

In today's episode, it is my pleasure to speak to James Rowland-Jones, principle program manager of Microsoft. James specializes in scale-out data solutions Azure SQL Data Warehouse, Microsoft Analytics Platform System and Hadoop. Welcome to the podcast.

JRJ: Thank you very much.

MR: I'd like to begin by asking you to tell us a little bit about your background, where you started out and the journey that led you to Microsoft.

JRJ: Sure. As you might get from my accent, I'm actually from the UK. My journey is not particularly orthodox. I actually trained in the legal field. I have a law degree and spent the early part of my career in that field. Around 1998, I started in IT as one of many people working in Y2K. My first job was with a company called Logica, and I started working on mainframes at Ford Motor Company in Europe where I learned a lot of my skills in the trade from colleagues who were working on very large databases and ecosystems at Ford. I spent about 5 years there before moving on to look at startups. I worked at a company call Cheapflights.com, now owned by Kayak, doing a real-time flight price comparison website and worked to build a lot of the analytics behind that as well as some of the core engine performance tuning.

But for me, a real seismic event was the 2005 launch of SQL server. That was a big thing for me because that's when I first met a chap named Simon Sabin, and he really introduced me to the community around SQL and this SQL ecosystem in a much broader way. I had a pretty deep knowledge of the product by then, but I was much more internally focused. So from that period until about 2008, I started doing more work with the community, and in 2008 I got my MPP (massively parallel processing) from Microsoft. I started publishing books on SQL and worked with a guy called Christian Bolton, and we published the first 2008 *Internals and Troubleshooting* book, which was very well received at the time.

This is when I really started to grow my influence with the community in and around SQL. One of the biggest achievements I have in that timescale is I started to work on the not-for-profit SQLBits Conference. Simon was one of the founders of that event, and I was asked to help organize it and started working in the partnership ecosystem and sponsorship of events and what I would largely consider to be fundraising for not-for-profit events that really focus on real-world education for customers. In my time doing SQLBits, which was up to about the end of last year, we had raised over a million dollars in sponsorship funding to put on events, training hundreds of people, thousands of hours—all the content is available on SQLBits.com for free with picture-in-picture video and stuff like that. So there were some really great outcomes from that experience. I spent an enormous amount of time in the community, and I sat on the board of directors for PASS as well, helping to grow the ecosystem globally in a number of different ways.

But at the same time I was looking at myself and how to differentiate my experience, and in about 2010, I started spending a disproportionate amount of my time on scale-out data warehousing—having really cut my teeth on the Fast Track Data Warehouse, the reference architectures, starting getting into

Parallel Data Warehouse then APS—and launched my own company called the Big Bang Data Company that really focused on this specific area. Really, through my community contributions and deep knowledge of MPP technologies, I kind of hit the radar of a number of people here at Microsoft. I was contributing to the product, I was writing IP for the product, I was doing a whole bunch of things, and in the end an offer came that was too good to be true to join the team itself, and at the end of 2014, I actually joined the team.

MR: That's awesome. I love the beginning of your journey because, I have to confess, I'm a recovering lawyer as well. So we both started in the same place and ended up in technology. SQLBits—I digress here a little bit—is that an annual event that still goes on?

JRJ: Yeah, the next one is actually coming up on Feb. 21 -24 in the UK. It's the second largest event in the world for SQL-based technology, and yeah, it's going to be a great event. I'm going to be doing a precon there with a partner company in the UK all around SQL data warehouse, and really helping to share and educate folks in the UK and all over the world. It really has grown—last time I looked, I think something like 10% of the attendees are international, at least. So it's pretty cool. I'm hoping we'll see up to about 3000 people.

MR: That's awesome. Now based on the journey that you've travelled, where have you seen the most significant changes in the data warehouse and BI space over, let's say, the last five years?

JRJ: In the last five years, I think it's pretty clear that it's actually the cloud that has really kind of transformed how people think about data warehousing, where they want their data warehouse to be, and what it really means for a cloud data warehouse architecture. The cloud allows you to do things both differently and more cost effectively than you can do in an on-premise. It changes how you think about budgeting and building a solution. For example, if you think about building a data warehouse on premise, you're thinking about sizing—how much data do I have—and that's one dimension. Then how many cores do I need, and you make a guess. Then you have to put some kind of growth projection on that because the accounting team are going to want to keep that asset around for 3-5 years. And then you want some contingency around that just in case you've got it wrong. So you end up with, what's the max I need, plus the max growth, plus the max contingency, and they end up having to put a big number in front of this solution before it's even been built. So these projects historically carry a huge weight, not only of expectation, but also funding as it's a massive strategic investment.

Contrast that to the cloud—I can spin up, for example, SQL Data Warehouse in five minutes or less, I can have as much compute as I want, when I want it, and then have as little as I need when I don't use it—or I can even turn the whole thing off. Now that is a transformational way of thinking about how you cost and move an analytical solution forward for a customer. And the cloud doesn't stop there. It's enabled us to do things with a product that you just aren't able to do on premise. Legacy data warehouses today are very much fixed in the mindset of ratios of compute and storage, which are coupled together, so typically you overpay in one of those dimensions and sometimes in both. You need more compute than you have storage or you have more storage than you have compute. The more

progressive cloud data warehouses are separating those two things out. So it's more a concept of, you store the amount of data that you need to have and then you apply the amount of compute, i.e., you just pay for the amount of performance that you need, and at the level and the time that you need it. Pricing is per second or per hour depending on your platform of choice, and that's a huge differentiator to enabling projects—you couldn't even begin to have that conversation five years ago.

MR: What are the main pain points you're now seeing with regard to data warehousing and ETL?

JRJ: I think that depends on whether you're looking at a new system or existing systems, and how customers want to go on their journey. I think that's a big differentiator. If you've already got a solution, the key questions are: what do I need to be able to do to leverage my existing assets, and what does that actually look like. I can tell you that there's a very clear and common situation with those customers—they are much less inclined to do what I consider a pure "lift and shift" of their solution. They're looking at the cloud and they can see the cloud economics, and they're looking to take advantage of other components and services available in the cloud. They're not just picking up that solution like an appliance-type model and saying, ok pick this up, plop it in the cloud, and we'll call it good. Customers are aspiring for more, but they need guidance in terms of how best to achieve that. So there's a lot of conversations about how to do that, what does that journey look like, how do you break that down into steps, etc. That's one type of customer profile.

Then you have the new customer and the "cloud won't work loads" and they are different in as much as they don't have the necessary metadata catalogue to understand those things, and then it's just a question of which services do I use and when, and how best to bring all those components together. And I think we need to do a better job of accelerating that part for customers.

MR: Based on that comment, do you think current data warehousing tools are adequately addressing these pain points that you raise?

JRJ: I think there are some that have very interesting capabilities around the knowledge they have accumulated through mappings of ETL—what the sources are, what the targets are, what the transformations are—so there's a level of very large accrued metadata in that space. But you stretch that like an on-premise into the cloud, there's a lot of change that's going on in that space. I work quite closely with companies like Informatica that are doing a lot to make sure they're able to connect to a lot of these services and to be able to provide either cloud bursting solutions or cloud migration solutions. But I think that overall we'll see more customers looking at simplification of their solutions and to make as much as they can out of self-service in the same way you see in business intelligence—that is, the front-end piece they want the data analyst to be doing. We're seeing more and more self-service coming into the data integration area, data wrangling tools, etc, but I've yet to see something that's completely blown me away in that area that makes that whole thing, end-to-end, a self-service.

MR: I've read in recent years some articles by pundits heralding the death of the data warehouse. What's your take on that?

JRJ: The first thing I say when people talk about the data warehouse is, it's a theme, it's not a product. Themes generally don't die, and neither does the driving force behind those themes. No one in their right mind goes out and says, today I'm going to build a data warehouse. They're looking to solve certain types of problems and they're looking to solve answers to questions—and to be able to do that in a repeatable way. That's a very different way to mentally frame the question in the first place, right? Because if you think about what data warehousing actually is, it's trying to make sense out of disparate data, to be able to curate that data, and to be able to provide that data in a form that can serve the enterprise or a business and to be able to share that and provide good, quality, clean data throughout the enterprise. None of that changes, whether you're using big data technologies and trying to do it in that space, or whether you're trying to use a SQL-based engine to try to deliver that type of experience.

What we have actually seen is a broad recognition that SQL as a language is actually a very, very good language for enabling very large bodies of users to consume data. It's a very easy, very intuitive language; if we look at how many people understand it conceptually, it's a very large body of people. So the management of change when you have a SQL-based language is much less. You can even see it in the big data space. The number of big data technologies that now espouse SQL as a capability. What those solutions have is SQL as a language—what they don't necessarily have is the many millions of hours of man-years in creating SQL as an engine to back that capability. And as the queries become more complex and the insights people are trying to get out of the data become more involved, that's where a SQL engine is much, much better suited to be able to answer the question, to be able to run that type of processing.

Both SQL engines and SQL as a language I can see a very rich and very vibrant future for. What I think will be even more interesting is how those technologies like SQL Data Warehouse will actually integrate into that ecosystem so that the data that is shared into the enterprise is accessible. That's a very interesting area to discuss and pursue.

MR: What's your take on the role automation can play in the making, design and build of data warehouses in a way that's faster and more cost-effective, especially when we look at the exponential increase in data flows and data complexity?

JRJ: I think this is an area that offers a rich vein of opportunity. As I said earlier, I don't think this has been knocked out of the park by anybody yet. I can see that there's a lot of investment in automation. One of the things I've seen at various companies is where you have very simple transformations required but across a large number of sources, but you don't want to spend a large amount of money basically taking those sources and mapping them out—if you can do that through cogeneration that's obviously a huge boon for everybody. You have technology stacks like Varigence and BIML, for example, in the SQL space, and you can see that integration coming through in Azure Data Factory. With those APIs and being able to build that stuff programmatically, cogeneration is a very attainable goal has been on the rise for some time. It's still at the point where you still need a data engineer rather than a data

analyst to build that stuff, but I can see a point where over time that will become more of a commoditized skill.

MR: Let's talk about dealing with businesses and different C-suites within those businesses. What we see at Dimodelo is a real knowledge gap between the IT department that really gets this stuff and the C-suite that has to be convinced to make the budgets available and to push the project forward. Is there one resource you could recommend to executives and middle management to help them on their BI journey?

JRJ: I still come from the data integration side of the house much more than the data visualization space, or even from the business intelligence side. I will say that that resources I use and reference—Chris webb's BI blog or the fine chaps at SQLBI, Marco Russo and Alberto Ferrari—I find that they have a very interesting perspective in terms of what the developments are in BI. They are generally good resources, and I know they've collaborated with Rafal on Project Boticelli to build videos and stuff like that, so there's educational material and their books are very well respected. Generally speaking, I find that the C-suite/exec level people have a much more simplified view of what they're actually trying to achieve, but not necessarily the understanding on how to get there. So the tools that are able to visualize that for them as quickly as possible are the ones that get the most traction.

MR: That's a great way of putting it. What I'd love is if you could send us the links for these resources you've mentioned, and I'll share them in the blog post when we create a link to this podcast, if that's ok.

JRJ: Yeah, no problem.

MR: So now let's switch gears to the segment of the show where we get to know our guests a little bit better, so some lighter fare. The first question: what is your current favourite TV show that you go to to binge or to relax?

JRJ: I'm not a huge TV watcher, but I'm now watching series 1 of American Gods. We're on episode 6 of 8. I'm not sure I can tell you what's going on, as I think the whole point of the show is that you actually have no idea what's going on right up until the very end. Don't ask me what it all means because I'm still struggling to find that out myself. I'm more of a comic book guy, and the fact that this is a Neil Gaiman show is the attraction for me. I'm a huge fan, so just seeing that all come to life is interesting.

MR: Well, you know, my kids are big fans. I haven't watched it yet—it might be too scary for me, but they rave about it.

JRJ: It makes Game of Thrones look tame on the blood front.

MR: Wow, that's a real marker there, wow. On to the next question: what's the best non-tech book that you've read in the past few months that you would recommend or give to your friends?

JRJ: Well, this one would depend very much on the friend, I would say. The book that I have most enjoyed reading and learned the most from, actually, is *Monetizing Innovation*, which is very much about product development and how different product methodologies and different companies have succeeded and failed with their way of bringing a product to market. They look at various case studies, and they're also looking at it very much from a point of view of value and what aspects do customers value, and how you actually think about valuing innovation and features and where they actually sit. So it's not actually non-technical, but for my role it's been a big area of growth for me over the last two years as I've come into a product management discipline after coming out of consultancy. This book has really spoken to me in a way that I would recommend it or share it across my peers to read.

MR: Who is the author?

JRJ: Ah, I knew you were going to say that.

MR: Don't worry, I'll find it and add a link to it in the blog post because if you got something out of it, I'm sure others would as well.

JRJ: Yeah, it's very interesting—I would encourage anyone to read it, and it's not expensive at all.

MR: So now to conclude the podcast—first of all I much appreciate your time—and I'd like to find out where our listeners can find out more about you on the web.

JRJ: I put a lot of my energy into our product documentation, so you'll see my ugly little face in the little circles on many of our product docs. If I'm writing blogs, I'll write them on the Azure blog, and any videos go up on Channel 9. I try to make as much of what I'm doing as public as possible such that it's available to everybody in that way and has the most impact on the thing I care passionately about, and that's the success of Azure SQL Data Warehouse. I really want to invest my exposure into that.

MR: Obviously we can find you on LinkedIn as well since you and I are linked up. What's your Twitter handle?

JRJ: Just @jrowlandjones.

MR: That's great. I'll share that as well.

JRJ: Most people just know me as JRJ . . .

MR: I know, I thought that might be your Twitter handle.

JRJ: Several million people beat me to that. Yeah, unfortunately it had been taken.

MR: We'll include yours in the notes as well. Thanks so much for your time today James. Much appreciated, and I hope we get a chance to do this again.

JRJ: Yeah, I look forward to it. Thanks very much.

MR: Have a great one, thank you.