



UNIVERZITET U NOVOM SADU
FAKULTET TEHNIČKIH NAUKA



Predmet: Projektovanje skladišta podataka

PROJEKAT

**Projektovanje skladišta podataka za set podataka o
fudbalskim utakmicama**

Profesor: dr Sonja Ristić

Asistent: Jelena Petrovački

Student: Filip Ilić IT7/2020

Novi Sad, 2024. godine

Sadržaj

1. Zadatak i ciljevi projekta	1
2. Opis postupka projektovanja DW sistema.....	2
3. Specifikacija zahteva korisnika.....	3
4. Specifikacija modela	4
4.1. Specifikacija izvora podataka.....	4
4.2. Specifikacija ciljanog <i>DataWarehouse</i> sistema	4
5. Opis ETL procesa.....	6
5.1. Kreiranje dimenzije <i>Stadium</i>	6
5.2. Kreiranje dimenzije <i>Club</i>	9
5.3. Kreiranje dimenzije <i>Competition</i>	12
5.4. Kreiranje dimenzije <i>Player</i>	13
5.5. Kreiranje dimenzije <i>Game</i>	15
5.6. Kreiranje dimenzije <i>Date</i>	17
5.7. Kreiranje tabele činjenica	19
6. Prikaz izveštaja.....	23
6.1. Izveštaj 1 - Koliko asistencija je imao fudbaler sa najviše asistencija u fudbalskom klubu Monako u sezoni 2021/22?	23
6.2. Izveštaj 2 - Koliko golova je postigao fudbaler Ante Rebić u dres fudbalskog Milan u sezoni 2020/21?	23
6.3. Izveštaj 3 - Koji fudbaleri su imali najviše žutih kartona za fudbalski klub Liverpool u sezoni 2019/20 u mesecu decembru?.....	24
6.4. Izveštaj 4 - Koliki je broj odigranih utakmica koje je fudbaler <i>Mohammed Salah</i> odigrao u sezoni 2018/19 u Ligi Šampiona?	25
6.5. Izveštaj 5 - Koji fudbaler je imao prosečno najviše odigranih minuta u fudbalskom klubu Borussia Dortmund u sezoni 2020/21?.....	26
7. Zaključak.....	27

Spisak slika

Slika 1 OLTP šema.....	4
Slika 2 OLAP šema	5
Slika 3 Script za kreiranje dimenzije Stadium.....	6
Slika 4 Odabir obeležja potrebnih za kreiranje dimenzije Stadium	7
Slika 5 Mapiranje obeležja shodno obeležjima u kreiranoj tabeli Stadium	8
Slika 6 Proces kreiranja dimenzije Stadium u Data Flow Task-u.....	8
Slika 7 Izgled tabele koja predstavlja dimenziju Stadium u bazi podataka	9
Slika 8 Script za kreiranje dimenzije Club.....	9
Slika 9 Odabir obeležja potrebnih za kreiranje dimenzije Club.....	10
Slika 10 Zamena nepostojećih vrednosti u dimenziji Club.....	10
Slika 11 Mapiranje obeležja shodno obeležjima u kreiranoj tabeli Club	11
Slika 12 Proces kreiranja dimenzije Club u Data Flow Task-u.....	11
Slika 13 Izgled tabele koja predstavlja dimenziju Club u bazi podataka	12
Slika 14 Script za kreiranje dimenzije Competition.....	12
Slika 15 Konverzija podataka prilikom kreiranja dimenzije Competition.....	13
Slika 16 Proces kreiranja dimenzije Competition u Data Flow Task-u.....	13
Slika 17 Izgled tabele koja predstavlja dimenziju Competition u bazi podataka	13
Slika 18 Script za kreiranje dimenzije Player	14
Slika 19 Uslov prilikom kreiranja dimenzije Player.....	14
Slika 20 Proces kreiranja dimenzije Player u Data Flow Task-u.....	14
Slika 21 Izgled tabele koja predstavlja dimenziju Player u bazi podataka.....	15
Slika 22 Script za kreiranje dimenzije Game	15
Slika 23 Mapiranje vrednosti stranog ključa u dimenziji Game pomoću komponente Lookup	16
Slika 24 Proces kreiranja dimenzije Game u Data Flow Task-u	16
Slika 25 Izgled tabele koja predstavlja dimenziju Game u bazi podataka	16
Slika 26 Script za kreiranje dimenzije Date	17
Slika 27 Uslov prilikom kreiranja dimenzije Date	17
Slika 28 Izvođenje novih kolona u dimenziji Date	18
Slika 29 Konverzija podataka prilikom kreiranja dimenzije Date	18
Slika 30 Proces kreiranja dimenzije Date u Data Flow Task-u	18
Slika 31 Izgled tabele koja predstavlja dimenziju Date u bazi podataka	19
Slika 32 Script za kreiranje tabele činjenica	19
Slika 33 Odabir mera za tabelu činjenica	20
Slika 34 Proces kreiranja tabele činjenica 1	20
Slika 35 Proces kreiranja tabele činjenica 2	21
Slika 36 Proces kreiranja tabele činjenica 3	21
Slika 37 Proces kreiranja tabele činjenica 4	21
Slika 38 Proces kreiranja tabele činjenica 5	22
Slika 39 Izgled tabele činjenica u bazi podataka	22

Slika 40 Izveštaj 1 – odgovor 11	23
Slika 41 Izveštaj 2 – odgovor 11	24
Slika 42 Izveštaj 3 – odgovor Fabinho i Virgil van Dijk.....	25
Slika 43 Izveštaj 4 – odgovor 12	25
Slika 44 Izveštaj 5 - odgovor Roman Burki	26

1. Zadatak i ciljevi projekta

Zadatak koji se postavlja u okviru ovog projekta je analiza i istraživanje podataka o fudbalskim utakmicama iz različitih fudbalskih liga širom sveta. Korišćen je set podataka koji se kontinuirano ažurira, omogućavajući pristup najnovijim informacijama na dnevnom nivou, ali i povećavajući rizik od grešaka u podacima.

Cilj projekta je višeslojan. Prvi korak je razumevanje osnovnih koncepata projektovanja skladišta podataka za izabrani set podataka o fudbalskim utakmicama. Ovo podrazumeva kreiranje šema koje su neophodne za realizaciju efikasnog i efektivnog skladišta podataka. Sledeći korak je sprovođenje ETL (*Extract, Transform, Load*) procesa, kojim će se podaci prikupiti, transformisati u potrebni format i učitati u skladište podataka.

Nakon završetka ETL procesa, fokus će biti na analizi podataka i kreiranju izveštaja. Cilj ovih izveštaja je da pruže odgovore na ključna pitanja vezana za fudbalske utakmice, kao što su statistika igrača i drugi relevantni parametri. Na ovaj način, izveštaji će omogućiti bolje razumevanje podataka i identifikaciju obrazaca koji mogu biti korisni za različite zainteresovane strane, uključujući trenere, analitičare, menadžere timova i ljubitelje fudbala.

Kroz detaljnu analizu i obradu podataka, ovaj projekat može značajno doprineti praćenju i unapređenju različitih aspekata fudbalskih liga, timova i strategija. Praćenje ključnih parametara može pomoći u donošenju informisanih odluka koje će doprineti efikasnosti i uspehu u sportu.

U nastavku će biti detaljno opisani sistem i osnovni elementi sistema koji se posmatraju (odnosno set podataka koji opisuje fudbalske utakmice širom sveta), ETL proces, pitanja na koja ćemo kroz analizu dati odgovore, kao i izveštaji koji predstavljaju završni segment izrade ovog projekta.

2. Opis postupka projektovanja DW sistema

Projektovanje skladišta podataka za kompleksne sisteme zahteva vreme, istraživanje i duboko razumevanje poslovnih procesa unutar sistema, kao i način na koji sistem funkcioniše. Potrebno je jasno definisati ciljeve na osnovu kojih se prikupljaju podaci, kako bi se na osnovu tih ciljeva mogli projektovati naredni segmenti projekta.

Proces projektovanja skladišta podataka obuhvata nekoliko ključnih koraka:

1. Istraživanje i prikupljanje podataka,
2. Analiza podataka,
3. Definisanje ciljeva,
4. Projektovanje i implementacije OLAP šeme,
5. ETL proces – izrada tabela dimenzija i tabela činjenica,
6. Izrada izveštaja i
7. Analiza dobijenih rezultata.

3. Specifikacija zahteva korisnika

Posmatrani set podataka je složen, te je bilo potrebno izdvojiti ključne stavke koje će biti u fokusu analize.

Na samom početku identifikovali smo nekoliko specifičnih pitanja na koja ćemo pokušati da odgovorimo kroz detaljnu analizu i istraživanje podataka. Pitanja su:

1. Koji fudbaler je imao prosečno najviše odigranih minuta u fudbalskom klubu Borussia Dortmund u sezoni 2020/21? *Roman Burki*
2. Koliko asistencija je imao fudbaler sa najviše asistencija u fudbalskom klubu Monaco u sezoni 2021/22? *11*
3. Koliko golova je postigao fudbaler Ante Rebić u dres fudbalskog Milan u sezoni 2020/21? *11*
4. Koji fudbaleri su imali najviše žutih kartona za fudbalski klub Liverpool u sezoni 2019/20 u mesecu decembru? *Fabinho, Virgil van Dijk*
5. Koliki je broj odigranih utakmica koje je fudbaler *Mohammed Salah* odigrao u sezoni 2018/19 u Ligi Šampiona? *12*

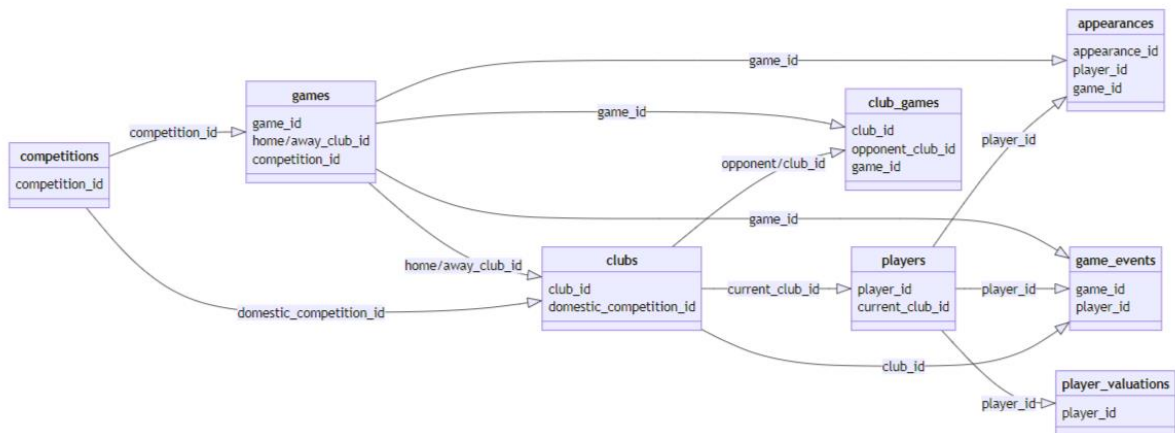
4. Specifikacija modela

4.1. Specifikacija izvora podataka

Izvorni podaci su incijalno preuzeti sa *Kaggle* sajta, na sledećem linku: <https://www.kaggle.com/davidcariboo/player-scores?select=appearances.csv>.

Podaci se nalaze u nekoliko .csv fajlova, koje su za potrebe projekta prebačeni u .txt format radi lakše obrade i smanjenja grešaka prilikom učitavanja podataka. Fajlovi su: *competition*, *game_events*, *games*, *club_games*, *appearances*, *players*, *clubs*, *player_valuations*.

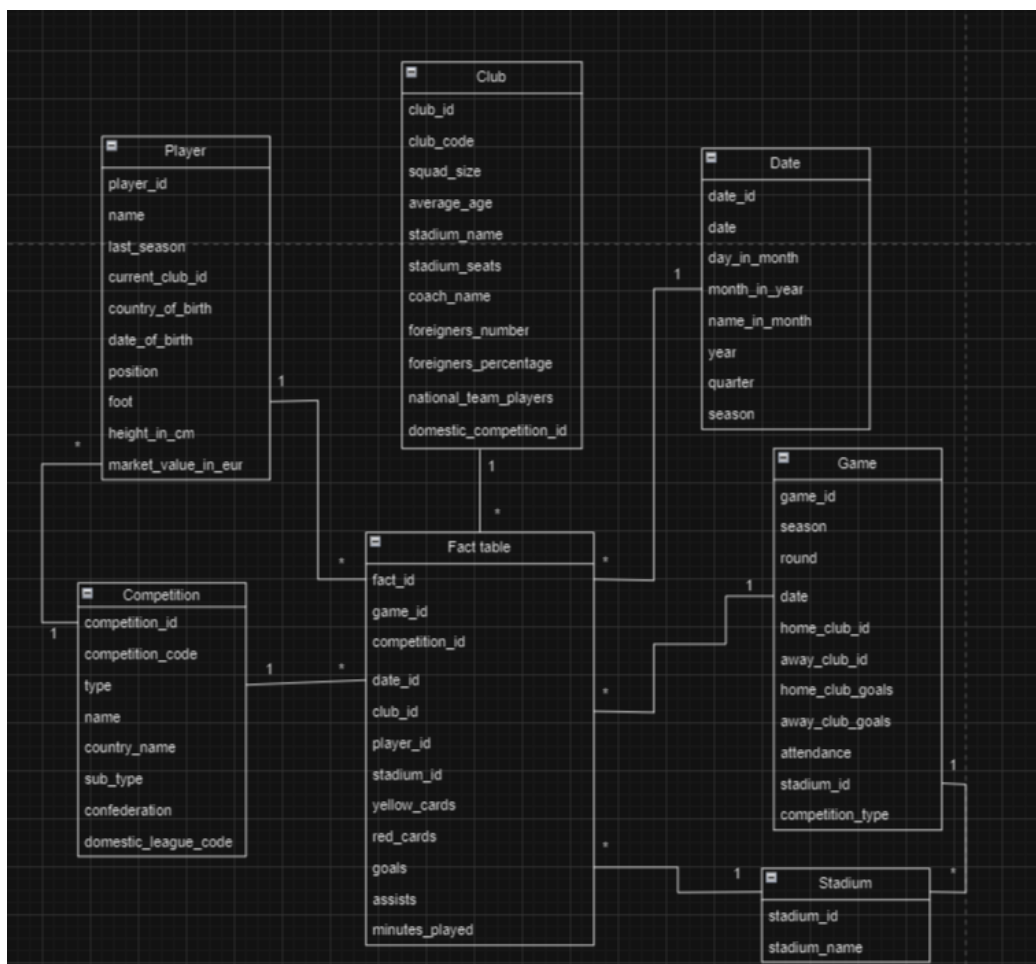
Na slici 1 prikazana je OLTP šema preuzeta iz izvora podataka.



Slika 1 OLTP šema

4.2. Specifikacija ciljanog *DataWarehouse* sistema

Na slici 2 prikazana je OLAP šema sa dimenzijama i tabelom činjenica.



Slika 2 OLAP šema

U okviru prikazane OLAP šeme izdvojene su sledeće dimenzije:

1. Dimenzija *Competition*,
2. Dimenzija *Game*,
3. Dimenzija *Stadium*,
4. Dimenzija *Player*,
5. Dimenzija *Club* i
6. Dimenzija *Date*.

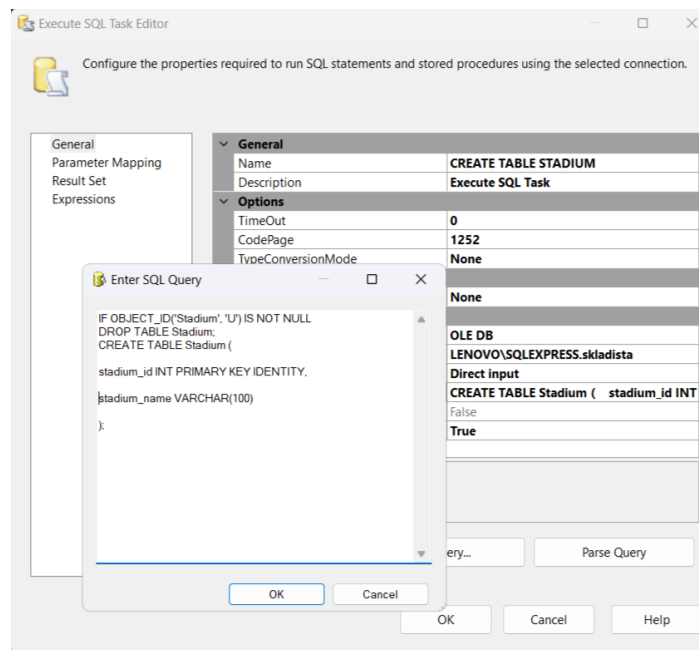
U okviru tabele činjenica izdvojeno je 5 mera, i to:

- Broj žutih kartona koji je jedan fudbaler zaradio na jednoj utakmici,
- Broj crvenih kartona koji je jedan fudbaler zaradio na jednoj utakmici,
- Broj golova koji je jedan fudbaler ostvario na jednoj utakmici,
- Broj asistencija koji je jedan fudbaler ostvario na jednoj utakmici i
- Broj odigranih minuta jednog fudbalera na jednoj utakmici.

5. Opis ETL procesa

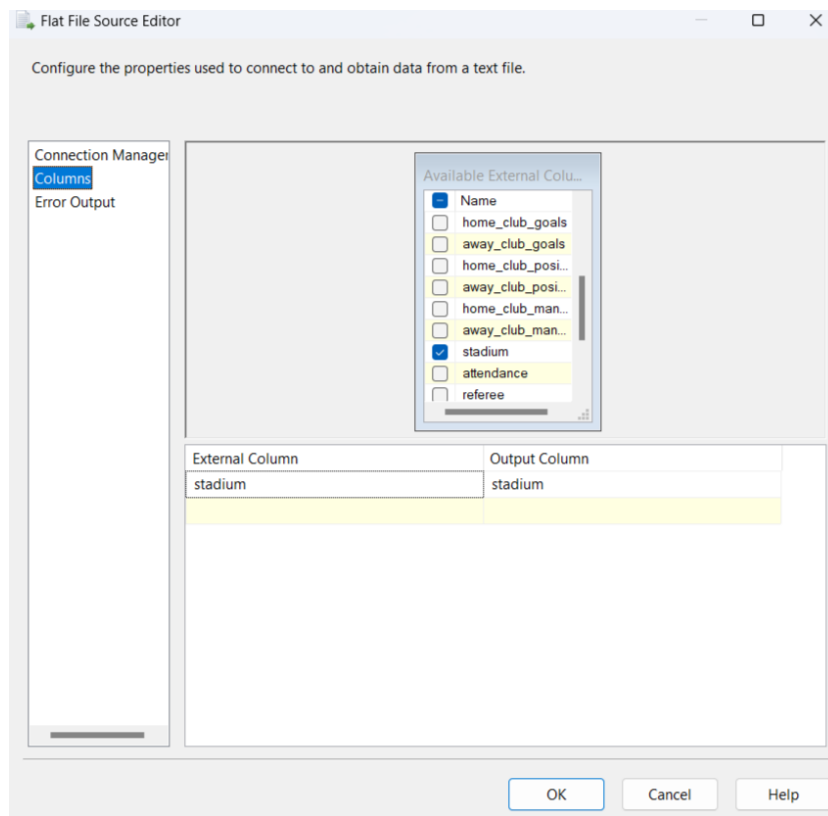
5.1. Kreiranje dimenzije *Stadium*

U okviru projekta prva kreirana dimenzija je dimenzija *Stadium*. Na samom početku prosleđen je script kako bi se kreirala tabela *skladiste.Stadium*. Pre samog kreiranja tabele vrši se provera da li već postoji tabela sa istim nazivom u bazi, ukoliko postoji ona biva izbrisana. Na slici 3 prikazan je *script* kojim se kreira dimenzija *Stadium*.



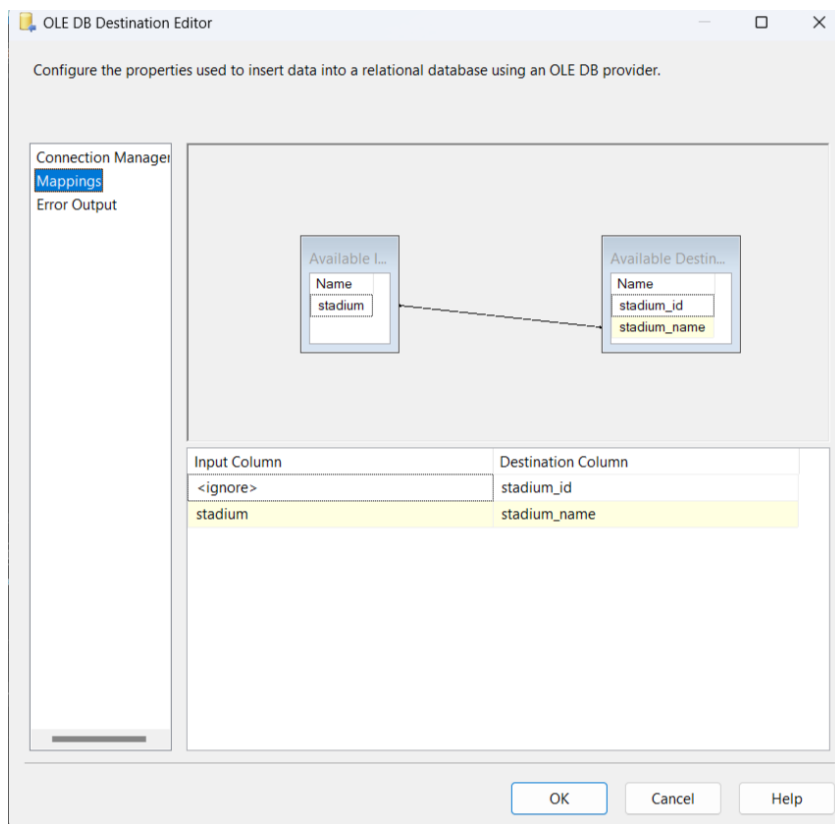
Slika 3 Script za kreiranje dimenzije *Stadium*

Nakon ubacivanja *.txt* fajla u kom se nalaze podaci, odabrana su obeležja potrebna za kreiranje dimenzije saglasno kreiranom *script*-u u prethodnom koraku u okviru *Execute SQL Task*-a. Na slici 4 prikazan je odabir obeležja potrebnih za kreiranje dimenzije *Stadium*.



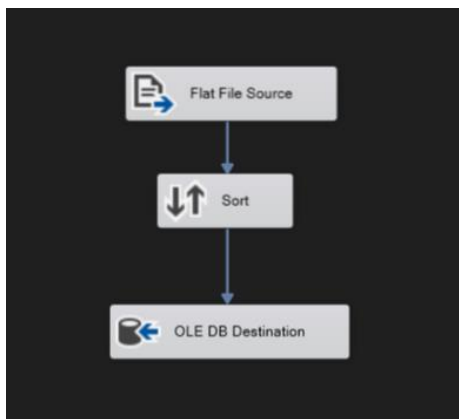
Slika 4 Odabir obeležja potrebnih za kreiranje dimenzije Stadium

Na kraju pre samog upisa podataka u bazu podataka, izvršeno je mapiranje obeležja gde će se upisati podaci. Program automatski prepoznava veze ukoliko su nazivi obeležja isti, a u suprotnom potrebno je manuelno mapirati obeležja. Na slici 5 prikazano je mapiranje obeležja shodno obeležjima u kreiranoj tabeli *Stadium*.



Slika 5 Mapiranje obeležja shodno obeležjima u kreiranoj tabeli Stadium

Na slici 6 vizuelno je predstavljen tok aktivnosti koji je izvršen prilikom kreiranja dimenzije *Stadium*.



Slika 6 Proces kreiranja dimenzije Stadium u Data Flow Task-u

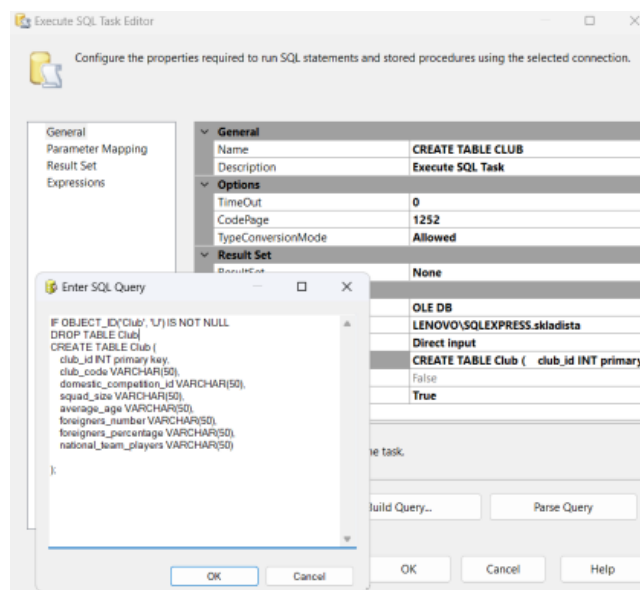
Nakon pokretanja ovog *Data Flow Task*-a podaci su uspešno upisani u bazu, te je samim tim uspešno kreirana dimenzija *Stadium*. Na slici 7 su prikazani upisani podaci.

	stadium_id	stadium_name
1	1	"BBSP Stadium"
2	2	"Capelli Sport Stadion"
3	3	"Comunale Renzo Barbera ""La Favorita"""
4	4	"Dimotiko Gipedo ""Gallou"""
5	5	"Dimotiko Gipedo Eginou ""Giannis Paralikidis"""
6	6	"Dimotiko Stadio ""Emilios Theofanidis"""
7	7	"Dimotiko Stadio Alexandroupolis ""Fotis Kosmas"""
8	8	"Dimotiko Stadio Lixouriou ""Agios Antonios"""
9	9	"Dimotiko Stadio Spaton ""Dimitris Dimitriou"""
10	10	"Gipedo ""Georgios Roumeliotis"""
11	11	"Gipedo ""Kleanthis Vikelidis"""
12	12	"Gipedo ""Pantelis Magoulas"""
13	13	"Gipedo Episkopis ""Theodoros Vardinogiannis"""
14	14	"Gipedo Evosmou ""Dimitris Kontikakis"""
15	15	"Gipedo Kesarianis ""Michalis Kritikopoulos"""
16	16	"Gipedo Panachaiki ""Kostas Davourlis"""
17	17	"Gipedo Rizoupolis ""Georgios Kamaras"""
18	18	"Gipedo Tavrou ""Spyros Gialabidis"""
19	19	"Kunst"
20	20	"Luzhniki"
21	21	"OlympiakAo StAadio Athinon ""Spyros Louis"""
22	22	"Pistas del HelmAantico ""Javier Sotomavor"""

Slika 7 Izgled tabele koja predstavlja dimenziju Stadium u bazi podataka

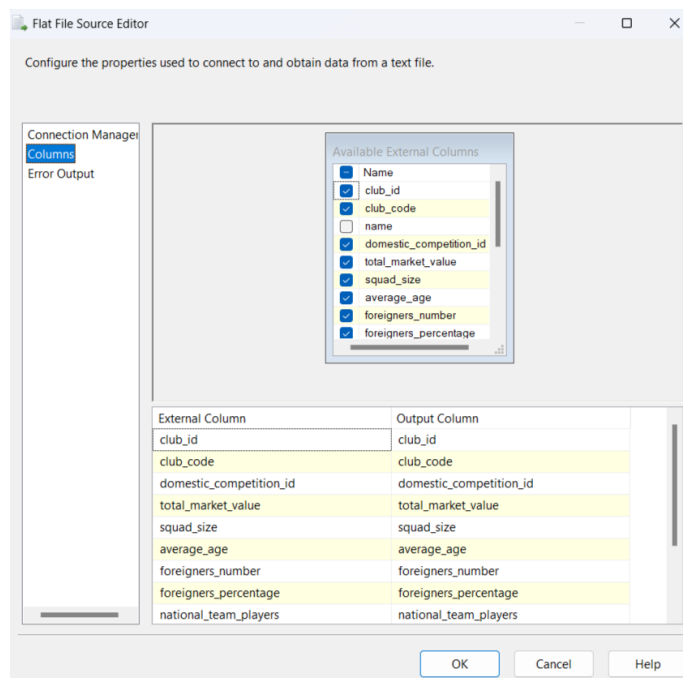
5.2. Kreiranje dimenzije Club

Analogno aktivnostima kojim je kreirana dimenzija Stadium, kreirana je i dimenzija Club. Na slici 8 prikazan je script kojim se kreira dimenzija Club.



Slika 8 Script za kreiranje dimenzije Club

Na slici 9 prikazan je odabir obeležja potrebnih za kreiranje dimenzije Stadium.



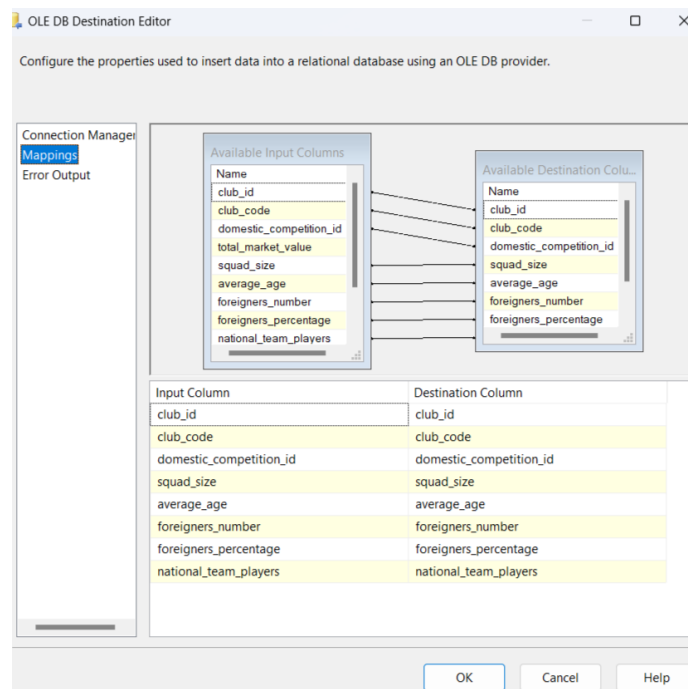
Slika 9 Odabir obeležja potrebnih za kreiranje dimenzije Club

Pošto u okviru kolona *average_age* i *foreigners_percentage* postoje vrednosti koje su *null*, zamenjene su im vrednosti tako da tamo gde ne postoji vrednost piše *Not available* uz pomoć komponente *Derived Column*. Na slici 10 prikazana je zamena vrednosti nad pomenutim kolonama.

Derived Column Name	Derived Column	Expression	Data
average_age	Replace 'average_age'	LEN(average_age) == 0 ? "Not available" : average_age	string
foreigners_percentage	Replace 'foreigners_...	LEN(foreigners_percentage) == 0 ? "Not available" : foreigners_perc...	string

Slika 10 Zamena nepostojećih vrednosti u dimenziji Club

Na slici 11 prikazano je mapiranje obeležja shodno obeležjima u kreiranoj tabeli *Club*.



Slika 11 Mapiranje obeležja shodno obeležjima u kreiranoj tabeli Club

Na slici 12 vizuelno je predstavljen tok aktivnosti koji je izvršen prilikom kreiranja dimenzije *Club*.



Slika 12 Proces kreiranja dimenzije Club u Data Flow Task-u

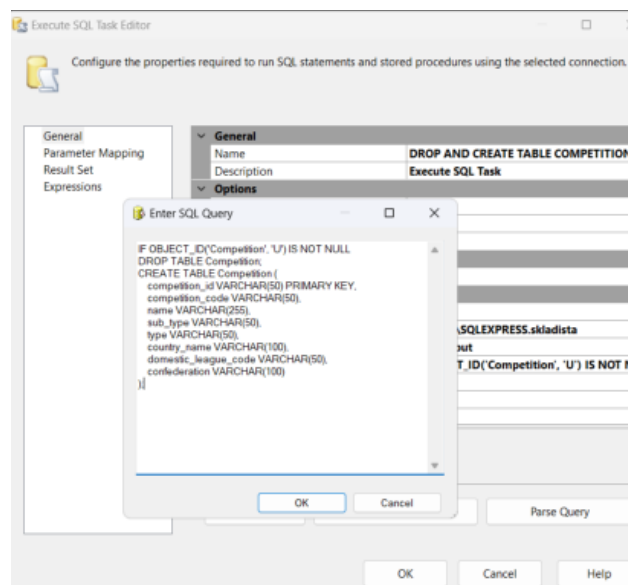
Na slici 13 su prikazani upisani podaci nakon pokretanja *Data Flow Task-a*.

	club_id	club_code	domestic_competition_id	squad_size	average_age	foreigners_number	foreigners_percentage	national_team_players
4	6	adanaspor	TR1	10	23.9	4	40.0	0
5	10	arminia-bielefeld	L1	27	25.3	15	55.6	4
6	11	fc-arsenal	GB1	25	25.7	18	72.0	20
7	12	as-rom	IT1	26	27.1	18	69.2	16
8	13	atletico-madrid	ES1	25	28.8	16	64.0	14
9	15	bayer-04-leverkusen	L1	27	24.9	22	81.5	20
10	16	borussia-dortmund	L1	29	25.5	13	44.8	16
11	18	borussia-monchengladbach	L1	31	25.7	16	51.6	8
12	19	brescia-calcio	IT1	28	25.1	12	42.9	1
13	20	bursaspor	TR1	26	21.4	0	Not available	0
14	23	eintracht-braunschweig	L1	30	25.6	11	36.7	2
15	24	eintracht-frankfurt	L1	34	24.5	20	58.8	8
16	27	fc-bayern-munchen	L1	27	26.8	16	59.3	20
17	29	fc-everton	GB1	24	27.5	11	45.8	7
18	31	fc-liverpool	GB1	28	25.8	21	75.0	18
19	33	fc-schalke-04	L1	27	26.4	12	44.4	4
20	36	fenerbahce-istanbul	TR1	27	28.6	16	59.3	16
21	38	fortuna-dusseldorf	L1	29	25.9	12	41.4	2
22	39	1-fsv-mainz-05	L1	27	26.7	11	40.7	6
23	40	fc-girondins-bordeaux	FR1	26	22.7	13	50.0	6
24	41	hamburger-sv	L1	30	23.6	15	50.0	4
25	42	hannover-96	L1	28	25.0	6	21.4	1

Slika 13 Izgled tabele koja predstavlja dimenziju Club u bazi podataka

5.3. Kreiranje dimenzije Competition

Analogno prethodnim aktivnostima, najpre je pokrenut *script* u okviru *Execute SQL Task*-a što je prikazano na slici 14.



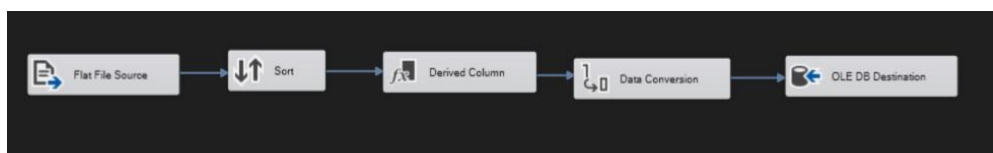
Slika 14 Script za kreiranje dimenzije Competition

U okviru narednih koraka izvršene su identične aktivnosti kao i za prethodne dimenzije, određena obeležja nisu imala vrednosti i pomoću komponente *Derived Column* umesto praznih polja upisane su vrednosti *Not available*. Nakon toga bilo je potrebno izvršiti konverziju tih podataka što je urađeno pomoću komponente *Data Conversion* što je prikazano na slici 15.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Country	Copy of Country	string [DT_STR]	50			1252 (ANS
Dom	Copy of Dom	string [DT_STR]	50			1252 (ANS

Slika 15 Konverzija podataka prilikom kreiranja dimenzije Competition

Pre samog pokretanja *Data Flow Task*-a izvršena su potrebna mapiranja. Na slici 16 prikazan je vizuelni pregled *Data Flow Task*-a.



Slika 16 Proces kreiranja dimenzije Competition u *Data Flow Task*-u

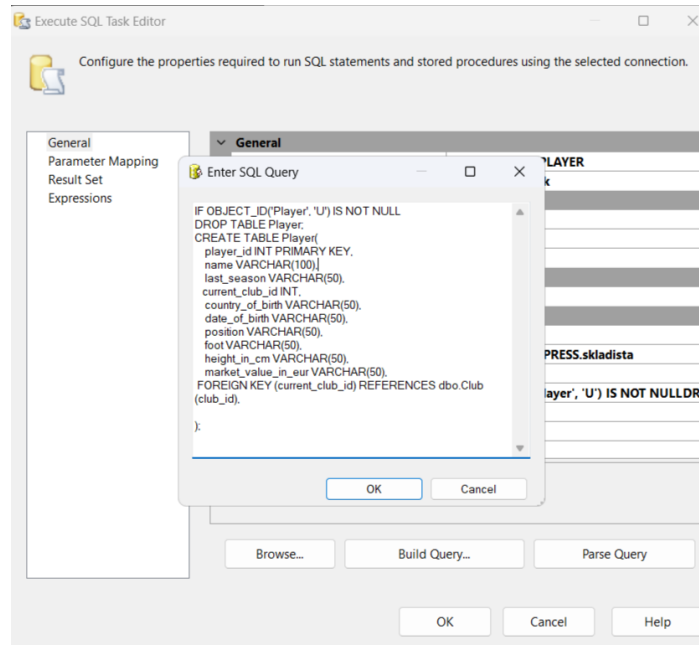
Na slici 17 su prikazani upisani podaci nakon pokretanja *Data Flow Task*-a.

	competition_id	competition_code	name	sub_type	type	country_name	domestic_league_code	confederation
1	BE1	jupiler-pro-league	jupiler-pro-league	first_tier	domestic_league	Belgium	BE1	europa
2	BESC	belgian-supercup	belgian-supercup	domestic_super_cup	other	Belgium	BE1	europa
3	CDR	copa-del-rey	copa-del-rey	domestic_cup	domestic_cup	Spain	ES1	europa
4	CGB	efl-cup	efl-cup	league_cup	other	England	GB1	europa
5	CIT	italy-cup	italy-cup	domestic_cup	domestic_cup	Italy	IT1	europa
6	CL	uefa-champions-league	uefa-champions-league	uefa_champions_league	international_cup	Not available	Not available	europa
7	CLQ	uefa-champions-league-qualifikation	uefa-champions-league-qualifikation	uefa_champions_league_qualifying	international_cup	Not available	Not available	europa
8	DFB	dfb-pokal	dfb-pokal	domestic_cup	domestic_cup	Germany	L1	europa
9	DFL	dfl-supercup	dfl-supercup	domestic_super_cup	other	Germany	L1	europa
10	DK1	superligaen	superligaen	first_tier	domestic_league	Denmark	DK1	europa
11	DKP	sydbank-pokalen	sydbank-pokalen	domestic_cup	domestic_cup	Denmark	DK1	europa
12	ECLQ	uefa-europa-conference-league-qualifikation	uefa-europa-conference-league-qualifikation	uefa_europa_conference_league_qualifiers	international_cup	Not available	Not available	europa
13	EL	europa-league	europa-league	europa_league	international_cup	Not available	Not available	europa
14	ELQ	europa-league-qualifikation	europa-league-qualifikation	europa_league_qualifying	international_cup	Not available	Not available	europa
15	ES1	laliga	laliga	first_tier	domestic_league	Spain	ES1	europa
16	FAC	fa-cup	fa-cup	domestic_cup	domestic_cup	England	GB1	europa
17	FR1	ligue-1	ligue-1	first_tier	domestic_league	France	FR1	europa
18	FRCH	trophee-des-champions	trophee-des-champions	domestic_super_cup	other	France	FR1	europa
19	GB1	premier-league	premier-league	first_tier	domestic_league	England	GB1	europa
20	GBCS	community-shield	community-shield	domestic_super_cup	other	England	GB1	europa
21	GR1	super-league-1	super-league-1	first_tier	domestic_league	Greece	GR1	europa
22	GRP	kyrillo-elladas	kyrillo-elladas	domestic_cup	domestic_cup	Greece	GR1	europa

Slika 17 Izgled tabele koja predstavlja dimenziju Competition u bazi podataka

5.4. Kreiranje dimenzije *Player*

Na slici 18 prikazan je *script* kojim se kreira dimenzija *Player* na kome se može videti da tabela sadrži strani ključ koji se odnosi na klub za koji nastupa fudbaler.



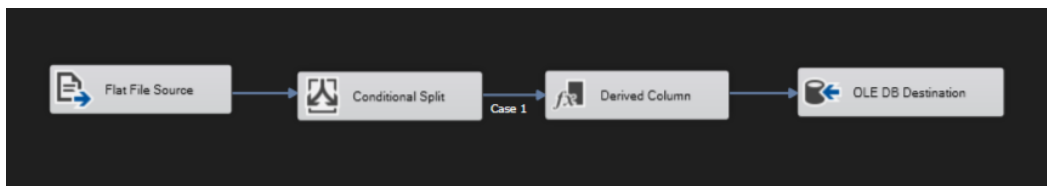
Slika 18 Script za kreiranje dimenzije Player

U okviru narednih koraka izvršene su identične aktivnosti kao i za prethodne dimenzije, određena obeležja nisu imala vrednosti i pomoću komponente *Derived Column* umesto praznih polja upisane su vrednosti *Not available*. Takođe, pre toga pomoću komponente *Conditional Split* je napravljen uslov na osnovu kog će se proveravati iz koje zemlje dolazi određeni fudbaler i na osnovu tog uslova će određeni fudbaleri otpasti od daljeg procesa zbog lošeg seta podataka. Na slici 19 je prikazan taj uslov.

Order	Output Name	Condition
1	Case 1	country_of_birth != "Canada" && country_of_birth != "UdSSR" && country_of_birth != "United States" && country_of_birth != "South Africa" && FINDSTRING(country_of_bir...

Slika 19 Uslov prilikom kreiranja dimenzije Player

Nakon izvršenih mapiranja pokrenut je *Data Flow Task* koji je prikazan na slici 20.



Slika 20 Proces kreiranja dimenzije Player u Data Flow Task-u

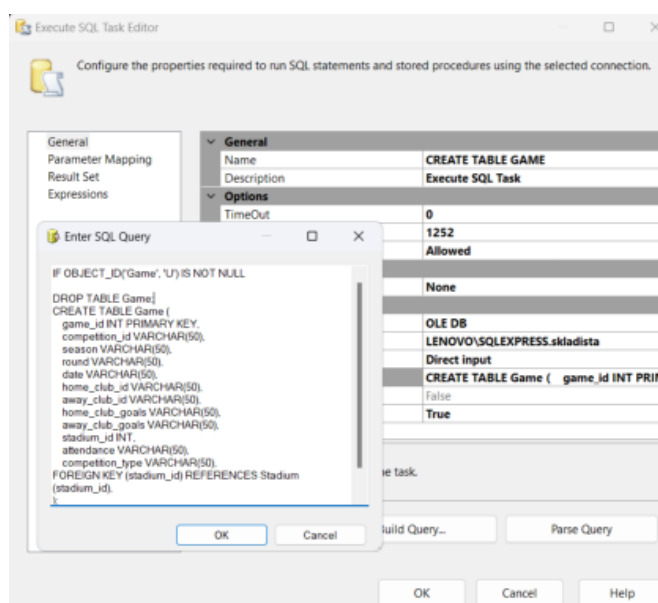
Na slici 21 su prikazani upisani podaci nakon pokretanja *Data Flow Task-a*.

	player_id	name	last_season	current_club_id	country_of_birth	date_of_birth	position	foot	height_in_cm	market_value_in_eur
1	10	Miroslav Klose	2015	398	Poland	1978-06-09	Attack	right	184	Not available
2	26	Roman Weidenfeller	2017	16	Germany	1980-08-06	Goalkeeper	left	190	Not available
3	65	Dimitar Berbatov	2015	1091	Bulgaria	1981-01-30	Attack	Not available	Not available	Not available
4	77	LAucio	2012	506	Brazil	1978-05-08	Defender	Not available	Not available	Not available
5	80	Tom Starke	2017	27	East Germany (GDR)	1981-03-18	Goalkeeper	right	194	Not available
6	109	DedAAt	2012	825	Brazil	1978-04-18	Defender	Not available	Not available	Not available
7	123	Christoph Metzelder	2012	33	Germany	1980-11-05	Defender	Not available	Not available	Not available
8	132	Tomas Rosicky	2015	11	CSSR	1980-10-04	Midfield	both	179	Not available
9	162	Marc Ziegler	2012	79	Germany	1976-06-13	Goalkeeper	right	193	Not available
10	215	Roque Santa Cruz	2015	1084	Paraguay	1981-08-16	Attack	right	193	250000
11	258	Gerald Asamoah	2013	33	Ghana	1978-10-03	Attack	right	180	Not available
12	264	TamAas Hajnal	2012	79	Hungary	1981-03-15	Midfield	both	167	Not available
13	276	Sladan Peric	2013	20100	Not available	1982-04-15	Defender	Not available	Not available	Not available
14	277	Sergio da Silva Pinto	2013	3368	Portugal	1980-10-16	Midfield	right	176	Not available
15	299	Fabian Ernst	2012	10484	Germany	Not available	Midfield	right	183	Not available
16	325	Tomas Ujfalusi	2012	141	CSSR	1978-03-24	Defender	Not available	Not available	Not available
17	332	Benjamin KA hier	2012	24	Germany	1980-08-04	Midfield	Not available	Not available	Not available
18	410	Sebastian Kehl	2014	16	Germany	1980-02-13	Midfield	left	187	Not available
19	447	Oka Nikolov	2012	24	Germany	1974-05-25	Goalkeeper	right	188	Not available
20	488	Gerhard Tremmel	2016	2288	Germany	1978-11-16	Goalkeeper	Not available	Not available	Not available
21	532	Claudio Pizarro	2019	86	Peru	1978-10-03	Attack	right	184	Not available
22	557	Simon Jentzsch	2012	167	Germany	1976-05-04	Goalkeeper	right	196	Not available

Slika 21 Izgled tabele koja predstavlja dimenziju Player u bazi podataka

5.5. Kreiranje dimenzije Game

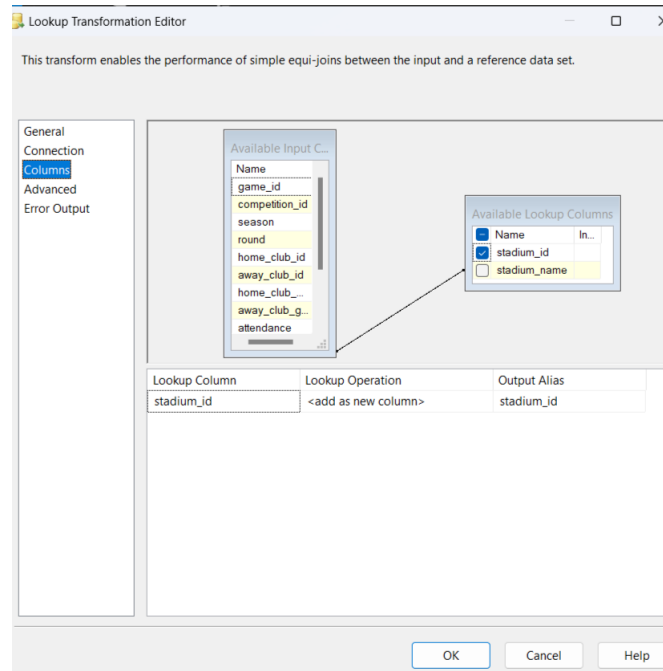
Kao kod prethodnih dimenzija i ovde je započeto pokretanjem script-a za kreiranje dimenzije Game što je prikazano na slici 22. Na slici je moguće videti da i ova tabela sadrži strani ključ koji se odnosi na stadion na kome je odigrana utakmica.



Slika 22 Script za kreiranje dimenzije Game

Pored komponente *Derived Column* koja je korišćenja u istu svrhu kao i kod prethodnih dimenzija, u okviru ovog procesa korišćene su komponente *Multicast* i *Lookup*

kako bi se pravilno mapirao strani ključ stadiona na postojeću kolonu stadiona u izvornom setu podataka. To je prikazano na slici 23.



Slika 23 Mapiranje vrednosti stranog ključa u dimenziji Game pomoću komponente Lookup

Pre samog pokretanja *Data Flow Task*-a izvršena su potrebna mapiranja kao i u prethodnim primerima. Na slici 24 prikazan je vizuelni pregled *Data Flow Task*-a.



Slika 24 Proces kreiranja dimenzije Game u Data Flow Task-u

Na slici 25 su prikazani upisani podaci nakon pokretanja *Data Flow Task*-a.

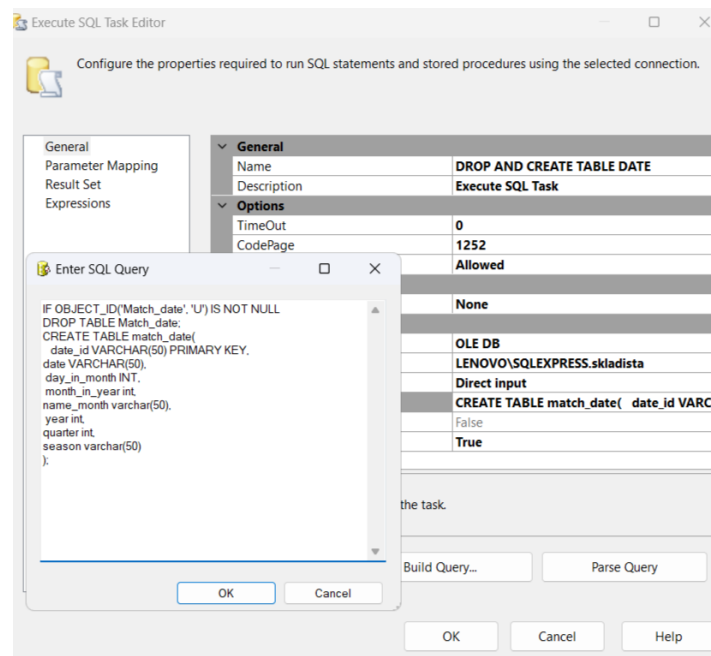
	game_id	competition_id	season	round	date	home_club_id	away_club_id	home_club_goals	away_club_goals	stadium_id	attendance	competition_type
15...	2508543	UKR1	2014	14. Match...	2014-11-30	6414	10690	4	2	1286	2811	domestic_leag...
15...	2508544	UKR1	2014	14. Match...	2014-11-29	23611	6992	2	1	1230	300	domestic_leag...
15...	2508546	UKR1	2014	14. Match...	2014-11-30	2740	338	0	3	2051	3500	domestic_leag...
15...	2508547	UKR1	2014	14. Match...	2014-11-30	6994	660	0	4	1525	2000	domestic_leag...
15...	2508548	UKR1	2014	14. Match...	2014-12-01	339	2783	1	1	478	1250	domestic_leag...
15...	2508549	UKR1	2014	14. Match...	2014-11-30	2477	6996	1	0	1825	2100	domestic_leag...
15...	2508629	SFA	2014	Third Rou...	2014-11-08	24232	3028	5	4	1203	1280	domestic_cup
15...	2508630	SFA	2014	Third Rou...	2014-11-08	3033	8929	15	14	655	567	domestic_cup
15...	2508631	SFA	2014	Third Rou...	2014-11-11	3020	1192	4	0	853	420	domestic_cup
15...	2508632	SFA	2014	Third Rou...	2014-11-11	2802	3026	0	2	1323	668	domestic_cup
15...	2508686	SFA	2014	Fourth Ro...	2014-11-29	2578	2759	2	1	1073	2383	domestic_cup
15...	2508687	SFA	2014	Fourth Ro...	2014-11-29	465	2451	1	1	1933	1957	domestic_cup
15...	2508688	SFA	2014	Fourth Ro...	2014-11-30	124	2553	3	0	847	14412	domestic_cup
15...	2508689	SFA	2014	Fourth Ro...	2014-11-30	43	371	0	4	1983	12676	domestic_cup
15...	2508690	SFA	2014	Fourth Ro...	2014-11-29	3001	3491	4	1	1295	1434	domestic_cup
15...	2508691	SFA	2014	Fourth Ro...	2014-11-29	3034	1270	2	2	1856	489	domestic_cup
15...	2508692	SFA	2014	Fourth Ro...	2014-11-29	1191	3027	1	0	1911	1237	domestic_cup
15...	2508693	SFA	2014	Fourth Ro...	2014-11-29	2801	3014	1	1	1500	439	domestic_cup
15...	2508694	SFA	2014	Fourth Ro...	2014-11-29	511	370	2	1	916	5956	domestic_cup
15...	2508695	SFA	2014	Fourth Ro...	2014-11-29	3015	3194	1	1	684	440	domestic_cup
15...	2508696	SFA	2014	Fourth Ro...	2014-11-29	987	1519	1	2	648	4827	domestic_cup
15...	2508697	SFA	2014	Fourth Ro...	2014-11-29	2760	2999	2	0	1910	2467	domestic_cuo

Slika 25 Izgled tabele koja predstavlja dimenziju Game u bazi podataka

5.6. Kreiranje dimenzije *Date*

Jedna od ključnih dimenzija u okviru OLAP šeme jeste *Date*, koja je kreirana poslednja od svih dimenzija. Iz fajla *appearances.txt* koji je najbliži kasnije kreiranoj tabeli činjenica, izdvojen je datum koji predstavlja primarni ključ u datoj dimenziji *Date*, a transformacijom tog datuma kasnije u procesu su izvedena i ostala obeležja u dimenziji.

Kao i do sad, pokrenut je *script* za kreiranje ove dimenzije u bazi podataka kroz *Execute SQL Task* koji je prikazan na slici 26.



Slika 26 Script za kreiranje dimenzije *Date*

Kako u izvornom fajlu postoji dosta redova sa istim datumom bilo je potrebno izdvojiti samo jednu vrednost tog datuma što je omogućeno upotrebom *Sort* komponente i čekiranjem dugmeta *Remove rows with duplicate sort values*. Nakon toga je upotrebom *Conditional Split* komponente obezbeđeno da ne prolaze redovi koji za vrednost datuma nemaju vrednost što je prikazano na slici 27.

Order	Output Name	Condition
1	Case 1	LEN(date) > 0

Slika 27 Uslov prilikom kreiranja dimenzije *Date*

U okviru komponente *Derived Column* izvedeno je 7 novih obeležja na osnovu kolone datum iz izvora. Kolone koje su izvedene se odnose na pojedinačno redni broj dana, meseca i godine, zatim kolona koja se odnosi na kvartal u godini, kolona koja se odnosi na godišnje doba, kolona koja se odnosi na naziv meseca u godini i na kraju kolona koja ispisuje izvorni datum u formatu *d-M-yyyy*. Izvorni datum koji predstavlja i primarni ključ je u formatu *yyyy-MM-dd*. Na slici 28 je prikazan način izvođenja pomenutih kolona.

Derived Column Name	Derived Column	Expression	Data Type
year	<add as new column>	YEAR((DT_DATE)date)	four-byt
month_in_year	<add as new column>	MONTH((DT_DATE)date)	four-byt
day	<add as new column>	DAY((DT_DATE)date)	four-byt
quarter	<add as new column>	MONTH((DT_DATE)date) <= 3 ? 1 : MONTH((DT_DATE)date) <= ...	four-byt
date	<add as new column>	(DT_WSTR,50)DAY((DT_DATE)date) + "-" + (DT_WSTR,50)MONT...	Unicode
season	<add as new column>	MONTH((DT_DATE)date) == 1 ? "Winter" : MONTH((DT_DATE)da...	Unicode
name	<add as new column>	MONTH((DT_DATE)date) == 1 ? "January" : MONTH((DT_DATE)d...	Unicode

Slika 28 Izvođenje novih kolona u dimenziji Date

Nakon izvođenja novih kolona izvršena je konverzija podataka pomoću *Data Conversion* komponente što je prikazano na slici 29.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Derived Column.date	Copy of Derived Col...	string [DT_STR]	50			1252 (ANS
season	Copy of season	string [DT_STR]	50			1252 (ANS
name	Copy of name	string [DT_STR]	50			1252 (ANS

Slika 29 Konverzija podataka prilikom kreiranja dimenzije Date

Pre samog pokretanja *Data Flow Task*-a izvršena su potrebna mapiranja. Na slici 30 prikazan je vizuelni pregled *Data Flow Task*-a.



Slika 30 Proces kreiranja dimenzije Date u Data Flow Task-u

Na slici 31 su prikazani upisani podaci nakon pokretanja *Data Flow Task*-a.

Results

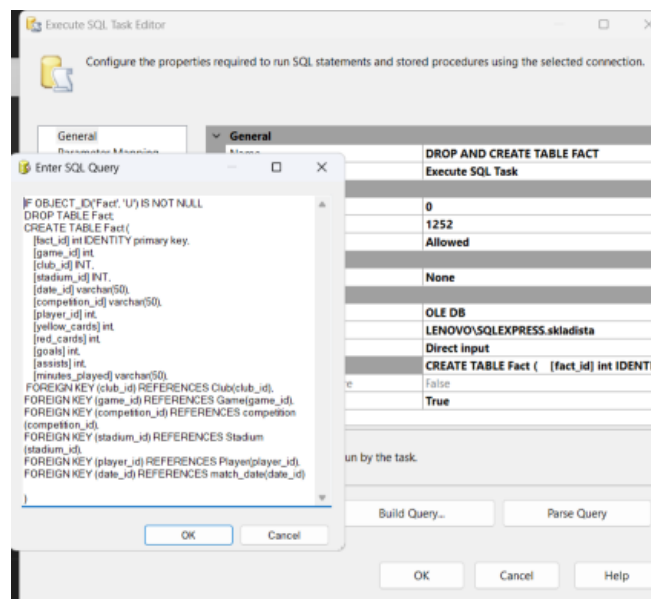
Messages

	date_id	date	day_in_month	month_in_year	name_month	year	quarter	season
1	2014-07-01	1-7-2014	1	7	July	2014	3	Summer
2	2014-07-03	3-7-2014	3	7	July	2014	3	Summer
3	2014-07-08	8-7-2014	8	7	July	2014	3	Summer
4	2014-07-10	10-7-2014	10	7	July	2014	3	Summer
5	2014-07-11	11-7-2014	11	7	July	2014	3	Summer
6	2014-07-15	15-7-2014	15	7	July	2014	3	Summer
7	2014-07-16	16-7-2014	16	7	July	2014	3	Summer
8	2014-07-17	17-7-2014	17	7	July	2014	3	Summer
9	2014-07-18	18-7-2014	18	7	July	2014	3	Summer
10	2014-07-19	19-7-2014	19	7	July	2014	3	Summer
11	2014-07-20	20-7-2014	20	7	July	2014	3	Summer
12	2014-07-21	21-7-2014	21	7	July	2014	3	Summer
13	2014-07-22	22-7-2014	22	7	July	2014	3	Summer
14	2014-07-23	23-7-2014	23	7	July	2014	3	Summer
15	2014-07-24	24-7-2014	24	7	July	2014	3	Summer
16	2014-07-25	25-7-2014	25	7	July	2014	3	Summer
17	2014-07-26	26-7-2014	26	7	July	2014	3	Summer
18	2014-07-27	27-7-2014	27	7	July	2014	3	Summer
19	2014-07-28	28-7-2014	28	7	July	2014	3	Summer
20	2014-07-29	29-7-2014	29	7	July	2014	3	Summer
21	2014-07-30	30-7-2014	30	7	July	2014	3	Summer
22	2014-07-31	31-7-2014	31	7	July	2014	3	Summer

Slika 31 Izgled tabele koja predstavlja dimenziju Date u bazi podataka

5.7. Kreiranje tabele činjenica

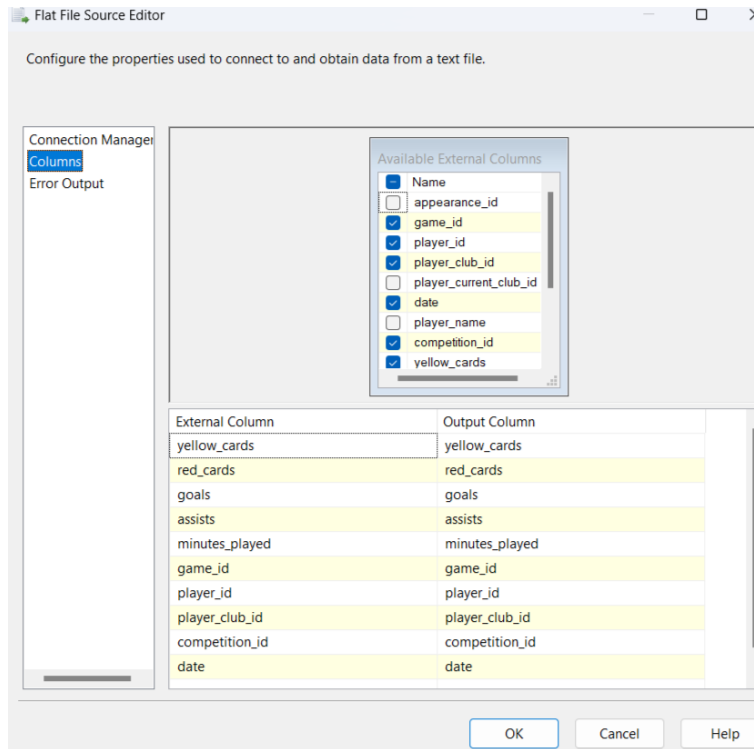
Kao i kod dimenzija, i ovde je prvo izvršeno pokretanje *script*-a koji prvo proverava postojanje tabele činjenica, a zatim i kreiranje iste putem *Execute SQL Task*-a. Na slici 32 prikazan je dati *script*.



Slika 32 Script za kreiranje tabele činjenica

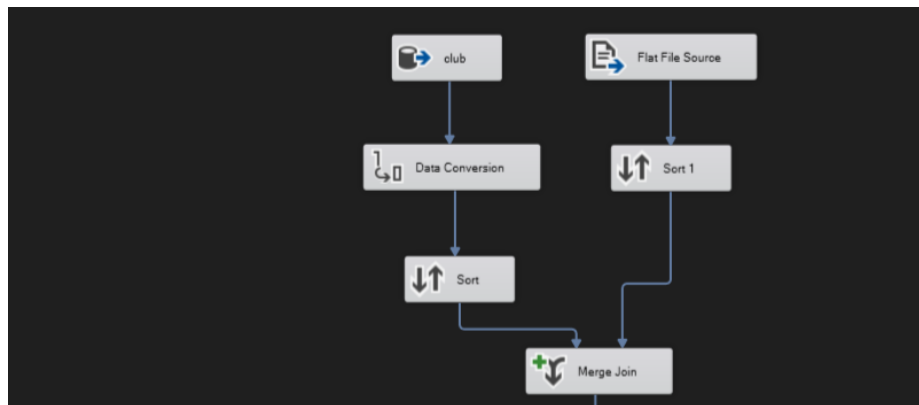
Naredni koraci predstavljaju učitavanje izvornih podataka i spajanje svih tabela po zajedničkim obeležjima, s tim što se mora voditi računa kojim redom treba povezivati.

Takođe, na samom početku prilikom učitavanja fajla, izabrane su kolone koje će predstavljati mere u tabeli činjenica što je prikazano na slici 33.

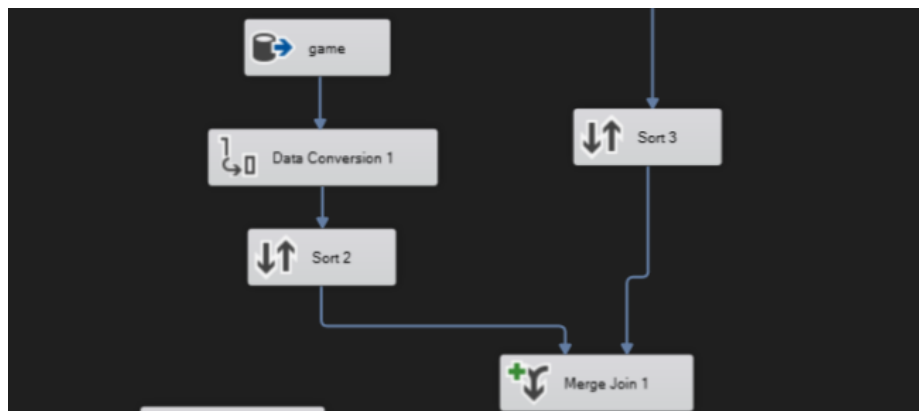


Slika 33 Odabir mera za tabelu činjenica

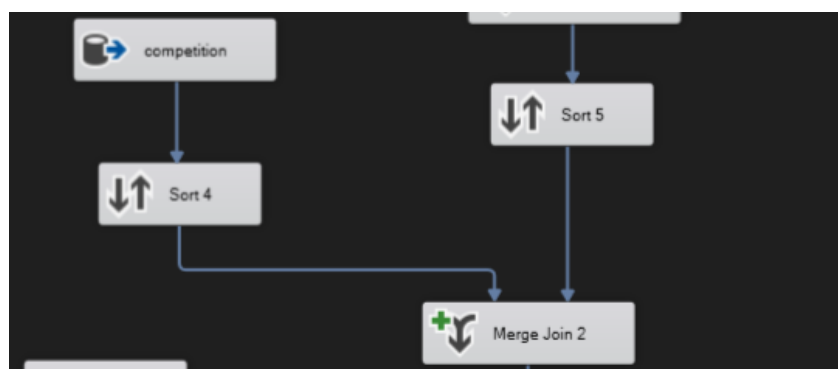
Na slikama 34, 35, 36, 37 i 38 prikazan je celokupan *Data Flow Task* za kreiranje tabele činjenica. U okviru ovog *Data Flow Task*-a korišćene su komponente *Sort*, *Data Conversion* i *Merge Join* (prilikom spajanja svih tabela korišćen je *inner join*).



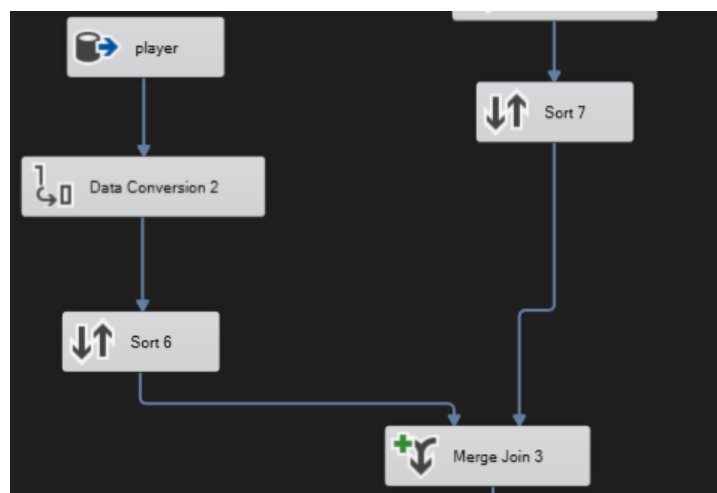
Slika 34 Proces kreiranja tabele činjenica 1



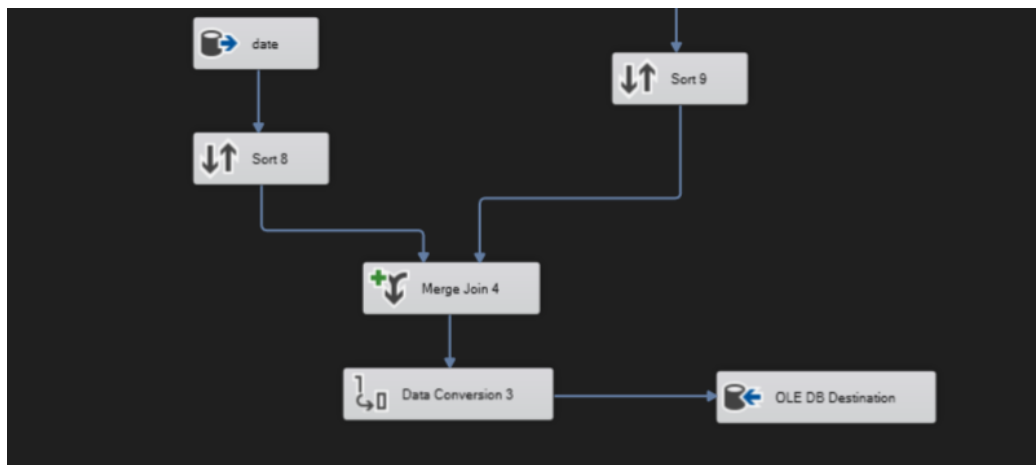
Slika 35 Proces kreiranja tabele činjenica 2



Slika 36 Proces kreiranja tabele činjenica 3



Slika 37 Proces kreiranja tabele činjenica 4



Slika 38 Proces kreiranja tabele činjenica 5

Na slici 39 su prikazani upisani podaci nakon pokretanja *Data Flow Task-a*.

Results		Messages										
	fact_id	game_id	club_id	stadium_id	date_id	competition_id	player_id	yellow_cards	red_cards	goals	assists	minutes_played
1	1	2460523	370	1339	2014-07-03	ELQ	12817	0	0	0	0	90
2	2	2460523	370	1339	2014-07-03	ELQ	149614	0	0	0	0	1
3	3	2460523	370	1339	2014-07-03	ELQ	158938	0	0	0	0	40
4	4	2460523	370	1339	2014-07-03	ELQ	21419	0	0	0	0	90
5	5	2460523	370	1339	2014-07-03	ELQ	276914	0	0	0	1	12
6	6	2460523	370	1339	2014-07-03	ELQ	28809	0	0	0	0	90
7	7	2460523	370	1339	2014-07-03	ELQ	37773	1	0	2	1	90
8	8	2460523	370	1339	2014-07-03	ELQ	12804	0	0	0	0	78
9	9	2460523	370	1339	2014-07-03	ELQ	52380	0	0	1	1	90
10	10	2460523	370	1339	2014-07-03	ELQ	58579	0	0	1	0	90
11	11	2460523	370	1339	2014-07-03	ELQ	60793	0	0	0	1	90
12	12	2460523	370	1339	2014-07-03	ELQ	12765	0	0	0	0	90
13	13	2460523	370	1339	2014-07-03	ELQ	62170	0	0	0	0	50
14	14	2460523	370	1339	2014-07-03	ELQ	68321	0	0	1	0	89
15	15	2460562	370	374	2014-07-10	ELQ	106721	0	0	0	0	90
16	16	2460562	370	374	2014-07-10	ELQ	95986	0	0	0	0	10
17	17	2460562	370	374	2014-07-10	ELQ	158938	0	0	0	0	28
18	18	2460562	370	374	2014-07-10	ELQ	21306	0	0	0	0	90
19	19	2460562	370	374	2014-07-10	ELQ	276914	0	0	0	0	10
20	20	2460562	370	374	2014-07-10	ELQ	28809	0	0	0	0	90
21	21	2460562	370	374	2014-07-10	ELQ	68321	0	0	0	0	80
22	22	2460562	370	374	2014-07-10	ELQ	37773	0	0	3	0	90

Slika 39 Izgled tabele činjenica u bazi podataka

6. Prikaz izveštaja

Za prikazivanje izveštaja korišćen je paket SSRS (*SQL Server Report System*) u sklopu *Visual Studio*-a.

6.1. Izveštaj 1 - Koliko asistencija je imao fudbaler sa najviše asistencija u fudbalskom klubu Monako u sezoni 2021/22?

Pretpostavka na ovo pitanje je bila 15 asistencija, a stvarni broj asistencija je prikazan na izveštaju na slici 40.

Prikaz asistencija u Ligi 1 sezona 2021	
as-monaco	
Fudbaler	Asistencije
Aleksandr Golovin	4
Caio Henrique	9
Gelson Martins	3
Ismail Jakobs	1
Jean Lucas	1
Kevin Volland	11
Myron Boadu	1
Ruben Aguilar	2
Vanderson	1
Wissam Ben Yedder	5

Slika 40 Izveštaj 1 – odgovor 11

6.2. Izveštaj 2 - Koliko golova je postigao fudbaler Ante Rebić u dres fudbalskog Milan u sezoni 2020/21?

Pretpostavka na ovo pitanje je bila 5 golova, a stvarni broj golova je prikazan na izveštaju na slici 41.

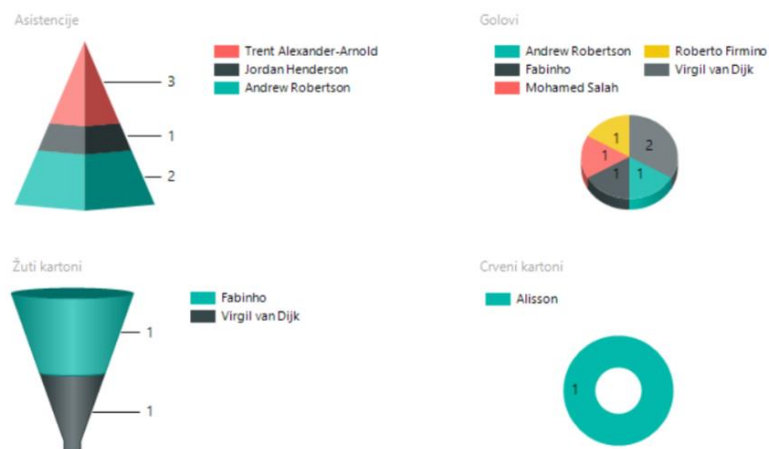
+	2015		
+	2016		
+	2017		
+	2018		
+	2019		
+	2020		
+	ac-florenz		
+	ac-mailand		
+	Alexis Saelemaekers		
+	Ante Rebic		
Broj zutih kartona	Broj crvenih kartona	Broj golova	Broj asistencija
4	1	11	7
+	Brahim DAA-az		
+	Ciprian Tatarusanu		
+	Davide Calabria		
+	IsmaAel Bennacer		
+	Mario Mandzukic		
+	Mateo Musacchio		
+	Matteo Gabbia		
+	Pierre Kalulu		
+	Rade Krunic		
+	Rafael LeAao		

Slika 41 Izveštaj 2 – odgovor 11

6.3. Izveštaj 3 - Koji fudbaleri su imali najviše žutih kartona za fudbalski klub Liverpool u sezoni 2019/20 u mesecu decembru?

Pretpostavka na ovo pitanje su bili fudbaleri *Fabinho*, *Roberto Firmino* i *Mohammed Salah*, a pravi odgovor je prikazan na izveštaju na slici 42.

Liverpul - statistika za sezonu 2019 mesec decembar



Slika 42 Izveštaj 3 – odgovor Fabinho i Virgil van Dijk

6.4. Izveštaj 4- Koliki je broj odigranih utakmica koje je fudbaler *Mohammed Salah* odigrao u sezoni 2018/19 u Ligi Šampiona?

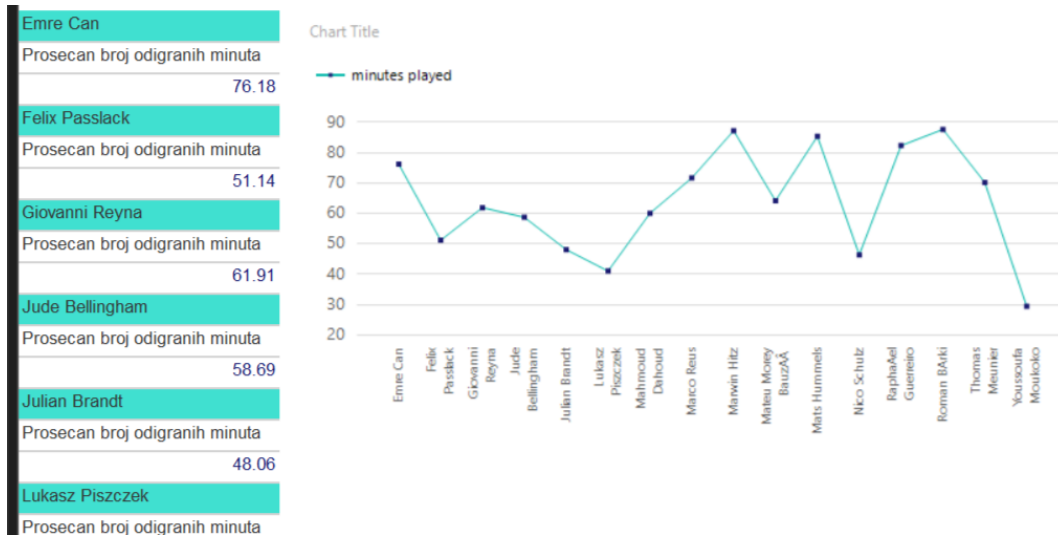
Pretpostavka na ovo pitanje bila je 10 utakmica, a pravi odgovor je prikazan na izveštaju na slici 43.

Statistika Mohameda Salaha kroz sezone u Liverpulu									
2017									
2018									
fa-cup									
premier-league									
uefa-champions-league									
	Broj zutih kartona		Broj crvenih kartona		Broj golova		Broj asistencija		Broj odigranih utakmica
	1		0		5		2		12
Ukupan broj zutih kartona	2	Ukupan broj crvenih kartona	0	Ukupan broj golova	27	Ukupan broj asistencija	12	Ukupan broj odigranih utakmica	51
2019									
2020									
2021									

Slika 43 Izveštaj 4 – odgovor 12

6.5. Izveštaj 5- Koji fudbaler je imao prosečno najviše odigranih minuta u fudbalskom klubu Borussia Dortmund u sezoni 2020/21?

Pretpostavka na ovo pitanje bila je *Mats Hummels*, a pravi odgovor je prikazan na izveštaju na slici 44.



Slika 44 Izveštaj 5 - odgovor Roman Burki

7. Zaključak

Primarni cilj realizacije skladišta podataka ostvaren je korišćenjem *Microsoft Visual Studio* razvojnog okruženja u kombinaciji sa *SQL Server Management Studio*-om. Pored implementacije skladišta podataka na temu fudbalskih utakmica u svetu, kreirani su i izveštaji koji odgovaraju specifikacijama zahteva korisnika. Izvorni podaci preuzeti sa sajta *Kaggle* bili su donekle dobro strukturirani, što je kao rezultat imalo veoma iscrpan rad na projektovanju skladišta. Postoji mogućnost dodavanja novih atributa i dimenzija u ovu šemu, čime bi se struktura učinila kompleksnijom, omogućavajući detaljniji uvid u analizirani sistem i bolju analizu fudbalskih utakmica i učinka fudbalera u svetu. Rad na projektu pružio mi je bolje razumevanje koncepata skladišta podataka, ističući ključne aspekte projektovanja kao što su specifikacije i namena sistema. Pored tehničkih aspekata, rad na ovom projektu mi je omogućio da razvijem veštine u analizi podataka i kreiranju sveobuhvatnih izveštaja koji su od praktične koristi korisnicima. Kroz ovaj proces, stekao sam i iskustvo u rešavanju problema i optimizaciji podataka za bolje performanse sistema. Projekat je, osim tehničkog znanja, doprineo i mom profesionalnom razvoju, pružajući mi praktično iskustvo koje će biti dragoceno u budućim radnim zadacima.