Informatica

Informatica® Cloud Data Integration

Microsoft Azure Data Lake Storage Gen1 V3 コネクタ

Informatica Cloud Data Integration Microsoft Azure Data Lake Storage Gen1 V3 コネクタ
2021 年 2 月

© 著作権 Informatica LLC 2018, 2021

本ソフトウェアおよびマニュアルは、使用および開示の制限を定めた個別の使用許諾契約のもとでのみ提供されています。本マニュアルのいかなる部分も、いかなる手段（電子的複写、写真複写、録音など）によっても、Informatica LLC の事前の承諾なしに複製または転載することは禁じられています。

米政府の権利プログラム、ソフトウェア、データベース、および関連文書や技術データは、米国政府の顧客に配信され、「商用コンピュータソフトウェア」または「商業技術データ」は、該当する連邦政府の取得規制と代理店固有の補足規定に基づきます。このように、使用、複製、開示、変更、および適応は、適用される政府の契約に規定されている制限およびライセンス条項に従うものとし、政府契約の条項によって適当な範囲において、FAR 52.227-19、商用コンピュータソフトウェアライセンスの追加権利を規定します。

Informatica、Informatica ロゴ、Informatica Cloud、および PowerCenter は、米国およびその他の国における Informatica LLC の商標または登録商標です。Informatica の商標の最新リストは、Web（https://www.informatica.com/trademarks.html ）にあります。その他の企業名および製品名は、それぞれの企業の商標または登録商標です。

本ソフトウェアまたはドキュメンテーション（あるいはその両方）の一部は、第三者が保有する著作権の対象となります。必要な第三者の通知は、製品に含まれています。

特許については、https://www.informatica.com/legal/patents.html を参照してください。

免責: 本文書は、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は他社の権利の非侵害、市場性および特定の目的への適合性の黙示の保証などを含めて、一切の明示的および黙示的保証の責任を負いません。Informatica LLC では、本ソフトウェアまたはドキュメントに誤りのないことを保証していません。本ソフトウェアまたはドキュメントに記載されている情報には、技術的に不正確な記述や誤植が含まれる場合があります。本ソフトウェアまたはドキュメントの情報は、予告なしに変更されることがあります。

NOTICES

この Informatica 製品（以下「ソフトウェア」）には、Progress Software Corporation（以下「DataDirect」）の事業子会社である DataDirect Technologies からの特定のドライバ（以下「DataDirect ドライバ」）が含まれています。DataDirect ドライバには、次の用語および条件が適用されます。

1. DataDirect ドライバは、特定物として現存するままの状態で提供され、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。
2. DataDirect または第三者は、予見の有無を問わず発生した ODBC ドライバの使用に関するいかなる直接的、間接的、偶発的、特別、あるいは結果的損害に対して責任を負わないものとします。本制限事項は、すべての訴訟原因に適用されます。訴訟原因には、契約違反、保証違反、過失、厳格責任、詐称、その他の不法行為を含みますが、これらに限るものではありません。

本マニュアルの情報は、予告なしに変更されることがあります。このドキュメントで問題が見つかった場合は、infa_documentation@informatica.com までご報告ください。

Informatica 製品は、それらが提供される契約の条件に従って保証されます。Informatica は、商品性、特定目的への適合性、非侵害性の保証等を含めて、明示的または黙示的ないかなる種類の保証をせず、本マニュアルの情報を「現状のまま」提供するものとします。

発行日: 2021-03-16

目次

序文 5

Informatica のリソース 5

Informatica マニュアル 5

Informatica Intelligent Cloud Services Web サイト 5

Informatica Intelligent Cloud Services コミュニティ 5

Informatica Intelligent Cloud Services マーケットプレース 6

Data Integration コネクタのドキュメント 6

Informatica ナレッジベース 6

Informatica Intelligent Cloud Services Trust Center 6

Informatica グローバルカスタマサポート 6

第 1 章 : Microsoft Azure Data Lake Storage Gen1 V3 コネクタの概要 7

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの概要 7

Microsoft Azure Data Lake Storage Gen1 の概要 8

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの例 8

Microsoft Azure Data Lake Storage Gen1 V3 コネクタのタスクタイプおよびオブジェクトタイプ . . . 8

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの管理 9

第 2 章 : Microsoft Azure Data Lake Storage Gen1 V3 接続 10

Microsoft Azure Data Lake Storage Gen1 V3 接続の概要 10

Microsoft Azure Data Lake Storage Gen1 V3 接続のプロパティ 10

第 3 章 : Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用したマッピングとマッピングタスク 12

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用したマッピングの概要 12

マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ソース 13

マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ターゲット 15

ターゲットの指定 18

ターゲットのタイムスタンプ 19

Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルのパラメータ化 20

パラメータファイルによる Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルのパラメータ化 20

マッピングおよびマッピングタスクに関するルールとガイドライン 20

第 4 章 : トラブルシューティング 22

マッピングまたはマッピングタスクのトラブルシューティング 22

第 5 章 : Microsoft Azure Data Lake Storage Gen1 V3 データ型リファレンス 23

データ型リファレンスの概要 23

Avro のデータ型とトランスフォーメーションデータ型 24

JSON のデータ型とトランスフォーメーションデータ型. 25
Parquet のデータ型とトランスフォーメーションデータ型. 25

索引. 27

序文

*Microsoft Azure Data Lake Storage Gen1 V3 コネクタ*を使用して、Cloud データ統合を使って Microsoft Azure Data Lake Storage Gen1 に対して読み取りまたは書き込みを行う方法を確認します。Cloud データ統合で Microsoft Azure Data Lake Storage Gen1 V3 接続を作成する方法や、マッピングとマッピングタスクを開発および実行する方法を確認します。

Informatica のリソース

Informatica は、Informatica Network やその他のオンラインポータルを通じてさまざまな製品リソースを提供しています。リソースを使用して Informatica 製品とソリューションを最大限に活用し、その他の Informatica ユーザーや各分野の専門家から知見を得ることができます。

Informatica マニュアル

Informatica マニュアルポータルでは、最新および最近の製品リリースに関するドキュメントの膨大なライブラリを参照できます。マニュアルポータルを利用するには、https://docs.informatica.com にアクセスしてください。

製品マニュアルに関する質問、コメント、ご意見については、Informatica マニュアルチーム（infa_documentation@informatica.com）までご連絡ください。

Informatica Intelligent Cloud Services Web サイト

Informatica Intelligent Cloud Services Web サイト（http://www.informatica.com/cloud）にアクセスできます。このサイトには、Informatica Cloud 統合サービスに関する情報が含まれます。

Informatica Intelligent Cloud Services コミュニティ

Informatica Intelligent Cloud Services コミュニティを使用して、技術的な問題について議論し、解決します。また、技術的なヒント、マニュアルの更新情報、FAQ（よくある質問）への答えを得ることもできます。

次の Informatica Intelligent Cloud Services コミュニティにアクセスします。

https://network.informatica.com/community/informatica-network/products/cloud-integration

開発者は、次の Cloud 開発者コミュニティで詳細情報を確認したり、ヒントを共有したりできます。

https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers

Informatica Intelligent Cloud Services マーケットプレース

Informatica マーケットプレイスにアクセスすると、Data Integration コネクタ、テンプレート、およびマップレットを試用したり購入したりできます。

https://marketplace.informatica.com/

Data Integration コネクタのドキュメント

Data Integration コネクタのドキュメントには、マニュアルポータルからアクセスできます。マニュアルポータルを利用するには、https://docs.informatica.com にアクセスしてください。

Informatica ナレッジベース

Informatica ナレッジベースを使用して、ハウツー記事、ベストプラクティス、よくある質問に対する回答など、製品リソースを見つけることができます。

ナレッジベースを検索するには、https://search.informatica.com にアクセスしてください。ナレッジベースに関する質問、コメント、ご意見の連絡先は、Informatica ナレッジベースチーム（KB_Feedback@informatica.com）です。

Informatica Intelligent Cloud Services Trust Center

Informatica Intelligent Cloud Services Trust Center は、Informatica のセキュリティポリシーおよびリアルタイムでのシステムの可用性について情報を提供します。

Trust Center（https://www.informatica.com/trust-center.html）にアクセスします。

Informatica Intelligent Cloud Services Trust Center にサブスクライブして、アップグレード、メンテナンス、およびインシデントの通知を受信します。Informatica Intelligent Cloud Services Status ページには、すべての Informatica Cloud 製品の実稼働ステータスが表示されます。メンテナンスの更新はすべてこのページに送信され、停止中は最新の情報が表示されます。更新と停止の通知がされるようにするには、Informatica Intelligent Cloud Services の 1 つのコンポーネントまたはすべてのコンポーネントについて更新の受信をサブスクライブします。すべてのコンポーネントにサブスクライブするのが、更新を逃さないようにするための最良の方法です。

登録するには、https://status.informatica.com/に移動し、**［更新を購読登録］**をクリックします。その後、電子メール、SMS テキストメッセージ、Webhook、RSS フィードとして、またはこの 4 つを任意に組み合わせて送信された通知を受信することを選択ができます。

Informatica グローバルカスタマサポート

電話またはオンラインでカスタマサポートセンターに連絡できます。

オンラインサポートについては、Informatica Intelligent Cloud Services の**［サポート要求の送信］**をクリックしてください。またオンラインサポートを使用して問題を記録することもできます。オンラインサポートを利用するには、ログインが必要です。https://network.informatica.com/welcome でログイン要求できます。

Informatica グローバルカスタマサポートの電話番号は、Informatica の Web サイト https://www.informatica.com/services-and-training/support-services/contact-us.html に掲載されています。

第 1 章

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの概要

この章では、以下の項目について説明します。

- Microsoft Azure Data Lake Storage Gen1 V3 コネクタの概要, 7 ページ
- Microsoft Azure Data Lake Storage Gen1 の概要, 8 ページ
- Microsoft Azure Data Lake Storage Gen1 V3 コネクタの例, 8 ページ
- Microsoft Azure Data Lake Storage Gen1 V3 コネクタのタスクタイプおよびオブジェクトタイプ, 8 ページ
- Microsoft Azure Data Lake Storage Gen1 V3 コネクタの管理, 9 ページ

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの概要

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用すると、データ統合から Data Integration に接続できます。

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用すると、.csv、バイナリ、Avro、JSON、Parquet の各ファイルを Microsoft Azure Data Lake Storage Gen1 から読み取ったり、書き込んだりできます。Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトをマッピングと mapping タスクのソースおよびターゲットとして使用することができます。

Avro または Parquet ファイルを使用する場合は、Secure Agent で有効になっている Cloudera CDH バージョン 5.8 以降のライセンスが 1 つであることを確認します。Avro または Parquet ファイルに対してデータの読み取りと書き込みを行う際に指定できるのは、Linux システムで実行されている Secure Agent のみです。

ネストされた、またはインデントされた複数行の JSON ファイルを読み書きすることはできません。

Microsoft Azure Data Lake Storage Gen1 の概要

Microsoft Azure Data Lake Storage Gen1 を使用すると、データをサイズ、構造、および形式に関係なく格納できます。Microsoft Azure Data Lake Storage Gen1 では、大量のデータを処理して、早期にビジネスの成果を上げることができます。

Microsoft Azure Data Lake Storage Gen1 は、以下のプライマリサービスを組み合わせたものです。

- Microsoft Azure Data Lake Storage Gen1: Microsoft Azure Data Lake Storage Gen1 は、組織の構造化データ、半構造化データ、および非構造化データをネイティブ形式で格納する、Hadoop 分散ファイルシステム（HDFS）互換のデータリポジトリです。データサイエンティストやデータアナリストは、Data Lake のデータを使用して特定のパターンがないか調べたうえで、分析したデータをデータウェアハウスに移動できます。
- Azure HDinstight: Azure HDInsight は、クラウドに設定される Apache Hadoop 分散システムです。HDinsight は、Hadoop システムの下で Apache HBase、Apache Spark、Apache Storm などの Hadoop コンポーネントをサポートしています。
- Data Lake Analytics: Data Lake Analytics は、分散インフラストラクチャを管理します。Microsoft Azure Data Lake Store Analytics では、リソースを動的に割り当てたり、割り当てを解除したりできるため、使用したサービスに対して料金を支払うことができます。

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの例

あなたは、大手金融企業でデータアナリストとして働いています。Microsoft Azure Data Lake Analytics を使用してリスク管理、不正検出、およびその他の分析を行っています。分析を行うには、区切りデータを Microsoft Azure Data Lake Storage Gen1 に書き込む必要があります。

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用して、リレーショナルデータベースやトランザクションデータベース、Salesforce などの他のアプリケーションなどのソースからデータを読み取り、Microsoft Azure Data Lake Storage Gen1 にデータを書き込むための mapping タスクを作成できます。Microsoft Azure Data Lake Storage Gen1 でデータを利用できるようになったら、データ分析を実行できます。

Microsoft Azure Data Lake Storage Gen1 V3 コネクタのタスクタイプおよびオブジェクトタイプ

次の表は、Data Integration タスクで使用できる Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトタイプを示しています。

タスクタイプ	ソース	ターゲット	ルックアップ
Mapping	○	はい	×
Mapping タスク	○	はい	×

Microsoft Azure Data Lake Storage Gen1 V3 コネクタの管理

タスクで Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトを使用する前に、管理者は以下のタスクを実行する必要があります。

- Microsoft Azure Data Lake Storage Gen1 アカウントを作成します。
- Microsoft Azure Data Lake Storage Gen1 および Microsoft Azure Data Lake Storage Gen1 のディレクトリへのアクセスをユーザーに許可します。
- Data Integration 組織に SDKPatch パッケージのライセンスがあることを確認する。
- Microsoft Azure Data Lake Storage Gen1 でのサービス間認証用に Azure AD Web アプリケーションを作成します。Web アプリケーションのサービス間認証については、https://kb.informatica.com/howto/6/Pages/20/512374.aspx を参照してください。

  注: コネクタを使用してアプリケーションで作成されたフォルダまたはファイルにアクセスするためのスーパーユーザー特権があることを確認します。
- Spring 2020 にアップグレードした後、変更せずに既存の接続を使用して既存のマッピングを実行できることを確認します。新しいマッピングで使用する新しい接続を作成します。既存のマッピングで使用されている接続では［**サブフォルダの表示**］を有効にしないでください。既存の接続で［**サブフォルダの表示**］を有効にすると、基盤のマッピングが失敗するか、既存のマッピングを実行する前にオブジェクトをインポートし直す必要があります。

第 2 章

Microsoft Azure Data Lake Storage Gen1 V3 接続

この章では、以下の項目について説明します。

- Microsoft Azure Data Lake Storage Gen1 V3 接続の概要, 10 ページ
- Microsoft Azure Data Lake Storage Gen1 V3 接続のプロパティ, 10 ページ

Microsoft Azure Data Lake Storage Gen1 V3 接続の概要

Data Integration から Microsoft Azure Data Lake Storage Gen1 データにアクセスするための Microsoft Azure Data Lake Storage Gen1 V3 接続を作成します。接続は、［接続］ページか、タスクの作成時に作成できます。接続を作成すると、組織へのアクセス権を持つすべてのユーザーがその接続を使用できるようになります。

mapping タスクを作成するときに接続を使用します。Microsoft Azure Data Lake Storage Gen1 V3 のソースとターゲットをマッピングまたは mapping タスクに指定できます。

Microsoft Azure Data Lake Storage Gen1 V3 接続のプロパティ

Microsoft Azure Data Lake Storage Gen1 V3 接続をセットアップするときは、接続プロパティを設定する必要があります。

以下の表に、Microsoft Azure Data Lake Storage Gen1 V3 接続のプロパティを示します。

接続プロパティ	説明
接続名	Microsoft Azure Data Lake Storage Gen1 V3 接続の名前。
説明	接続の説明。説明は、765 文字を超えることはできません。
タイプ	接続タイプ。Microsoft Azure Data Lake Storage Gen1 V3 接続を選択します。

接続プロパティ	説明
ランタイム環境	タスクを実行するランタイム環境の名前。Secure Agent、Hosted Agent、またはサーバーレスランタイム環境を指定します。
ADLS アカウント名	Microsoft Azure Data Lake Storage Gen1 アカウントの名前。
クライアント ID	Active Directory で OAuth 認証を完了するためのアプリケーションの ID。
クライアントシークレット	Active Directory で OAuth 認証を完了するためのクライアント秘密鍵。
ディレクトリ	データの読み取りまたは書き込みに使用する Microsoft Azure Data Lake Storage Gen1 ディレクトリ。デフォルトはルートディレクトリです。
AuthEndpoint	クライアント ID およびクライアントシークレットに基づく認証が完了する OAuth 2.0 トークンエンドポイント。
サブフォルダの表示	指定したディレクトリのサブフォルダからオブジェクトをインポートするために、サブフォルダを有効にします。

クライアント ID、クライアントシークレット、AuthEndpoint の作成の詳細については、*Microsoft Azure Data Lake Store のドキュメント*を参照してください。

第 3 章

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用したマッピングとマッピングタスク

この章では、以下の項目について説明します。

- Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用したマッピングの概要, 12 ページ
- マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ソース, 13 ページ
- マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ターゲット, 15 ページ
- Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルのパラメータ化, 20 ページ
- マッピングおよびマッピングタスクに関するルールとガイドライン, 20 ページ

Microsoft Azure Data Lake Storage Gen1 V3 コネクタを使用したマッピングの概要

マッピングを作成するには Mapping Designer を使用します。マッピングを作成するときには、単一の Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトを表すようにソースまたはターゲットを設定します。

Secure Agent がターゲットにデータを書き込む前に、ソースとターゲットからのデータのフローと必要なトランスフォーメーションについて説明します。mapping タスクを作成するときに、使用するマッピングを選択します。Mapping タスクウィザードを使用して、mapping タスクを作成します。mapping タスクは、マッピングで定義したデータフローロジックに基づいてデータを処理します。

マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ソース

マッピングでは、単一の Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトを表すようにソーストランスフォーメーションを設定できます。

以下の表に、ソーストランスフォーメーションで設定できる Microsoft Azure Data Lake Storage Gen1 V3 ソースプロパティを示します。

プロパティ	説明
接続	ソース接続の名前。
ソースタイプ	［単一オブジェクト］または［パラメータ］を選択します。

プロパティ	説明
オブジェクト	ソースオブジェクトの名前。指定したディレクトリまたはサブディレクトリからオブジェクトを選択できます。また、ソースオブジェクトとして圧縮 gzip フラットファイルを選択することもできます。
形式オプション	必須。Microsoft Azure Data Lake Storage Gen1 V3 形式オプション。**［形式オプション］** ダイアログボックスを開き、ファイルの形式を定義します。 次の形式オプションを設定します。 - **入力形式**: ファイル入力形式を選択します。［なし］、［区切りファイル］、［Avro］、［JSON］、［Parquet］ のファイル形式を選択できます。デフォルトは［なし］です。 **注**: バイナリファイルを読み取るには、**［入力形式］** で **［なし］** を選択する必要があります。 - **スキーマソース**: スキーマのソースを指定します。**［データファイルから読み取る］** または **［スキーマファイルからインポートする］** を選択できます。 ［入力形式］ で Avro、JSON、または Parquet を選択し、**［データファイルから読み取る］** オプションを選択した場合は、区切り文字、エスケープ文字、修飾子のオプションを設定できません。 ［入力形式］ で Avro、JSON、または Parquet を選択し、**［スキーマファイルからインポートする］** オプションを選択した場合、**［スキーマファイル］** プロパティフィールドでアップロードできるのはスキーマファイルのみです。区切り記号、エスケープ文字、および修飾子の各オプションを設定することはできません。 - **スキーマファイル**: スキーマファイルをアップロードできます。 - **区切り文字**: データのカラムを区切るために使用される文字。カンマ、タブ、コロン、セミコロンなどのパラメータを設定できます。 **注**: **［区切り文字］** フィールドで区切り文字としてタブを直接設定することはできません。タブを区切り文字として設定するには、任意のテキストエディタでタブ文字を入力する必要があります。その後、そのタブ文字をコピーして **［区切り文字］** フィールドに貼り付けます。 - **エスケープ文字**: 引用符で囲まれていない文字列におけるカラム区切り文字の直前の文字、または引用符で囲まれた文字列内における引用符の直前の文字。 - **修飾子**: テキスト文字列の境界を定義する引用符。単一引用符や二重引用符などのパラメータを設定できます。 **注**: データに区切り文字の値が存在する場合は、出力テキスト修飾子を使用できます。 - **コードページ**: Secure Agent がデータの読み取りまたは書き込みに使用する必要があるコードページを選択します。 Microsoft Azure Data Lake Storage Gen1 V3 残りのコードページは無視します。 - **［ヘッダー行番号］** : Microsoft Azure Data Lake Storage Gen1 からのデータの読み取り時にヘッダーとして使用する行番号を指定します。ヘッダーのないファイルからデータを読み取ることもできます。ヘッダーのないファイルからデータを読み取るには、**［ヘッダー行番号］** フィールドの値を 0 に指定します。

プロパティ	説明
	- **［最初のデータ行］**：Secure Agent がデータの読み取りを開始する行番号を指定します。1 以上の値を入力する必要があります。 ヘッダーからデータを読み取るには、**［ヘッダー行番号］**と**［最初のデータ行］**のフィールドに同じ値を指定する必要があります。デフォルトは 1 です。 - **［ターゲットヘッダー］**：このプロパティは、Microsoft Azure Data Lake Storage Gen1 ソースからデータを読み取る場合には適用されません。 - **［分散カラム］**：このプロパティは、Microsoft Azure Data Lake Storage Gen1 ソースからデータを読み取る場合には適用されません。 - **［プレビューする最大行数］**：このプロパティは、Microsoft Azure Data Lake Storage Gen1 ソースからデータを読み取る場合には適用されません。 - **［行区切り文字］**：このプロパティは、Microsoft Azure Data Lake Storage Gen1 ソースからデータを読み取る場合には適用されません。

以下の表に、Microsoft Azure Data Lake Storage Gen1 V3 ソース詳細プロパティを示します。

プロパティ	説明
ディレクトリパスのオーバーライド	データの読み取りに使用する Microsoft Azure Data Lake Storage Gen1 ディレクトリ。 実行時に指定されたディレクトリパスで、接続の作成時に指定されたパスが上書きされます。
ファイル名のオーバーライド	ソースオブジェクト。データの読み取り元のファイルを選択します。 実行時に指定したファイルによって、オブジェクトに指定したファイルが上書きされます。
トレースレベル	なし

マッピングでの Microsoft Azure Data Lake Storage Gen1 V3 ターゲット

マッピングでは、Microsoft Azure Data Lake Storage Gen1 V3 オブジェクトをターゲットとして使用できます。

Microsoft Azure Data Lake Storage Gen1 V3 ターゲットオブジェクトを使用する場合は、Microsoft Azure Data Lake Storage Gen1 V3 接続をターゲットとして選択できます。マッピングウィザードの［ターゲット］ページで、Microsoft Azure Data Lake Storage Gen1 V3 ターゲットプロパティを設定できます。Microsoft Azure Data Lake Storage Gen1 にデータを書き込む場合は、［ターゲットの作成］フィールドを使用して、実行時にターゲットを作成できます。ソースに基づいて新しいターゲットを作成する場合は、フィールドマッピングからすべてのバイナリフィールドを削除する必要があります。

以下の表に、ターゲットトランスフォーメーションで設定できる Microsoft Azure Data Lake Storage Gen1 V3 ターゲットプロパティを示します。

プロパティ	説明
接続	ターゲット接続の名前。
ターゲットタイプ	**[単一オブジェクト]** を選択します。
オブジェクト	ターゲットオブジェクトの名前。指定したディレクトリまたはサブディレクトリから既存のオブジェクトを選択するか、実行時に新規ターゲットを作成できます。 **[実行時に新規作成]** を選択した場合は、ターゲットオブジェクトの名前とパスを入力し、使用するソースフィールドを選択します。デフォルトでは、すべてのソースフィールドが使用されます。**[パス]** 属性には適用されません。 ターゲット名に英数字を含めることはできません。ファイル名では、ピリオド（.）、アンダースコア（_）、アットマーク（@）、ドル記号（$）、パーセント記号（%）のみを特殊文字として使用できます。 ターゲット名では、パラメータファイルに定義したパラメータを使用できます。

プロパティ	説明
形式オプション	必須。Microsoft Azure Data Lake Storage Gen1 V3 形式オプション。**［形式オプション］**ダイアログボックスを開き、ファイルの形式を定義します。 次の形式オプションを設定します。 - **入力形式**: ファイル入力形式を選択します。［なし］、［区切りファイル］、［Avro］、［JSON］、［Parquet］のファイル形式を選択できます。デフォルトは［なし］です。 **注**: バイナリファイルを読み取るには、**［入力形式］**で**［なし］**を選択する必要があります。 - **スキーマソース**: スキーマのソースを指定します。**［データファイルから読み取る］**または**［スキーマファイルからインポートする］**を選択できます。 ［入力形式］で Avro、JSON、または Parquet を選択し、**［データファイルから読み取る］**オプションを選択した場合は、区切り文字、エスケープ文字、修飾子のオプションを設定できません。 ［入力形式］で Avro、JSON、または Parquet を選択し、**［スキーマファイルからインポートする］**オプションを選択した場合、**［スキーマファイル］**プロパティフィールドでアップロードできるのはスキーマファイルのみです。区切り文字、エスケープ文字、および修飾子の各オプションを設定することはできません。 - **スキーマファイル**: スキーマファイルをアップロードできます。［ターゲットの作成］オプションを選択した場合は、スキーマファイルをアップロードできません。 - **区切り文字**: データのカラムを区切るために使用される文字。カンマ、タブ、コロン、セミコロンなどのパラメータを設定できます。 **注**: **［区切り文字］**フィールドで区切り文字としてタブを直接設定することはできません。タブを区切り文字として設定するには、任意のテキストエディタでタブ文字を入力する必要があります。その後、そのタブ文字をコピーして**［区切り文字］**フィールドに貼り付けます。 - **エスケープ文字**: 引用符で囲まれていない文字列におけるカラム区切り文字の直前の文字、または引用符で囲まれた文字列内における引用符の直前の文字。 - **修飾子**: テキスト文字列の境界を定義する引用符。単一引用符や二重引用符などのパラメータを設定できます。 **注**: ターゲットにデータを書き込むために、データに区切り文字の値があるソースからデータを読み取る場合に、テキスト修飾子を使用します。それ以外の場合、Secure Agent はターゲットにテキスト修飾子を追加しません。 - **コードページ**: Secure Agent がデータの読み取りまたは書き込みに使用する必要があるコードページを選択します。 Microsoft Azure Data Lake Storage Gen1 V3 残りのコードページは無視します。 - **［ヘッダー行番号］**: Microsoft Azure Data Lake Stor からのデータの読み取り時にヘッダーとして使用する行番号を指定します。ヘッダーのないファイルからデータを読み取ることもできます。ヘッダーのないファイルからデータを読み取るには、**［ヘッダー行番号］**フィールドの値を 0 に指定します。 **注**: このプロパティは、データプレビューを実行する場合に適用されます。 - **［最初のデータ行］**: Secure Agent がデータの読み取りを開始する行番号を指定します。1 以上の値を入力する必要があります。 ヘッダーからデータを読み取るには、**［ヘッダー行番号］**と**［最初のデータ行］**のフィールドに同じ値を指定する必要があります。デフォルトは 1 です。 **注**: このプロパティは、データプレビューを実行する場合に適用されます。 - **［ターゲットヘッダー］**: このプロパティは、Microsoft Azure Data Lake Storage Gen1 ターゲットからデータを読み取る場合には適用されません。 - **［分散カラム］**: このプロパティは、Microsoft Azure Data Lake Storage Gen1 ターゲットにデータを書き込む場合には適用されません。 - **［プレビューする最大行数］**: このプロパティは、Microsoft Azure Data Lake Storage Gen1 ターゲットにデータを書き込む場合には適用されません。

プロパティ	説明
	- **[行区切り文字]**：データの行を区切るために使用される文字。値には、\r、\n、\r\n を設定できます。
操作	ターゲット操作。**[挿入]** を選択します。Microsoft Azure Data Lake Storage Gen1 ターゲットにデータを挿入することのみが可能です。

以下の表に、Microsoft Azure Data Lake Storage Gen1 V3 の詳細ターゲットプロパティを示します。

詳細ターゲットプロパティ	説明
ディレクトリパスのオーバーライド	データの書き込みに使用する Microsoft Azure Data Lake Storage Gen1 ディレクトリ。 ディレクトリが存在しない場合には、Secure Agent によって作成されます。実行時に指定されたディレクトリパスで、接続の作成時に指定されたパスが上書きされます。
ファイル名上書き	ターゲットオブジェクト。データの書き込み元のファイルを選択します。 実行時に指定したファイルによって、オブジェクトに指定したファイルが上書きされます。
新たに作成されたファイルの圧縮	True に設定すると、新たに作成されたファイルを gzip 形式で圧縮します。フラットファイルにのみ適用されます。
ファイルが存在する場合	Microsoft Azure Data Lake Storage Gen1 にファイルが存在する場合は、既存のファイルを上書きするか、追加するかを選択できます。現在のタスクを失敗させることもできます。 **注**: 追加はフラットファイルのみに適用されます。
Forward Rejected Rows	該当せず

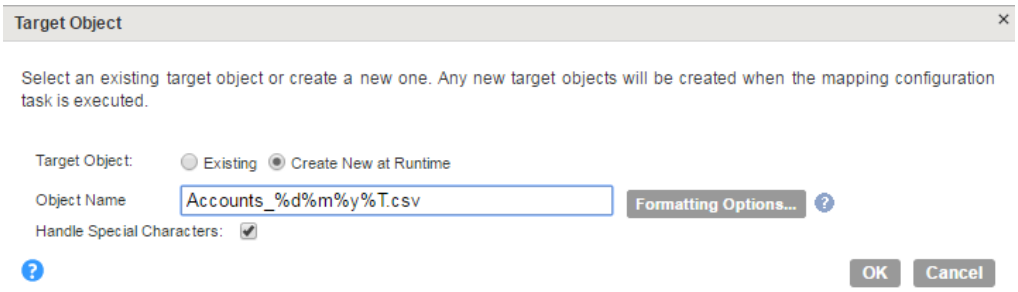
ターゲットの指定

既存のターゲットを使用するか、ターゲットを作成して、マッピングの結果を格納するためにことができます。ターゲットを作成することを選択した場合、エージェントはタスクの実行時にターゲットを作成します。

ターゲットのプロパティを指定するには、次の手順に従います。

1. マッピングでターゲットトランスフォーメーションを選択します。
2. **[受信フィールド]** タブで、ターゲットに含めるフィールドを指定するフィールドルールを設定します。
3. ターゲットを指定するには、**[ターゲット]** タブをクリックします。
4. ターゲット接続を選択します。
5. ターゲットタイプとして **[単一オブジェクト]** または **[パラメータ]** を選択します。
6. ターゲットオブジェクトまたはパラメータを指定します。
   - 実行時にターゲットファイルを作成するには、Accounts.csv のように、拡張子を含めてターゲットファイルの名前を入力します。
   - ファイル名にタイムスタンプを含める場合は、**[特殊文字の処理]** をクリックして、ファイル名に特殊文字を追加します。例えば、すべてのタイムスタンプ情報を含める場合は、Accounts_%d%m%y%T.csv のように指定します。

7. ファイルの形式オプションを設定する場合は、**[形式オプション]** をクリックし、**[OK]** をクリックします。
8. **[選択]** をクリックしてターゲットオブジェクトを選択します。既存のターゲットオブジェクトを選択するか、実行時に新しいターゲットオブジェクトを作成し、オブジェクト名を指定することができます。



9. 必要に応じて、ターゲットの詳細プロパティを指定します。

ターゲットのタイムスタンプ

実行時にターゲットを作成する場合、タイムスタンプ情報をファイル名に追加して、ファイルがいつ作成されたのかを示すことができます。

ターゲットファイルのファイル名を指定する場合は、Linux STRFTIME 関数は、mapping タスクでファイル名にタイムスタンプ情報を含めるために使用される形式です。タイムスタンプは、組織のタイムゾーンに基づいています。

次の表に、マッピングまたは mapping タスクで使用される一般的な STRFTIME 関数の形式を示します。

特殊文字	説明
%d	2 桁の 10 進数で表される日（01～31）。
%m	2 桁の 10 進数で表される月（01～12）。
%y	2 桁の 10 進数で表される、世紀を含まない年（00～99）。
%Y	世紀を含む年（2015 など）。
%T	24 時間表記の時刻（%H:%:M:%S に相当）。
%H	24 時間表記の時間（00～24）。
%l	12 時間表記の時間（01～12）。
%M	10 進数で表される分（00～59）。
%S	10 進数で表される秒（00～60）。
%p	AM または PM。

Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルのパラメータ化

Microsoft Azure Data Lake Storage Gen1 V3 ターゲットオブジェクトのファイル名とターゲットフォルダの場所をパラメータ化すると、実行時にファイル名とフォルダの場所を渡すことができます。フォルダがない場合は、Secure Agent によってフォルダ構造が動的に作成されます。

パラメータファイルによる Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルのパラメータ化

パラメータファイルを使用して、Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルをパラメータ化できます。

パラメータファイルを使用して Microsoft Azure Data Lake Storage Gen1 V3 ターゲットファイルをパラメータ化するには、Microsoft Azure Data Lake Storage Gen1 V3 ターゲットオブジェクトを作成し、ターゲットオブジェクト名とターゲットオブジェクトパスにパラメータを追加します。ターゲットオブジェクトに追加したパラメータをパラメータファイルで定義します。次に、パラメータファイルを次の場所に配置してマッピングタスクを実行します。

```
<Informatica Cloud Secure Agent\apps\Data_Integration_Server\data\userparameters>
```

マッピングおよびマッピングタスクに関するルールとガイドライン

マッピングおよびマッピングタスクを設定するときは、次のルールとガイドラインを考慮します。

- メタデータを編集すると、ネイティブデータ型はすべて Bigint 型に変わり、文字列データ型を除くデータ型のスケールと精度を変更できなくなります。
- JSON ファイルを Microsoft Azure Data Lake Storage Gen1 に書き込む場合は、カラム名に Unicode 文字が含まれていないことを確認します。
- Microsoft Azure Data Lake Storage Gen1 との間でデータの読み取りまたは書き込みを行う場合は、ターゲットファイル内で double データ型のスケールが 0 に設定されていないことを確認します。
- 実行時に新規オブジェクトを作成し、オブジェクト名に%T 時刻形式を使用した場合、作成されたオブジェクトの名前が%H_%M_%S 形式で表示されます。
- Avro、JSON および Parquet のファイル形式を書き込み、実行時に新しいオブジェクトを作成するときは、オブジェクト名に特殊文字が含まれていないことを確認します。
- ソースオブジェクトまたはターゲットのフィールド名に特殊文字や Unicode 文字が含まれていないことを確認します。
- 実行時にターゲットの作成を使用して JSON ファイルに書き込むと、ブール値はソースファイル内の true または false でなく、1 または 0 として書き込まれます。
- JSON ファイルを Microsoft Azure Data Lake Storage Gen1 ソースから読み取り、JSON ファイルを Microsoft Azure Data Lake Storage Gen1 ターゲットに書き込むと、Decimal データ型フィールドの精度がターゲットファイル内で変更されます。

- Avro ファイルまたは Parquet ファイルを書き込む場合は、ファイルに NULL 値が含まれていないことを確認します。NULL 値がある場合、ターゲットの NULL 値が含まれるフィールドに誤ったデータが書き込まれます。
- Microsoft Azure Data Lake Storage Gen1 ターゲットにデータを書き込むときに形式オプションで区切り文字としてタブを選択した場合は、データをプレビューできません。

第 4 章

トラブルシューティング

- マッピングまたはマッピングタスクのトラブルシューティング, 22 ページ

マッピングまたはマッピングタスクのトラブルシューティング

JSON ファイルを書き込むと、Java ヒープ領域エラーが発生してマッピングタスクが失敗します。

サイズが 1 GB 以上の JSON ファイルを書き込むと、Java ヒープ領域エラーが発生してタスクが失敗します。

タイプ DTM の JVM オプションを設定して、Secure Agent のシステム構成の詳細で-Xms および-Xmx 値を増やします。

[実行時に新規作成] オプションを使用して Avro ファイルを書き込むと、スキーマは NULL 値を含めるオプションを持たないプリミティブデータ型を使用して作成されます。

スキーマを手動で編集し、必要に応じて NULL 値を許可する必要があります。以下に例を示します。

```
{"type":"record","name":"Azure_Avro_CT_N","fields":[
{ "name":"c_custkey" , "type":["int","null"]},
{ "name":"c_name" , "type":"string"}
{"name":"c_address" , "type":"string"},
{ "name":"c_nationkey" , "type":["long","null"]}
```

マッピングが失敗するたびに、同じエラーメッセージが表示されます。

エラーメッセージはセッションログで確認できます。

第 5 章

Microsoft Azure Data Lake Storage Gen1 V3 データ型リファレンス

この章では、以下の項目について説明します。

- データ型リファレンスの概要, 23 ページ
- Avro のデータ型とトランスフォーメーションデータ型, 24 ページ
- JSON のデータ型とトランスフォーメーションデータ型, 25 ページ
- Parquet のデータ型とトランスフォーメーションデータ型, 25 ページ

データ型リファレンスの概要

Data Integration では、Microsoft Azure Data Lake Storage Gen1 V3 によるマッピングおよび mapping タスクで次のデータ型を使用します。

Microsoft Azure Data Lake Storage Gen1 V3 **ネイティブデータ型**

フィールドのメタデータを編集しようとすると、Microsoft Azure Data Lake Storage Gen1 V3 データ型がソーストランスフォーメーションとターゲットトランスフォーメーションに表示されます。

トランスフォーメーションデータ型

トランスフォーメーションで表示されるデータ型のセットです。ANSI SQL-92 汎用データ型に基づく内部データ型で、プラットフォーム間でデータを移動するとき Secure Agent によって使用されます。トランスフォーメーションデータ型は、マッピングおよび mapping タスクのすべてのトランスフォーメーションに表示されます。

Data Integration は、ソースデータを読み取るときに、ネイティブデータ型を対応するトランスフォーメーションデータ型に変換してから、データのトランスフォームを実行します。Data Integration は、ターゲットに書き込むときに、トランスフォーメーションデータ型を対応するネイティブデータ型に変換します。

次の表は、Data Integration でサポートされる Microsoft Azure Data Lake Storage Gen1 V3 データ型と、対応するトランスフォーメーションデータ型を示しています。

Microsoft Azure Data Lake Storage Gen1 ネイティブデータ型	トランスフォーメーションデータ型	説明
String	String	1〜104,857,600 文字

Avro のデータ型とトランスフォーメーションデータ型

Avro ファイルのデータ型は、Secure Agent がプラットフォーム間でデータを移動するために使用するトランスフォーメーションデータ型にマップされます。

次の表に、Secure Agent がサポートしている Avro ファイルのデータ型と、対応するトランスフォーメーションデータ型の一覧を示します。

Avro データ型	トランスフォーメーションデータ型	範囲と内容
Boolean	Integer	TRUE（1）または FALSE（0）
Byte	Binary	精度 4000
Double	Double	精度 15
Float	Double	精度 15
Int	Integer	-2,147,483,648〜2,147,483,647 精度 10、位取り 0
Long	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 精度 19、位取り 0
Null	Integer	-2,147,483,648〜2,147,483,647 精度 10、位取り 0
String	String	-1〜104,857,600 文字

JSON のデータ型とトランスフォーメーションデータ型

JSON 複合ファイルのデータ型は、Secure Agent がプラットフォーム間でデータを移動するために使用するトランスフォーメーションデータ型にマップされます。

以下の表は、Data Integration でサポートされる JSON 複合ファイルのデータ型と、対応するトランスフォーメーションデータ型を示しています。

JSON データ型	トランスフォーメーションデータ型	範囲と内容
ブール	整数	Boolean のデフォルトのトランスフォーメーション型は Integer です。True および False の値を持つ String データ型を指定することもできます。True は Integer の 1 に相当し、False は Integer の 0 に相当します。
Number（double）	double	-1.79769313486231570E+308～1.79769313486231570E+308。精度 15。
Number（float）	double	-1.79769313486231570E+308～1.79769313486231570E+308。精度 15。
Number（int）	integer	-2,147,483,648～2,147,483,647。精度 10、位取り 0。
Number（long）	bigint	-9,223,372,036,854,775,808～9,223,372,036,854,775,807。精度 19、位取り 0。
string	string	1～104,857,600 文字。

Parquet のデータ型とトランスフォーメーションデータ型

Parquet ファイルのデータ型は、Secure Agent がプラットフォーム間でデータを移動するために使用するトランスフォーメーションデータ型にマップされます。

次の表に、Secure Agent がサポートしている Parquet ファイルのデータ型と、対応するトランスフォーメーションデータ型の一覧を示します。

Parquet データ型	トランスフォーメーションデータ型	範囲と内容
Boolean	Integer	TRUE（1）または FALSE（0）
Byte_Array	Binary	任意の長バイト配列
Double	Double	精度 15

Parquet データ型	トランスフォーメーションデータ型	範囲と内容
Float	Double	精度 15
Int32	Integer	-2,147,483,648～ +2,147,483,647
Int64	Bigint	-9,223,372,036,854,775,808 ～ +9,223,372,036,854,775,807 8 バイト符号付き整数
Int96	Binary	12 バイト符号付き整数

Parquet ファイルの読み取りまたは書き込みを行うために指定する Parquet スキーマは小文字である必要があります。Parquet は大文字と小文字を区別するスキーマではありません。

索引

C

Cloud Application Integration コミュニティ
URL 5
Cloud 開発者コミュニティ
URL 5

I

Informatica Intelligent Cloud Services
Web サイト 5
Informatica グローバルカスタマサポート
連絡先情報 6

M

Microsoft Azure Data Lake Storage Gen1
概要 8
Microsoft Azure Data Lake Storage Gen1 V3
ソーストランスフォーメーション 13
ターゲットトランスフォーメーション 15
マッピングでのソース 13
マッピングのターゲット 15
接続プロパティ 10
Microsoft Azure Data Lake Storage Gen1 V3 コネクタ
概要 7, 10
例 8

W

Web サイト 5

あ

アップグレード通知 6

し

システムステータス 6

す

ステータス
Informatica Intelligent Cloud Services 6

そ

ソース
マッピングにおける Microsoft Azure Data Lake Storage Gen1 V3 13
ソーストランスフォーメーション
Microsoft Azure Data Lake Storage Gen1 V3 のプロパティ 13

た

ターゲット
Microsoft Azure Data Lake Storage Gen1 V3 マッピング 15
ターゲットトランスフォーメーション
Microsoft Azure Data Lake Storage Gen1 V3 のプロパティ 15
ターゲットの作成
ターゲットファイルのパラメータ化 19, 20
タイムスタンプの追加 19
ターゲットの指定 18

て

データ型
Avro 24
Parquet 25
データ型リファレンス
概要 23

は

パラメータファイルを使用したパラメータ化 20

ま

マッピング
Microsoft Azure Data Lake Storage Gen1 V3 ソースのプロパティ 13
Microsoft Azure Data Lake Storage Gen1 V3 ターゲットプロパティ 15

め

メンテナンスの停止 6