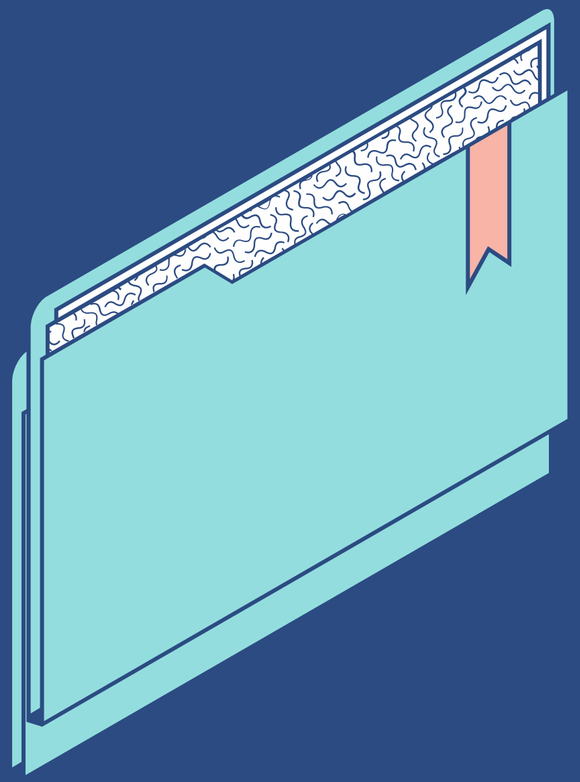
Supercharge DataEngineering with AWS Wrangler

Hands on Python Code to make AWS S3 as your Data Warehouse

Made By

Kamalraj M M



Pre-Requisites

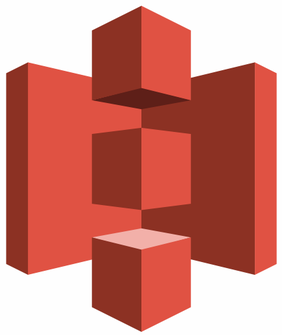
- Active AWS Credentials to create Boto3 sessions in Python.
- Example Files that will be used for explanation.
- Willingness to Learn the Basics

Uploading Files to S3

PSEUDOCODE

Steps To upload the Files

1. Create a Jupyter notebook in the same path where the file to be uploaded is kept.
2. Make a new S3 bucket in AWS console.
3. Use the AWS Wrangler Upload method to upload the file
4. Verify the file has been uploaded to S3 bucket



File Read Methods

0 - Creating and connecting to AWS Session

1 - wr.s3.upload(bucket_location, your_local_file)

2 - Using wr.s3.list_objects(bucket_location)

3 - Using wr.s3.download(bucket_location, bucket_file)

4 - Reading json file using wr.s3.read_json(s3_bucket/folder)

5 - Reading csv file using wr.s3.read_csv(s3_bucket/folder)

6 - Reading parquet file using wr.s3.read_parquet(your_local_file)

File/Data Write Method

0 - Creating and connecting to AWS Session

1 - Writing csv file using wr.s3.to_csv(df, path=s3_destination/file.csv)

2 - Reading csv file using wr.s3.to_parquet(df, path=s3_destination/file.parquet)

3 - Reading fwf file using wr.s3.to_json(df, path=s3_destination/file.json)

4 - Reading fwf file using wr.s3.to_excel(df, path=s3_destination/file.xlsx)

Whats Next

GLUE CATALOG & ATHENA

Working with Glue Catalog & Athena

1. What is Glue Catalog?
2. Reading the tables in Glue Catalog
3. Writing the files to S3, and to creating tables in the Catalog
4. Querying with Athena using AWS Wrangler
5. Executing table joins from AWS Wrangler