**Lab Exercise- Using AWS Glue Notebooks for AWS Glue Catalog to Amazon S3**

To perform a lab exercise using an AWS Glue Notebook where you read data from the AWS Glue Data Catalog and then write it to an S3 bucket, follow these steps:

**Prerequisites**

* AWS Account: Ensure you have an active AWS account with appropriate permissions.
* AWS Glue Data Catalog: You should have a table or data catalog set up in AWS Glue.
* S3 Bucket: Make sure you have an S3 bucket available where you can write the data.

**Step 1: Set Up AWS Glue Notebook**

* Navigate to AWS Glue Console: Log in to the AWS Management Console, and navigate to the AWS Glue service.
* Create an AWS Glue Notebook: Go to the "Notebooks" section and create a new Glue Notebook.
* Start the Notebook: Once created, start the notebook instance.

**Step 2: Connect to AWS Glue Data Catalog**

Import Necessary Libraries:

import sys

from awsglue.context import GlueContext

from pyspark.context import SparkContext

from awsglue.utils import getResolvedOptions

from pyspark.sql import SQLContext

from awsglue.context import GlueContext from pyspark.context import SparkContext glueContext = GlueContext(SparkContext.getOrCreate()) spark = glueContext.spark\_session

**Read Data from AWS Glue Data Catalog and Write to S3:**

from awsglue.context import GlueContext

from pyspark.context import SparkContext

# Use the existing SparkContext

glueContext = GlueContext(SparkContext.getOrCreate())

spark = glueContext.spark\_session

# Reading from Glue Data Catalog

database\_name = "your\_database\_name"

table\_name = "your\_table\_name"

dyf = glueContext.create\_dynamic\_frame.from\_catalog(

database=database\_name,

table\_name=table\_name

)

# Convert to DataFrame and display

df = dyf.toDF()

df.show()

# Write to S3 Bucket

s3\_output\_path = "s3://your-bucket-name/folder/"

df.write.format("parquet").mode("overwrite").save(s3\_output\_path)

**Step 4: Run the Notebook**

* Run the cells in your Glue Notebook to execute the code. This will read the data from the Glue Data Catalog and write it to the specified S3 bucket.

**Step 5: Verify the Output**

* Navigate to the S3 bucket in the AWS Management Console and verify that the data has been written as expected.