

MR: Mike Rabinovici here and this is Dimodelo’s “Conversations with Data Warehouse Experts” podcast series. Our mission in these podcasts is to speak to the best minds in the data warehouse and BI space, and to get their take on the state of the business and find out what they think the future holds. We are also committed to do it in a way that adds value to tech professionals, while at the same time is clear and compelling for the business managers and executives they collaborate with every day.

In today's episode, it is my pleasure to speak to Dr. Claudia Imhof, a real rock star in our corner of the technology universe. She not only has the technical chops, but also the in-depth understanding of the business landscape. In 1992, she founded her consulting company, Intelligent Solutions Inc. And in 2006, she founded the Boulder BI Brain Trust, which has gained a reputation as being one of the finest collections of smarts in the industry. Welcome Claudia!

CI: Well, thank you Michael, for a lovely introduction.

MR: That's how we roll at Dimodelo. So, I want to start by asking you a little bit about your background for our listeners who have not had a chance to work with you.

CI: You covered it. I started my company in 1992, so that means I have, oh gosh, 25 years of building and designing and everything about data warehouses. In recent years, my work has focused more on the vendors themselves—looking at where they fit into the architectures. I've written 5 books, more than 150 articles, and again, my focus has been on the architecture supporting business intelligence and analytics—all the infrastructure that it takes to make a good, solid, modern BI and analytics environment.

MR: Since you have such a good overview of the industry from both the tech and business end, where have you seen the most significant changes in the data warehouse space over the last, let's say, 5 years?

CI: The last five years have been an incredible period of innovation. And unfortunately, with tremendous innovation comes a lot of disruption. That's certainly shaken up everybody's ideal architecture. But let me start with what I see as some of the changes. The first one, of course, has got to be a focus on the different types of data that we now can use for analytics. It used to be all structured data, and mostly internally-derived data coming out of our operational systems. But that has gone bye-bye! We certainly do still need the operational data, no doubt about that—it tells us what we're doing, but now we have the access to all sorts of external data and it's not well-behaved operational data or structured data. It's pretty misbehaving, if you will—variably structured, odd formats, odd sources, and that type of thing. So that's the first one.

The second one is that data storage itself has undergone dramatic changes. With the advent of Hadoop and Spark and so forth, analytic appliances, analytic databases, we now have tremendous performance and storage capabilities for all of that data that I just talked about.

And the third one are the different deployment options that we have today. We have to look at the cloud. The cloud has become a major deployment factor. Do we go off-premises or do we keep everything on-premises?

And then the last one are the advances in the analytics themselves. Obviously, data visualization was a big trend 5, 6, 7 years ago, but we also now have, of course, the advent of the much-vaulted data scientist and all of the wizardry that goes behind the advanced analytics that are now capable of being performed.

So those are the major trends, changes I've seen, and they've had a dramatic impact on our business.

MR: One of the topics that I see discussed almost exclusively is the quantity of data—that there's so much of it. What I want to ask you to address is, do you think current data warehouse products are adequately addressing the increasing complexity—not just the quantity—of data, as well as the exponential increase in data flows?

CI: The answer is yes and no. I hate to equivocate that way, but yes and no. First of all, as I mentioned, data storage and the performance of analytics—absolutely yes. They have kept up with tremendous performance and tremendous ability to store now terabytes, petabytes, you name it, all kinds of data. Where the no comes in is really around the administration and the governance of that data. It's still new to a lot of companies, and therefore they don't quite know how to manage it, they don't know how to administer it. And in particular, the governance side is very questionable. How do I govern massive amounts of data? And that leads to another problem which is, of course, if you can't manage it, then maintaining it is very difficult. The adage of, gee, let's just throw all the data into the data lake and somebody else will figure it out is a road to failure. That is not the way we want to do this. We do need to design it, we do need to understand it, and we do need to govern it.

MR: Some pundits over the last few years have heralded, in their own words, the death of the data warehouse. What's your take on that?

CI: Really? That's a fallacy, of course. The data warehouse is anything but dead. Its definition has certainly become more crystallized. If people are interested, I wrote a paper for the Data Warehousing Institute on the extended data warehouse architecture. Let me just spend a minute or two on what that is. My focus, as I said, has always been on architectures. The extended data warehouse architecture takes into account this disruption to the traditional architectures—whether it's a corporation information factory or a bus architecture—and recognizes that analytics are now found in three different areas.

The first one, of course, is the traditional EDW—enterprise data warehouse. Its definition has crystallized though to be a production analytics environment. Don't throw the data warehouse out. It is a source of wonderful, certified, trusted, reliable data, and it's the production environment. Where else can we produce trusted analytics—the analytics that we run our companies on—we have to believe that

they are created from trusted data. Things like key performance indicators, all of the customer analytics, profitability analytics, and those used for forecasting. So the EDW is a wonderful place for those production analytics.

But we also have another need, and that is the exploration or experimental side of the house. That's where a lot of the new technologies focus their attention. Things like Hadoop and columnar storage, in memory, the analytic engines or the analytic appliances—all of these really work well in this experimental area. And that's where we do these general unplanned queries. That is something that is just impossible to do in a well-planned data warehouse. So now we've got a place to do that.

But there's also a third area for analytics, and that's in operations, if you will. It's real-time analytics on real-time data; in other words, streaming analytics. We are analyzing data as it comes into our organization before we even store it. Both the investigative computing platform—that experimental area—and the data warehouse... you have to store the data and then analyze it. But in streaming analytics, we're going to analyze it as it's coming in and then we may or may not store that information.

So what's happened is this new architecture, if you will, does include a data warehouse—it is anything but dead. It certainly does take that into account. Don't throw it out. It is certainly a good place for these production-type analytics, but it doesn't mean that we don't have other areas where analytics take place. Now, the good news is the new technologies have helped speed up the data warehouse. We certainly want to bring some of those technologies to bear on some of the trusted analytics. They do certainly work in the world of investigative computing platforms. And some even help us out with the streaming analytics. So, good news. We've now got technologies in all those areas.

MR: A moment ago you spoke briefly about the cloud and I want to come back to that. As more and more companies are adopting the cloud, what do you think will be the lasting impact on data warehouse projects because of that move?

CI: Well, there are some challenges and some benefits. Let me just go through them quickly. The first challenge I see is the IT governance and control. A lot of IT departments are leery of letting go of control over their environments. You know, putting the data in the cloud means you no longer have strict control over it. So there are concerns about security and privacy, and that ranks pretty high on their lists.

Another challenge is actually a change to the IT workflows. These workflows that deal with compliance and security are certainly more complicated in these hybrid environments, and I do think it's going to be hybrid well into the future—hybrid meaning that some data is on-premises and some data is in the cloud.

And then the third one is interesting. It seems like one cloud instance isn't good enough, and now we have multiple cloud deployments. We may have some data in private clouds, some in public clouds. So now we have to handle the flow of data into these multiple cloud deployments.

And then of course managing costs is another one. The on-demand and scalable nature of cloud computing services can make it pretty hard to determine and predict all of the costs. And of course different cloud computing companies have different cost plans. Some may charge by data volume, others by number of active users, and others by cluster size, for example. So those are some of the challenges.

Now, some of the benefits of the cloud... the first and foremost is the lowered operating costs. This is a capex vs. an opex determination. When you are renting your infrastructure, there's a monthly or yearly spend, but you're not paying for everything up front like you do in an on-premises environment. So that's the first benefit you see.

Then, of course, the company doesn't have to worry about maintenance or upgrade hassles, or all of that kind of stuff. The company offering the software and hardware has to keep up with that.

Ease of implementation is another one. Boy, you could get a cloud deployment just by swiping your credit card, right? And a lot of companies do that.

And then, of course, the best thing, I think, about the whole cloud deployment is its elastic scalability, I guess you'd call it. You can quickly scale up—and down—depending on your needs.

So those are just some of the things I see that are pretty important about moving to the cloud. And their impact on data warehouse projects—certainly it is easier to spool up an instance and get data into it than it is to build your own, and that's probably the biggest impact on data warehousing.

MR: In addition to your consulting practice, you're also the founder of the Boulder BI Brain Trust. What led you to start it in 2006?

CI: What led me to creating it was kind of a frustration. I'm an independent analyst. I don't work for a big analyst firm. I am pretty much a sole proprietorship. And if you are an independent analyst, or consultant, or practitioner, you don't get a lot of the vendors' attention unless you reach a certain level of, shall we say, notoriety.

MR: I was going to use that word, yes 😊.

CI: I'm fortunate that I reached that level of notoriety and the vendors began to invite me to their analyst events. But so many of my independent analyst buddies and practitioner buddies weren't invited. So I created the BBBT to give them access to vendor briefings. It started out in 2006 with about 6 of us. I invited the vendor to Boulder and I said: look, do a deep dive for my friends here in this local area, and let's spend about 3 hours going through all your technology and even some NDA stuff at the end of it. And from those humble beginnings—I think we did maybe 4 or 5 that year—the word got out, and everyone started saying: well, wait a minute. I'm independent. I want access too. We're now well

over 220 members representing I think around 25 different countries. So that's pretty remarkable right there. We have done, in the past, about 37 events in a year. We're cutting that back because it's just way too many for us to handle. So we're now down to around 22, 24, something like that.

If people are interested, I do hope that they will either become a member—if you're an independent analyst or practitioner—and independent means you work either for a small company, and certainly don't take payments from vendors for selling their software. That would not be an appropriate usage. Or if you're not an independent person like that, then you can become a subscriber and you get to watch the videos that we produce from each and every event, or hear a podcast, or watch a 5-minute trailer.

So it was more out of frustration that caused me to form it but, boy, I must have hit a nerve because there's a lot of demand.

MR: Well, it sure looks like it. I'm a subscriber. But, just for our listeners, we will have links to Claudia's contact information and website in our show notes, so everyone will be able to see it and go ahead and subscribe to or join the BBBT.

Now, I want to switch gears to an area that's closer to our hearts: automation. How do you think data warehouse automation is changing the delivery of information to the business side of the equation?

CI: Excellent question. I really like data warehouse automation. One of the things that has hit every IT shop throughout the world, and every business, is let's do more with less. We have to do more, we have to be more productive, we have to cost less, and yet still get the same quality of product out the door. So, to me, the first thing that data warehouse automation brings to the table is the ability to create an environment for less money, make it cheaper. The reason for that is that it makes the implementers much more productive, much more efficient. Reusability has always been something that we've said you must have in a data warehouse environment. Well, data warehouse automation ensures reusability by making sure that we do have certain standards, certain configurations that we use over and over.

I think the second advantage is we have to do more with less, but we also have to do it faster. Faster, faster, faster, cheaper, cheaper. And the faster part is the agility that we get by using a data warehouse automation technology. Far more agility in the team itself. We can create prototypes faster. We can change prototypes faster. We can change existing production things faster. So I think that's the second big advantage.

Then the third one: data warehouse or business intelligence analytics environments are here for the long haul. They are constantly changing. They have to be. That's the sign of a good, robust, and healthy analytics environment. But changes are very difficult if you don't have good documentation, if you don't have a good way to pick up the pieces and change them and put them back together again and put it into production. So I think the third thing that data warehouse automation gives us is that robust documentation that allows us to manage the environment far better to manage and sustain it into the

future. So good on you, automation people! That's a tremendous boost to anybody that's in this environment.

MR: Thank you, and we are seeing that pain point out there, so it was interesting to get your take on it. I want to also talk about the business side. In our projects at Dimodelo, we often find that the technical folks are quite knowledgeable. It's on the business side that we often find the knowledge gap. And because of the pressure to get good, cost effective, almost real time BI these days, the executives and business managers we deal with find themselves in a situation where they have to integrate a lot of knowledge very quickly. If there was a resource or resources that you could recommend to help them on their BI journey, what would it be?

CI: Yes, we all have the adult onset ADD. So I guess... where to start? First off, I would probably look into TED Talks. Normally they're about 15 minutes, they're very, very good talks by incredible speakers. So if you want to get a taste of what analytics can do for your company, what it's all about, and so forth, I suggest you start with a TED Talk. Now that's going to be very high level and very much not into the details that most people might want. If you do get an interest in it and want to go into more depth, then I would say there are any number of online courses—linda.com, coursera.com—go to those free online courses and dive into it a little more deeply. And then the last suggestion, I don't know where people live—if you live in a city that's got a university, I can just about guarantee that that university has some kind of analytics course. And again, nice, good, practical examples of analytics and how you can use them in your everyday life, generally in most university courses.

MR: Thank you. Although I'm a big TED consumer, I never even thought of that. What I especially love about TED Talks is that, as you said, they're short and sweet, and really packed with knowledge. So I appreciate you bringing that up.

Now I want to move onto the second part of the podcast where we try to get to know our guests a little bit better with some totally unrelated non-tech questions.

CI: This will be difficult!

MR: First one. What is your current go-to TV show that you're either watching or binging on, as the case may be?

CI: Oh, good question. You know, I love *Madam Secretary*. I know it's a political show, but it's a feel-good political show that shows a side of our government perhaps that can do some good in the world. I love the lady, Tia Leone, who stars as Madam Secretary. So that's my go-to show, Sunday nights!

MR: I'm a fan as well, and if you loved the *West Wing* and you have not watched *Madam Secretary*, you will love *Madam Secretary*. One other question, before we conclude. Which non-tech book have you recommended or given away as a gift most often?

CI: Wow, another good one. I'm actually reading it right now so I haven't given it away, but, boy, if the opportunity arises, I certainly will. It's a book called *Salt*, by Mark Kurlansky. It's the world history of salt—the salt on your table—and how it has shaped civilization throughout the millennia. Believe it or not, it's an absolutely fascinating book. So I'd recommend it to anyone and if I get a chance, I'll give it away.

MR: Thank you, that's a great recommendation. I'm going to add it to my list. Last but not least, where can our audience find out more about you and the work you do?

CI: Probably the best place is to go to the BBBT website, and that's bbbt.us, and you can see all the videos we've produced, you can see the vendors that have come, you can read a little bit about me and any of our members for that matter. That's my pride and joy. I'm so happy the BBBT has become such a successful entity. By the way, being a subscriber or member is free—I forgot to mention that—so if that entices anyone, I hope it does, go to BBBT and click on the "Get Involved" tab. It's really easy if you'd like to either be a member or subscriber, and off you go.

MR: That's awesome. Having joined it as a subscriber myself, in maybe the last month, I can definitely tell folks out there listening that the content is real, top-notch quality. So thank you for putting that out there.

CI: Thank you! I appreciate that.

MR: So this concludes our podcast today with Dr. Claudia Imhof. Thank you so much for your time today and I think there are so many more questions I would love to ask, so maybe we can do a second round at some point.

CI: I'd love to. That would be wonderful. Thank you Michael, and thank you Dimodelo. I appreciate the time.

MR: Thank you and for our listeners out there, you'll be able to find a replay of this podcast at www.dimodelo.com, and as always, we'll be sending out a link to the podcast with a whole bunch of the notes for the resources people have mentioned. Thanks again and until next time.