PROJECT SUMMARY

Maximizing Fan Satisfaction and Revenue by Increasing Offensive Production

INSIDE

CREATED BY

Priyanka Mitra

Sion Kim

Dal Price

Table of Contents

Organizational Overview

Vision and Objectives

Real Salt Lake, also known as RSL, is an American professional soccer franchise and a member of the United States Major League Soccer (MLS) association. Real Salt Lake represents the state of Utah within the MLS and started in 2004. The team is currently owned by Dell Loy Hansen who is also the owner of Wasatch Property Management, a company that acquires, manages and develops commercial properties. Real Salt Lake has two primary types of operations including Business Operations and Soccer Operations. AbracaData has been hired to focus efforts on Real Salt Lake's soccer operations. Including the following key objectives being outlined from RSL's executive leadership:

1. Strategic solutions for increasing revenues
2. Maintaining RSL's profit margins
3. Increasing RSL's fan satisfaction levels
4. Deepen RSL fan and brand loyalty

Products and Services Offered

As a professional soccer organization, RSL ultimately provides entertainment to their fans. They do this through acquiring and investing in key players that can be developed to fit the organization's winning strategy. RSL also trains players to prepare for the season and upcoming matches while coaches carefully manage the game to provide the best entertainment to RSL's increasingly loyal fans.

Real Salt Lake's Utilization of a Data Warehouse

Data warehousing projects are driven by business needs such as increased revenue, improve customer satisfaction which are some of the strategic objectives for the soccer organization. Datawarehouse planning drives efficient management and the enterprise data warehouse bus architecture decomposes the DW/BI planning process into manageable pieces by focusing

on business processes and objectives, while delivering integration via standardized conformed dimensions that are reused across processes. The bus architecture is technology and database platform independent.

Bus Matrix and Prioritized Requirements

The organization's requirements are to increase revenue, maintain profit margins, increase fan satisfaction, and increase fan loyalty. The business processes to achieve these requirements are to acquire/divest in players, develop players, prepare for the season, prepare for the match, manage the team, and to entertain fans. The conformed dimensions that support our business requirements are as follows:

- Date
- Player
- Event
- Location
- Player position
- Game Location
- Shooter
- Goalie

The initial bus matrix is used in the requirement phase of the data warehouse lifecycle for communicating requirements and also for prioritization of tasks across the team. The bus matrix provides a graphical overview of RSL's core processes and sub-processes. Currently, our project will focus on creating the initial data warehouse using the conformed dimensions EVENT and PLAYER. The implementation of the other dimensions is planned for the future.

Logical Model Strategy

For the logical design of RSL's organization, we have created a dimensional model that presents data in a framework that is intuitive for data access and also allows high performing queries. The advantage of using this model is that we can store data in such a way that it is easier to store and retrieve the data once stored in a data warehouse. The dimensional model is designed to read, summarize, analyze numeric information like values, balances, counts, weights, etc. in the data warehouse.

The following diagram represents the basic dimensional model for the soccer organization which was created using MySQL workbench:

Logical Model Diagram

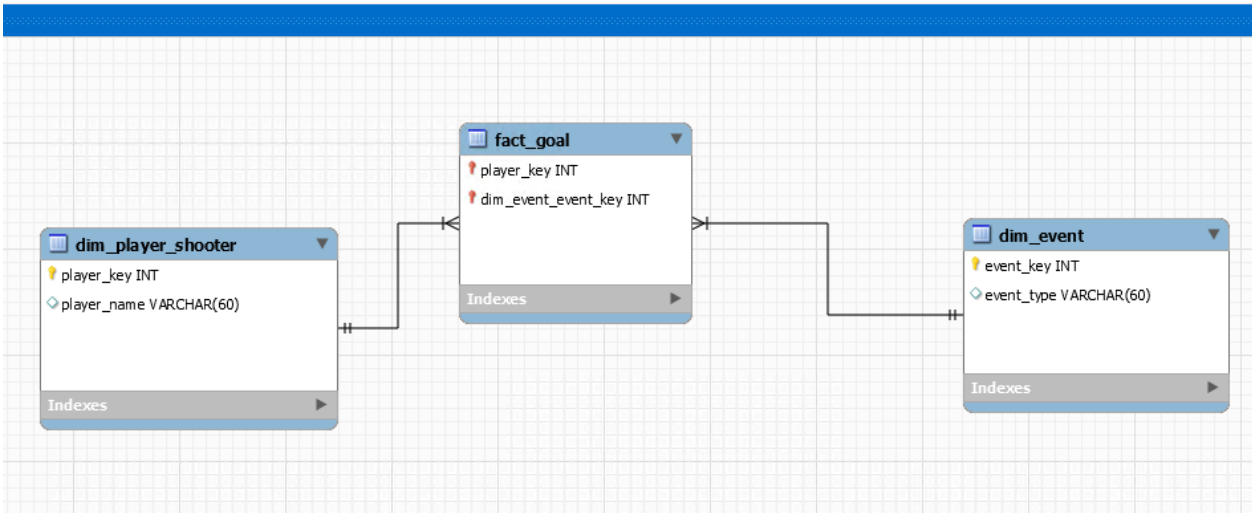

Figure: Logical Model of soccer organization

Elements of the Logical Data Model

Fact Table

In our data warehouse, the fact table is the Fact_Goal. It contains the measurements/metrics from the business processes and also foreign keys to dimensional tables for our warehouse, the measurement would be count of goals.

Dimension Tables

Dimensions provide the context surrounding a business process event. In simple terms, they give the who, what, where of a fact. In other words, a dimension is a window to view information in the facts. Dimensions offer descriptive characteristics of the facts with the help of their attributes.

For our data warehouse, there are 2 dimensions:

- Event Dimension: It contains all the unique events taken place during a soccer game such as Goals, Pass, Fouls, etc.
- Player Dimension: It contains the unique values of all the players which are described by a player_key.

Attributes

The attributes are the various characteristics of the dimension. Attributes are used to search, filter, or classify facts. In the event dimension, the most important attribute present is the 'Event Type'.

Physical Design and ETL Processes

The physical design phase deals with the effective way of storing and retrieving the data. In the physical design, the logical design needs to be converted into a description of the physical database structure. At this time, the following mapping takes place:

- Entities to tables
- Relationships to foreign key constraints
- Attributes to columns
- Primary unique identifiers to primary key constraints
- Unique identifiers to unique key constraints

End-to-end Architecture

Data Sources

The data sources for the warehouse are flat files called "class event" and "class location" which are tab separated. These flat files are loaded into the database using Pentaho Data Integration.

ETL Processes

The ETL process for the warehouse has been divided into 2 parts.

1. Dimensional Tables Creation: The below diagram represents the ETL process for populating the dimension tables.

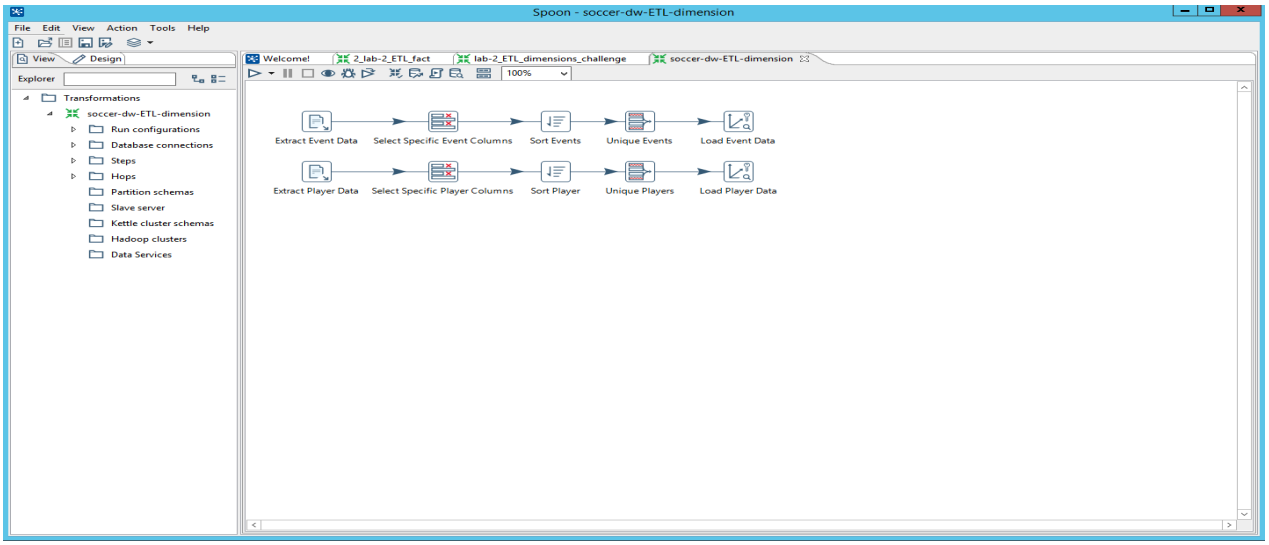

Figure: ETL Process for Dimensional Tables Creation

Both the event and player dimensions have been created using the "class event" flat file. For creating the event dimension, the 'event type' attribute was selected in the step 'Select Specific Event Columns' which was then sorted, and only unique values were selected and loaded into the dimension table. Each event type has its own unique event key. Similarly, for creating the player dimension, only the player names were selected and loaded into the dimension table. Each player name has its own player key.

2. Fact Table Creation:
   The below diagram represents the ETL process for fact tables:

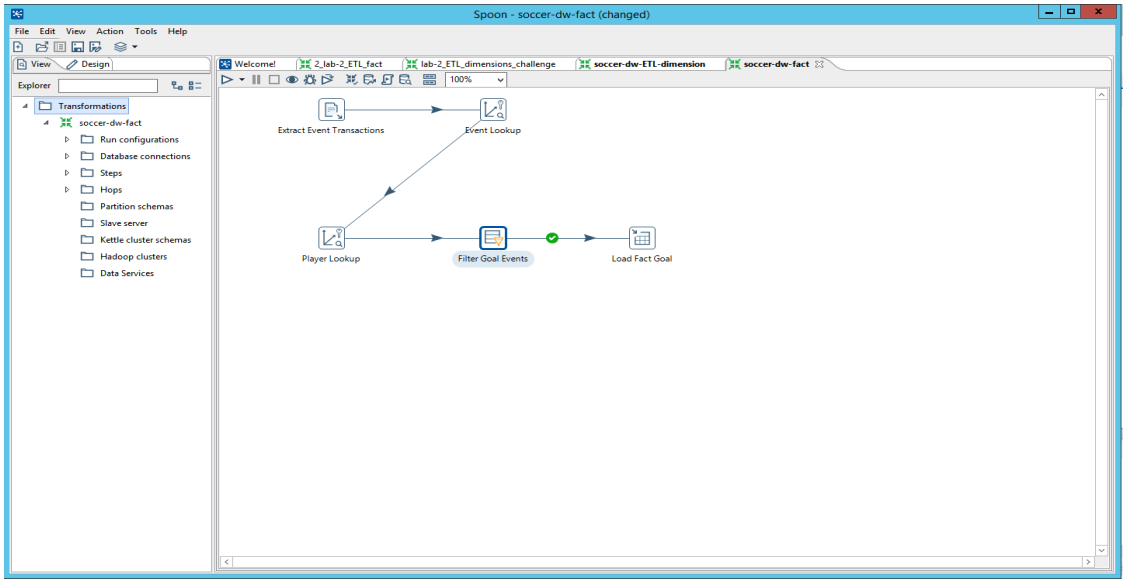

Figure: ETL Process for Fact Table Creation

Fact table creation requires a combination of "class event" data and lookup from event and player dimensions. Once both lookups have been completed, we have filtered the data to be loaded into fact based on event key equal to '31' as the event type 'Goal' has the event key 31. Based on the filter, 108 rows get loaded into the table 'fact_goal'.

Physical Dimensional Model

Once the dimensional table and fact table ETL processes have been created, we ran the SQL query generated for each process. This resulted in the creation of the physical tables in the schema 'soccer_dw' allowing us to load the data from the source to the target database. The physical model is required for the actual implementation of the database.

OLAP Cube

An OLAP cube is a method of storing data in a multidimensional form, generally for reporting purposes. In OLAP cubes, data (measures) are categorized by dimensions. OLAP cubes are often pre-summarized across dimensions to drastically improve query time over relational databases. Although it stores data like a traditional database does, an OLAP cube is structured very differently. OLAP cubes are used by business users for advanced analytics.

The OLAP cube for RSL has been created using the Pentaho Schema Workbench. Currently, in the scope of the project, the OLAP cube 'goal' has been created which measures the total count of goals for each player based on the goal event key. The fact table 'fact_goal' has been used for the creation of the cube GOAL.

The schema 'soccer_dw' also contains the dimension 'player' which has been created using the table 'dim_player'. The dimension player contains the names of all the players present at the target database.

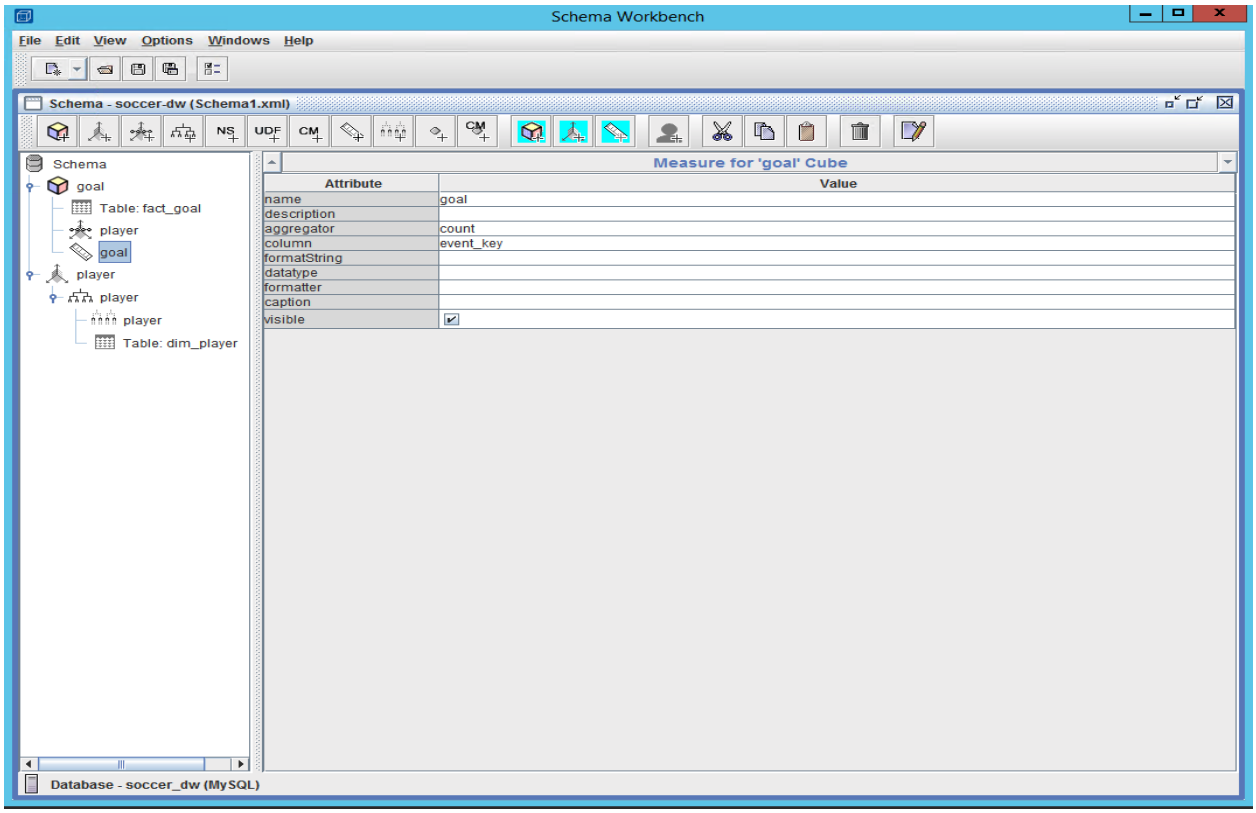

Figure: OLAP Configuration in Schema Workbench

User Interaction Components

There are various user interaction components created for ease of access and query performance.

JPivot View

An existing schema can be explored in Pentaho using JPivot. It allows users the ability to perform typical OLAP navigations such as slice and dice, drill down and roll up. It allows users to dynamically explore a database via an intuitive web interface.

Based on the warehouse schema ‘dw_soccer‘, a Jpivot has been created which contains the number of goals for each player. This can be used to perform OLAP navigations such as rollup, like total number of goals for all the players, or drill down to the number of goals for a particular player.

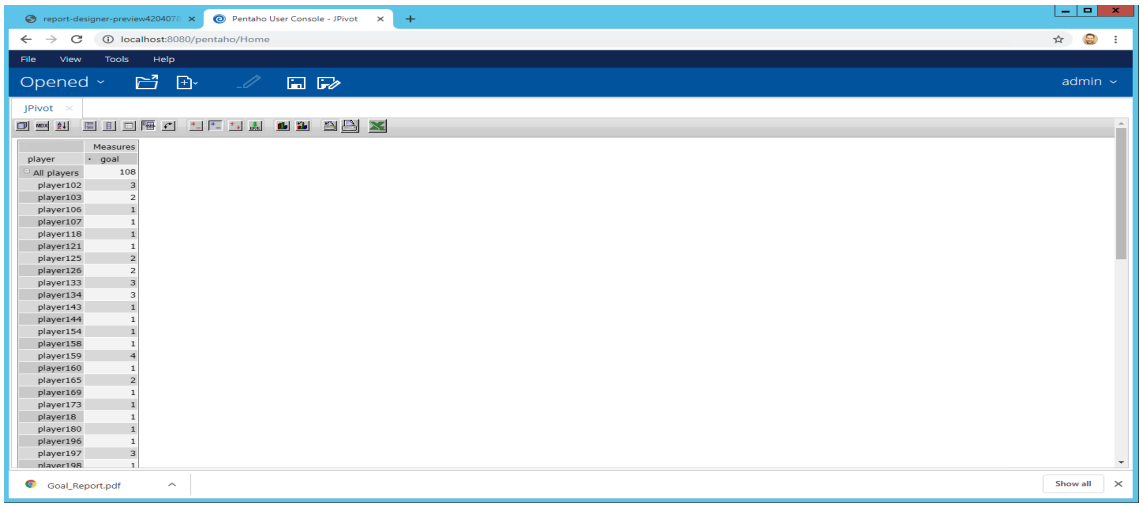

Figure: JPivot view for count of goals

Reports

Using Pentaho Report Designer, sophisticated highly detailed, print quality reports can be created for the end users. These computer-generated reports easily refine data from various data sources into a readable form.

For RSL data reporting, firstly, the report designer wizard has been used to create a basic goal report containing number of goals grouped by each player. A query containing player key, event key and count of event key i.e. goal was used to generate the data in the report.

PDF Goal Report

A view of the goal report can be found below.

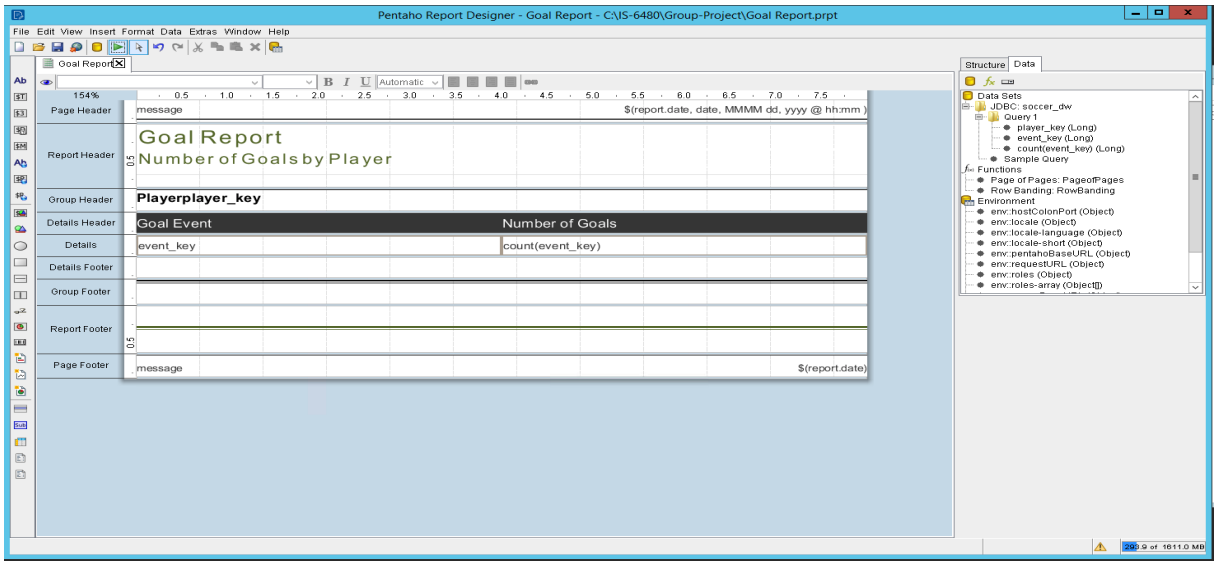

Figure: Goal Report Structure

The report was then published to the Pentaho BI server under the folder 'soccer-reports'. The HTML output of the report is as shown below:

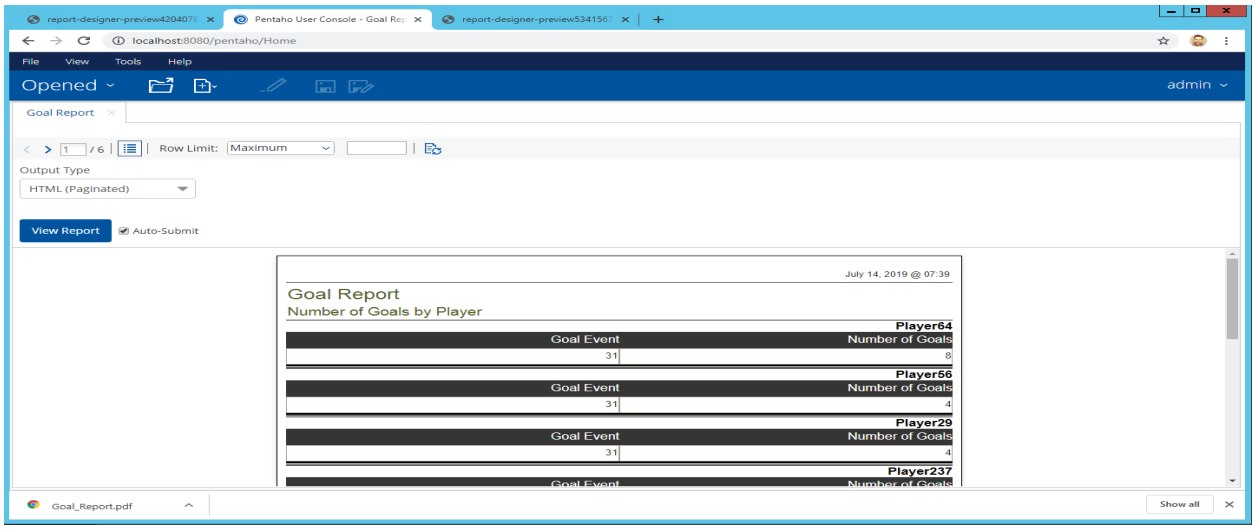

Figure: HTML output of Goal Report

Security Features

To provide security and restrict data access to the warehouse, role-based security has been added to the web instance. A role can be defined as the constraint that restricts access on a row level for specific users or roles.

For the RPL data warehouse, there are two primary roles which have been created:

1. Executive: Executive role has been created to restrict data access to the RPL data warehouse. Only the users which have role as 'Executive' have the access to read the content of the report and execute the report. However, they do not have permission to delete the report.

   For example, the owner of RSL Dell Hansen has been given executive access to the report and is able to view the Goal Report.

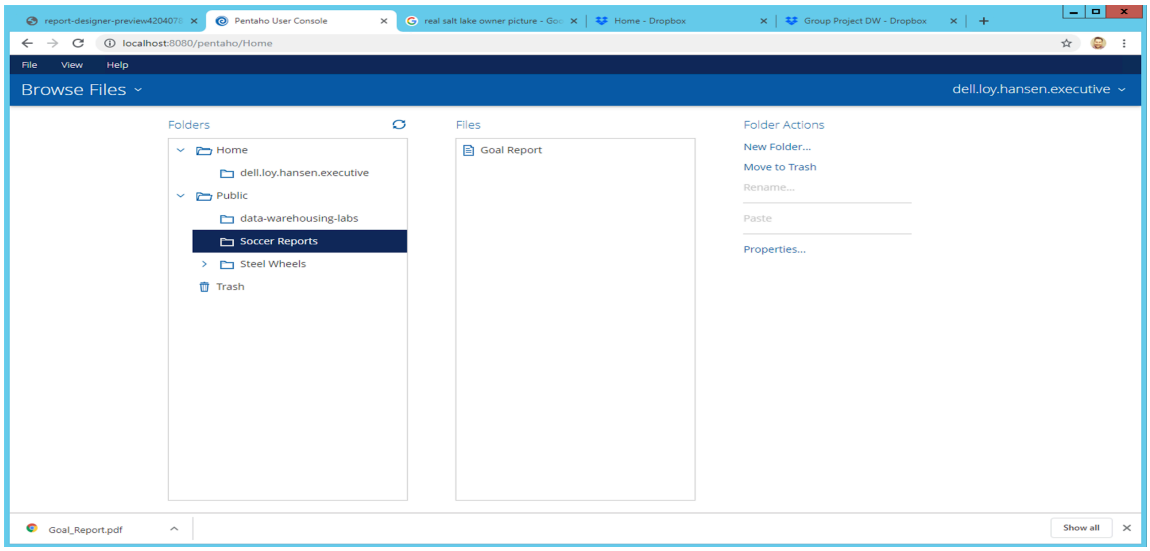

Figure: Executive Role Access to Goal Report

2. Non-Executive: Non-executive users are not authorized to access the data of the RPL data warehouse. Therefore, they do not have permission to view the report or its contents.

For example, the user 'leo-lion' does not belong to the executive team and does not have access to the goal report in the soccer reports folder.

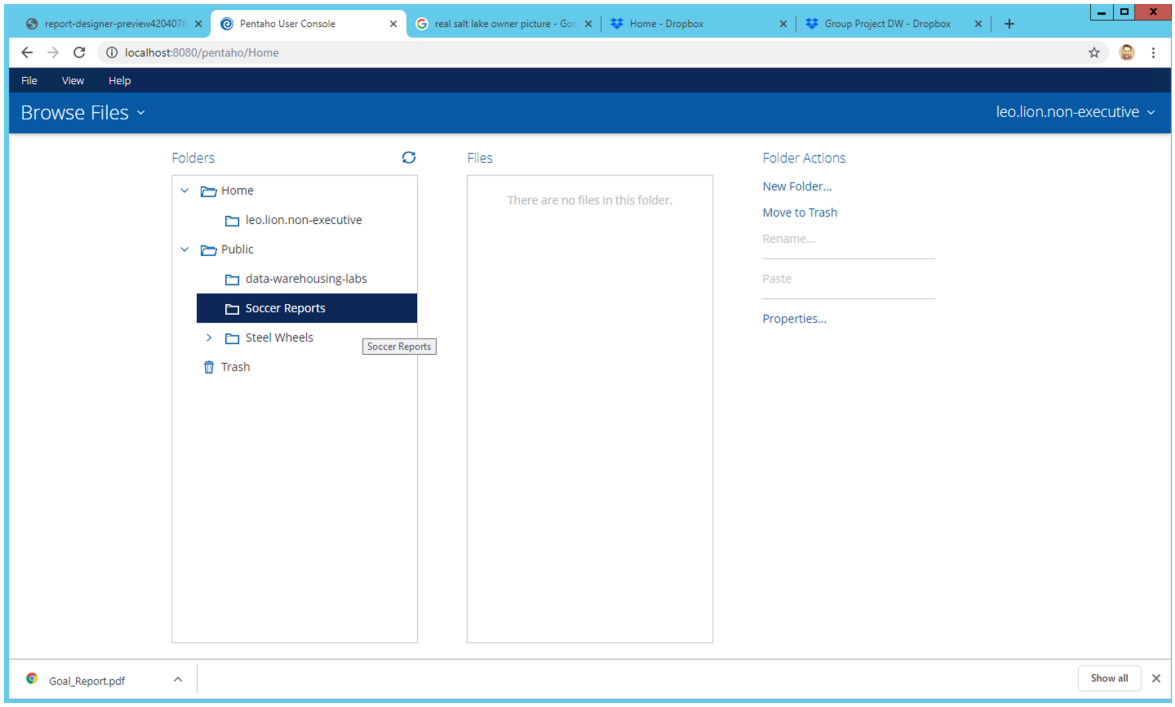

Figure: Non-Executive roles restricted from accessing Goal Report

Solution Approach

Ralph Kimball's dimensional design approach was used for implementing the data warehouse. The design of the data warehouse includes selection of appropriate software and hardware, selection of data and building a logical interface between the warehouse and the analysis tools. For our project, initially a bus matrix was designed which describes the main processes of the organization and its corresponding sub-processes. Based on the bus matrix, the current scope of the project was decided to include dimensions 'event' and 'player' and fact table corresponding to the event 'goal'. The data warehouse lifecycle was used to reach the solution i.e. implement a successful warehouse so that the end users are able to view reports based on their requirements. Each step of the lifecycle was implemented in the following manner:

1. Planning: Discussing the objectives and vision of the organizations and performing an initial feasibility analysis. It also included developing a project plan task list.
2. Requirements: Initial bus matrix was created to discuss the business processes and sub-processes and consequently decide the business process to focus on.
3. Logical Design: Dimensional Model was developed to understand the relation between fact and dimensions and also design an appropriate table structure and key relationships.
4. Physical Design: Based on the Dimensional Model, SQL scripts were created for the table structure creation at the target database.
5. Data Integration: It included implementation of the ETL processes so that data can be extracted from flat files, transformed based on business logic and loaded into target database.
6. Analysis: OLAP cube have been created so that users can perform standard and adhoc analyses.

Technologies Used to Achieve Data Warehouse

- MYSQL Workbench: MYSQL Workbench is an integrated tool environment used for Database Design & Modelling, SQL Development, Database Administration and Database Migration.
- Pentaho Data Integration: Pentaho Data Integration (PDI) is the component of Pentaho responsible for the Extract, Transform and Load (ETL) processes. It is also used for data cleansing, loading data massively into databases and migrating data between applications or databases.
- Pentaho BI Server: Pentaho Server is a business intelligence (BI) software that provides data integration, OLAP services, reporting, information dashboards, data mining capabilities.
- Schema Workbench: The schema workbench is a designer interface that allows users to create and test OLAP schema cubes visually.It provides the following functionality:
    - Schema editor integrated with the underlying data source for validation. (See above)
    - Test MDX queries against schema and database
    - Browse underlying databases structure
- Pentaho Report Designer: Pentaho Report Designer is a sophisticated report creation tool that you can use standalone, or as part of the larger Pentaho Business Analytics distribution. It enables professionals to create highly detailed, "pixel-perfect" reports based on adequately prepared data from virtually any data source.

Sample Data Used to Populate the Data Warehouse

There were 2 sets of data used to populate the data warehouse:

- Event dataset
- Player Location dataset

In the current scope of the project, we have used the event dataset. The event dataset is a tab separated flat file containing 10 attributes. It contains the game date with the names of teams playing example: team18 and team2.

Each event type taken place at the game along with the event time and which player performed the event is present in the dataset.

Current and Future Reports / Analyses

Currently Functioning Reports and Analysis

The report which is currently functional is called the 'Goal Report'. This report provides the number of goals in total and can be drilled down to the number of goals for each player. This report can be developed further to analyze whether good offensive production leads to increase in fan satisfaction and revenue.

Future Functional Reports and Analysis

Event dataset contains information related to event type, players, game date and number of team played. Based on that, making slight modifications in the ETL process, reports such as maximum number of events occurring in a game can be created which can be useful in examining whether a particular event type helps in increasing fan satisfaction. Also, other reports/analyses related to event distribution in each game period, teams having most defensive/offensive play and the relation of the play with the revenue generated for the team can also be designed.

On adding additional dimensions such as location and time and specific players such as shooter and goalie to the data warehouse, many reports can be created. Next step of the project includes implementation of dimension table location and adding it to the OLAP schema and based on that reports for business processes such as 'Manage Injuries' and 'Entertain Fans' can be developed.

Executive Summary

In the current scope of the project, a data warehouse has been implemented based on the steps of the Kimball data warehouse architecture starting from loading the data from the source system such as flat file up to creation of report. ETL processes have been created to load the dimension tables 'Player' and 'Event' and fact table 'fact_goal' at the target database. For ease of access and increased query performance, OLAP cube 'Goal' provides the count of goals for each player. This OLAP cube has been used to create user interaction components such as Pivot Views and Reports. To prevent illegal data access, additional data security roles have been designed at the report level so that it can be viewed by authorized executives only.

Based on the initial data warehouse, future plan of the project includes increasing the functionality of the warehouse by adding dimensions and facts based on location, date and game details. These added functionalities will widen the scope of the project tremendously and will help in answering questions related to the organization's functionality.

Appendix

Bus Matrix

	Date	Player	Shooter	Goalie	Event	Location	Event-time	Game-location
Manage Players								
acquire	x	x	x		x	x	x	
divest	x	x	x		x	x	x	
Develop Players								
experience	x	x	x	x	x	x	x	
skills	x	x	x	x	x	x	x	
Manage Injuries								
substitutions	x	x	x		x	x	x	x
Manage Fitness								
conditioning	x	x	x		x	x	x	x
Prepare for Season								
fitness test	x	x	x		x	x	x	x
Prepare for Match								
practice	x	x	x		x	x	x	x

Manage Player Personnel Tactics								
Coach calls plays	x	x	x	x	x	x	x	x
Manage Game / Opponent tactics								
Player match-ups	x	x	x	x	x	x	x	
Entertain Fans								
goals	x	x	x	x	x	x	x	x

Group Project Approach and Course Material

The following course material used to define logical model, physical model, ETL process, OLAP cube, Reports, Web Reports and Permissions.

Group Project Deliverable	Course Material Used	Description
Real Salt Lake's Utilization of the Data Warehouse	DWL Planning Phase Lecture Slides	The lecture slides related to DWL planning phase were referred to for creating a relationship between DW/BI and organization's objectives.
Data Warehouse Bus Matrix	DWL Requirement Phase Lecture Slides	The lecture slides helped in understanding the need for bus matrix and also provided an example of a bus matrix and how it should be created.
Logical Model	Lab 1 & DWL Logical Design	Lab 1 exercises were utilized to assist by example and practice on how to create the logical model that would later be used to build the physical tables.
Physical Model	Lab 1 & DWL Physical Design	Lab 1 exercises were utilized to assist by example and

		practice on how to create the physical model that would later be used by the ETL process.
ETL Process	Lab 2 & DWL data Integration	Lab 2 exercises were used to assist in creating the ETL process that would extract player and event data.
OLAP Cube	Lab 3	Lab 3 exercises were used to assist in creating the OLAP cubes and accessing the drill down JPivot report.
Reports, Web Reports, and Permissions	Lab 4 & DWL Analysis Phase	Lab 4 exercises were used to help in generating PDF reports, establishing web reports and setting up user/role permissions.

Hours Spent Details

Date	Team Member	Hours Spent	Description of work	Additional Comments
Priyanka Mitra				
06/18/2019	Priyanka	1.5	- AWS Redshift tutorial	NA
07/08/19	Priyanka Mitra	5.5	Working session with group. Redshift walkthrough, re-assess strategy, begin outlining work, start on dimensional models, physical models, project summary and slides.	

07/08/19	Priyanka Mitra	1	Started working on the write up of the Organizational details for the project summary	
07/13/19	Priyanka Mitra	1.5	Worked on the Logical Design part of the Project Summary	
07/14/19	Priyanka Mitra	2	Refine bus matrix, review slides, discuss assigned slides	
07/15/19	Priyanka Mitra	5.5	Worked on the Project Summary Report	
07/16/19	Priyanka Mitra	1.5	Completed the Project Summary Report	
07/16/19	Priyanka Mitra	1.5	Meet with group to rehearse presentation and review project summary.	
07/17/19	Priyanka Mitra	1.5	Final preparation and assignment turn in	
Sion Kim				
06/18/2019	Sion Kim	1.5	AWS Redshift tutorial	
07/08/19	Sion Kim	5.5	Working session with group. Redshift walkthrough, re-assess strategy, begin outlining work, start on	NA

			dimensional models, physical models, project summary and slides.	
07/08/19	Sion Kim	1	Started working on the write up of the Organizational details and BUS matrix for the project summary	
07/14/19	Sion Kim	1	Created the template for the slide presentation	
07/14/19	Sion Kim	1	Added content to slides opening slide, RSL, business problem, tools, process	
07/16/19	Sion Kim	1.5	Final review of project summary and Group Project Approach and Course Material section.	
Darrell (Dal) Price				
06/16/19	Darrell Price	1.25	Redshift walkthrough	
07/08/19	Darrell Price	5.5	Working session with group. Redshift walkthrough, re-assess strategy, begin outlining work, start on dimensional models, physical models, project summary	

			and slides.	
07/08/19	Darrell Price	1	Continue working on dimensional models and physical models	
07/09/19	Darrell Price	4	Got the basic fact table working. Adding additional dimensions and start building OLAP cubes.	
07/14/19	Darrell Price	2	Completed building Goal Report via Report Builder and published report to Pentaho web interface. Setup roles for two users (exec and non-exec) with view permissions.	
07/14/19	Darrell Price	2	Refine bus matrix, review slides, discuss assigned slides	
7/15	Darrell Price	3	Add to Slides: Logical Model Physical Model / ETL Process OLAP Cubes Sample Reports	
7/15	Darrell Price	1.5	Add and Refine Project Summary - Add cover page add table of contents modify headers	
7/16	Darrell Price	1.5	Meet with group to rehearse presentation and review project	

			summary.	
7/16	Darrell Price	1.5	Final review of project summary and Group Project Approach and Course Material section.	
7/17	Darrell Price	1.5	Final preparation and turn in assignment.	