

Eon モード 版のデータロード

原文は[こちら](#)

概要

Vertica 9.0.x 以降、Eon モードのベータ版でデータベースを操作できます。Eon モードのベータ版は、計算処理をデータベースのストレージレイヤーから分離します。Eon モードのベータ版の実装は Amazon Web Services に限られています。いくつかの点で、Eon モードのベータ版は従来の Vertica Enterprise モードとは異なります。データを効果的にロードすることで、さまざまな需要のワークロードをどのように活用するかが大きな違いになります。Eon モードのベータ版のロードをよりよく理解するために、ユースケースを見てみましょう。

S3 からのロード

この投稿は、ソースデータが S3 上にあることを前提としています。この投稿は、S3 からロードするユーザーを対象としており、ローカルディスクからロードするユーザーを対象としていません。

Enterprise モードと同じように、入力データのロードは Eon モードのベータ版でも同じですが、注意すべき点があります。この例でそれらについて見ていきましょう。デフォルトの US East Amazon リージョンにある 1 つの大きなファイルをロードするとします。これを行うには次の 3 つの方法があります。

1. ファイルをダウンロードし、COPY FROM STDIN を使用してロード
 - a. このオプションは機能しますが、遅くなる可能性があり、推奨されません。このオプションを使用すると、S3 からローカルディスクに 1 回、ディスクから Vertica にもう 1 回、データ転送のコストを 2 回支払うことになります。
2. COPY t WITH SOURCE S3 (file = 's3://bucket/file...') のフォーマットで AWS UDX を使用
 - a. このオプションはオプション 1 より優れていますが、UDx は Apportioned load をサポートしていません。このオプションには、UDSession パラメーターであるセッションパラメータも必要です。詳細については、[Configuring the Vertica Library for Amazon Web Services](#) を参照してください。
3. COPY t FROM 's3://bucket/file...' を使って S3 から COPY するように UDFS を使用
 - a. これは、異なるノードからのファイルの Apportioned load をサポートするため、最良のオプションです。Vertica のエンジニアが実行するパフォーマンステストに基づいて、このオプションでロードすることは、オプション 2 を使用するよりも少なくとも 2 倍高速です。このオプションにはセッションパラメーターは必要ありません。

これらの方法のいずれかが機能しますが、いくつかの異なるシナリオに遭遇した場合、特定の方法を選択することができます。

シナリオ 1: US West などの別の地域にファイルをロードしたい場合はどうなりますか？

ファイルが別の地域にある場合、オプション 2 を使用する必要があります。前述の理由により、オプション 1

は使用しないでください。リージョン設定がグローバルパラメーター(AWSRegion)であるため、ファイルシステムは 2 つの異なるリージョンのファイルに同時にアクセスできないため、オプション 3 は使用できません。マルチリージョンではオプション 3 はサポートされていません。

シナリオ 2: AWS エンドポイントが異なる場合はどうなりますか？

異なる AWS エンドポイントに遭遇した場合、いくつかの調整が必要です。オプション 2 の場合、UDSession パラメーターである `aws_endpoint` を設定する必要があります。オプション 3 の場合、`AWSEndpoint` 構成パラメーターを設定する必要があります。これもグローバルパラメーターです。詳細については、[Vertica documentation](#) の [AWS Parameters](#) を参照してください。

シナリオ 3: ファイルが圧縮されている場合はどうなりますか？

ファイルが圧縮されている場合、フィルターを使用する必要があります。これにより、異なるノードからファイルの異なる部分をロードする UDFS の機能が無効になります。オプション 2 では、SDK サンプルで GZIP フィルターをコンパイルする必要があります。オプション 3 では、ネイティブフィルターを使用してファイルを解凍することができます。この場合、オプション 3 はオプション 2 よりも利点がないため、どちらのオプションも使用できます。詳細については、COPY の [examples](#) を参照してください。

マネージメントコンソールからのロード

マネージメントコンソール(MC)を使用して、S3 から Vertica テーブルにデータをロードし、データロードの結果を表示することもできます。

Overview ページの下部にある「Load」タブから「Load」ページにアクセスできます。「Load」ページでは、テーブルの 2 種類のデータを表示できます。S3 データの場合、「Instance」タブを使用する必要があります。このタブでは、データの単一のロード処理が可能です。

「Instance」タブから「New S3 Data Load」ボタンを選択し、フォームに記入します。IAM ロール認証(同じアカウントまたは法人アカウントで作成または所有されている S3 バケットの場合)を使用するか、AWS キークレデンシャルを入力するか(認証によって別のユーザーのバケットにアクセスする場合)を選択できます。

データのタイプに応じて、生成される COPY ステートメントのパラメーターを追加することができます。裏では、MC は UDFS を使用して S3 から COPY t FROM 's3:// bucket / file ...'でコピーします。

Create New Amazon S3 Loading Job X

Note: you must be the owner of the table in order to perform data load.

AWS S3 Authentication Use IAM Role Authentication Use AWS Key Credentials
To use IAM, please make sure IAM role on the instance is enabled

AWS Source URL *

AWS Region *

Vertica Target Database VerticaDB Your query will be executed as database user `dbadmin`

Schema name *

Table name *

Direct (bypass WOS)

Copy Parameters

Handling of rejected data

Capture rejected data in a table Yes No
You must have CREATE privilege on schema if this table doesn't exist

Rejections Table (autogenerated)

Reject Max

* Required field.

ロード処理の結果は、ロード履歴テーブルに表示されます。コマンドラインやバッチスクリプトを含むすべてのロード処理の結果を確認できます。MC から開始されたデータロードは、ストリーム名「MC_S3_Load」を持つことに注意してください。詳細については、[Loading Data Using MC](#) を参照してください。

Vertica Management Console

uidadmin Log out 2.00 ?

Databases and Clusters > VerticaDB > Data Load Activity

Auto Refresh Last update: 18 Jan 2018 17:51:41

Continuous Instance

Load history for the past: 1 hour

New S3 Data Load

Status	Source File	Time Started	User	Schema Name	Table Name	Execution Time (ms)	Accepted Rows	Rejected Rows	Stream Name
Failure	N/A	Jan 18, 2018 5:30:45 PM	uidadmin	public	towns	59	0	0	MC_S3_Load
Success	nimbusdb/natalia-s3-data/t...	Jan 18, 2018 5:29:03 PM	uidadmin	public	towns	429	238	0	MC_S3_Load
Success	mcqabucket/csvfiles/comma...	Jan 18, 2018 5:26:52 PM	uidadmin	public	towns	287	238	0	MC_S3_Load
Failure	N/A	Jan 18, 2018 5:25:39 PM	uidadmin	public	towns	177	0	0	MC_S3_Load
Success	nimbusdb/natalia-s3-data/t...	Jan 18, 2018 5:22:34 PM	uidadmin	public	examples	148	0	239	MC_S3_Load
Success	nimbusdb/natalia-s3-data/t...	Jan 18, 2018 5:12:38 PM	uidadmin	public	towns	132	0	239	MC_S3_Load

Overview Activity Manage Design Load Query Execution Query Plan License Settings

ロード処理が失敗した場合、「Failure」ステータスをクリックして何がうまくいかなかったのかの詳細を確認できます。次の例では、バケット内のファイルの名前が正しくありませんでした。

Load history for the past: 1 hour

Failed Load Details

For query with sessionId: v_verticadb_node0001-6109:0x72, transactionId: 45035996273705956, statementId: 1.

Copy Command

```
COPY public.towns FROM 's3://nimbusdb/badFile' DELIMITER ',' ENCLOSED BY '''' SKIP 1 STREAM NAME 'MC_S3_Load' DIRECT REJECTED DATA AS TABLE public.s3_load_rejections_towns
```

Error Time

18 Jan 2018 17:30:45

Error Code

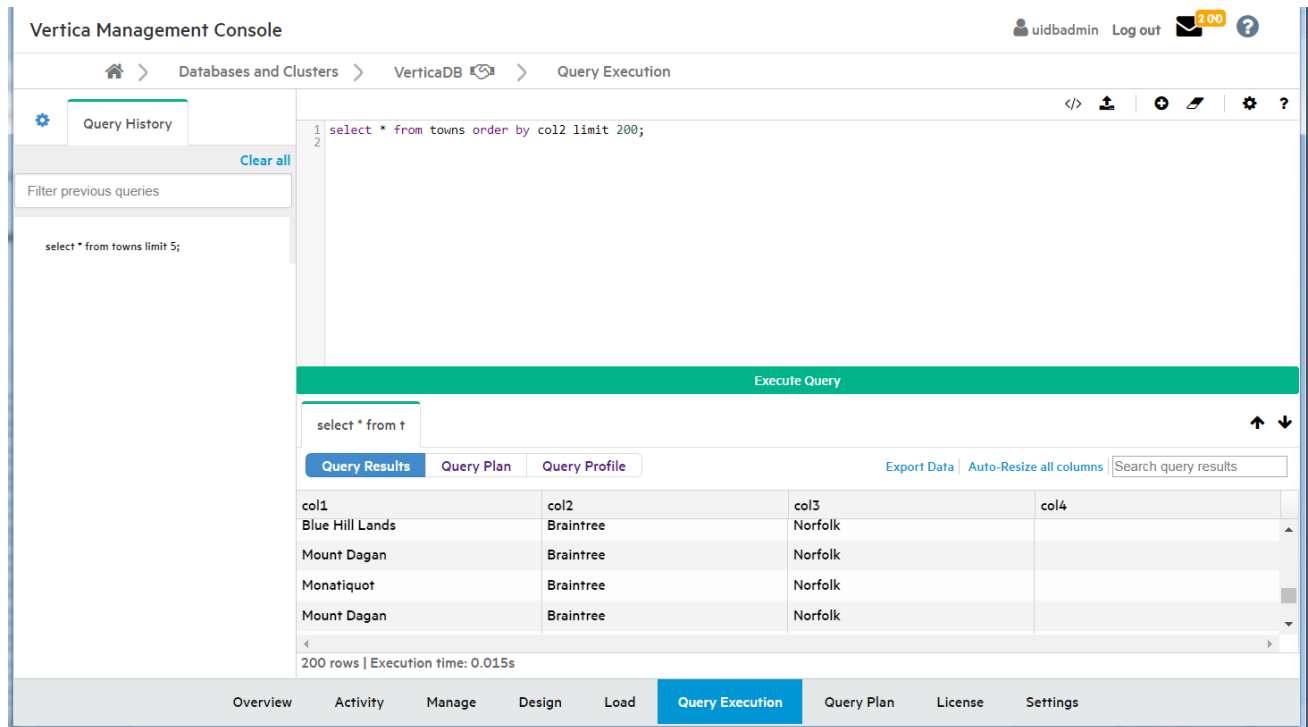
7756

Error Details

```
COPY(MC_S3_Load): Could not open file [s3://nimbusdb/badFile] for reading; File not found
```

Close

MCを使用してロード直後にデータへのクエリを開始できます。「Query Execution」タブを選択し、クエリを入力します。



Vertica Management Console

uidbadmin Log out 2:00 ?

Databases and Clusters > VerticaDB > Query Execution

Query History

Clear all

Filter previous queries

select * from towns limit 5;

Execute Query

select * from t

Query Results Query Plan Query Profile

Export Data Auto-Resize all columns Search query results

col1	col2	col3	col4
Blue Hill Lands	Braintree	Norfolk	
Mount Dagan	Braintree	Norfolk	
Monatiquot	Braintree	Norfolk	
Mount Dagan	Braintree	Norfolk	

200 rows | Execution time: 0.015s

Overview Activity Manage Design Load **Query Execution** Query Plan License Settings

詳細については、[Vertica documentation](#) の [Using Eon Mode Beta](#) を参照してください。

www.vertica.com

© 2018 Micro Focus. All rights reserved. Micro Focus and the Micro Focus logo, among others, are trademarks or registered trademarks of Micro Focus or its subsidiaries or affiliated companies in the United Kingdom, United States and other countries. All other marks are the property of their respective owners.