Notebook to run lake ingestion from File(s3), RDS(MySql), NoSQL(DynamoDB)

This notebook will load lake tables player_info, matches, and deliveries

Assumptions

- **Ideally we should be doing incremental loads fetching new data since the last lake load using watermark, but in this example we will doing full load everytime.**

In [6]:

```
%idle_timeout 20
%glue_version 3.0
%worker_type G.1X
%number_of_workers 2
```

```
Current idle_timeout is 2800 minutes.
idle_timeout has been set to 20 minutes.
Setting Glue version to: 3.0
Previous worker type: G.1X
Setting new worker type to: G.1X
Previous number of workers: 5
Setting new number of workers to: 2
```

In [7]:

```
# Run this cell if you want to write data in delta lake format
# %%configure
# {
#     "--datalake-formats":"delta",
#     "--conf":"spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension --conf spark
.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
# }
```

Run this cell to set up and start your interactive session.

In [4]:

```
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

sc = SparkContext.getOrCreate()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
```

In [5]:

```
import boto3

# Create a Glue client
glue_client = boto3.client('glue')
```

Variables

In [6]:

```
####################
##### Change variables
####################
# Replace with your IAM role ARN to be used by crawler
crawler_role = 'arn:aws:iam::687003041478:role/orka-glue-role'

# Change variable for MySQL deliveries source
host = 'mydbinstance.ctf19flbptnt.us-east-1.rds.amazonaws.com'
port = 3306
user = 'admin'
password = 'MyPassword123'
mssql_database = 'ipl'
mssql_table='deliveries'

# Change variables for s3 player info source
player_info_input_file_s3_path = "s3://iplcricketinfo/input_files/Players_info.csv" # Change to the path where you have your Players_info.csv

# Change variable for dynamodb matches source
macthes_dynamodb_table_name = "MatchesTable"

# DB variables
bronze_database_name = "orka_warehouse_bronze" # Specify the lake/bronze glue catalog database name
bronze_database_loc = "s3://iplcricketinfo/datalake/bronze" # S3 location for bronze table

# Target Player info table
player_info_table_name = "player_info" # Specify player info target table name in lake
# Target Deliveries table
deliveries_table_name = "deliveries"
# Target matches table
matches_table_name = "matches"
```

Utility functions

In [12]:

```
import boto3

# Function to read data from MYSQL database
def read_from_mysql(host, port, user, password, database, table):
    mysql_url = f"jdbc:mysql://{host}:{port}"
    mysql_properties = {
        "user": user,
        "password": password,
        "driver": "com.mysql.jdbc.Driver"
    }

    return spark.read \
        .format("jdbc") \
        .option("url", mysql_url) \
        .option("dbtable", f"{database}.{table}") \
        .options(**mysql_properties) \
        .load()

# Function to crawl data stored in lake
def crawl_tables(glue_client, df_dic, data_loc, database_name):
    for table, data_df in df_dic.items():
        crawler_name = f"{table}_crwl"
        # Specify the S3 target for the crawler
        s3_target_path = f"{data_loc}/{table}"

        # Create the crawler
        response = glue_client.create_crawler(
            Name=crawler_name,
            Role=crawler_role,
            Targets={
                'S3Targets': [
                    {
                        'Path': s3_target_path
                    }
                ]
            },
            DatabaseName= database_name,
            SchemaChangePolicy={
                'UpdateBehavior': 'UPDATE_IN_DATABASE',
                'DeleteBehavior': 'DEPRECATE_IN_DATABASE'
            }
        )

        # Start the crawler
        glue_client.start_crawler(Name=crawler_name)


def read_from_dynamodb(table_name):
    # Read data from the DynamoDB table
    dynamodb_options = {
        "dynamodb.input.tableName": table_name,
        "dynamodb.throughput.read.percent": "1.0"
    }
    dynamic_frame = glueContext.create_dynamic_frame.from_options(
        connection_type="dynamodb",
        connection_options=dynamodb_options
    )

    # Convert the dynamic frame to a Spark DataFrame and return
    return dynamic_frame.toDF()
```

Create database in catalog

In [7]:

```
# Create the database
glue_client.create_database(
    DatabaseInput={
        'Name': database_name
    }
)
```

```
{'ResponseMetadata': {'RequestId': '7c34261c-eced-4bcd-96ed-2191951b32ee', 'HTTPStatusCod
e': 200, 'HTTPHeaders': {'date': 'Sun, 09 Jul 2023 05:57:44 GMT', 'content-type': 'applic
ation/x-amz-json-1.1', 'content-length': '2', 'connection': 'keep-alive', 'x-amzn-request
id': '7c34261c-eced-4bcd-96ed-2191951b32ee'}, 'RetryAttempts': 0}}
```

**

File ingestion example - Loading csv file in s3 to lake/silver layer

**

Ingesting table from s3 into df

In [8]:

```
# Created my own clean header from the csv
player_info_header_clean = "index,id,name,country,full_name,birthdate,birthplace,died,dat
e_of_death,age,major_teams,batting_style,bowling_style,other,awards,batting_tests_mat,bat
ting_tests_inns,batting_tests_no,batting_tests_runs,batting_tests_hs,batting_tests_ave,ba
```

```
tting_tests_bf,batting_tests_sr,batting_tests_100,batting_tests_50,batting_tests_4s,batti
ng_tests_6s,batting_tests_ct,batting_tests_st,batting_odis_mat,batting_odis_inns,batting_
odis_no,batting_odis_runs,batting_odis_hs,batting_odis_ave,batting_odis_bf,batting_odis_s
r,batting_odis_100,batting_odis_50,batting_odis_4s,batting_odis_6s,batting_odis_ct,battin
g_odis_st,batting_t20is_mat,batting_t20is_inns,batting_t20is_no,batting_t20is_runs,battin
g_t20is_hs,batting_t20is_ave,batting_t20is_bf,batting_t20is_sr,batting_t20is_100,batting_
t20is_50,batting_t20is_4s,batting_t20is_6s,batting_t20is_ct,batting_t20is_st,batting_firs
t_class_mat,batting_first_class_inns,batting_first_class_no,batting_first_class_runs,batt
ing_first_class_hs,batting_first_class_ave,batting_first_class_bf,batting_first_class_sr,
batting_first_class_100,batting_first_class_50,batting_first_class_4s,batting_first_class
_6s,batting_first_class_ct,batting_first_class_st,batting_list_a_mat,batting_list_a_inns,
batting_list_a_no,batting_list_a_runs,batting_list_a_hs,batting_list_a_ave,batting_list_a
_bf,batting_list_a_sr,batting_list_a_100,batting_list_a_50,batting_list_a_4s,batting_list
_a_6s,batting_list_a_ct,batting_list_a_st,batting_t20s_mat,batting_t20s_inns,batting_t20s
_no,batting_t20s_runs,batting_t20s_hs,batting_t20s_ave,batting_t20s_bf,batting_t20s_sr,ba
tting_t20s_100,batting_t20s_50,batting_t20s_4s,batting_t20s_6s,batting_t20s_ct,batting_t2
0s_st,bowling_tests_mat,bowling_tests_inns,bowling_tests_balls,bowling_tests_runs,bowling
_tests_wkts,bowling_tests_bbi,bowling_tests_bbm,bowling_tests_ave,bowling_tests_econ,bowl
ing_tests_sr,bowling_tests_4w,bowling_tests_5w,bowling_tests_10,bowling_odis_mat,bowling_
odis_inns,bowling_odis_balls,bowling_odis_runs,bowling_odis_wkts,bowling_odis_bbi,bowling
_odis_bbm,bowling_odis_ave,bowling_odis_econ,bowling_odis_sr,bowling_odis_4w,bowling_odis
_5w,bowling_odis_10,bowling_t20is_mat,bowling_t20is_inns,bowling_t20is_balls,bowling_t20i
s_runs,bowling_t20is_wkts,bowling_t20is_bbi,bowling_t20is_bbm,bowling_t20is_ave,bowling_t
20is_econ,bowling_t20is_sr,bowling_t20is_4w,bowling_t20is_5w,bowling_t20is_10,bowling_fir
st_class_mat,bowling_first_class_inns,bowling_first_class_balls,bowling_first_class_runs,
bowling_first_class_wkts,bowling_first_class_bbi,bowling_first_class_bbm,bowling_first_cl
ass_ave,bowling_first_class_econ,bowling_first_class_sr,bowling_first_class_4w,bowling_fi
rst_class_5w,bowling_first_class_10,bowling_list_a_mat,bowling_list_a_inns,bowling_list_a
_balls,bowling_list_a_runs,bowling_list_a_wkts,bowling_list_a_bbi,bowling_list_a_bbm,bowl
ing_list_a_ave,bowling_list_a_econ,bowling_list_a_sr,bowling_list_a_4w,bowling_list_a_5w,
bowling_list_a_10,bowling_t20s_mat,bowling_t20s_inns,bowling_t20s_balls,bowling_t20s_runs
,bowling_t20s_wkts,bowling_t20s_bbi,bowling_t20s_bbm,bowling_t20s_ave,bowling_t20s_econ,b
owling_t20s_sr,bowling_t20s_4w,bowling_t20s_5w,bowling_t20s_10"

# Read the CSV file and into a DataFrame
player_info_raw = spark.read.csv(player_info_input_file_s3_path, header=True, inferSchema
=True)

# Use the custom header
player_info_raw = player_info_raw.toDF(*player_info_header_clean.split(','))

# Show the dataframe
player_info_raw.show(5)
```

```
+-----+------+----------------+-------+--------------------+----------+---------------
----+-----+-------------+---+--------------------+--------------+-------------------+-
----+------+-----------------+------------------+----------------+-----------------+--
--------------+-----------------+----------------+----------------+----------------+--
--------------+----------------+----------------+----------------+---------------+----
------------+-----------------+---------------+-----------------+---------------+-----
----------+---------------+---------------+----------------+---------------+----------
----+---------------+---------------+---------------+-----------------+---------------
--+----------------+------------------+----------------+-----------------+------------
---+----------------+-----------------+----------------+----------------+-------------
--+----------------+----------------+-----------------------+-----------------------+-
---------------------+------------------------+----------------------+----------------
------+----------------------+----------------------+----------------------+----------
------------+----------------------+----------------------+----------------------+----
-----------------+------------------+-------------------+-----------------+-----------
-------+-----------------+------------------+-----------------+-----------------+-----
------------+-----------------+-----------------+-----------------+----------------+--
---------------+----------------+-----------------+---------------+----------------+--
-------------+----------------+---------------+---------------+----------------+------
--------+---------------+---------------+---------------+---------------+-------------
---+------------------+-------------------+------------------+------------------+-----
-----------+-----------------+-----------------+------------------+---------------+---
-------------+----------------+----------------+----------------+----------------+----
--------------+-----------------+-----------------+----------------+---------------+--
--------------+-----------------+---------------+---------------+---------------+-----
---------+-----------------+------------------+-------------------+-----------------+-
-----------------+-----------------+----------------+------------------+--------------
---+----------------+----------------+----------------+----------------+--------------
```

```
--------+------------------------+-------------------------+------------------------+--
----------------------+-----------------------+-----------------------+----------------
-------+------------------------+----------------------+----------------------+--------
--------------+----------------------+------------------+-------------------+----------
----------+-------------------+-------------------+------------------+-----------------
-+------------------+-------------------+-----------------+-----------------+----------
-------+-----------------+----------------+-----------------+------------------+-------
----------+-----------------+----------------+----------------+----------------+-------
----------+---------------+---------------+---------------+---------------+
|index|    id|            name|country|           full_name| birthdate|          birthpl
ace| died|date_of_death|age|         major_teams| batting_style|       bowling_style|othe
r|awards|batting_tests_mat|batting_tests_inns|batting_tests_no|batting_tests_runs|batting
_tests_hs|batting_tests_ave|batting_tests_bf|batting_tests_sr|batting_tests_100|batting_t
ests_50|batting_tests_4s|batting_tests_6s|batting_tests_ct|batting_tests_st|batting_odis_
mat|batting_odis_inns|batting_odis_no|batting_odis_runs|batting_odis_hs|batting_odis_ave|
batting_odis_bf|batting_odis_sr|batting_odis_100|batting_odis_50|batting_odis_4s|batting_
odis_6s|batting_odis_ct|batting_odis_st|batting_t20is_mat|batting_t20is_inns|batting_t20i
s_no|batting_t20is_runs|batting_t20is_hs|batting_t20is_ave|batting_t20is_bf|batting_t20is
_sr|batting_t20is_100|batting_t20is_50|batting_t20is_4s|batting_t20is_6s|batting_t20is_ct
|batting_t20is_st|batting_first_class_mat|batting_first_class_inns|batting_first_class_no
|batting_first_class_runs|batting_first_class_hs|batting_first_class_ave|batting_first_cl
ass_bf|batting_first_class_sr|batting_first_class_100|batting_first_class_50|batting_firs
t_class_4s|batting_first_class_6s|batting_first_class_ct|batting_first_class_st|batting_l
ist_a_mat|batting_list_a_inns|batting_list_a_no|batting_list_a_runs|batting_list_a_hs|bat
ting_list_a_ave|batting_list_a_bf|batting_list_a_sr|batting_list_a_100|batting_list_a_50|
batting_list_a_4s|batting_list_a_6s|batting_list_a_ct|batting_list_a_st|batting_t20s_mat|
batting_t20s_inns|batting_t20s_no|batting_t20s_runs|batting_t20s_hs|batting_t20s_ave|batt
ing_t20s_bf|batting_t20s_sr|batting_t20s_100|batting_t20s_50|batting_t20s_4s|batting_t20s
_6s|batting_t20s_ct|batting_t20s_st|bowling_tests_mat|bowling_tests_inns|bowling_tests_ba
lls|bowling_tests_runs|bowling_tests_wkts|bowling_tests_bbi|bowling_tests_bbm|bowling_tes
ts_ave|bowling_tests_econ|bowling_tests_sr|bowling_tests_4w|bowling_tests_5w|bowling_test
s_10|bowling_odis_mat|bowling_odis_inns|bowling_odis_balls|bowling_odis_runs|bowling_odis
_wkts|bowling_odis_bbi|bowling_odis_bbm|bowling_odis_ave|bowling_odis_econ|bowling_odis_s
r|bowling_odis_4w|bowling_odis_5w|bowling_odis_10|bowling_t20is_mat|bowling_t20is_inns|bo
wling_t20is_balls|bowling_t20is_runs|bowling_t20is_wkts|bowling_t20is_bbi|bowling_t20is_b
bm|bowling_t20is_ave|bowling_t20is_econ|bowling_t20is_sr|bowling_t20is_4w|bowling_t20is_5
w|bowling_t20is_10|bowling_first_class_mat|bowling_first_class_inns|bowling_first_class_b
alls|bowling_first_class_runs|bowling_first_class_wkts|bowling_first_class_bbi|bowling_fi
rst_class_bbm|bowling_first_class_ave|bowling_first_class_econ|bowling_first_class_sr|bow
ling_first_class_4w|bowling_first_class_5w|bowling_first_class_10|bowling_list_a_mat|bowl
ing_list_a_inns|bowling_list_a_balls|bowling_list_a_runs|bowling_list_a_wkts|bowling_list
_a_bbi|bowling_list_a_bbm|bowling_list_a_ave|bowling_list_a_econ|bowling_list_a_sr|bowlin
g_list_a_4w|bowling_list_a_5w|bowling_list_a_10|bowling_t20s_mat|bowling_t20s_inns|bowlin
g_t20s_balls|bowling_t20s_runs|bowling_t20s_wkts|bowling_t20s_bbi|bowling_t20s_bbm|bowlin
g_t20s_ave|bowling_t20s_econ|bowling_t20s_sr|bowling_t20s_4w|bowling_t20s_5w|bowling_t20s
_10|
+-----+------+----------------+-------+--------------------+----------+----------------
----+-----+-------------+---+--------------------+--------------+--------------------+-
----+------+-----------------+------------------+----------------+------------------+--
--------------+-----------------+----------------+----------------+-----------------+--
--------------+----------------+----------------+----------------+----------------+----
------------+-----------------+---------------+-----------------+---------------+------
----------+---------------+---------------+----------------+---------------+-----------
----+---------------+---------------+---------------+-----------------+----------------
--+----------------+------------------+----------------+-----------------+-------------
---+----------------+-----------------+----------------+----------------+--------------
--+----------------+----------------+-----------------------+------------------------+-
---------------------+------------------------+----------------------+-----------------
------+----------------------+----------------------+-----------------------+----------
------------+----------------------+----------------------+----------------------+-----
-----------------+------------------+-------------------+-----------------+------------
-------+-----------------+------------------+-----------------+-----------------+------
------------+-----------------+-----------------+-----------------+-----------------+--
---------------+----------------+-----------------+---------------+-----------------+--
-------------+----------------+---------------+---------------+----------------+-------
--------+---------------+---------------+---------------+---------------+--------------
---+------------------+-------------------+------------------+------------------+------
-----------+-----------------+-----------------+------------------+----------------+---
-------------+----------------+----------------+----------------+-----------------+----
--------------+-----------------+-----------------+----------------+----------------+--
--------------+-----------------+---------------+---------------+---------------+------
---------+-----------------+------------------+-------------------+------------------+-
```

```
-----------------+-----------------+-----------------+-----------------+---------------
---+----------------+----------------+----------------+----------------+---------------
--------+------------------------+-------------------------+-------------------------+--
----------------------+-----------------------+-----------------------+----------------
-------+------------------------+----------------------+----------------------+--------
--------------+----------------------+------------------+--------------------+---------
----------+-------------------+-------------------+-------------------+----------------
-+------------------+-------------------+-----------------+------------------+---------
-------+-----------------+-----------------+-----------------+------------------+------
----------+------------------+---------------+----------------+-----------------+------
----------+---------------+----------------+---------------+-------------+
|    0|  8772|    Henry Arkell|England|Henry John Denham...|1898-06-26| Edmonton, Middles
ex| Dead|     12/03/82| 84|    Northamptonshire|Right-hand bat|              null| null
|    []|              null|              null|            null|              null|
null|             null|            null|            null|             null|
null|            null|            null|            null|            null|           nul
l|             null|           null|             null|           null|            null|
null|           null|            null|           null|           null|           null|
null|           null|             null|              null|            null|
null|            null|             null|            null|            null|
null|            null|            null|            null|            null|           nul
l|                      2|                      2.0|                     0|
11.0|                     6|                    5.5|                  null|
null|                      0|                   0.0|                  null|
null|                     0|                   0.0|              null|              nu
ll|             null|               null|             null|              null|
null|             null|              null|             null|             null|
null|             null|             null|            null|             null|
null|             null|           null|            null|           null|           null|
null|           null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|                      2|                    null
|                     null|                    null|                    null|
null|                   null|                   null|                    null|
null|                  null|                  null|                  null|
null|               null|                null|               null|               null|
null|              null|              null|               null|             null|
null|             null|             null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
|    1|532565|   Richard Nyren|England|       Richard Nyren|1734-04-25|     Eartham, Suss
ex| Dead|   1797-04-25| 63|        Hampshire XI| Left-hand bat|Left-arm bowler (...| null
|    []|              null|              null|            null|              null|
null|             null|            null|            null|             null|
null|            null|            null|            null|            null|           nul
l|             null|           null|             null|           null|            null|
null|           null|            null|           null|           null|           null|
null|           null|             null|              null|            null|
null|            null|             null|            null|            null|
null|            null|            null|            null|            null|           nul
l|                     49|                    90.0|                    11|
1026.0|                    97|                  12.98|                  null|
null|                      0|                   2.0|                  null|
null|                    24|                   0.0|              null|              nu
ll|             null|               null|             null|              null|
null|             null|              null|             null|             null|
null|             null|             null|            null|             null|
null|             null|           null|            null|           null|           null|
null|           null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|                     49|                      57
|                       0+|                      0+|                     104|
5/?|                    6/?|                   null|                    null|
null|                     7|                     1|                     0|
null|               null|                null|               null|               null|
null|              null|              null|               null|             null|
null|             null|             null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
|    2| 16856|Sydney Maartensz|England|Sydney Gratien Ad...|1882-04-14|     Colombo, Ceyl
on| Dead|     10/09/67| 85|           Hampshire|Right-hand bat|                null| nul
l|    []|             null|              null|            null|              null|
null|             null|            null|            null|             null|
null|            null|            null|            null|            null|            nul
l|             null|           null|             null|           null|            null|
null|           null|            null|           null|           null|           null|
null|           null|             null|              null|            null|
null|            null|             null|            null|            null|
null|            null|            null|            null|            null|            nul
l|                     12|                    17.0|                     2|
283.0|                    60|                  18.86|                  null|
null|                      0|                   1.0|                  null|
null|                    21|                   4.0|              null|               nu
ll|             null|               null|             null|              null|
null|             null|              null|             null|             null|
null|             null|             null|            null|             null|
null|             null|           null|            null|           null|           null|
null|           null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|                     12|                    null
|                     null|                    null|                    null|
null|                   null|                   null|                    null|
null|                  null|                  null|                  null|
null|               null|                null|               null|               null|
null|              null|              null|               null|             null|
null|             null|             null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
|    3| 16715|    Brian Lander|England|Brian Richard Lander|  09/01/42|Bishop Auckland, .
..|Alive|         null| 77|['Durham,', 'Mino...|Right-hand bat|    Right-arm medium| null
|    []|             null|              null|            null|              null|
null|             null|            null|            null|             null|
null|            null|            null|            null|            null|            nul
l|             null|           null|             null|           null|            null|
null|           null|            null|           null|           null|           null|
null|           null|             null|              null|            null|
null|            null|             null|            null|            null|
null|            null|            null|            null|            null|            nul
l|                   null|                    null|                  null|
null|                  null|                   null|                  null|
null|                   null|                  null|                  null|
null|                  null|                  null|                26|
21|               3|               122.0|               28|              6.77|
null|             null|                 0|              0.0|             null|
null|                2|              0.0|            null|             null|
null|             null|           null|            null|           null|           null|
null|           null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|                   null|                    null
|                     null|                    null|                    null|
null|                   null|                   null|                    null|
null|                  null|                  null|                  null|
26|               null|                1402|                859|                 25|
May-15|            May-15|             34.36|               3.67|               56|
0|               1|                 0|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
|    4| 15989|Derek Kenderdine|England|Derek Charles Ken...|1897-10-28|   Chislehurst, Ke
nt| Dead|     28/08/47| 50|          Royal Navy|Right-hand bat|Right-arm medium-...| null
|    []|             null|              null|            null|              null|
null|             null|            null|            null|             null|
null|            null|            null|            null|            null|            nul
l|             null|           null|             null|           null|            null|
null|           null|            null|           null|           null|           null|
null|           null|             null|              null|            null|
null|            null|             null|            null|            null|
null|            null|            null|            null|            null|            nul
l|                      2|                     4.0|                     1|
7.0|                     6|                   2.33|                  null|
null|                      0|                   0.0|                  null|
null|                     1|                   0.0|              null|             nu
ll|             null|               null|             null|              null|
null|             null|              null|             null|             null|
null|             null|             null|            null|             null|
null|             null|           null|            null|           null|           null|
null|           null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|            null|             null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
null|              null|               null|              null|              null|
null|             null|             null|              null|            null|
null|            null|            null|                      2|                    null
|                      198|                     121|                       2|
Jan-46|                   null|                   60.5|                    3.66|
99|                  null|                     0|                     0|              n
ull|               null|                null|               null|               null|
null|              null|              null|               null|             null|
null|             null|             null|            null|              null|
null|             null|             null|            null|            null|
null|             null|           null|           null|           null|           null|
+-----+------+----------------+-------+--------------------+----------+----------------
----+-----+-------------+---+--------------------+--------------+--------------------+-
----+------+-----------------+------------------+----------------+------------------+--
--------------+-----------------+----------------+----------------+-----------------+--
--------------+----------------+----------------+----------------+----------------+----
------------+-----------------+---------------+-----------------+--------------+------
----------+---------------+---------------+----------------+---------------+-----------
----+---------------+---------------+---------------+-----------------+----------------
--+----------------+------------------+----------------+-----------------+-------------
---+----------------+-----------------+----------------+----------------+--------------
--+----------------+----------------+-----------------------+------------------------+-
---------------------+------------------------+----------------------+-----------------
------+----------------------+----------------------+-----------------------+----------
------------+----------------------+----------------------+----------------------+-----
-----------------+------------------+-------------------+-----------------+------------
-------+-----------------+------------------+-----------------+----------------+-------
------------+-----------------+-----------------+-----------------+-----------------+--
---------------+----------------+-----------------+---------------+------------------+-
-------------+----------------+---------------+---------------+----------------+-------
--------+---------------+---------------+---------------+---------------+--------------
---+------------------+-------------------+------------------+-----------------+-------
-----------+-----------------+-----------------+------------------+----------------+---
-------------+----------------+----------------+----------------+-----------------+----
--------------+-----------------+-----------------+----------------+-----------------+--
--------------+-----------------+---------------+---------------+--------------+-------
---------+-----------------+-----------------+--------------------+------------------+-
-----------------+-----------------+-----------------+-----------------+---------------
---+----------------+----------------+----------------+----------------+---------------
--------+------------------------+-------------------------+------------------------+--
----------------------+-----------------------+-----------------------+----------------
```

```
-------+------------------------+----------------------+----------------------+--------
--------------+----------------------+------------------+-------------------+----------
----------+-------------------+-------------------+------------------+-----------------
-+------------------+-------------------+-----------------+-----------------+----------
-------+-----------------+----------------+-----------------+------------------+-------
----------+-----------------+----------------+----------------+----------------+-------
----------+---------------+---------------+---------------+---------------+
only showing top 5 rows
```

Check the schema

Loading the dataframe to lake

In [12]:

```
player_info_raw.write.format("parquet").mode("overwrite").save(f"{bronze_database_loc}/{
player_info_table_name}")
```

**

Relational database example - Loading tablle from MySQL DB to lake/silver layer

**

Reading deliveries from mysql

In [17]:

```
deliveries_df = read_from_mysql(host, port, user, password, mssql_database, mssql_table)
deliveries_df.show()
```

```
+--------+------+--------------------+--------------------+----+----+--------------+---
-----------+------------+-------------+---------+--------+-----------+-----------+-----
-------+------------+----------+----------+----------------+--------------+-------+
|match_id|inning|        batting_team|        bowling_team|over|ball|       batsman|   no
n_striker|      bowler|is_super_over|wide_runs|bye_runs|legbye_runs|noball_runs|penalty_r
uns|batsman_runs|extra_runs|total_runs|player_dismissed|dismissal_kind|fielder|
+--------+------+--------------------+--------------------+----+----+--------------+---
-----------+------------+-------------+---------+--------+-----------+-----------+-----
-------+------------+----------+----------+----------------+--------------+-------+
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  11|   6|    SC Ganguly|
BJ Hodge|  SK Trivedi|            0|        0|       0|          0|          0|
0|           0|         0|         0|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   1|      BJ Hodge|
SC Ganguly|    SK Warne|            0|        0|       0|          0|          0|
0|           1|         0|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   2|    SC Ganguly|
BJ Hodge|    SK Warne|            0|        0|       0|          1|          0|
0|           0|         1|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   3|      BJ Hodge|
SC Ganguly|    SK Warne|            0|        0|       0|          0|          0|
0|           0|         0|         0|            null|          null|   null|
|     566|     2|Royal Challengers...|     Kings XI Punjab|   4|   1|      CH Gayle|AB de
Villiers|    R Dhawan|            0|        0|       0|          0|          0|
0|           1|         0|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   4|      BJ Hodge|
SC Ganguly|    SK Warne|            0|        0|       0|          0|          0|
0|           4|         0|         4|            null|          null|   null|
|     377|     1|    Rajasthan Royals|      Mumbai Indians|  12|   6|       OA Shah|
SR Watson|  KA Pollard|            0|        0|       0|          0|          0|
0|           1|         0|         1|            null|          null|   null|
|     566|     2|Royal Challengers...|     Kings XI Punjab|   4|   2|AB de Villiers|
CH Gayle|    R Dhawan|            0|        0|       0|          0|          0|
0|           1|         0|         1|            null|          null|   null|
0|          1|          0|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   5|       BJ Hodge|
SC Ganguly|    SK Warne|            0|        0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     377|     1|    Rajasthan Royals|      Mumbai Indians|  13|   1|        OA Shah|
SR Watson|JEC Franklin|           0|         0|        0|          0|           0|
0|          0|          0|         0|            null|          null|   null|
|     566|     2|Royal Challengers...|     Kings XI Punjab|   4|   3|       CH Gayle|AB de
Villiers|    R Dhawan|           0|         0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  12|   6|     SC Ganguly|
BJ Hodge|    SK Warne|           0|         0|        0|          0|           0|
0|          0|          0|         0|            null|          null|   null|
|     377|     1|    Rajasthan Royals|      Mumbai Indians|  13|   2|        OA Shah|
SR Watson|JEC Franklin|           0|         0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     566|     2|Royal Challengers...|     Kings XI Punjab|   4|   4|AB de Villiers|
CH Gayle|    R Dhawan|           0|         0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|       1|     1| Sunrisers Hyderabad|Royal Challengers...|   1|   1|      DA Warner|
S Dhawan|    TS Mills|           0|         0|        0|          0|           0|
0|          0|          0|         0|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  13|   1|       BJ Hodge|
SC Ganguly|   YK Pathan|            0|        0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     377|     1|    Rajasthan Royals|      Mumbai Indians|  13|   3|      SR Watson|
OA Shah|JEC Franklin|          0|         0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     566|     2|Royal Challengers...|     Kings XI Punjab|   4|   5|       CH Gayle|AB de
Villiers|    R Dhawan|           0|         0|        0|          0|           0|
0|          1|          0|         1|            null|          null|   null|
|     187|     2|Kolkata Knight Ri...|    Rajasthan Royals|  13|   2|     SC Ganguly|
BJ Hodge|   YK Pathan|           0|         0|        0|          0|           0|
0|          0|          0|         0|            null|          null|   null|
|       1|     1| Sunrisers Hyderabad|Royal Challengers...|   1|   2|      DA Warner|
S Dhawan|    TS Mills|           0|         0|        0|          0|           0|
0|          0|          0|         0|            null|          null|   null|
+--------+------+--------------------+--------------------+----+----+---------------+---
-----------+------------+-------------+---------+--------+-----------+-----------+-----
-------+------------+----------+----------+----------------+--------------+--------+
only showing top 20 rows
```

Storing deliveries in lake

In [23]:

```
deliveries_df.write.format("parquet").mode("overwrite").save(f"{bronze_database_loc}/{del
iveries_table_name}")
```

**

NOSQL ingestion example - Loading DynamoDB Table in s3 to lake/silver layer

**

Loading matches from dynamodb

In [20]:

```
matches_df = read_from_dynamodb(macthes_dynamodb_table_name)
matches_df.show()
```

```
+-------------+--------------------+----------+----------------+--------------------+--
------+------+--------------+--------------+------------+----------+---------------+----
```

```
-+----------+--------------------+--------------------+--------------------+-----------
---------+
|toss_decision|              winner|      city| player_of_match|             umpire3|  S
eason|result|       umpire2|win_by_wickets|win_by_runs|dl_applied|        umpire1|   id|
date|               team1|               team2|               venue|         toss_winner
|
+-------------+--------------------+----------+----------------+--------------------+--
------+------+--------------+--------------+-----------+----------+---------------+----
-+----------+--------------------+--------------------+--------------------+-----------
---------+
|        field| Chennai Super Kings|    Mumbai|        SK Raina|                    |IPL
-2011|normal|    SJA Taufel|             6|          0|         0|      Asad Rauf|  304|
24-05-2011|Royal Challengers...| Chennai Super Kings|    Wankhede Stadium| Chennai Super
Kings|
|        field|Kolkata Knight Ri...|   Kolkata|        UT Yadav|                    |IPL
-2015|normal|     M Erasmus|             0|         35|         0|   AK Chaudhary|  555|
04-05-2015|Kolkata Knight Ri...| Sunrisers Hyderabad|        Eden Gardens| Sunrisers Hyde
rabad|
|        field|      Mumbai Indians|   Chennai|       RG Sharma|     Vineet Kulkarni|IPL
-2019|normal|Anil Chaudhary|             0|         46|         0|    Nigel Llong|11335|
26-04-2019|      Mumbai Indians| Chennai Super Kings|M. A. Chidambaram...| Chennai Super
Kings|
|          bat| Sunrisers Hyderabad| Hyderabad|     Rashid Khan|Virender Kumar Sh...|IPL-
2018|normal|      O Nandan|             7|          0|         0| Bruce Oxenford| 7929|0
5-05-2018|    Delhi Daredevils| Sunrisers Hyderabad|Rajiv Gandhi Inte...|    Delhi Darede
vils|
|          bat|      Mumbai Indians|     Delhi|      SL Malinga|                    |IPL
-2011|normal|     RB Tiffin|             8|          0|         0|      AM Saheba|  238|
10-04-2011|    Delhi Daredevils|      Mumbai Indians|    Feroz Shah Kotla|    Delhi Dared
evils|
|        field|Kochi Tuskers Kerala|     Kochi|        BJ Hodge|                    |IPL
-2011|normal|     RJ Tucker|             0|         17|         0|         S Ravi|  278|
05-05-2011|Kochi Tuskers Kerala|Kolkata Knight Ri...|       Nehru Stadium|Kolkata Knight
Ri...|
|          bat| Chennai Super Kings|    Ranchi|       RA Jadeja|                    |IPL
-2014|normal|       NJ Llong|             0|         34|         0|   AK Chaudhary|  478|
02-05-2014| Chennai Super Kings|Kolkata Knight Ri...|JSCA Internationa...| Chennai Super
Kings|
|          bat|Kolkata Knight Ri...|     Delhi|       G Gambhir|                    |IPL
-2014|normal| C Shamshuddin|             8|          0|         0|   BNJ Oxenford|  485|
07-05-2014|    Delhi Daredevils|Kolkata Knight Ri...|    Feroz Shah Kotla|    Delhi Dared
evils|
|          bat|      Mumbai Indians|    Mumbai|      SL Malinga|                    |IPL
-2015|normal|   CB Gaffaney|             0|         20|         0|HDPK Dharmasena|  541|
25-04-2015|      Mumbai Indians| Sunrisers Hyderabad|    Wankhede Stadium|      Mumbai In
dians|
|        field|Royal Challengers...| Bangalore|        CH Gayle|                    |IPL
-2015|normal|   VA Kulkarni|             0|        138|         0| RK Illingworth|  557|
06-05-2015|Royal Challengers...|     Kings XI Punjab|M Chinnaswamy Sta...|     Kings XI P
unjab|
|        field|     Kings XI Punjab|Chandigarh|  AD Mascarenhas|                    |IPL
-2012|normal|   SK Tarapore|             7|          0|         0|    VA Kulkarni|  321|
12-04-2012|       Pune Warriors|     Kings XI Punjab|Punjab Cricket As...|     Kings XI P
unjab|
|        field|Kolkata Knight Ri...| Bangalore|      AD Russell|                    |IPL
-2016|normal|        S Ravi|             5|          0|         0|      M Erasmus|  606|
02-05-2016|Royal Challengers...|Kolkata Knight Ri...|M Chinnaswamy Sta...|Kolkata Knight
Ri...|
|        field|    Delhi Daredevils| Cape Town|      DL Vettori|                    |IPL
-2009|normal|     SD Ranade|            10|          0|         1|      MR Benson|  120|
19-04-2009|     Kings XI Punjab|    Delhi Daredevils|            Newlands|    Delhi Dared
evils|
|          bat|     Kings XI Punjab|   Kolkata|DPMD Jayawardene|                    |IPL
-2010|normal|     DJ Harper|             8|          0|         0|       S Asnani|  208|
04-04-2010|Kolkata Knight Ri...|     Kings XI Punjab|        Eden Gardens|Kolkata Knight
Ri...|
|        field|      Delhi Capitals| Hyderabad|          K Paul|     Chris Gaffaney|IPL
-2019|normal|Bruce Oxenford|             0|         39|         0| Anil Chaudhary|11321|
14-04-2019|      Delhi Capitals| Sunrisers Hyderabad|Rajiv Gandhi Intl...| Sunrisers Hyde
rabad|
|          bat|Kolkata Knight Ri...|   Kolkata|   Shoaib Akhtar|                    |IPL
-2008|normal|      IL Howell|             0|         23|         0|      Asad Rauf|   95|
13-05-2008|Kolkata Knight Ri...|    Delhi Daredevils|        Eden Gardens|Kolkata Knight
```

```
13-05-2008|Kolkata Knight Ri...|    Delhi Daredevils|        Eden Gardens|Kolkata Knight
Ri...|
|        field|    Delhi Daredevils|      Pune|       SV Samson|                    |IPL
-2017|normal|        S Ravi|             0|         97|         0|    AY Dandekar|    9|
11-04-2017|    Delhi Daredevils|Rising Pune Super...|Maharashtra Crick...|Rising Pune Sup
er...|
|        field|Royal Challengers...| Bangalore|         P Kumar|                    |IPL
-2008|normal|    SL Shastri|             0|          3|         0|    BR Doctrove|  111|
03-05-2008|Royal Challengers...|     Deccan Chargers|M Chinnaswamy Sta...|     Deccan Cha
rgers|
|          bat| Chennai Super Kings|     Delhi|      MEK Hussey|                    |IPL
-2013|normal|     RJ Tucker|             0|         48|         0|       NJ Llong|  454|
21-05-2013| Chennai Super Kings|      Mumbai Indians|    Feroz Shah Kotla| Chennai Super
Kings|
|          bat|     Kings XI Punjab|    Mohali|        CH Gayle|     Vineet Kulkarni|IPL
-2018|normal|Anil Chaudhary|             0|         15|         0|    Nigel Llong| 7909|
19-04-2018|     Kings XI Punjab| Sunrisers Hyderabad|Punjab Cricket As...|     Kings XI P
unjab|
+-------------+--------------------+----------+----------------+--------------------+--
------+------+--------------+--------------+-----------+----------+---------------+----
-+----------+--------------------+--------------------+--------------------+-----------
---------+
only showing top 20 rows
```

Loading matches in lake

In [21]:

```
matches_df.write.format("parquet").mode("overwrite").save(f"{bronze_database_loc}/{matches_table_name}")
```

Creating crawlers to generate metadata for ingested lake tables

In [23]:

```
bronze_tables_to_load = {player_info_table_name:player_info_raw,
                         deliveries_table_name:deliveries_df,
                         matches_table_name:matches_df}

crawl_tables(glue_client, bronze_tables_to_load, bronze_database_loc, bronze_database_name)
```

**

Verifying lake ingestion

**

In [8]:

```
loaded_player_info_dyf = glueContext.create_dynamic_frame.from_catalog(database=bronze_database_name, table_name=player_info_table_name)
loaded_deliveries_dyf = glueContext.create_dynamic_frame.from_catalog(database=bronze_database_name, table_name=deliveries_table_name)
loaded_matches_dyf = glueContext.create_dynamic_frame.from_catalog(database=bronze_database_name, table_name=matches_table_name)

# Convert the dynamic frame to a Spark DataFrame
lplayerinfo_df = loaded_player_info_dyf.toDF()
ldeliveries_df = loaded_deliveries_dyf.toDF()
lmatches_df = loaded_matches_dyf.toDF()
```

For player info

In [10]:

```
lplayerinfo_df.count()
```

```
90308
```

In [11]:

```
lplayerinfo_df.printSchema()
```

```
root
 |-- index: integer (nullable = true)
 |-- id: integer (nullable = true)
 |-- name: string (nullable = true)
 |-- country: string (nullable = true)
 |-- full_name: string (nullable = true)
 |-- birthdate: string (nullable = true)
 |-- birthplace: string (nullable = true)
 |-- died: string (nullable = true)
 |-- date_of_death: string (nullable = true)
 |-- age: integer (nullable = true)
 |-- major_teams: string (nullable = true)
 |-- batting_style: string (nullable = true)
 |-- bowling_style: string (nullable = true)
 |-- other: string (nullable = true)
 |-- awards: string (nullable = true)
 |-- batting_tests_mat: string (nullable = true)
 |-- batting_tests_inns: string (nullable = true)
 |-- batting_tests_no: string (nullable = true)
 |-- batting_tests_runs: string (nullable = true)
 |-- batting_tests_hs: string (nullable = true)
 |-- batting_tests_ave: string (nullable = true)
 |-- batting_tests_bf: string (nullable = true)
 |-- batting_tests_sr: string (nullable = true)
 |-- batting_tests_100: string (nullable = true)
 |-- batting_tests_50: string (nullable = true)
 |-- batting_tests_4s: string (nullable = true)
 |-- batting_tests_6s: string (nullable = true)
 |-- batting_tests_ct: string (nullable = true)
 |-- batting_tests_st: string (nullable = true)
 |-- batting_odis_mat: string (nullable = true)
 |-- batting_odis_inns: string (nullable = true)
 |-- batting_odis_no: string (nullable = true)
 |-- batting_odis_runs: string (nullable = true)
 |-- batting_odis_hs: string (nullable = true)
 |-- batting_odis_ave: string (nullable = true)
 |-- batting_odis_bf: string (nullable = true)
 |-- batting_odis_sr: string (nullable = true)
 |-- batting_odis_100: string (nullable = true)
 |-- batting_odis_50: string (nullable = true)
 |-- batting_odis_4s: string (nullable = true)
 |-- batting_odis_6s: string (nullable = true)
 |-- batting_odis_ct: string (nullable = true)
 |-- batting_odis_st: string (nullable = true)
 |-- batting_t20is_mat: string (nullable = true)
 |-- batting_t20is_inns: string (nullable = true)
 |-- batting_t20is_no: string (nullable = true)
 |-- batting_t20is_runs: string (nullable = true)
 |-- batting_t20is_hs: string (nullable = true)
 |-- batting_t20is_ave: double (nullable = true)
 |-- batting_t20is_bf: string (nullable = true)
 |-- batting_t20is_sr: double (nullable = true)
 |-- batting_t20is_100: string (nullable = true)
 |-- batting_t20is_50: double (nullable = true)
 |-- batting_t20is_4s: string (nullable = true)
 |-- batting_t20is_6s: double (nullable = true)
 |-- batting_t20is_ct: string (nullable = true)
 |-- batting_t20is_st: double (nullable = true)
 |-- batting_first_class_mat: string (nullable = true)
 |-- batting_first_class_inns: double (nullable = true)
 |-- batting_first_class_no: integer (nullable = true)
 |-- batting_first_class_runs: double (nullable = true)
 |-- batting_first_class_hs: string (nullable = true)
 |-- batting_first_class_ave: string (nullable = true)
 |-- batting_first_class_bf: string (nullable = true)
 |-- batting_first_class_sr: string (nullable = true)
 |-- batting_first_class_100: string (nullable = true)
 |-- batting_first_class_50: double (nullable = true)
 |-- batting_first_class_4s: string (nullable = true)
 |-- batting_first_class_6s: double (nullable = true)
 |-- batting_first_class_ct: string (nullable = true)
 |-- batting_first_class_st: double (nullable = true)
 |-- batting_list_a_mat: integer (nullable = true)
 |-- batting_list_a_inns: string (nullable = true)
 |-- batting_list_a_no: string (nullable = true)
 |-- batting_list_a_runs: double (nullable = true)
 |-- batting_list_a_hs: string (nullable = true)
 |-- batting_list_a_ave: string (nullable = true)
 |-- batting_list_a_bf: string (nullable = true)
 |-- batting_list_a_sr: string (nullable = true)
 |-- batting_list_a_100: string (nullable = true)
 |-- batting_list_a_50: double (nullable = true)
 |-- batting_list_a_4s: string (nullable = true)
 |-- batting_list_a_6s: double (nullable = true)
 |-- batting_list_a_ct: string (nullable = true)
 |-- batting_list_a_st: double (nullable = true)
 |-- batting_t20s_mat: string (nullable = true)
 |-- batting_t20s_inns: double (nullable = true)
 |-- batting_t20s_no: string (nullable = true)
 |-- batting_t20s_runs: double (nullable = true)
 |-- batting_t20s_hs: string (nullable = true)
 |-- batting_t20s_ave: string (nullable = true)
 |-- batting_t20s_bf: string (nullable = true)
 |-- batting_t20s_sr: double (nullable = true)
 |-- batting_t20s_100: string (nullable = true)
 |-- batting_t20s_50: double (nullable = true)
 |-- batting_t20s_4s: string (nullable = true)
 |-- batting_t20s_6s: double (nullable = true)
 |-- batting_t20s_ct: string (nullable = true)
 |-- batting_t20s_st: double (nullable = true)
 |-- bowling_tests_mat: string (nullable = true)
 |-- bowling_tests_inns: double (nullable = true)
 |-- bowling_tests_balls: string (nullable = true)
 |-- bowling_tests_runs: double (nullable = true)
 |-- bowling_tests_wkts: integer (nullable = true)
 |-- bowling_tests_bbi: string (nullable = true)
 |-- bowling_tests_bbm: string (nullable = true)
 |-- bowling_tests_ave: string (nullable = true)
 |-- bowling_tests_econ: string (nullable = true)
 |-- bowling_tests_sr: string (nullable = true)
 |-- bowling_tests_4w: string (nullable = true)
 |-- bowling_tests_5w: string (nullable = true)
 |-- bowling_tests_10: string (nullable = true)
 |-- bowling_odis_mat: string (nullable = true)
 |-- bowling_odis_inns: string (nullable = true)
 |-- bowling_odis_balls: string (nullable = true)
 |-- bowling_odis_runs: string (nullable = true)
 |-- bowling_odis_wkts: string (nullable = true)
 |-- bowling_odis_bbi: string (nullable = true)
 |-- bowling_odis_bbm: string (nullable = true)
 |-- bowling_odis_ave: string (nullable = true)
 |-- bowling_odis_econ: string (nullable = true)
 |-- bowling_odis_sr: string (nullable = true)
 |-- bowling_odis_4w: string (nullable = true)
 |-- bowling_odis_5w: string (nullable = true)
 |-- bowling_odis_10: string (nullable = true)
 |-- bowling_t20is_mat: string (nullable = true)
 |-- bowling_t20is_inns: string (nullable = true)
 |-- bowling_t20is_balls: string (nullable = true)
 |-- bowling_t20is_runs: string (nullable = true)
 |-- bowling_t20is_wkts: string (nullable = true)
 |-- bowling_t20is_bbi: string (nullable = true)
 |-- bowling_t20is_bbm: string (nullable = true)
 |-- bowling_t20is_ave: string (nullable = true)
 |-- bowling_t20is_econ: string (nullable = true)
 |-- bowling_t20is_sr: string (nullable = true)
 |-- bowling_t20is_4w: string (nullable = true)
 |-- bowling_t20is_5w: string (nullable = true)
 |-- bowling_t20is_10: string (nullable = true)
 |-- bowling_first_class_mat: string (nullable = true)
 |-- bowling_first_class_inns: string (nullable = true)
 |-- bowling_first_class_balls: string (nullable = true)
 |-- bowling_first_class_runs: string (nullable = true)
 |-- bowling_first_class_wkts: string (nullable = true)
 |-- bowling_first_class_bbi: string (nullable = true)
 |-- bowling_first_class_bbm: string (nullable = true)
 |-- bowling_first_class_ave: string (nullable = true)
 |-- bowling_first_class_econ: string (nullable = true)
 |-- bowling_first_class_sr: string (nullable = true)
 |-- bowling_first_class_4w: string (nullable = true)
 |-- bowling_first_class_5w: string (nullable = true)
 |-- bowling_first_class_10: string (nullable = true)
 |-- bowling_list_a_mat: string (nullable = true)
 |-- bowling_list_a_inns: string (nullable = true)
 |-- bowling_list_a_balls: string (nullable = true)
 |-- bowling_list_a_runs: string (nullable = true)
 |-- bowling_list_a_wkts: string (nullable = true)
 |-- bowling_list_a_bbi: string (nullable = true)
 |-- bowling_list_a_bbm: string (nullable = true)
 |-- bowling_list_a_ave: string (nullable = true)
 |-- bowling_list_a_econ: string (nullable = true)
 |-- bowling_list_a_sr: string (nullable = true)
 |-- bowling_list_a_4w: string (nullable = true)
 |-- bowling_list_a_5w: string (nullable = true)
 |-- bowling_list_a_10: string (nullable = true)
 |-- bowling_t20s_mat: string (nullable = true)
 |-- bowling_t20s_inns: string (nullable = true)
 |-- bowling_t20s_balls: string (nullable = true)
 |-- bowling_t20s_runs: string (nullable = true)
 |-- bowling_t20s_wkts: string (nullable = true)
 |-- bowling_t20s_bbi: string (nullable = true)
 |-- bowling_t20s_bbm: string (nullable = true)
 |-- bowling_t20s_ave: string (nullable = true)
 |-- bowling_t20s_econ: string (nullable = true)
 |-- bowling_t20s_sr: string (nullable = true)
 |-- bowling_t20s_4w: string (nullable = true)
 |-- bowling_t20s_5w: string (nullable = true)
 |-- bowling_t20s_10: string (nullable = true)
```

For deliveries

In [12]:

```
ldeliveries_df.count()
```

```
179078
```

In [13]:

```
ldeliveries_df.printSchema()
```

```
root
 |-- match_id: integer (nullable = true)
 |-- inning: integer (nullable = true)
 |-- batting_team: string (nullable = true)
 |-- bowling_team: string (nullable = true)
 |-- over: integer (nullable = true)
 |-- ball: integer (nullable = true)
 |-- batsman: string (nullable = true)
 |-- non_striker: string (nullable = true)
 |-- bowler: string (nullable = true)
 |-- is_super_over: integer (nullable = true)
 |-- wide_runs: integer (nullable = true)
 |-- bye_runs: integer (nullable = true)
 |-- legbye_runs: integer (nullable = true)
 |-- noball_runs: integer (nullable = true)
 |-- penalty_runs: integer (nullable = true)
 |-- batsman_runs: integer (nullable = true)
 |-- extra_runs: integer (nullable = true)
 |-- total_runs: integer (nullable = true)
 |-- player_dismissed: string (nullable = true)
 |-- dismissal_kind: string (nullable = true)
 |-- fielder: string (nullable = true)
```

In [11]:

```
# For matches
```

In [14]:

```
lmatches_df.count()
```

```
756
```

In [15]:

```
lmatches_df.printSchema()
```

```
root
 |-- toss_decision: string (nullable = true)
 |-- winner: string (nullable = true)
 |-- city: string (nullable = true)
 |-- player_of_match: string (nullable = true)
 |-- umpire3: string (nullable = true)
 |-- Season: string (nullable = true)
 |-- result: string (nullable = true)
 |-- umpire2: string (nullable = true)
 |-- win_by_wickets: long (nullable = true)
 |-- win_by_runs: long (nullable = true)
 |-- dl_applied: long (nullable = true)
 |-- umpire1: string (nullable = true)
 |-- id: long (nullable = true)
 |-- date: string (nullable = true)
 |-- team1: string (nullable = true)
 |-- team2: string (nullable = true)
 |-- venue: string (nullable = true)
 |-- toss_winner: string (nullable = true)
```