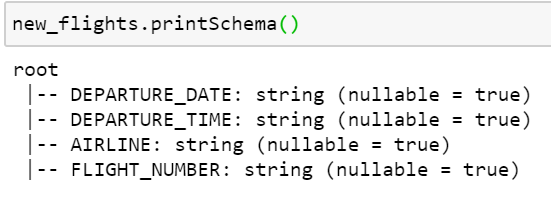
from pyspark.sql import SparkSession

spark = SparkSession.builder.getOrCreate()

flights = spark.read.csv('flights.csv', header=True)

flights.createOrReplaceTempView("tableFlights")

new\_flights = spark.sql("select IFNULL(date\_format(CAST(UNIX\_TIMESTAMP(concat(YEAR, '-', MONTH, '-', DAY), 'yyyy-MM-dd') as TIMESTAMP),'yyyy-MM-dd'), date\_format(to\_timestamp('2015-01-01', 'yyyy-MM-dd'), 'yyyy-MM-dd')) as DEPARTURE\_DATE, IFNULL(date\_format(TO\_TIMESTAMP(CAST(UNIX\_TIMESTAMP(DEPARTURE\_TIME, 'HHmm') AS TIMESTAMP)), 'HH:mm:ss'), date\_format(to\_timestamp('00:00:00', 'HH:mm:ss'), 'HH:mm:ss')) as DEPARTURE\_TIME, AIRLINE, FLIGHT\_NUMBER from tableFlights")



sequence = pyspark.sql.Window.partitionBy("DEPARTURE\_DATE").orderBy("DEPARTURE\_TIME")

new\_flights.withColumn("flight\_serial\_number", pyspark.sql.functions.row\_number().over(sequence)).show(10)

