Amazon Athena ACID transaction workshop

2022/04/15

シニアエバンジェリスト 亀田

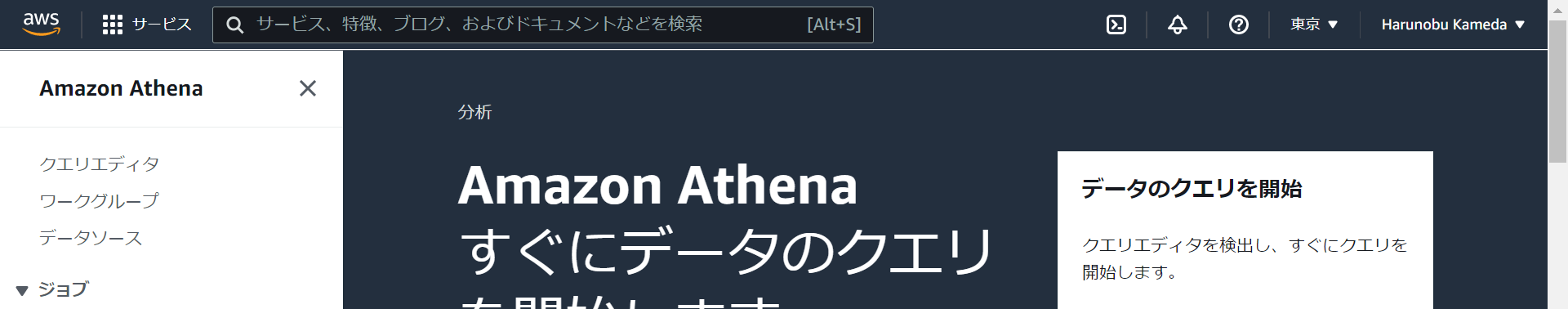
はじめに：

Amazon AthenaはサーバレスでＳ３などのデータに対してSQLクエリーが発行可能なインタラクティブな分析サービスです。従来、AthenaはSELECT 及びCTAS (Create Table As Select)のみに対応していましたが、新しくACIDトランザクションをサポートし、INSERT, DELETE, UPDATE、よして過去のデータに対してクエリーが発行可能なタイムトラベルクエリに対応しました。この機能を使うためには、S3に保存されているデータはApache Iceberg形式である必要があります。

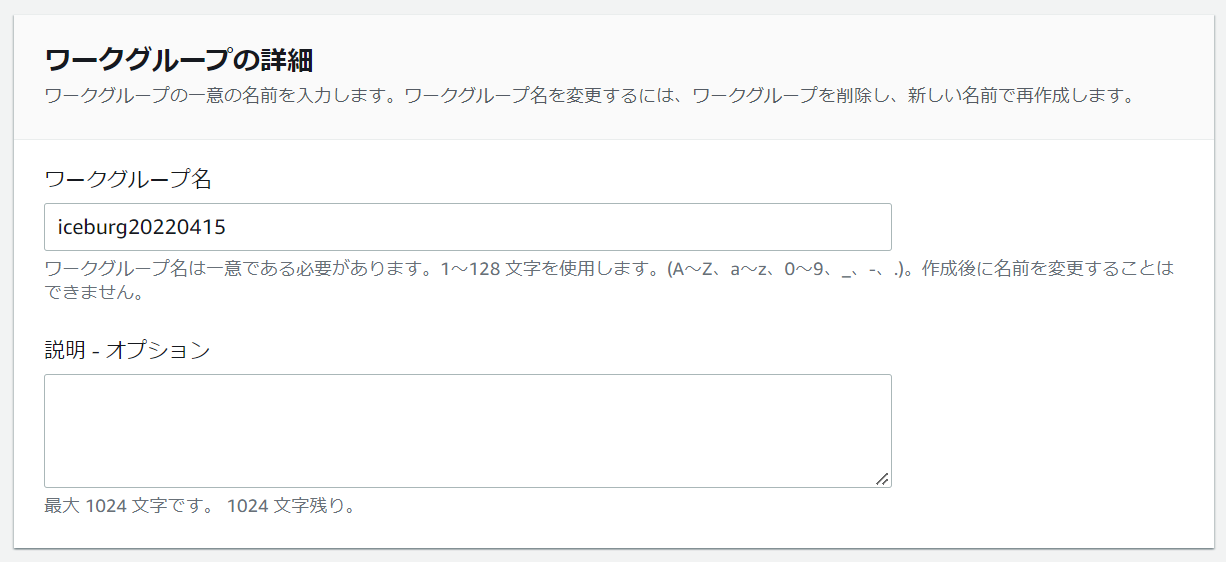
1. Ｓ３のマネージメントコンソールに移動します
2. [バケットの作成]をおします



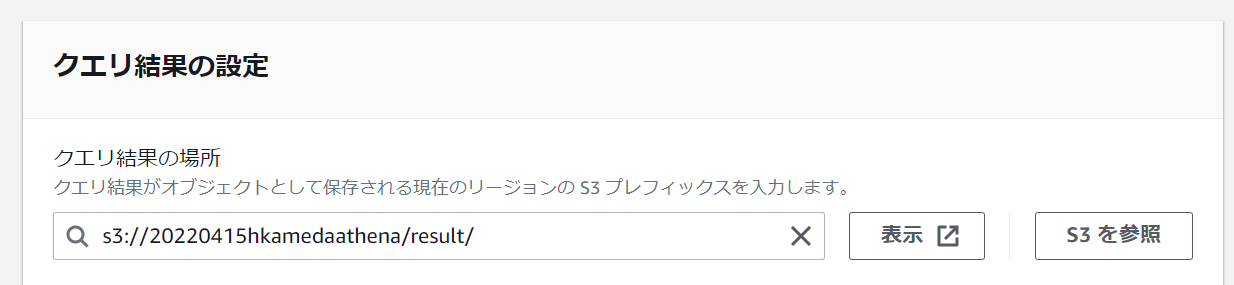
1. 適当な名前を入力しその他はデフォルトのまま[バケットを作成]をおします。作成したバケットの名前はコピーしておいてください
2. Athenaのマネージメントコンソールを開きワークグループをクリックします。ワークグループとはユーザーごとの作業環境でありクエリの発行結果や一時計算に用いられるS3バケットを指定します。



1. [ワークグループを作成]をおします
2. 適当な名前を設定します



1. 先程作成したバケットを指定し [/result/]を付与します



1. [AWS CloudWatch にクエリメトリクスを発行]のチェックを外し、[上書き]のチェックをつけます

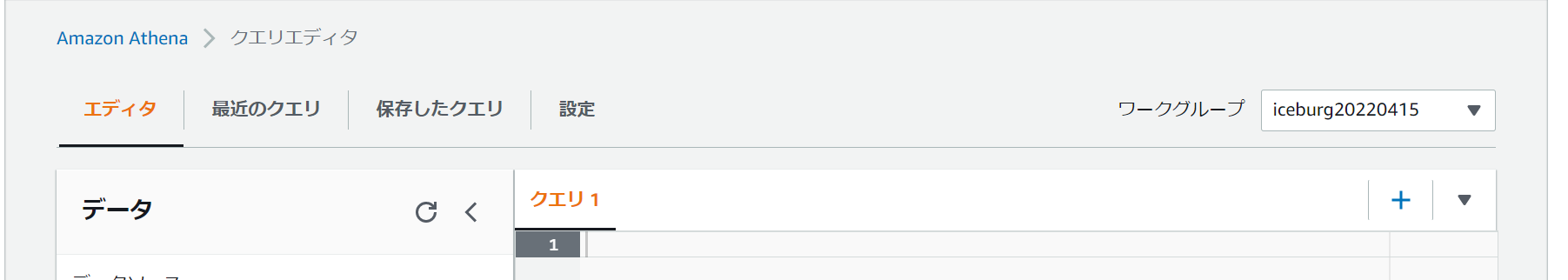


1. [ワークグループを作成]をおします。以下の通りワークグループが作成されていれば完了です



これで実行の準備ができました。早速クエリーを発行していきます。

1. 左ペインの[クエリエディタ]をおします
2. 画面右上のドロップダウンから先程作成した[ワークグループ]を選びます。追加で新しいウインドウが表示された場合は、オレンジ色のボタンを押します



1. 以下のクエリーを２回に分けて実行します。<mybucket>は先程作成したS3バケットへ置き換えてください

１回目：

CREATE DATABASE IF NOT EXISTS my\_database;

２回目：

CREATE TABLE iceberg\_table20220415 (

id int,

data string,

category string)

PARTITIONED BY (category, bucket(16,id))

LOCATION 's3://**<mybucket>**/iceberg\_table/'

TBLPROPERTIES (

'table\_type'='ICEBERG',

'format'='parquet',

'write\_target\_data\_file\_size\_bytes'='536870912'

)



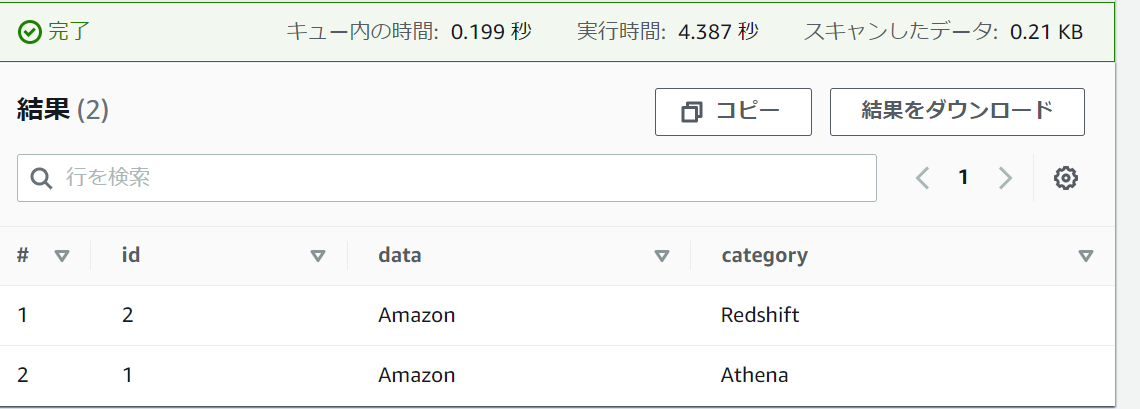
1. テーブルができましたので以下のクエリーを順番に1行ずつじっこうします

INSERT INTO my\_database.iceberg\_table20220415 (id, data, category) values(1, 'Amazon', 'Athena');

INSERT INTO my\_database.iceberg\_table20220415 (id, data, category) values(2, 'Amazon', 'Redshift');

1. 以下のクエリーで、テーブルの中身を確認します。2行Insertが成功している鵜ことがわかります。

SELECT \* from my\_database.iceberg\_table20220415;



1. 次にタイムトラベルクエリを試してみます。以下のクエリーは上で実行したクエリーを1日前のデータに対して実行します

SELECT \* FROM my\_database.iceberg\_table20220415 FOR VERSION AS OF (current\_timestamp - interval '1' day);

1日前はテーブルが存在していないため、以下の通りエラーが発生します



タイムトラベルクエリの詳細はこちらをご覧ください

<https://docs.aws.amazon.com/athena/latest/ug/querying-iceberg-table-data.html>

1. 今度はUpdateクエリーを実行してみます

UPDATE my\_database.iceberg\_table20220415 SET data='Redshift Spectrum' WHERE id = 2;

1. 全件検索をかけると以下の通り値が書き換わっていることがわかります

SELECT \* from my\_database.iceberg\_table20220415;



1. 次に以下のクエリーでレコードの削除を行います

DELETE FROM my\_database.iceberg\_table20220415 WHERE id=1;

1. 全件検索をかけると以下の通りレコードが1件削除されていることがわかります

SELECT \* from my\_database.iceberg\_table20220415;



では最後に以下のクエリを数分ずつ遡って実行してください。通常では出てこない過去のスナップショットに対するクエリ結果が出てきます。

SELECT \* FROM iceberg\_table20220415 FOR TIMESTAMP AS OF TIMESTAMP '2025-01-26 08:40:00 UTC'

おつかれさまでした！

削除は以下を行ってください

S3バケット

Athenaワークグループ