MSIS 537

Major League Baseball Advanced Media

Statcast Database

by Kyle Capuani

Title Page Logos Credit to MLB: https://d1.awsstatic.com/logos/customers/MLBAM-logo.1ffc8319f402f547b3da8a5cddfc88bf8a836294.jpg
AWS:https://upload.wikimedia.org/wikipedia/commons/thumb/1/1d/AmazonWebservices_Logo.svg/2000px-AmazonWebservices_Logo.svg.png

Table of Contents

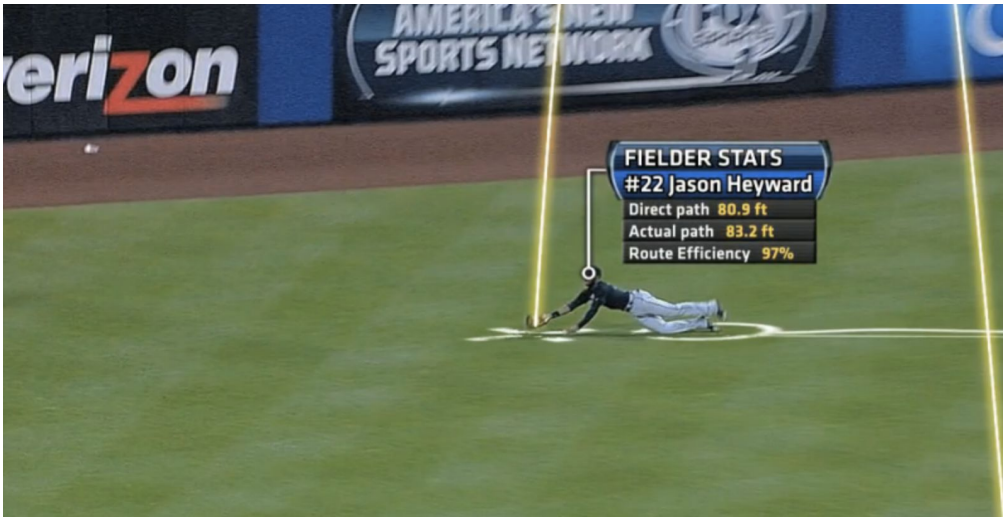
Executive Summary

Overview

While data analytics have revolutionized the sport of baseball, the analytics behind defense (more specifically fielding) have always been a mystery. Many baseball professionals could see that a player was defensively gifted but could not explain how, why or put a value on an individual's talents. In 2014, Major League Baseball Advanced Media (MLBAM), the internet and interactive branch of Major League Baseball (MLB)[1], implemented a player tracking system called *Statcast* that can, "generate three-dimensional snapshots of every movement on a baseball field, some 40,000 frames per second converted into digital data."[2] This data could then be converted into information about how well a player reads and reacts to any given play. MLBAM Statcast was instantly proclaimed as the Holy Grail of Baseball Analytics.

Statcast relies on, "two data-acquisition systems at … stadiums that provide coordinate information. A Doppler radar system sits behind home plate, sampling the ball position 2,000 times a second. Two stereoscopic imaging devices, usually positioned above the third-base line, sample the positions of players on the field 30 times a second."[3] Instead of using expensive on-premise information systems, though, MLBAM relies on Amazon Web Services (AWS) to store and distribute their data to teams around the league. "The AWS cloud offered an ideal alternative that could support as many as 15 games on a single day—(then) some days with just one or two… and then be shut down during the off season."[3]

[1] "MLB Advanced Media" Wikipedia https://en.wikipedia.org/wiki/MLB_Advanced_Media
[2]Schoenfeld, Bruce. "Can New Technology Bring Baseball's Data Revolution to Fielding?" New York Times Magazine. September 30th, 2016. https://www.nytimes.com/2016/10/02/magazine/can-new-technology-bring-baseballs-data-revolution-to-fielding.html?_r=1
[3] "MLB Advanced Media Case Study" Amazon Web Services https://aws.amazon.com/solutions/case-studies/major-league-baseball-mlbam/



Statcast tracks the path of the baseball, the movement of the players and how efficiently they executed the play.[4]

Statcast records data on every movement that occurs on the baseball field during a single play. For example, in the event of a sacrifice fly, the tracking system is able to record the spin rate of the pitcher's curveball, the exit velocity of the batted ball, the time the outfielder took to react, his route efficiency in reaching the hit ball and even the velocity of that fielder's throw to home plate. Over time, data collection from Statcast allows all major league teams the ability to better evaluate a player's talent as well as strategize for upcoming opponents.

With greater data collection, Statcast can better evaluate the hidden talents of MLB players[5]

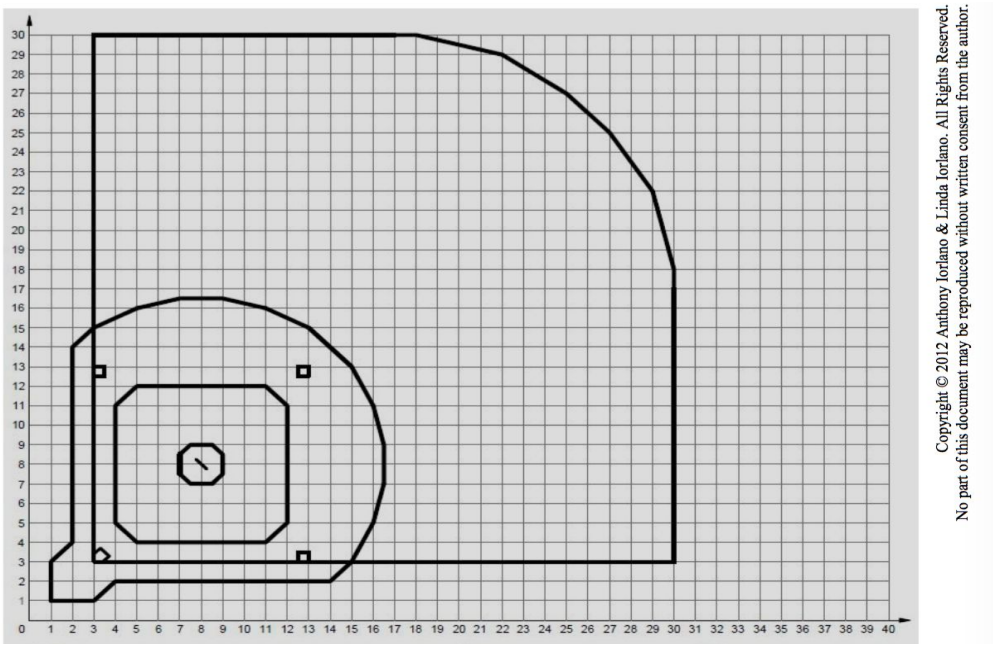
[4] https://deadspin.com/mlb-announces-revolutionary-new-fielding-tracking-syste-1534200504
[5] https://www.sporttechie.com/wp-content/uploads/2016/04/mlbf_607328483_th_45.jpg

Objectives

The purpose of this database is to store the immense amount of data being recorded by MLBAM Statcast. The database can be used to evaluate Major League Baseball players and how efficiently they use the field. By interpreting the baseball field as a giant (X,Y) grid, data analysts can better assess a fielder's range. This information can then be deciphered to understand a player's overall fielding ability.

This report will explain how the database:

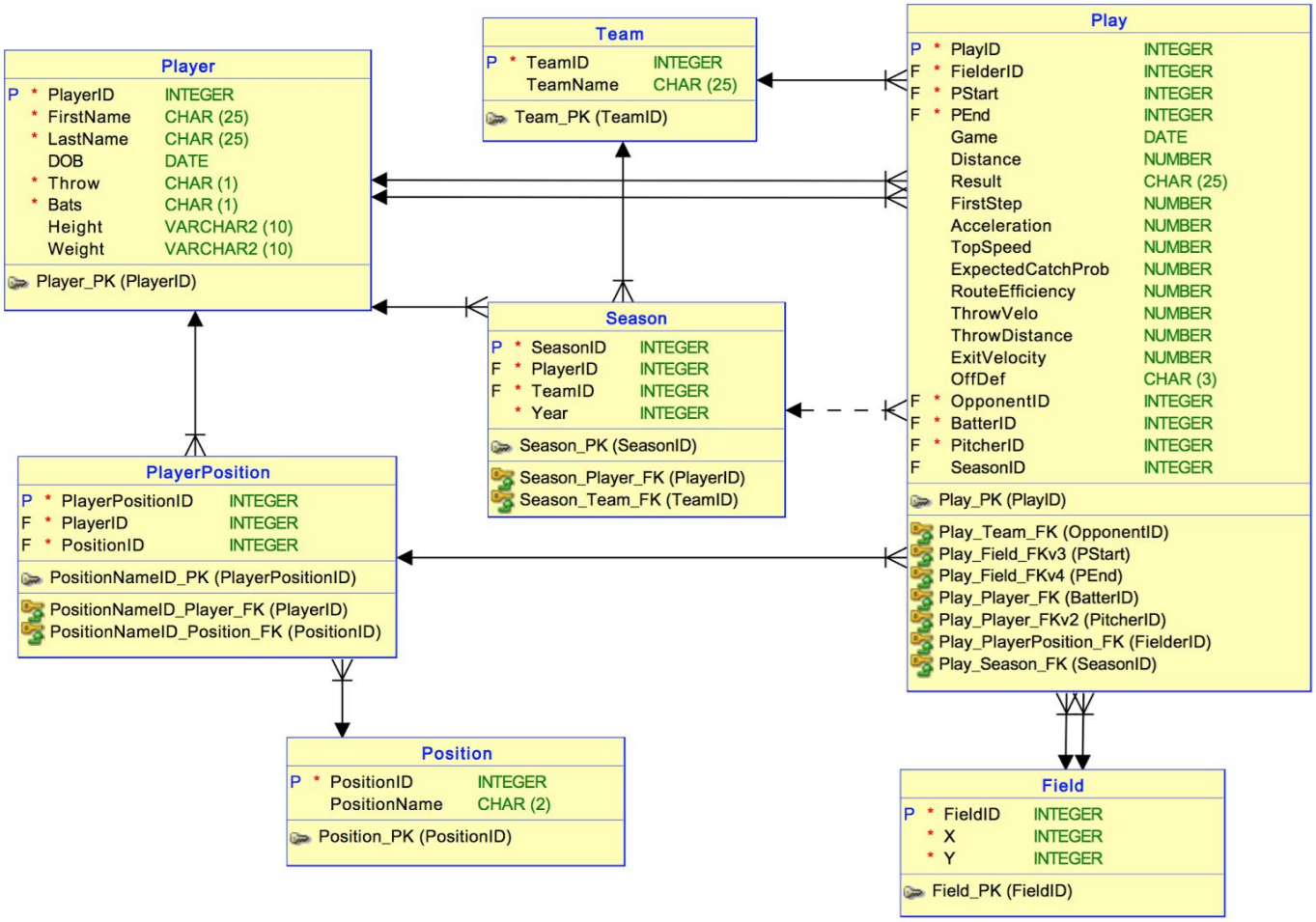
- Records individual plays
- Tracks a Player's movements on the field
- Evaluates players' route efficiency and range compared to difficulty of catch
- Gauges a pitcher's effect on balls batted into play
- How each table is able to produce valuable information



Turning the baseball field into an (X,Y) grid allows plays stored in the database some context.[6]

[6] I purchased the rights to use said grid. Credit to Anthony and Linda Iorlano of TeachersPayTeachers

Entity Relationship Diagram (ERD)



Tables

Player

Purpose

The purpose of this table is to house defining data on individual MLB players such as the person's first and last name, date of birth, height and weight. The table also stores data regarding which way a player bats and throws (R signifying Right, L signifying Left.) The primary key for this table is *PlayerID* and is used to tie the Player table to numerous other tables in the database.

Create Statement

```
CREATE TABLE player
(
   playerid                INTEGER NOT NULL,
   firstname               CHAR(25) NOT NULL,
   lastname                CHAR(25) NOT NULL,
   dob                     DATE,
   throw                   CHAR(1) NOT NULL,
   bats                    CHAR(1) NOT NULL,
   height                  VARCHAR2(10),
   weight                  VARCHAR2(10)
);

ALTER TABLE player ADD CONSTRAINT player_pk PRIMARY KEY ( playerid );
```

Functional Dependencies

PlayerID → FirstName, LastName, DOB, Throw, Bats, Height, Weight

Sample Data

	PLAYERID	FIRSTNAME	LASTNAME	DOB	THROW	BATS	HEIGHT	WEIGHT
1	1	Bondie	Wahlberg	23-JAN-86	R	R	6 ft	200 lbs
2	2	Donnie	Miller	02-MAR-89	L	L	6 ft 2 in	240 lbs
3	3	Kenneth	Greene	10-JUN-82	R	L	6 ft 3 in	235 lbs
4	4	Clyde	Dog	22-JUN-92	R	R	5 ft 11 in	165 lbs
5	5	Derrick	Lynn	11-AUG-96	R	L	6 ft	185 lbs
6	6	Kris	Travis	13-DEC-82	L	L	6 ft	190 lbs
7	7	David	Roberts	09-MAY-91	R	R	6 ft 4 in	220 lbs
8	8	Michael	Young	18-APR-94	L	L	6 ft 6 in	240 lbs
9	9	Chris	Teal	18-JUN-91	R	L	6 ft 3 in	195 lbs
10	10	Jacob	Harvey	13-JUL-93	R	L	6 ft 1 in	215 lbs
11	11	Matt	Roberto	15-SEP-89	L	R	6 ft 2 in	200 lbs
12	12	Juan	Carlos	07-DEC-87	R	R	6 ft 1 in	215 lbs

Position

Purpose

The purpose of the Position table is to house the distinct positions found on the baseball field. The table contains the Primary Key *PositionID* and column PositionName. The PositionID is mapped to the same number conventionally associated with the position name by standard baseball scoring.

Create Statement

```
CREATE TABLE position
(
   positionid                INTEGER NOT NULL,
   positionname              CHAR(2)
);

ALTER TABLE position ADD CONSTRAINT position_pk PRIMARY KEY ( positionid );
```

Functional Dependencies

PositionID → PositionName

Sample Data

	POSITIONID	POSITIONNAME
1	1	P
2	2	C
3	3	1B
4	4	2B
5	5	3B
6	6	SS
7	7	LF
8	8	CF
9	9	RF
10	10	DH

PlayerPosition

Purpose

The PlayerPosition table allows the database to account for a player who can play multiple positions on the field. Its Primary Key *PlayerPositionID* is later used in the Play table to tie the data back to the Player table and the Position table. This allows the individual play to have the context of who and where. The PlayerPosition table contains the foreign keys of PlayerID and PositionID.

Create Statement

```
CREATE TABLE playerposition
(
   playerpositionid          INTEGER NOT NULL,
   playerid                  INTEGER NOT NULL,
   positionid                INTEGER NOT NULL
);

ALTER TABLE playerposition ADD CONSTRAINT positionnameid_pk PRIMARY KEY (
playerpositionid );
```

Functional Dependencies

PlayPositionID → PlayerID, PositionID

Sample Data

	PLAYERPOSITIONID	PLAYERID	POSITIONID
1	1	1	6
2	2	1	4
3	3	1	5
4	4	2	3
5	5	3	2
6	6	4	4
7	7	4	5
8	8	5	8
9	9	5	7
10	10	5	9
11	11	6	9
12	12	6	7
13	13	7	5
14	14	8	1
15	15	9	1

Team

Purpose

The Team table's purpose is to house all Major League Baseball organizations. It will later be used in the Season table in connecting an individual player to a specific team. The Primary Key is *TeamID* and the table also contains the column, TeamName.

Create Statement

```
CREATE TABLE team (
   teamid                  INTEGER NOT NULL,
   teamname                CHAR(25)
);

ALTER TABLE team ADD CONSTRAINT team_pk PRIMARY KEY ( teamid );
```

Functional Dependencies

TeamID → TeamName

Sample Data

	TEAMID	TEAMNAME
1	1	Baltimore Orioles
2	2	Boston Red Sox
3	3	Chicago White Sox
4	4	Cleveland Indians
5	5	Detriot Tigers
6	6	Houston Astros
7	7	Kansas City Royals
8	8	Los Angeles Angels
9	9	Minnesota Twins
10	10	New York Yankees
11	11	Oakland Athletics
12	12	Seattle Mariners
13	13	Tampa Bay Rays
14	14	Texas Rangers

Season

Purpose

The purpose of the Season table is to connect the Player table and the Team table. The Year is also provided. This allows the database to anticipate for players changing teams as well refer to when the player was a member of such team. Its Primary Key is *SeasonID* and the table also contain two Foreign Keys in PlayerID and TeamID. As mentioned above, the field Year is also present.

Create Statement

```
CREATE TABLE season
(
   seasonid                INTEGER NOT NULL,
   playerid                INTEGER NOT NULL,
   teamid                  INTEGER NOT NULL,
   year                    INTEGER NOT NULL
);

ALTER TABLE season ADD CONSTRAINT season_pk PRIMARY KEY ( seasonid );
```

Functional Dependencies

SeasonID → PlayerID, TeamID, Year

Sample Data

	SEASONID	PLAYERID	TEAMID	YEAR
1	1	1	24	2017
2	2	2	24	2017
3	3	3	10	2017
4	4	4	10	2017
5	5	4	7	2017
6	6	5	24	2017
7	7	6	10	2017
8	8	7	10	2017
9	9	8	10	2017
10	10	9	10	2017
11	11	10	24	2017

Field

Purpose

The Field table houses all (X,Y) coordinate possibilities on the baseball field. For example, the FieldID 431 is connected to coordinates (15,7) on the baseball grid. This table helps to provide context on where the play occurs. It is tied to the Play table twice to signify where the player is when the play begins and where the player ends up by time the play is made. The Field table's Primary Key is *FieldID* and its following columns are X and Y.

Create Statement

```
CREATE TABLE field
(
   fieldid                    INTEGER NOT NULL,
   x                          INTEGER NOT NULL,
   y                          INTEGER NOT NULL
);

ALTER TABLE field ADD CONSTRAINT field_pk PRIMARY KEY ( fieldid );
```

Functional Dependencies

FieldID → X, Y

Sample Data

	FIELDID	X	Y
1	331	12	1
2	332	12	2
3	333	12	3
4	334	12	4
5	335	12	5
6	336	12	6
7	337	12	7
8	338	12	8
9	339	12	9
10	340	12	10

Play

Purpose

The Play table contains all the data fields recorded during a given play in the game. Essentially all the data that Statcast records is stored here. All of the other tables simply give this table context. Its Primary Key is *PlayID* which signifies the distinct play. Its Foreign Keys are FielderID, which is an alias for PlayerPositionID and is tied to the PlayerPosition table, PStart (play starting point) and PEnd (play ending point), which are two separate sets of data and aliases for FieldID, OpponentID which is another alias for TeamID, BatterID and PitcherID, both aliases for PlayerID. A final Foreign Key is SeasonID which can tie a play to a specific season/team. Other interesting fields in the Play table are Game, which is used to 'date' a play, Result, which allows for leeway context on if the play was recorded as an out, an error, or a safe play, among others. Finally, the OffDef column defines whether the play was offensively or defensively driven (meaning that if the play spawned from a batted ball or if the specific play originated from a runner tagging up, forcing a throw. These are considered separate plays and can be defined as such in the table). ExpectedCatchProb(ability) is a reference on difficulty of play based on how often it is made by other players in the league.

Create Statement

```
CREATE TABLE play (
    playid              INTEGER NOT NULL,
    fielderid           INTEGER NOT NULL,
    pstart              INTEGER NOT NULL,
    pend                INTEGER NOT NULL,
    game                DATE,
    distance            NUMBER,
    result              CHAR(25),
    firststep           NUMBER,
    acceleration        NUMBER,
    topspeed            NUMBER,
    expectedcatchprob   NUMBER,
    routeefficiency     NUMBER,
    throwvelo           NUMBER,
    throwdistance       NUMBER,
    exitvelocity        NUMBER,
    offdef              CHAR(3),
    opponentid          INTEGER NOT NULL,
    batterid            INTEGER NOT NULL,
    pitcherid           INTEGER NOT NULL,
    seasonid            INTEGER
);

ALTER TABLE play ADD CONSTRAINT play_pk PRIMARY KEY ( playid );
```

Functional Dependencies

PlayID → Game, Distance, Result, FirstStep, Acceleration, TopSpeed, ExpectedCatchProb, Route Efficiency, ThrowVelo, ThrowDistance, ExitVelocity, OffDef

Measurements

Distance → *Feet* FirstStep → *Seconds* Acceleration → *Feet/Second (F/S)*

TopSpeed → *Miles per Hour (MPH)* ExpectedCatchProb → *Percentage (%)*

RouteEfficiency → *%* ThrowVelo(city) → *MPH* ThrowDistance → *Feet*

ExitVelocity → *MPH*

Sample Data

	PLAYID	FIELDERID	PSTART	PEND	GAME	DISTANCE	RESULT	...	FIRSTSTEP	ACCE...	TOPSPEED	EXPECTEDCATCHPROB	ROUTEEFFICIENCY	THROWVELO	THROWDISTANCE	EXITVELOCITY	OFFDEF	OPPONENTID	BATTERID	PITCHERID	SEASONID
1	25	18	398	431	26-JUN-17	16	Out	...	0.12	9.4	11.3	92.2	98.2	65.4	(null)	54.3	off	23	15	20	13
2	26	20	254	196	01-OCT-17	22	Safe	...	0.5	6.7	8.7	97.6	89.3	45.6	(null)	81.2	off	27	1	8	14
3	27	20	254	344	02-OCT-17	28	Safe	...	0.3	9.4	12.3	86.7	90.3	60.7	(null)	72.2	off	15	2	8	14
4	28	22	696	693	02-AUG-17	12	Out	...	0.3	4.4	4.6	94.5	98.9	(null)	(null)	67.8	off	1	1	9	15
5	29	22	696	704	23-AUG-17	9	Out	...	0.02	15	17.2	75.5	97.9	(null)	(null)	90.6	off	1	1	9	15
6	30	24	324	142	27-AUG-17	43	Out	...	0.04	16.2	18.1	62.3	99.2	(null)	(null)	75.2	off	15	15	8	15
7	31	24	324	63	22-JUL-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	101.7	235	67.8	def	12	17	9	15
8	32	22	696	63	24-JUL-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	104.7	223.7	80.6	def	12	15	20	15
9	33	22	696	63	27-JUN-17	(null)	Safe	...	(null)	(null)	(null)	(null)	(null)	101.7	242.3	89.2	def	18	7	20	15
10	34	25	63	373	27-JUN-17	(null)	Safe	...	(null)	(null)	(null)	(null)	(null)	82.2	(null)	(null)	def	20	1	20	17
11	35	25	63	96	27-AUG-17	13	Out	...	0.2	7.6	8	95.4	99.2	67.3	(null)	34.5	off	13	17	8	17
12	36	26	134	194	21-AUG-17	15	Out	...	0.02	10.4	8.4	85.6	98.3	68.4	(null)	67.8	off	1	17	9	18
13	37	26	134	221	01-AUG-17	19	Safe	...	0.06	8.2	11.5	91.3	88.4	54.4	(null)	30.4	off	16	12	8	18
14	38	27	395	398	02-MAY-17	15	Out	...	0.02	9.4	5.6	94.6	98.6	76.5	(null)	73.4	off	17	1	9	19
15	39	28	709	716	02-JUN-17	73	Out	...	0.02	15.6	19.3	43.5	98.6	(null)	(null)	92.1	off	11	17	11	20
16	40	28	709	588	05-JUN-17	57	Out	...	0.04	17.2	19.6	67.2	99.6	(null)	(null)	64.7	off	14	5	9	20
17	41	31	296	300	05-SEP-17	63	Out	...	0.07	14.3	15.3	82.3	94.6	(null)	(null)	96.4	off	15	1	8	21
18	42	31	296	141	08-SEP-17	53	Out	...	0.02	12.1	12.3	91.4	95.6	(null)	(null)	83.4	off	19	2	9	21
19	43	31	141	63	08-SEP-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	103.4	198	(null)	def	19	2	9	21
20	44	34	399	371	12-JUL-17	15	Out	...	0.05	13.4	11.4	92.2	98.4	83.3	(null)	76.6	off	20	1	9	23
21	45	34	399	371	21-JUL-17	15	Safe	...	0.7	10.4	9.4	92.2	89.2	69.9	(null)	50.1	off	23	19	11	23
22	46	36	395	399	21-JUN-17	12	Error	...	0.1	10.6	12.6	95.4	91.2	(null)	(null)	101.3	off	25	1	9	23
23	47	37	395	399	22-JUN-17	12	Out	...	0.03	12.5	9.5	95.4	99.2	86.5	(null)	91.4	off	27	1	11	24
24	48	38	103	194	22-SEP-17	10	Out	...	0.01	9.6	7.5	92.1	98.3	65.5	(null)	79.1	off	1	5	25	25
25	49	38	104	75	25-OCT-17	9	Error	...	0.5	3.4	5.4	98.7	82.3	(null)	(null)	67.1	off	4	17	20	25
26	50	40	399	399	25-MAY-17	1	Out	...	0.02	0.4	0.4	98.9	99.9	(null)	(null)	103.4	off	5	1	11	26
27	1	1	429	401	23-JUN-17	10	Out	...	0.02	6.3	12.2	97.1	95.9	74.5	(null)	99.3	off	12	14	8	1
28	2	1	369	186	10-SEP-17	27	Out	...	0.04	9.2	14.1	63.6	98.9	83.8	(null)	52.5	off	17	16	25	1
29	3	1	431	487	02-MAY-17	18	Safe	...	0.5	4.9	9.1	58.8	78.2	(null)	(null)	69.8	off	1	12	8	1
30	4	4	104	74	09-JUL-17	4	Out	...	0.02	4.1	4.9	92.3	100	(null)	(null)	101.2	off	25	17	11	2
31	5	4	104	194	21-AUG-17	9	Safe	...	1.1	3.1	5.6	93.4	89.2	(null)	(null)	96.6	off	22	17	8	2
32	6	4	133	102	21-JUL-17	7	Safe	...	0.92	3.4	6.2	88.4	88.6	(null)	(null)	98.4	off	21	1	9	2
33	7	5	63	124	27-MAY-17	79	Out	...	0.3	4.7	6.8	78.9	97.6	82.2	100	40.1	off	22	1	8	3
34	8	5	63	4	14-APR-17	12	Out	...	0.25	2.3	2.6	96.7	97.9	(null)	(null)	37.9	off	29	17	20	3
35	9	6	284	346	14-SEP-17	24	Out	...	0.06	9.6	15.4	42.3	99.2	90.5	80	90.2	off	27	17	11	4
36	10	7	396	393	22-SEP-17	8	Out	...	0.04	7.8	5.6	37.8	99.8	95.8	103	110.1	off	24	1	9	5
37	11	7	366	482	27-SEP-17	17	Out	...	0.12	12.3	14.3	33.4	96.7	(null)	(null)	56.7	off	3	14	8	4
38	12	8	712	716	04-JUL-17	79	Out	...	0.02	15.3	17.3	26.7	98.7	(null)	(null)	93.2	off	14	14	25	6
39	13	8	708	527	18-JUL-17	53	Out	...	0.04	12.4	15.4	45.7	98.9	(null)	(null)	76.7	off	17	23	20	6
40	14	8	708	373	20-JUL-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	94.3	145	(null)	def	18	21	8	6
41	15	8	716	373	09-APR-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	101.5	145	(null)	def	19	15	11	6
42	16	8	716	63	12-MAY-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	105.5	235	(null)	def	20	11	9	6
43	17	8	704	63	15-MAY-17	(null)	Out	...	(null)	(null)	(null)	(null)	(null)	103.5	247	(null)	def	18	3	20	6
44	18	11	176	63	17-AUG-17	(null)	Safe	...	(null)	(null)	(null)	(null)	(null)	88.7	203	(null)	def	19	3	20	7
45	19	11	176	178	17-JUL-17	17	Safe	...	1.2	8.6	12.3	94.4	83.2	(null)	(null)	101.3	off	13	1	9	7
46	20	13	396	399	11-JUL-17	5	Out		0.4	5.4	2.4	98.3	99.5	76.6	(null)	100.4	off	15	1	11	8

Views

Plays_View

Purpose

This View provides a simple look at all the offensively-driven plays with context as to who, where and when. It is sorted by GAME (Date). The view allows for quick access without the need to see all other tables.

Create Statement

```
CREATE OR REPLACE VIEW Plays_view as
Select
p.playid as Play,
p.game as Game,
b.firstname as FirstName,b.lastname as LastName,
po.positionname as Position,
s.fieldid as StartingFieldID,s.x as StartingX,s.y as StartingY,
initcap(s.x) || ',' || initcap(s.y) as StartingXY,
e.fieldid as EndingFieldID,e.x as EndingX,e.y as EndingY,
initcap(e.x) || ','|| initcap(e.y) as EndingXY,
p.distance as DistanceTraveled,p.firststep as ReactionSpeed,
p.acceleration as Acceleration,p.topspeed as TopSpeed,
p.routeefficiency as RouteEfficiency,
p.expectedcatchprob,p.exitvelocity as BattedBallExitVelocity,
p.result as PlayResult,
ba.lastname as Batter,pi.lastname as Pitcher
from play p

join playerposition pp
on p.fielderid=pp.playerpositionid
join player b
on pp.playerid=b.playerid
join position po
on pp.positionid=po.positionid
join field s
on p.pstart=s.fieldid
join field e
on p.pend=e.fieldid
join player ba
on p.batterid=ba.playerid
join player pi
on p.pitcherid=pi.playerid
where offdef='off';
```

Sample Data

	PL...	GAME	FIRSTNAME	LASTNAME	P...	STARTING...	STARTINGX	STARTINGY	STARTINGXY	ENDING...	ENDINGX	ENDINGY	ENDINGXY	DISTANCE...	REACTION...	ACCE...	TOPSPEED	ROUTEEFFI...	EXPECTEDCA...	BATTEDBALLEXIT...	PLAYRESULT	PITCHER	BATTER
1	8	14-APR-17	Kenneth	Greene	C	63	3	3	3,3	4	1	4	1,4	12	0.25	2.3	2.6	97.9	96.7	37.9	Out	Ramirez	Tyler
2	3	02-MAY-17	Bondie	Wahlberg	SS	431	15	11	15,11	487	17	7	17,7	18	0.5	4.9	9.1	78.2	58.8	69.8	Safe	Young	Carlos
3	38	02-MAY-17	Taylor	Tyler	3B	395	14	5	14,5	398	14	8	14,8	15	0.02	9.4	5.6	98.6	94.6	73.4	Out	Teal	Wahlberg
4	50	25-MAY-17	Todd	Matthews	SS	399	14	9	14,9	399	14	9	14,9	1	0.02	0.4	0.4	99.9	98.9	103.4	Out	Roberto	Wahlberg
5	7	27-MAY-17	Kenneth	Greene	C	63	3	3	3,3	124	5	4	5,4	79	0.3	4.7	6.8	97.6	78.9	40.1	Out	Young	Wahlberg
6	39	02-JUN-17	George	Michaels	CF	709	24	19	24,19	716	24	26	24,26	73	0.02	15.6	19.3	98.6	43.5	92.1	Out	Roberto	Tyler
7	40	05-JUN-17	George	Michaels	CF	709	24	19	24,19	588	20	18	20,18	57	0.04	17.2	19.6	99.6	67.2	64.7	Out	Teal	Lynn
8	23	13-JUN-17	Juan	Carlos	LF	422	15	2	15,2	434	15	14	15,14	15	0.12	6.7	8.9	92.2	88.9	94.4	Error	Young	Miller
9	46	21-JUN-17	Pablo	Sanchez	3B	395	14	5	14,5	399	14	9	14,9	12	0.1	10.6	12.6	91.2	95.4	101.3	Error	Teal	Wahlberg
10	47	22-JUN-17	Ricco	LaRouche	3B	395	14	5	14,5	399	14	9	14,9	12	0.03	12.5	9.5	99.2	95.4	91.4	Out	Roberto	Wahlberg
11	24	22-JUN-17	Juan	Carlos	LF	422	15	2	15,2	427	15	7	15,7	11	0.02	8.7	10.2	97.6	93.3	87.4	Out	Bonds	Wahlberg
12	1	23-JUN-17	Bondie	Wahlberg	SS	429	15	9	15,9	401	14	11	14,11	10	0.02	6.3	12.2	95.9	97.1	99.3	Out	Young	Cuddy
13	25	26-JUN-17	Juan	Carlos	LF	398	14	8	14,8	431	15	11	15,11	16	0.12	9.4	11.3	98.2	92.2	54.3	Out	Ramirez	Diaz
14	12	04-JUL-17	Derrick	Lynn	CF	712	24	22	24,22	716	24	26	24,26	79	0.02	15.3	17.3	98.7	26.7	93.2	Out	Bonds	Cuddy
15	4	09-JUL-17	Donnie	Miller	1B	104	4	14	4,14	74	3	14	3,14	4	0.02	4.1	4.9	100	92.3	101.2	Out	Roberto	Tyler
16	20	11-JUL-17	David	Roberts	3B	396	14	6	14,6	399	14	9	14,9	5	0.4	5.4	2.4	99.5	98.3	100.4	Out	Roberto	Wahlberg
17	44	12-JUL-17	Pablo	Sanchez	SS	399	14	9	14,9	371	13	11	13,11	15	0.05	13.4	11.4	98.4	92.2	76.6	Out	Teal	Wahlberg
18	22	13-JUL-17	David	Roberts	3B	399	14	9	14,9	399	14	9	14,9	1	0.01	(null)	(null)	99.9	99.8	103.3	Out	Harvey	Wahlberg
19	21	13-JUL-17	David	Roberts	3B	399	14	9	14,9	459	16	9	16,9	11	0.3	6.7	5.7	93.4	93.4	97.4	Out	Teal	Diaz
20	19	17-JUL-17	Kris	Travis	RF	176	6	26	6,26	178	6	28	6,28	17	1.2	8.6	12.3	83.2	94.4	101.3	Safe	Teal	Wahlberg
21	13	18-JUL-17	Derrick	Lynn	CF	708	24	18	24,18	527	18	17	18,17	53	0.04	12.4	15.4	98.9	45.7	76.7	Out	Ramirez	Lucas
22	45	21-JUL-17	Pablo	Sanchez	SS	399	14	9	14,9	371	13	11	13,11	15	0.7	10.4	9.4	89.2	92.2	50.1	Safe	Roberto	Quenton
23	6	21-JUL-17	Donnie	Miller	1B	133	5	13	5,13	102	4	12	4,12	7	0.92	3.4	6.2	88.6	88.4	98.4	Safe	Teal	Wahlberg
24	37	01-AUG-17	Michael	Tyler	1B	134	5	14	5,14	221	8	11	8,11	19	0.06	8.2	11.5	88.4	91.3	30.4	Safe	Young	Carlos
25	28	02-AUG-17	Michael	Cuddy	LF	696	24	6	24,6	693	24	3	24,3	12	0.3	4.4	4.6	98.9	94.5	67.8	Out	Teal	Wahlberg
26	36	21-AUG-17	Michael	Tyler	1B	134	5	14	5,14	194	7	14	7,14	15	0.02	10.4	8.4	98.3	85.6	67.8	Out	Teal	Tyler
27	5	21-AUG-17	Donnie	Miller	1B	104	4	14	4,14	194	7	14	7,14	9	1.1	3.1	5.6	89.2	93.4	96.6	Safe	Young	Tyler
28	29	23-AUG-17	Michael	Cuddy	LF	696	24	6	24,6	704	24	14	24,14	9	0.02	15	17.2	97.9	75.5	90.6	Out	Teal	Wahlberg
29	35	27-AUG-17	Mike	Diaz	C	63	3	3	3,3	96	4	6	4,6	13	0.2	7.6	8	99.2	95.4	34.5	Out	Young	Tyler
30	30	27-AUG-17	Michael	Cuddy	RF	324	11	24	11,24	142	5	22	5,22	43	0.04	16.2	18.1	99.2	62.3	75.2	Out	Young	Diaz
31	41	05-SEP-17	Gregg	Quenton	RF	296	10	26	10,26	300	10	30	10,30	63	0.07	14.3	15.3	94.6	82.3	96.4	Out	Young	Wahlberg
32	42	08-SEP-17	Gregg	Quenton	RF	296	10	26	10,26	141	5	21	5,21	53	0.02	12.1	12.3	95.6	91.4	83.4	Out	Teal	Miller
33	2	10-SEP-17	Bondie	Wahlberg	SS	369	13	9	13,9	186	7	6	7,6	27	0.04	9.2	14.1	98.9	63.6	52.5	Out	Bonds	Tyler
34	9	14-SEP-17	Clyde	Dog	2B	284	10	14	10,14	346	12	16	12,16	24	0.06	9.6	15.4	99.2	42.3	90.2	Out	Roberto	Tyler
35	10	22-SEP-17	Clyde	Dog	3B	396	14	6	14,6	393	14	3	14,3	8	0.04	7.8	5.6	99.8	37.8	110.1	Out	Teal	Wahlberg
36	48	22-SEP-17	Kirk	Lucas	1B	103	4	13	4,13	194	7	14	7,14	10	0.01	9.6	7.5	98.3	92.1	79.1	Out	Bonds	Lynn
37	11	27-SEP-17	Clyde	Dog	3B	366	13	6	13,6	482	17	2	17,2	17	0.12	12.3	14.3	96.7	33.4	56.7	Out	Young	Cuddy
38	26	01-OCT-17	Jose	Flores	2B	254	9	14	9,14	196	7	16	7,16	22	0.5	6.7	8.7	89.3	97.6	81.2	Safe	Young	Wahlberg
39	27	02-OCT-17	Jose	Flores	2B	254	9	14	9,14	344	12	14	12,14	28	0.3	9.4	12.3	90.3	86.7	72.2	Safe	Young	Miller
40	49	25-OCT-17	Kirk	Lucas	1B	104	4	14	4,14	75	3	15	3,15	9	0.5	3.4	5.4	82.3	98.7	67.1	Error	Ramirez	Tyler

ThrowsToHome_View

Purpose

The creation of this view allows the analyst to see the Throw Velocity on specific throws. In this distinct case, this view are all the plays in which a player throws to home plate. These types of throws are closely tied to the baseball concepts of runs and sacrifice flies. For context, most of these throws are of maximum effort by the player so it is a legitimate measure of arm strength. Because of our field grid, we know that home plate is located at FieldID 63. So all throws to this FieldID are considered throws to home plate.

Create Statement

```
CREATE OR REPLACE VIEW THROWSTOHOME_VIEW AS
Select
Player.FirstName as FirstName,
Player.LastName as LastName,
Position.PositionName as Position,
Play.ThrowVelo as ThrowVelocity
FROM Player,PlayerPosition,Play,Position
WHERE
Player.PlayerID=PlayerPosition.PlayerID
and PlayerPosition.PlayerPositionID=Play.FielderID
and PlayerPosition.PositionID=Position.PositionID
and Play.pend=63;
```

Sample Data

	FIRSTNAME	LASTNAME	POSITION	THROWVELOCITY
1	Derrick	Lynn	CF	105.5
2	Michael	Cuddy	LF	104.7
3	Derrick	Lynn	CF	103.5
4	Gregg	Quenton	RF	103.4
5	Michael	Cuddy	RF	101.7
6	Michael	Cuddy	LF	101.7
7	Kris	Travis	RF	88.7

ReactionTime_View

Purpose

This view provides information about reaction speed. By utilizing our FirstStep data, the view shows an average reaction speed on batted balls for players. Reaction time is measured in seconds. Sorted by Reaction Time.

Create Statement

```
CREATE OR REPLACE VIEW REACTIONTIME_VIEW AS
Select Play.FielderID as FielderID,
Player.FirstName as FirstName,
Player.LastName as LastName,Position.PositionName as Position,
ROUND (AVG(CAST(Play.FirstStep as FLOAT)), 2) as ReactionTime
FROM Play,PlayerPosition,Player,Position
WHERE
Play.FielderID=PlayerPosition.PlayerPositionID
and PlayerPosition.PlayerID=Player.PlayerID
and PlayerPosition.PositionID=Position.PositionID
GROUP BY play.fielderid, Player.FirstName, Player.LastName, Position.PositionName;
```

Sample Data

	FIELDERID	FIRSTNAME	LASTNAME	POSITION	REACTION...
1	40	Todd	Matthews	SS	0.02
2	27	Taylor	Tyler	3B	0.02
3	8	Derrick	Lynn	CF	0.03
4	37	Ricco	LaRouche	3B	0.03
5	28	George	Michaels	CF	0.03
6	26	Michael	Tyler	1B	0.04
7	24	Michael	Cuddy	RF	0.04
8	31	Gregg	Quenton	RF	0.05
9	6	Clyde	Dog	2B	0.06
10	7	Clyde	Dog	3B	0.08
11	18	Juan	Carlos	LF	0.09
12	36	Pablo	Sanchez	3B	0.1
13	22	Michael	Cuddy	LF	0.16
14	1	Bondie	Wahlberg	SS	0.19
15	25	Mike	Diaz	C	0.2
16	13	David	Roberts	3B	0.24
17	38	Kirk	Lucas	1B	0.26
18	5	Kenneth	Greene	C	0.28
19	34	Pablo	Sanchez	SS	0.38
20	20	Jose	Flores	2B	0.4
21	4	Donnie	Miller	1B	0.68
22	11	Kris	Travis	RF	1.2

PitcherCatchProb_View

Purpose

PitcherCatchProb_View provides a different use for the MLBAM Statcast database. In this case, the View takes into consideration all plays attached to a distinct PitcherID (alias for PlayerID) and puts out an average catch probability of each batted ball. This is to judge difficulty of catches, not results of those plays. It is meant to show if a pitcher is giving his fielders a proper chance to make a play.

Create Statement

```
CREATE OR REPLACE VIEW PITCHERCATCHPROB_VIEW AS
Select
Player.FirstName as FirstName,
Player.LastName as LastName,
Player.Throw as Throws,
ROUND (AVG(CAST(Play.ExpectedCatchProb as FLOAT)), 2) as ExpectedCatchProbability
FROM Player,Play,PlayerPosition
WHERE
Player.PlayerID=Play.PitcherID
and Player.PlayerID=PlayerPosition.PlayerID
and PositionID=1
GROUP BY Player.LastName, Player.FirstName, Player.Throw;
```

Sample Data

	FIRSTNAME	LASTNAME	THROWS	EXPECTEDCATCHPROBABI...
1	Jacob	Harvey	R	99.8
2	Chris	Teal	R	84.2
3	George	Ramirez	L	83.33
4	Michael	Young	L	80.51
5	Matt	Roberto	L	80.41
6	Tyler	Bonds	R	68.93

InfPlaysOver15_View

Purpose

This view provides a look at infielders range. The distance is set to any plays made in which the player traveled more than 15 feet. It takes into account the player's position, route efficiency and the result of said play.

Create Statement

```
CREATE OR REPLACE VIEW INFPLAYSOVER15_VIEW AS
Select
Player.FirstName as FirstName,
Player.LastName as LastName,
Position.PositionName as Position,
Play.Distance as Distance,
Play.RouteEfficiency as RouteEfficiency,
Play.Result as PlayResult
FROM Player,PlayerPosition,Play,Position
WHERE
Player.PlayerID=PlayerPosition.PlayerID
and PlayerPosition.PlayerPositionID=Play.FielderID
and PlayerPosition.PositionID=Position.PositionID
and Position.PositionID BETWEEN 3 and 6
and Play.Distance>15;
```

Sample Data

	FIRSTNAME	LASTNAME	POSITION	DISTA...	ROUTEEFFICIENCY	PLAYRESULT
1	Jose	Flores	2B	28	90.3	Safe
2	Bondie	Wahlberg	SS	27	98.9	Out
3	Clyde	Dog	2B	24	99.2	Out
4	Jose	Flores	2B	22	89.3	Safe
5	Michael	Tyler	1B	19	88.4	Safe
6	Bondie	Wahlberg	SS	18	78.2	Safe
7	Clyde	Dog	3B	17	96.7	Out

BatterExitVelocity_View

Purpose

Another view testing the flexibility of the Statcast Database: given the Exit Velocity of each play and how that play is attached to a BatterID (alias for PlayerID), we can create a view in which we compare all of the batter's exit velocity. The view shows the Player, the Average Batted Exit Velocity and which team the player is on.

Create Statement

```
CREATE OR REPLACE VIEW BATTEREXITVELOCITY_VIEW AS
Select
Player.FirstName as FirstName,
Player.LastName as LastName,
Team.TeamName as Team,
ROUND (AVG(CAST(Play.ExitVelocity as FLOAT)), 2) as ExitVelocity
FROM Player,Team,Play,Season
WHERE
Player.PlayerID=Play.BatterID
and Player.PlayerID=Season.PlayerID
and Season.TeamID=Team.TeamID
GROUP BY Player.FirstName, Player.LastName, Team.TeamName;
```

Sample Data

	FIRSTNAME	LASTNAME	TEAM	EXITVELO...
1	David	Roberts	New York Yankees	89.2
2	Bondie	Wahlberg	New York Mets	88.94
3	Donnie	Miller	New York Mets	83.33
4	Michael	Cuddy	New York Yankees	83.07
5	Mike	Diaz	New York Yankees	76.88
6	Mike	Diaz	Boston Red Sox	76.88
7	Kirk	Lucas	Toronto Blue Jays	76.7
8	Taylor	Tyler	Minnesota Twins	72.8
9	Derrick	Lynn	New York Mets	71.9
10	Michael	Tyler	New York Mets	52.5
11	Juan	Carlos	New York Mets	50.1
12	Gregg	Quenton	New York Mets	50.1

Implementation Notes

November 8th, 2017 ⟶ Research and Development of Basic ERD tables

November 15th, 2017 ⟶ Decided upon making the Field into a Grid system for Player Tracking

November 22nd, 2017 ⟶ Introduced fields such as RouteEfficiency, FirstStep, and ExpectedCatchProb

December 2nd, 2017 ⟶ Completed ERD and uploaded into SQL Developer

December 4th-6th, 2017 ⟶ Inserted all data into the database

December 7th-9th, 2017 ⟶ Created and Replaced Views to Test Database Usability and Flexibility

December 10th-12th, 2017 ⟶ Drafted Final Report on MLBAM Statcast Database

Known Problems

- I had trouble connecting the Play table with the Field table in SQL Developer. In Data Modeler, I attached both PStart (play starting point) and PEnd (play ending point) in my Play table to FieldID of the Field table. In my data, I recorded PStart and PEnd as different integers. But when I queried PStart=FieldID and PEnd=FieldID, I only received one result: a play in which the baseball was hit directly at the fielder. In the end, I was able to define PStart and PEnd as querying two separate FieldIDs. But, looking back, I could have created a separate junction table to mitigate this problem.
- In the Player table, I could have created a separate junction table for "Bats". This would have allowed for a batter who bats Switch. Right now, it only allows for Left or Right handed batters.
- The Play table does not account for every possible play that can occur on the baseball field. For example, if a runner gets into a rundown and the ball is thrown multiple times. Another further enhancement can be tracking a base stealer. It would be interesting to see reaction time compared to a pitcher's "move" to the plate.