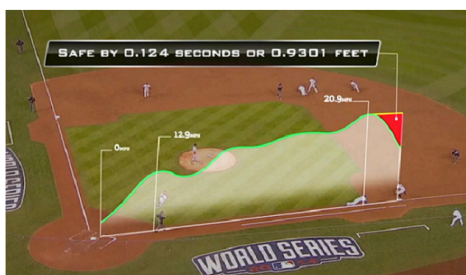


Major League Baseball Fields Big Data, and Excitement, with AWS

MLB Advanced Media (MLBAM) wanted a new way to capture and analyze every play using data-collection and analysis tools. It needed a platform that could quickly ingest data from ballparks across North America, provide enough compute power for real-time analytics, produce results in seconds, and then be shut down during the off season. It turned to AWS to power its revolutionary Player Tracking System, which is transforming the sport by revealing new, richly detailed information about the nuances and athleticism of the game—information that's generating new levels of excitement among fans, broadcasters, and teams.



Data from the Player Tracking System (Statcast) overlaid on video of the Panik-Hosmer play. The red section on the right shows that if Hosmer had maintained his speed instead of diving to the bag, he would have been safe by about a foot.

Revealing Greater Subtleties of the Sport

Data plays a huge role in baseball, with fat volumes of statistics cataloging the game's arc over the seasons. This information, however, is historical and static. MLBAM wanted to change its approach to statistics by capturing and analyzing data in real time to reveal greater subtleties about the sport.

MLBAM considered an on-premises IT solution, but ultimately ruled it out. "We looked at using compute capabilities in all the stadiums," says Dirk Van Dal, MLBAM's vice president for multimedia technology development. "But distributing the data efficiently and from so many locations would have involved a lot of time and investment in expensive IT resources that would sit idle for about half the year."

The AWS cloud offered an ideal alternative that could support as many as 15 games on a single day—and some days with just one or two.

"AWS provides nationwide coverage for reasonable round-trip times for sending data between the game sites and the cloud, and multiple services that we use for building Statcast," Van Dal says. "It provides great scalability, so we can burst when we need it the most, manage just one, two, or many games on a single day, and then shut down the resources during the off season."

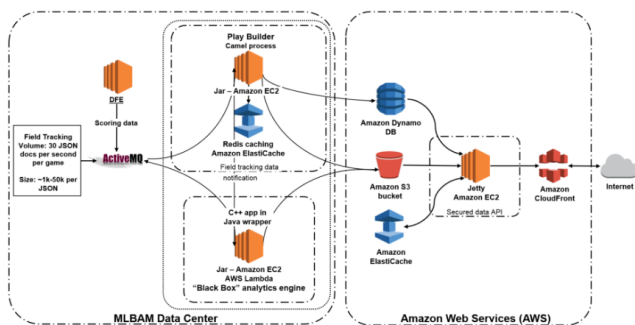
The workflow begins with two data-acquisition systems at the stadiums that provide coordinate information. A Doppler radar system sits behind home plate, sampling the ball position 2,000 times a second. Two stereoscopic imaging devices, usually positioned above the third-base line, sample the positions of players on the field 30 times a second. Data from these systems is augmented by brief written descriptions of each play entered by personnel on the field after the action is over.

Ten to 15 seconds after a play is completed, the data is transmitted over private networks at the stadiums, aggregated, and then sent to the AWS cloud using [AWS Direct Connect](#), which provides a dedicated network connection for rapid data delivery. MLBAM uses [Amazon Elastic Compute Cloud \(Amazon EC2\)](#) for the compute power behind the solution. The coordinate data from each play is stored in [Amazon Simple Storage Service \(Amazon S3\)](#), which will expand to hold the vast amount of information generated through the solution. MLBAM anticipates that an average of 7 TB of data will be generated per game. With 2,430 games in a season, that's about 17 petabytes of data each season."

MLBAM uses [Amazon ElastiCache](#) to temporarily store game information in memory caches instead of on hard drives, which enables fast retrieval of the data for analysis tasks. [Amazon DynamoDB](#) powers queries and supports the fast data retrieval required, while [Amazon CloudFront](#) delivers a scalable solution to serve up the APIs.

[AWS Lambda](#), a serverless computing service that runs code in response to events, supports analysis of data feeds in the solution's metrics engine. "Lambda is really clever. It's where we take the raw data, do some cleaning up and error detection, then create the metrics that bring more insights into plays—the throws, the player's acceleration rate, the top running speeds," Van Dal says. "We're accessing a truly big data mine, and have yet to scratch the surface."

The analysis happens within milliseconds after the data is received, a key for broadcasters to take delivery of the raw metrics and video within 12 seconds after a play is complete.



The Statcast architecture powered by AWS. Click on the image to enlarge.

Delivering a New Level of Excitement

Speed, scalability, and the ability to capture, analyze, and deliver large quantities of data in different ways are central to MLBAM's efforts to innovate for the benefit of everyone who loves the game, especially fans who now have reliable metrics for those arguments about who, for example, runs the bases most efficiently or has the fastest reaction times to fielding line drives.

"We're giving fans empirical information to power that conversation, which is a huge part of what sports is all about," says Inzerillo.

Broadcasters also have new information to use for on-air analysis, further enhancing viewer engagement, while clubs have new data and tools to analyze and coach players.

"We believe the Player Tracking System powered by AWS will deliver new and more exciting information to apps and devices, and that will appeal to a younger generation of fans, who are used to video games and who have a lot of expectations about the viewing experience," Van Dal says. "It delivers a new level of excitement to baseball."