Modern at the North America AI & Big Data Expo

This past week, we had the chance to present a new, modern approach to the data development lifecycle at the North America AI & Big Data Expo in Santa Clara, California. We were joined by 5,000 like-minded attendees, including developers, designers, data experts, innovators, and many more. All came together with the common goal of building a smarter future with AI/ML and big data.

On the first day of the conference our very own head of AI, Ash Damle, gave a keynote talk on a novel vision for empowering a new generation of data developers: a decision ops platform. He explained the current challenges and what's needed — an entire AI/ML data lifecycle covering data discovery, cleaning, cataloging, observation, and governance. This not only accelerates the everyday for data scientists and engineers, it also unlocks data's full potential for an organization.

How? We share an overview below.
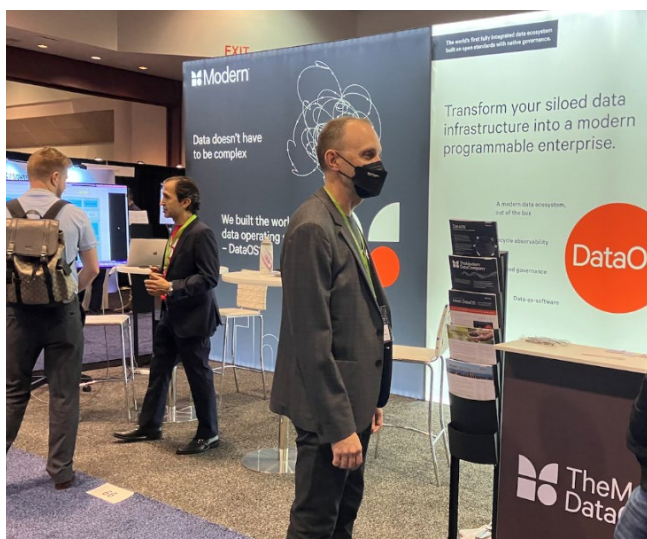
AI/ML Is Changing the World

There are seemingly innumerable examples of how AI is changing the world. It helps make the impossible possible, all the way to outer space with SpaceX. It facilitates scaling and reengineering our everyday lives, including Cruise's autonomous vehicles. It empowers us to see the world in new, clearer ways, demonstrated by Google's novel NeRF technology. AI superpowers productivity and creativity, with OpenAI DALL-E as a great example. Lastly, it revolutionizes our capacity, illustrated by innovations such as production AI from Fraunhofer-Institut für Optronik.

What is the difference between this world-changing AI and ineffective AI? It's data.

As Good as the Sum of Your Data

The beating heart of all good AI/ML is data. Data that is clean, connected, tall, and wide is what transforms "just a model" into amazing technology. It may sound simple, but paradoxically data is often the most complicated part of building AI/ML. For most organizations, 90% of time, energy, and cost is spent on getting data ready to plug into modern AI platforms, while a mere 10% is spent generating and extracting value.

The core of this issue is commonly the lack of comprehensive data abstraction to make sense of a set of disparate and heterogeneous data sources. This equates to an army of data engineers with a spaghetti of ETLs using Airflow, Python scripts, Talend, and other technology; the rigidity of the system creates fragility that requires constant attention.

What is the root of this issue, and what are the barriers to fixing it?

Barriers to a Truly Data-driven Organization

Operational, strategic, and tactical barriers currently stand in the way of organizations being truly data driven. A lack of native governance and observability, labor intensive data management, a shortage of high-skilled resources, shrinking budgets, and increasing requests translate into a majority of time being spent on operational tasks. Combined with costly data TCO, high failure rates, extensive business disruption, and remaining competitive in a field with increasingly advanced analytics leaves organizations with little time for proactive strategic initiatives.

Perhaps most importantly, needlessly complex data infrastructure, lack of data agility for experimentation, and the rise of a labyrinth of "point solutions" leading to overly rigid data architectures restricts the amount of value that can be extracted from data; therefore, limiting the power of AI/ML across the organization.

Ultimately, these obstacles amount to a fragmented data landscape, where IT is forced to act primarily as system integrators and modern data needs aren't met. It doesn't have to be this way.

The New Generation of Data Devs

Data developers are the ones designing and building this data-driven future: they are the experts molding data into decision products. A decision product is a data asset used to affect decisions by humans or machines. Data composability, visibility, and quality are the differences between a flawed model and a fruitful decision product.

However, given the barriers described above, data developers don't currently have tools they need. In order to unlock these tools, we must apply concepts from engineering to data and data development. These processes will empower data developers to see, use, and connect data differently with less time, effort, and maintenance.

A platform like Modern's DataOS provides this shift as a modern data layer that automates a significant portion of these tasks while reducing time, risk, and cost — accelerating data value extraction for data developers and the organization at large.

DataOS Is a Decision Ops Platform, Addressing the Needs of Data Developers

DataOS, the world's first data operating system, makes all facets of data use simple. It is built on the principles of data clarity, true data democratization, and ease of use. Four conceptual pillars ensure an infrastructure that reduces time-to-decision-product: flexibility, composability, outcome-based data engineering, and data as software.

Modern's tools for data developers embody a powerful shift for the entire AI/ML data lifecycle. DataOS provides a unified way of accessing, joining, governing, and observing all of an enterprise's data. It delivers loosely coupled, tightly integrated building blocks that enable organizations to compose flexible, comprehensive data architectures.

DataOS provides an integrated data layer that automates a significant portion of data processing tasks while reducing time, risk, and cost. It provides data developers with the tools they need to treat data as software and spend less time processing data, with more time

operationalizing data. It's like Git for data, a thoughtfully built platform with the UI, CLI, and other tools data developers need to manage the entire data development lifecycle. Meanwhile, built-in data governance and observability makes it easy to discover, share, and secure assets.

All of this is possible while significantly reducing time-to-value to a matter of 6-8 weeks. We've enabled this across double digits of customers with 200+ data sources, tens of thousands of tables, an average of 2.3K pipelines, and with data ranging from terabytes to petabytes. Today is the day to give your data developers the tools they need to build tomorrow's AI/ML.

Then, take the next steps to schedule a demo to see it work in real time.

Schedule a demo →

BY C. BOSTIAN