Big Data for Small and Medium-Sized Businesses

The term “big data” paints a picture of the terabytes, or even petabytes, of data that large enterprises use to glean sharper insight and make better decisions. But how relevant is big data to a small or medium-sized business? How much of the big data promise is accessible to a company with limited resources?

According to Paul Zikopoulos, Vice President of Information Management Technical Sales and Big Data at IBM, big data is highly relevant to small and medium-sized businesses. Zikopoulos, who works with companies both large and small to incorporate the benefits of big data into their businesses, believes that smaller companies have been analyzing their existing data for years and that big data simply takes that analysis to a more advanced level, giving these businesses the opportunity to broaden the aperture on data they may already have or data they could easily get.

ForwardView spoke with Zikopoulos about the advantages that big data presents, as well as the latest related technology and tools that are available for small and medium-sized businesses. Read excerpts from the interview below.

Forward View:
How would you define “big data”?

Paul Zikopoulos:
I like to always respond to that question with my personal disclaimer: I really dislike this term because it brings with it an implicit myth – that big data, as the name implies, is just about large data sets. I want to start out by noting that big data isn’t only about the volume of data, it’s about bringing data together that hasn’t been correlated in the past. Big data can indeed be about more data, but it doesn’t have to be.

On the “more data” theme, I would suggest that it’s about data sets that are larger than what you have today. And especially for the SMB community, that’s important. You don’t have to ingest petabytes of data to get some value. I think a lot of people think that big data just means Hadoop and that’s it. That’s absolutely not true. I’ve seen a client with a 100-node Hadoop cluster that housed over 300 TBs of data and zero analytics – that isn’t big data, that’s the ability to store a lot of data.

If I had to define big data for you, I’d zone in on four [components] that I think will give everyone a really good framework in which to understand big data. The first is pretty obvious – it’s volume. Another one is variety – you start to look at analyzing different kinds of data; imagine the effect of being able to correlate a blog post I made about your product with my customer record when I call you up with a problem! The next one is veracity, [which] really stands for the trustworthiness of the data; “where did this data come from?” is the first question that leads you down into this genre. Finally, and I think perhaps one of the most overlooked components in the way that I would define big data, is velocity. Velocity is a game changer, because it’s not just how fast data is produced or changed – it’s the speed that it has to be received, understood and processed. Most people talk about velocity as how fast your data volumes are growing. I see velocity as about moving your business from a forecasting model to a “now-casting” one.
Forward View:
How relevant is big data to small and medium-sized businesses?

Paul Zikopoulos:
I think it’s incredibly relevant, because when we think big data, we actually want to think analytics. Whenever I talk to customers of any size, I suggest [that they’ve] probably been doing big data for a long time. And the reason why I’ll suggest that is because if we can come to the agreement that [big data] is about analytics and deriving value and monetizing data, then my suggestion is that you have an analytics IQ. You’ve [already] been landing data and cleansing it and maybe aggregating it into cubes; you’ve done that to establish your IQ. What big data is in the new modern era, if you will, is an opportunity for you to increase your analytics IQ. You can increase your IQ by looking into raw data to find things you never dreamed about; you can connect the dots between your systems of record and systems of engagement; and more.

Forward View:
What are the competitive advantages that big data creates for a small or medium-sized business?

Paul Zikopoulos:
The competitive advantages are the same advantages they can create for large businesses. [All companies from] SMBs to [large] enterprises can benefit from a 360-degree view of your customer. They can also benefit from log analytics – something I often refer to as “data exhaust.” Data exhaust is all those click streams that go on. I bet you 95 percent of SMBs take this gold mine of data and throw it away like it’s the exhaust of an engine – their operational engine. What big data allows you to do is get more out of it and recycle it into valuable insight.

Forward View:
Can you talk a little bit about IBM’s big data analytics capabilities?

Paul Zikopoulos:
If I look across vendors that are out there from a breadth and depth perspective, it’s just really unparalleled. We’ve done a number of acquisitions as well as [all kinds of] homegrown technologies and really integrated them together into an analytics platform.

And, of course, we have our data warehouses – so you think of things like Netezza or DB2® or BLU Acceleration – and that’s where I can store all kinds of traditional data at rest in a “Load and Go” fashion. And then we provide an engine to perform analytics on data in motion called InfoSphere® Streams. And the neat thing is a lot of the analytics that you build at rest can be instantly deployed in motion using Streams.

Of course, let’s not forget about Hadoop. IBM has brought to market a nonforked Hadoop engine we call InfoSphere BigInsights. It has other features that make it easy to get analytics out of it: for example, a number of accelerators to get you going faster with log data and social media analytics, a spreadsheet data discovery tool, management tooling, and a set of features to harden it for enterprise use, among other things.

Finally, data exploration is becoming all the rage for big data. IBM made a major investment in this space a number of years ago when we purchased the leader in this space, Vivisimo, and integrated it into our big data platform as InfoSphere Data Explorer.

And, [whether] you are a large or small company, you need to be governing this data, so our suite of products – for example, InfoSphere Guardian® and InfoSphere Information Server – is integrated as well. Now, consider the rest of IBM and you really get an eye-opener. SPSS® models can be deployed on our streaming engine, Cognos® is integrated, hardware reference architectures are built, we’ve “appliancized” our at-rest engines, and more; it’s really a great story.

Forward View:
IBM PureData™ System for Analytics is designed specifically to help with big data. Can you talk a little bit about that?

Paul Zikopoulos:
When you start to look at what folks are trying to do with their data, this technology plays an incredible role because it provides a very deep analysis of data at literally immense amount of speeds. IBM PureData™
System for Analytics is going to give you an incredible opportunity to take data that you know to be trusted and true – or that you want to be trusted and true – and provide deep analytics on detailed data. You’re just going to plug the machine in. You’re not going to create indexes and you’re not going to create partitioning keys. You’re just going to start writing analytics. In fact, this analytics powerhouse has more built-in functions for mathematical, geospatial, time series, predictive and other analytic genres than its next three competitive offerings combined! It’s like turnkey analytics for immense volumes of data.

Forward View:
How convenient would it be for a small or medium-sized business to get started or switch over to IBM’s big data analytics?

Paul Zikopoulos:
You know, I think it’s probably pretty easy. We’ve seen a number of customers really knocking at our doors complaining of skyrocketing costs and maintenance costs. We have business partners and SMBs that work together. And we have an entire ecosystem that gives you lab quality–like development skills with direct communication lines back to our development labs [that will] be there as a partner.

Forward View:
How quickly have companies successfully achieved ROI using big data?

Paul Zikopoulos:
I can tell you with IBM PureData for Analytics, we walked into one client, did a proof of concept, and two days later they bought the box because they didn’t want us to take it out.

Just the other day, [we met with] this company that does fraud detection on credit cards. They were struggling with a competitor’s in-memory columnar technology. It had been there for two weeks, they couldn’t get much going – that’s a really steep time-to-value curve. Our technical team just happened to be on-site when feelings of exasperation were shared and they said, “Listen, do you mind if we use this box and put BLU Acceleration on it?” Well, in 48 hours, we were running a query. The very first query we ran found this organization tens of thousands of dollars in fraudulent transactions. The customer was just over the moon – we call it flattening the time to value. ROI has to be quick these days.