



Red Hat OpenShift Data Science Self-managed 1.29

Amazon S3 からのデータの統合

Amazon Web Services (AWS) の Simple Storage Service (S3) バケットに保存されているデータの使用

Red Hat OpenShift Data Science Self-managed 1.29 Amazon S3 からのデータの統合

Amazon Web Services (AWS) の Simple Storage Service (S3) バケットに保存されているデータの使用

法律上の通知

Copyright © 2023 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

概要

Amazon Web Services (AWS) の Simple Storage Service (S3) バケットに保存されているデータの使用方法を説明します。

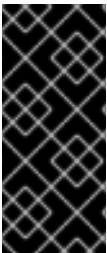
目次

第1章 RED HAT ドキュメントへのフィードバック (英語のみ)	3
第2章 前提条件	4
第3章 ノートブックセルを使用した AMAZON S3 クライアントの作成	5
第4章 ノートブックセルを使用して利用可能な AMAZON S3 バケットを一覧表示	6
第5章 ノートブックセルを使用して利用可能な AMAZON S3 バケットのファイルを一覧表示	7
第6章 ノートブックセルを使用して利用可能な AMAZON S3 バケットからファイルをダウンロード	9
第7章 ノートブックセルを使用した利用可能な AMAZON S3 バケットへのファイルのアップロード	10
第8章 関連情報	11

第1章 RED HAT ドキュメントへのフィードバック (英語のみ)

ドキュメントの改善点は、Red Hat にお知らせください。以下の手順に従って、ドキュメントページから直接フィードバックを送信できます。

1. カスタマーポータルにログインしていることを確認します。
2. このドキュメントが **Multi-page HTML** 形式で表示されていることを確認します。
3. フィードバックを送信する文章を強調表示します。**Add Feedback** プロンプトが表示されます。
4. **Add Feedback** をクリックします。
5. **Feedback** テキストボックスにコメントを入力し、**Submit** をクリックします。



重要

一部の広告ブロッカーは、Red Hat ドキュメントに関するフィードバック機能を妨げる可能性があります。広告ブロッカーが有効になっている Web ブラウザーを使用していることが原因でフィードバックできない場合は、広告ブロッカーを無効にすることを検討してください。広告ブロッカーを無効にする方法の詳細は、Web ブラウザーのドキュメントを参照してください。

フィードバックが送信されると、Red Hat は自動的に問題の追跡を開始します。Submit をクリックした後に表示されるリンクを開き、問題の監視を開始するか、さらにコメントを追加して問題に関する詳細情報を送信してください。

フィードバックの提供にご協力いただきありがとうございました。

Jupyter ノートブックで作業する場合は、Amazon Web Services (AWS) Simple Storage Service (S3) バケットに保存されているデータを使用することが必要な場合があります。このセクションでは、Amazon S3 に保存されているデータを使用するコマンドおよび手順を説明します。

第2章 前提条件

- Red Hat OpenShift Data Science で実行する Jupyter サーバー。
- Amazon Web Services S3 バケットへのアクセス。
- Amazon S3 アカウントの **AWS Access Key ID** および **AWS Secret Access Key** を特定する。
- Jupyter ノートブック。

第3章 ノートブックセルを使用した AMAZON S3 クライアントの作成

Amazon S3 バケットでデータを操作するには、そのサービスに対する要求を処理するローカルクライアントを作成する必要があります。

前提条件

- Red Hat OpenShift Data Science で実行している Jupyter ノートブックサーバーへのアクセス。
- My Security Credentials** の Amazon Web Services アカウントの値を使用して、ノートブックサーバーの起動時に **AWS_ACCESS_KEY_ID** および **AWS_SECRET_ACCESS_KEY** 環境変数の値を定義している。

手順

- 新しいノートブックセルで、以下を追加して必要なライブラリーをインポートします。

```
import os
import boto3
from boto3 import session
```

- 別の新規ノートブックセルで、以下を定義してセッションとクライアントを作成します。

- 認証情報を定義します。

```
key_id = os.environ.get('AWS_ACCESS_KEY_ID')
secret_key = os.environ.get('AWS_SECRET_ACCESS_KEY')
```

- クライアントセッションを定義します。

```
session = boto3.session.Session(aws_access_key_id=key_id,
aws_secret_access_key=secret_key)
```

- クライアント接続を定義します。

```
s3_client = boto3.client('s3', aws_access_key_id=key_id,
aws_secret_access_key=secret_key)
```

検証

- 新しいセルを作成し、以下のような Amazon S3 コマンドを実行します。

```
s3_client.list_buckets()
```

正常な応答には、**HTTPStatusCode 200** と以下のような **Buckets** の一覧が含まれます。

```
'Buckets': [{'Name': 'my-app-asdf3-image-registry-us-east-1-wbmlcvbasdfasdgvtsmkpt',
'CreationDate': datetime.datetime(2021, 4, 21, 6, 8, 52, tzinfo=tzlocal())},
{'Name': 'cf-templates-18rxasdfggawsvb-us-east-1',
'CreationDate': datetime.datetime(2021, 2, 15, 18, 35, 34, tzinfo=tzlocal())}]
```

第4章 ノートブックセルを使用して利用可能な AMAZON S3 バケットを一覧表示

アカウントで利用可能なバケットを一覧表示して、アクセス可能なバケットを確認できます。

前提条件

- ノートブックの前のセルで Amazon S3 クライアントを設定する。詳細は、[ノートブックセルを使用した Amazon S3 クライアントの作成](#) を参照してください。

手順

1. 新しいノートブックセルを作成し、**s3_client** を使用して利用可能なバケットを一覧表示します。

```
s3_client.list_buckets()
```

2. 完全な応答ではなく、名前のみを出力することで、このバケット一覧を読みやすくすることができます。次に例を示します。

```
for bucket in s3_client.list_buckets()['Buckets']:
    print(bucket['Name'])
```

これにより、以下のような出力が返されます。

```
my-app-asdf3-image-registry-us-east-1-wbmlcvbasdgasdgtkpt
cf-templates-18rxuasgasgvb-us-east-1
```

関連情報

- [ノートブックセルを使用した Amazon S3 クライアントの作成](#)
- [Amazon Web Services list buckets command reference](#)

第5章 ノートブックセルを使用して利用可能な AMAZON S3 バケットのファイルを一覧表示

バケットのオブジェクトを一覧表示し、アクセス可能なバケットにある利用可能なファイルを確認できます。バケットは通常のファイルシステムの代わりにオブジェクトストレージを使用するため、オブジェクトの命名は通常のファイル命名と異なります。バケットのオブジェクトはキーによって常に認識されます。これはバケットの完全パスとファイル自体の名前で設定されます。

前提条件

- ノートブックの前のセルで Amazon S3 クライアントを設定する。詳細は、[ノートブックセルを使用した Amazon S3 クライアントの作成](#) を参照してください。

手順

- 新しいノートブックセルを作成し、バケットのオブジェクトを一覧表示します。以下に例を示します。

```
bucket_name = 'std-user-bucket1'
s3_client.list_objects_v2(Bucket=bucket_name)
```

以下の形式で複数のオブジェクトが返されます。

```
{'Key':
'docker/registry/v2/blobs/sha256/00/0080913dd3f10aadb34asfgsgsdgasdga072049c93606b98b
ec84adb259b424f/data',
'LastModified': datetime.datetime(2021, 4, 22, 1, 26, 1, tzinfo=tzlocal()),
'ETag': '"6e02fad2deassadfsf900a4bd7344ffe"',
'Size': 4052,
'StorageClass': 'STANDARD'}
```

- 完全な応答ではなくキーのみを出力するなどして、この一覧の読み取りを容易化できます。

```
bucket_name = 'std-user-bucket1'
for key in s3_client.list_objects_v2(Bucket=bucket_name)['Contents']:
    print(key['Key'])
```

これにより、以下のような出力が返されます。

```
docker/registry/v2/blobs/sha256/00/0080913dd3f10aadb34asfgsgsdgasdga072049c93606b98b
ec84adb259b424f/data
```

- また、クエリーをフィルターして、特定の "パス" またはファイル名などを一覧表示できます。以下はその例です。

```
bucket_name = 'std-user-bucket1'
for key in s3_client.list_objects_v2(Bucket=bucket_name, Prefix='<start_of_file_path>')
['Contents']:
    print(key['Key'])
```

前の例で、<start_of_file_path> を独自の値に置き換えます。

関連情報

- [ノートブックセルを使用した Amazon S3 クライアントの作成](#)
- [Amazon Web Services list objects command reference](#)

第6章 ノートブックセルを使用して利用可能な AMAZON S3 バケットからファイルをダウンロード

`download_file` メソッドを使用して、ファイルをノートブックサーバーにダウンロードできます。

前提条件

- ノートブックの前のセルで Amazon S3 クライアントを設定する。詳細は、[ノートブックセルを使用した Amazon S3 クライアントの作成](#) を参照してください。

手順

- ノートブックセルに以下の詳細を定義します。

- ファイルが置かれているバケット。<name_of_the_bucket> を独自の値に置き換えます。

```
bucket_name = '<name_of_the_bucket>'
```

- ダウンロードするファイルの名前。<name_of_the_file_to_download> を独自の値に置き換えます。

```
file_name = '<name_of_the_file_to_download>' # Full path from the bucket
```

- ダウンロード後にファイルに使用される名前。完全パス、相対パス、または新しいファイル名のみ指定できます。<name_of_the_file_when_downloaded> を独自の値に置き換えます。

```
new_file_name = '<name_of_the_file_when_downloaded>'
```

- 以前の変数を引数として指定し、ファイルをダウンロードします。

```
s3_client.download_file(bucket_name, file_name, new_file_name)
```



注記

ファイルをオブジェクトとして取得し、`read()` メソッドを使用して標準ファイルとしてストリーミングする場合は、[Amazon Sev Services get object command reference](#) を参照してください。

関連情報

- [ノートブックセルを使用した Amazon S3 クライアントの作成](#)
- [Amazon Web Services download file command reference](#)

第7章 ノートブックセルを使用した利用可能な AMAZON S3 バケットへのファイルのアップロード

`upload_file` メソッドを使用して、ノートブックサーバーから Amazon S3 バケットにファイルをアップロードできます。

前提条件

- ノートブックの前のセルで Amazon S3 クライアントを設定する。詳細は、[ノートブックセルを使用した Amazon S3 クライアントの作成](#) を参照してください。

手順

- ノートブックセルに以下の詳細を定義します。

- アップロードするファイルの名前。ファイルへの完全なローカルパスを含める必要があります。`<name_of_the_file_to_upload>` を独自の値に置き換えます。

```
file_name = '<name_of_the_file_to_upload>'
```

- ファイルをアップロードするバケットの名前。`<name_of_the_bucket>` を独自の値に置き換えます。

```
bucket_name = '<name_of_the_bucket>'
```

- ファイルをバケットに保存するために使用する完全なキー。`<full_path_and_file_name>` を独自の値に置き換えます。

```
key = '<full_path_and_file_name>'
```

- 以前の変数を引数として指定し、ファイルをアップロードします。

```
s3_client.upload_file(file_name, bucket_name, key)
```

関連情報

- [ノートブックセルを使用した Amazon S3 クライアントの作成](#)
- [Amazon Web Services upload file command reference](#)

第8章 関連情報

- [Red Hat OpenShift Data Science ドキュメント](#)