

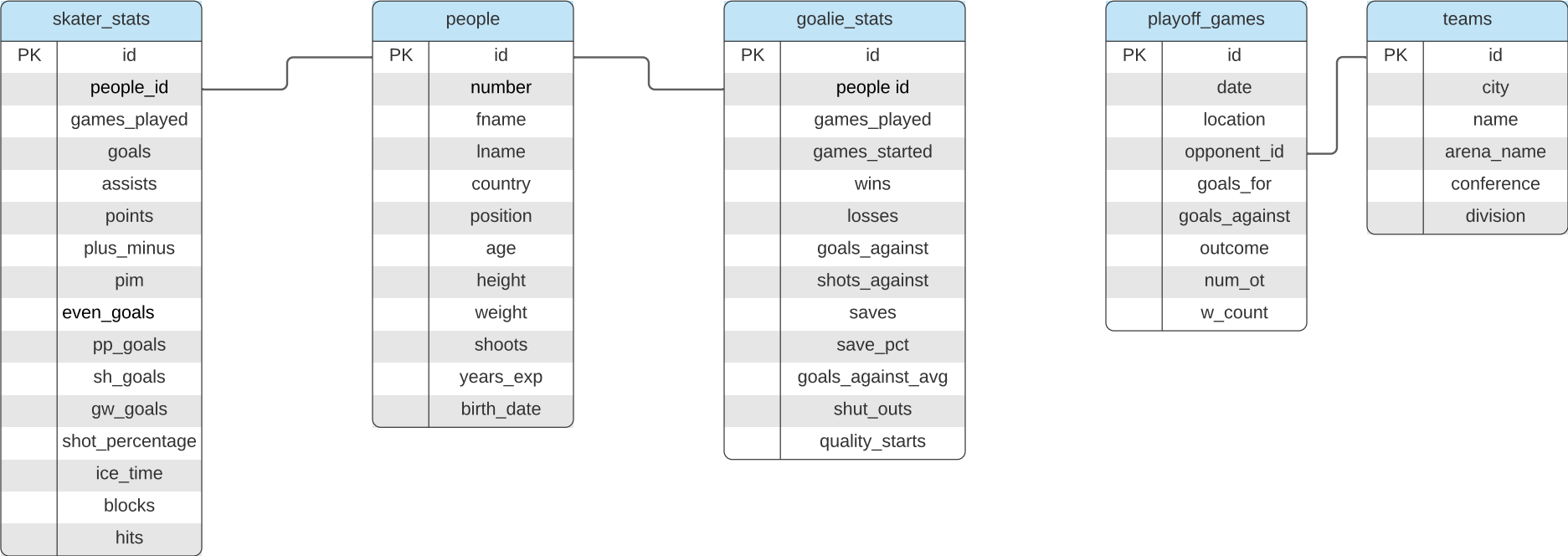
# Detroit Red Wings

## 2007-2008

Final Project

SI 564

Mike VerHulst



# Database Outline

## Data

- The data I chose was sourced from [hockey-reference.com](http://hockey-reference.com)
  - [Link here](#)
  - Downloaded in .xls format then converted to CSV
- The data didn't need much cleaning so I cleaned it by hand before importing it into SQL
  - Deleted columns that were more advanced statistics that most people aren't familiar with
  - Format for the "shoots" column in the final database was "L/-" for skaters and "-/R" for goalies. To make it easier to query, I changed the values to just a single letter in Excel.
  - Date format had to be changed to YYYY-MM-DD so that was converted in Excel before importing

## Table layout

- Chose to put all the players demographic information in one table. Each player had data relating to height, weight, years in the league, country of origin, and more that weren't relevant to their game statistics.
- To avoid making the statistics tables too complicated I included just the people\_id assigned to each player so things like names and jersey numbers weren't duplicated in the database.
- The goalies have their own table because goalie stats are very different than regular player stats.
  - However, goalies can still score points (goals and assists) so they also are represented on the skater\_stats table.
- I also decided to place all the team info such as arena name, conference and more in its own table. This way I wouldn't have that information duplicated over and over in the playoff\_games table.
  - Each playoff series is best-of-seven so there could potentially be a lot of duplicate data if this wasn't done.

## If I had more time

- Would have included more information on the other teams
  - Arena capacity
  - Coaches
  - Their record for the year
  - Rosters
  - Scoring/goalies statistics for the year
  - Basically everything I did for the Red Wings but for the whole league

From: Sarah Dimatto

To: You

Subject: Help me impress my partner's parents!!

Hi Friend,

Christmas is quickly approaching and I just found out my partner's dad will be attending for sinner this year. As you know, since me and my partner are becoming more serious, I want to make a good impression with their parents. I found out that their dad is a HUGE Red Wings fan. I know we have a Red Wings database here at Borromean so could you find the answers to the following questions that my partner says will impress their dad?

1. What are the names of the 2 defensemen with the most assists for the season?
2. What position did our highest goal scorer play and how many goals did they score?
3. Which players scored 20 or more goals AND had more than 50 penalty minutes?
4. Which goalie had the best goals against average?
5. Which city were the Red Wings playing when they scored the most goals in a single playoff game?
6. It takes 16 wins to win the Stanley Cup. What's the name of the arena and the city it's in when the Red Wings won the Stanley Cup?
7. Which player (that is NOT a goalie) had the most ice time this season and how many points did they have?
8. Which players were 35 years or older? Give me their names, age, points, and position.

Thank you so much for your help. I'll have to buy you a couple drinks when we get back and I can tell you how it went!

Sarah

Sarah Dimatto

Salesperson

Borromean Digital

[sdimmatto@borromean.digital](mailto:sdimmatto@borromean.digital)

## Letter from answer key

- What are the names of the 2 defensemen with the most assists for the season?
  - Select p.fname, p.lname, s.assists from people p join skater\_stats s on p.id = s.people\_id where p.position = 'D' order by s.assists desc limit 1
  - Niklas Lindstrom (60 Assists)
- What position did our highest goal scorer play and how many goals did they score?
  - Select p.fname, p.lname, p.position, s.goals from people p join skater\_stats s on p.id = s.people\_id order by s.goals desc limit 1
  - LW (Henrik Zetterberg, 40 Goals)
- Which players scored 20 or more goals AND had more than 50 penalty minutes?
  - Select p.fname, p.lname, s.pim, s.goals from people p join skater\_stats s on p.id = s.people\_id where s.goals > 19 AND pim > 49
  - Tomas Holmstrom (58 pim, 20 goals) and Johan Franzen (51 pim, 27 goals)
- Which goalie had the best goals against average?
  - Select p.fname, p.lname, g.goals\_against\_avg from people p join goalie\_stats g on p.id = g.people\_id order by goals\_against\_avg limit 1
  - Chris Osgood (2.09 GAA)
- Which city were the Red Wings playing when they scored the most goals in a single playoff game?
  - Select t.city, p.goals\_for from playoff\_record p join teams t on p.opponent\_id = t.id order by goals\_for desc limit 1
  - Colorado (8 goals)
- It takes 16 wins to win the Stanley Cup. What's the name of the arena and the city it's in when the Red Wings won the Stanley Cup?
  - Select t.city, t.arena\_name from teams t join playoff\_record p on t.id = p.opponent\_id where p.w\_count = 16
  - Mellon Arena in Pittsburgh
- Which player (that is NOT a goalie) had the most ice time this season and how many points did they have?
  - Select p.fname, p.lname, s.points, p.position, s.ice\_time from people p join skater\_stats s on p.id = s.people\_id where p.position != 'G' order by s.ice\_time desc limit 1
  - Nicklas Lindstrom, 70 points (D, 2031 minutes)
- Which players were 35 years or older? Give me their names, age, points, and position.
  - select p.fname, p.lname, p.age, ss.points, position from people p join skater\_stats ss on p.id = ss.people\_id where p.age > 34 order by ss.points desc

	fname	lname	age	points	position
1	Nicklas	Lindstrom	37	70	D
2	Tomas	Holmstrom	35	40	RW
3	Kris	Draper	36	17	C
4	Chris	Chelios	46	12	D
5	Kirk	Maltby	35	10	LW
6	Dallas	Drake	38	6	RW
7	Chris	Osgood	35	3	G
8	Dominik	Hasek	43	1	G
9	Darren	McCarty	35	1	RW

# Final Project Query Log

## Creating tables:

- create table people

```
(
  id int auto_increment,
  number int null,
  fname varchar(255) null,
  lname varchar(255) null,
  country varchar(10) null,
  position varchar(10) null,
  age int null,
  height int null,
  weight int null,
  shoots varchar(10) null,
  years_exp int null,
  birth_date date null,
  constraint people_pk
    primary key (id)
);
```

- create table skater\_stats

```
(
  id int auto_increment,
  people_id int null,
  games_played int null,
  goals int null,
  assists int null,
  points int null,
  plus_minus int null,
  pim int null,
  even_goals int null,
  pp_goals int null,
  sh_goals int null,
  gw_goals int null,
  shot_percent float null,
  ice_time int null,
  blocks int null,
  hits int null,
  constraint skater_stats_pk
    primary key (id)
);
```

- create table goalie\_stats

```
(
```

```

    id int auto_increment,
    people_id int null,
    games_played int null,
    games_started int null,
    wins int null,
    losses int null,
    goals_against int null,
    shots_against int null,
    save_percent float null,
    goals_against_avg float null,
    shut_outs int null,
    quality_starts int null,
    quality_start_percent float null,
    constraint goalie_stats_pk
        primary key (id)
);
• create table playoff_record
(
    id int auto_increment,
    date date null,
    location varchar(10) null,
    opponent_id int null,
    goals_for int null,
    goals_against int null,
    outcome varchar(1) null,
    num_ot int null,
    w_count int null,
    temp int null,
    constraint playoff_record_pk
        primary key (id)
);
• create table teams
(
    id int auto_increment,
    city varchar(100) null,
    name varchar(100) null,
    arena_name varchar(100) null,
    conference varchar(100) null,
    division varchar(100) null,
    constraint teams_pk
        primary key (id)
);

```

## Add temp columns to connect tables

- alter table goalie\_stats  
add temp varchar(255) null;
- alter table skater\_stats

- add temp varchar(255) null;  
 • alter table playoff\_record  
 add temp varchar(255) null;

## Import data

Used the “Import data from file” feature of Datagrip. The data came from CSVs that I created myself after cleaning the raw data by hand in Excel.

## Populate people\_id columns

- UPDATE goalie\_stats gs  
 SET people\_id = (select p.id from people p WHERE p.lname = gs.temp )
- UPDATE skater\_stats ss  
 SET people\_id = (select p.id from people p WHERE p.lname = ss.temp)
- UPDATE playoff\_record pr  
 SET opponent\_id = (select t.id from teams t WHERE pr.temp = t.city)

## Drop temp columns

- alter table skater\_stats drop column temp;
- alter table goalie\_stats drop column temp;
- alter table playoff\_record drop column temp;

## Update values in goalie\_stats (col quality starts)

For whatever reason this field didn't populate properly during the data import

- UPDATE red\_wings.goalie\_stats t SET t.quality\_starts = 26 WHERE t.id = 1
- UPDATE red\_wings.goalie\_stats t SET t.quality\_starts = 1 WHERE t.id = 3
- UPDATE red\_wings.goalie\_stats t SET t.quality\_starts = 19 WHERE t.id = 2

## Add foreign keys

- alter table skater\_stats  
 add constraint skater\_stats\_people\_id\_fk  
 foreign key (people\_id) references people (id)  
 on update cascade;
- alter table goalie\_stats  
 add constraint goalie\_stats\_people\_id\_fk  
 foreign key (people\_id) references people (id)  
 on update cascade;
- alter table playoff\_record  
 add constraint playoff\_record\_teams\_id\_fk  
 foreign key (opponent\_id) references teams (id)  
 on update cascade;

## SQL Dump

- mysqldump -h 35.239.81.122 --port 13501 -u verhulst-rw -p red\_wings > red\_wings.sql