**9 investors discuss hurdles, opportunities and the impact of cloud vendors in enterprise data lakes.**

**By Ron Miller, February 13, 2021**

About a decade ago, I remember having a conversation with a friend about big data. At the time, we both agreed that it was the purview of large companies like Facebook, Yahoo and Google, and not something most companies would have to worry about.

As it turned out, we were both wrong. Within a short time, everyone would be dealing with big data. In fact, it turns out that huge amounts of data are the fuel of machine learning applications, something my friend and I didn’t foresee.

Frameworks were already emerging like Hadoop and Spark and concepts like the data warehouses were evolving. This was fine when it involved structured data like credit card info, but data warehouses weren’t designed for unstructured data you needed to build machine learning algorithms, and the concept of the data lake developed as a way to take unprocessed data and store until needed. It wasn’t sitting neatly in shelves in warehouses all labeled and organized, it was more amorphous and raw.

Over time, this idea caught the attention of the cloud vendors like Amazon, Microsoft and Google. What’s more, it caught the attention of investors as companies like Snowflake and Databricks built substantial companies on the data lake concept.

Even as that was happening startup founders began to identify other adjacent problems to attack like moving data into the data lake, cleaning it, processing it and funneling to applications and algorithms that could actually make use of that data. As this was happening, data science advanced outside of academia and became more mainstream inside businesses.

At that point there was a whole new modern ecosystem and when something like that happens, ideas develop, companies are built and investors come. We spoke to nine investors about the data lake idea and why they are so intrigued by it, the role of the cloud companies in this space, how an investor finds new companies in a maturing market and where the opportunities and challenges are in this lucrative area.

To learn about all of this, we queried the following investors:

Caryn Marooney, general partner, Coatue Management

Dharmesh Thakker, general partner, Battery Ventures

Casey Aylward, principal, Costanoa Ventures

Derek Zanutto, general partner, CapitalG

Navin Chaddha, managing director, Mayfield

Jon Lehr, co-founder and general partner, Work-Bench

Peter Wagner, founding partner, Wing Ventures

Nicole Priel, managing director, Ibex Ventures

Ilya Sukhar, partner, Matrix Partners

Where are the opportunities for startups in the data lakes space with players like Snowflake and the cloud infrastructure vendors so firmly established?

Caryn Marooney: The data market is very large, driven by the opportunity to unlock value through digital transformation. Both the data lake and data warehouse architectures will be important over the long term because they solve different needs.

For established companies (think big banks, large brands) with significant existing data infrastructure, moving all their data to a data warehouse can be expensive and time consuming. For these companies, the data lake can be a good solution because it enables optionality and federated queries across data sources.

Dharmesh Thakker: Databricks (which Battery has invested in) and Snowflake have certainly become household names in the data lake and warehouse markets, respectively. But technical requirements and business needs are constantly shifting in these markets — and it’s important for both companies to continue to invest aggressively to maintain a competitive edge. They will have to keep innovating to continue to succeed.

Regardless of how this plays out, we feel excited about the ecosystem that’s emerging around these players (and others) given the massive data sprawl that’s occurring across cloud and on-premise workloads, and around a variety of data-storage vendors. We think there is a significant opportunity for vendors to continue to emerge as “unification layers” between data sources and different types of end users (including data scientists, data engineers, business analysts and others) in the form of integration middleware (cloud ELT vendors); real-time streaming and analytics; data governance and management; data security; and data monitoring. These markets shouldn’t be underestimated.

Casey Aylward: There are a handful of big opportunities in the data lake space even with many established cloud infrastructure players in the space:

Read more at:

<https://techcrunch.com/2021/02/13/9-investors-discuss-challenges-opportunities-and-the-impact-of-cloud-vendors-in-enterprise-data-lakes/?tpcc=ecfb2020&fbclid=IwAR10eNVfXbozkRyqn4QADpzh0Pv1---DAJ_biCIyOfy5emZ497dlbrIKoXY>