



**Data analytics
with Google Cloud:
a **winning strategy**
for the NCAA[®]
and your business**

Google Cloud

Introduction

As the NCAA's official cloud partner, Google Cloud helps turn decades of game data into valuable insights and real-time predictions. Beyond modeling matchups and forecasting free throws, it's a scalable, serverless approach to data preparation and analytics that has numerous applications for all kinds of organizations.

Does defense really travel better than offense? What factors contribute most to optimal athlete health? How many times has a team with three freshmen starters shot better than 50 percent from a three-point range when they were losing by eight points with nine minutes to go?

The answers to these questions lay in the NCAA's millions of data points tracing back more than 100 years. But coaches, officials, and fans needed a way to glean actionable insights about team and player performance – and even build simulations for tournament analysis and forecasting – from this statistical treasure trove. So in 2017, the NCAA leveraged Google Cloud Platform to create a smart, serverless, and fully managed architecture for descriptive and predictive data analytics workflows. The goal was to elevate the sophistication of the NCAA's analytics, allowing it to derive real value from its data.



Getting data ready for game time

As their pilot initiative, the NCAA and the Google Cloud team tackled historical and play-by-play data from the Division I Men's and Women's Basketball Championships. Like that of most large organizations, the NCAA's data resided in a variety of locations and existed in multiple formats. The first step to this large-scaled transformation was getting everything into one centralized repository and preparing it for analysis.

A group of Googlers (known as "The Wolfpack") migrated 80-plus years of play-by-play game and tournament information into [BigQuery](#), Google Cloud's powerful data warehouse with built-in machine learning. A combination of JSON, XML, and CSV files were ingested via RESTful services or pulled from FTP and sent to [Cloud Storage](#). Then the team used [Cloud Dataflow](#) to extract the data from Cloud Storage, transform it into actionable structures, and load it into BigQuery.

Making sense of past performance, one query at a time

To extract insights from the data in BigQuery, The Wolfpack used standard SQL to build hundreds of different queries, each focused on a specific analytic goal. For example, they investigated whether players are shooting more three-pointers than they used to. Prior to the migration, answering this question would have required opening tens of thousands of files on different servers and writing a query against each one. The Wolfpack got their answer – yes – with just a few lines of code. Other queries sought to quantify qualitative measures to better analyze them, including variables such as pressure, performance, competitiveness, and experience.



Next, the team used two different tools to create informative and descriptive visuals. [Google Data Studio](#) enabled them to rapidly develop dashboards, drawing on various data sources in addition to BigQuery. [Cloud Datalab](#) helped them build more complex, data-science-focused visualizations.

Staying one step ahead of March Madness®

In addition to querying historical data to report on past performance, the NCAA and the Google Cloud team wanted to experiment with anticipating events and outcomes as live games unfolded. For the 2018 March Madness tournament, The Wolfpack used machine learning to make pregame and halftime predictions for metrics like shots, turnovers, dunks, and momentum, based on both historical data and real-time data from live broadcasts. They shared these real-time predictions with [halftime ads](#) on television and online. The Google Cloud team later proved that the ability to accurately predict specific key game metrics often contributed to better prediction models for entire game outcomes.

Getting students in on the action

For the 2019 March Madness tournament, the Google Cloud team has invited 30 college sophomores with an interest in computer science, data science, statistics, and, of course, basketball to participate in the NCAA data analysis project. Continuing last year's focus on quantifying qualitative variables, some of the students are exploring hypotheses and queries for concepts such as competitiveness, [grittiness](#), [clutchness](#), and explosiveness. During March Madness, the students will help with pregame analysis of matchups. They will also write queries based on that analysis that allow them to take a subjective concept and turn it into an objective metric. For the concept of explosiveness, for example, the Wolfpack might post to the [campaign website](#) a definition and analysis of explosiveness and state that based on

“Every player wants to be the best they can be. Every coach wants to be the best they can be. And to have more tools around the ability to learn about your game and to improve because of those tools and that data is incredibly powerful.”

— Dan Gavitt, Senior Vice President of Basketball, NCAA

this analysis, here is how two teams compare to each other in terms of explosiveness. Visitors to the Hub could drill down into a metric that contributes to the explosiveness definition. In this case explosiveness is focusing on understanding scoring margin within runs.

Beyond the basketball court

The partnership between Google Cloud and the NCAA shows that there's so much to learn from liberating your data from cumbersome legacy processes and combining it with cloud technology – and that's true whether you're a basketball coach, a stat-obsessed sports fan, or an enterprise organization. Google Cloud makes it easy with a serverless, end-to-end solution for seamless ETL, powerful queries, data visualization, predictive analytics, and machine learning. Armed with these tools, data scientists and analysts can work with modest datasets or run complex queries over petabytes of information, testing out creative new hypotheses and making their data actionable – and valuable.

For example, using an analytics architecture like the NCAA's, a retail business hoping to glean a specific insight from multiple customer touchpoints can get answers in seconds. A manufacturer can run predictive maintenance on machines in multiple plants, powered by data-intensive workflows that update the data warehouse in real time.

In today's information-filled world, organizations are more likely to succeed when they can tap their data to gain business insights and improve the customer experience. To put it in basketball terms, it may even be the difference between winning and losing.

Find out how your organization can advance in tomorrow's tournaments with Google Cloud.





Google Cloud