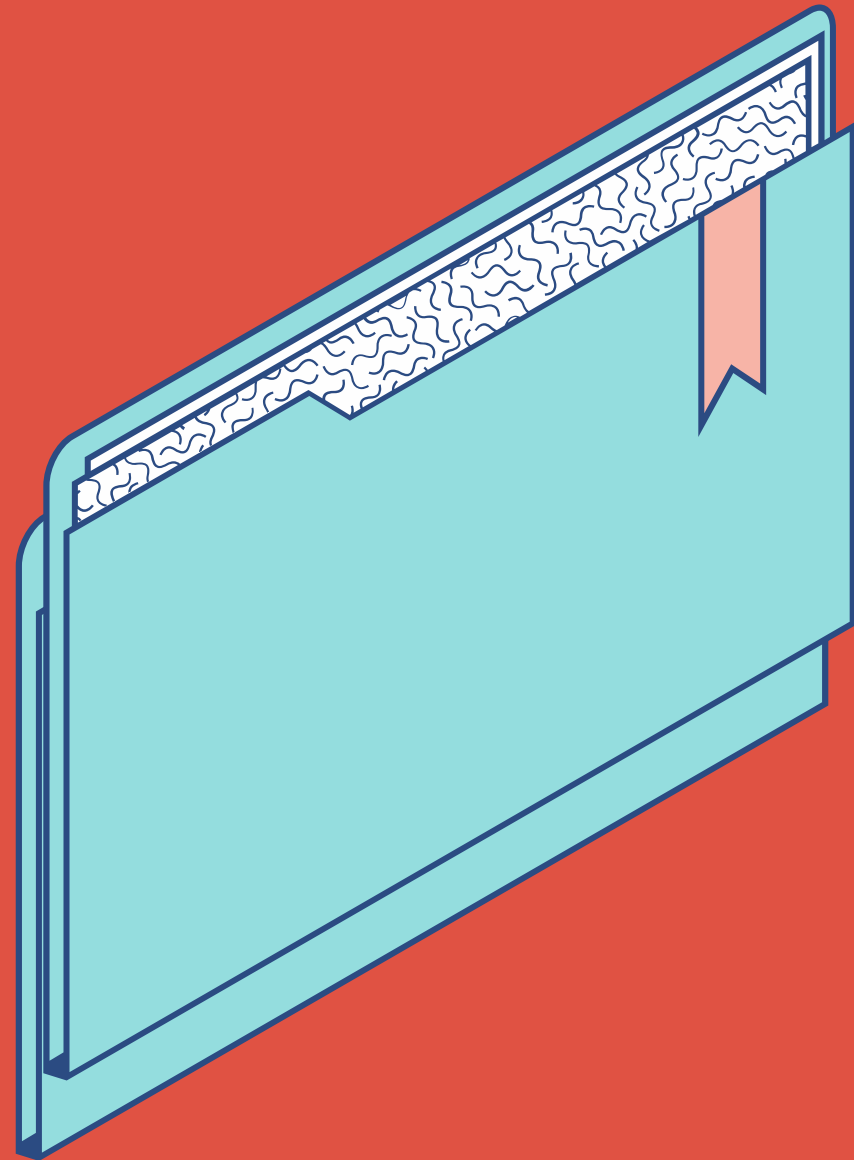




Supercharge Glue - With AWS Wrangler

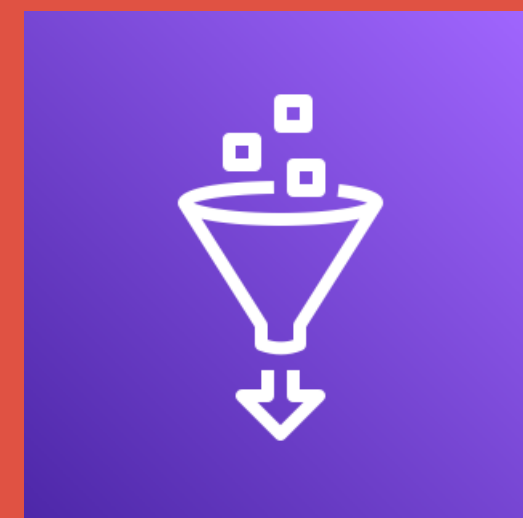
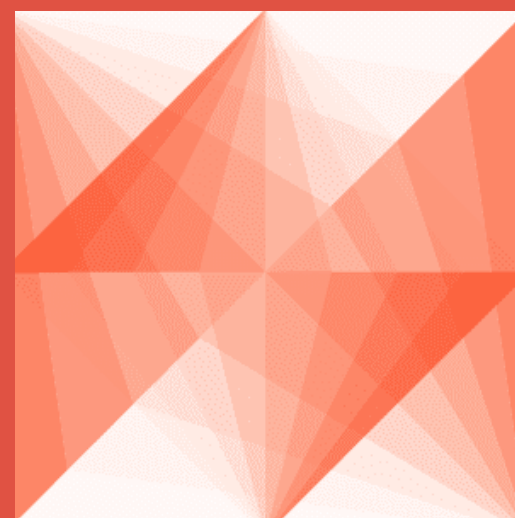
Connecting to Glue using AWS Wrangler &
Boto3

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



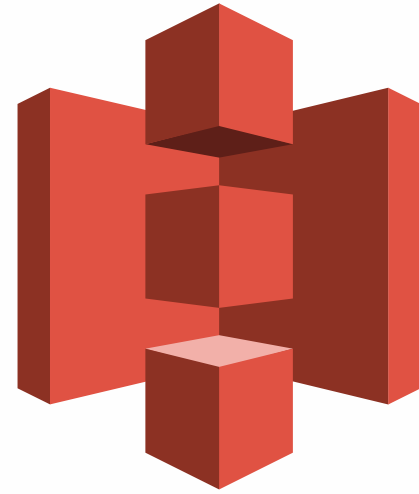
Firing up the Glue Catalog

PSEUDOCODE



Working with Glue Catalog

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



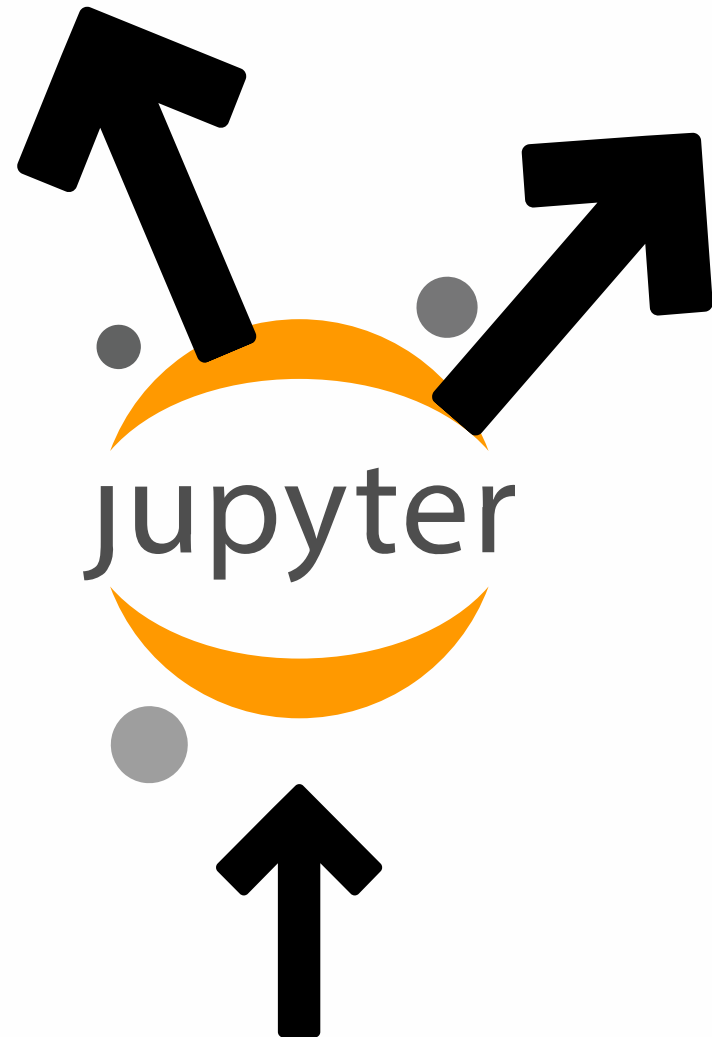
Glue Catalog Methods

0 - Creating and connecting to AWS Session

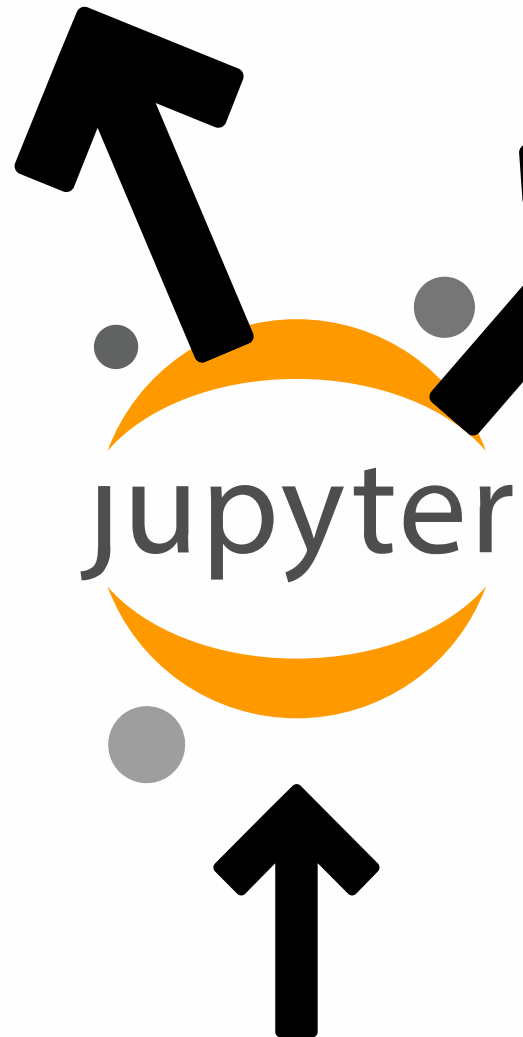
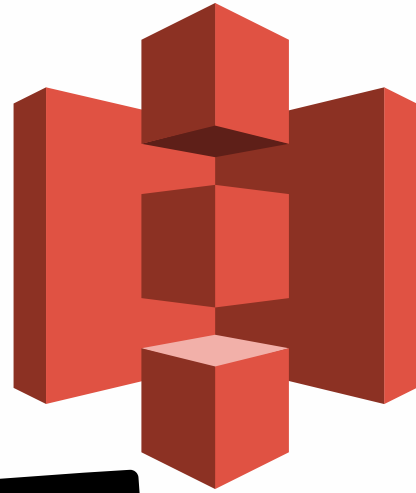
1 - `wr.catalog.databases(boto3_session=your_session)`

2 - `wr.catalog.tables(database='name', boto3_session=your_session)`

3 - Using `wr.catalog.create_database(name='db_name')`



These methods are used for
setting up the database
inside the Glue Catalog



File/Data Write Method - With Glue

The same write method, with addition of database, dataset, and table name

1 - Writing csv file using `wr.s3.to_csv(df, path=s3_destination/csv, dataset=True, database='db_name', table='table_name')`

2 - Writing parquent file using `wr.s3.to_parquet(df, path=s3_destination/parquet, dataset=True, database='db_name', table='table_name')`

3 - WritingJson file using `wr.s3.to_json(df, path=s3_destination/json, dataset=True, database='db_name', table='table_name')`

Note: There is no method to write to excel file or text file with the dataset, database and table parameters

Whats Next

GLUE CATALOG & ATHENA



Working with Athena

1. What is Athena, for Data Engineers?
2. Querying Athena from AWS Wrangler session
3. Executing table joins from AWS Wrangler

