

Task 2

Objective: Assign unique key to each record

Approach 1:

```
from pyspark.sql.functions import monotonically_increasing_id  
df2=df.withColumn("uniqueID",monotonically_increasing_id()).show()
```

Approach 2:

```
rdd1=df.rdd.zipWithIndex()
```

How to convert back to DF?

Approach 3:

```
df.createOrReplaceTempView("stage_tbl")  
df2 = spark.sql("select row_number() over(order by some_column) as id from stage_tbl")  
Can any other function instead of row_number help here?
```

Can you relate this functionality to any business case?