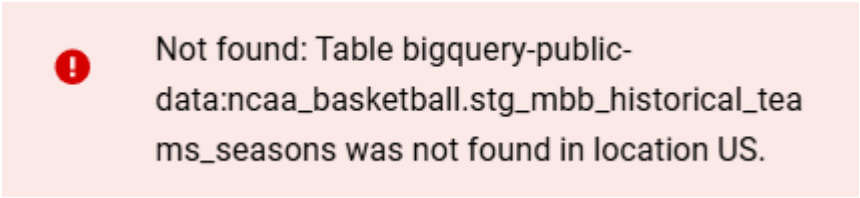
Optimal Approach:

- Source Declaration (Easier referencing, ensure linearity)
- Data Transformation:
    - Retrieve data from source, conduct small alterations to tidy up raw data
- Modular Transformation:
    - Cleaning up data and merging necessary tables to ensure modularity and readability
- Query from the modular tables

Reason optimal approach was not achieved:



Due to the error above, unable to utilize reference point to conduct modularity, thus star schema table was not produced

Setup Intructions:

1. Open the dataset in Google BigQuery
2. Go to IAM, ensure the service account has the following roles:
    a. BigQuery Admin
    b. BigQuery Data Editor
    c. BigQuery Data Viewer
    d. BigQuery Job User
    e. BigQuery User
3. Ensure *workflow_settings.yaml* has the following settings:

```
defaultProject: bigquery-public-data
defaultLocation: US
defaultDataset: ncaa_basketball
defaultAssertionDataset: dataform_assertions
dataformCoreVersion: 3.0.0
```

4. Declare source
5. Conduct data and modular transformation
6. Query from the transformed tables

Star Schema:

Fact Table:

mbb_pbp_sr

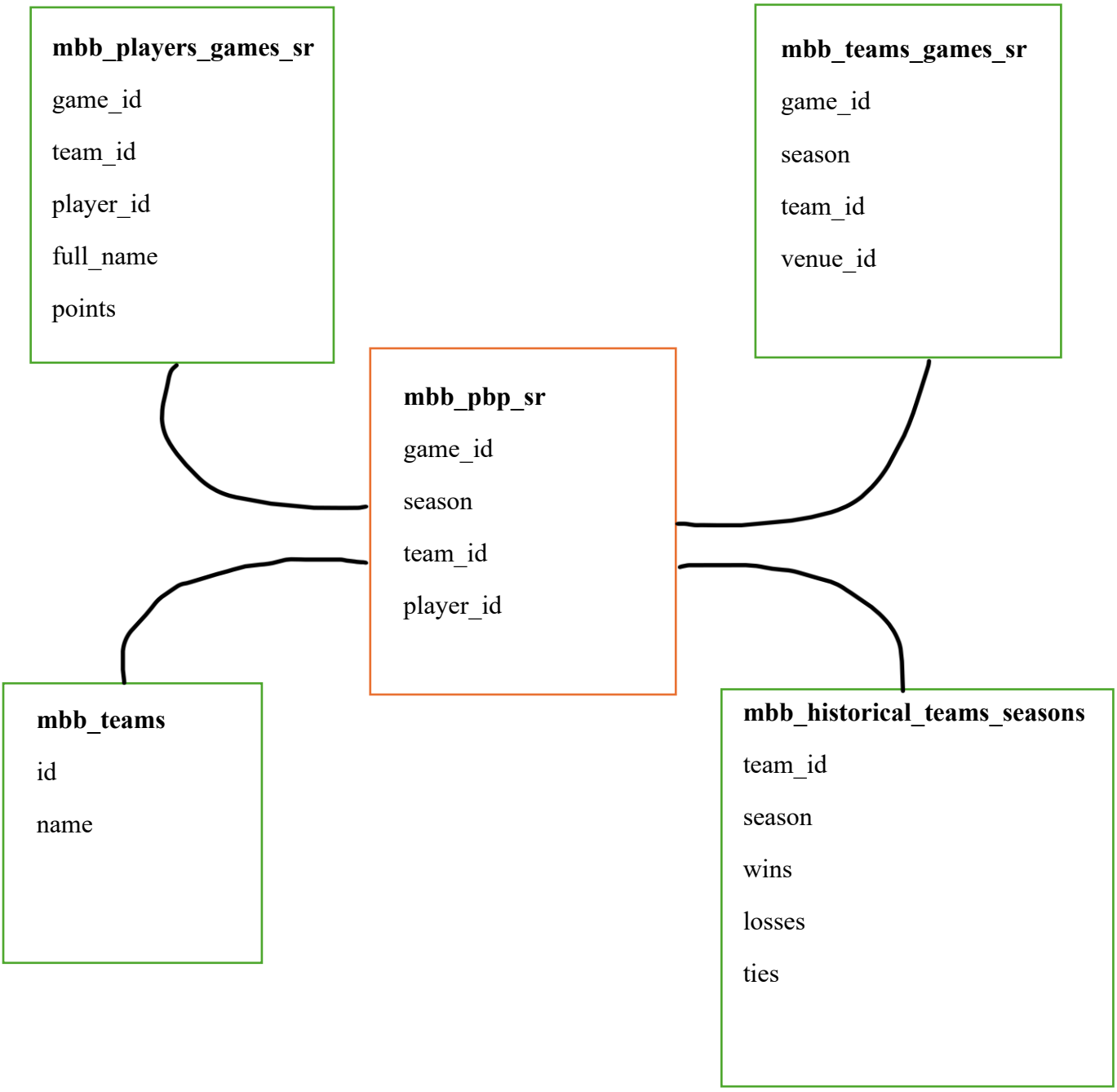
Dimension Table:

mbb_players_games_sr

mbb_teams

mbb_teams_games_sr

mbb_historical_teams_seasons



mbb_php_sr is chosen as fact table because it contains the most foreign keys to different primary keys.

Directory Structure Explanation:

source_declarations folder: Contains source declarations files

staging folder : Contains data transformed files (supposedly to include modular transformed files, and then to create star schema tables from there)

Queries folder : Contains query files(SQL) for the 3 selected questions (q1-3)

Tables folder : Contains output tables from queries

README.pdf : Contains explanations, approaches and setup intructions