**SVKM’s NMIMS**

**School of Technology Management & Engineering, Chandigarh**

A.Y. 2023 - 24

**Course: Database Management Systems**

**Project Report**

|  |  |  |
| --- | --- | --- |
| Program | MBA(Tech) DS | |
| Semester | 4th | |
| Name of the Project: | Olympics Games 2020 Database | |
|  | | |
| Details of Project Members |  |  |
| Batch | Roll No. | Name |
| J2 | S025 | Kian Bilawala |
| J2 | S026 | Kinjal Mandlecha |
| J2 | S027 | Krutant Shah |
| J2 S031 Mausam Shah | | |
| J2 S042 Ranapratap Deshmukh | | |
|  | | |
|  | | |
| Date of Submission: 1/04/2024 | | |
|  | | |
|  | | |
|  | | |
|  | | |
|  | | |
|  | | |

**Contribution of each project Members:**

|  |  |  |
| --- | --- | --- |
| Roll No. | Name: | Contribution |
| S025 | Kian Bilawala | Sql queries, Report, ER diagram, Relation model,etc. |
| S026 | Kinjal Mandlecha | Sql queries, Report, ER diagram, Relation model,etc. |
| S027 | Krutant Shah | Sql queries, Report, ER diagram, Relation model,etc. |
| S031 | Mausam Shah | Sql queries, Report, ER diagram, Relation model,etc. |
| S042 | Ranapratap Deshmukh | Sql queries, Report, ER diagram, Relation model,etc. |

**Github link of your project:**

**Note:**

1. Create a readme file if you have multiple files
2. All files must be properly named (Example:R004\_DBMSProject)
3. Submit all relevant files of your work ( Report, all SQL files, Any other files)
4. **Plagiarism is highly discouraged (Your report will be checked for plagiarism)**

**Rubrics for the Project evaluation:**

|  |  |
| --- | --- |
| First phase of evaluation:  Innovative Ideas (5 Marks)  Design and Partial implementation (5 Marks) | 10 marks |
| Final phase of evaluation  Implementation, presentation and viva, Self-Learning and Learning Beyond classroom | 10 marks |

**Project Report**

**Olympics Games 2020 Database**

**by**

Kian Bilawala, S025

Kinjal Mandlecha, S026

Krutant Shah, S027

Mausam Shah, S031

Ranapratap Deshmukh, S042

**Course: DBMS**

**AY: 2023-24**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr no.** | **Topic** | **Page no.** |
| **1** | Storyline |  |
| **2** | Components of Database Design |  |
| **3** | Entity Relationship Diagram |  |
| **4** | Relational Model |  |
| **5** | Normalization |  |
| **6** | SQL Queries |  |
| **7** | Learning from the Project |  |
| **8** | Project Demonstration |  |
| **9** | Self-learning beyond classroom |  |
| **10** | Learning from the project |  |
| **8** | Challenges faced |  |
| **9** | Conclusion |  |

**I. Storyline**

**INTRODUCTION:**

The Olympic Games is the world's most prestigious and celebrated international multi-sport event, featuring athletes from all over the world. The event is held every four years, and the next edition will take place in 2026.

1. The Olympic Games 2020 Database Project aims to design a database system to manage and organize the data related to the event. The database will be used to store information about athletes, countries, events, venues, records, medals, and sports, among other things.
2. This project will provide a comprehensive and integrated system to manage the data and provide accurate and relevant information to the users. The system will be designed to provide accurate, up-to-date information about athletes, countries, events, venues, records, medals, and sports, among other things. The system will also be designed to provide customized reports and analysis to the users, allowing them to analyze and understand the data in more detail.
3. The database will be designed using the entity-relationship (ER) model, which is a data model used to represent the relationships between different entities in a database system.

**MOTIVATION:**

1. Interest in sports: Being a fan of the Olympics or sports in general, we find it engaging to work with data related to the Games. We could use the database to explore trends and patterns across different sports, countries, and time periods, or to analyze the performance of specific athletes or teams.
2. Academic research: Interested in analyzing the Olympics database to explore questions about the impact of the Games on host cities, the economics of sports, or the role of sports in society.
3. Overall, the Olympics database can be a rich and rewarding topic for a project, with applications in a wide range of fields and contexts.

**OBJECTIVES:**

1. To design a database system that can effectively store, manage, and retrieve data related to the Olympic Games.
2. To analyze the existing methods for designing Olympic Databases.
3. To design the ER diagram of Olympic Games.
4. To design the Relational Schema of Olympic Games.
5. To implement the Olympic Database using SQL.

**II. Components of Database Design**

**ENTITIES:**

1. Athlete: This table will contain information about the athletes participating in the Olympic Games 2020, including their name , unique ID and the event they are participating in.
2. Country: This table will contain information about the countries participating in the Olympic Games 2020, including their name unique ID.
3. Events: This table will contain information about the events in the Olympic Games 2020, including the name of the event, the date and unique ID.
4. Venue: This table will contain information about the venues hosting the Olympic Games 2020, including the name and unique ID.
5. Performance: This table will contain information about the records set during the Olympic Games 2020, including the records made and unique ID.
6. Medals: This table will contain information about the medals awarded during the Olympic Games 2020, including the medal won and unique ID.
7. Sports: This table will contain information about the different sports featured in the Olympic Games 2020, including the name and unique ID.

**RELATION:**

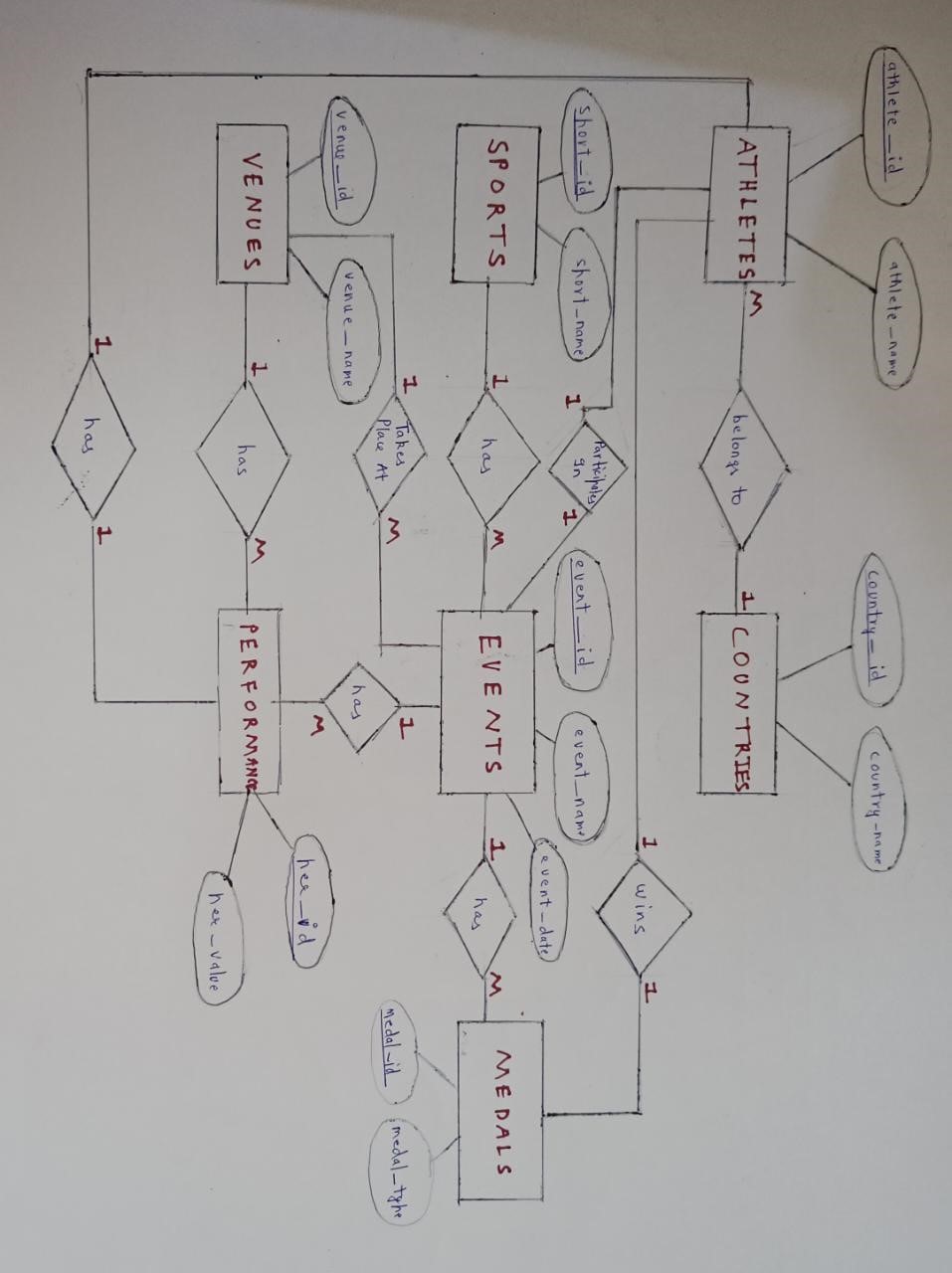
1. Athletes belongs to a country. (M-1)
2. An Athlete has a Performance. (1-1)
3. An Athlete wins a Medal. (1-1)
4. An Athlete participates in Event. (1-1)
5. Sport has many Events. (1-M)
6. Event has many Performances. (1-M)
7. Event has many Medals. (1-M)
8. Events takes place at Venue. (M-1)
9. Venue has many Performances. (1-M)

**DESCRIPTION:**

1. The Countries table contains information about countries participating in the Olympics, such as the country ID and name.
2. The Athletes table contains information about the athletes participating in the Olympics, such as athlete ID, name, the event they are participating in and the country they represent. The country\_id and event\_id column in this table is a foreign key referencing the Countries and Events table respectively.
3. The Sports table contains information about the sports played in the Olympics, such as sport ID and name.
4. The Events table contains information about the events taking place in the Olympics, such as event ID, name, date, the venue and the sport it belongs to. The sport\_id and venue\_id column in this table is a foreign key referencing the Sports and Venue Table respectively.
5. The Venues table contains information about the venues where the Olympic events take place, such as venue ID and name.
6. The Performance table contains information about the performance of the athletes in the respective events , such as performance ID, value The event\_id , venue\_id , athlete\_id column in this table is a foreign key referencing the Events , Venues , Athletes table respectively.

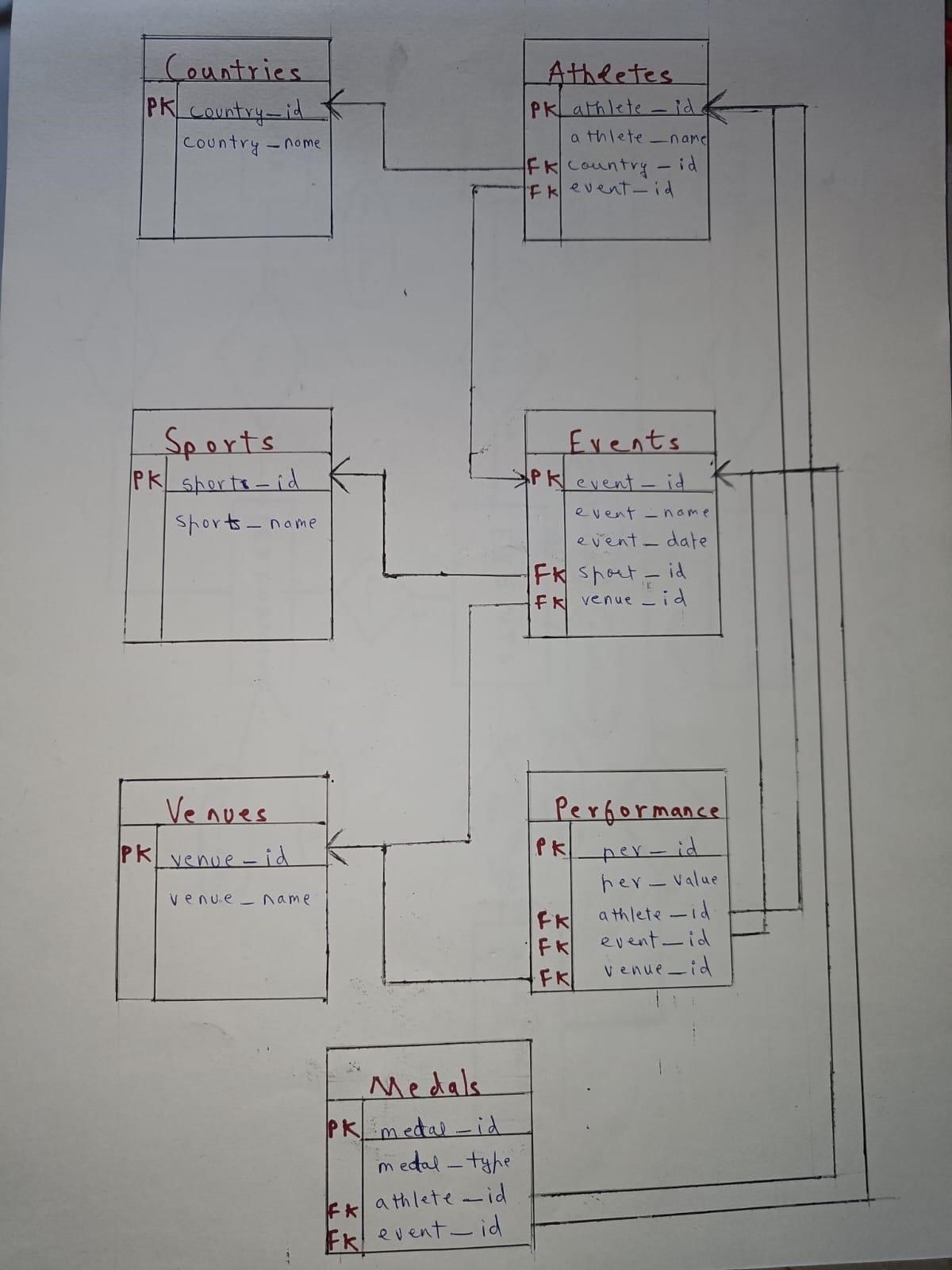
The Medals table contains information about the medals won by the athletes in the respective events , such as medal ID and medal type. The event\_id and athlete\_id column in this table is a foreign key referencing the Events and Athletes Table respectively.

**III. Entity Relationship Diagram**



**IV. Relational Model**

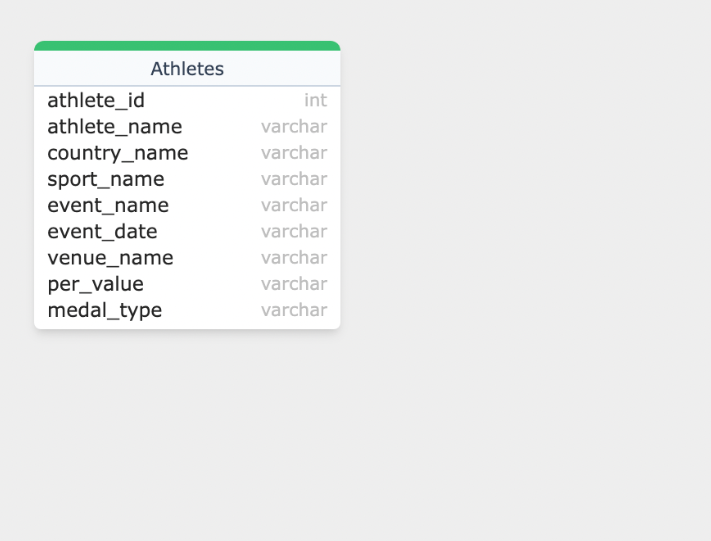
.



**V. Normalization**

Perform normalization (1NF, 2NF, 3NF, BCNF) as applicable for the entire database.

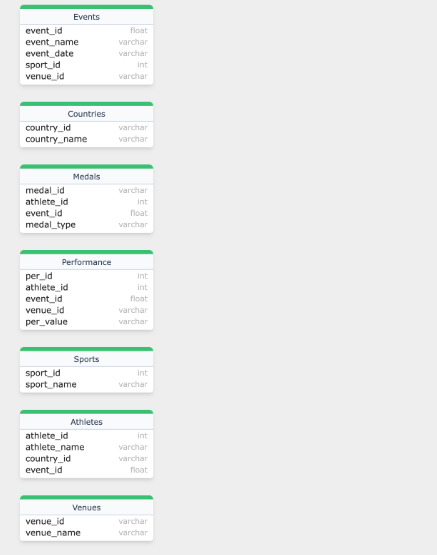
Before Normalization:

****

1. **First Normal Form (1NF)**:
   1. In 1NF, each column should contain atomic values, meaning no multi-valued attributes.
   2. To achieve 1NF, we need to ensure that each column contains only single values, and we don't have repeating groups.
   3. In our denormalized table, each column seems to contain atomic values, so it's already in 1NF.
2. **Second Normal Form (2NF)**:
   1. In 2NF, the table should be in 1NF and all non-key attributes should be fully functionally dependent on the primary key.
   2. To achieve 2NF, we need to remove partial dependencies.
   3. In our denormalized table, there seem to be partial dependencies. For example, country\_name, sport\_name, event\_name, event\_date, and venue\_name are all functionally dependent on athlete\_id.
   4. To remove partial dependencies, we need to split the table into smaller tables.
3. **Third Normal Form (3NF)**:
   1. In 3NF, the table should be in 2NF and there should be no transitive dependencies.
   2. To achieve 3NF, we need to remove transitive dependencies.
   3. In our denormalized table, event\_date, sport\_name, and venue\_name seem to be functionally dependent on event\_name. This indicates a transitive dependency.
   4. To remove this transitive dependency, we further split the tables to ensure that each non-key attribute is directly dependent on the primary key.

By following these steps, we can normalize the denormalized table into multiple smaller tables that eliminate data redundancy and ensure data integrity up to 3NF

After Normalization:



**VI. SQL Queries**

.

alter table Athletes add constraint FK\_country\_id

foreign key(country\_id) references Countries (country\_id);



alter table Athletes add constraint FK\_country\_id2

foreign key(event\_id) references Events (event\_id);



alter table Events add constraint FK\_sport\_id

foreign key(sport\_id) references Sports(sport\_id);

error.

alter table Events add constraint FK\_venue\_id

foreign key(venue\_id) references Venues(venue\_id);



alter table Performance add constraint FK\_athlete\_id

foreign key(athlete\_id ) references Athletes(athlete\_id );



alter table Performance add constraint FK\_event\_id

foreign key(event\_id) references Events(event\_id);



alter table Performance add constraint FK\_venue\_id2

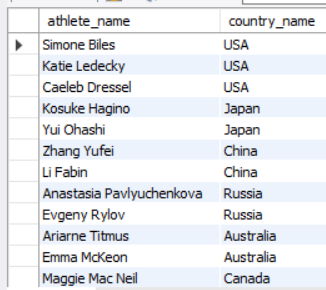
foreign key(venue\_id) references Venues(venue\_id);



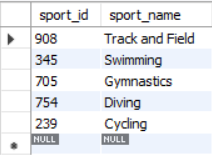
SELECT \* FROM Countries ORDER BY country\_name;



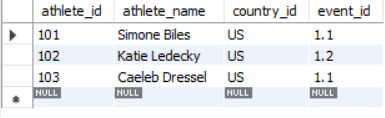
SELECT Athletes.athlete\_name, Countries.country\_name FROM Athletes JOIN Countries ON Athletes.country\_id = Countries.country\_id;



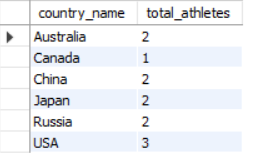
SELECT \* FROM Sports ORDER BY sport\_name DESC;



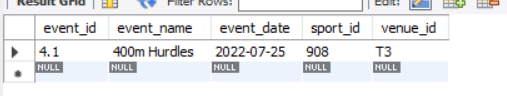
SELECT \* FROM Athletes WHERE country\_id = 'US';



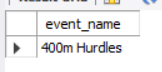
SELECT Countries.country\_name, COUNT(Athletes.athlete\_id) as total\_athletes FROM Countries JOIN Athletes ON Countries.country\_id = Athletes.country\_id GROUP BY Countries.country\_name;



SELECT \* FROM Events WHERE event\_date = '2022-07-25';



SELECT Events.event\_name FROM Events,Venues WHERE Venues.venue\_name = 'Tokyo Stadium' and Events.venue\_id = Venues.venue\_id;



SELECT Athletes.athlete\_name, Events.event\_name, Events.event\_date FROM Athletes, Events where Athletes.event\_id = Events.event\_id ORDER BY Events.event\_date;

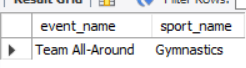


SELECT Events.event\_name, Venues.venue\_name FROM Events,Venues where Events.venue\_id = Venues.venue\_id;



SELECT Events.event\_name, Sports.sport\_name FROM Events JOIN Sports ON Events.sport\_id = Sports.sport\_id JOIN Venues ON Events.venue\_id = Venues.venue\_id

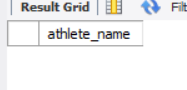
WHERE Venues.venue\_name = 'Saitama Super Arena';



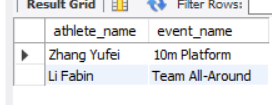
SELECT venue\_name, COUNT(\*) AS num\_events FROM Events JOIN Venues ON Events.venue\_id = Venues.venue\_id GROUP BY venue\_name;



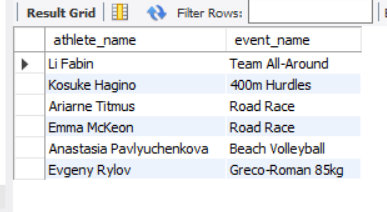
SELECT athlete\_name FROM Athletes JOIN Events ON Athletes.event\_id=Events.event\_id JOIN Sports ON Events.sport\_id=Sports.sport\_id WHERE sport\_name='Basketball' OR sport\_name='Soccer';



SELECT athlete\_name, event\_name FROM Athletes JOIN Events ON Athletes.event\_id=Events.event\_id JOIN Countries ON Athletes.country\_id=Countries.country\_id WHERE country\_name='China';



SELECT athlete\_name, event\_name FROM Athletes JOIN Events ON Athletes.event\_id=Events.event\_id WHERE event\_date BETWEEN '2022-07-25' AND '2022-07-30';



**VI. Project demonstration**

MySQL Workbench: MySQL Workbench is a visual database design tool that integrates SQL development, administration, database design, creation and maintenance into a single integrated development environment for the MySQL database system.

ERDPlus: ERDPlus is a user-friendly online tool designed for creating and modeling Entity-Relationship Diagrams (ERDs). These diagrams serve as a visual representation of the relationships between different entities within a database. ERDPlus simplifies the process of designing databases by allowing users to easily create, edit, and share ERDs without the need for complex software or technical expertise..

**VII. Self -Learning beyond classroom**

:

1. Data Science and Analytics: Start by learning the basics of data science, including collecting, cleansing, analyzing, and visualizing data. Dig deeper into machine learning techniques for predictive modeling and pattern recognition that can be used to analyze athlete performance trends and predict medal outcomes.
2. Database Management: Learn about database management systems (DBMS), such as SQL and NoSQL, and their role in efficiently storing, retrieving, and manipulating data. Understand concepts such as data normalization, indexing, and queries that are important for designing and optimizing databases, such as the 2020 Olympic Database.
3. Software Development: Gain knowledge of software development methodologies, tools, and best practices.
4. Explore concepts such as version control using Git, agile development methodologies, and collaborative coding using platforms such as GitHub.
5. Sports Analytics: Immerse yourself in the field of Sports Analytics. We combine statistical analysis, machine learning, and expertise to provide insights into athlete performance, strategy optimization, and match results. Discover the use of key performance indicators (KPIs) across a variety of sports, advanced statistical models for player evaluation, and tracking data for tactical analysis.
6. Ethical and legal considerations: Consider the ethical and legal implications of working with sensitive data, especially in the context of sports analytics. Learn about privacy regulations such as GDPR and HIPAA, and explore ethical frameworks for responsible data collection, use, and sharing. Understand the importance of transparency, fairness, and accountability in data-driven decision-making.
7. Communication and Visualization: Develop effective communication skills to convey complex technical concepts and insights to diverse audiences. Learn about data storytelling techniques, effective visualization principles. Practice presenting your results and recommendations clearly, concisely, and persuasively.
8. Continuous Learning and Collaboration: Adopt a mindset of lifelong learning and collaboration and seek opportunities to expand your knowledge and skills in areas related to the 2020 Olympic Database Project.
9. Join online community, attend conferences and workshops, collaborate with colleagues and leaders to share ideas, solve problems, and stay up to date on the latest developments in data science and sports analytics.

**VIII. Learning from the Project**

Project offers several benefits and ways it can help us:

1. Athlete performance insights: By analyzing data on athletes' biometrics, training regimens and performance results, the project can provide valuable insights about the factors that contribute to sports success. This information can be used by coaches and athletes to optimize training programs and improve performance.
2. Strategic Decision Making: By providing comprehensive data on event logistics, resource allocation and public feedback, the project enables organizers to make informed decisions to improve improve the overall experience of the Olympics. This includes decisions about location selection, programming, and marketing strategy.
3. Advances in Sports Analytics: The project contributes to the growing field of sports analytics by providing data sets to researchers and research analysts. Insights gained from data can lead to advances in performance analysis, and strategic decision-making in sports.
4. Heritage and Documentation: The database provides a valuable resource for recording the history of the 2020 Olympic Games, preserving data on athletes' achievements, event results and impacts cultural heritage for future generations. This helps create a lasting legacy for the Games and provides a basis for future research and analysis.
5. Technological innovation: Database development involves leveraging advanced technologies and methods in data science, database management, and software development. This contributes to technological innovation and advances in areas such as data processing, analysis and visualization.

**IX. Challenges Faced**

Challenges faced by us:

Information Volume and Assortment: Overseeing and preparing the endless sum of information created by the Olympics, counting competitor execution measurements, occasion plans, and social media intelligent, displayed a noteworthy challenge. Taking care of different information sorts and guaranteeing information quality were significant but complex errands.  
  
Specialized Framework: Building and keeping up the specialized foundation required to back the database, and capacity frameworks, postured calculated and asset challenges. Guaranteeing versatility, unwavering quality, and execution were basic to taking care of the requests of the Olympics.  
  
Integration of Information Sources: Joining information from different sources, counting official Olympic databases, and outside information, displayed integration challenges. Guaranteeing consistency, precision, and compatibility over dissimilar information sources required cautious coordination and specialized skill.  
  
Complexity of Olympic Occasions: The assorted nature of Olympic occasions, traversing different sports, disciplines, and scenes, included complexity to the database extend. Creating comprehensive information models and patterns able of capturing the complexities of each occasion whereas keeping up consistency over the database was a noteworthy challenge.

**X. Conclusion**

The database designed and implemented in this project provides a comprehensive solution for managing the data related to the athletes, countries, sports, events, and venues involved in the 2020 Tokyo Olympics. The database has been constructed with four main tables: Countries, Athletes, Sports, and Events, each containing their specific attributes and relationships with each other. The Venues table has also been added to provide further details about the event locations.

The Countries table holds data related to all the countries participating in the Olympics. Each country has been assigned a unique identifier, and their names have been stored in this table. The Athletes table contains information about all the athletes participating in the games, including their names, country of origin, and the events in which they will participate.

The Sports table stores information about the sports included in the games, including their unique identifiers and names. The Events table provides information about the specific events that will take place during the games, such as their unique identifiers, names, dates, and locations.

Finally, the Venues table contains data about the locations where the events will take place, including their unique identifiers and names.

The relationships between these tables are crucial to the functioning of the database, as they allow for efficient and accurate retrieval of data. For instance, the Athletes table has a foreign key reference to the Countries table, allowing for easy retrieval of an athlete's country of origin. Similarly, the Events table has foreign key references to the Sports and Venues tables, allowing for efficient queries regarding the sport and location of a specific event.

In conclusion, this database provides a well-structured and organized solution for managing the data related to the 2020 Tokyo Olympics. Its design allows for efficient retrieval and management of data, making it an excellent tool for anyone involved in organizing, managing, or reporting on the games.

**References:**

* [**https://en.m.wikipedia.org/wiki/2020\_Summer\_Olympics**](https://en.m.wikipedia.org/wiki/2020_Summer_Olympics)
* [**https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-intokyo**](https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo)
* [**https://www.bbc.com/sport/olympics/58109921**](https://www.bbc.com/sport/olympics/58109921)
* [**https://olympics.com/en/olympic-games/tokyo-2020**](https://olympics.com/en/olympic-games/tokyo-2020)