

S3Tables に AWS Lambda から DuckDB でクエリする

nk（えぬけー）

自己紹介

- データ好きなエンジニア
- 最近は Iceberg と DuckDB に注目中！

今回の内容の記事

DuckDB の公式ブログにて

AWS S3 Tables に Iceberg REST Catalog API 経由で接続する方法が紹介されました！

 [DuckDB Blog - 2025/03/14](#)

データカタログにアクセスしつつ料金も抑えられる？

- Iceberg REST Catalog API 経由 の組み合わせなら コストを抑えつつ分析が可能！
(のはず)
- Iceberg 内のカタログにも接続できるようになった。

① S3Tables バケット の準備

- 最初は Re:Invent 直後に作成した S3Tables で試したが失敗...
- 東京リージョンで一から再作成して成功！

👉 [参考記事](#)

② Athena で Iceberg テーブル作成

```
CREATE TABLE test1.nk (  
    id INT,  
    name STRING  
)  
TBLPROPERTIES ('table_type' = 'ICEBERG');
```

```
INSERT INTO test.nk  
VALUES  
(1, 'nk')
```

③ IAM ポリシー設定（抜粋）

```
{  
  "Action": [  
    "s3tables:GetTableBucket",  
    "s3tables:ListTableBuckets",  
    "s3tables:GetNamespace",  
    "s3tables:ListNamespaces",  
    "s3Tables:GetTable",  
    "s3Tables:ListTables",  
    "s3Tables:GetTableData",  
    "s3Tables:PutTableData",  
    "s3Tables:GetTableMetadataLocation",  
    "s3Tables:UpdateTableMetadataLocation"  
  ],  
  "Resource": ["*"],  
  "Effect": "Allow"  
}
```

④ Lambda Layer に DuckDB を追加

```
mkdir python
```

```
pip install -t python --platform manylinux2014_x86_64  
--only-binary=:all: duckdb
```

```
zip -r layer.zip python
```


⑤ Lambda 関数の実行コード

```
import json
import duckdb

def lambda_handler(event, context):
    # 1. DuckDB 初期化
    con = duckdb.connect(database=':memory:')
    con.execute("SET home_directory='/tmp';")

    # 2. 拡張モジュールをインストール&ロード
    con.sql("FORCE INSTALL aws FROM core_nightly")
    con.sql("FORCE INSTALL httpfs FROM core_nightly")
    con.sql("FORCE INSTALL iceberg FROM core_nightly")
    con.sql("LOAD aws")
    con.sql("LOAD httpfs")
    con.sql("LOAD iceberg")
```

⑤ AWS Lambda 関数実行

```
# 3
con.sql("""
CREATE SECRET (
  TYPE s3,
  KEY_ID 'XXXXXXXXXX',
  SECRET 'XXXXXXXXXX',
  REGION 'ap-northeast-1'
);
""")
```

```
# 4
con.sql("""
ATTACH 'XXXXXXXXXX'
AS s3_tables_db (
  TYPE iceberg,
  ENDPOINT_TYPE s3_tables
);
""")
```

⑥ 実行結果

```
# 5
print(con.sql("SHOW ALL TABLES"))
print(con.sql("FROM s3_tables_db.test.nk"))
```

database varchar	schema varchar	name varchar	column_names varchar[]	column_types varchar[]	temporary boolean
s3_tables_db	test	nk	[__]	[INTEGER]	false

id int32	name varchar
1	nk

まとめ

iceberg のテーブル表示までなんとかできました。

Lambda 上で使用するために毎回リポジトリやシークレットの実行をやるのはちょっと...

自分の設定意外にも他にもやり方があるのかもしれませんが。