INTERVIEW WITH CONOR O’MAHONEY

Eric Green: Hello and welcome to a new podcast series from IBM software that explores the challenges IT managers and business professionals are facing today. I’m Eric Green and I’ll be talking with a range of experts to discover new perspectives, approaches and examples that can help meet these challenges and introduce you to the capabilities of smarter software from IBM. So let’s get started.

Welcome back to our next episode focused on data management. As organizations process more and more information, the value of the systems and availability and need to organize that data become much more critical. Here to discuss that and many other things with us today is Conor O’Mahoney, with Information Management at IBM. Thanks so much for joining us today, Conor.

Conor O’Mahoney: Thanks Eric, it’s my great pleasure.

Eric Green: So I thought to start with, I’d ask how has data management been evolving over the last few years.

Conor O’Mahoney: Well the past few years have been a really fascinating time for data management, and some of the highlights have included integrated hardware/software systems for data management. And they’ve captured a lot of the headlines with IBM and Oracle bringing some very interesting offerings like the IBM SmartAnalytic system, the IBM Pure Scale Application system, Oracle ExoData and Oracle Exologic to market. And then other vendors have been following suit.

You know, another big recent news story is the emergence of what Forrester is calling the database compatibility layer. It’s a set of features that allows customers to easily migrate from one database vendor to another. For instance, you know, in 2010 more than 1,000 Oracle database customers chose IBM DBII instead. But probably the hottest story right now is big data. Big data, it’s a set of data management challenges that are too big for approaches like traditional relational database management systems.

Eric Green: So those are all extremely interesting, obviously, and we could probably have an entire conversation on each one, but kind of – I was wondering if you could talk a little bit more around big data. I mean, it sounds a bit like organizations are running into limitations with their current approaches and how is this kind of coming into play?

Conor O’Mahoney: Well, a moment ago, I used the term too big. Well, too big typically refers to either the volume of data or the velocity of data. So by volume, I mean that there’s simply too much data to manage in a reliable or performant manner, using traditional systems. You know, it’s not going to be long before pedobyte systems are commonplace instead of a relative rarity. It might be difficult for some people to imagine pedobyte systems in their environment. But it shouldn’t be, because if you just think back five years ago and consider the data volumes in your environment and compare those with today – when you do this, you know, you’re quickly going to quickly realize that data volumes are growing at an incredible pace, and this pace is only going to accelerate as more data points are generated and pulled into systems.

Now the other aspect of too big was velocity. Actually, this is probably the most prominent aspect of big data today. By velocity I mean that the data is being generated too quickly to be processed in a timely manner using traditional systems. With the key phrase here being timely manner, there’s a certain class of data that’s generated at extreme velocity making a real time or near real time analysis by traditional systems impossible. So with the increase of machine generated data, and it’s the webscale collection of data that are the primary drivers of these extreme velocities of data.

Now there’s one more characteristic of big data that begins with V and that’s variety. Big data often includes a variety of data sources, including structured data, unstructured documents, weblocked data, streaming sensor readings, image data and so on. So to sum up, big data solutions address volume, velocity and variety limitations of traditional systems.

Eric Green: So just to expand on that, when it comes to the great increase in the amount of data that’s needing to be processed and these companies are just increasing in data – I mean, where is that coming from? And do we expect that that is going to continue in that direction? Do we see it to start consolidating at some point, or how is that piece of it – you know, what’s driving this demand? Like where is all of this data coming from on an increased scale?

Conor O’Mahoney: So there’s a couple of driver’s here. So one is a lot of organizations are becoming more sophisticated with regards what information they are pulling into decision-making processes. So in the past they may have stored a lot of information, but not necessarily leveraged it for decision-making. Well now, they want to leverage whatever information they can. So that’s one factor here, is simply bringing more data into decision-making.

And another factor is, there’s a lot of environments where there’s a lot more machine generated data. Now this could be, for instance, a smart meter environment. You know, traditionally your energy company would have come around and read the meter once a month. Now that’s not a lot of data, really, that’s one meter reading per month. But when you move to smart meters, you know, perhaps those smart meters are communicating back with a central facility once a day. Well, straightaway there you’ve got 30 times the amount of data. Now what if they’re communicating every 15 minutes back to some sort of central location? And so as you can see, this multiplies up very quickly, and smart meters are just one example. There are numerous other examples of either metering or sensoring or logging systems that are generating just tremendous amounts of data.

Eric Green: Excellent. Thanks for that. Well, so talking about examples, why don’t we sort of dig in a little bit on an example or some examples that you’ve seen real time through your experience at IBM with I guess big data, and in general, success stories around organizations that have really needed to roll out a good data management strategy.

Conor O’Mahoney: Sure. Yeah. One example that jumps to mind is a telecommunications providers. And IBM had been working with them to offer promotions to their subscribers. Now this provider needs to process 100,000 records per second with a 10 millisecond decision making latency. So as you can imagine, before big data solutions, analyzing this volume of data at this type of velocity, it was impractical. Another is a smart traffic system. It processes data from 250,000 GPS probes and 630,000 segments every second. Again, a tremendous volume and velocity of machine generated data. Another is an agency that processes 600,000 records per second with a 1 to 2 millisecond latency for decision making. This is simply astounding. And last but not least, one of my favorites – IBM Watson. You know, it’s system that computed on the Jeopardy game show in the US. This system processes a variety – you know, that V word from our original definition, of unstructured and structured data with a latency of less than 3 seconds for highly complex decision-making.

Eric Green: To extend this conversation, do you think you can talk a bit about the technologies organizations can apply here?

Conor O’Mahoney: Absolutely. So, when people talk about big data technologies, they can actually refer to a number of different technologies, including grid-based map produced systems, like Apache Hadoop or IBM Big Insights. These systems, they can process large volumes of data with low latency. They could also be referring to systems that analyze streaming data, like IBM Infosphere Streams. These process extreme volumes of data with low latency. It could also be talking about massively parallel processing systems like IBM Netiza that process large volumes of data with low latency. Another common big data technology is in-memory database systems like IBM Solid DB that process data with extremely low latency. So there are actually a number of different technologies from map produced systems, to streaming systems to MPP systems to in-memory systems and beyond.

Eric Green: So this whole talk around big data sounds a whole lot like it’s – in a sense, it’s the end of traditional database systems as we know it. I mean, is that the reality of this?

Conor O’Mahoney: You know, that’s a great question. It’s a very common misconception that this is an either-or situation. But the reality is, and we’re seeing this reality from, you know, actually working with our clients with these solutions – the reality is that many real world situations see the two work hand in hand. So with big data technologies tackling the part of the challenge that they’re best suited to and relationship systems tackling the part of the challenge that they’re best suited to.

So if you remember, a few minutes ago, you know, I briefly described some real world big data systems. Well one really interesting aspect of all of those big data systems that they all have a relational database as part of the overall solution. So the solution at the telecommunications provider, well that actually has a streaming analytic solution that works in unison with the relational Data Warehouse. Whereas, you know, both the smart traffic system and the agency that I mentioned, they both bring data from traditional relational systems into the decision making process. And IBM Watson, it actually uses DBII to quickly work with the answers that the big data component of its system generates.

So expect both traditional relational systems and big data systems to work hand in hand in your data centers as we move forward. And of course, something that you’ll want to be cognizant of as you work with big data solutions is their integration into your existing relational systems.

Eric Green: Very interesting. And we’re running out of time here, but I mean, it seems that in this conversation, if you look at good database management and the like here, I mean you have cost savings involved, you have, you know, resource savings involved, you have customer experience being, you know, affected. So many parts of the enterprise are affected by this. I thought maybe as a close you could give us maybe your top two, three, four suggestions on things that organizations should be thinking about when they’re managing their data.

Conor O’Mahoney: Well, I think the most pressing topic for everybody in today’s environment is controlling costs. Data management in general, what we’re seeing is a need to do more with less and to keep costs relatively steady. There’s a number of strategies for doing this. You know, one of which is looking at alternative database management systems that are lower in cost. Another is leveraging technologies like data compression that can really bring huge savings to data storage related costs. Another is being cognizant of the staffing overhead associated with certain products. You know, staffing is a significant cost factor in a lot of IT environments, and doing some level of evaluation of the relative staffing needs of respective products, it’s a very wise thing to do to keep that manageable. Because also, not only are you coping with staffing resources but also with, you know, the happiness of those staff resources.

Eric Green: Excellent. Thank you very much for that. And thank you, very much, Conor, for joining us for the show today.

Conor O’Mahoney: It’s been my great pleasure, thanks for having me.

Eric Green: Thanks for listening. Please do visit IBM.com/software to connect with our experts, continue the conversation, and to learn more about smarter software from IBM. Let’s build a smarter planet.