Data Book: Part 1, Database Narrative

Apoorv Awasthi

Indiana University-Purdue University, Indianapolis, IN 46202, USA

1. Description:

When we think of any sport, we usually think of something that involves physical activity. Sports is always thought as something to test physical ability of a person. But what about the mental ability? Many sports do include tactics and strategies to achieve results but chess is one of the few games that solely tests the mental ability of a player. Chess is often referred to as the battle of the minds where one player makes strategies, use tactics and prior knowledge to outwit their opponent.

Chess is a game that is played all around the world. Most chess tournaments that are organized have players from variety of countries come and participate. There are lot of casual and amateur players that play the game as well. Lot of online chess platforms have risen over the last decade to facilitate the players to learn as well as play the game online. During the lockdown in 2020 due to COVID-19, chess gained immense popularity on Twitch and YouTube. This resulted in many new people picking up chess as a game to relieve the stress caused due to the lockdown. As more and more people are coming to chess, chess data is increasing as well. So I thought to design a chess database.

I have prepared the chess database. In this database, there are 10 entities - PLAYER, COACH, GAME, TIME_CONTROL, OPENINGS, TOURNAMENTS, TITLE, PLATFORM (ONLINE, OFFLINE), ENGINE and ANALYSIS.

PLAYER - This entity will have all the details of the players playing chess be it casual or professional.
COACH - This entity will have the details of the coaches of players playing chess.
GAME - This entity will have details of all the chess games taking place all over the world as well as in online platforms.
TIME_CONTROL - This entity will give details of all the time controls used in the game of chess. Basically, each chess game is timed. Both players are given fixed time at the start of the game and they have to make all their moves in the allotted time.

OPENINGS - This entity will have the details of all the different openings in chess. Opening in chess is basically a combination of starting moves in a chess game. Different openings will lead to different types of games (Tactical, complicated or positional)
TOURNAMENTS - This entity will have all the details about various tournaments in the world of chess. Most games in chess are played in a tournament.
TITLE - This entity contains the details of all the titles or awards or trophies that have been earned by the players. For example - World championship title, world cup title, world U20 champion etc.
PLATFORM - This entity contains the details of all the platforms in which chess is played be it online or offline. This is a supertype for two subtypes - ONLINE and OFFLINE.
ENGINE - This entity contains the details of all the engines used in the game of chess. A chess engine is basically a computer program designed to solve chess problems and evaluate a position and generate the best move in that position. It is a great tool for chess analysis.
ANALYSIS - This table contains the details of analysis of chess games. Chess engines are used for analysis. The analysis will give all the details of the game. It will also point out the mistakes made in the game.

As the no. of players are increasing day by day, the chess data that is being generated also is increasing tremendously. Chess data usually includes the details of the players playing, the details of the game and the analysis of the game. This data can be used by players to prepare against their opponent in advance by looking at their previous games and learning about their styles. They can also see their own mistakes and can improve upon them. This data comes to good use to the tournament organizers as well. For example - they can invite the players with a lower draw rate and aggressive styles to promote fighting chess (when players play for a win regardless of the situation). The data will also help the online chess platforms to get a feedback based on the player count and will help them to improve their platform to attract more players. Chess journalists will also be able to use this data and can write their points with facts.

As the computing power is increasing day by day, there are many developers trying to design new versions of the engines that are required in chess for analysis. Notable examples of few strongest chess engines are - Stockfish NNUE, Leela chess Zero, Alpha Zero, Houdini, Komodo etc. The developers of these engines also require the chess data to feed in some prior knowledge to the engine so that it can understand few patterns and positions better. This will reduce the computing time and will make the engine more efficient. Development of engines will eventually give humans more correct perspective and will improve the level of top level chess games as well.

To sum it up, we see that there are a lot of end users of the chess database. The players, their coaches, the tournament organizers, chess platforms, chess journalists as well as the chess engine developers are some notable end users. The need for this data for these end users has given me motivation to design the ultimate chess database that can help players to learn, play and to eventually grow the game of chess

2. Data Dictionary:

ENTITY	ATTRIBUTE NAME	CONTENTS	DATA TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FOREIGN KEY REFERENCED TABLE
PLAYER	PLAYER_ID	Player ID	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	PLAYER_FNAME	Player First Name	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	PLAYER_LNAME	Player Last Name	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	PLAYER_DOB	Player's date of birth	DATE	yyyy-mm-dd		Y		
	PLAYER_RATING	Player's chess rating	CHAR(4)	9999	0000-9999	Y		
	TOTAL_GAMES	Total games played	CHAR(6)	999999	000000-999999	Y		
	PLAYER_WINPERCENT	Player Win percent	CHAR(3)	999	000-999	Y		
	PLAYER_DRAWPERCENT	Player draw percent	CHAR(3)	999	000-999	Y		
	PLAYER_LOSSPERSCET	Player loss percent	CHAR(3)	999	000-999	Y		
	COACH_ID	Coach ID	VARCHAR(10)	XXXXXXXXXX			FK	COACH
COACH	COACH_ID	Coach ID	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	COACH_FNAME	Coach's First Name	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	COACH_LNAME	Coach's Last Name	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		

ENTITY	ATTRIBUTE NAME	CONTENTS	DATA TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FOREIGN KEY REFERENCED TABLE
	COACH_DOB	Coach's date of birth	DATE	yyyy-mm-dd		Y		
	COACH_EXP	Coach's Experience in terms of no. of years	CHAR(2)	99	00-99			
GAMES	GAME_ID	Game ID	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	OPENING_ID	Opening ID	VARCHAR(10)	XXXXXXXXXX		Y	FK	OPENINGS
	GAME_PGN	Portable Game Notation (Moves of the game)	VARCHAR(1000)	Xxxxxxxxxxxxxxxxxxx		Y		
	GAME_DATE	Date of the game	DATE	yyyy-mm-dd		Y		
	WHITE_PLAYER_ID	Player ID of player playing white	VARCHAR(10)	XXXXXXXXXX		Y	FK	PLAYER
	BLACK_PLAYER_ID	Player ID of player playing black	VARCHAR(10)	XXXXXXXXXX		Y	FK	PLAYER
	GAME_RESULT	Game result	CHAR(3)	9-9		Y		
	TIME_CONTROL_ID	ID of the time control of the game (Blitz, Rapid, Classical)	VARCHAR(10)	XXXXXXXXXX		Y	FK	TIME_CONTROL
	TOURNAMENT_ID	Tournament ID in which the game is played	VARCHAR(10)	XXXXXXXXXX			FK	TOURNAMENTS
	ANALYSIS_ID	Analysis ID of games	VARCHAR(10)	XXXXXXXXXX			FK	ANALYSIS
TIME_CONTROL	TIME_CONTROL_ID	ID of the time control of the game (Blitz, Rapid, Classical)	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	TIME_CONTROL_TYPE	Type of the game(Rapid, blitz, classical, Bullet)	VARCHAR(10)	Xxxxxxxxxx		Y		
	TIME_CONTROL_TIME	Total Time of the game	CHAR(3)	999	000-999	Y		
	TIME_CONTROL_INCREMENT	Time increment per move	CHAR(2)	99	00-99			
	TIME_CONTROL_DESCRIPTION	Description	VARCHAR(1000)	Xxxxxxxxxxxxxxxxxxxx				

ENTITY	ATTRIBUTE NAME	CONTENTS	DATA TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FOREIGN KEY REFERENCED TABLE
OPENINGS	OPENING_ID	Opening ID	VARCHAR(10)	XXXXXXXXX		Y	PK	
	OPENING_NAME	Name of the opening played	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	OPENING_PGN	Portable Game Notation of the opening	VARCHAR(1000)	Xxxxxxxxxxxxxxxxxxx		Y		
	WHITE_WIN_PERCENT	Win Percentage of white playing the opening	CHAR(3)	999	000-999	Y		
	BLACK_WIN_PERCENT	Win Percentage of Black playing the opening	CHAR(3)	999	000-999	Y		
	DRAW_PERCENT	Draw percentage while playing the opening	CHAR(3)	999	000-999	Y		
TOURNAMENTS	TOURNAMENT_ID	Tournament ID in which the game is played	VARCHAR(10)	XXXXXXXXX		Y	PK	
	TOURNAMENT_NAME	Name of the tournament	VARCHAR(20)	Xxxxxxxxxxxxxxxxxxxx		Y		
	TOURNAMENT_YEAR	Date in which the tournament was held	DATE	yyyy-mm-dd		Y		
	TIME_CONTROL_ID	Time control of the games in the tournament	VARCHAR(10)	XXXXXXXXX		Y	FK	TIME_CONTROL
	PLATFORM_ID	ID of the platform in which the game is being played	VARCHAR(10)	XXXXXXXXX		Y	FK	PLATFORM
TITLE	TITLE_ID	ID of the title	VARCHAR(10)	XXXXXXXXX		Y	PK	
	TITLE_NAME	Name of the Title	VARCHAR(20)	Xxxxxxxxxxxxxxxxxxxx		Y		
	TITLE_YEAR	Date of the title	DATE	yyyy-mm-dd		Y		
	PLAYER_ID	Player ID of player who achieved this title	VARCHAR(10)	XXXXXXXXX		Y	FK	PLAYER
	TIME_CONTROL_ID	Time control of the games required to win the title	VARCHAR(10)	XXXXXXXXX		Y	FK	TIME_CONTROL
PLATFORM	PLATFORM_ID	ID of the platform in which the game is being played	VARCHAR(10)	XXXXXXXXX		Y	PK	

ENTITY	ATTRIBUTE NAME	CONTENTS	DATA TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FOREIGN KEY REFERENCED TABLE
	PLATFORM_NAME	Name of the online platform	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	PLATFORM_TYPE	Platform type (online/offline)	VARCHAR(8)	Xxxxxxxx		Y		
ONLINE	PLATFORM_ID	ID of the platform in which the game is being played	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	ONLINE_PLATFORM_LAUNCH	Launch date of the platform	DATE	yyyy-mm-dd		Y		
	ONLINE_PLATFORM_OWNER	Owner of the platform	VARCHAR(20)	Xxxxxxxxxxxxxxxxxxxx				
OFFLINE	PLATFORM_ID	ID of the platform in which the game is being played	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	OFFLINE_PLATFORM_DOB	Date of establishment of the platform	DATE	yyyy-mm-dd		Y		
	OFFLINE_PLATFORM_DIRECTOR	Name of the director of the platform	VARCHAR(20)	Xxxxxxxxxxxxxxxxxxxx				
ENGINE	ENGINE_ID	ID of the engine used for analysis	VARCHAR(10)	XXXXXXXXXX		Y	PK	ENGINE
	ENGINE_NAME	Name of the engine	VARCHAR(15)	Xxxxxxxxxxxxxx		Y		
	RATING	Rating of the Engine used	CHAR(4)	9999	0000-9999	Y		
	ENGINE_LAUNCH_DATE	Launch date of the engine	DATE	yyyy-mm-dd		Y		
	ENGINE_TYPE	Type of engine used	VARCHAR(10)	Xxxxxxxxxx		Y		
ANALYSIS	ANALYSIS_ID	Analysis ID	VARCHAR(10)	XXXXXXXXXX		Y	PK	
	ENGINE_ID	ID of the engine used for analysis	VARCHAR(10)	XXXXXXXXXX		Y	FK	ENGINE
	ANALYSIS_DEPTH	Depth of game analysis	CHAR(2)	99	00-99	Y		
	BEST_MOVES	No. Of best moves	CHAR(3)	999	000-999	Y		
	GOOD_MOVES	No. of good moves	CHAR(3)	999	000-999	Y		

ENTITY	ATTRIBUTE NAME	CONTENTS	DATA TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FOREIGN KEY REFERENCED TABLE
	INNACURACY	No. of Innaccuracies	CHAR(3)	999	000-999	Y		
	MISTAKES	No. of mistakes	CHAR(3)	999	000-999	Y		
	BLUNDERS	No. of Blunders	CHAR(3)	999	000-999	Y		
	FINAL_EVAL	Engine Evaluation of final position	VARCHAR(4)	XXXX				
	ANALYSIS_DESCRIPTION	Game analysis description	VARCHAR(1000)	Xxxxxxxxxxxxxxxxxxxx				

3. Business Rules

RELATIONSHIP	TYPE OF RELATIONSHIP	BI-DIRECTIONAL BUSINESS RULE
PLAYER - COACH	(1:M)	One player may be coached by one coach. One coach may coach many players
GAME - PLAYER	(M:N)	One game is played by two players. One player plays many games.
GAME - TIME_CONTROL	(1:M)	One game has one time control. One time control can be used in many games
GAME - TOURNAMENTS	(1:M)	One game is included in one tournament. One tournament includes many games.
GAME - ANALYSIS	(1:1)	One game has one analysis. One analysis can be done for one game.
GAME - OPENINGS	(1:M)	One opening can be played in many games. One game can have one opening
TOURNAMENTS - TIME_CONTROL	(1:M)	One tournament can have one time control. One time control can be used in many tournaments
TOURNAMENTS - PLATFORM	(1:M)	One tournament can be hosted in one platform. One platform can host many tournaments.

TITLE - PLAYER	(1:M)	One title belongs to one player. One player may have many titles.
TITLE - TIME_CONTROL	(1:M)	One title has one time control. One time control can be used for many titles.
ANALYSIS - ENGINE	(M:N)	One analysis can be done by many engines. One engine can do many analysis.

4. Entity Relationship Model components:

ENTITY	RELATIONSHIP	CONNECTIVITY	ENTITY
COACH	Coaches	1:M	PLAYER
PLAYER	Plays	M:N	GAME
TIME_CONTROL	Used in	1:M	GAME
TOURNAMENTS	Includes	1:M	GAME
GAME	Has	1:1	ANALYSIS
OPENINGS	Played in	1:M	GAME
TIME_CONTROL	Used in	1:M	TOURNAMENTS
PLATFORM	Host	1:M	TOURNAMENTS
PLAYER	Has	1:M	TITLE
TIME_CONTROL	Used for	1:M	TITLE
ANALYSIS	Done by	M:N	ENGINE

DATA BOOK Part 2,

By Apoorv Awasthi

Relational schema of entities in the Chess Database

PLAYER(**PLAYER_ID**, PLAYER_FNAME, PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING,TOTAL_GAMES, PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT, COACH_ID)

COACH(**COACH_ID**, COACH_FNAME, COACH_LNAME, COACH_DOB, COACH_EXP)

GAMES(**GAME_ID**, OPENING_ID, GAME_PGN, GAME_DATE, WHITE_PLAYER_ID, BLACK_PLAYER_ID, GAME_RESULT,

TIME_CONTROL_ID, TOURNAMENT_ID, ANALYSIS_ID)

TIME_CONTROL(**TIME_CONTROL_ID**, TIME_CONTROL_TYPE, TIME_CONTROL_TIME, TIME_CONTROL_INCREMENT, TIME_CONTROL_DESCRIPTION)

OPENINGS(**OPENING_ID**, OPENING_NAME, OPENING_PGN, WHITE_WIN_PERCENT, BLACK_WIN_PERCENT, DRAW_PERCENT)

TOURNAMENTS(**TOURNAMENT_ID**, TOURNAMENT_NAME, TOURNAMENT_YEAR, TIME_CONTROL_ID, PLATFORM_ID)

TITLE(**TITLE_ID**, TITLE_NAME, TITLE_YEAR, PLAYER_ID, TIME_CONTROL_ID)

PLATFORM(**PLATFORM_ID**, PLATFORM_NAME, PLATFORM_TYPE)

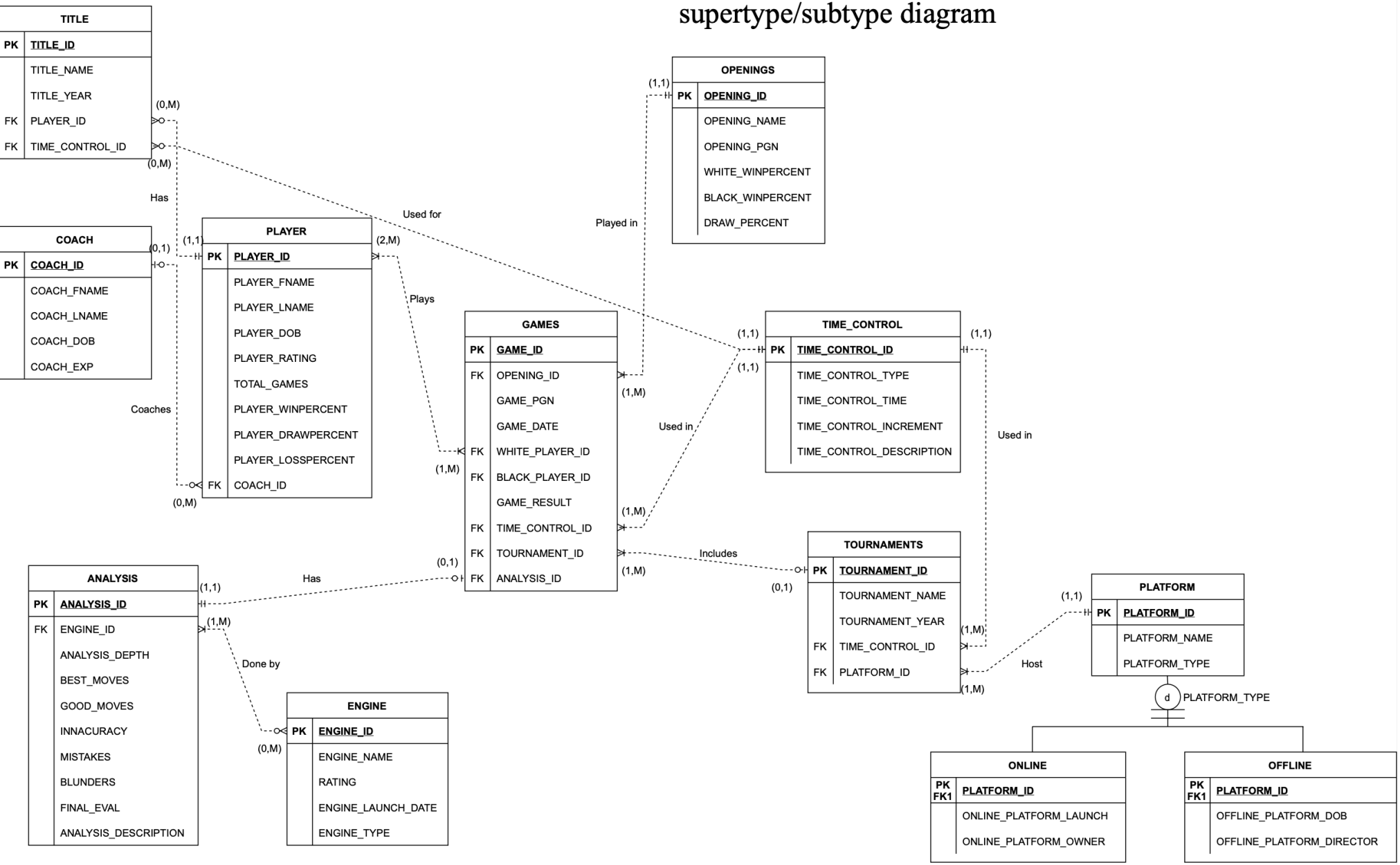
ONLINE(**PLATFORM_ID**, ONLINE PLATFORM_LAUNCH, ONLINE_PLATFORM_OWNER)

OFFLINE(**PLATFORM_ID**, OFFLINE PLATFORM_DOB, OFFLINE_PLATFORM_DIRECTOR)

ENGINE(**ENGINE_ID**, ENGINE_NAME, RATING, ENGINE_LAUNCH_DATE, ENGINE_TYPE)

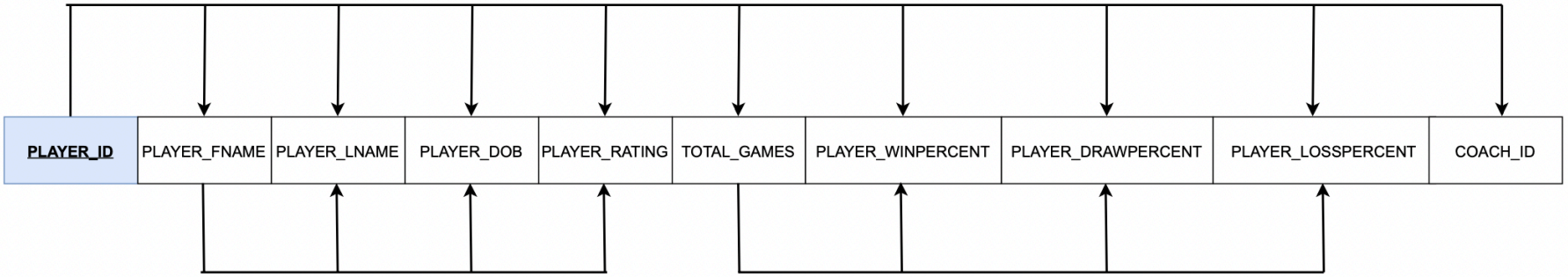
ANALYSIS(**ANALYSIS_ID**, ENGINE_ID, ANALYSIS_DEPTH, BEST_MOVES, GOOD_MOVES, INACCURACY, MISTAKES, BLUNDERS, FINAL_EVAL, ANALYSIS_DESCRIPTION)

ERM of the chess databook with supertype/subtype diagram



Normalization of the Table: PLAYER

1NF:



Relational Schema

1NF:(**PLAYER_ID**, PLAYER_FNAME, PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING,TOTAL_GAMES, PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT, COACH_ID)

Transitive Dependencies:(PLAYER_FNAME--> PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING),(TOTAL_GAMES--> PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT)

2NF:

Same as 1NF as there are no partial dependencies due to the absence of composite primary key (only single valued primary key is present

2NF:(**PLAYER_ID**, PLAYER_FNAME, PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING,TOTAL_GAMES, PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT, COACH_ID)

Transitive Dependencies:(PLAYER_FNAME--> PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING),(TOTAL_GAMES--> PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT)

3NF:

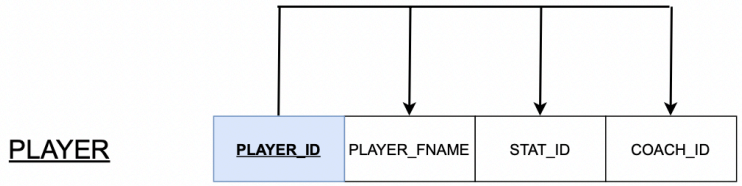
3NF Relationship Schema:

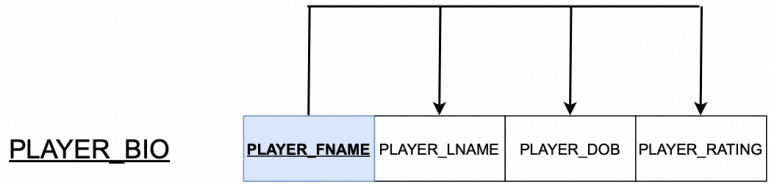
PLAYER(**PLAYER_ID**, PLAYER_FNAME,STAT_ID, COACH_ID)

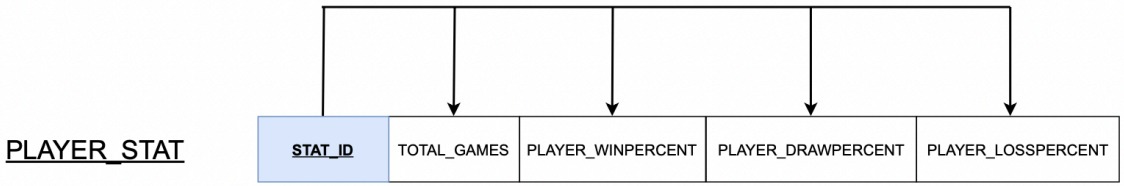
PLAYER_BIO(**PLAYER_FNAME**, PLAYER_LNAME, PLAYER_DOB, PLAYER_RATING)

PLAYER_STATS(**STAT_ID**, TOTAL_GAMES, PLAYER_WINPERCENT, PLAYER_DRAWPERCENT, PLAYER_LOSSPERCENT)

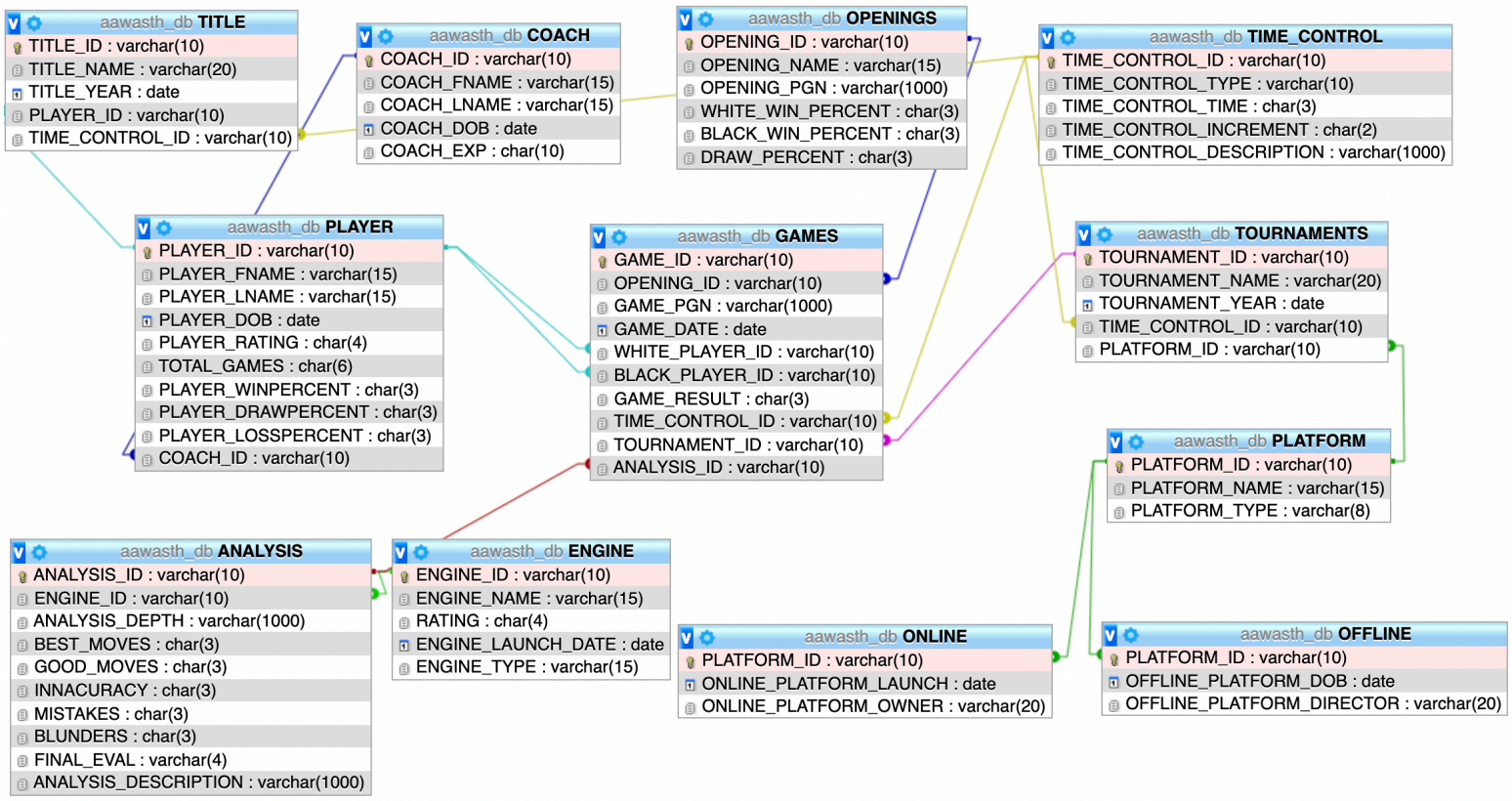
Note: Added STAT_ID in the PLAYER entity as a Foreign key and added it to the PLAYER_STAT entity as a Primary key to avoid anomalies and for referential integrity







Relationship of entities



Queries:

Question 1:
What are the details of the ‘STOCKFISH NNUE’ engine.

Query:
```
SELECT * FROM ENGINE WHERE ENGINE_NAME = 'STOCKFISH NNUE';
```

Question 2:
What is the information of the games played in the ‘world championship’.

Query:
```
SELECT GAMES.* FROM GAMES INNER JOIN TOURNAMENTS ON GAMES.TOURNAMENT_ID =
TOURNAMENTS.TOURNAMENT_ID WHERE
TOURNAMENT_NAME = 'World Championship';
```

Question 3:
What are the games played by the player ‘Viswanathan Anand’
.

Query:
```
SELECT * FROM GAMES WHERE WHITE_PLAYER_ID = (SELECT PLAYER_ID
FROM PLAYER WHERE PLAYER_FNAME = 'Viswanathan' AND PLAYER_LNAME =
'Anand') OR BLACK_PLAYER_ID = (SELECT PLAYER_ID FROM PLAYER WHERE PLAYER_FNAME = 'Viswanathan' AND
PLAYER_LNAME = 'Anand');
```