

The purpose of this project is to migrate tables between S3 buckets. We accomplished this with Crawler and Glue.

Uploading the table to an S3 bucket:

H01	Total income	Financial performance	757,504	ANZSIC06 divisions A-S (excluding classes K6330, L6
H04	Sales, government funding, grants and subsidies	Financial performance	674,890	ANZSIC06 divisions A-S (excluding classes K6330, L6
H05	Interest, dividends and donations	Financial performance	49,593	ANZSIC06 divisions A-S (excluding classes K6330, L6
H07	Non-operating income	Financial performance	33,020	ANZSIC06 divisions A-S (excluding classes K6330, L6
H08	Total expenditure	Financial performance	654,404	ANZSIC06 divisions A-S (excluding classes K6330, L6
H09	Interest and donations	Financial performance	26,138	ANZSIC06 divisions A-S (excluding classes K6330, L6
H10	Indirect taxes	Financial performance	6,991	ANZSIC06 divisions A-S (excluding classes K6330, L6
H11	Depreciation	Financial performance	27,801	ANZSIC06 divisions A-S (excluding classes K6330, L6
H12	Salaries and wages paid	Financial performance	123,620	ANZSIC06 divisions A-S (excluding classes K6330, L6
H13	Redundancy and severance	Financial performance	275	ANZSIC06 divisions A-S (excluding classes K6330, L6
H14	Salaries and wages to self employed commission agents	Financial performance	2,085	ANZSIC06 divisions A-S (excluding classes K6330, L6
H19	Purchases and other operating expenses	Financial performance	452,963	ANZSIC06 divisions A-S (excluding classes K6330, L6
H20	Non-operating expenses	Financial performance	14,806	ANZSIC06 divisions A-S (excluding classes K6330, L6
H21	Opening stocks	Financial performance	68,896	ANZSIC06 divisions A-S (excluding classes K6330, L6
H22	Closing stocks	Financial performance	69,127	ANZSIC06 divisions A-S (excluding classes K6330, L6
H23	Surplus before income tax	Financial performance	103,330	ANZSIC06 divisions A-S (excluding classes K6330, L6
H24	Total assets	Financial position	2,512,677	ANZSIC06 divisions A-S (excluding classes K6330, L6
H25	Current assets	Financial position	730,587	ANZSIC06 divisions A-S (excluding classes K6330, L6
H26	Fixed tangible assets	Financial position	591,351	ANZSIC06 divisions A-S (excluding classes K6330, L6
H29	Other assets	Financial position	1,190,739	ANZSIC06 divisions A-S (excluding classes K6330, L6
H30	Total equity and liabilities	Financial position	2,512,677	ANZSIC06 divisions A-S (excluding classes K6330, L6
H31	Shareholders funds or owners equity	Financial position	813,949	ANZSIC06 divisions A-S (excluding classes K6330, L6
H32	Current liabilities	Financial position	933,093	ANZSIC06 divisions A-S (excluding classes K6330, L6

Making a Crawler to create a table in the Glue database:

Review and update

Step 1: Set crawler properties

Edit

Set crawler properties

Name	Description	Tags
annual_reports_crawler	-	-

Step 2: Choose data sources and classifiers

Edit

Data sources (1) [Info](#)
The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
S3	s3://aws-glue-demo-tutorial123/data-store/annual_reports/	Recrawl all

Step 3: Configure security settings

Edit

Configure security settings

IAM role	Security configuration	Lake Formation configuration
AWS-DEMO-GLUE	-	-

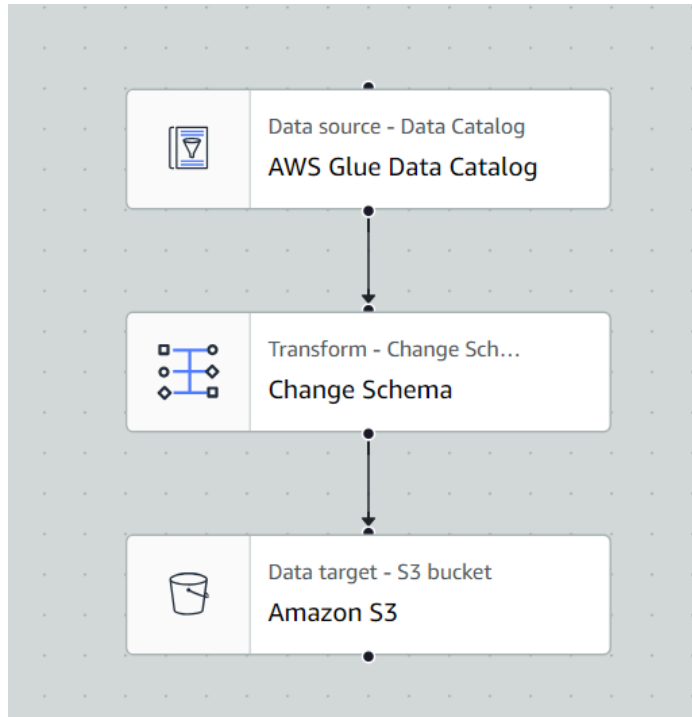
Step 4: Set output and scheduling

Edit

Set output and scheduling

Database	Table prefix - <i>optional</i>	Maximum table threshold - <i>optional</i>	Schedule
annual_reports_practice	-	-	On demand

Either use the visual ETL or write a script for the Glue job to perform the necessary mappings:



```
import sys
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job

args = getResolvedOptions(sys.argv, ["JOB_NAME"])
sc = SparkContext()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args["JOB_NAME"], args)

# Script generated for node AWS Glue Data Catalog
AWSGlueDataCatalog_node1709107214924 =
glueContext.create_dynamic_frame.from_catalog(
    database="annual_reports_practice",
    table_name="annual_reports",
    transformation_ctx="AWSGlueDataCatalog_node1709107214924",
```

)

Script generated for node Change Schema

```
ChangeSchema_node1709107250066 = ApplyMapping.apply(  
    frame=AWSGlueDataCatalog_node1709107214924,  
    mappings=[  
        ("year", "long", "year", "long"),  
        (  
            "industry_aggregation_nzsic",  
            "string",  
            "industry_aggregation_nzsic",  
            "string",  
        ),  
        ("industry_code_nzsic", "string", "industry_code_nzsic", "string"),  
        ("industry_name_nzsic", "string", "industry_name_nzsic", "string"),  
        ("units", "string", "units", "string"),  
        ("variable_code", "string", "variable_code", "string"),  
        ("variable_name", "string", "variable_name", "string"),  
        ("variable_category", "string", "variable_category", "string"),  
        ("value", "string", "value", "string"),  
        ("industry_code_anzsic06", "string", "industry_code_anzsic06", "string"),  
        ("partition_0", "string", "partition_0", "string"),  
    ],  
    transformation_ctx="ChangeSchema_node1709107250066",  
)
```

Script generated for node Amazon S3

```
AmazonS3_node1709107342754 = glueContext.write_dynamic_frame.from_options(  
    frame=ChangeSchema_node1709107250066,  
    connection_type="s3",  
    format="glueparquet",  
    connection_options={  
        "path": "s3://aws-glue-demo-tutorial123/target-data-store/parquet-reports/",  
        "partitionKeys": [],  
    },  
    format_options={"compression": "snappy"},  
    transformation_ctx="AmazonS3_node1709107342754",  
)
```

job.commit()

Running the Glue job results in the creation of the transformed table in the designated S3 bucket:

H01	Total income	Financial performance	757,504	ANZSIC06
H04	Sales, government funding, grants and subsidies	Financial performance	674,890	ANZSIC06
H05	Interest, dividends and donations	Financial performance	49,593	ANZSIC06
H07	Non-operating income	Financial performance	33,020	ANZSIC06
H08	Total expenditure	Financial performance	654,404	ANZSIC06
H09	Interest and donations	Financial performance	26,138	ANZSIC06
H10	Indirect taxes	Financial performance	6,991	ANZSIC06
H11	Depreciation	Financial performance	27,801	ANZSIC06
H12	Salaries and wages paid	Financial performance	123,620	ANZSIC06
H13	Redundancy and severance	Financial performance	275	ANZSIC06
H14	Salaries and wages to self employed commission agents	Financial performance	2,085	ANZSIC06
H19	Purchases and other operating expenses	Financial performance	452,963	ANZSIC06
H20	Non-operating expenses	Financial performance	14,806	ANZSIC06