

Import Postgres data into RDS using S3 and aws_s3

Asked 3 years, 4 months ago Modified 6 months ago Viewed 21k times  Part of [AWS](#) Collective

I'm having a hard time importing data from S3 into an RDS postgres instance. [According to the docs](#), you can use this syntax:

23

```
aws_s3.table_import_from_s3 (  
    table_name text,  
    column_list text,  
    options text,  
    bucket text,  
    file_path text,  
    region text,  
    access_key text,  
    secret_key text,  
    session_token text  
)
```

So, in pgAdmin, I did this:

```
SELECT aws_s3.table_import_from_s3(  
    'contacts_1',  
    'firstname,lastname,imported',  
    '(format csv)',  
    'com.foo.mybucket',  
    'mydir/subdir/myfile.csv',  
    'us-east-2',  
    'AKIAYYXUMxxxxxxxxxx',  
    '3zB4S5jb1xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'  
);
```

I also tried it with an explicit NULL for the last parameter.

The error message I get is:

```
NOTICE:  CURL error code: 51 when attempting to validate pre-signed URL, 1  
attempt(s) remaining  
NOTICE:  CURL error code: 51 when attempting to validate pre-signed URL, 0  
attempt(s) remaining  
  
ERROR:  Unable to generate pre-signed url, look at engine log for details.  
SQL state: XX000
```

I checked the server logs and there was no further information.

I have triple-checked the correctness of all the parameters. How do I make this work?

UPDATE:

I can confirm that I can do an `s3.getObject()` in the Java aws sdk using these same credentials.

Share Follow

edited Oct 30, 2019 at 2:34

asked Oct 29, 2019 at 22:02



John Rotenstein

228k 21 352 431



ccleve

14.8k 27 88 153

- 1 This could be the periods (.) in the bucket name. This question's a bit old now, but worth trying from a different bucket without periods in the name. I hit this a few weeks back. – [SteveChapman](#) Sep 28, 2020 at 12:57

stackoverflow.com/a/64594301/6718580 – [Vladyslav Diachenko](#) Jun 18, 2021 at 8:38

Sorted by:

Highest score (default)

8 Answers



25



The main issue here is that you need to 1) add a IAM role to the RDS instance to access the S3 bucket and 2) add an S3 endpoint to the VPC where the RDS instance run in order to allow communications.

This is the procedure I followed to make it work, using AWS cli commands in a shell (take care of value properly the environmental variables involved), hope it can help:

1. Create the IAM role:

```
$ aws iam create-role \
  --role-name $ROLE_NAME \
  --assume-role-policy-document '{"Version": "2012-10-17", "Statement":
[{"Effect": "Allow", "Principal": {"Service": "rds.amazonaws.com"}, "Action":
"sts:AssumeRole"}]}'
```

2. Create the IAM policy that will be attached to the IAM role:

```
$ aws iam create-policy \
  --policy-name $POLICY_NAME \
  --policy-document '{"Version": "2012-10-17", "Statement": [{"Sid":
"s3import", "Action": ["s3:GetObject", "s3:ListBucket"], "Effect": "Allow",
"Resource": ["arn:aws:s3:::${BUCKET_NAME}",
"arn:aws:s3:::${BUCKET_NAME}/*"]}']}'
```

3. Attach the policy:

```
$ aws iam attach-role-policy \
  --policy-arn arn:aws:iam::$AWS_ACCOUNT_ID:policy/$POLICY_NAME \
  --role-name $ROLE_NAME
```

4. Add the role to a specific instance - this step need to be repeated for every new instance:

```
$ aws rds add-role-to-db-instance \
  --db-instance-identifier $RDS_INSTANCE_NAME \
  --feature-name s3Import \
  --role-arn arn:aws:iam::$AWS_ACCOUNT_ID:role/$ROLE_NAME \
```

--region \$REGION

5. Create the VPC endpoint for the S3 service:

```
$ aws ec2 create-vpc-endpoint \
  --vpc-id $VPC_ID \
  --service-name com.amazonaws.$REGION.s3
  --route-table-ids $ROUTE_TABLE_ID
```

The route table id related to the VPC where the endpoint is created can be retrieved through the command

```
$ aws ec2 describe-route-tables | jq -r '.RouteTables[] | "\(.VpcId) \
(.RouteTableId)"'
```

Share Follow

answered Jan 22, 2020 at 17:08



Giuseppe Broccolo

415 4 8

2 OMG. Thank you very much for posting this answer. – Joe Harris Mar 3, 2020 at 0:54



10



Currently (2020-10-16) at least with RDS Postgres 12.4, importing files from S3 does not work if the file is not in root of S3 bucket.

Importing file myfile.csv works, importing file mydir/subdir/myfile.csv doesn't work. The latter will give these kinds of errors if permissions and everything else is working:

```
[XX000] ERROR: HTTP 403. Permission denied. Check bucket or provided
credentials as they may no longer be valid.
```

This is known issue to AWS and according to AWS Support they are working on the issue.

Also it seems that you need to reboot the RDS instance after adding the roles described in [documentation](#). Otherwise the roles do not have any effect. The need for reboot is not mentioned in the documentation.

In general RDS s3Import gives very confusing error messages. For example, importing 0 length file, gives this error:

```
[Amazon](500310) Invalid operation: HTTP 416. Check your arguments and try
again.;
```

Share Follow

edited Oct 16, 2020 at 11:19

answered Oct 16, 2020 at 10:38



Juha Syrjälä

33k 31 130 183

I had exactly the same problem. It worked when I moved the file to the root of the bucket. What do you mean by it is a known issue to AWS? Is there a ticket, forum, page, or anything showing they are working on a solution? – Fábio Paiva Jun 24, 2021 at 12:44

To reproduce your situation, I did the following:

5

- Launched an Amazon RDS PostgreSQL instance in a **public subnet**
- Under **Manage IAM Roles**, I assigned `AmazonRDSServiceRolePolicy` for `s3import`
- Created a table
- Put a CSV file in S3
- Used the `SELECT aws_s3.table_import_from_s3()` command (as above) to load the data

It worked fine for me.

Given that your error message relates to a pre-signed URL, it suggests that the credentials you provided did not have permission to access the CSV file in S3. However, you then say that you used those credentials successfully to retrieve the object. So, this is unlikely to be the cause.

Based on [Reddit: Having issue with AWS RDS Postgres 11+ import from S3 using RDS s3Import feature : aws](#), the issue might be related to the fact that **the Amazon RDS instance is unable to access Amazon S3**. This could be because it is in a **private subnet** with no NAT Gateway in the VPC. If this is the case, then you could either add a NAT Gateway to provide Internet connectivity or, as mentioned in the link, add a **VPC Endpoint for S3**.

Another comment in that post reported the same problem with **a missing Outbound rule in the Security Group**, which stopped the RDS instance from accessing Amazon S3.

Share Follow

edited Oct 30, 2019 at 6:08

answered Oct 30, 2019 at 5:37



John Rotenstein
228k 21 352 431

- 1 You've gone above and beyond the call of duty here. Thank you. My RDS instance does have an internet gateway because I can reach it from my laptop on the Postgres port. My S3 bucket is also accessible from the public internet, as I mentioned. My RDS instance is in a security group with an outbound rule of "All traffic" on 0.0.0.0/0. Other thoughts?
– [ccleve](#) Oct 30, 2019 at 16:57
- 1 Perhaps you could try launching a new temporary RDS instance and follow the above steps to see whether it works. Then, try and identify the differences with the existing system. – [John Rotenstein](#) Oct 30, 2019 at 20:26
- 1 @ccleve could you solve it? I am facing the same issue. – [Aakash Basu](#) Nov 12, 2019 at 12:15
- 4 @AakashBasu No, I didn't solve it. Gave up. Life's too short to deal with AWS configuration problems. I elected to use INSERT instead of COPY. When you insert large batches of records in a single statement the performance isn't bad.
– [ccleve](#) Nov 12, 2019 at 16:17
- 1 I'm trying to do that using Python Dataframe, can you suggest how to go about it for a faster solution? The psycopg2 library isn't performant enough for bigger data. – [Aakash Basu](#) Nov 13, 2019 at 6:37

3 ▲ Check your RDS and S3 are at the same region. I had same problem and fixed it using a bucket in the same region my Aurora RDS.

▼ Share Follow

answered Feb 22, 2020 at 8:49



[morci7](#)

31 2



2 ▲ I solved the same problem when deployed production clusters under private subnets.

▼ Please check your cluster **security group's outbound** (my case)

And also add rds-import-role to "Manage IAM roles" with select feature is s3import



I hope it helps.



Share Follow

answered Oct 31, 2019 at 6:41



[namiops](#)

51 3

1 I do not see an rds-import-role in Manage IAM roles. Also, my outbound rule is All Traffic/All/All/0.0.0.0/0. Is that correct?
– [ccleve](#) Oct 31, 2019 at 17:41

1 Inbound should not allow all traffic to your RDS, but only your services IP Whitelist. Outbound, whatever your requests from RDS go out, normally RDS or any Database have no business to send the request out. But this aws_s3 plug-in need send a request out. rds_import_role you have to create yourself in IAM. Please follow this link to create it, and don't forget to assign a policy access to your S3 bucket for that rds_import_roe
[docs.aws.amazon.com/AmazonRDS/latest/UserGuide/...](https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/) – [namiops](#) Nov 4, 2019 at 10:31



I had the same issue.

1

ERROR: Unable to generate pre-signed url, look at engine log for details



that issue was related to error:



:LOG: S3 bucket names with a period (.) are not supported

The root cause of the issue, in my case, was . (dot) in bucket's name.

Share Follow

answered Oct 29, 2020 at 15:30



[Vladyslav Diachenko](#)

605 4 9

Weird because the documentation doesn't prohibit dot in the name:

[docs.aws.amazon.com/AmazonS3/latest/userguide/...](https://docs.aws.amazon.com/AmazonS3/latest/userguide/) – [thoroc](#) Jul 19, 2022 at 16:25

Yes, but as you can see, error message says opposite)) – [Vladyslav Diachenko](#) Jul 20, 2022 at 17:14



0

One cause of this issue, assuming you've otherwise configured everything correctly, may be Network ACLs. Make sure there are NACL rules explicitly allowing connectivity between the DB and S3.



In our case, we used the `terraform-aws-modules/vpc/aws` module for VPC configuration, which provides for broad VPC configuration, including NACL. After manually inspecting the resulting configuration in AWS Console, we found that the rules generated by the configuration provided to the module were too restrictive. This wasn't reflected in any error messages, as the requests from RDS to S3 were simply blocked at the network level.



Share Follow

answered Nov 23, 2021 at 22:05



[wrsenn](#)

23 4



Had a similar problem, Postgres in a private subnet

0

Solved it with a VPC endpoint for S3



Share Follow

answered Aug 20, 2022 at 5:21



[primordius](#)

11 1 2

