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 the tech profession, 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 welcome Hans Michiels, a highly respected business intelligence and data warehouse consultant based in Utrecht in the Netherlands. He's also founder and moderator of datawarehouseautomation.guide which we will be talking about in this podcast. Welcome Hans.

Hans: Thank you Mike.

MR: I'd like to start by asking you to tell us a little bit about your background, for listeners who have not had the chance to hear you speak, work with you or come across the datawarehouseautomation.guide.

Hans: My name is Hans Michiels. I'm an independent IT consultant since 1998. I work and live in the Netherlands, and I work in the area of business intelligence, more precisely Microsoft stack using accelerators or data warehouse automation tools like Dimodelo Architect when I can—it depends on the customer. I have a number of certifications on Microsoft SQL Server and Data Vault modelling, and as you already mentioned, I'm founder of the website datawarehouseautomation.guide, a website that provides comparisons between data warehouse automation tools and tells you where to get more information on them. It also has a short domain name which is dwa.guide. That's about it in a nutshell.

MR: In the past five years, where have you seen the most significant changes in the data warehousing space?

Hans: The most significant changes—I think there are two important changes, first, broader acceptance of modelling techniques such as Data Vault and to a lesser extent anchor modelling. Data Vault especially has become an accepted mainstream system for data warehousing and business intelligence. Another significant change is in the way new technologies arose or have grown bigger, for example, cloud technologies such as Amazon Web Services, Microsoft Azure, and PowerBI.com. And also data-based technologies in the cloud such as Snowflake Data Warehouse. Also, in the last year there's been a lot of talk about big data; I think it's a bit of a buzz word and a bit overrated. Still, technologies like Spark and Hadoop can be used in certain cases where there are huge volumes of data.

MR: Do you think that the current slew of data warehouse products are adequately addressing the increasing complexity of data, and, I should add, the exponential increase in data flows and sources?

Hans: It's a difficult question. For increasing complexity there's no simple answer. It just means you have to do more effort to gain useful insights from the source data. You also mentioned exponential increase in data flows—if you mean increasing numbers of sources, I think that classic ETL tools will fail big-time when new sources are added to data warehouses. It's all manual labour to add all those sources, and it

just takes too much time and is too expensive for most companies. Apart from that, the time to consumption of the data will be far too long. Business users need insights from new data sources quickly. You should be able to absorb a new source in a few weeks to a few months, but with classic ETL tools this will take much longer. There's no need to think that there is no other way because with data warehouse automation you can do all this much faster.

MR: Before we get into the automation aspect of it, I want you to comment on some authors that have put out articles saying that the data warehouse is dead. What's your take on that?

Hans: I disagree with that. I don't think the data warehouse is dead or will be dead within the next ten years or so. A lot of small- and medium-sized enterprises still use CRM, ERP and other systems that still contain data, and a data warehouse is a good way to get insights from those systems. There are also reporting tools that work directly on the source data or on an exact copy of that, but I think those tools will have a difficult time when the number of sources grows and the complexity increases.

I remember when I was an independent consultant working for a consulting firm in the Netherlands, Capgemini, and we already made reports using Microsoft Access and Crystal Reports on source systems, and we found out it was not the best way. It had a performance impact on operational systems and it didn't work for the data we needed integrated. So that's why data warehouses became more popular. People today seem to have forgotten that, or don't know it at all, and continue to do things directly on source systems—and history repeats itself.

In the end, for structured data, I think the data warehouse still has added value and that it will change in the near future, Mike.

MR: Speaking of areas that have been written about a lot, one of them is the cloud. As more and more companies adopt it, what do you think will be the impact on their data warehousing projects?

Hans: In the technical sense, I don't think the impact will necessarily be that big. Instead of storing the data on premises on the server, you just send it to Amazon or any other cloud storage. One impact I can think of is the issue of how to transfer the data fast. One solution is to use "Change Data Capture" so you only send the inserted, modified data, and the deleted data if you want to detect deletion. You can also make sure that between sources at the warehouse there is a high bandwidth internet connection, for example using optical fibre technologies. So you can imagine if you have dashboards and mobile devices, it's an advantage if data is stored in the cloud, but with products like Power BI Gateway, you can achieve almost the same result while having your data on an on-premise server. So the cloud is a natural evolution for BI. It works and the impact is reasonable.

But perhaps you didn't keep the legislation aspect in mind when asking the question, because in the EU we have a new law, the General Data Protection Regulation, that will come into effect in May 2018. It will impact what you can do when storing customer data. I'm not sure how it applies if you have a server

in the US, but people should be careful to find out how this might apply to you if you're storing data in the cloud.

MR: And this regulation applies to data stored in the cloud within the EU?

Hans: I believe so. I'm not sure if it's applicable to citizens in the EU if the storage is physically outside the EU, but again that's something that's important for companies to find out if you're using cloud technologies.

MR: That's a very important point and we'll be sure to add a link about that in the show notes [https://www.eugdpr.org/]. That's a very important piece of legislation to pay attention to. Getting back to data warehousing, in addition to your consulting practice, you are also the founder of datawarehouseautomation.guide. How did it come about?

Hans: As a consultant, I was doing projects with Microsoft technologies and some (of my) own scripting, but I was looking for a better way. I knew there were a number of those tools, but I didn't know them by name and I didn't know them all. So I started to do serious investigation just for my own use to find out which tools are available and what are the features, but I couldn't find the information on the internet, except vendor websites that were more like marketing brochures and not really useful. So I started to contact all the vendors and send them a questionnaire, and that gave me a good overview of the capabilities of all the data warehouse automation tools around so I could give good advice to my customers. Then I thought, I have all this information, I shouldn't keep it to myself but instead share it with the business intelligence community. But it was too much for a blog post—fifty questions and 30 tools—so instead I put it on a website which has become dwa.guide. And that's how it went.

MR: In terms of the impact of automation, how do you think it's affecting the delivery of information to businesses?

Hans: I think it will have a big impact because with data warehouse automation, you will be able to deliver information to the business much faster. To compare with the automotive industry, there are no manufacturers that don't use automation in their manufacturing plants because using only manual labour, the car would be much too expensive to make and you wouldn't sell it. But in BI, we still do the work manually, but if we would automate some of that effort, things would go much more efficiently. And it's been proven that with data warehouse automation, a productivity gain of 50% or more is easily feasible—in some cases up to 80%, depending on the complexity of your source systems. You can imagine what that means in terms of human resource costs. The advantage is huge, so the license cost for a data warehouse automation tool is earned back very quickly. It's almost a no-brainer. But in addition to the cost gains, the increased speed in delivering information to the business is much faster. It's a win – win, yet companies are hesitant to apply it. I think it's just awareness that's missing.

MR: In our projects at Dimodelo, we often find that technical teams are quite knowledgeable when it comes to this; it's on the business side that we find a knowledge gap. Based on your experience, where

could executive and business people go to educate themselves a little more or to integrate the knowledge they have more quickly so they can work more smoothly with their tech teams? It could be a book, a resource, or just a piece of advice you could share with them.

Hans: This is a difficult question because there is no single source that has all the answers. You can find some interesting videos on YouTube looking at data warehouse concepts. If I may give some advice, one thing that is important to understand for executives is dimension effects. Dimension effects, introduced by Ralph Kimball, are the cornerstone of the modern data warehouse, at least on the presentation side of things. And if you understand them, it will be easier to communicate with the technical guys because you will be speaking the same language.

It's also important to understand the concept of data integration, data cleansing and master data management. Data integration means that data from different systems is integrated so that you can have, for example, a merged list of customers. It's important to have a data warehouse that loads and integrates the data and prepares it to be useful in reports and dashboards. As I said earlier, if you do things directly on source systems, some things are very difficult or just not possible, and the data warehouse is a solution for that.

Last but not least, it's important to understand that the data warehouse is not what you see. Often people refer to a data warehouse when they look at a dashboard or a report, but these are only the outcomes of the systems behind it—the tip of the iceberg. Sometimes the team has to do work on the system that you don't really see at the front end on the report or the dashboard, but work still needs to be done and sometimes that is not well understood by managers.

MR: That is an excellent point Hans. We see when we speak to C-suites that their focus is on the dashboards, as it should be, but I think bringing to the table the understanding that dashboards need to be, not just be pretty, but accurate, and they are really dependent on what underlies them. That's where the data warehouse is playing a key role.

Hans: I have also in my experience seen systems where the managers were very keen on having quick answers in the dashboard, but you have to realize that if you put too much pressure on the team, they may think they have to do a shortcut. As a manager you should make sure the structure is good for maintenance and long-term benefits.

MR: I'd like to switch gears now for something completely non-technical, so the audience can get to know our guest a little bit better. So my first question to you Hans, if you're ready, is: what is your go-to TV show that you like to watch or binge on, as the case may be?

Hans: I like to watch Netflix because the video-on-demand concept lets you watch things when you have the time. My problem with that is that when I start watching a series, I always want to know how it ends. I'm currently watching *Bloodline*, it's a three-season show about a very upper-class family with a lot of problems that are getting worse and worse. It's all about hypocrisy.

MR: If any of us are ever in Utrecht, what is the one restaurant we absolutely have to visit?

Hans: Well, my wife and I like to go to a sushi restaurant which is really good. It's always full and it's very good. It's not in the city of Utrecht, but south of Utrecht in Vianen. It's called ShiZen and it's really nice. Good food, quick service, excellent.

MR: My final question Hans, where can the audience find out more about you, the work you do and the datawarehouseautomation.guide, if you could mention that URL again?

Hans: First my personal website is hansmichiels.com. It contains my blog and some other information such as my LinkedIn and Twitter details. For datawarehouseautomation.guide, the short URL is dwa.guide. I also blog there now and then, and you can subscribe there to an upcoming newsletter. Last but not least, the Twitter account for that website is dwaguide, and I tweet there about all the news on data warehouse automation tools. Also, when there is a new blog post on the website, you will find a tweet there.

MR: And where can folks sign up for the newsletter?—on which site?

Hans: That's on dwa.guide.

MR: Well thank you so much for you time today, Hans, it was very informative and hopefully we can get to do a second round at some point.

Hans: Yes, it was nice for me too, Mike, thanks for inviting me for the interview.

MR: My pleasure, and thanks for listening out there. We'll upload the podcast to our blog, as well as some notes from the show regarding the resources and URLs Hans mentioned. Thanks for listening—'til next time.