

# How business taught scientists about big data

Traditionally, scientists and researchers develop the latest and greatest techniques in computing, which then trickle down corporate data centers where they're relevant. But with big data — the process of analyzing voluminous quantities of data in new, unique ways — it's the industry that's driving innovation. Look at Google; look at Walmart; look anywhere you like. Big data means dollars and cents to companies, so they take it very seriously.

Last month, I sat down with former Yahoo chief data officer and current [ChoozOn](#) CTO Usama Fayyad, who gave his explanation of how this happened. Essentially, he said, businesses took the big data reins by necessity when the proliferation of web-derived data hit its stride several years ago. Companies suddenly had all this data, and competitive pressures forced them to find ways to use it to their advantage.

Indeed, at the [ACM SIGKDD Conference on Knowledge Discovery and Data Mining](#) where I met Fayyad (he's the ACM SIGKDD chair), he estimated the demographics have flipped from about 70 percent academicians to only about 40 percent.

Researchers, of course, had mined and analyzed huge scientific data sets for years and were always working on new techniques for improving their results. But it was the difference between academic culture and corporate culture that resulted in a relatively rare industry-led technological revolution once companies got interested in the problem. Whereas researchers sometimes took process shortcuts or ignored certain difficult-to-solve problems, businesses invested the resources to do all sorts of new things and do them accurately, Fayyad said.

And, he explained, it wasn't just mega web companies such as Google and Yahoo that drove analytic innovation. Whatever a company's focus — travel, retail, gaming, etc. — they all invested in data-analysis techniques unique to their own fields. In fact, he noted, an upstart Amazon [bought data-mining and](#)

[recommendation-engine pioneer Jungle](#) in 1998 while Google was still an early-phase company.

Google certainly made important contributions to the big data space, but Fayyad noted that, in some ways, its early work on search engines was relatively easy because search users have a relatively high tolerance for error (i.e., how would they even notice an error?). For companies such as casinos, retailers and others, though, customer loyalty is critical, so using data to optimize the customer experience has always been of the utmost importance.

Industry also had one big advantage over scientists when it comes to analytics: They have lots and lots of relevant data. In fact, Fayyad noted, it was easier than anticipated for him to hire data scientists to work for Yahoo because Yahoo could give them access to massive hardware and data sets the likes of which they couldn't find at a university.

By now, of course, cloud computing and [research testbeds](#) from many entities have democratized access to resources, but data is still relatively hard to come by. Yes, there are public data sets and many more [available for a fee](#), but that doesn't mean it's all relevant or useful for data researchers, or that it's particularly timely. As Fayyad explained, scientists can be just as protective of their data as business people are, so they often won't release it until they've finished with it and have published something based on it.

For all these reasons, it seems unlikely we'll see a return to the status quo with researchers driving big data innovation. But Fayyad seems confident that academicians and computer science students can continue to do important work as long as companies such as Google, Yahoo and Microsoft are open about their own research efforts and techniques, and if the various stakeholders can figure out a way to feed more and better data into universities.

By: [Derrick Harris](#), Sep. 26, 2011