Data and Applications Monsoon 2022 - Project Phase 1

Team 7 : Chirag Jain, Pratham Thakkar, Vineeth Bhat

October 2022

1 Introduction - Tokyo 2020 Olympics

1.1 Miniworld

- The mini world considered by our team is the 2020 Summer Olympics, officially known as the Games of the XXXII Olympiad and also known as Tokyo 2020. It was an international multi-sport event held in Tokyo, Japan, from July 23 to August 8, 2021.
- At the 2020 Summer Olympics, several countries were represented by a diverse range of athletes competing in a variety of sports. In our project, the Universe of Discourse focuses on the performance of these contestants and countries and how various factors influence it.

1.2 Purpose of Database

- The database will track the various sports played at the 2020 Summer Olympics, as well as the athletes who competed, their coaches, and the countries they represented, as well as the countries' current sports budget and overall performance.
- It also stores additional information about the Overseers in charge of the various matches being played, as well as the Sponsors and venue.

1.3 Users of the Database

- The database is subject to separate views for different entities, for example the contestants can use to to analyse their performance at different venues and time and thus conduct future training accordingly.
- Countries can use it to analyse their players' attributes and determine which attributes influenced their wins and losses and therefore do the future selections accordingly.

1.4 Applications of the Database

- We would try to find answers to the following questions using this database.:
  1. What effect does the host country have in the medals won at the Olympics?
  2. Is the performance of countries in Olympic games affected by the economic factors of the country?
  3. Is the age of winning Olympics changing?
- This database will also help various countries identify their strengths and weaknesses, as well as areas where they need to improve.

2 Database Requirements

2.1 Entity Types

1. Contestant

   Attributes -

   - Olympic ID (Primary Key - Alphanumeric)
   - Name (Alphabets)
   - Height (Float)

- Sex (Character)
- Weight (Float)
- DOB (Date)
- Age (Integer - derived attribute)
- Career (Multi-valued attribute listing previous years in which the athlete played in the Olympics, can be null)
- Previous Wins (Multi-valued attribute listing previous years in which the athlete won in the Olympics, can be null)

2. Country

   Attributes -

   - Name (Primary Key - alphabets)
   - Representational Name (Candidate key - alphabets)
   - Continent (alphabets)
   - Participation (Multi-valued attribute storing the previous years in which the country participated in the Olympics, can be null)
   - Previous Olympic Rank (Integer, can be null)
   - Investment (Integer)

3. Match (Weak Entity)

   Attributes -

   - Start Timing (Date)
   - End Timing (Date)
   - Configuration of Contestants (Integer - example: mixed doubles)
   - Match Specification (text - example: 100m race) Partial key - composite key of timings

4. Sport

   Attributes -

   - Name (Primary Key)
   - Configurations (multi-valued attribute of texts. Example - doubles, singles, etc.)

5. Coach

   Attributes -

   - Olympic ID (Primary Key - alphanumeric)
   - Name (Alphabets)
   - DOB (Date)
   - Age (Integer - derived attribute)

6. Overseer

   Attributes -

   - Olympic ID (Primary Key - alphanumeric)
   - Name (Alphabets)
   - DOB (Date)
   - Age (Integer - derived attribute)
   - Type (Alphabets - judge, referee, etc.)

7. Medal (Weak Entity)

   Attributes

   - Type (Gold, Silver, Bronze)

- Time of conferring (date)
  Partial key - composite key of Type and Conferring time

8. Sponsor
   Attributes -
   - Name (Primary Key - alphanumeric)
   - Donation value (Integer)
9. Venue
   Attributes -
   - Name (Alphabets - can be null)
   - Location (Text)
     Primary key - Composite Key of Name and Location

2.2 Relationship Types

1. Contestant PLAYS in a Match in sport at venue
   - Participating Entities - Contestant, Match, Sport, Venue
   - Degree - 4
   - Min, Max - Contestant : (1, N); Match : (1, 1), Sport : (1, N), Venue : (1, N)
   - Salient Features - Identifying relationship for match with partial key as the match duration
2. Contestant is COACHED by Coach
   - Participating Entities - Contestant, Coach
   - Degree - 2
   - Min, Max - Contestant : (1, N); Coach : (1, N)
   - Salient Features - A contestant can be coached by more than one coach
3. Match is PLAYED among Countries
   - Participating Entities - Match, Country
   - Degree - 2
   - Min, Max - Match : (1, N); Countries : (2, N)
   - Salient Features - None
4. Overseer OVERSEES a match
   - Participating Entities - Overseer, Match
   - Degree - 2
   - Min, Max - Overseer : (1, N); Match : (1, N)
   - Salient Features - None
5. Contestant REPRESENTS Country
   - Participating Entities - Contestant, Country
   - Degree - 2
   - Min, Max - Contestant : (1, 1); Country : (0, N)
   - Salient Features - None
6. Country GOT Medal in Match
   - Participating Entities - Country, Medal, Match
   - Degree - 3
   - Min, Max - Country : (0, N); Medal : (1, 1); Match : (1, N)

- Salient Features - Ternary and identifying relationship for medal. Further, there can even be draws so multiple winners can be given the same medal.

7. Coach is FROM Country
    - Participating Entities - Coach, Country
    - Degree - 2
    - Min, Max - Coach : (1,N); Country : (1, 1)
    - Salient Features - None
8. Contestant is SPONSORED by Sponsor
    - Participating Entities - Contestant, Sponsor
    - Degree - 2
    - Min, Max - Contestant : (1,N); Sponsor : (1, N)
    - Salient Features - None

3 Functional Requirements

3.1 Retrieval

3.1.1 Selection

- Retrieve all Contestants belonging to a Country
- Retrieve all coaches from a country

3.1.2 Projection

- List all players with a gold medal
- List all matches played between more than 2 countries

3.1.3 Aggregate

- List the number of medals won by a country
- List the country with the maximum number of medals
- List the medals won by a player
- List the numbers of matches played in a sport
- Average number of medals won by countries in a continent

3.1.4 Search

- List all players whose name starts with a particular letter
- List all matches whose duration was more than 2 hours

3.1.5 Analysis

- We can generate the ranking of the countries based on the number of medals won
- We can analyse how much investment from a sponsor lead to how many medals from the sponsored
- We can also see the correlation of investment of a country and the number of medals won
- We can see the amount of participation from different sexes from different countries
- We can relate countries within the same continent based on number of medals won
- We can derive which athlete won the maximum number of medals from a country

3.2 Modification

- Insert data about Players and Coaches.
- Insert data about Match and Venue.
- Insert data about Sponsors.
- Insert data about Participating Countries.
- Update Match results and Medals won by Player.
- Update Venue and Timings for a Match if changed.
- Update Overseer for a Match.
- Delete Player from database
- Delete Sponsor from database