

# How DataOps Makes Information Age Companies Smarter

**Brian Dudley**

Partner, Growth Equity

**Fred Wang**

Partner, Growth Equity

**Ali Cliff**

Principal, Growth Equity

**Caoimhe MacRunnels**

Associate, Growth Equity

**KEY TAKEAWAYS**

- Cloud-based DataOps make the gathering, storing and analysis of data more efficient, reliable and cost effective than outmoded, monolithic data architectures
- Companies that leverage their data more effectively have the potential to improve product innovation and cost management
- The emergent DataOps ecosystem is set to grow as companies adopt modern architectures that integrate more seamlessly with SaaS systems

Organizations across all industries have realized that effectively leveraging their data is fundamentally tied to accelerating product innovation, optimizing financial management, realizing customer insights, and building agile operations. It's for these reasons that the global market for data infrastructure software has been one of the hottest spaces within tech over the last decade. As the volume of data created and collected by companies continues to accelerate, the global data infrastructure market is expected to surpass \$140 billion by 2027.

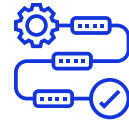
The trouble is, connecting, organizing, and making sense of mountains of data can be daunting and highly complex because the information often resides across dozens, sometimes thousands, of sources and in multiple formats in monolithic, legacy data stack architectures. Adoption of cloud-based software to analyze surging volumes of relevant data has sparked demand for flexible tools that allow data-driven organizations to reap competitive advantages from real-time analysis of the information they gather.

A new generation of open, distributed solutions has emerged that decrease the technical barriers to gathering data-driven insights with easier-to-use infrastructure for accessing, connecting, and analyzing, and challenges the traditional method of storing and processing data in proprietary, costly ways.

## Backbone of the Modern Data Stack

As companies increasingly realize that the old method of moving data away from its source to extract value is poorly suited for the modern organization, the re-emergence of data lakes and a blossoming ecosystem of cloud-based analytical tools are helping to power the open and distributed nature of the modern data stack.

Companies such as Dremio are reimagining the “modern cloud-based data stack” and promoting open architecture by moving the analytics engine to the data lake or connecting directly to warehouse sources. The cloud eliminates the need for expensive hardware to house massive amounts of data, and instead enables resource-based and flexible storage. Because cloud-based solutions help companies to easily capture, analyze and leverage the full scope of their data, it’s no surprise that there has been an explosion in tools that help organizations to manage every aspect of their cloud data lake. These tools offer better and easier-to-use infrastructure for accessing, analyzing, and furthering the use of data. It’s the collaborative data management process called “DataOps.”



**By focusing on the integration, communication, and automation of data pipelines, DataOps can significantly improve how quickly and reliably organizations gain insights to make better, more informed business decisions**

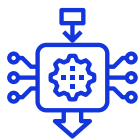
## Better Data, Improved Decisions

By focusing on the integration, communication, and automation of data pipelines, DataOps can significantly improve how quickly and reliably organizations gain insights to make better, more informed business decisions. It can be thought of as bringing the flow and organization of DevOps – the process of streamlining the development cycle – to the modern analytics team.

These new technologies give companies the ability to further leverage available data in use cases such as business intelligence, machine learning, and data science by improving and re-inventing how data is moved, manipulated, and analyzed. While DataOps encompasses tools such as testing, privacy, governance, and analytics, we believe three core categories are instrumental in capturing the power and value of DataOps in a dynamic, data-driven organization – orchestration, observability and reverse ETL pipelining (extract, transform, load).

### Three Core Categories in Capturing the Power and Value of DataOps

#### ORCHESTRATION



Orchestration makes data available to analysis tools from multiple siloed locations

#### OBSERVABILITY



Observability platforms protect the integrity and reliability of information, and identify anomalies or inconsistencies across an organization’s data stack

#### REVERSE ETL



Reverse ETL allows organizations to create a single source of useful information by transforming assets from their data lake or warehouse

## DATA ORCHESTRATION

Many companies have hundreds or thousands of “data pipelines” that transport data from one system to another. Most data pipelines have numerous dependencies and require complex sequencing with regards to when data should be transmitted. Data orchestration helps with this problem by combining and organizing siloed data from multiple storage locations and making it available to analysis tools. The latest generation of data aware orchestration solutions, such as Astronomer, use artificial intelligence and machine learning (AI/ML) to differentiate between the type of data and programmatically author, schedule, and monitor pipelines.

The ability of data orchestration tools to power the cleansing, syncing, and troubleshooting of increasingly complex and interconnected data pipelines makes them critical to the modern stack. It can be thought of as a single command center for monitoring all data pipelines. As data pipelines have greater levels of interconnectedness, a central data orchestration layer is critical in managing this complexity. Additionally, data pipelining is notoriously time-consuming for engineers but orchestration platforms allow them to focus on higher value-add tasks by dramatically simplifying the process.

## DATA OBSERVABILITY

As data grows in scale and complexity, tracing its flow across the organization is becoming increasingly difficult. This is where data observability solutions come in. New regulation and the reputational risk that can be caused by data leaks mean companies are more focused than ever on stricter management and visibility over their data flow. Players like Monte Carlo, Superconductive, Bigeye and Soda are helping companies to ensure the quality of and trust in data as it moves throughout the organization. Data observability platforms monitor data assets across an organization’s entire data stack to protect integrity and reliability, and to identify any anomalies or inconsistencies.

It is critical for companies to have a modern data observability platform that protects five measured data qualities — freshness, distribution, volume, schema and lineage. Using data for business intelligence and AI/ML is a strategic priority for most organizations and data observability platforms help to ensure that these efforts are not derailed by faulty data, the implications of which can be significant.

## REVERSE ETL

In recent years, there has been a shift in data management from traditional ETL to a new category of tools called reverse ETL or ELT (extract, load, transform). These tools allow organizations to import raw data into their data lake or data warehouse and then transform the data from there, creating a single source of useful information.

Reverse ETL solutions such as Hightouch and Census allow organizations to pull assets out of a data lake or warehouse and push them into third party tools. By offering out-of-the-box connectors to various SaaS tools, these platforms largely eliminate the need to build and maintain numerous data pipelines. This data flow allows organizations to get additional leverage out of their tools by having enriched data at a user’s fingertips. Companies invest significant resources into SaaS applications and leveraging reverse ETL tools allows organizations to get the most out of each platform while simultaneously reducing the engineering hours needed to do so.



---

**Data observability platforms monitor data assets across an organization’s entire data stack to protect integrity and reliability, and to identify any anomalies or inconsistencies**



## Emerging Opportunities

Though we are still in the early days of the evolution of the data stack, the market is showing a clear shift of workloads from proprietary warehouses to open architectures. The right tools are just beginning to emerge, and we expect to see several areas of further innovation and enhancement within the DataOps layer as the modern data stack is more widely embraced. That's why we've invested in companies such as Dremio, Astronomer, and Alation, and will seek to continue to partner with category leaders who share our view that DataOps is critical to the next generation of data. ■



LEADING WITH FORESIGHT™

**Adams Street Partners** is a global private markets investment manager with investments in more than thirty countries across five continents. The firm is 100% employee-owned and has \$50 billion in assets under management.\* Adams Street strives to generate actionable investment insights across market cycles by drawing on 50 years of private markets experience, proprietary intelligence, and trusted relationships. Adams Street has offices in Austin, Beijing, Boston, Chicago, London, Menlo Park, Munich, New York, Seoul, Singapore, and Tokyo. [adamsstreetpartners.com](https://adamsstreetpartners.com)

---

**Important Considerations:** This information (the "Paper") is provided for educational purposes only and is not investment advice or an offer or sale of any security or investment product or investment advice. Offerings are made only pursuant to a private offering memorandum containing important information. Statements in this Paper are made as of the date of this Paper unless stated otherwise, and there is no implication that the information contained herein is correct as of any time subsequent to such date. All information has been obtained from sources believed to be reliable and current, but accuracy cannot be guaranteed. References herein to specific sectors or companies are not to be considered a recommendation or solicitation for any such sector or company. Any references to Adams Street portfolio companies are provided for illustrative purposes and are not intended as a complete list of Adams Street's investments in the DataOps sector. Past performance is not a guarantee of future results. Projections or forward-looking statements contained in the Paper are only estimates of future results or events that are based upon assumptions made at the time such projections or statements were developed or made. There can be no assurance that the results set forth in the projections or the events predicted will be attained, and actual results may be significantly different from the projections. Also, general economic factors, which are not predictable, can have a material impact on the reliability of projections or forward-looking statements.

\*Firmwide AUM as of December 31, 2021 was \$49.3 billion.