



August / September 2012

zData Perspectives

The Big Deal About Big Data

by Craig S. Mullins

amazon



DB2 Developer's Guide: A...

\$53.99

Shop now

Big Data... I'm sure you've heard the term. It is all over the place these days. But what is Big Data? The cynic in me wants to say that there is no universal definition because the marketers want to keep it nebulous. You know the drill -- every product adapts, at least in the marketing literature, to become part of the next big thing. In this case the next big thing is Big Data.

Although there is a fair amount of this type of bandwagon adaptation going on, this is far too simplistic to be the entire answer. Indeed, Big Data has grown somewhat holistically over time, driven in part by very large data requirements

with extreme availability needs, such as on large web sites or the streaming data measurements taken by medical devices. At the same time the field of analytics has exploded with newer and more sophisticated tools being delivered for deriving useful observations with sophisticated algorithms on large data sets.

storage mechanisms include key-value stores, graph databases, and document stores.

Stream computing is another concept that gets tied into Big Data. Stream computing involves the ingestion of data (structured or unstructured) from arbitrary sources and the processing of it without necessarily persisting it. Any digitized data is fair game for stream computing. As the data streams it is analyzed and processed in a problem-specific manner. Stream computing is adopted in situations where data is difficult for humans to interpret easily and is likely to be too voluminous to be stored in a database. Examples of types of data include healthcare and stock trades.

The final aspect of Big Data is **data analytics**. We are storing all of this data for a purpose. By analyzing large amount of data and looking for trends, patterns, and "interesting" data, analytics can discover issues and solve problems that were not practical, or even possible, using traditional computing methods.

Now couple these trends with the NoSQL database movement and advanced analytics and we see the makings of a meme... a Big Data meme! But what exactly is Big Data? Forrester Research defines Big Data in the context of what it calls the 4 V's: Volume, Velocity, Variety, Variability.

The first V is **Volume** and that is the obvious one, right? In order for "data" to be "big" you have to have a lot of it. And most of us do in some form or another. A recent survey published by IDC claims that the volume of data under management by the year 2020 will be 44 times greater than what was managed in 2009.

But volume is only the first dimension of the big data challenge; the others are velocity, variety, and variability. **Velocity** refers to the increased speed of data arriving in our systems along with the growing number and frequency of business transactions being conducted. **Variety** refers to the increasing growth in both structured and unstructured data being managed. And the fourth V, **Variability**, refers to the increasing variety of data formats (as opposed to just relational data). Others have tried to add more V's to the Big Data definition, as well. I've seen and heard people add Verification, Value, and Veracity to this discussion.

Frequently, Big Data is coupled with **NoSQL** database systems. The biggest difference between a NoSQL DBMS and a relational DBMS is that NoSQL does not rely on SQL for accessing data. Additionally, a NoSQL DBMS typically does not require a fixed table schema, does not provide ACID properties (instead delivering "eventually consistent" data), and are highly scalable. There are no hard-and-fast rules as to how NoSQL databases store data. Some of the more popular NoSQL

So what is Big Data? We've talked about a lot of different things but we haven't really pinned down a definition yet. Personally, I think all this talk about V's and NoSQL just muddies the water. To me Big Data is so simple that it needs no definition. It is like saying Big Dog... you immediately know what I am talking about. Big Data is all about a lot of data. Big Data doesn't have to be NoSQL. And you don't have to sit there counting up your V's to see if you are doing it. Real-time analytics on large relational data warehouses qualifies as Big Data to me. And it should to you, too.

As a data bigot, I see the Big Data trend as a good thing. A lot of the more recent computing trends have been process orientated (e.g. object-oriented programming, web services, SOA). But the data is more important than the code and it always will be. As I've said before, applications are temporary, but data is forever! And if the Big Data trend helps us to be able to better protect, administer, and use our data, then I'm all in favor of it.

Although I am usually skeptical of marketing trends and industry memes, this one is different. We can use the rise of Big Data to the forefront of computing as a means to improve data quality, institute data governance, and pay more attention to our data management infrastructure. After all, if you are going to have Big Data it had better be *good* Big Data.

Big Data Forever!

From [zJournal](#), August / September 2012.

© 2012 Craig S. Mullins,

