



1	Catcher	(none)	Out			C	Called Strike
2	Pitcher	S	Single			S	Swinging Strike
3	First Base	D	Double			B	Ball
4	Second Base	T	Triple			F	Foul Ball
5	Third Base	HR	Home Run	G	Ground Ball	L	Foul Bunt
6	Shortstop	HP	Hit by pitch	L	Line Drive	M	Missed Bunt
7	Left Field	W	Walk	P	Pop Fly	P	Pitchout
8	Center Field	IW	Int. Walk	F	Fly Ball	Q	Pitchout (strike)
9	Right Field	WP	Wild Pitch	B	Prefix indicating bunt	R	Pitchout (foul)
	(a) Positions	(b) Play Type				I	Int. Ball
						H	Hit by pitch
						K	Strike (unknown)
						U	Unknown
						n <sup>1</sup>	Pickoff
						+n <sup>2</sup>	Pickoff by catcher
							(d) Pitches

<sup>1</sup>n is 1, 2, or 3, corresponding to base thrown to. This should not be confused with player number.

<sup>2</sup>See above

## Sample Records

	<u>Player ID</u>	<u>Player Name</u>	<u>Team</u>	<u>Batting Order</u>	<u>Position</u>	
<i>start,</i>	<i>grudm001,</i>	<i>"MarkGrudzielanek",</i>	<i>0 ,</i>	<i>1</i>	<i>,</i>	<i>6</i>
	<u>Inning</u>	<u>Player ID</u>	<u>Count</u>	<u>Pitches</u>	<u>Play</u>	
<i>play,</i>	<i>1,0 ,</i>	<i>aaroh101,</i>	<i>12 ,</i>	<i>BF1CX,</i>	<i>S</i>	<i>7 / L7S . 3 - H; 1 - 2</i>
					<u>Play Type Location</u>	<u>Description</u> <u>Runner movement</u>

**Hadoop** A platform for storage and processing of large data sets. Hadoop consists of a distributed filesystem, Java interfaces, and services for managing distribution of tasks. Core code is maintained by the Apache Software Foundation.

**MapR distribution for Hadoop** MapR provides an enterprise-ready distribution for Hadoop including the core open source components as well as a high-performance distributed filesystem, high-availability services, and easy-to-use management interface.

**MapReduce** A programming model for parallel processing, introduced in Google's 2004 paper<sup>3</sup>. Hadoop provides a framework for running MapReduce jobs written in Java.

**Pig** A high level data flow language for processing data. Pig programs describe steps to be executed on sets of tuples, and are executed by running one or more MapReduce jobs.

**Hive** A framework for executing SQL-like queries on large data sets. The queries are executed as MapReduce jobs.

---

<sup>3</sup><http://research.google.com/archive/mapreduce.html>