



Informatica® Intelligent Cloud Services
April 2024

Model Serve

© 著作権 Informatica LLC 2020, 2024

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 Cloud、Informatica Intelligent Cloud Services、PowerCenter、PowerExchange、および Informatica ロゴは、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。

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: 2024-06-18

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 マーケットプレイス	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 : Informatica Model Serve の紹介	7
ホームページ	8
アセットタイプ	8
デプロイメントプロセス	9
第 2 章 : クイックスタートモデル	10
画像分類モデル	10
テキスト翻訳モデル	11
クイックスタートモデルのデプロイ	12
第 3 章 : ユーザー定義モデル	13
機械学習モデルの登録	13
機械学習モデルの定義	13
モデルアーティファクトのアップロード	14
入力フィールドと出力フィールドの定義	14
機械学習モデルのデプロイ	15
モデルデプロイメントの設定	15
モデルデプロイメントの開始	16
モデルデプロイメントのテスト	17
第 4 章 : 監視デプロイメント	18
ログのダウンロード	19
第 5 章 : プロジェクトおよびアセットの管理	20
プロジェクト、フォルダ、およびアセットのコピー	20
フォルダとアセットの移動	21
プロジェクト、フォルダ、およびアセットの名前変更	22
プロジェクト、フォルダ、およびアセットの削除	22
タグ	23
タグの作成	23

タグの割り当て.....	24
タグの編集と削除.....	25
アセット依存関係.....	25
権限.....	26
権限のルールおよびガイドライン.....	27
権限の設定.....	27
第 6 章 : トラブルシューティング.....	29
索引.....	30

序文

「*Model Serve*」を使用して、Model Serve で機械学習モデルを登録、デプロイ、および監視する方法を学習します。

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 マーケットプレイス

Informatica マーケットプレイスにアクセスすると、データ統合コネクタ、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/>

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。

サブスクライブするには、[Informatica Intelligent Cloud Services Status](#) ページで **【サブスクライブして更新】** をクリックします。電子メール、SMS テキストメッセージ、Webhook、RSS フィード、またはこの 4 つの任意に組み合わせとして送信される通知を受信するという選択ができます。

Informatica グローバルカスタマサポート

グローバルサポートセンターには、Informatica Network または電話でお問い合わせください。

Informatica Network でオンラインサポートリソースを検索するには、Informatica Intelligent Cloud Services のヘルプメニューで **【サポートにお問い合わせください】** をクリックして、**Cloud Support** ページに移動します。**Cloud Support** ページには、システムステータス情報とコミュニティディスカッションが記載されています。追加のリソースを検索する場合や電子メールで Informatica グローバルカスタマサポートに問い合わせる場合は、Informatica Network にログインし、**【サポートが必要な場合】** をクリックしてください。

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

Informatica Model Serve の紹介

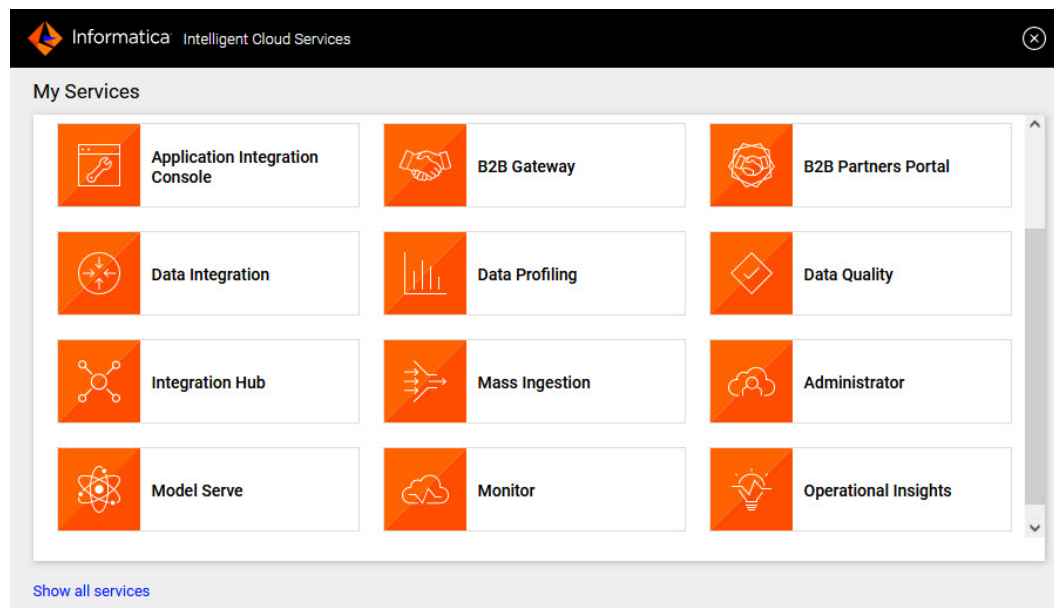
Informatica Model Serve へようこそ! Model Serve を使用すると、機械学習モデルをデプロイし、履歴データに基づいてビジネスの予測を生成できます。クイックスタートモデルをデプロイして標準モデルを開始するか、独自のファイルをアップロードしてユーザー定義モデルをデプロイすることができます。

Model Serve は、Informatica Intelligent Cloud Services で利用可能な複数のサービスの 1 つです。Informatica Intelligent Cloud Services にログインすると、**【マイサービス】** ページが画面上に開きます。**【マイサービス】** ページには、Model Serve に加えて、サブスクライブしている他のサービスとトライアル版の他のサービスも含まれる場合があります。

Model Serve ユーザーは、次のサービスを使用することができます。

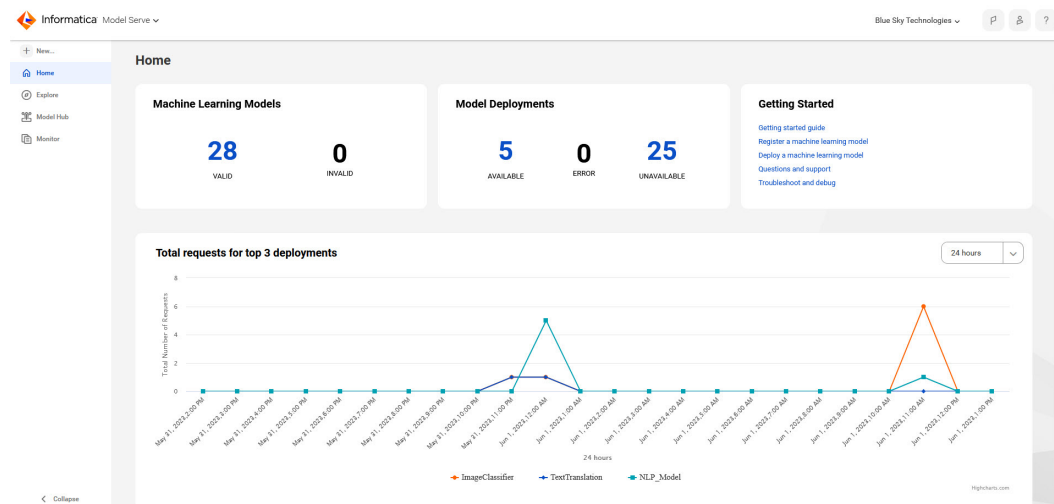
- Model Serve。機械学習モデルを登録してデプロイし、デプロイメントを監視します。
- 管理者。特権の管理などの管理タスクを実行します。

次のイメージは **【マイサービス】** ページの例を示しています。



ホームページ

【マイサービス】 ページで Model Serve を選択すると、Model Serve の【ホーム】 ページが表示されます。次の図に、【ホーム】 ページを示します。



このページには、次のパネルが表示されます。

- 機械学習モデル。組織内の機械学習モデルの数。
 - モデルデプロイメント。【モニタ】 ページに表示される組織内のモデルデプロイメントの数。
 - ビギナーズガイド。ドキュメントとサポートへのリンク。
 - 上位3つのデプロイメントの合計要求数。最もよく使用されるモデルデプロイメントの予測要求の数。
- 注: グラフでは、ユーザープロファイルで設定されたタイムゾーンが使用されます。

ナビゲーションバーから、次のページにアクセスできます。

- 新規。新しい機械学習モデルまたはモデルデプロイメントを作成します。
- ホーム。ホームページに戻ります。
- エクスプローラ。プロジェクトとアセットを作成および管理します。
- モデルハブ。クイックスタートモデルを表示およびデプロイします。
- モニタ。モデルデプロイメントのステータスを表示および管理します。

アセットタイプ

Model Serve でアセットを使用して、機械学習アルゴリズムを登録およびデプロイします。

次のようなタイプのアセットを使用できます。

クイックスタートモデル

クイックスタートモデルは、Informatica がすぐにデプロイメントを行うことができるようにするための機械学習アルゴリズムです。クイックスタートモデルを使用して、一般的なユースケース向けのアルゴリズムを迅速にデプロイします。

機械学習モデル

機械学習モデルには、入力データを分類するように機械学習手法を使用してトレーニングされたユーザー定義のアルゴリズムのアーティファクトが含まれています。独自のアルゴリズムをデプロイする場合は、機械学習モデルを使用します。

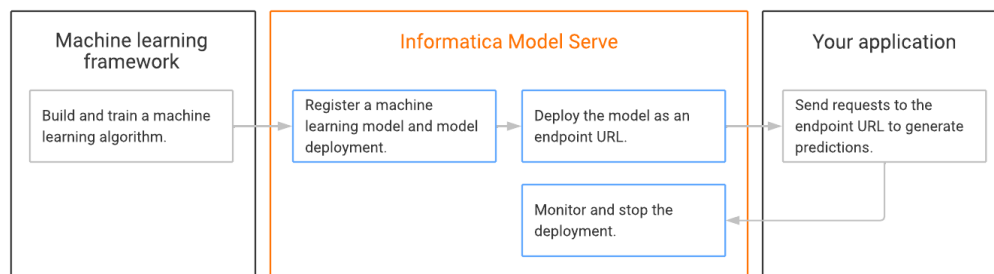
モデルデプロイメント

モデルデプロイメントは、デプロイメントで使用するコンピューティングユニットの最大数など、機械学習モデルをデプロイするためのランタイム設定を定義します。

デプロイメントプロセス

Model Serve でのモデルのデプロイには、複数のアプリケーションで実行するタスクがあります。プロセスは、使用するモデルのタイプに応じて異なります。

次の図に、モデルをデプロイするためのプロセスを示します。



次の表に、デプロイメントプロセスのタスクを示します。

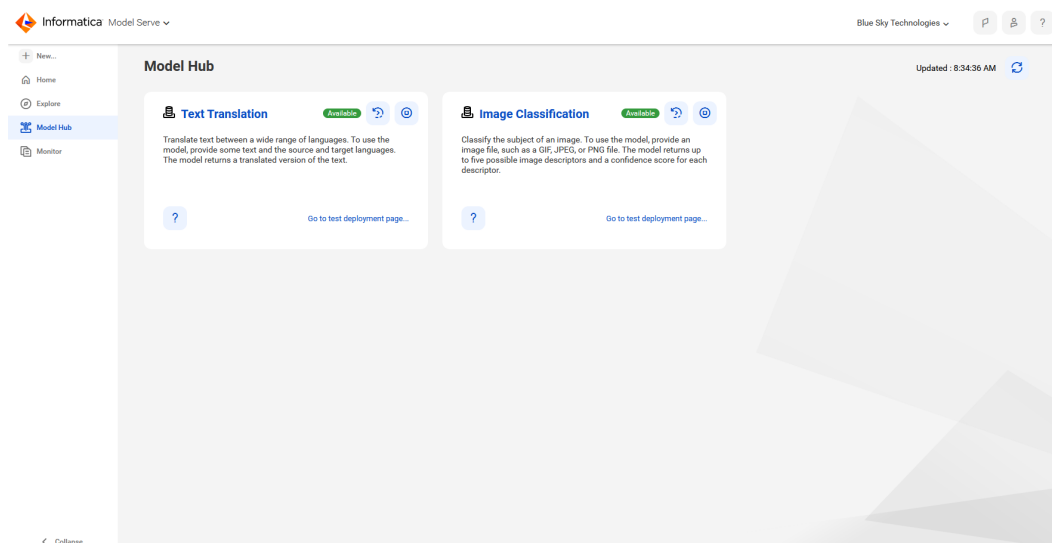
タスク	必須	説明
機械学習アルゴリズムを構築してトレーニングする。	ユーザー定義モデル	TensorFlow などのサードパーティの機械学習フレームワークを使用して、アルゴリズムを構築してトレーニングします。
機械学習モデルとモデルデプロイメントを登録する。	ユーザー定義モデル	Model Serve で、アルゴリズムファイルをアップロードする機械学習モデルを設定します。次に、モデルデプロイメントを作成して、モデルをデプロイするためのランタイムプロパティを定義します。
モデルをエンドポイント URL としてデプロイする。	クイックスタートモデルとユーザー定義モデル	クイックスタートモデルまたはモデルデプロイメントを開始して、エンドポイント URL を要求に使用できるようにします。
要求をエンドポイントに送信して予測を生成する。	クイックスタートモデルとユーザー定義モデル	アプリケーションで、API 呼び出しをエンドポイント URL に送信して、デプロイされたモデルからの予測を要求します。
デプロイメントを監視して停止する。	クイックスタートモデルとユーザー定義モデル	Model Serve を使用して、モデルデプロイメントのステータスの監視、トラブルシューティング用のログのダウンロード、およびクラウドリソースを解放するためのデプロイメントの停止を行います。

第 2 章

クイックスタートモデル

クイックスタートモデルを使用して、すでにトレーニングおよび登録されている機械学習アルゴリズムを迅速にデプロイします。クイックスタートモデルには、**【モデルハブ】** ページからアクセスできます。

次の図に、**【モデルハブ】** ページを示します。



クイックスタートモデルが使用可能な場合は、**【モデルハブ】** ページまたは **【監視】** ページでモデルの名前をクリックすると、モデルに関する情報の確認やモデルのテストができます。クイックスタートモデルのステータスは、**【モデルハブ】** ページおよび **【監視】** ページで確認できます。

画像分類モデル

画像分類モデルにより、画像の被写体を分類するための画像記述子を生成します。

モデルにより、[ResNet-50 v1.5 model from Hugging Face](#) をデプロイします。入力として、最大 5 MB の JPEG、JPG、または PNG 画像ファイルが必要です。モデルは、5 つの分類および各分類の信頼度スコアを返します。

画像分類モデルの結果は、ResNet-50 v1.5 モデルの Hugging Face で表示される結果とは異なる場合があります。Hugging Face では Web サイト上のモデルが変更される場合がありますが、画像分類モデルは静的です。

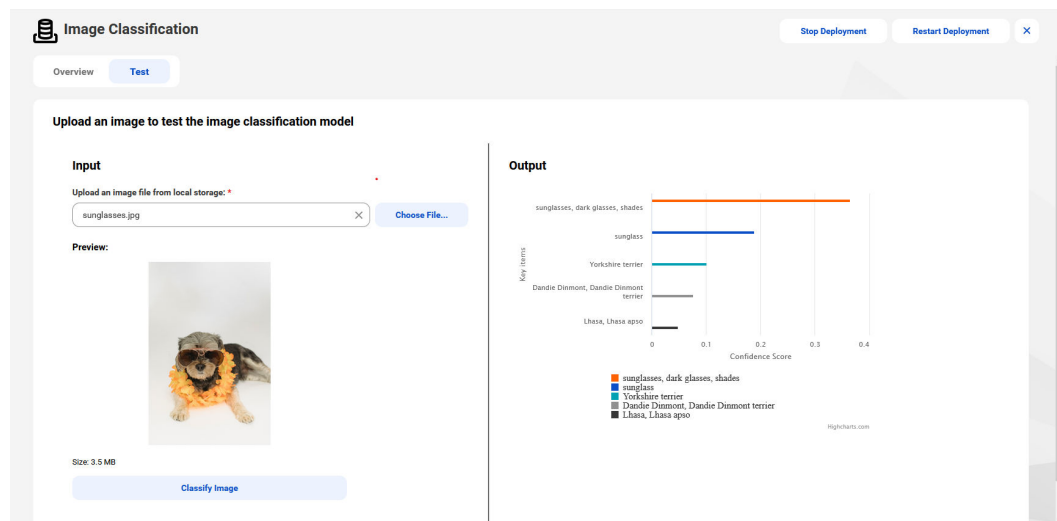
注: このモデルはサードパーティから供給されています。Informatica は、その動作、機能、または関連する責務について責任を負いません。このアーティファクトに関連する問題、更新、または懸念事項は、元のプロバ

イダまたは作成者に問い合わせる必要があります。当社はこのアーティファクトを現状のまま、保証なしで提供します。

例

オンラインストアを運営しており、顧客からの商品レビュー内に写真が添付されることがあるとします。写真を Web サイトに掲載する前に、その写真がレビュー対象の商品に関連しているかどうかを確認する必要があります。最初の確認事項として、画像分類モデルを使用して写真を分類します。写真の信頼度スコアが 0.5 以上で正しい商品タイプが返された場合は、レビューを公開することができます。信頼度スコアが 0.5 未満の値である場合は、コンテンツモデレーターが確認を行うように写真にフラグを立てます。

次の図に、画像分類モデルのテストの例を示します。



テキスト翻訳モデル

テキスト翻訳モデルにより、幅広い言語間でテキストの翻訳を行います。

モデルにより、[M2M100 418M model from Hugging Face](#) をデプロイします。モデルにはソース言語コード、ターゲット言語コード、およびソーステキストが必要です。また、モデルによって、翻訳されたテキストが返されます。最大 3,000 文字までのソーステキストを指定できます。モデルでサポートされる言語のリストと言語コードを確認するには、Hugging Face のマニュアルを参照してください。

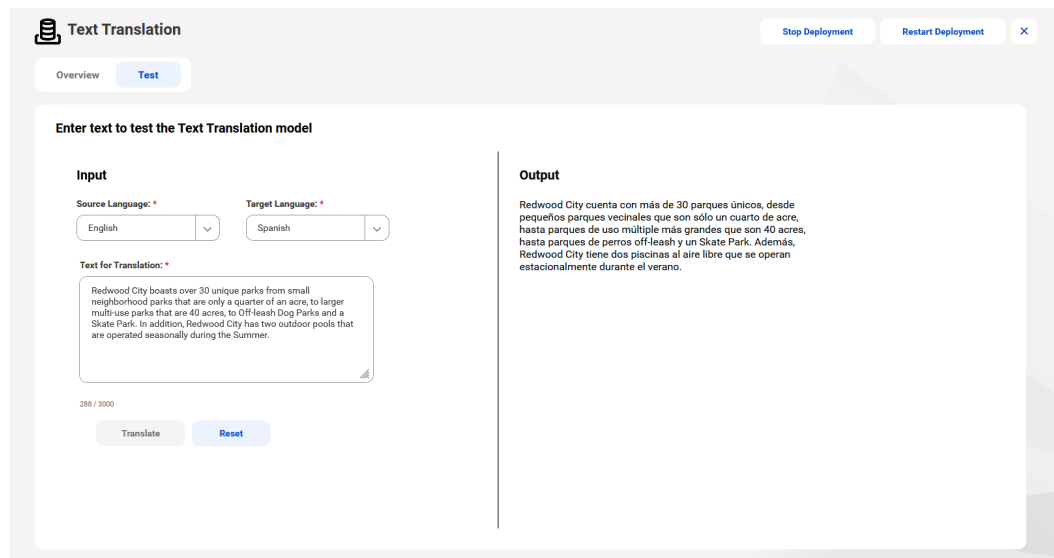
テキスト翻訳モデルの結果は、M2M100 418M モデルの Hugging Face で表示される結果とは異なる場合があります。Hugging Face では Web サイト上のモデルが変更される場合がありますが、テキスト翻訳モデルは静的です。

注: このモデルはサードパーティから供給されています。Informatica は、その動作、機能、または関連する責務について責任を負いません。このアーティファクトに関連する問題、更新、または懸念事項は、元のプロバイダまたは作成者に問い合わせる必要があります。当社はこのアーティファクトを現状のまま、保証なしで提供します。

例

地方自治体の Web サイトのデザインを担当しているとします。さまざまな言語を話すユーザーが情報にアクセスできるようにするには、テキスト翻訳モデルを使用して、ユーザーのブラウザに設定されている言語で Web サイトをレンダリングします。

次の図に、テキスト翻訳モデルのテストの例を示します。



Text Translation [Stop Deployment] [Restart Deployment] X

Overview **Test**

Enter text to test the Text Translation model

Input

Source Language: * English ▼ Target Language: * Spanish ▼

Text for Translation: *

Redwood City boasts over 30 unique parks from small neighborhood parks that are only a quarter of an acre, to larger multi-use parks that are 40 acres, to Off-leash Dog Parks and a Skate Park. In addition, Redwood City has two outdoor pools that are operated seasonally during the Summer.

288 / 3000

Translate Reset

Output

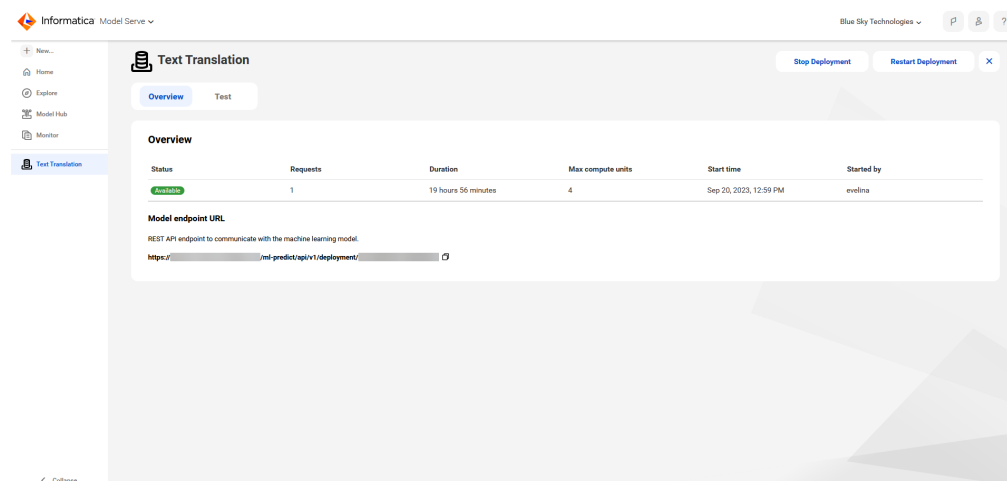
Redwood City cuenta con más de 30 parques únicos, desde pequeños parques vecinales que son sólo un cuarto de acre, hasta parques de uso múltiple más grandes que son 40 acres, hasta parques de perros off-leash y un Skate Park. Además, Redwood City tiene dos piscinas al aire libre que se operan estacionalmente durante el verano.

クイックスタートモデルのデプロイ

クイックスタートモデルを使用するには、【モデルハブ】ページからモデルをデプロイし、API 経由で要求を行います。

1. 【モデルハブ】ページで、目的のクイックスタートモデルを見つけて、【デプロイメントを開始】をクリックします。
2. ステータスが *使用可能* になった後に、【モデルハブ】ページまたは【監視】ページでクイックスタートモデルの名前をクリックします。概要ページが開き、モデルのエンドポイント URL など、デプロイされたクイックスタートモデルに関する情報を確認できます。

次の図に、概要情報を示します。



Informatica Model Serve Blue Sky Technologies P A ?

Text Translation [Stop Deployment] [Restart Deployment] X

Overview **Test**

Overview

Status	Requests	Duration	Max compute units	Start time	Started by
Ready	1	19 hours 56 minutes	4	Sep 20, 2023, 12:59 PM	evelina

Model endpoint URL

REST API endpoint to communicate with the machine learning model.

https://ml-predict/api/v1/deployment/

3. エンドポイント URL を使用して、API 経由で予測要求を作成します。API 要求の作成の詳細については、「[Model Serve REST API](#)」を参照してください。
4. リソースを解放するには、不要な場合にデプロイメントを停止します。

第 3 章

ユーザー定義モデル

ユーザー定義モデルは、Model Serve の外部で構築およびトレーニングを行い、Model Serve にアップロードしてデプロイする機械学習モデルです。ユーザー定義モデルをデプロイするには、機械学習モデルを登録し、モデルのデプロイを作成して開始します。

機械学習モデルの登録

Model Serve の使用を開始するには、トレーニング済みの機械学習アルゴリズムに基づいた機械学習モデルを登録します。

機械学習モデルを登録すると、Model Serve はモデルファイルを内部のファイルストアにアップロードします。モデルを登録すると、1 つまたは複数のモデルデプロイメントでそのモデルを使用できます。

機械学習モデルを設定するには、次の手順を実行します。

1. モデルの全般プロパティを定義します。
2. モデルのアーティファクトを ZIP ファイルとしてアップロードします。
3. モデルが想定する入力フィールドと、モデルが返す出力フィールドを定義します。

機械学習モデルを保存すると、概要ページが表示されます。組織内のユーザーは、概要ページからモデルアーティファクトをダウンロードできます。

機械学習モデルを編集する必要がある場合は、そのモデルを使用するすべてのモデルデプロイメントを停止します。実行中のモデルデプロイメントで使用されている場合、機械学習モデルを編集することはできません。

機械学習モデルの定義

機械学習モデルを作成し、全般プロパティを設定します。

1. ナビゲーションバーで、**[新規]** > **[機械学習モデル]** を選択します。

2. モデルの全般プロパティを設定します。次の表に、設定するプロパティを示します。

プロパティ	説明
モデル名	必須。機械学習モデルの名前。 名前は英字で始まる必要があり、英数字、ハイフン (-)、およびアンダースコア (_) を含めることができます。最大長は 100 文字です。
説明	機械学習モデルの説明。
場所	必須。機械学習モデルが存在するプロジェクトフォルダ。 [参照] ページが現在アクティブになっていてプロジェクトまたはフォルダが選択されている場合、機械学習モデルのデフォルトの場所は、選択されているプロジェクトまたはフォルダです。それ以外の場合、デフォルトの場所は <i>Default</i> フォルダになります。

モデルアーティファクトのアップロード

コードテンプレートをダウンロードして編集し、モデルアーティファクトをアップロードします。

1. フレームワークとして [カスタム] を選択します。
2. コードテンプレートをダウンロードして、解凍します。
コードテンプレートにはファイル `predict.py` と `requirements.txt` が含まれています。
3. `predict.py` を編集して、カスタムアルゴリズムファイルへのエントリポイントを定義します。モデルデプロイメントの開始時にモデルファイルをロードするように、`__init__()` メソッドを定義します。JSON 形式の要求を取得してモデルから予測を返すように、`predict()` メソッドを定義します。

注: `__init__()` メソッドの定義は必須ではありません。このメソッドが定義されていない場合、モデルデプロイメントは要求のたびにモデルファイルをロードするため、パフォーマンスが低下します。

4. `requirements.txt` ファイルを編集して、機械学習アルゴリズムの実行に必要なサードパーティライブラリをリストします。

例えば、このファイルには次のリストが含まれている場合があります。

```
scikit-learn
scipy
tensorflow
numpy==1.19.5
```

注: ライブラリのバージョンが指定されていない場合、Model Serve は最新バージョンを使用します。使用するライブラリのバージョンが KServe 0.8.0 と互換性があることを確認してください。KServe 0.8.0 と互換性のあるライブラリバージョンに、モデルに必要な別のライブラリとの互換性がない場合は、Informatica グローバルカスタマーサポートに連絡して、KServe のバージョンをアップグレードしてください。

5. `predict.py`、`requirements.txt`、およびアルゴリズムファイルを ZIP ファイルに圧縮します。ファイルの最大サイズは 3 GB です。`predict.py` ファイルと `requirements.txt` ファイルは、ZIP ファイルのルートレベルにある必要があります。
6. ZIP ファイルをモデルファイルとしてアップロードします。

入力フィールドと出力フィールドの定義

モデルが API 要求で想定する入力フィールドと、モデルが API 応答で返す出力フィールドを定義します。

[フィールド] タブで、各入力フィールドと出力フィールドの名前とデータ型を追加します。少なくとも 1 つの入力フィールドと出力フィールドを設定します。

次の図は、[フィールド] タブを示しています。

New Machine Learning Model

< Back Next > Save Cancel X

① General ② Model Artifacts ③ Fields

Specify the Input and Output fields for your model

File Name : MNIST.zip

Input Fields

Name: * Data Type: * Shape: * ⓘ

pixels Array of Float [28,28] ⓘ

Add Field

Output Fields

Name: * Data Type: * Shape: * ⓘ

classify Array of Float [1,1] ⓘ

Add Field

注: ユーザー定義オブジェクトを出力する場合は、predict()メソッドがシリアル化された JSON 文字列を返すことを確認してから、文字列出力フィールドを設定してください。

機械学習モデルのデプロイ

機械学習モデルをデプロイして、API 要求のエンドポイント URL を公開します。

機械学習モデルをデプロイするには、まず、デプロイの実行方法を定義するモデルデプロイメントを設定します。モデルデプロイメントを開始すると、Model Serve は必要なクラウドリソースをプロビジョニングし、機械学習モデルをデプロイします。モデルデプロイメントが使用可能になると、エンドポイント URL に要求を送信できるようになります。

機械学習モデルをデプロイするには、次のタスクを実行します。

1. モデルデプロイメントでランタイムプロパティを設定します。
2. モデルデプロイメントを開始します。
3. 必要に応じて、モデルが想定どおりの予測を生成しているかテストします。

モデルデプロイメントの設定

モデルデプロイメントを設定して、機械学習モデルのデプロイに使用するランタイムプロパティを定義します。

1. **【新規】 > 【モデルデプロイメント】** を選択します。

2. [全般] タブで、モデルデプロイメントの全般プロパティを設定します。次の表に、設定するプロパティを示します。

プロパティ	説明
モデルデプロイメント名	必須。モデルデプロイメントの名前。 名前は英字で始まる必要があり、英数字、ハイフン (-)、およびアンダースコア (_) を含めることができます。最大長は 100 文字です。
説明	モデルデプロイメントの説明。
場所	必須。モデルデプロイメントが存在するプロジェクトフォルダ。 [参照] ページが現在アクティブになっていてプロジェクトまたはフォルダが選択されている場合、モデルデプロイメントのデフォルトの場所は、選択されているプロジェクトまたはフォルダです。それ以外の場合、 <i>Default</i> フォルダがデフォルトの場所です。
機械学習モデル	必須。デプロイする機械学習モデル。

3. [コンピューティング] タブで、[コンピューティングユニット] プロパティを、モデルデプロイメントで使用するコンピューティングユニットの最大数に設定します。

1つのコンピューティングユニットには、4つのCPUと16GBのメモリが含まれています。デフォルトは4個のコンピューティングユニットです。

4. モデルデプロイメントを保存します。

モデルデプロイメントを設定して保存したら、デプロイを開始するか、エディタを閉じてモデルデプロイメントの概要ページを表示できます。

注: 実行中のモデルデプロイメントを編集することはできません。編集するには、モデルデプロイメントを停止し、使用できなくなるまで待ちます。

モデルデプロイメントの開始

モデルデプロイメントを開始すると、Model Serve は機械学習モデルをデプロイし、URL エンドポイントを要求でできるようにします。

以下のいずれかの方法で、モデルデプロイメントを開始できます。

- モデルデプロイメントエディタで、有効なデプロイメントを保存した後に、[デプロイメントを開始] を選択します。
- モデルデプロイメントの概要ページで、[デプロイメントを開始] を選択します。
- [モニタ] ページで、モデルデプロイメントの [アクション] メニューを開き、[開始] を選択します。

モデルデプロイメントを開始するときに、他に利用可能なデプロイメントがない場合、Model Serve は必要なクラウドリソースを起動します。これには最大 10 分かかる場合があります。

モデルデプロイメントから予測を要求する場合、要求は最大 5 分間実行できます。モデルが 5 分後に応答を返さない場合、要求はタイムアウトになります。

モデルデプロイメントのテスト

要求を送信して、モデルが想定どおりの予測を生成しているか確認します。利用可能なモデルデプロイメントのテスト予測を生成できます。

1. モデルデプロイメントの概要で、**【テスト】** タブに移動します。
2. 要求呼び出しを JSON 形式で入力します。呼び出しに必要な入力フィールドがすべて含まれていることを確認してください。
3. **【要求の送信】** を選択します。

Model Serve は要求を送信し、モデルからの応答を表示します。Model Serve は、可能であれば JSON 形式で応答を返します。それ以外の場合は、応答を文字列として返します。

第 4 章

監視デプロイメント

【監視】 ページを使用して、デプロイメントのステータスを確認および管理し、ログをダウンロードします。

【監視】 ページには、クイックスタートモデルおよびユーザー定義デプロイメントが表示されます。

次の表に、【監視】 ページに表示される各デプロイメントのプロパティを示します。

プロパティ	説明
名前	デプロイメントの名前。
ステータス	デプロイメントのステータス。デプロイメントは、次のいずれかの状態になります。 <ul style="list-style-type: none">- 開始中。Model Serve が機械学習モデルのデプロイを準備中です。- リソースのプロビジョニング中。Model Serve がクラウドリソースをプロビジョニング中です。- デプロイ中。Model Serve がモデルをデプロイ中です。- 使用可能。デプロイメントが正常に開始され、エンドポイント URL を要求で使えるようになりました。- 停止中。デプロイメントは停止中です。- 使用不可能。デプロイメントが停止し、すべてのリソースが解放されました。- 再起動中。Model Serve がデプロイメントを停止して再開しています。- 再デプロイ中。Model Serve が再起動後にモデルを再デプロイ中です。- エラー。エラーが発生したため、デプロイメントを開始できませんでした。エラーが発生する前に Model Serve がクラウドリソースをプロビジョニングしていた場合は、ステータスが「エラー」であってもリソースはまだ実行されている可能性があります。- 終了。Model Serve
場所	モデルデプロイメントが存在するプロジェクトフォルダ。 クイックスタートモデルには適用されません。
開始したユーザー	デプロイメントを開始したユーザーの名前。
開始時刻	デプロイメントが前回「デプロイ中」ステータスになった日時。 注: デプロイメントを開始し、「開始中」または「リソースのプロビジョニング中」ステータスのときにエラーが発生した場合、開始時刻は表示されません。
終了時刻	デプロイメントが停止した日時。
継続時間	デプロイメントが実行されている時間。
要求	デプロイメントが受信した予測要求の合計数。

このページの情報が最新であることを確認するために、Model Serve は Informatica Intelligent Cloud Services リポジトリを定期的にポーリングし、ユーザーがアクティブである場合は【監視】 ページで監視の詳細を更新します。

【監視】 ページに表示されていない新しいデプロイメントをユーザーが開始した場合、このデプロイメントはページに表示されません。ページ上部のステータスメッセージは、情報が最新ではないことを示し、ユーザーはページを更新して新しいデプロイメントを表示するように求められます。

ログのダウンロード

モデルデプロイメントのトラブルシューティングに役立つログファイルをダウンロードできます。ステータスが **【開始中】**、**【使用可能】**、**【停止】**、または **【失敗】** のデプロイメントのログファイルをダウンロードできます。

1. **【アクション】** メニューを選択し、**【ログのダウンロード】** を選択します。
2. **【ログのダウンロード】** ダイアログボックスで、ダウンロードするログの時間範囲を設定します。過去 30 日間に作成されたログをダウンロードできます。

第 5 章

プロジェクトおよびアセットの管理

プロジェクトとアセットは、**[参照]** ページで管理できます。

Informatica Intelligent Cloud Services のプロジェクトやアセットは、次の方法で管理できます。

- プロジェクト、フォルダ、またはアセットをコピーして、新しいバージョンを作成します。
- フォルダまたはアセットを他の場所に移動します。
- プロジェクト、フォルダ、またはアセットの名前を変更します。
- プロジェクト、フォルダ、またはアセットを削除します。
- タグを適用して、**[エクスプローラ]** ページで関連アセットをフィルタ処理できるようにします。
- プロジェクト、フォルダ、またはアセットの権限を構成します。
- 組織間でアセットを移行します。

プロジェクト、フォルダ、およびアセットのコピー

[エクスプローラ] ページでは、プロジェクト、フォルダ、およびアセットをコピーできます。テンプレートとして使用するプロジェクトをコピーすることも、バックアップコピーを作成することもできます。

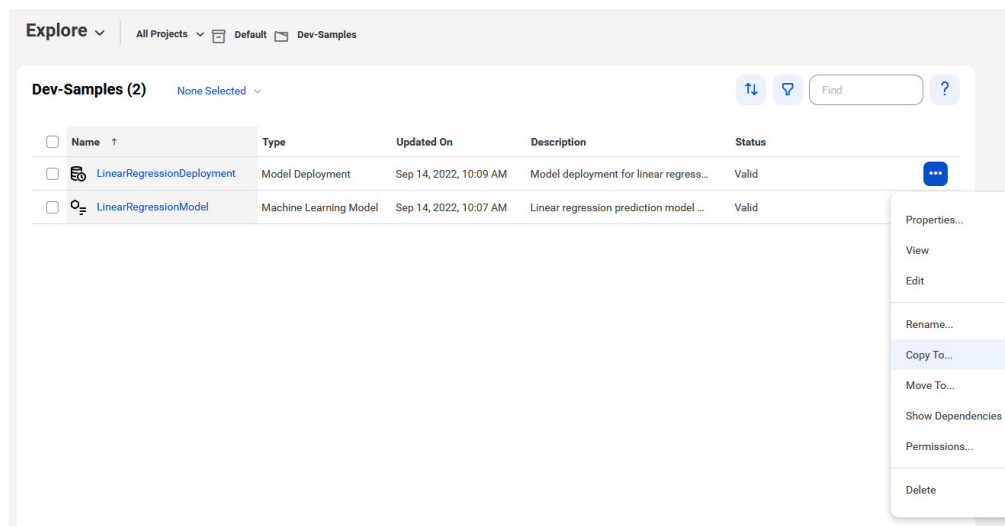
オブジェクトをコピーする場合、Informatica Intelligent Cloud Services は次のインスタンスに子要素を保持します。

- プロジェクトをコピーすると、新しいプロジェクトには、元のプロジェクトにあったすべてのフォルダとアセットが含まれます。同様に、フォルダをコピーすると、新しいフォルダには、元のフォルダにあったすべてのアセットが含まれます。
- アセットまたはアセットを含むフォルダを別の場所にコピーすると、アセットは依存オブジェクトへの参照を保持します。
例えば、Development プロジェクトにはアセット `primary_asset` と `reference_asset` が含まれており、`primary_asset` には `reference_asset` が必要だとします。`primary_asset` を Production プロジェクトにコピーすると、Production プロジェクトの新しい `primary_asset` は、Development プロジェクトの `reference_asset` を参照します。

フォルダ内でアセットをコピーすると、両方のアセットを保持できます。同じ名前のアセットを含む別のフォルダにアセットをコピーする場合は、フォルダ内のアセットを上書きするか、または両方を保持できます。アセットを上書きすることを選択した場合は、アセットタグをマージすることができます。両方のアセットを保持する場合、Informatica Intelligent Cloud Services は新しいアセット名に「コピー x」を付加します。x は、コピーの連番です。

注: 重複するアセットとの名前の競合を避けるには、「コピー x」接尾辞の付いたアセットの名前を変更します。
オブジェクトをコピーするには、次の手順を実行します。

1. **【参照】** ページで、コピーするオブジェクトに移動します。
2. コピーするオブジェクトが含まれる行で、**【アクション】** をクリックし、**【コピー先】** を選択します。

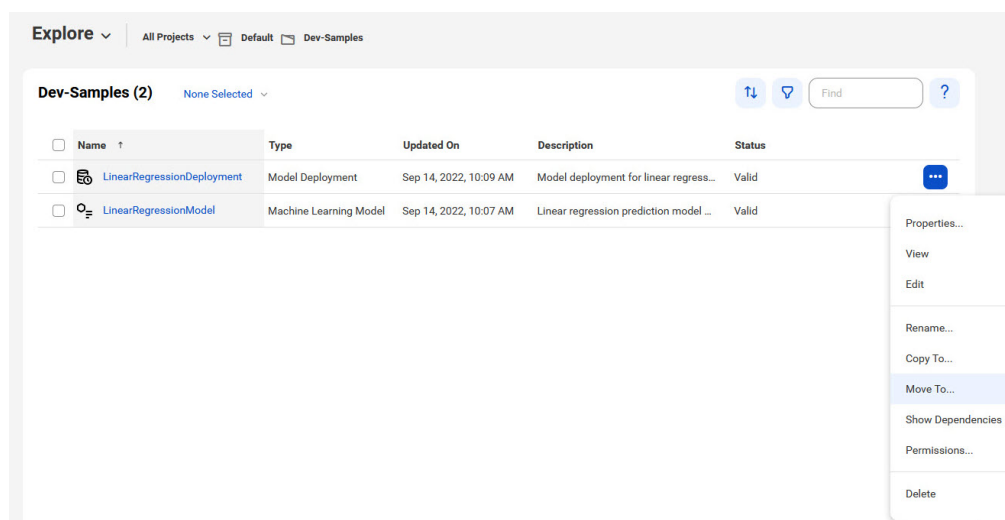


3. 新しい場所を参照し、**【選択】** をクリックします。

フォルダとアセットの移動

フォルダとアセットは、**【エクスプローラ】** ページで移動できます。

1. **【エクスプローラ】** ページで、移動するフォルダまたはアセットに移動します。
2. フォルダまたは単一のアセットを移動するには、フォルダまたはアセットを含む行で **【アクション】** をクリックして **【移動先】** を選択し、新しい場所を参照してから **【選択】** をクリックします。



プロジェクト、フォルダ、およびアセットの名前変更

使用するオブジェクトへの参照を保持したまま、プロジェクト、フォルダ、およびアセットの名前を変更できます。

例えば、primary_asset という名前のアセットが reference_asset という名前のアセットを使用している場合、primary_asset の名前を main_asset に変更しても、reference_asset は引き続き参照されます。

プロジェクト、フォルダ、アセット、またはタグの名前に次の文字を使用することはできません。

? ' | { } " ^ & [] / \

1. **【エクスプローラ】** ページで、名前を変更するオブジェクトに移動します。
2. アセットの名前を変更するには、**【アクション】** をクリックし、**【名前の変更】** を選択します。
3. プロジェクトまたはフォルダの名前を変更するには、プロジェクトまたはフォルダが含まれる行で、**【アクション】** をクリックし、**【プロパティ】** を選択します。
4. 新しい名前を入力し、**【保存】** をクリックします。

名前に特殊文字を使用したり、同じフォルダ内の別のオブジェクトと同じ名前を使用したりすることはできません。

プロジェクト、フォルダ、およびアセットの削除

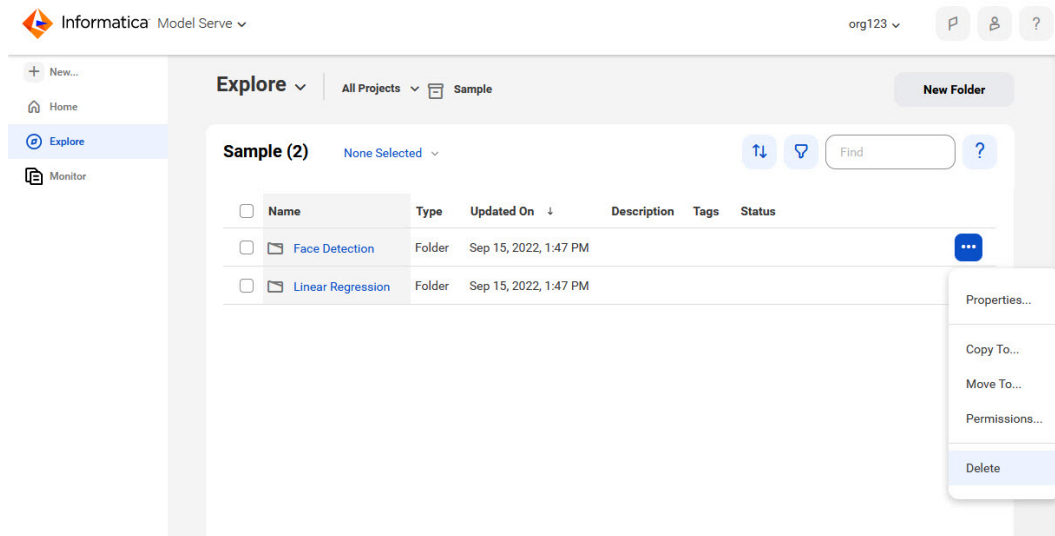
必要なくなった場合には、プロジェクト、フォルダ、またはアセットを削除できます。ただし、削除する前に、組織内のユーザーがその使用を計画していないことを確認してください。プロジェクト、フォルダ、またはアセットは、削除後に取得することはできません。

次の状況では、アセットを削除できません。

- アセットが現在使用可能なモデルデプロイメントである。
- アセットが別のアセットによって使用されている。先にアセットの依存関係を削除してから、アセットを削除する必要があります。

アセットの依存関係の表示については、「[「アセット依存関係」 \(ページ 25\)](#)」を参照してください。

次の図に示すように、**【エクスプローラ】** ページからプロジェクト、フォルダ、またはアセットを削除します。



1. **【エクスプローラ】** ページで、削除するオブジェクトに移動します。
2. プロジェクト、フォルダ、またはアセットを含む行で、**【アクション】** をクリックし、**【削除】** を選択します。

タグ

タグは、アセットをグループ化するために使用可能なアセットプロパティです。**【エクスプローラ】** ページで共通の属性を共有するアセットをフィルタリングするためにタグを作成します。

例えば、組織の各アセットに、そのアセットを管理する支社を識別するタグが含まれます。南西地域の支社が管理するすべてのアセットを表示するには、**【エクスプローラ】** ページで、タグ別に参照を行い、**【南西地域】** タグをクリックします。

タグはすべてのアセットタイプに割り当てることができます。1つのアセットには最大 64 個のタグを設定できます。

次のいずれかの方法を使用して、特定のタグを含むすべてのアセットを検索できます。

- 任意の行で、**【タグ】** カラムのタグ名をクリックします。
- タグを基準に検索し、ページに表示されるタグの一覧で、タグの名前をクリックします。

タグの作成

アセットの**【プロパティ】** ダイアログボックスを使用して、そのアセットのタグを作成して割り当てるか、将来使用できるようにタグを作成します。

アセットに割り当てずに複数のタグを作成するには、次の手順を実行します。

1. **【エクスプローラ】** ページで、アセットタイプを参照します。
2. アセットを含む行で、**【アクション】** をクリックし、**【プロパティ】** を選択します。
3. **【タグ】** フィールドに、作成するタグの名前を入力し、Enter キーを押します。

タグの最大文字数は 255 文字です。

プロジェクト、フォルダ、アセット、またはタグの名前に次の文字を使用することはできません。

? ' | { } " ^ & [] / \

4. 引き続き、他のタグを入力します。タグリストに追加するタグ名ごとに Enter キーを押します。

Properties: LinearRegressionDeployment



Name: LinearRegressionDeployment

Description: Model deployment for linear regression with one variable

Tags:

Classification ×

Linear_Regression ×



Save

Cancel

5. タグを入力したら、**【タグ】** フィールドからタグを削除して、アセットがタグに関連付けられないようにします。タグはまだ、使用可能なタグの一覧に表示されたままとなります。
6. **【保存】** をクリックします。

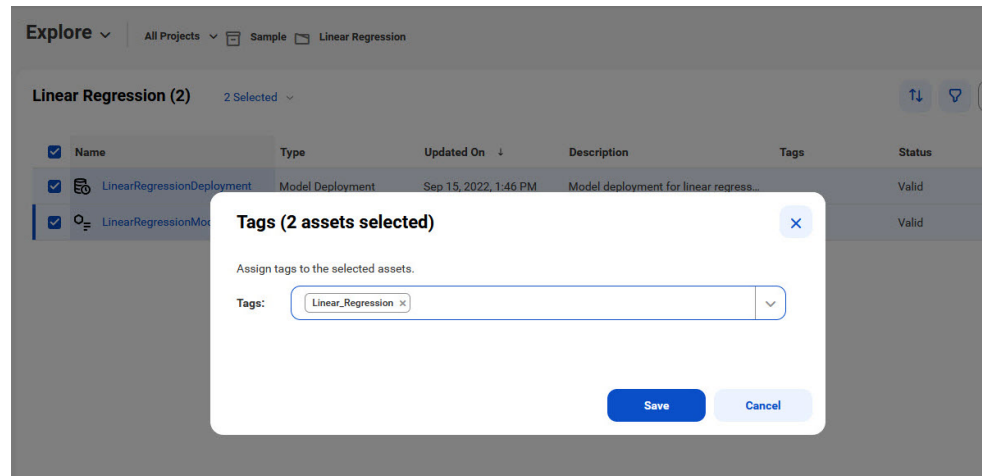
タグの割り当て

アセットにタグを一对で割り当てたり、複数のアセットに1つのタグを同時に割り当てることができます。複数のタグを1つのアセットに割り当てすることもできます。

アセットにタグを割り当てる場合は、既存のタグを選択するか、新規作成できます。

1. **【エクスプローラ】** ページで、1つ以上のアセットに移動します。
2. 1つのアセットにタグを割り当てるか、複数のアセットにタグを同時に割り当てるかに応じて、次のいずれかのタスクを実行します。
 - 1つのアセットにタグを割り当てるには、アセットを含む行で、**【アクション】** をクリックし、**【プロパティ】** を選択します。

- 複数のアセットにタグを同時に割り当てるには、各アセットの行でチェックボックスをオンにします。すべてのアセットを選択した後、選択メニューから **【タグ】** を選択します。



3. 既存のタグを選択するか、新しいタグの名前を入力します。
必要なタグがすべて割り当てられるまで、タグの追加や新しいタグの作成を続けます。
4. **【保存】** をクリックします。

タグの編集と削除

タグは **【エクスプローラ】** ページで編集または削除できます。

タグのプロパティでタグ名または説明を編集します。タグを編集すると、関連するアセットのプロパティも更新されます。例えば、m_sales アセットに NorthWest タグがあり、このタグの名前を NW に変更すると、m_sales アセットプロパティのタグの名前は NW に変更されます。

タグを削除すると、そのタグはアセットプロパティに表示されなくなります。

1. **【エクスプローラ】** ページで、タグを基準にして参照します。
2. タグが含まれている行で、次のいずれかのタスクを実行します。
 - タグを編集するには、**【アクション】** をクリックし、**【編集】** を選択します。変更を行ったら **【保存】** をクリックします。
 - タグを削除するには、**【アクション】** をクリックし、**【編集】** を選択します。

アセット依存関係

アセットのオブジェクトの依存関係を表示できます。アセットに対して特定の操作を実行する前に、オブジェクトの依存関係を確認することをお勧めします。

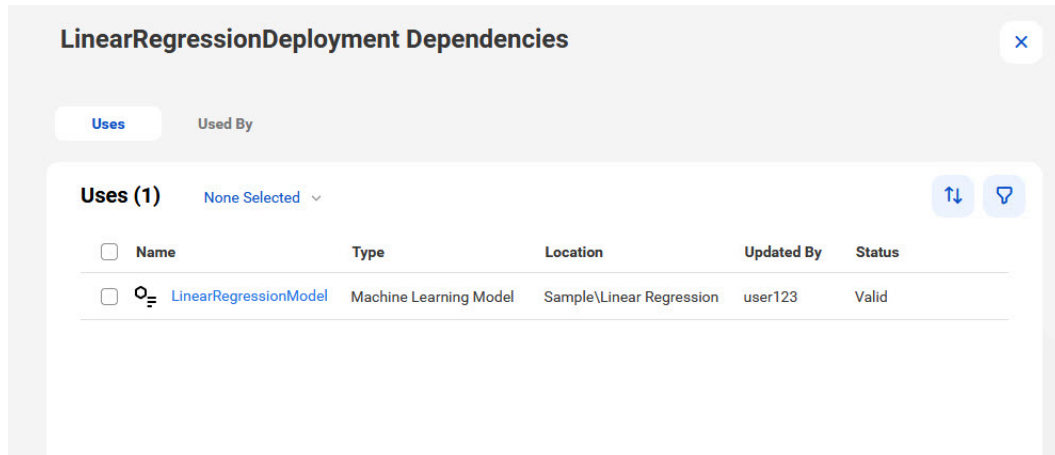
例えば、別のオブジェクトがアセットに依存している場合、アセットを削除できません。先に依存オブジェクトを削除してから、アセットを削除する必要があります。アセットの依存関係を確認することで、依存オブジェクトを見つけることができます。

【エクスプローラ】 ページでアセットのオブジェクト依存関係を表示できます。アセットのオブジェクト依存関係を表示するには、目的のアセットを含む行で、**【アクション】** をクリックして **【依存関係の表示】** を選択します。**【アセット依存関係】** ページが開き、デフォルトで **【次を使用】** タブが表示されます。

【次を使用】 タブには、選択したアセットが使用するオブジェクトのリストが表示されます。

【次により使用】 タブには、選択したアセットを使用するオブジェクトのリストが表示されます。

次の図は、モデルデプロイメントの【アセット依存関係】 ページを示しています。【次を使用】 タブが表示されています。



Uses		Used By			
Uses (1)		None Selected			
<input type="checkbox"/>	Name	Type	Location	Updated By	Status
<input type="checkbox"/>	LinearRegressionModel	Machine Learning Model	Sample\Linear Regression	user123	Valid

最も低いレベルの依存関係にドリルダウンするには、【依存関係】 ページに表示される各アセットの依存関係をさらに表示します。【依存関係】 ページの上部にあるパンくずリストには、依存関係のチェーンが表示されます。

適切な権限がある場合は、【依存関係】 ページで、アセットの表示や削除などのアクションを実行できます。アセットを表示または削除するには、目的のアセットを含む行で、【アクション】 をクリックしてアクションを選択します。

権限

権限によって、Secure Agent、Secure Agent グループ、接続、スケジュール、またはアセットに対するユーザーのアクセス権が決まります。また、オブジェクトに対する追加またはカスタムのセキュリティを追加します。権限によって、オブジェクトに対する権限の読み取り、更新、削除、実行、および変更が可能なユーザーおよびグループが定義されます。

オブジェクトの権限を構成するには、次のライセンスと特権が必要です。

- プロジェクト内のすべてのアセットについてプロジェクトレベルで権限を構成するには、プロジェクトレベルでセキュリティ権限の設定または設定解除を行うためのライセンスが組織に必要です。
- プロジェクト内のすべてのアセットについてフォルダレベルで権限を構成するには、フォルダレベルでセキュリティ権限の設定または設定解除を行うためのライセンスが組織に必要です。
- 個々のアセットの権限を構成するには、セキュリティを詳細に設定するためのライセンスが組織に必要です。
- ユーザーアカウント、または管理者がメンバとなっているグループに割り当てられたロールには、オブジェクトタイプに対する権限の設定特権が必要です。例えば、Secure Agent の権限を構成するには、Secure Agent に対する権限の設定特権を持つロールが割り当てられる必要があります。

オブジェクトの権限を構成するには、オブジェクトに移動して適切な権限を設定します。例えば、開発チームのユーザーグループのユーザーだけが開発データフォルダのアセットにアクセスできるようにします。フォルダに移動し、権限を編集し、フォルダに開発チームのユーザーグループの権限を付与します。

権限は、オブジェクトのコピーではなく、設定したオブジェクトに適用されます。したがって、アセットをコピーまたはエクスポートする場合、その権限はアセットと一緒にコピーまたはエクスポートされません。例えば、ユーザー `rjones` のみが実行権限を持っているモデルデプロイメントをコピーするとします。デプロイメントのコピーには権限が割り当てられていないため、モデルデプロイメントを実行する特権を持つユーザーは、デプロイメントのコピーを開始できます。

オブジェクトに対して次の権限を構成できます。

権限	説明
読み取り	オブジェクトを開いて表示します。
更新	オブジェクトを編集します。 読み取り権限が必要です（自動的に付与される）。
削除	オブジェクトを削除します。
実行	オブジェクトを実行します。 モデルデプロイメントに適用されます。
権限の変更	オブジェクトに割り当てられている権限を変更します。

注: これらの権限は、Informatica Intelligent Cloud Services 内で制御されます。Windows や Linux で Secure Agent を起動、停止、または設定する場合の権限のような、オペレーティングシステムの権限を制御するものではありません。

権限のルールおよびガイドライン

権限には、次の規則とガイドラインを使用します。

- オブジェクトの権限を構成するときに、権限を付与するユーザーまたはグループに、そのオブジェクトタイプに対する適切な特権を持つロールが割り当てられていることを確認します。例えば、ユーザーに特定のフォルダに対する監視ロールの更新特権があっても、監視ロールにはフォルダの更新特権がないため、ユーザーはフォルダを更新できません。
- アセットを編集するには、アセット内で使用されているすべてのアセットに対する読み取り権限がユーザーに与えられている必要があります。例えば、モデルデプロイメントに対する読み取り権限および更新権限をユーザーに割り当てた場合、そのユーザーにデプロイメントで使用されている機械学習モデルに対する読み取り権限もあることを確認します。

権限の設定

オブジェクトタイプに対する権限の設定特権を持つロールが割り当てられている場合は、オブジェクトの権限を構成できます。例えば、フォルダの権限を構成するには、フォルダの権限の設定特権を持つロールが割り当てられている必要があります。

1. 権限を構成するオブジェクトに移動します。

例:

- 機械学習モデルまたはモデルデプロイメントに対する権限を設定するには、Model Serve で、アセットを含むプロジェクトとフォルダを開きます。
2. オブジェクトを含む行で、**【アクション】** をクリックして **【権限】** を選択するか、**【権限の変更】** アイコンをクリックします。

【権限】 ダイアログボックスには、オブジェクトに対する権限を持つユーザーとグループが一覧表示されます。

【権限】 ダイアログボックスにユーザーまたはグループが一覧表示されない場合は、そのオブジェクトに対して権限が構成されていません。オブジェクトタイプに対して適切な特権を持つユーザーは、オブジェクトにアクセスできます。

3. オブジェクトに対するユーザー権限を構成するには:

- a. **【ユーザー】** を選択します。
- b. ユーザーが **【ユーザー】** の一覧に表示されていない場合は、**【追加】** をクリックし、ユーザーを選択します。
- c. ユーザーに対する適切な権限を有効または無効にします。

注: オブジェクトに対するユーザー権限を付与すると、Informatica Intelligent Cloud Services によって管理者もオブジェクトに対するアクセス権限を持つユーザーとして追加されます。これにより、権限を構成するときにオブジェクトへのアクセスが失われるのを防ぎます。

4. オブジェクトに対するユーザーグループの権限を構成するには:

- a. **【グループ】** を選択します。
- b. グループが **【グループ】** の一覧に表示されていない場合は、**【追加】** をクリックし、グループを選択します。
- c. グループに対する適切な権限を有効または無効にします。

注: オブジェクトに対するグループの権限を付与すると、Informatica Intelligent Cloud Services によって管理者もオブジェクトに対するアクセス権限を持つユーザーとして追加されます。これにより、権限を構成するときにオブジェクトへのアクセスが失われるのを防ぎます。

5. オブジェクトの権限の制限をすべて削除するには、**【権限】** ダイアログボックスからすべてのユーザーとグループを削除します。

すべてのユーザーとグループを削除すると、そのオブジェクトタイプに対して適切な特権を持つすべてのユーザーがオブジェクトにアクセスできるようになります。

6. **【保存】** をクリックします。

第 6 章

トラブルシューティング

Apple のアーカイブユーティリティで圧縮されたモデルファイルをアップロードできません。

アーカイブユーティリティがディレクトリを圧縮すると、ファイル名に複数のピリオドが含まれたファイルを含む、__MACOSX という名前のディレクトリが作成されます。セキュリティ上の理由から、これらのファイルは許可されていません。別のアプリケーションを使用してモデルファイルを圧縮するか、モデルファイルを圧縮するときに__MACOSX ディレクトリを除外してから、モデルファイルを再度アップロードしてください。

注: macOS はデフォルトで__MACOSX ディレクトリを隠しているため、このディレクトリを表示するには隠しフォルダを表示することが必要になる場合があります。

索引

C

Cloud アプリケーション統合コミュニティ
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](#)

W

Web サイト [5](#)

あ

アセット
コピー [20](#)
タグ [23](#)
タグの割り当て [24](#)
タグの作成 [23](#)
移動 [21](#)
削除 [22](#)
名前の変更 [22](#)
アップグレード通知 [6](#)

く

クイックスタートモデル
テキスト翻訳モデル [11](#)
デプロイ [12](#)
画像分類モデル [10](#)
概要 [10](#)

こ

コピー
アセット [20](#)
フォルダ [20](#)
プロジェクト [20](#)

し

システムステータス [6](#)

す

スケジュール設定
タスクとタスクフローのコピー [20](#)
ステータス
Informatica Intelligent Cloud Services [6](#)

た

タグ
プロパティ [25](#)
割り当て [24](#)
作成 [23](#)
削除 [25](#)
編集 [25](#)
タスク
削除 [22](#)

て

テキスト翻訳モデル [11](#)
デプロイメント
設定 [15](#)

と

トラブルシューティング [19](#), [29](#)

ふ

フォルダ
コピー [20](#)
移動 [21](#)
削除 [22](#)
プロジェクト
コピー [20](#)
削除 [22](#)

め

メンテナンスの停止 [6](#)

も

モデルアーティファクト [14](#)
モデルデプロイメント
テスト [17](#)
設定 [15](#)
モデルデプロイメントのテスト [17](#)
モデルハブ [10](#)

モデルファイルのアップロード [14](#)
モニタ [18](#)

ろ

ログのダウンロード [19](#)