```
IPL cricket database
              Abstract:
             The motive of this assignment is to build a physical model for IPL cricket database. Raw data is
              collected from Kaggle as a CSV format and data is audited for null values, cleaned and reformatted
             to fit the conceptual model.
              Data extracted from the Source
                1. Match – Match ID, Teams played, Date of match, Season, venue, City name, Country, Toss
                   winner, Match winner, Toss Decision, win type, Outcome, Man of the match, Win margin
                2. Player Role - Match ID, Player ID, Player Name, Date of Birth, batting Hand, Bowling Skill,
                   Country name, Role Description
                3. Ball by Ball details - Match ID, Innings ID, Over ID, Ball ID, Runs scored, Extra type, Out type
                4. Player Details - Player ID, Player Name, Country
                5. Team - Team ID, Team Name
              Normalization of Dataset:
             Normalization is a database design technique which organizes tables in a manner that reduces
              redundancy and dependency of data. In this dataset we are going to perform first three stages of
              normalization.
              1) First normal form:
             The dataset that we have downloaded is already in the first normal form because every table has a
             unique primary key and each attribute contain only atomic values. The ER form of the first normal
             form is shown below
In [11]: from PIL import Image
              from IPython.display import display
              display(Image.open('1st Normal Form.png'))
                     Ball by Ball
                                                            Player Match
                                                                                                              Match
               Match ID
                                                      Match ID
                                                                                                     Match ID
               Innings ID
                                                       Player ID
                                                                                                     Team 1
               Over ID
                                                                                                     Team 2
                                                       Player Name
                                                       Date of Birth
                                                                                                     Match Date
               Team Batting
                                                      Batting hand
                                                                                                     Season Year
               Team Bowling
                                                                                                     Venue name
                                                      Bowling Skill
               Runs Scored
                                                       Role Description
                                                                                                     City Name
               Extra Type
                                                       Country
                                                                                                     Country
               Out Type
                                                                                                     Toss Name
               Match date
                                                                                                     Toss winner
               Season Year
                                                                                                     Match winner
                                                                                                     win type
                                                               Player
                                                                                                     Outcome
                                                       Player ID
                                                                                                     Man of the match
                                                       Player
                                                                                                     win margin
                                                      Country Name
                                                                                                     Country ID
              2) Second normal form:
             In the second normal form, primary key in the role description table is Match ID and player ID but
             the attributes Player name, Date of birth, Batting hand, Bowling skill is partially dependent on the
             player ID. So these attributes are moved to the Player table.
             The ER form of the first normal form is shown below
In [14]: from PIL import Image
              from IPython.display import display
              display(Image.open('2nd Normal Form.png'))
                                                                                                               Match
                                                          Ball by Ball
                                                                                                       Match ID
                                                     Match ID
                                                                                                       Team 1
                                                     Innings ID
                                                                                                       Team 2
                                                     Over Number
                                                                                                       Venue name
                                                     Ball Number
                                                                                                       City name
                                                     Runs Scored
                                                                                                       Country
                                                     Extra Type
                                                                                                       Match Date
                                                     Out Type
                                                                                                       Season Year
                                                                                                       Toss name
                                                                                                       Toss Winner
                       Player
                                                                                                       Match Winner
               Player ID
                                                                                                       Win type
               Player name
                                                               Role Description
                                                                                                       Outcome
                                                             Match ID
               Date of Birth
                                                                                                       Man of the match
                                                             Player ID
               Batting hand
                                                                                                       Win margin
                                                            Role Description
               Bowling skill
                                                                                                       Country ID
               Country
             3) Third normal form
             In the third normal form, the attributes which has transitive functional dependencies are divided.
             The ER form of the third normal form is shown below
In [15]: from PIL import Image
              from IPython.display import display
              display(Image.open('3rd Normal Form.png'))
                                                                                                           Toss
                         Extra_Runs
                                                 Wicket_type
                                                                                                     Toss_ID
                                                                                Match
                    Extra_ID
                                                                                                     Toss Description
                                                                           Match ID
                    Extra_type
                                             Out_type
                                                                           Team 1
                                                                           Team 2
                                                                                                                   Win
                                                        Team
                              Ball by Ball
                                                                           Match Date
                                                 Team_ID
                          Match ID
                                                                           Season Year
                                                 Team Name
                                                                           Venue name
                          Over Number
                                                                           City Name
                                                    Country
                                                                                                                  Outcome
                          Ball Number
                                               Country ID
                           Runs Scored
                                                                           Toss Name
                           Extra Type
                                                                           Toss winner
                          Out Type
                                                                           Match winne
                                                                           win type
                                                                                                                 Venue
                                                                           Outcome
                                            City
                                     City_ID
                                                                           Man of the match
                                     City Name
                                                                           win margin
                                                           Batting_hand
                               Player
                                                                                                                 Player role
                                                                                                             Role_ID
                          Player ID
                         Date of Birth
                                                          Bowling_Skill
                         Batting hand
                         Bowling skill
                                                                                 Role Description
                                                                               Role Description
              The conceptual model
             The below model displays the conceptual entities and their relationship.
 In [3]: from PIL import Image
              from IPython.display import display
              display(Image.open("C:\\Users\\kumaran\\Downloads\\Conceptual model.png"))
                        Ball by Ball
                                            Extra_Runs
                                                                                                         oss_ID
                                                                                        Match
                     Match ID
                                         Extra_ID
                                                             Team_ID
                                                                                    Match ID
                                                                                                        Toss Descriptio
                     nnings ID
                     Over Number
                                                                                    Team 2
                    Ball Number
                                                                                    Match Date
                                                                                                                    Win
                    Runs Scored
                                                                                    Season Year
                                                                  Country
                     Extra Type
                                                                                                              Win type
                                                                                    Venue name
                     Out Type
                                                                                    City Name
                                                                                    Country
                                                                                                                   Outcome
                                                                                    Toss Name
                       Wicket_type
                                                                                    Toss winner
                                                                                                               Outcome
                     Out_ID
                                                                                    Match winne
                     Out_type
                                                                                    vin type
                                                                                                                  Venue
                                                                                    Outcome
                 Batting_hand
                                                                                                              enue_ID
                                                                                    Man of the match
                                                                  City
                                                                                                             Venue name
                                                                                    win margin
               Batting hand
                                               Player
                                                             City_ID
                                           Player ID
                                           Player name
                                                                                                                  Player role
                                           Date of Birth
                                                                                      Role Description
                                                                                                               Role_ID
                 Bowling_Skill
                                          Batting hand
                                                                                     Match ID
                                                                                                               Role Description
                                                                                     Player ID
                                           Bowling skill
                                                                                    Role_ID
              ER Diagram
 In [4]: from PIL import Image
              from IPython.display import display
              display(Image.open("C:\\Users\\kumaran\\Downloads\\ERD.png"))
                                                                                                             Toss
                        Ball by Ball
                                            Extra_Runs
                                                                                                         Toss_ID (PK)
                                                                  Team
                                                                                        Match
                     Match ID (PK)
                                         Extra_ID (PK)
                                                             Team_ID (PK)
                                                                                   Match ID (PK)
                                                                                                        Toss Description
                     nnings ID (PK)
                                         Extra_type
                     Over Number (PK)
                                                                                    Team 2
                    Ball Number (PK)
                                                                                    Match Date
                                                                                                                    Win
                     Runs Scored
                                                                                                              D (PK)
                                                                                    Season Year
                                                                  Country
                     Extra Type
                                                                                                              Win type
                                                                                    Venue name
                    Out Type
                                                                                    City Name
                                                                                    Country
                                                                                                                   Outcome
                                                                                    Toss Name
                                                                                                               Outcome ID (PK)
                       Wicket_type
                     Out_ID
                                                                                    Toss winner
                                                                                                               Outcome
                                                                                    Match winne
                    Out_type
                                                                                    win type
                                                                                                                  Venue
                                                                                    Outcome
                Batting_hand
                                                                                                              /enue_ID (PK)
                                                                                    Man of the match
               ID (PK)
                                                                  City
                                                                                                             Venue name
                                                                                    win margin
                                                            City_ID (PK)
               atting hand
                                               Player
                                                            City Name
                                          Player ID (FK)
                                          Player name
                                                                                                                   Player role
                                           Date of Birth
                                                                                     Role Description
                                                                                                               Role_ID (PK)
                 Bowling_Skill
                                          Batting hand
                                                                                     Match ID (PK)
               ID (PK)
                                                                                     Player ID (PK)
                                           Bowling skill
                                                                                    Role_ID
                                           Country
              Relationship between the entities
              Physical model
 In [5]: from PIL import Image
              from IPython.display import display
              display(Image.open("C:\\Users\\kumaran\\Downloads\\Physical model.png"))
                                                                                      Match_ID
                    Wicket type
                                                                                    IX_Relationship13 (IX2)
IX_Relationship14 (IX3)
                                                                  Role Descript
                                     IX_Relationship23 (IX3)
                  ling Skill Char(45
              Questions about the Conceptual model
              1) What are the ranges, data types and format of all of the attributes in your entities?
             1. Ball by Ball:
             Match ID: INT, Innings ID: INT, Over number: INT, Ball number: INT, Runs scored: INT,
             Extra_type: INT, Out_type: INT,
             2. Extra_Runs
              Extra_ID: INT, Extra_type: VarChar(45)
              3. Wicket_type
              Out_ID : INT , Out_type :VarChar(45)
```

## 7. Outcome ID: INT, Outcome:char(45) 8. Venue

Match winner: INT, Win type: INT, Outcome: INT,

Toss\_ID: INT, Toss Description: Char(45)

ID: INT, Win type: CHar(45)

Match ID: INT, Team 1: Varchar(45), Team 2: Varchar(45), Match Date: DATE, Season year :INT, Venue name:INT, City name:INT, Country:INT, Toss name:INT, Toss Winner:INT,

Player ID: INT, Player name: char(45), Date of birth: DATE, Batting hand: INT, Bowling skill:

Entity is the object and attribute is the property of the Object. When there are certain characteritic

for an object we will be considering that Object as the entity and its charactertic as attribute.

3) When should you use an entity or relationship, and placement of attributes? (Example: a

Entity is the Object which has characteritic as attributes. Relationship is formed where is an

Keys were choosed based on the uniqueness of the attribute which can identify a particular record.

ID: INT, Batting hand: char(45), 14. Bowling\_Skill ID: INT, Bowling skill: char(45)

4. Match

5. Toss

6. Win

ID: INT,

9. Team

10. Country

11. City

12. Player

INT, Country: INT,

13. Batting hand

15. Role\_Description

16. Player\_Role

Venue name: char(45)

Team\_ID: INT, Team name: char(45)

Country ID: INT, Country: char(45)

City\_ID: INT, City name: char(45)

Win margin :INT

Role\_ID: INT, Description: char(45) 2) When should you use an entity versus attribute? (Example: address of a person could be modeled as either)

manager could be modeled as either)

dependency between two entities.

4) How did you choose your keys? Which are unique?

in Match table , Match\_ID is unique which will identify a particualar record.

7) Where are you going find real-world data populate your model?

The tables are in 2NF because there are partial dependency.

Questions you must answer about your physical model:

The tables are in 3NF because transitive dependency are eliminated by splitting the table.

The above files was already audited for null values and cleaned. So there is no specific need to

Normalization was carried out on the dataset to reduce redundancy, insertion anamoly, deletion

obtaining a copy of this software and associated documentation files (the "Software"), to deal in the

FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE

Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software. THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,

Files used: Match.csv, Ball by Ball.csv, Player details.csv, Player role.csv

5) Did you model hierarchies using the "ISA" design element? Why or why not?

Match ID: INT, Player ID: char(45), Role Description: char(45),

6) Were there design alternatives? What are their tradeoffs: entity vs. attribute, entity vs. relationship, binary vs. ternary relationships? There should be no design alternative because the data can be linked only with Match ID and Played ID.

1) Are all the tables in 1NF? Yes, every value in each table is atomic. Are all the tables in 2NF?

perform auditing for null values again.

anamoly and updation anamoly.

Conclusion

Contribution

Self efforts: 50%

External source: 20%

Guidance by the Professor: 10%

Guidance by Teaching Assistant: 20%

Are all the tables in 3NF?

Report

Real world data was found on Kaggle.

Data downloaded from Kaggle had lots of redundancy. Normalization techniques were carried out to reduce redundancy, insertion anamoly, deletion anamoly and updation anamoly. Also, Conceptual, physical and ERD diagram for the dataset is created.

Data source: <a href="https://www.kaggle.com/raghu07/ipl-data-till-2017">https://www.kaggle.com/raghu07/ipl-data-till-2017</a> LICENSE: Copyright 2020 Kumaran Nehru Uthra. Permission is hereby granted, free of charge, to any person

**Citations** 

AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

In [ ]: