価格として提示しているもので、現行価格であり、通知なしに変更されるものです。卸価格は、異なる場合があります。

本書はプランニング目的としてのみ記述されています。記述内容は製品が使用可能になる前に変更になる場合があります。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

© (お客様の会社名) (年). このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

© Copyright IBM Corp. _年を入れる_. All rights reserved.

この情報をソフトコピーでご覧になっている場合は、写真やカラーの図表は表示されない場合があります。

商標

IBM、IBM ロゴおよび ibm.com は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Adobe、Adobe ロゴ、PostScript、PostScript ロゴは、Adobe Systems Incorporated の米国およびその他の国における登録商標または商標です。

IT Infrastructure Library は AXELOS Limited の登録商標です。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子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。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

ITIL は AXELOS Limited の登録商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Cell Broadband Engine は、Sony Computer Entertainment, Inc. の米国およびその他の国における商標であり、同社の許諾を受けて使用しています。

Linear Tape-Open、LTO、LTO ロゴ、Ultrium および Ultrium ロゴは、HP、IBM Corp. および Quantum の米国およびその他の国における商標です。



Printed in Japan