

Why Global Data Hubs Are Critical for Your Data Governance Success

Find the transcript to our Q&A with Nisum Content Specialist [Courtney Boone](#) and Project Lead [Vyasraj Vaidya](#) below where they share information about what a data hub is and how it can be used to provide actionable business insights that give companies a competitive advantage. Find the original blog [here](#).

Courtney Boone (00:10):

Thank you for joining us today. Today, we are going to do a quick Q&A session and learn about the global data catalog. We're going to learn what it is and why businesses need it. And today, we have one of our SMEs joining us. Could you give our listeners some background information? What's your name, and what is your role here at Nisum?

Vyasraj Vaidya (00:32):

Hello, my name's Vyasraj Vaidya, and I've been working as a Project Lead for data science and engineering projects at Nisum. I've been in this project for more than 18 months and with Nisum for almost five years.

Courtney Boone (00:46):

Okay. And today's Q&A will cover the topic of modern data catalog. Could you provide us with some information about this topic? What is it, and why do organizations need it?

Vyasraj Vaidya (00:57):

Yeah. A modern data catalog is a repository of metadata that provides data for discovery, observability, and governance. Organizations need the data catalog because it provides improved data quality. It provides improved data context. It reduces the risk of data and provides scope for improved data analysis.

Courtney Boone (01:20):

And how can utilizing a data catalog provide a competitive advantage for an organization?

Vyasraj Vaidya (01:27):

Yeah, the simplest advantage that a data catalog can provide is data security, confidentiality, access control, availability, and integrity of data.

Courtney Boone (01:41):

So we have an understanding of what a data hub is, but how does this differ from a data lake? And in which situations would using a data hub or a data lake be preferred?

Vyasraj Vaidya (01:54):

The difference between a data hub and a data lake is that a data hub is where a company's datasets are stored in a homogenous format with no duplicates. It includes de-duplicated, standardized data from different sources that have been cleaned and enriched with additional value adds, such as standard

definition and consistent nomenclature. A data lake, on the other hand, is a storage repository optimized for storing large amounts of raw data, then making it available for further processing at a later time. A data lake can be useful if you want to store historical data for long periods or if you need to make it accessible to multiple users across locations. However, if you need real-time insights from your data and you want to make those insights actionable, then you'll want to use a data hub instead.

Courtney Boone (02:42):

And could you provide a real-world example of how a data catalog is used?

Vyasraj Vaidya (02:48):

Yeah. The most famous example that I can provide is the LinkedIn data hub project. Yeah, yeah. The LinkedIn data hub project was created for scaling productivity and innovation in data to create a generalized metadata and search discovery tool. *(Reference the video to view Vyasraj's presentation.)* So, the front-end of the LinkedIn data hub platform provides such viewing and editing metadata information interactively. The UI would look something like this where you have, uh, dataset information. You have the origin of data, the platform or the source of data, and owner information. And yeah, this would be the entity relationship diagram or what we call lineage, where users can understand owner information, data set information, user information, membership information group, and profile information. Yeah.

Courtney Boone (04:05):

Okay. Thank you for sharing that with us. And what are some indicators that businesses need to use a data catalog and if the organizations have the architecture to do so?

Vyasraj Vaidya (04:18):

Some indicators that organizations will have to understand if they want to implement a data catalog would be if there's a need for data discovery, if there's a need for understanding the context of data, and if there's a need for elimination of storage of historical snapshots of data with no dependency constraint on other data sources to provide consistency and security.

Courtney Boone (04:48):

And what would you say is the most challenging aspect of managing a global data hub? And what is one way that companies can overcome this challenge?

Vyasraj Vaidya (04:57):

Yeah. There are several challenges in building and managing a data hub. Some of them are managing the fragility in crawling data sources, network connections, firewalls, and configuration. Non-responsiveness or croaking in data ingestion is also another challenge. Non-availability of change logs when something goes wrong, and high dependency on centralized IT teams for supporting consumers downstream are other challenges. We at Nisum help businesses gain an edge with the utilization of data hub. The implementation of a data hub can enable you to capture analytics and data insights, achieve better collaboration across teams, and sharpened decision-making abilities by collecting and utilizing data effectively. Nisum can help you incorporate the use of global data hub into your business practice. By working with us, your business will have access to the data it needs in order to take advantage of market opportunity or better plan for the future.

Courtney Boone (05:55):

Okay. So, thank you for your in-depth analysis about global data hubs, and if you'd like to learn more or are interested in incorporating the use of global data hubs in your business practice, feel free to contact us or visit us online at www.nisum.com. Thank you, and thank you for joining us today.

**The transcript has been edited for clarity.*