**Student Name: Bianca-Denise Lazar**

**Student Number: C12706389**

**DT228/4**

**Relational and Dimensional Model, ETL Assignment**

**Instructions:**

I have each of my scripts numbered in the order they should be executed.

Once procedures are created they can be run using:

* exec etl1;
* exec etl2;

1. **Create tables and insert initial data**

I used the script below to create the tables which will hold data. The tables/inserts can be found in the **CreateTables.Insert** file.

Each table has relevant primary keys and foreign keys.

Moreover, I inserted the data provided (into the Teams and Stadium tables).

--Create tables

--Table Player

**create** **table** Player

**(**

Player\_ID **int** **primary** **key,**

Pl\_name **varchar2(**30**),**

Pl\_surname **varchar2(**30**),**

Team\_ID **int** **references** Teams**(**Team\_ID**)**

**);**

--Table Stadium

**create** **table** Stadium

**(**

Stadium\_ID **int** **primary** **key,**

St\_name **varchar2(**30**),**

City **varchar2(**30**),**

Capacity **int,**

Team\_ID **int** **references** Teams**(**Team\_ID**)**

**);**

--Table Teams

**create** **table** Teams

**(**

Team\_ID **int** **primary** **key,**

Team\_name **varchar2(**30**),**

YearOfFound **number(**4**)**

**);**

--Table Matches

**create** **table** Matches

**(**

Team\_A **int,**

Team\_B **int,**

M\_date **date,**

**primary** **key** **(**Team\_A**,** Team\_B**,** M\_date**)**

**);**

--Table Players\_stats

**create** **table** Players\_stats

**(**

Team\_A\_ID **int,**

Team\_B\_ID **int,**

M\_date **date,**

Player\_ID **int** **references** Player**(**Player\_ID**),**

MinPlayed **number,**

Goals **number,**

Shot\_on **number,**

Shot\_off **number,**

Penalty **number,**

Pass\_OK **number,**

Pass\_KO **number,**

**foreign** **key** **(**Team\_A\_ID**,** Team\_B\_ID**,** M\_date**)** **references** Matches**(**Team\_A\_ID**,** Team\_B\_ID**,** M\_date**)**

**);**

1. **Insert from CSV file**

I converted the data from the Premier.csv to SQL insert statements by using an online converter: <http://www.convertcsv.com/csv-to-sql.htm>

I inserted all of this data into **data\_table**, and then moved the data to the relevant tables.

All this process can be found in **CreateTable.Insert(Premier.csv)** file:

--Create a large table to hold all data

**create** **table** data\_table

**(**

m\_date **date,**

player\_id **int,**

pl\_surname **varchar2** **(**30**),**

pl\_firstname **varchar2(**30**),**

team\_name **varchar2(**30**),**

team\_id **int,**

opposition\_name **varchar2(**30**),**

opposition\_id **int,**

min\_played **number,**

goals **number,**

shots\_on\_target **number,**

Shots\_off\_target **number,**

penalty\_goals **number,**

pass\_ok **number,**

pass\_ko **number**

**);**

--Large amount of inserts

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'26/11/2011'**,** 'DD-MM-YYYY'**),**28566**,**'Abou Diaby'**,**'Vassiriki'**,**'Arsenal'**,**3**,**'Fulham'**,**54**,**21**,**0**,**0**,**0**,**0**,**18**,**5**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'21/04/2012'**,** 'DD-MM-YYYY'**),**28566**,**'Abou Diaby'**,**'Vassiriki'**,**'Arsenal'**,**3**,**'Chelsea'**,**8**,**25**,**0**,**0**,**0**,**0**,**18**,**3**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'13/08/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Sunderland'**,**56**,**90**,**0**,**1**,**1**,**0**,**45**,**18**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'05/11/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Swansea City'**,**80**,**90**,**0**,**0**,**0**,**0**,**65**,**20**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'06/02/2012'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Tottenham Hotspur'**,**6**,**90**,**0**,**0**,**0**,**0**,**62**,**13**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'14/01/2012'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Stoke City'**,**110**,**90**,**0**,**0**,**2**,**0**,**64**,**17**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'30/12/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Newcastle United'**,**4**,**59**,**0**,**0**,**1**,**0**,**21**,**14**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'22/10/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Norwich City'**,**45**,**90**,**0**,**0**,**0**,**0**,**47**,**6**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'27/11/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Manchester City'**,**43**,**90**,**0**,**1**,**1**,**0**,**46**,**19**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'15/10/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Manchester United'**,**1**,**90**,**0**,**0**,**0**,**0**,**31**,**11**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'10/12/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Queens Park Rangers'**,**52**,**90**,**0**,**0**,**0**,**0**,**59**,**24**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'24/09/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Wolverhampton Wanderers'**,**39**,**90**,**0**,**0**,**1**,**0**,**29**,**11**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'03/03/2012'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Arsenal'**,**3**,**90**,**0**,**0**,**0**,**0**,**49**,**27**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'27/08/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Bolton Wanderers'**,**30**,**90**,**1**,**1**,**1**,**0**,**32**,**13**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'26/12/2011'**,** 'DD-MM-YYYY'**),**20208**,**'Adam'**,**'Charlie'**,**'Liverpool'**,**14**,**'Blackburn Rovers'**,**5**,**69**,**0**,**1**,**1**,**0**,**50**,**12**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'03/12/2011'**,** 'DD-MM-YYYY'**),**17500**,**'Adebayor'**,**'Emmanuel'**,**'Tottenham Hotspur'**,**6**,**'Bolton Wanderers'**,**30**,**90**,**0**,**2**,**2**,**0**,**21**,**7**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'29/04/2012'**,** 'DD-MM-YYYY'**),**17500**,**'Adebayor'**,**'Emmanuel'**,**'Tottenham Hotspur'**,**6**,**'Blackburn Rovers'**,**5**,**90**,**0**,**0**,**0**,**0**,**39**,**7**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'18/12/2011'**,** 'DD-MM-YYYY'**),**17500**,**'Adebayor'**,**'Emmanuel'**,**'Tottenham Hotspur'**,**6**,**'Sunderland'**,**56**,**90**,**0**,**2**,**1**,**0**,**18**,**13**);**

*--Moving the data in the Player, Matches and Players\_stats tables*

**insert** **into** Player**(**Player\_ID**,** Pl\_name**,** Pl\_surname**,** Team\_ID**)**

**(select** **unique** player\_id**,** pl\_firstname**,** pl\_surname**,** team\_id **from** data\_table**);**

**insert** **into** Matches**(**Team\_A**,** Team\_B**,** M\_date**)**

**(select** **unique** team\_id**,** opposition\_id**,** m\_date **from** data\_table**);**

**insert** **into** Players\_stats**(**Team\_A\_ID**,** Team\_B\_ID**,** M\_date**,** Player\_ID**,** MinPlayed**,** Goals**,**Shot\_on**,**Shot\_off**,** Penalty**,** Pass\_OK**,**Pass\_KO **)**

**(select** team\_id**,** opposition\_id**,** m\_date**,** player\_id**,** min\_played**,** goals**,** shots\_on\_target**,** shots\_off\_target**,** penalty\_goals**,** pass\_ok**,** pass\_ko **from** data\_table**);**

1. **Create dimensional and staging tables**

The dimensional tables were created using the model in the specification. Also, the staging tables were created using a similar model, to allow the data to be easily transformed and loaded. The tables can be found in CreateDimensional&FactsTables file:

--Drop tables

**drop** **table** fact\_stats**;**

**drop** **table** dimplayer**;**

**drop** **table** dimteam**;**

**drop** **table** dimtime**;**

**drop** **table** dimstadium**;**

--Create dimensional and fact tables

**create** **table** DimPlayer

**(**

player\_sk **int** **primary** **key,**

player\_name **varchar2(**30**),**

player\_surname **varchar2(**30**)**

**);**

**create** **table** DimTeam

**(**

team\_sk **int** **primary** **key,**

team\_name **varchar2(**30**),**

year\_founded **int**

**);**

**create** **table** DimTime

**(**

date\_sk **date** **primary** **key,**

**year** **number(**4**),**

**month** **number(**2**),**

**day** **number(**2**)**

**);**

**create** **table** DimStadium

**(**

stadium\_sk **int** **primary** **key,**

stadium\_name **varchar2(**30**),**

stadium\_city **varchar2(**30**),**

capacity **int**

**);**

**create** **table** Fact\_Stats

**(**

date\_sk **date** **references** DimTime**(**date\_sk**),**

player\_sk **int** **references** DimPlayer**(**player\_sk**),**

team\_sk **int** **references** DimTeam**(**team\_sk**),**

opponent\_sk **int** **references** DimTeam**(**team\_sk**),**

stadium\_sk **int** **references** DimStadium**(**stadium\_sk**),**

min\_played **number,**

goals **number,**

shot\_on **number,**

shot\_off **number,**

penalty **number,**

pass\_ok **number,**

pass\_ko **number**

**);**

--Staging tables

**drop** **table** fact\_stage**;**

**drop** **table** player\_stage**;**

**drop** **table** team\_stage**;**

**drop** **table** stadium\_stage**;**

**drop** **table** time\_stage**;**

**create** **table** player\_stage **(**

player\_sk **integer** **primary** **key,**

player\_name **varchar2(**30**),**

player\_surname **varchar2(**30**)**

**);**

**create** **table** team\_stage

**(**

team\_sk **int,**

team\_name **varchar2(**30**),**

year\_founded **number(**4**)**

**);**

**create** **table** stadium\_stage

**(**

stadium\_sk **int,**

stadium\_name **varchar2(**30**),**

stadium\_city **varchar2(**30**),**

capacity **int,**

team\_id **int**

**);**

**create** **table** time\_stage

**(**

date\_sk **date,**

**year** **number(**4**),**

**month** **number(**2**),**

**day** **number(**2**)**

**);**

**create** **table** fact\_stage

**(**

date\_sk **date,**

player\_sk **int,**

team\_sk **int,**

opponent\_sk **int,**

stadium\_sk **int,**

min\_played **number,**

goals **number,**

shot\_on **number,**

shot\_off **number,**

penalty **number,**

pass\_ok **number,**

pass\_ko **number**

**);**

1. **ETL1**

Below is ETL1. It is placed inside a procedure, which inserts the relevant data into staging tables, performs and modifications needed, and then inserts into the dimensional tables. The procedure can be executed using **exec etl1;**

**exec** etl1**;**

**CREATE** **OR** **REPLACE** **PROCEDURE** ETL1

**IS**

**BEGIN**

--Player Staging

**insert** **into** player\_stage **(**player\_sk**,** player\_name**,** player\_surname**)**

**select** Player\_ID**,** Pl\_name**,** Pl\_surname **from** Player**;**

**insert** **into** dimplayer **select** player\_sk**,** player\_name**,** player\_surname **from** player\_stage**;**

--Teams Staging

**insert** **into** team\_stage **(**team\_sk**,** team\_name**,** year\_founded**)**

**select** Team\_ID**,** Team\_name**,** YearofFound **from** Teams**;**

**insert** **into** dimteam **select** team\_sk**,** team\_name**,** year\_founded **from** team\_stage**;**

--Stadium Staging

**insert** **into** stadium\_stage **(**stadium\_sk**,** stadium\_name**,** stadium\_city**,** capacity**,** team\_id**)**

**select** Stadium\_ID**,** St\_name**,** City**,** Capacity**,** Team\_ID **from** Stadium**;**

**insert** **into** dimstadium **select** stadium\_sk**,** stadium\_name**,** stadium\_city**,** capacity **from** stadium\_stage**;**

--Time Staging

**insert** **into** time\_stage **(**date\_sk**,** **year,** **month,** **day)**

**select** **unique** M\_date**,** **extract** **(year** **from** M\_date**),** **extract** **(month** **from** M\_date**),** **extract** **(day** **from** M\_date**)** **from** Matches**;**

**insert** **into** dimtime **select** date\_sk**,** **year,** **month,** **day** **from** time\_stage**;**

--Fact Staging

**insert** **into** fact\_stage **(**date\_sk**,** player\_sk**,** team\_sk**,**opponent\_sk**,**stadium\_sk**,**min\_played**,**goals**,**shot\_on**,**shot\_off**,**penalty**,** pass\_ok**,** pass\_ko**)**

**select** M\_date**,** Player\_ID**,** Team\_A\_ID**,** Team\_B\_ID**,** s**.**stadium\_sk**,** MinPlayed**,** Goals**,** Shot\_on**,** Shot\_off**,** Penalty**,** Pass\_OK**,** Pass\_KO **from** Players\_stats**,** stadium\_stage s

**where** Team\_A\_ID **=** s**.**TEAM\_ID**;**

**insert** **into** fact\_stats **(**date\_sk**,** player\_sk**,** team\_sk**,** opponent\_sk**,** stadium\_sk**,** min\_played**,** goals**,** shot\_on**,** shot\_off**,** penalty**,** pass\_ok**,** pass\_ko**)**

**select** date\_sk**,** player\_sk**,** team\_sk**,** opponent\_sk**,** stadium\_sk**,** min\_played**,** goals**,** shot\_on**,** shot\_off**,** penalty**,** pass\_ok**,** pass\_ko **from** fact\_stage**;**

**commit;**

**END;**

1. **Insert new data**

New data is inserted in bulk into the **data\_table**. This table is initially truncated as previous data is no longer needed. Data is then again inserted into tables, using slightly different method to the first time.

The process can be found in InsertNewData file:

--Insert bulk data

**truncate** **table** data\_table**;**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'06/05/2012'**,** 'DD-MM-YYYY'**),**11378**,**'Williamson'**,**'Michael'**,**'Newcastle United'**,**4**,**'Manchester City'**,**43**,**90**,**0**,**0**,**0**,**0**,**24**,**8**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'21/04/2012'**,** 'DD-MM-YYYY'**),**11378**,**'Williamson'**,**'Michael'**,**'Newcastle United'**,**4**,**'Stoke City'**,**110**,**90**,**0**,**0**,**0**,**0**,**37**,**7**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'15/01/2012'**,** 'DD-MM-YYYY'**),**11378**,**'Williamson'**,**'Michael'**,**'Newcastle United'**,**4**,**'Queens Park Rangers'**,**52**,**90**,**0**,**0**,**0**,**0**,**43**,**14**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'18/03/2012'**,** 'DD-MM-YYYY'**),**11378**,**'Williamson'**,**'Michael'**,**'Newcastle United'**,**4**,**'Norwich City'**,**45**,**90**,**0**,**0**,**0**,**0**,**26**,**15**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'01/04/2012'**,** 'DD-MM-YYYY'**),**11378**,**'Williamson'**,**'Michael'**,**'Newcastle United'**,**4**,**'Liverpool'**,**14**,**90**,**0**,**1**,**2**,**0**,**22**,**5**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'01/05/2012'**,** 'DD-MM-YYYY'**),**32318**,**'Wilson'**,**'Marc'**,**'Stoke City'**,**110**,**'Everton'**,**11**,**90**,**0**,**0**,**0**,**0**,**16**,**18**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'24/09/2011'**,** 'DD-MM-YYYY'**),**32318**,**'Wilson'**,**'Marc'**,**'Stoke City'**,**110**,**'Manchester United'**,**1**,**90**,**0**,**1**,**1**,**0**,**20**,**10**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'28/04/2012'**,** 'DD-MM-YYYY'**),**32318**,**'Wilson'**,**'Marc'**,**'Stoke City'**,**110**,**'Arsenal'**,**3**,**90**,**0**,**0**,**0**,**0**,**27**,**10**);**

**INSERT** **INTO** data\_table **VALUES** **(to\_date(**'07/04/2012'**,** 'DD-MM-YYYY'**),**32318**,**'Wilson'**,**'Marc'**,**'Stoke City'**,**110**,**'Wolverhampton Wanderers'**,**39**,**90**,**0**,**0**,**1**,**0**,**20**,**11**);**

--Insert

**insert** **into** Player**(**Player\_ID**,** Pl\_name**,** Pl\_surname**,** Team\_ID**)**

**(select** **unique** player\_id**,** pl\_firstname**,** pl\_surname**,** team\_id **from** data\_table

**where** player\_id **not** **in** **(select** Player\_ID **from** Player**));**

**insert** **into** Players\_stats**(**Team\_A\_ID**,** Team\_B\_ID**,** M\_date**,** Player\_ID**,** MinPlayed**,** Goals**,** Shot\_on**,** Shot\_off**,** Penalty**,** Pass\_OK**,** Pass\_KO**)**

**(select** team\_id**,** opposition\_id**,** m\_date**,** player\_id**,** min\_played**,** goals**,** shots\_on\_target**,** shots\_off\_target**,** penalty\_goals**,** pass\_ok**,** pass\_ko **from** data\_table**);**

1. **ETL2**

The second ETL is again bundled into a procedure similar to the first. Extra filtering is added to the inserts to ensure duplicate players are not inserted into **dimplayer**, only new players. This procedure can be executed using **exec etl2;**

**exec** etl2**;**

**CREATE** **OR** **REPLACE** **PROCEDURE** ETL2

**IS**

**BEGIN**

--Player Staging

**insert** **into** player\_stage **(**player\_sk**,** player\_name**,** player\_surname**)**

**(select** Player\_ID**,** Pl\_name**,** Pl\_surname **from** Player

**where** Player\_ID **not** **in** **(select** player\_sk **from** player\_stage**));**

**insert** **into** dimplayer

**(select** player\_sk**,** player\_name**,** player\_surname **from** player\_stage

**where** player\_sk **not** **in** **(select** player\_sk **from** dimplayer**));**

--Fact Staging

**execute** **immediate** 'truncate table fact\_stage'**;**

**insert** **into** fact\_stage **(**date\_sk**,** player\_sk**,** team\_sk**,**opponent\_sk**,**stadium\_sk**,**min\_played**,**goals**,**shot\_on**,**shot\_off**,**penalty**,** pass\_ok**,** pass\_ko**)**

**select** M\_date**,** Player\_ID**,** Team\_A\_ID**,** Team\_B\_ID**,** s**.**stadium\_sk**,** MinPlayed**,** Goals**,** Shot\_on**,** Shot\_off**,** Penalty**,** Pass\_OK**,** Pass\_KO **from** Players\_stats**,** stadium\_stage s

**where** Team\_A\_ID **=** s**.**TEAM\_ID**;**

**execute** **immediate** 'truncate table fact\_stats'**;**

**insert** **into** fact\_stats **(**date\_sk**,** player\_sk**,** team\_sk**,** opponent\_sk**,** stadium\_sk**,** min\_played**,** goals**,** shot\_on**,** shot\_off**,** penalty**,** pass\_ok**,** pass\_ko**)**

**select** date\_sk**,** player\_sk**,** team\_sk**,** opponent\_sk**,** stadium\_sk**,** min\_played**,** goals**,** shot\_on**,** shot\_off**,** penalty**,** pass\_ok**,** pass\_ko **from** fact\_stage**;**

**commit;**

**END;**

1. **Queries**

--Query 1

--For all the shots that team took, what percentage of the shots were on target

**select** **sum(**shot\_on **+** shot\_off**)** **as** Total\_Shots\_Made**,** **round(** **sum(**shot\_on**)** **/** **sum(**shot\_on **+** shot\_off**)\*** 100 **)** **||**'%' **as** Shots\_on\_Target

**from** fact\_stats

**where** team\_sk **=** 1

**and** date\_sk **>=** **to\_date(**'01/10/2011'**,** 'DD-MM-YYYY'**)**

**and** date\_sk **<=** **to\_date(**'31/12/2011'**,** 'DD-MM-YYYY'**);**

**Output:**



--Query 2

--List all goals for a certain team

**select** t**.**team\_name**,** **sum(**f**.**goals**)** **from** fact\_stats s

**join** dimteam t **on** f**.**team\_sk **=** t**.**team\_sk

**where** f**.**team\_sk **=** 1

**group** **by** team\_name**;**

**Output:**

