

# Data Warehouse for Soccer Team

Gaurav Kutemate  
Abhijay Sharma  
Angel Yang

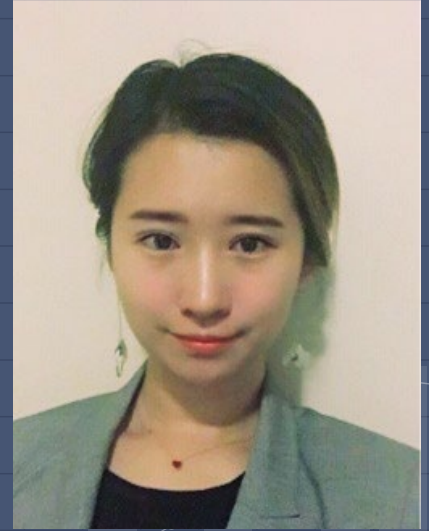




Abhijay Sharma



Gaurav Kutemate



Angel Yang

# Strategic Objective

Having high offensive production has a high impact on fan satisfaction and a small impact on revenue



# Approach Description

- ▣ Use of Pentaho Ecosystem
- ▣ Analyze the impact of various metrics
- ▣ Provide organised information
- ▣ Support Management to achieve goals



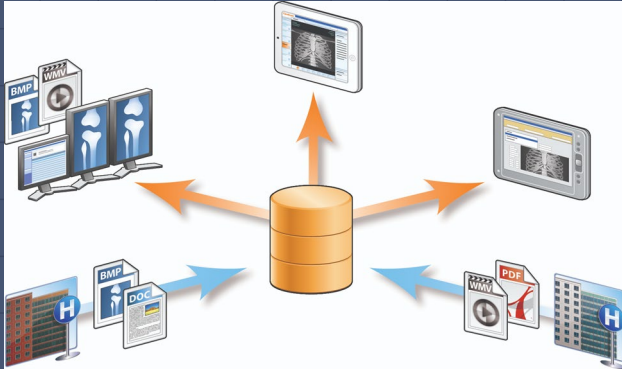
# The EPL

- ▣ Hotly Contested
- ▣ World's best
- ▣ Heavy viewership
- ▣ Huge fanbase
- ▣ Offensive production a key



# Why a Data Warehouse?

Need for Centralised storage



Analytical abilities



Support Soccer operations

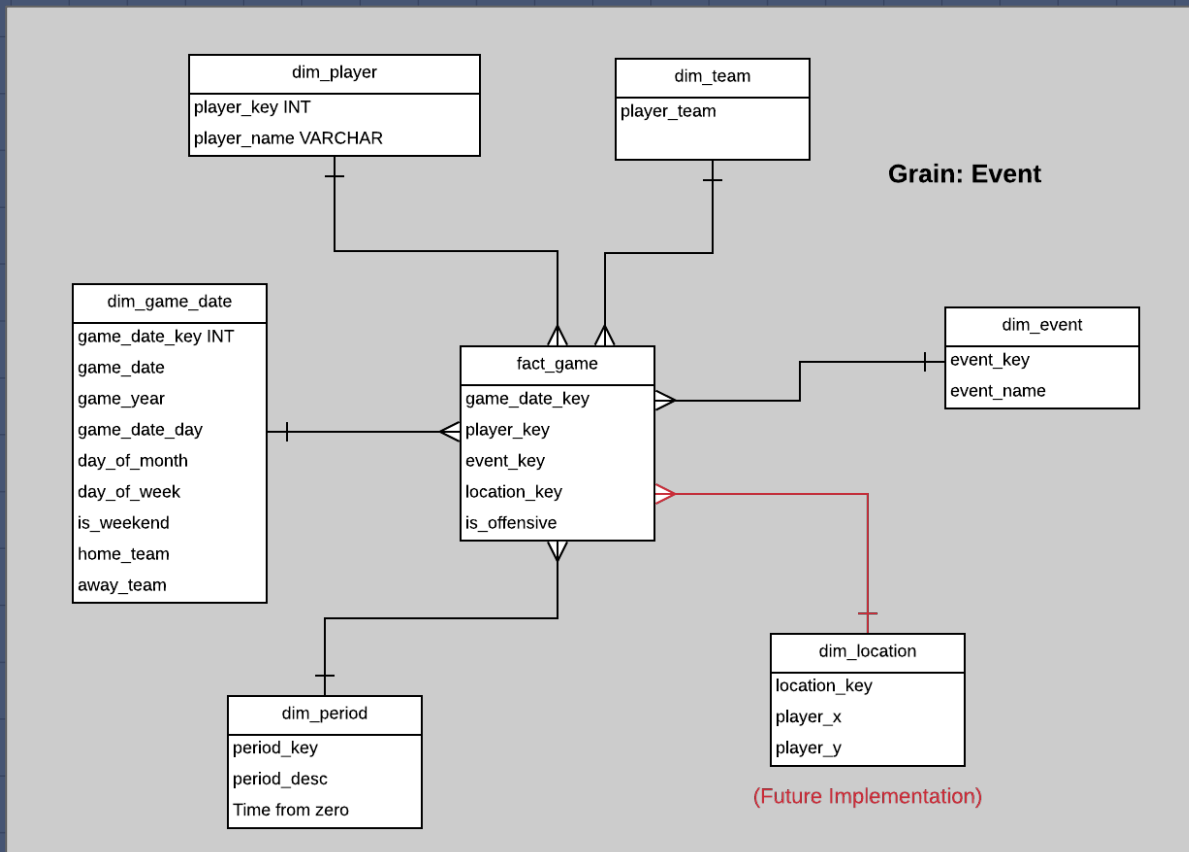


# Bus Matrix

Business processes	Game Date	Team	Player	Location	Period	Event
<b>Manage player personnel strategy</b>						
Acquire players		✓	✓			
Divest players		✓	✓			
<b>Develop players</b>						
On pitch, game time training	✓	✓	✓			✓
Off pitch Training			✓	✓		✓
<b>Manage injuries</b>						
Physical condition management			✓			
<b>Manage fitness</b>						
Diet management			✓			
<b>Manage player personnel tactics</b>						
Goalie Tactics			✓		✓	
Forward Tactics			✓	✓	✓	
Midfield Tactics			✓	✓	✓	
Defensive Tactics			✓	✓	✓	
<b>Manage game/opponent tactics</b>						
Formations managements			✓	✓		
Player instructions			✓	✓		
Fouls management	✓					✓
<b>Entertain fans</b>						
Offense productions	✓	✓	✓	✓	✓	✓
Goal productions	✓		✓			✓



# Logical Model





# What have we accomplished?

- ▣ Created Dimensional Model
- ▣ Conceptualized strategic objectives
- ▣ Built Player analysis
- ▣ Team selection procedure



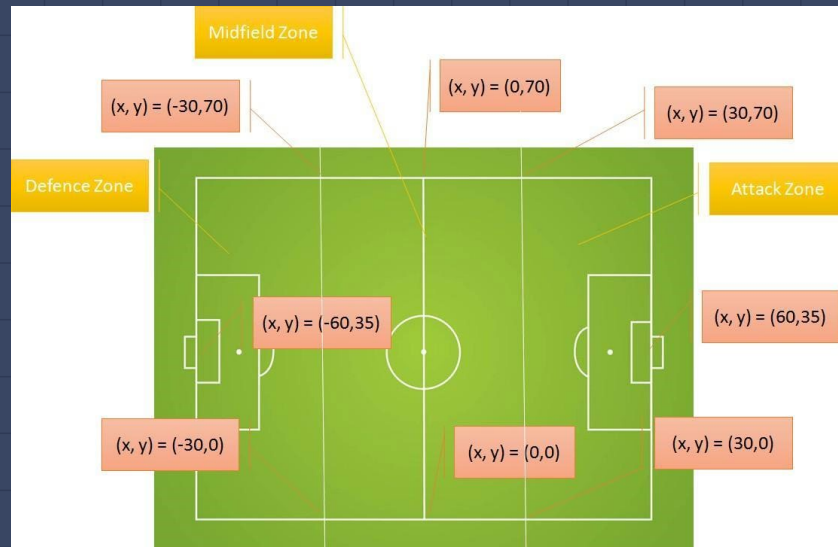
# What needs refinement?

- ▣ Injury management
- ▣ Team utilization



# Planned for future

- ▣ Combine event and location
- ▣ Zoned categorization



# Physical Design

- ▣ Creating tables using MySQL
- ▣ ETL in Pentaho Data Integration
- ▣ OLAP cubes in Schema Workbench
- ▣ Reports and Analyses



# Creating SQL tables

13

The screenshot displays the SQL Server Enterprise Manager interface. The left pane shows the 'SCHEMAS' tree with 'soccer\_data' expanded. The center pane shows a SQL query in 'SQL File 11'.

```
1 use soccer_data;
2 SELECT player_name, count(event_type), event_type
3 from fact_event f
4 join dim_team d on f.team_key=d.team_key
5 join dim_event e on f.event_key = e.event_key
6 join dim_period p on f.period_key = p.period_key
7 join dim_player py on py.player_key=f.player_key
8 where player_1_team_gen='team16'
9 group by player_name,event_type
10 order by player_name desc;
```

The right pane shows the 'SQLAdditions' tab with a message: 'Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.'

Below the query, the 'Result Grid' shows the results of the query:

player_name	count(event_type)	event_type
player96	13	Block
player96	1	Clearance
player96	1	Corner Cross
player96	1	Cross
player96	1	Direct Free Kick Cross
player96	12	Direct Free Kick Pass
player96	3	Dribble
player96	7	Foul
player96	1	Handball
player96	15	Header
player96	2	Header Shot
player96	2	Indirect Free Kick Pass
player96	287	Pass

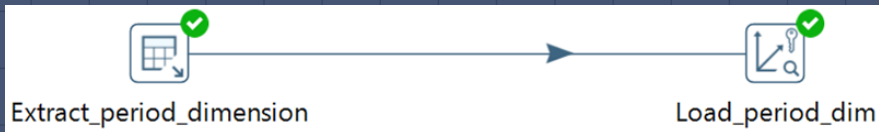
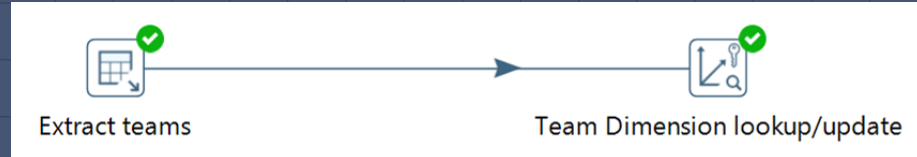
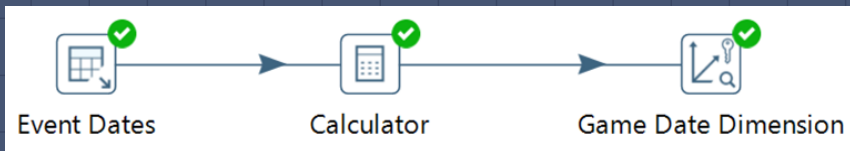
At the bottom, the 'Output' pane shows the 'Action Output' table:

#	Time	Action	Message	Duration / Fetch
1	02:56:58	SELECT player_name, count(event_type), event_type from fact_event f join dim_team d on f.team_key=d.team...	Error Code: 1046. No database selected Select the default DB to be used by double-clicking its name in the SCH...	0.016 sec
2	02:57:17	use schema soccer_data	Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds to your MySQL serv...	0.000 sec
3	02:57:23	use soccer_data	0 row(s) affected	0.000 sec
4	02:57:23	SELECT player_name, count(event_type), event_type from fact_event f join dim_team d on f.team_key=d.team...	224 row(s) returned	0.454 sec / 0.000 sec



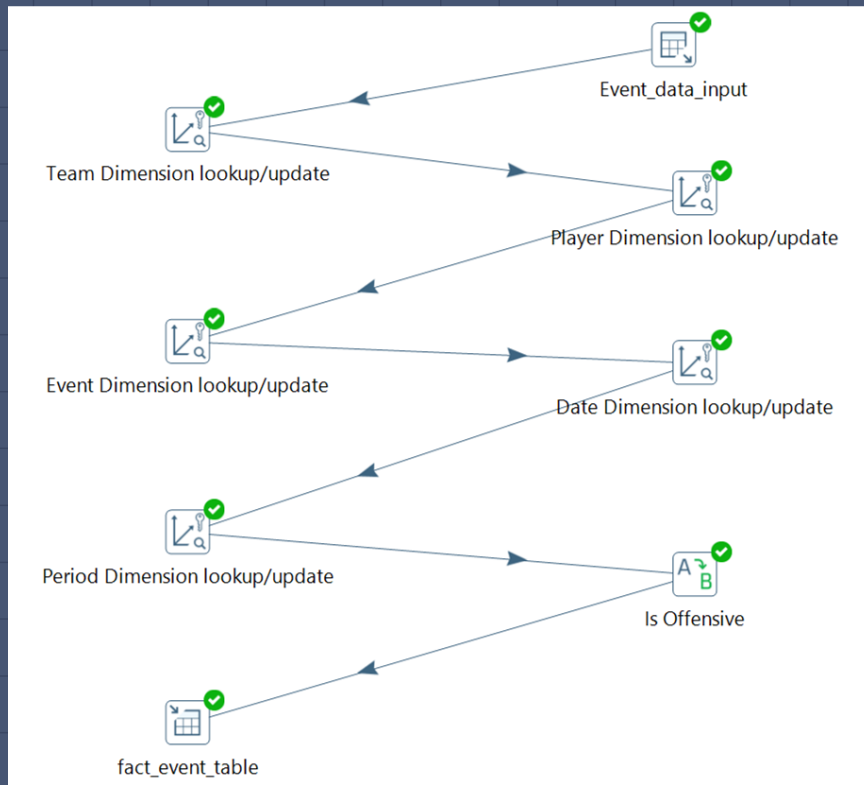
# ETL TransformationsDimensions

14



# ETL Transformations Fact Table


15



# OLAP Cubes

16

JPivot x



dim player	dim game date	dim event	dim period	dim team	All is offensives	Not Offensive	Offensive
All dim players	All dim game dates	All dim events	All dim periods	team16	8,610	8,220	390
player107	All dim game dates	All dim events	All dim periods	team16	946	916	30
player140	All dim game dates	All dim events	All dim periods	team16	568	555	13
player151	All dim game dates	All dim events	All dim periods	team16	603	595	8
player19	All dim game dates	All dim events	All dim periods	team16	2	2	
player202	All dim game dates	All dim events	All dim periods	team16	1,361	1,275	86
player206	All dim game dates	All dim events	All dim periods	team16	471	457	14
player217	All dim game dates	All dim events	All dim periods	team16	405	401	4
player260	All dim game dates	All dim events	All dim periods	team16	371	360	11
player264	All dim game dates	All dim events	All dim periods	team16	178	174	4
player296	All dim game dates	All dim events	All dim periods	team16	1,217	1,130	87
			First Half	team16	563	522	41
			Second Half	team16	654	608	46
player297	All dim game dates	All dim events	All dim periods	team16	589	565	24
player43	All dim game dates	All dim events	All dim periods	team16	98	90	8
player49	All dim game dates	All dim events	All dim periods	team16	539	481	58
player63	All dim game dates	All dim events	All dim periods	team16	530	510	20
player80	All dim game dates	All dim events	All dim periods	team16	4	4	
player96	All dim game dates	All dim events	All dim periods	team16	728	705	23

Slicer: [(All)=All home teams]





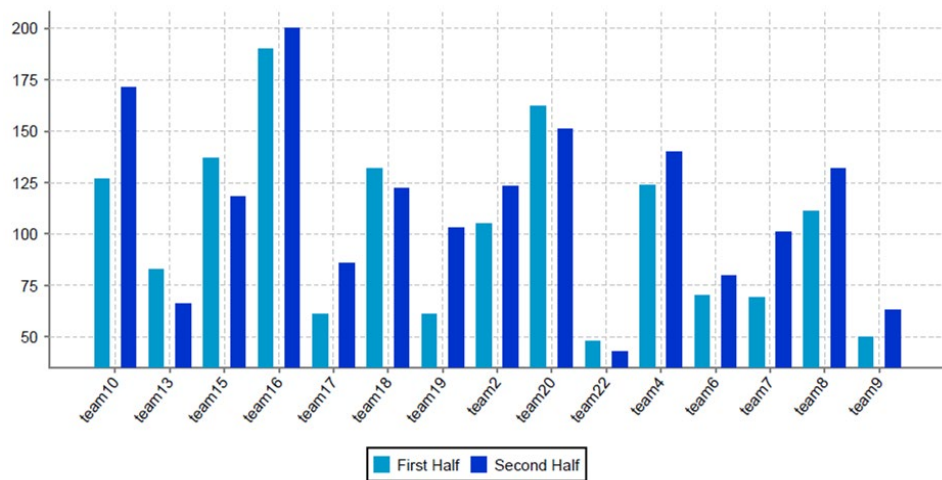
# Reports

17

July 13, 2019 @ 11:15

## Report on Offensive Events of each team

Teams	Period	Number of offensive events
team10	First Half	127
team10	Second Half	171
team13	First Half	83
team13	Second Half	66
team15	First Half	137
team15	Second Half	118
team16	First Half	190
team16	Second Half	200
team17	First Half	61
team17	Second Half	86
team18	First Half	132
team18	Second Half	122
team19	First Half	61
team19	Second Half	103
team2	First Half	105
team2	Second Half	123



# Conclusion....

- Small but a crucial step towards creating an analytics environment
- Capable of being customized
- Serves different processes



# Challenges

- Conceptualizing the project approach
  - Designing the Bus matrix
- EDA of data source due to inconsistencies
  - Finalizing of Dimensions and grain
- Navigating on Pentaho Report Designer



# Questions?

