

# Game-Changing NFL Analytics with KEL

North Carolina State University



**NC STATE**  
UNIVERSITY

# Acknowledgements

— — —

## **LexisNexis Risk Solutions**

Jesse Shaw

Ken Kurzweil

Flavio Villanustre

Trish McCall

## **NCSU**

Daniel Mills

Ankit Kumar

Will Eason

**Goal:  
Promote HPCC Systems**



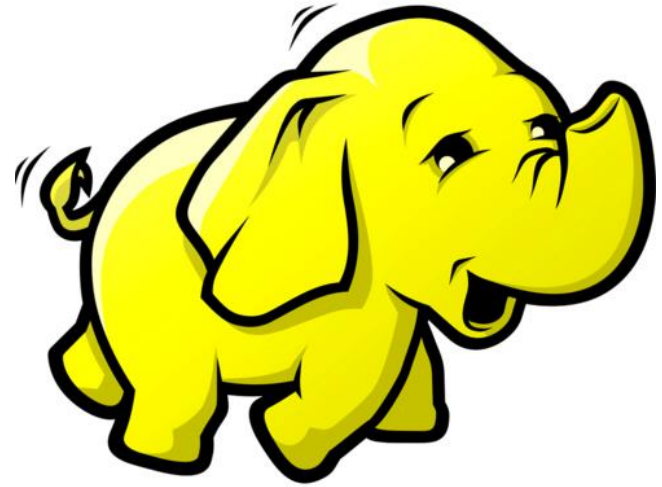
# The problem

---

In my class: students  
choose to use Hadoop over  
HPCC Systems.

Why?

- Better known
- Better documentation
- More jobs





# The solution

Excite.

Lower the barrier.

Bribe.

Blackmail.

---

# How it works

— — —

## Fetch and parse

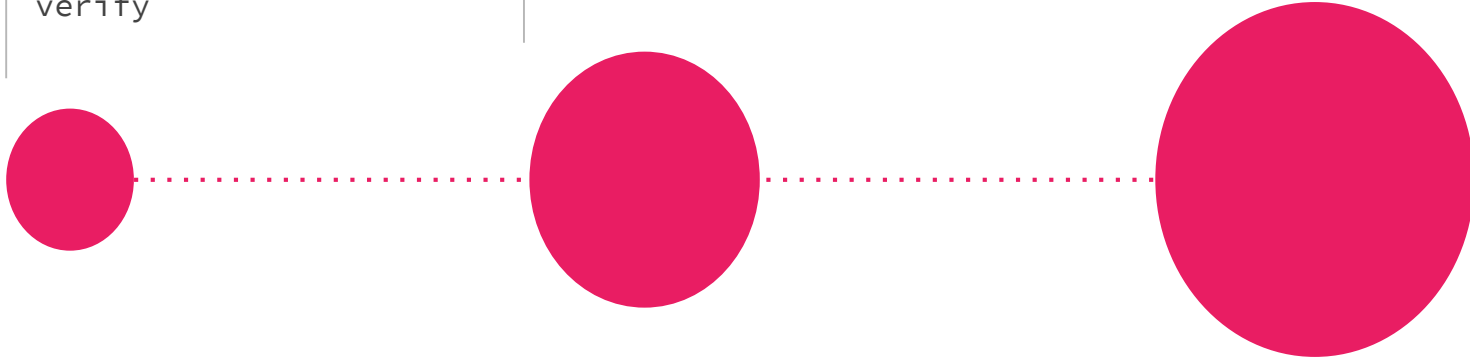
Get data, format, normalize, and verify

## Upload and transform

Upload to cluster, transform for analysis

## Analyze

Run custom KEL and ECL programs



# Model

Using KEL and ECL

## KEL

- Entities
- Associations

## ECL

- Records
- Transformations

---

# Football is discrete

## Carolina Panthers at 09:11

1-10-CAR 43 (9:11) (Shotgun) 1-C.Newton pass short right to 88-G.Olsen to MIN 39 for 18 yards (54-E.Kendricks; 34-A.Sendejo).

1-10-MIN 39 (8:30) (No Huddle, Shotgun) 1-C.Newton pass incomplete deep right to 17-D.Funchess. **PENALTY** on MIN-23-T.Newman, **Illegal Contact**, 5 yards, enforced at MIN 39 - No Play.

1-10-MIN 34 (8:23) 19-T.Ginn right end pushed ob at MIN 30 for 4 yards (22-H.Smith). **PENALTY** on CAR-13-K.Benjamin, **Offensive Holding**, 10 yards, enforced at MIN 30.

1-16-MIN 40 (8:00) 34-C.Artis-Payne right guard to MIN 37 for 3 yards (98-L.Joseph; 55-A.Barr).

2-13-MIN 37 (7:20) (No Huddle, Shotgun) 1-C.Newton scrambles right end pushed ob at MIN 36 for 1 yard (55-A.Barr).

3-12-MIN 36 (6:43) (No Huddle, Shotgun) 1-C.Newton pass deep left to 10-C.Brown to MIN 9 for 27 yards (24-C.Munnerlyn).

**WATCH HIGHLIGHT**

1-9-MIN 9 (5:58) (Shotgun) 43-F.Whittaker right end pushed ob at MIN 3 for 6 yards (23-T.Newman).

2-3-MIN 3 (5:25) (Shotgun) 1-C.Newton right end for 3 yards, **TOUCHDOWN**.

**WATCH HIGHLIGHT**

9-G.Gano extra point is **GOOD**, Center-44-J.Jansen, Holder-8-A.Lee. **PENALTY** on MIN-93-S.Stephen, **Leverage**, 15 yards, enforced between downs. **Penalty** on MIN-92-Tom.Johnson, **Unsportsmanlike Conduct**, declined.

MIN 0 CAR 10 Plays: 7 Possession: 3:51

From NFL.com (<http://www.nfl.com/gamecenter/2016092501/2016/REG3/vikings@panthers>)



# Football is discrete

— — —

- Game is a series of plays
- Start of play defined by
  - Score delta
  - Time remaining
  - Line of scrimmage
  - Down & distance



# Football is discrete



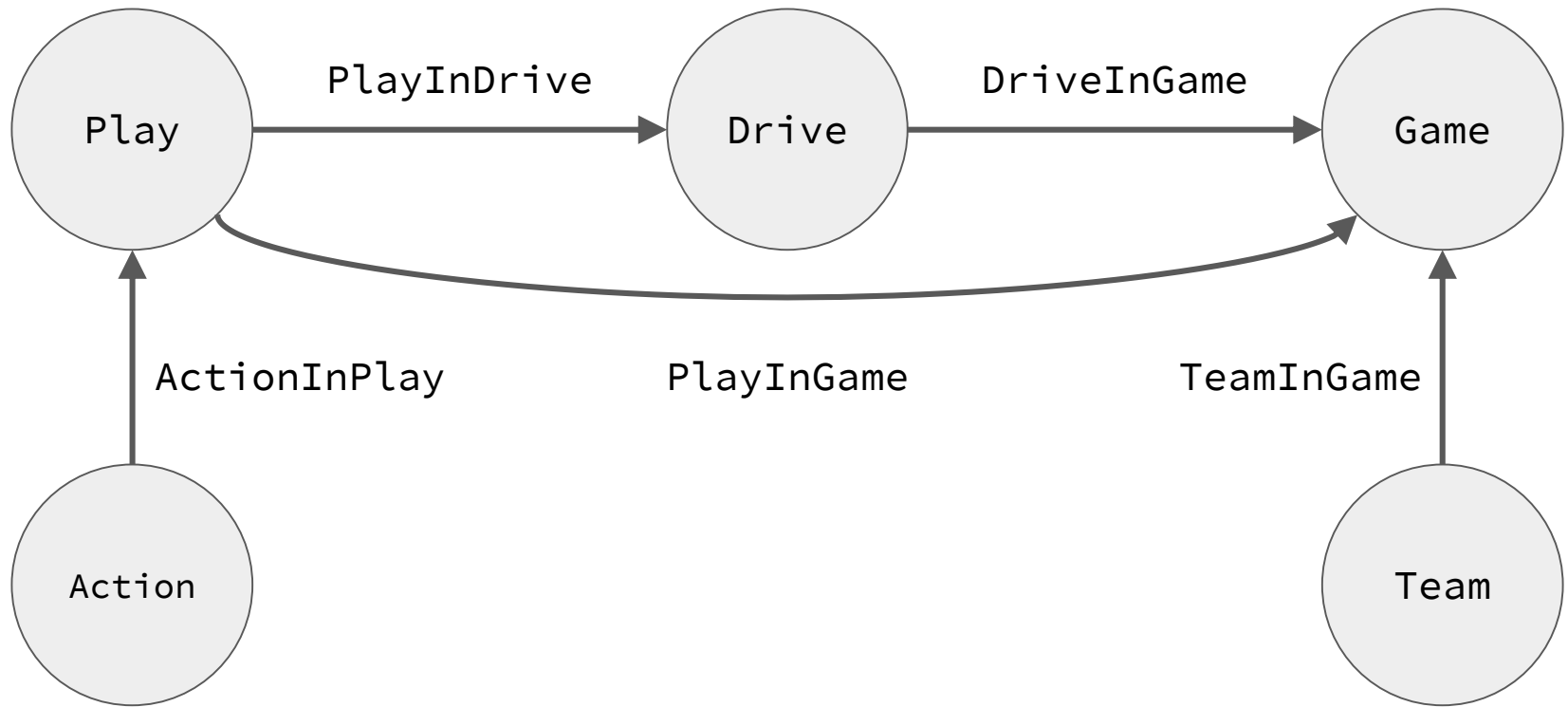
ECL Record →

## RECORD

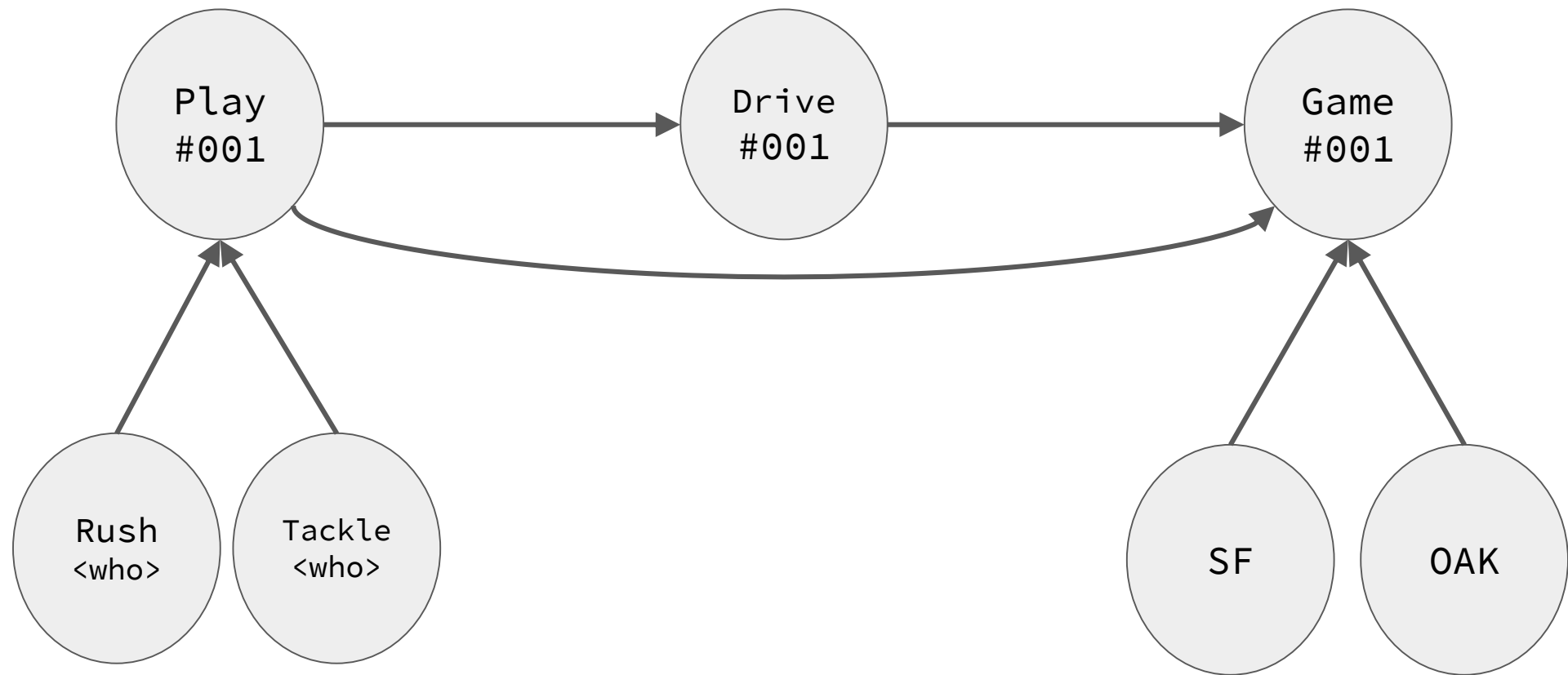
**UNSIGNED4** GameID;  
**UNSIGNED1** DriveNum;  
**UNSIGNED2** Sequence;  
**STRING** PlayType;  
**STRING1** Possession;  
**STRING1** Victor;  
**UNSIGNED2** Time;  
**UNSIGNED1** Quarter;  
**UNSIGNED1** Down;  
**UNSIGNED1** Yardline;  
**UNSIGNED1** Distance;  
**BOOLEAN** ScoringPlay;  
**UNSIGNED1** AwayScore;  
**UNSIGNED1** HomeScore;  
**UNSIGNED1** AwayTimeouts;  
**UNSIGNED1** HomeTimeouts;  
**STRING** Description;  
**END;**

2016092501|3|15|PASS|H|A|349|1|1|57|10|0|0|3|3|3|(9:11) (Shotgun) C.Newton pass sh  
2016092501|3|16|NoPlay|H|A|390|1|1|39|10|0|0|3|3|3|(8:30) (No Huddle, Shotgun) C.N  
2016092501|3|17|RUSH|H|A|397|1|1|34|10|0|0|3|3|3|(8:23) T.Ginn right end pushed ob  
2016092501|3|18|RUSH|H|A|420|1|1|40|16|0|0|3|3|3|(8:00) C.Artis-Payne right guard  
2016092501|3|19|RUSH|H|A|460|1|2|37|13|0|0|3|3|3|(7:20) (No Huddle, Shotgun) C.New  
2016092501|3|20|PASS|H|A|497|1|3|36|12|0|0|3|3|3|(6:43) (No Huddle, Shotgun) C.New  
2016092501|3|21|RUSH|H|A|542|1|1|9|9|0|0|3|3|3|(5:58) (Shotgun) F.Whittaker right  
2016092501|3|22|RUSH|H|A|575|1|2|3|3|1|0|3|3|3|(5:25) (Shotgun) C.Newton right end  
2016092501|3|23|EXTRA POINT|H|A|580|1|0|15|0|1|0|9|3|3|G.Gano extra point is GOOD,

# Model



# Example: Rushing play



# model.kel – entities

— — —

**Game** := ENTITY(FLAT(UID=**gameid**, INTEGER **date**, INTEGER **season**, INTEGER **week**, STRING **away**, STRING **home**, INTEGER **awayscore**, INTEGER **homescore**, STRING **victor**), MODEL(\*));

**Team** := ENTITY(FLAT( UID=**teamid**, STRING **short**, STRING **city**, STRING **nickname**, STRING **conference**, STRING **division**), MODEL(\*));

**Play** := ENTITY(FLAT( UID=**playid**, INTEGER **sequence**, STRING **playtype**, STRING **possession**, STRING **victor**, INTEGER **time**, INTEGER **quarter**, INTEGER **down**, INTEGER **yardline**, BOOLEAN **scoringplay**, INTEGER **awayScore**, INTEGER **homeScore**, INTEGER **awayTimeouts**, INTEGER **homeTimeouts**, STRING **description**), MODEL(\*));

**Action** := ENTITY(FLAT(UID=**actionid**, INTEGER **sequence**, INTEGER **statid**, STRING **playerid**, INTEGER yards, STRING team), MODEL(\*));

**Drive** := ENTITY(FLAT(UID=**driveid**, INTEGER **sequence**, STRING **possession**, INTEGER **starttime**, INTEGER **endtime**, INTEGER **startyard**, INTEGER **endyard**, STRING **result**, BOOLEAN **redzone**, INTEGER **yardspenalized**, INTEGER **yardsgained**, REAL **vf**, REAL **vfdelta**, INTEGER **ScoreDelta**), MODEL(\*));

# model.kel – associations and FLAT files

— — —  
**/\* \*\*\*\*\* ASSOCIATIONS \*\*\*\*\* \*/**

**PlayInGame** := ASSOCIATION(FLAT( Play  
play=playid, Game game=gameid));

**ActionInPlay** := ASSOCIATION(FLAT( Action  
action=actionid, Play play=playid));

**TeamInGame** := ASSOCIATION(FLAT( Team  
team=teamid, Game game=gameid, BOOLEAN  
ishome));

**DriveInGame** := ASSOCIATION(FLAT( Drive  
drive=driveid, Game game=gameid));

**PlayInDrive** := ASSOCIATION(FLAT( Play  
play=playid, Drive drive=driveid));

**/\* \*\*\*\*\* USE \*\*\*\*\* \*/**

USE ECL\_Model.Game.Dataset\_All(FLAT,  
**Game**);

USE ECL\_Model.Play.Dataset\_All(FLAT,  
**Play, PlayInGame, PlayInDrive**);

USE ECL\_Model.Action.Dataset\_All(FLAT,  
**Action, ActionInPlay**);

USE ECL\_Model.Team.Dataset\_All(FLAT,  
**Team**);

USE ECL\_Model.Team.Dataset\_TIG(FLAT,  
**TeamInGame**);

USE ECL\_Model.Drive.Dataset\_All(  
FLAT, **Drive, DriveInGame**);

# Victory Forecast

Probability of victory

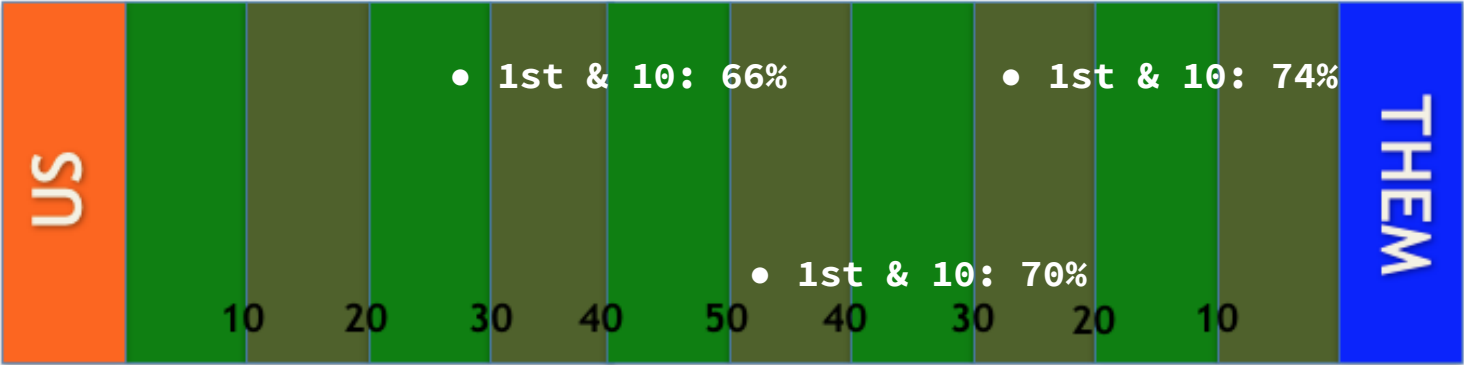
- Data from 1000+ games
- Each play
- Game state: (score\_diff, time, scrimmage, down, distance)
- Append outcome of game (W, L, or T)
- Using R, compute logistic regression to create the Victory Forecast (VF)

---

$$VF = f(s, t, S, d, D)$$

# VF Example: Line of Scrimmage

---

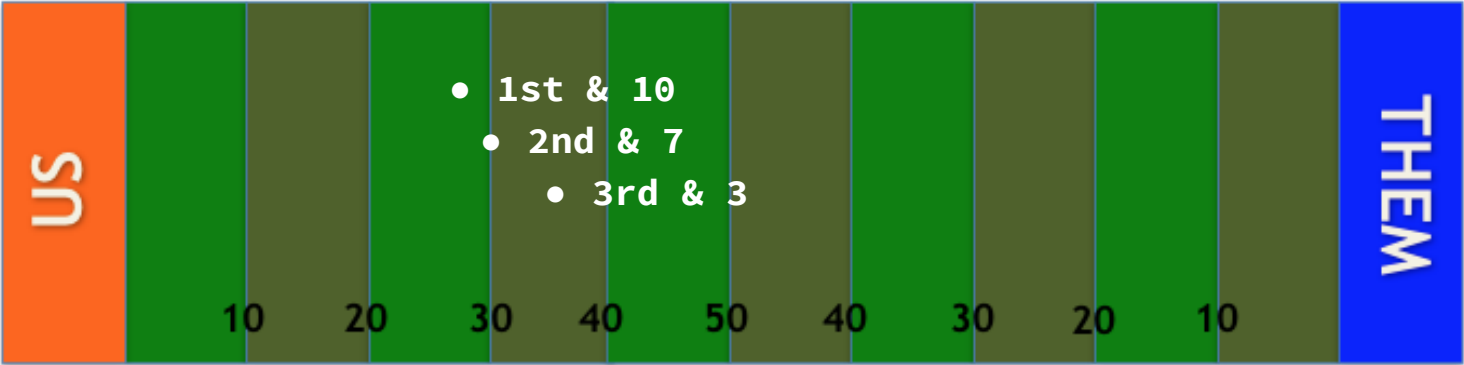


Middle 3rd Q, ahead by 3  
points



# VF Example: Neutral

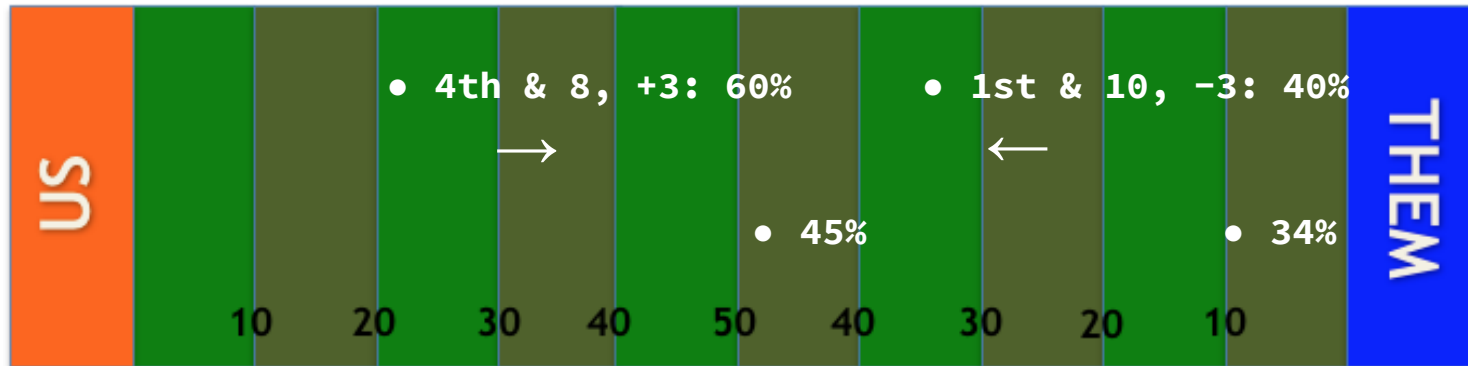
— — —



1st Quarter, tied: 52%

# VF Example: Punt

— — —



Start of 4th Quarter

# Transform

Upload and format for KEL

- Upload
- Spray
- Transform
- Persist as FLAT files

---

# Create FLAT file from CSVs

— — —

- Create additional information
  - Play: Add VF and VFdelta
  - Drive: Add VF, VFdelta, and ScoreDelta
- Set up for KEL
  - Add IDs
  - Use same files for entities and associations

# Fetch

Read data from web

- Multiple sources
- Compare and correct
- Parse into CSV

---

# Fetch and parse

— — —

- Multiple sources
- Python programs
- Fetch: HTTP crawl
- Parse
  - Walk DOM
  - Emit information as CSV
- Files
  - Games: date, home, winner, ...
  - Plays: time, down, type, ...
  - Drives: start/end, outcome, ...
  - Actions: rush, catch, tackle, ...
- Data
  - Games (~16 per week; 256 or 271/yr)
  - Plays (168 per game; 46,000/year)
  - Drives (24 per game; 6400/year)
  - Actions (4 per play; 188,000/year)



## Logical Files

Refresh
 Open
 Delete
 Remote Copy
 Copy
 Rename
 Add To Superfile
 Despray
 Filter

<input type="checkbox"/>				Logical Name	Owner	Description	Cluster	Records	Size	Parts	Modified (UTC/GMT)
<input type="checkbox"/>				└─ nfl					0		
<input type="checkbox"/>				└─ csvs					0		
<input type="checkbox"/>				└─ 2013					0		
<input type="checkbox"/>				└─ 2014					0		
<input type="checkbox"/>				└─ 2015					0		
<input type="checkbox"/>				└─ 2016					0		
<input type="checkbox"/>				└─ 1					0		
<input type="checkbox"/>				└─ drive					0		
<input type="checkbox"/>				└─ pa					0		
<input type="checkbox"/>				└─ pbb					0		
<input type="checkbox"/>				<a href="#">2016090800.csv</a>	sprayer		mythor		21,637	1	2016-09-09 20:23:02
<input type="checkbox"/>				<a href="#">2016091100.csv</a>	sprayer		mythor		22,042	1	2016-09-12 23:09:35
<input type="checkbox"/>				<a href="#">2016091101.csv</a>	sprayer		mythor		20,065	1	2016-09-12 23:09:15
<input type="checkbox"/>				<a href="#">2016091102.csv</a>	sprayer		mythor		22,432	1	2016-09-12 23:11:15
<input type="checkbox"/>				<a href="#">2016091103.csv</a>	sprayer		mythor		23,304	1	2016-09-12 23:10:35
<input type="checkbox"/>				<a href="#">2016091104.csv</a>	sprayer		mythor		25,626	1	2016-09-12 23:10:45
<input type="checkbox"/>				<a href="#">2016091105.csv</a>	sprayer		mythor		26,345	1	2016-09-12 23:11:05
<input type="checkbox"/>				<a href="#">2016091106.csv</a>	sprayer		mythor		22,530	1	2016-09-12 23:09:25
<input type="checkbox"/>				<a href="#">2016091107.csv</a>	sprayer		mythor		21,801	1	2016-09-12 23:10:15
<input type="checkbox"/>				<a href="#">2016091108.csv</a>	sprayer		mythor		21,933	1	2016-09-12 23:10:55
<input type="checkbox"/>				<a href="#">2016091109.csv</a>	sprayer		mythor		22,829	1	2016-09-12 23:11:25
<input type="checkbox"/>				<a href="#">2016091110.csv</a>	sprayer		mythor		24,360	1	2016-09-12 23:10:25
<input type="checkbox"/>				<a href="#">2016091111.csv</a>	sprayer		mythor		24,942	1	2016-09-12 23:09:45
<input type="checkbox"/>				<a href="#">2016091112.csv</a>	sprayer		mythor		22,586	1	2016-09-12 23:09:55
<input type="checkbox"/>				<a href="#">2016091200.csv</a>	sprayer		mythor		22,492	1	2016-09-19 05:24:06
<input type="checkbox"/>				<a href="#">2016091201.csv</a>	sprayer		mythor		24,965	1	2016-09-16 16:38:19
<input type="checkbox"/>				<a href="#">games.csv</a>	sprayer		mythor		717	1	2016-09-13 00:03:06

# Analysis



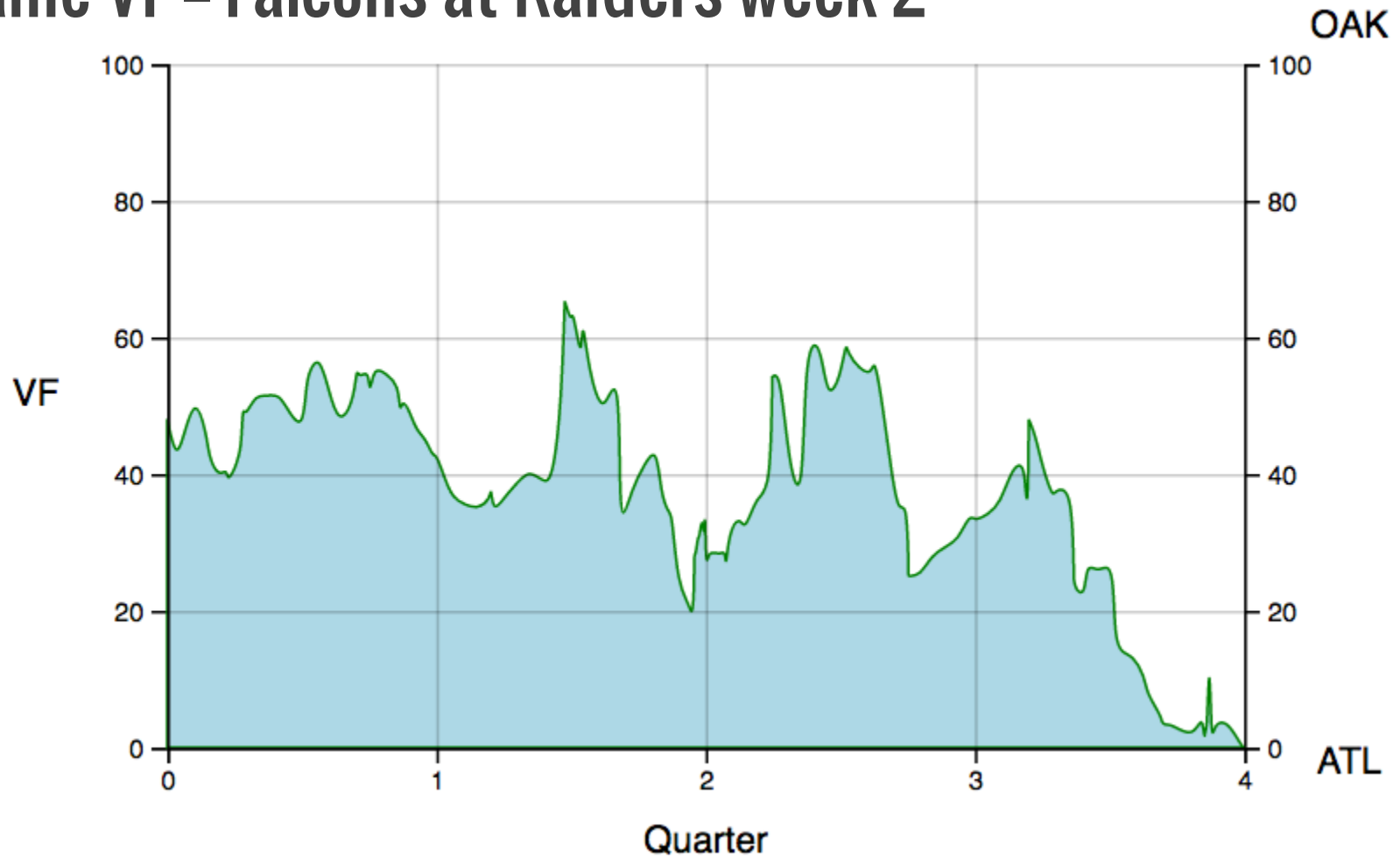


# Top plays – week 1

— — —

- 1.-73% NE v ARI: Catanzaro misses 47-yd FG.
- 2.-61% CAR v DEN: Gano misses 50-yd FG.
- 3.-53% GB v JAX: Jax fails to convert at 4&1 on the 14.
- 4.+44% DET v IND: Stafford passes to Jones for 22 yards.
- 5.+40% OAK v NO: the 2-pt conversion.
- 6.-36% MIN v TEN: Kendricks' pick-six at end of 3d.
- 7.+32% SD v KC: Smith's TD run in OT.
- 8.+27% DET v IND: Luck to Dorsett for 33 yds to DET 26.
- 9.+24% DET v IND: Prater's 43-yd FG.
- 10.+22% CIN v NYJ: Dalton to Green 54-yd TD.

# Game VF – Falcons at Raiders week 2



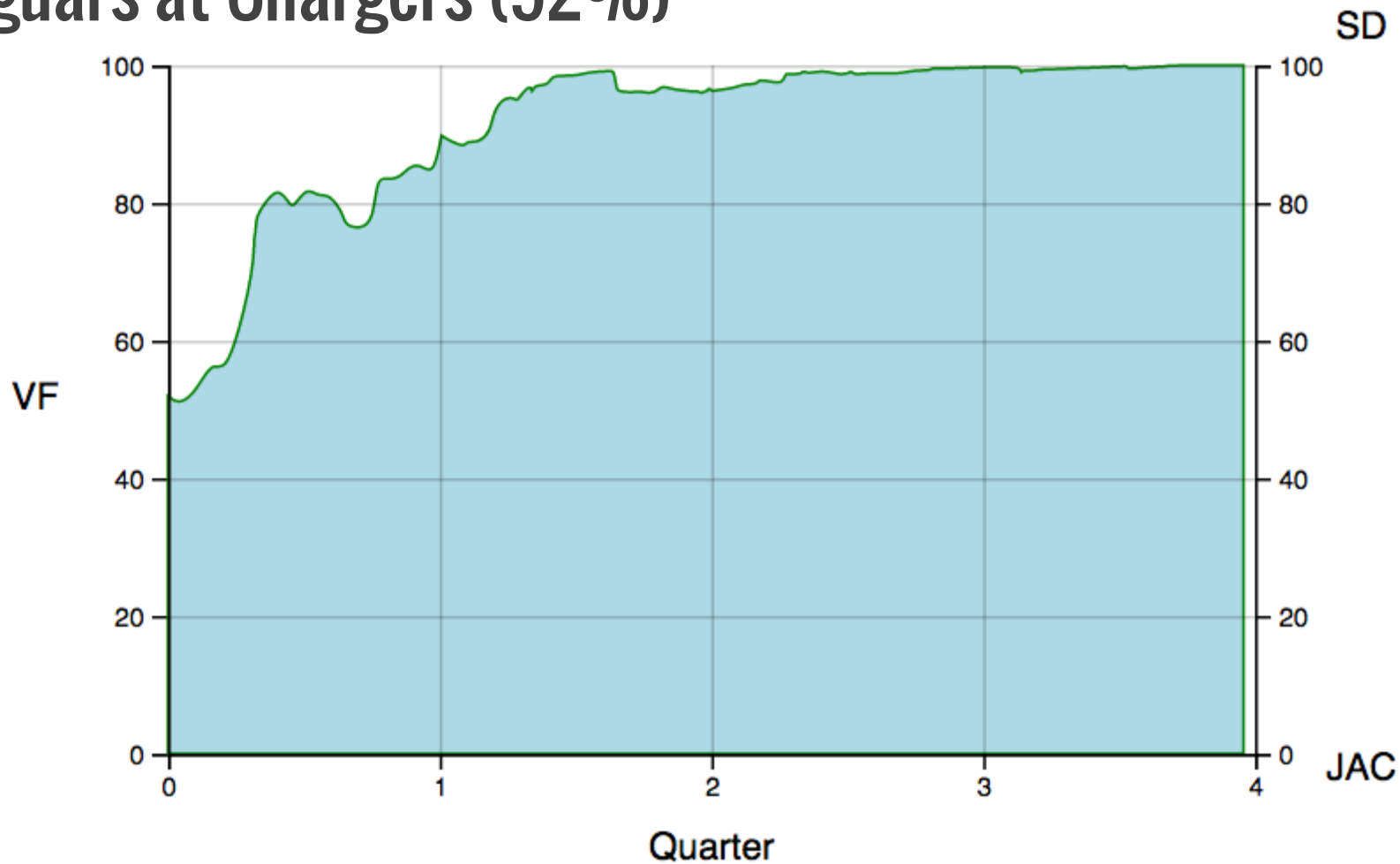
# Strength of Victory

- Trapezoidal integration of VF
- Magnitude of victory



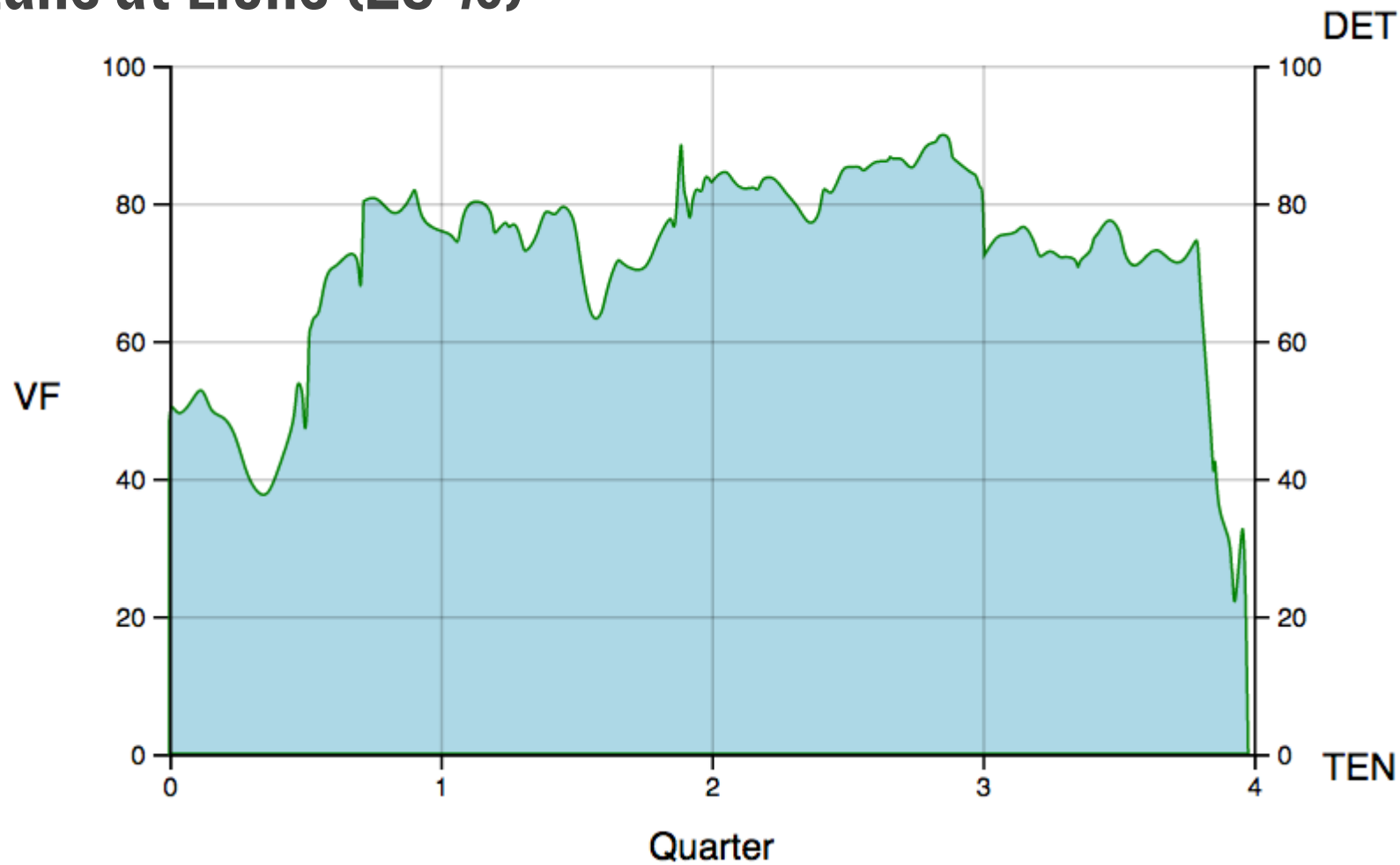
Game	Strength of victory
JAC@SD	92%
MIA@NE	91%
TB@ARI	83%
KC@HOU	77%
PIT@CIN	73%
SF@CAR	72%
IND@DEN	66%
NYJ@BUF	65%
PHI@CHI	65%
SEA@LA	63%
ATL@OAK	62%
NO@NYG	58%
DAL@WAS	58%
GB@MIN	54%
CLE@BAL	34%
TEN@DET	28%

# Jaguars at Chargers (92%)<sup>VF Graph</sup>



# Titans at Lions (28%)

VF Graph



# Competitiveness

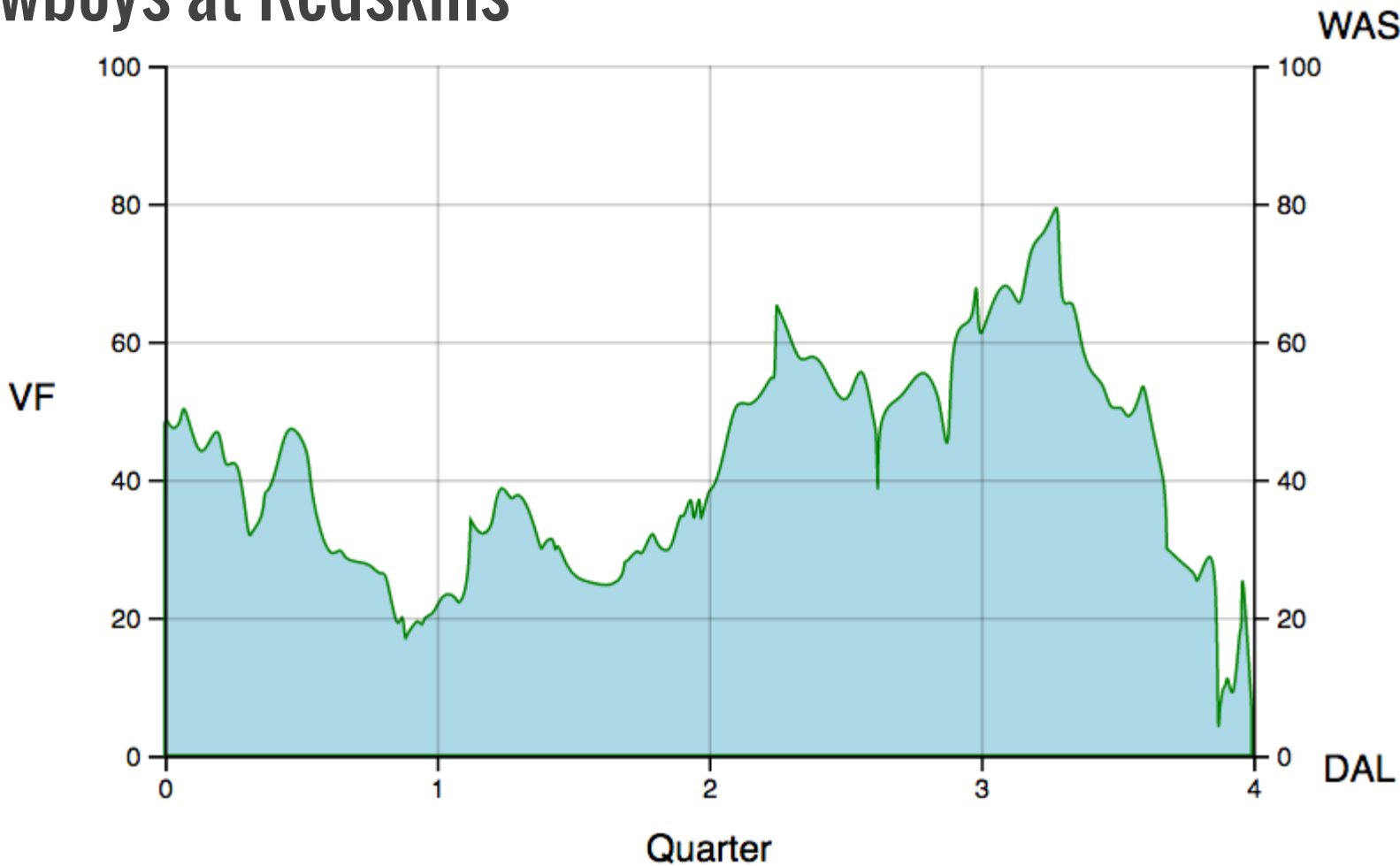
Sum absolute value  
of all VF gains.



Game	Strength of victory	Competi- tiveness
JAC@SD	92%	0.99
MIA@NE	91%	1.75
TB@ARI	83%	2.27
KC@HOU	77%	3.69
PIT@CIN	73%	3.90
SF@CAR	72%	4.39
IND@DEN	66%	4.89
NYJ@BUF	65%	4.98
PHI@CHI	65%	3.80
SEA@LA	63%	5.16
ATL@OAK	62%	5.71
NO@NYG	58%	5.61
DAL@WAS	58%	<b>5.98</b>
GB@MIN	54%	5.81
CLE@BAL	34%	4.19
TEN@DET	28%	4.68

# Cowboys at Redskins

VF Graph



# Website

See at the  
HPCC Systems booth

- View games, VF, and other stats.
- Front-end interface to Roxie queries on our HPCC cluster.

---



# Conclusion

Thanks