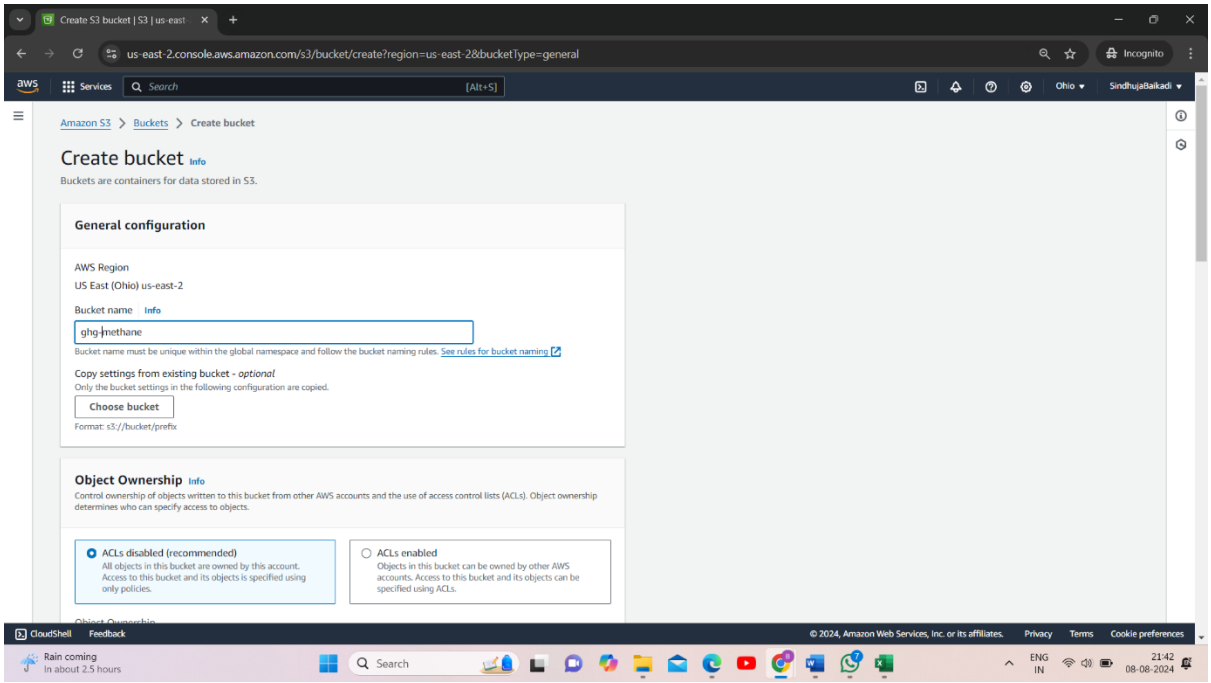
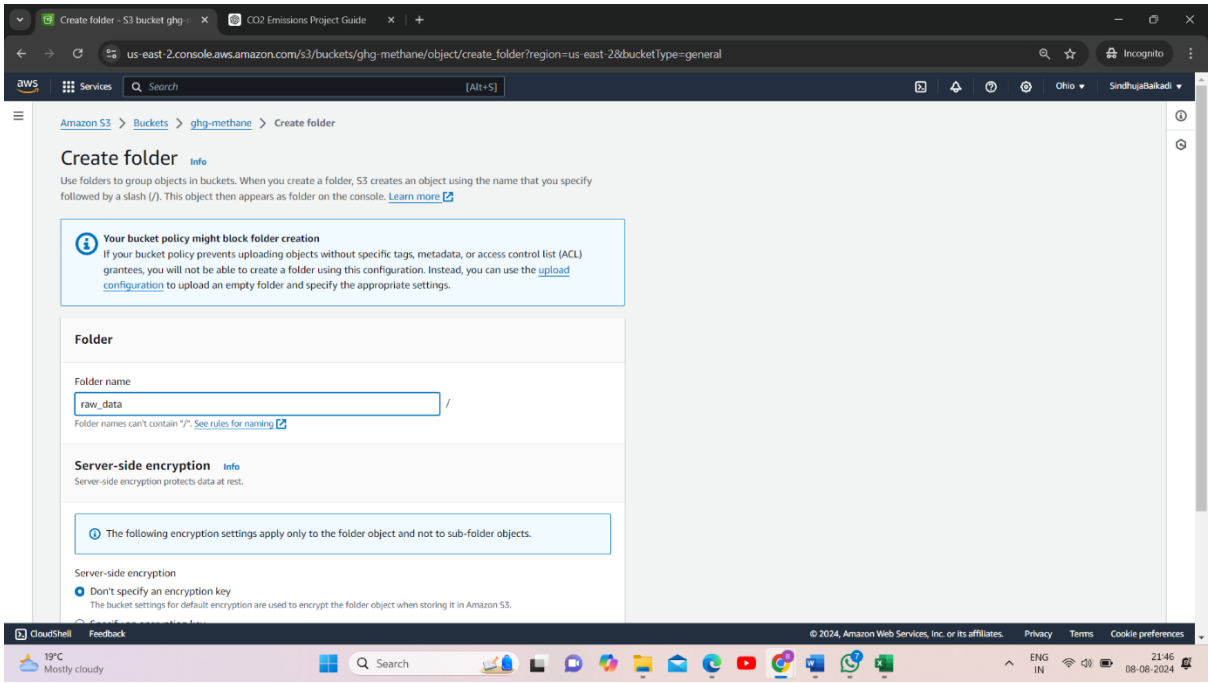
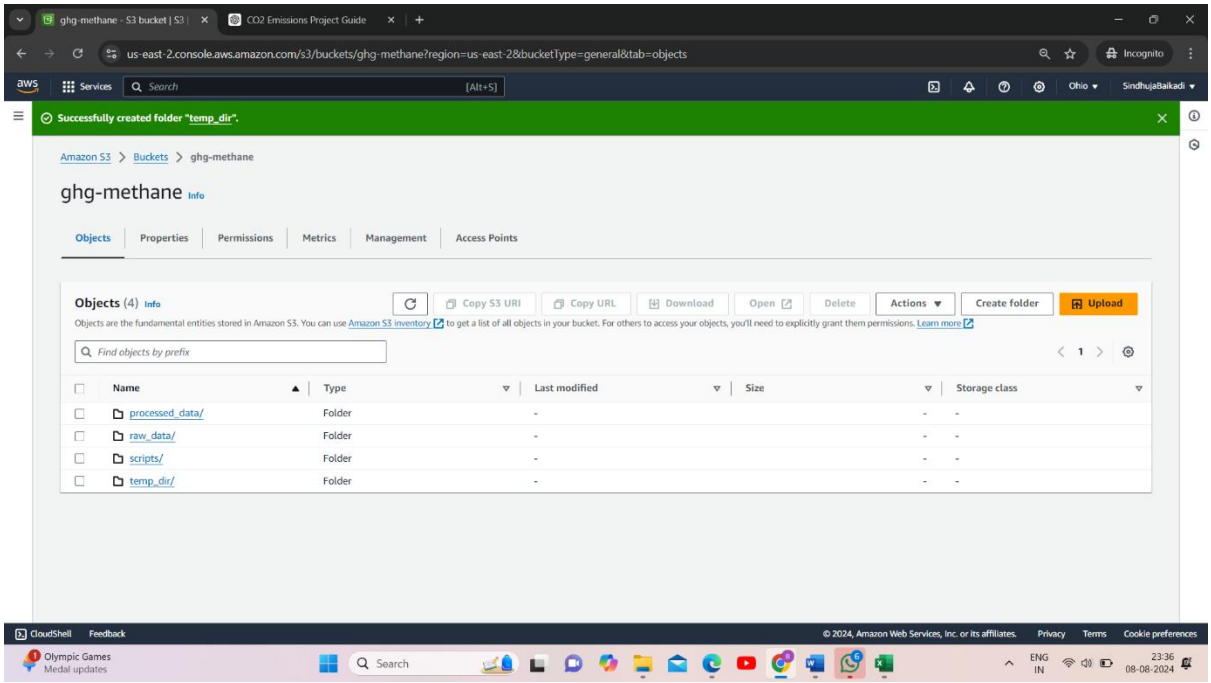

Create S3 bucket | S3 | us-east
us-east-2.console.aws.amazon.com/s3/bucket/create?region=us-east-2&bucketType=general
Incognito
aws
Services
Search
[Alt+S]
Ohio
SindhujaBaikadi
Amazon S3 > Buckets > Create bucket
Create bucket Info
Buckets are containers for data stored in S3.
General configuration
AWS Region
US East (Ohio) us-east-2
Bucket name Info
ghg-methane
Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming
Copy settings from existing bucket - optional
Only the bucket settings in the following configuration are copied.
Choose bucket
Format: s3://bucket/prefix
Object Ownership Info
Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.
ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.
ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Rain coming
In about 2.5 hours
Search
ENG
IN
21:42
08-08-2024


Create folder - S3 bucket ghg
CO2 Emissions Project Guide
us-east-2.console.aws.amazon.com/s3/buckets/ghg-methane/object/create_folder?region=us-east-2&bucketType=general
Incognito
aws
Services
Search
[Alt+S]
Ohio
SindhujaBaikadi
Amazon S3 > Buckets > ghg-methane > Create folder
Create folder Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
raw_data
/
Folder names can't contain "/". See rules for naming
Server-side encryption Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
19°C
Mostly cloudy
Search
ENG
IN
21:46
08-08-2024


ghg-methane - S3 bucket | S3
CO2 Emissions Project Guide
us-east-2.console.aws.amazon.com/s3/buckets/ghg-methane?region=us-east-2&bucketType=general&tab=objects
Successfully created folder "temp_dir".
Amazon S3 > Buckets > ghg-methane
ghg-methane Info
Objects Properties Permissions Metrics Management Access Points
Objects (4) Info
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
Find objects by prefix
Name Type Last modified Size Storage class
processed_data/ Folder
raw_data/ Folder
scripts/ Folder
temp_dir/ Folder
CloudShell Feedback
© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

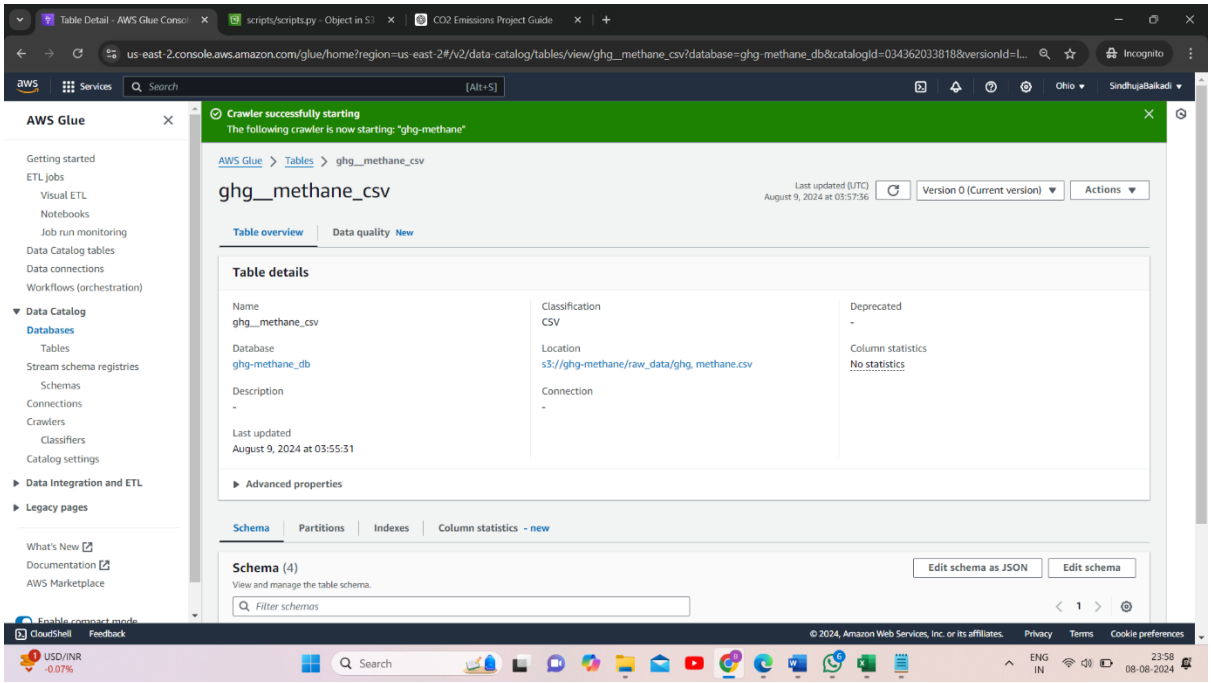
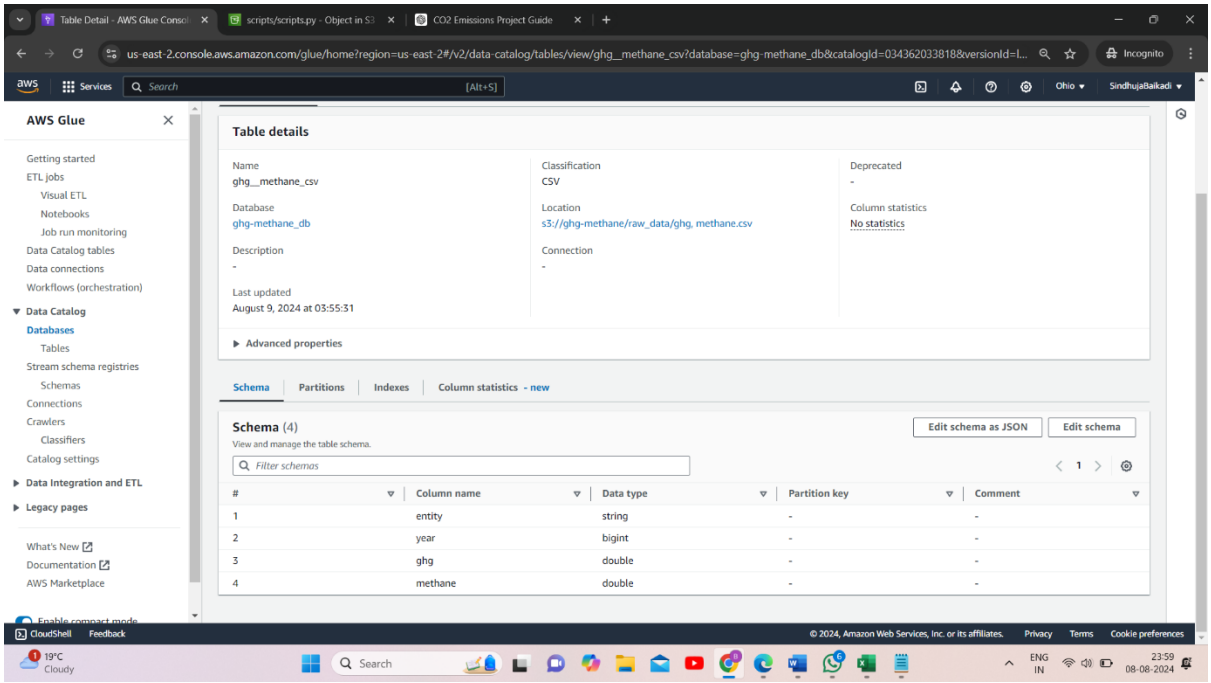
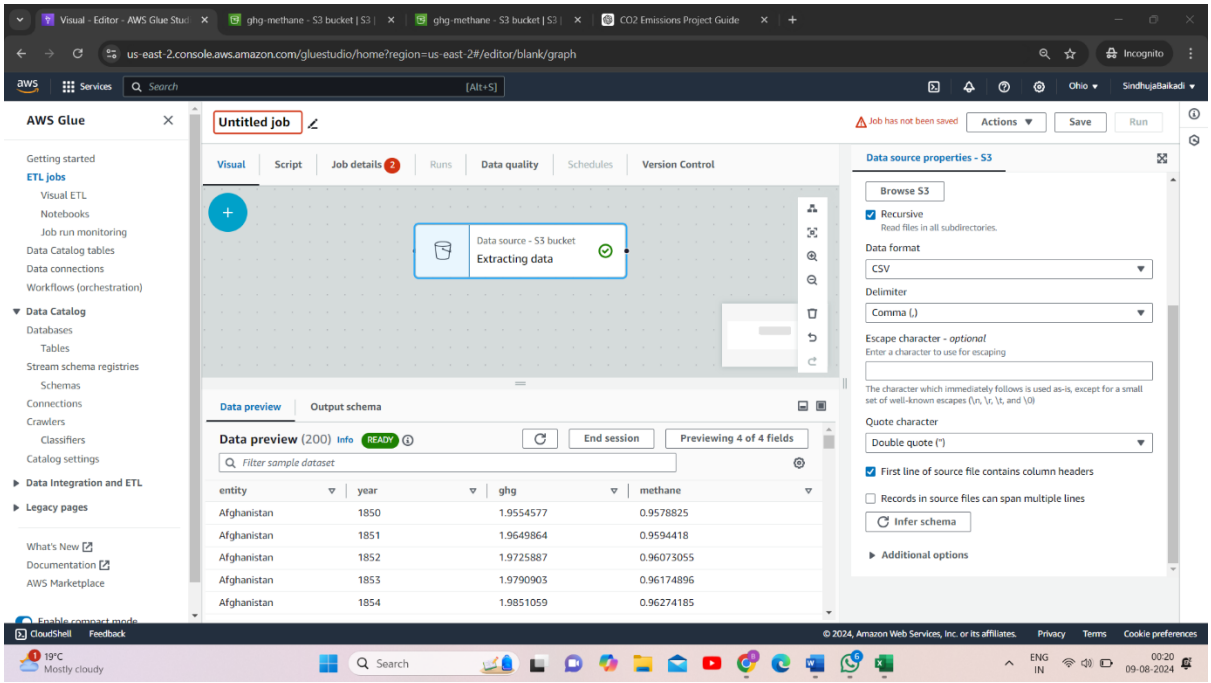
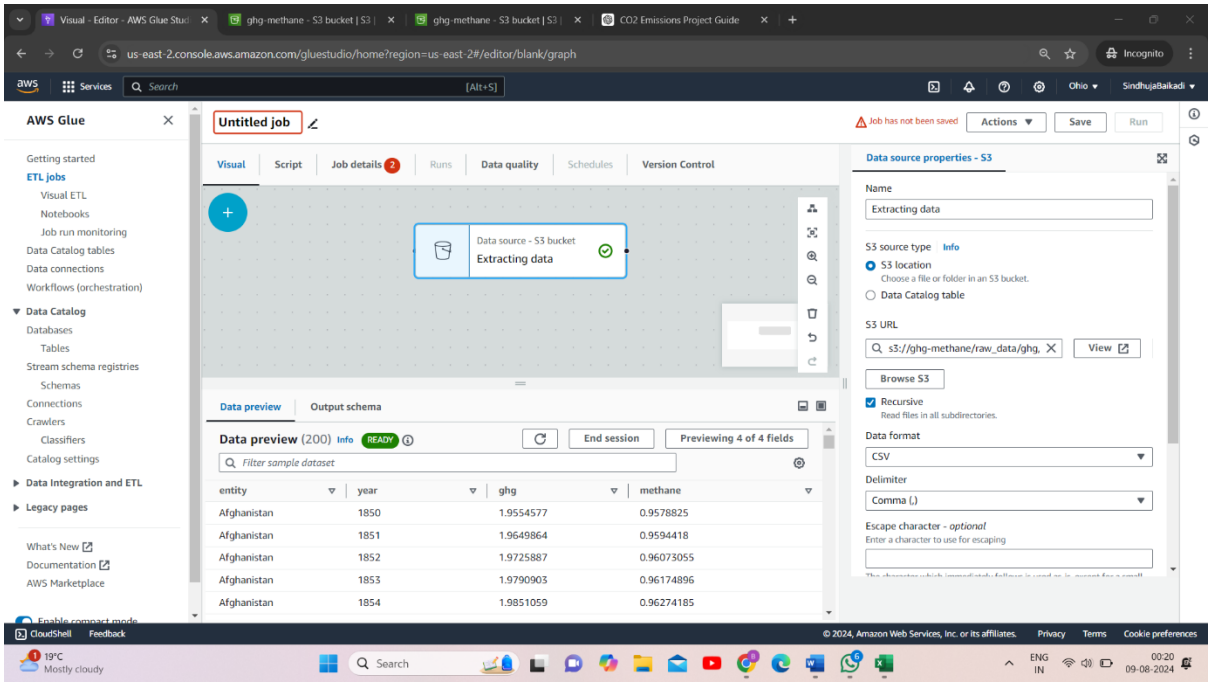
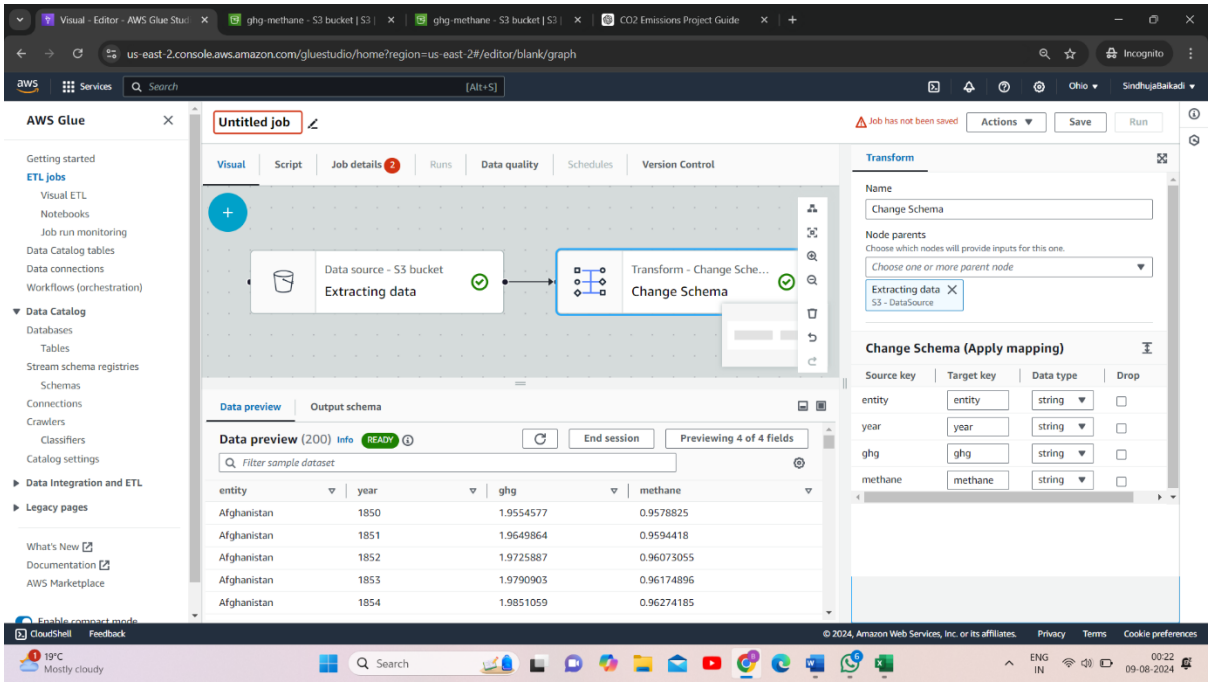


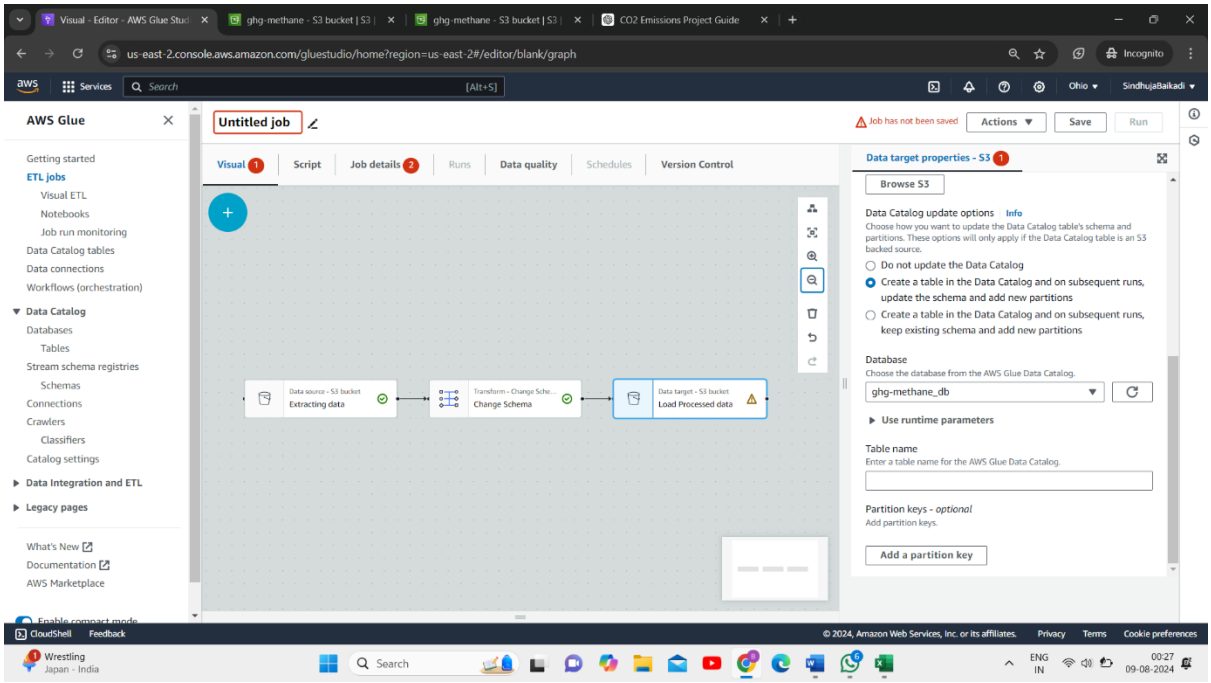
Table Detail - AWS Glue Consol
scripts/scripts.py - Object in S3
CO2 Emissions Project Guide
us-east-2.console.aws.amazon.com/glue/home?region=us-east-2#/v2/data-catalog/tables/view/ghg__methane_csv?database=ghg-methane_db&catalogId=034362033818&versionId=l...
AWS Glue
Crawler successfully starting
The following crawler is now starting: "ghg-methane"
AWS Glue > Tables > ghg__methane_csv
ghg__methane_csv
Last updated (UTC) August 9, 2024 at 03:57:36
Version 0 (Current version) Actions
Table overview Data quality New
Table details
Name ghg__methane_csv
Classification CSV
Deprecated -
Database ghg-methane_db
Location s3://ghg-methane/raw_data/ghg_methane.csv
Column statistics No statistics
Description -
Connection -
Last updated August 9, 2024 at 03:55:31
Advanced properties
Schema Partitions Indexes Column statistics - new
Schema (4)
View and manage the table schema.
Edit schema as JSON Edit schema
Filter schemas
CloudShell Feedback
© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

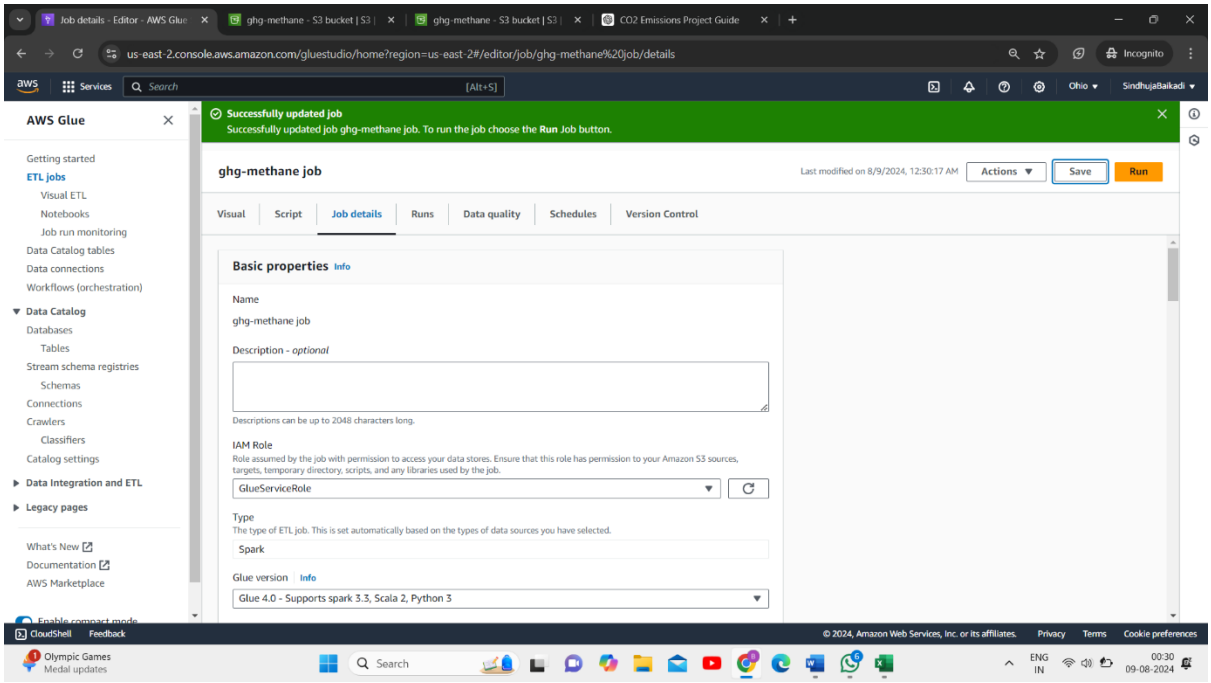

AWS Glue
Table details
Name
ghg__methane_csv
Database
ghg-methane_db
Description
-
Last updated
August 9, 2024 at 03:55:31
Classification
CSV
Location
s3://ghg-methane/raw_data/ghg_ methane.csv
Connection
-
Deprecated
-
Column statistics
No statistics
Advanced properties
Schema
Partitions
Indexes
Column statistics - new
Schema (4)
View and manage the table schema.
Edit schema as JSON
Edit schema
Filter schemas
| Column name | Data type | Partition key | Comment
1 | entity | string | - | -
2 | year | bigint | - | -
3 | ghg | double | - | -
4 | methane | double | - | -

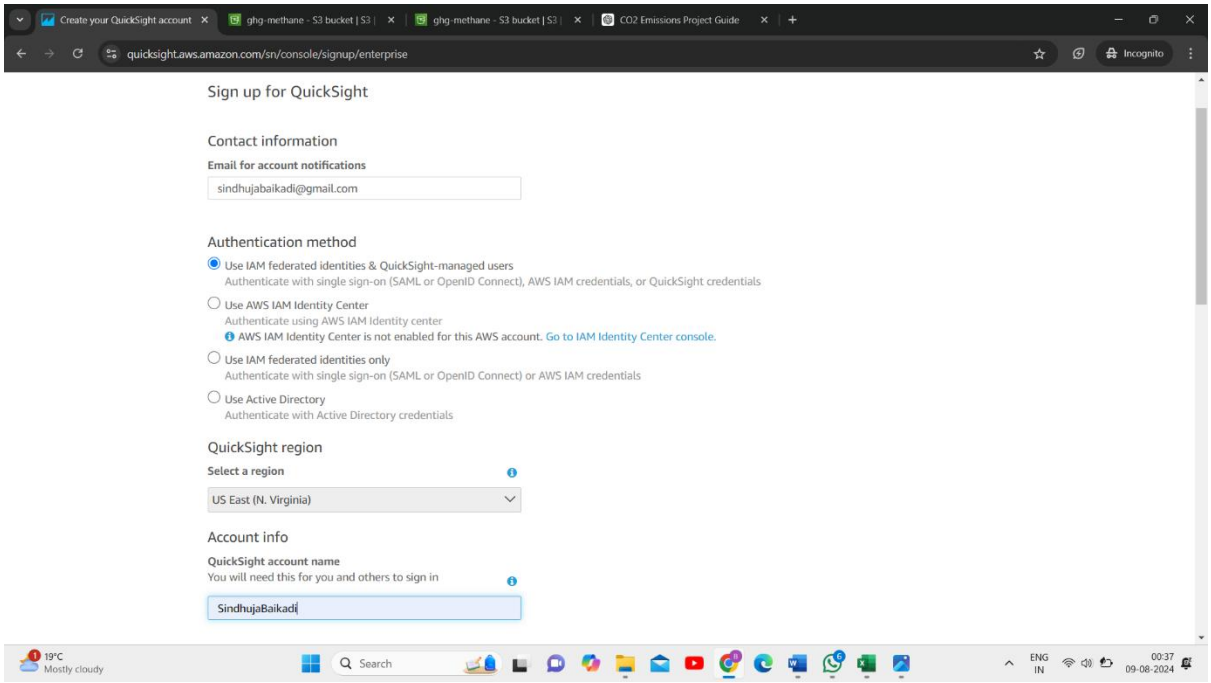

AWS Glue
Untitled job
Job has not been saved
Actions
Save
Run
Visual
Script
Job details 2
Runs
Data quality
Schedules
Version Control
Data source - S3 bucket
Extracting data
Data source properties - S3
Browse S3
Recursive
Read files in all subdirectories.
Data format
CSV
Delimiter
Comma (,)
Escape character - optional
Enter a character to use for escaping
Quote character
Double quote (")
First line of source file contains column headers
Records in source files can span multiple lines
Infer schema
Additional options
Data preview
Output schema
Data preview (200) Info
READY
End session
Previewing 4 of 4 fields
Filter sample dataset
entity | year | ghg | methane
Afghanistan | 1850 | 1.9554577 | 0.9578825
Afghanistan | 1851 | 1.9649864 | 0.9594418
Afghanistan | 1852 | 1.9725887 | 0.96073055
Afghanistan | 1853 | 1.9790903 | 0.96174896
Afghanistan | 1854 | 1.9851059 | 0.96274185

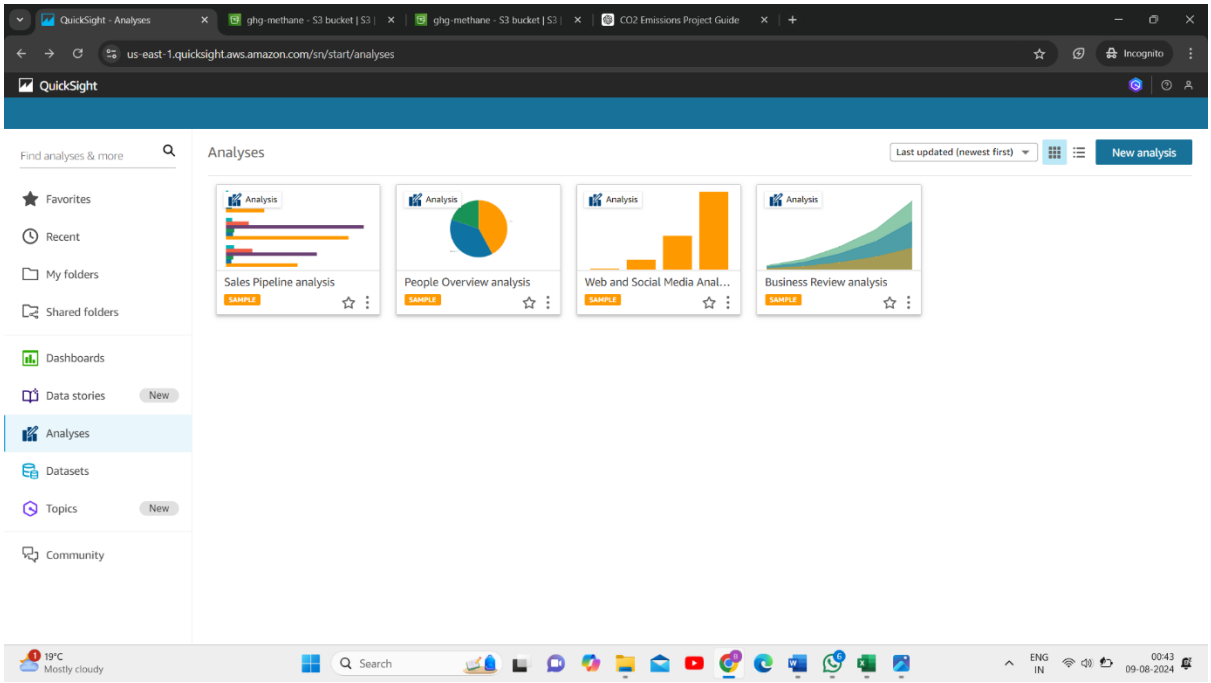

AWS Glue
Untitled job
Job has not been saved
Actions
Save
Run
Visual
Script
Job details 2
Runs
Data quality
Schedules
Version Control
Data source - S3 bucket
Extracting data
Data source properties - S3
Name
Extracting data
S3 source type Info
S3 location
Choose a file or folder in an S3 bucket.
Data Catalog table
S3 URL
s3://ghg-methane/raw_data/ghg,
View
Browse S3
Recursive
Read files in all subdirectories.
Data format
CSV
Delimiter
Comma (,)
Escape character - optional
Enter a character to use for escaping
Data preview
Output schema
Data preview (200) Info
READY
End session
Previewing 4 of 4 fields
Filter sample dataset
entity
year
ghg
methane
Afghanistan 1850 1.9554577 0.9578825
Afghanistan 1851 1.9649864 0.9594418
Afghanistan 1852 1.9725887 0.96073055
Afghanistan 1853 1.9790903 0.96174896
Afghanistan 1854 1.9851059 0.96274185
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

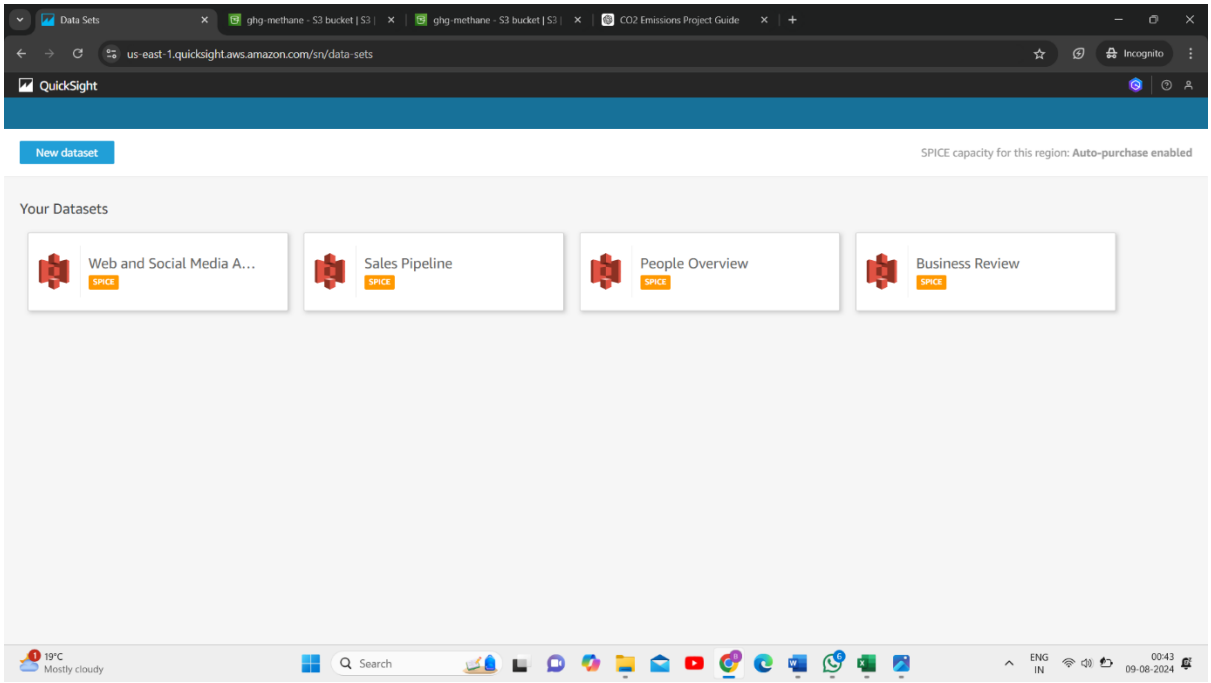

AWS Glue
Untitled job
Job has not been saved
Actions
Save
Run
Visual
Script
Job details 2
Runs
Data quality
Schedules
Version Control
Data source - S3 bucket
Extracting data
Transform - Change Sche...
Change Schema
Transform
Name
Change Schema
Node parents
Choose one or more parent node
Extracting data
S3 - DataSource
Change Schema (Apply mapping)
Source key
Target key
Data type
Drop
entity entity string
year year string
ghg ghg string
methane methane string
Data preview
Output schema
Data preview (200) Info
READY
End session
Previewing 4 of 4 fields
Filter sample dataset
entity
year
ghg
methane
Afghanistan 1850 1.9554577 0.9578825
Afghanistan 1851 1.9649864 0.9594418
Afghanistan 1852 1.9725887 0.96073055
Afghanistan 1853 1.9790903 0.96174896
Afghanistan 1854 1.9851059 0.96274185
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

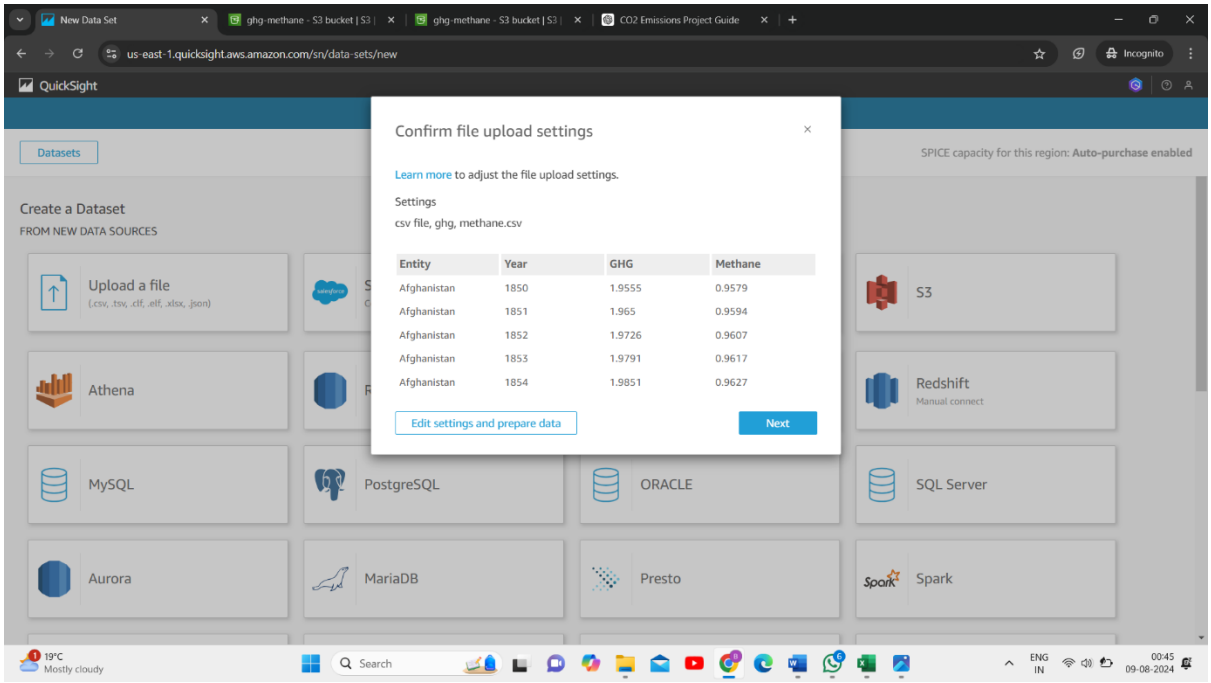

Visual - Editor - AWS Glue Stud
ghg-methane - S3 bucket | S3 |
ghg-methane - S3 bucket | S3 |
CO2 Emissions Project Guide
us-east-2.console.aws.amazon.com/gluestudio/home?region=us-east-2#/editor/blank/graph
Incognito
Services
Search
[Alt+S]
Ohio
SindhujaBaikadi
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
AWS Marketplace
Untitled job
Job has not been saved
Actions
Save
Run
Visual 1
Script
Job details 2
Runs
Data quality
Schedules
Version Control
Data source - S3 bucket
Extracting data
Transform - Change Sche...
Change Schema
Data target - S3 bucket
Load Processed data
Data target properties - S3 1
Browse S3
Data Catalog update options Info
Choose how you want to update the Data Catalog table's schema and partitions. These options will only apply if the Data Catalog table is an S3 backed source.
Do not update the Data Catalog
Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions
Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions
Database
Choose the database from the AWS Glue Data Catalog.
ghg-methane_db
Use runtime parameters
Table name
Enter a table name for the AWS Glue Data Catalog.
Partition keys - optional
Add partition keys.
Add a partition key
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Wrestling
Japan - India
Search
ENG
IN
00:27
09-08-2024

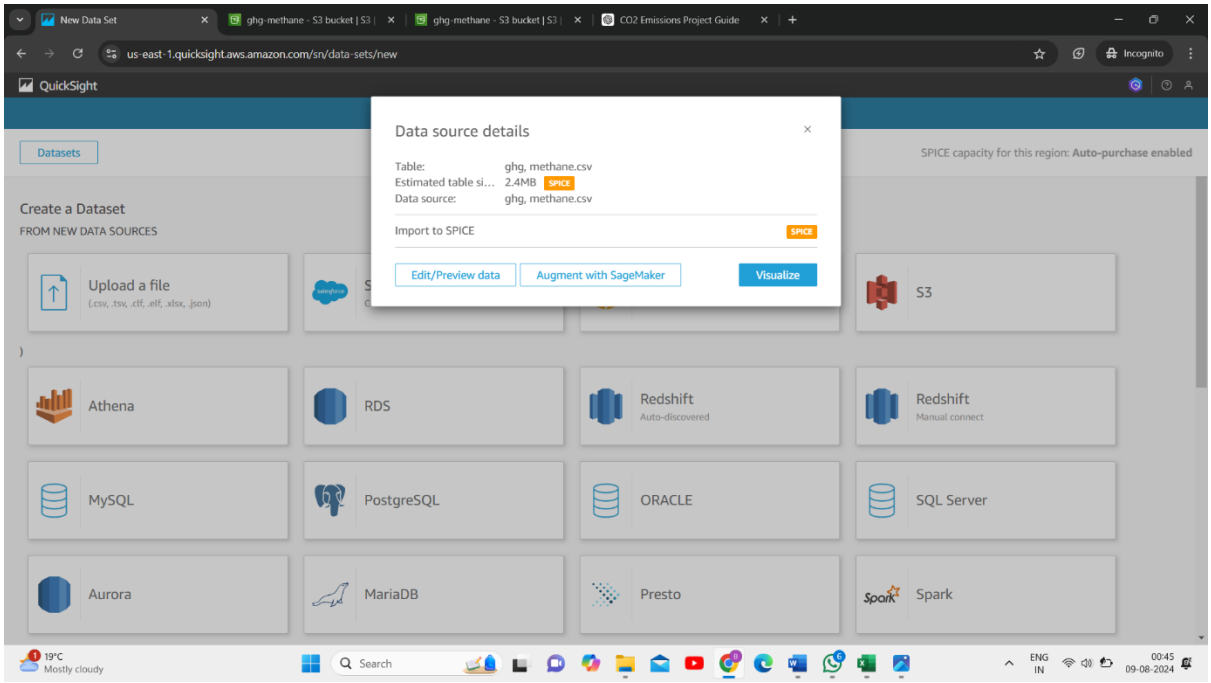

Job details - Editor - AWS Glue
ghg-methane - S3 bucket | S3 |
ghg-methane - S3 bucket | S3 |
CO2 Emissions Project Guide
us-east-2.console.aws.amazon.com/gluestudio/home?region=us-east-2#/editor/job/ghg-methane%20job/details
Incognito
Services
Search
[Alt+S]
Ohio
SindhujaBaikadi
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
AWS Marketplace
Successfully updated job
Successfully updated job ghg-methane job. To run the job choose the Run Job button.
ghg-methane job
Last modified on 8/9/2024, 12:30:17 AM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Basic properties Info
Name
ghg-methane job
Description - optional
Descriptions can be up to 2048 characters long.
IAM Role
Role assumed by the job with permission to access your data stores. Ensure that this role has permission to your Amazon S3 sources, targets, temporary directory, scripts, and any libraries used by the job.
GlueServiceRole
Type
The type of ETL job. This is set automatically based on the types of data sources you have selected.
Spark
Glue version Info
Glue 4.0 - Supports spark 3.3, Scala 2, Python 3
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Olympic Games
Medal updates
Search
ENG
IN
00:30
09-08-2024


Create your QuickSight account
ghg-methane - S3 bucket | S3
ghg-methane - S3 bucket | S3
CO2 Emissions Project Guide
quicksight.aws.amazon.com/sn/console/signup/enterprise
Incognito
Sign up for QuickSight
Contact information
Email for account notifications
sindhujabaikadi@gmail.com
Authentication method
Use IAM federated identities & QuickSight-managed users
Authenticate with single sign-on (SAML or OpenID Connect), AWS IAM credentials, or QuickSight credentials
Use AWS IAM Identity Center
Authenticate using AWS IAM Identity center
AWS IAM Identity Center is not enabled for this AWS account. Go to IAM Identity Center console.
Use IAM federated identities only
Authenticate with single sign-on (SAML or OpenID Connect) or AWS IAM credentials
Use Active Directory
Authenticate with Active Directory credentials
QuickSight region
Select a region
US East (N. Virginia)
Account info
QuickSight account name
You will need this for you and others to sign in
SindhujaBaikadi
19°C
Mostly cloudy
Search
ENG
IN
00:37
09-08-2024


QuickSight - Analyses
ghg-methane - S3 bucket | S3
ghg-methane - S3 bucket | S3
CO2 Emissions Project Guide
us-east-1.quicksight.aws.amazon.com/sn/start/analyses
Incognito
QuickSight
Find analyses & more
Favorites
Recent
My folders
Shared folders
Dashboards
Data stories
New
Analyses
Datasets
Topics
New
Community
Analyses
Last updated (newest first)
New analysis
Analysis
Sales Pipeline analysis
SAMPLE
People Overview analysis
SAMPLE
Web and Social Media Anal...
SAMPLE
Business Review analysis
SAMPLE
19°C
Mostly cloudy
Search
ENG
IN
00:43
09-08-2024


Data Sets
ghg-methane - S3 bucket | S3
ghg-methane - S3 bucket | S3
CO2 Emissions Project Guide
us-east-1.quicksight.aws.amazon.com/sn/data-sets
Incognito
QuickSight
New dataset
SPICE capacity for this region: Auto-purchase enabled
Your Datasets
Web and Social Media A...
SPICE
Sales Pipeline
SPICE
People Overview
SPICE
Business Review
SPICE
19°C
Mostly cloudy
Search
ENG
IN
00:43
09-08-2024


New Data Set
ghg-methane - S3 bucket | S3
ghg-methane - S3 bucket | S3
CO2 Emissions Project Guide
us-east-1.quicksight.aws.amazon.com/sn/data-sets/new
Incognito
QuickSight
Datasets
SPICE capacity for this region: Auto-purchase enabled
Create a Dataset
FROM NEW DATA SOURCES
Upload a file
(.csv, .tsv, .clf, .elf, .xlsx, .json)
Athena
MySQL
PostgreSQL
ORACLE
SQL Server
Aurora
MariaDB
Presto
Spark
S3
Redshift
Manual connect
Confirm file upload settings
Learn more to adjust the file upload settings.
Settings
csv file, ghg, methane.csv
Entity Year GHG Methane
Afghanistan 1850 1.9555 0.9579
Afghanistan 1851 1.965 0.9594
Afghanistan 1852 1.9726 0.9607
Afghanistan 1853 1.9791 0.9617
Afghanistan 1854 1.9851 0.9627
Edit settings and prepare data
Next
19°C
Mostly cloudy
Search
ENG
IN
00:45
09-08-2024



Retrieve all columns (limit to the first 5 rows)

SELECT * FROM s3object s LIMIT 5

Retrieve specific columns (Entity and Year)
SELECT s."Entity", s."Year" FROM s3object s LIMIT 5

Filter rows where Year is a specific value (e.g., 2020)
SELECT * FROM s3object s WHERE s."Year" = '2020' LIMIT 5

Count the number of rows in the file
SELECT COUNT(*) FROM s3object s

s3://ghg-methane/raw_data/ghg, methane.csv

s3://ghg-methane/raw_data/

Athena

1. SELECT * FROM "ghg methane db"."raw_data" limit 10;
2. **How many times each country appears in the entity column. (no of entries for each country)**

SELECT s."entity", COUNT(*) as EntityCount

FROM "ghg methane db"."raw_data" s

GROUP BY s."entity"

3. **Showing countries in the database**

WITH DistinctCountries AS (

SELECT DISTINCT s."entity" AS Country

```
    FROM "ghg methane db"."raw_data" s
)
SELECT Country
FROM DistinctCountries
ORDER BY Country;
```

4. **Finding the Max and min GHG Emissions by Entity**

This query finds the maximum GHG emissions value recorded for each entity:

```
SELECT s."entity", MIN(s."GHG") as MIN_GHG, MAX(s."GHG") as MAX_GHG
FROM "ghg methane db"."raw_data" s
GROUP BY s."entity"
ORDER BY s."entity"
```

5. **Finding the Max and min methan Emissions by Entity**

This query finds the maximum GHG emissions value recorded for each entity:

```
SELECT s."entity", MIN(s."Methane") as MIN_GHG, MAX(s."Methane") as MAX_GHG
FROM "ghg methane db"."raw_data" s
GROUP BY s."entity"
ORDER BY s."entity"
```

6. **Find the Minimum GHG and Corresponding Year:**

```
SELECT s."entity",
    MIN(s."GHG") AS MIN_GHG,
    MIN(s."Year") AS YearWithMinGHG
FROM "ghg methane db"."raw_data" s
WHERE s."GHG" = (
   SELECT MIN(s2."GHG")
   FROM "ghg methane db"."raw_data" s2
   WHERE s2."entity" = s."entity"
)
GROUP BY s."entity"
ORDER BY MIN_GHG DESC;
```

7. **Find the Minimum methane and Corresponding Year:**

```
SELECT s."entity",
```

```
    MIN(s."Methane") AS MIN_Methane,
    MIN(s."Year") AS YearWithMinMethane
FROM "ghg methane db"."raw_data" s
WHERE s."Methane" = (
  SELECT MIN(s2."Methane")
  FROM "ghg methane db"."raw_data" s2
  WHERE s2."entity" = s."entity"
)
GROUP BY s."entity"
ORDER BY MIN_Methane DESC;
```

8. **Top 5 Entities by Total Methane Emissions**

This query finds the top 5 entities with the highest total Methane emissions:

```
SELECT s."entity", SUM(s."Methane") as TotalMethane
FROM "ghg methane db"."raw_data" s
GROUP BY s."entity"
ORDER BY TotalMethane DESC
LIMIT 5
```

9. **Distinct Count of Years Recorded for Each Entity which are below 150**

This query counts the number of distinct years available in the dataset for each entity which are below 150

```
SELECT s."entity", COUNT(DISTINCT s."year") as YearCount
FROM "ghg methane db"."raw_data" s
GROUP BY s."entity"
HAVING COUNT(DISTINCT s."year") < 150
ORDER BY YearCount DESC
```