Data science skills: Is NoSQL better than SQL?

15 Dec 201614.5k Views

Big data is one of the hottest sectors in tech right now, but how do you stay on top of the changing technologies? David Pardoe of Hays Recruitment talks about the differences between SQL and NoSQL in data.

How often have you heard that, in this (new) big data world, the newer NoSQL data sources and structures are the key to effective data science? And further, that relational data that can be queried with SQL is old-fashioned and traditional, and is no longer fit for purpose?

Why spend time and investment in building ETL processes that shift data from one database to another and enforce a rigid, non-scalable data model?

Why not just dump all of the data in an unstructured, or schema-less, model? Surely that gives you the most flexibility to really find what you are looking for in the petabytes of data that your organisation collects.

The reality is far more complex; as is usually the case in the field of data science. In fact, the discussion is moot, as it always has been when talking about which technology is best for solving business problems. I heard nearly identical debates when I started my career over 20 years ago, and I have found it odd to see such similar themes re-emerging.

The most critical aspect of data science is not the technology or the data structures; it is doing things that can result in better (or quicker) decisions being made. If you focus on that for just a second, you will realise that most, if not all, business decisions are made about things (or to get technical "entities").

If data science is going to help you make a better (or quicker) decision about something, then you had better make sure your data science output maps back to that thing. To be a little less abstract; if I am going to use data science to help me match a candidate on my database to the job I am being paid by my client to fill, then I better be able to map my data science outputs back to the candidate and job "things" (entities).

I can, of course, do this multiple ways, but the reality is that the candidate and job in my operational system are going to have a unique identifier of some kind, and I will need to link my insights back to those unique identifiers.

So, whether I choose to extract the data from my operational system, transform it and then load it to a NoSQL repository or a relational database, I am still going to need to write and execute ETL processes of some kind. The decision on which technology and data structure I use does, however, still need to be made.

My point here is that the decision about SQL or NoSQL should be driven by what skills and technology you will have or need in your job, rather than by what the newest, shiniest technology is.

The reality for most organisations is that a hybrid solution is almost always going to provide the greatest returns and biggest impact on the business. In essence, focus more on understanding the business decision you want to influence and less on the technology you are going to use.

By David Pardoe

A version of this article originally appeared on Hays' Viewpoint blog.

David Pardoe is the group head of data science at Hays. He joined in 2015 as Hays embarked on establishing data science as a core component of decision-making across the group.

Looking for jobs in tech or science? Check out our Employer Profiles for information on companies hiring right now and sign up for our Career Republic e-zine for a weekly digest of sci-tech careers news and advice.

Tweet
Share
Email
Share
More
Share
Pin This
Reddit

In-Depth: More on Hays

Related: data, skills, big data, Hays, SQL developers, data analysts

Make sure you never miss an opportunity

Get our weekly newsletter for award-winning news, features and advice on sci-tech careers

| First Name | Last Name | email@example.com |