ABC file reference (by Bhargav B.)

this is a reference to abc file procedure

```
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
from  pyspark.sql.functions import concat
from  pyspark.sql.functions import lit
from awsglue.dynamicframe import DynamicFrame
sc = SparkContext.getOrCreate()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
dyf_out = glueContext.create_dynamic_frame.from_catalog(
database = "bh_glue_test",
table_name = "abc_out_csv")
df_out1 = dyf_out.toDF()
df_out = df_out1.drop("statusid_dec")
df_out_concat = df_out.withColumn("temp",concat("subnum",concat(lit('-'),"visitid"),concat(lit('-'),"visitseq"))
dyf_ref1 = glueContext.create_dynamic_frame.from_catalog(
database = "bh_glue_test",
table_name = "ref1_csv")
df_ref1 = dyf_ref1.toDF()
df_ref1_concat = df_ref1.withColumn("temp1",concat("subnum",concat(lit('-'),"visitid"),concat(lit('-'),"visitsec
select1 = df_ref1_concat.select("eamnum","temp1")
df_final1 = df_out_concat.join(select1,df_out_concat.temp == select1.temp1,"inner")
df_final_1=df_final1.drop("temp1")
dyf_ref2 = glueContext.create_dynamic_frame.from_catalog(
database = "bh_glue_test",
table_name = "ref2_csv")
df_ref2 = dyf_ref2.toDF()
df_ref2_concat = df_ref2.withColumn("temp1",concat("subnum",concat(lit('-'),"visitid"),concat(lit('-'),"visitsec
select2 = df_ref2_concat.select("aenum","temp1")
df_final2 = df_final_1.join(select2,df_final_1.temp == select2.temp1,"inner")
df_final_2=df_final2.drop("temp1")
dyf_ref3 = glueContext.create_dynamic_frame.from_catalog(
database = "bh_glue_test",
table_name = "ref3_csv")
df_ref3 = dyf_ref3.toDF()
df_ref3_concat = df_ref3.withColumn("temp1",concat("subnum",concat(lit('-'),"visitid"),concat(lit('-'),"visitsec
select3 = df_ref3_concat.select("statusid_dec","temp1")
df_final3 = df_final_2.join(select3,df_final_2.temp == select3.temp1,"inner")
df_final_3_=df_final3.drop("temp1")
df_final_3 = df_final_3_.withColumn("Primary",concat("temp",concat(lit('-'),"aenum")))
DF_Final = df_final_3.select("reldeva_dec","reldevb_dec","reldevc_dec","relproca_dec","relprocb_dec","relprocc_d
DYF_semi_final = DynamicFrame.fromDF(DF_Final, glueContext, "convert")
Dyf_final = DYF_semi_final.repartition(1)
output = glueContext.write_dynamic_frame.from_options(
                                frame = Dyf_final,
                                connection_type="s3",
                                connection_options = {"path": "s3://bh-glue-testing/Final Output/"},
```

```
                                format = "csv",
                                format_options={
                                    "separator": ","
                                    },
                                transformation_ctx = "output")
import boto3
client = boto3.client('s3')
BUCKET_NAME= "bh-glue-testing"
PREFIX="Final Output/"
response = client.list_objects(
    Bucket=BUCKET_NAME,
    Prefix=PREFIX,
)
# print(response)
```