

Requirement

Hi Sudarshan,

As discussed today. In the current extract, we facing some issue in the target file. The NULL data from the source is showing as "" in the target file as given in the screenshot below. Please find the screenshots below and go-through the code and let me know the solution for the same. Please call me if you have any questions. Thanks

Issue detail: NULL values are coming as Empty string with Quotes in the file. The below example is from a CSV file. Check & fix all file format. To re-produce the Issue, Run the extract 3019 in Dev.

```
pri_diag_cd,diag_cd_001,diag_cd_002,diag_cd_003,eacpls_cd,eacpt_cd,eacplm_pri_c  
2930,4019,7291,78079,A,1500.0,00,"",2014-07-14T00:00  
8472,"","","",A,1500.0,00,"",2014-07-14T00:00  
7061,70909,"","",A,1500.0,00,"",2014-07-14T00:00  
30002,3003,"","",A,1500.0,WD,"",2014-07-14T00:00
```

Original Code

```
#display(df)
print(f"file_delivery_format: {file_delivery_format}")
print(f"header_row_ind: {header_row_ind}")
print(f"compress_ind: {compress_ind}")
df_write = df.write.format("csv")
if(file_delivery_format=="| Pipe Delimited"):
    df_write = df_write.option("sep","|")
elif(file_delivery_format=="~ Tilde Delimited"):
    df_write = df_write.option("sep","~")
elif file_delivery_format == "XML":
    df_write = df.write.format("com.databricks.spark.xml") \
        .option("rootTag", "root") \
        .option("rowTag", "row") \
        .option("charset", "UTF-8") \
        .option("version", "1.0")
elif file_delivery_format == "JSON":
    df_write = df.write.format("json")
else:
```

The original code was without this parameter added: `.option("emptyValue", "")`

to prevent pyspark from replacing empty values with an empty string, we can explicitly instruct it to treat empty values as null values instead using the option parameter

Fix I added:

change made

```
print(f"compress_ind: {compress_ind}")
# pyspark replaces empty values with "", to prevent this we can state it explicitly to
# use empty values instead of ""
df_write = df.write.format("csv").option("emptyValue", "")
if(file_delivery_format=="| Pipe Delimited"):
    df_write = df_write.option("sep", "|")
elif(file_delivery_format=="~ Tilde Delimited"):
    df_write = df_write.option("sep", "~")
else:
    df_write = df_write.option("sep", ",")
if(header_row_ind):
    df_write = df_write.option("header", "true")
else:
    df_write = df_write.option("header", "false")

if(compress_ind):
    df_write = df_write.option("codec", "org.apache.hadoop.io.compress.GzipCodec")

df_write = df_write.option("parentProject", project).option("path", destination_uri)
df_write.save()
```