The 2015 Cricket World Cup Database

David Ebert

Tarleton State University

29 November 2016

GitHub: `dpebert7/2015cwc`

Introduction

Cricket Basics

- Two teams of 11 players
- Toss determines who *bats* and *fields* first
- The fielding team *bowls* 50 *overs* of 6 *balls* each
- Two *batsmen* at a time from the batting team score as many runs as possible without losing their *wicket*
- After the batting team loses 10 wickets fall or 50 overs are completed, the teams switch and the fielding team *chases* the batting team's score
- After the second team has batted, the team with the highest score wins

Cricket Basics

- Two teams of 11 players
- Toss determines who *bats* and *fields* first
- The fielding team *bowls* 50 *overs* of 6 *balls* each
- Two *batsmen* at a time from the batting team score as many runs as possible without losing their *wicket*
- After the batting team loses 10 wickets fall or 50 overs are completed, the teams switch and the fielding team *chases* the batting team's score
- After the second team has batted, the team with the highest score wins

  E.g. "South Africa 227/7 (50 overs) defeated
   Pakistan 209/10 (48.5 overs) by 17 runs"

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tournament

- February-March 2015 Cricket World Cup
- Hosted in Australia and New Zealand
- Tournament Structure
  - 2 Pools of 7 teams
  - Stage 1: Round-robin play within pools
  - Stage 2: Seeded 8-team single-elimination playoffs
- 49 matches

2015 Cricket
World Cup
Database

David Ebert

Overview

**Data
Collection**

Database

Extensions

ESPNcricinfo
LIVE SCORES

👍 Like 443 | Tweet

ICC Cricket World Cup, 11th Match, Pool A: Australia v Bangladesh at Brisbane, Feb 21, 2015

Scorecard | **Commentary** | Almanack report | Report | Articles (7) | Photos (8) | Videos (2)

ICC Cricket World Cup - 11th match, Pool A
Australia v Bangladesh

Match abandoned without a ball bowled

ODI no. 3608a | 2014/15 season

Played at Brisbane Cricket Ground,
Woolloongabba, Brisbane
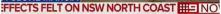
21 February 2015 - day/night match (50-
over match)

Andrew Ramsey ✔
@ARamseyCricket

🐦 Follow

Sad scene at Gabba turnstiles where water not patrons
flooding in. Rain heavier now than for past hour #CWC15

11:18 PM - 20 Feb 2015

↩   🔁 275   ♥ 162

2015 Cricket World Cup Database

David Ebert

Overview

Data Collection

Database

Extensions

CYCLONE MARCIA
EFFECTS FELT ON NSW NORTH COAST NO...

Alt Cricket
@AltCricket

Follow

AUSTRALIA WEATHER UPDATE: Cloudy with a chance of Sharknado.

12:35 AM - 21 Feb 2015

69   49

●●●●● vodafone AU 4G **11:32 am**   68%

underwater cricke...   3

Scott Craig
@Scott_CraigEnig

Follow

The @CricketAus boys are getting some practice in at The Gabba before tonight's game #CWC15 #AUSvBAN

7:35 PM - 20 Feb 2015

15   11

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tournament

- 2015 Cricket World Cup
- Hosted in Australia and New Zealand
- Tournament Structure
  - 2 Pools of 7 teams
  - Stage 1: Round-robin play within pools
  - Stage 2: Seeded 8-team single-elimination playoffs
- ~~49~~ **48** matches
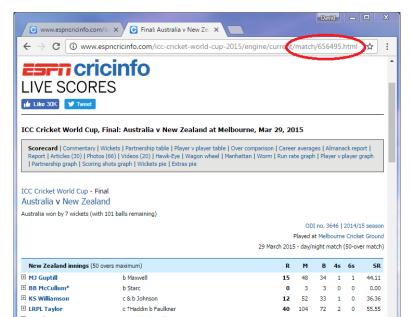- Data Collected from ESPNcricinfo website

Web Scraping - Match Page

David

www.espncricinfo.com/ic...   Final: Australia v New Ze...

www.espncricinfo.com/icc-cricket-world-cup-2015/engine/current/match/656495.html

ESFncricinfo
LIVE SCORES

Like 30K   Tweet

ICC Cricket World Cup, Final: Australia v New Zealand at Melbourne, Mar 29, 2015

Scorecard | Commentary | Wickets | Partnership table | Player v player table | Over comparison | Career averages | Almanack report | Report | Articles (30) | Photos (66) | Videos (20) | Hawk-Eye | Wagon wheel | Manhattan | Worm | Run rate graph | Player v player graph | Partnership graph | Scoring shots graph | Wickets pie | Extras pie

ICC Cricket World Cup - Final
Australia v New Zealand

Australia won by 7 wickets (with 101 balls remaining)

ODI no. 3646 | 2014/15 season
Played at Melbourne Cricket Ground
29 March 2015 - day/night match (50-over match)

New Zealand innings (50 overs maximum)		R	M	B	4s	6s	SR
MJ Guptill	b Maxwell	15	48	34	1	1	44.11
BB McCullum*	b Starc	0	3	3	0	0	0.00
KS Williamson	c & b Johnson	12	52	33	1	0	36.36
LRPL Taylor	c †Haddin b Faulkner	40	104	72	2	0	55.55

Web Scraping - Match Page

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Data Processing

- Match data from ESPNcricinfo website

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Data Processing

- Match data from ESPNcricinfo website
- `python-espncricinfo` module from GitHub used to scrape data

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Data Processing

- Match data from ESPNcricinfo website
- `python-espncricinfo` module from GitHub used to scrape data
- Python list comprehensions and `pandas` used to organize data and add missing columns

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Data Processing

- Match data from ESPNcricinfo website
- python-espncricinfo module from GitHub used to scrape data
- Python list comprehensions and pandas used to organize data and add missing columns
- sqlite3 module used for conversion to SQL

2015 Cricket
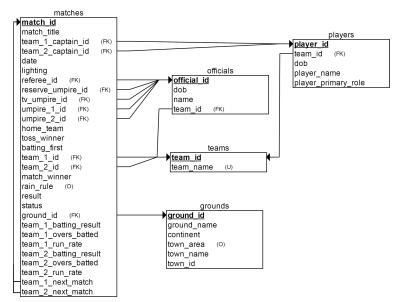World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

ER Diagram

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

matches

```
match_id
match_title
team_1_captain_id     (FK)
team_2_captain_id     (FK)
date
lighting
referee_id        (FK)
reserve_umpire_id     (FK)
tv_umpire_id      (FK)
umpire_1_id       (FK)
umpire_2_id       (FK)
home_team
toss_winner
batting_first
team_1_id     (FK)
team_2_id     (FK)
match_winner
rain_rule     (O)
result
status
ground_id     (FK)
team_1_batting_result
team_1_innings
team_1_overs_batted
team_1_run_rate
team_2_batting_result
team_2_innings
team_2_overs_batted
team_2_run_rate
team_1_next_match
team_2_next_match
```

Tables: Matches

- `match_id` is primary key

- 2 `team_id` keys point to `teams` table

- `match_id` is primary key

- 5 official ids point to `officials` table

- 2 `captain_id` keys point to `players` table

- `ground_id` points to `ground` table

- 2 `next_match` keys point back to `matches` table

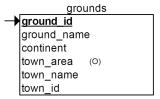- all other columns provide details about the match time, conditions, match status and result

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tables: Grounds

grounds
ground_id
ground_name
continent
town_area   (O)
town_name
town_id

- ground_id is primary key
- town_area is optional
- town_id is presumably a foreign key to another ESPNcricinfo table.

2015 Cricket
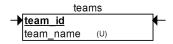World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tables: Teams

```
                    teams
          →▮team_id                 ◄─
           team_name      (U)      ◄─
```

- team_id is primary key
- team_name is team abbreviation
- ESPNcricinfo's database includes additional attributes including team full name, logo, etc.

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tables: Players

players

player_id
team_id   (FK)
dob
player_name
player_primary_role

- player_id is primary key
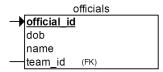- player_primary_role is usually batsman, bowler, or all-rounder

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Tables: Officials

```
              officials
→ official_id
  dob
  name
  team_id    (FK)
```

- official_id is primary key

- team_id matches to teams table.
  (Officials may not officiate their home country's match.)

2015 Cricket
World Cup
Database

David Ebert

Overview

Data
Collection

Database

Extensions

Extensions

- Extend to pool A and B results
- Include more match details for players, including runs scored and wickets takeen
- Scrape ball-by-ball detail
- Scrape ball tracking data
- Streamline data collection and organization