



"With Ascend, more people can build sophisticated pipelines to an extent I'd never have believed was possible."

Miguel Alvarado, CTO

Lumiata is focused on making healthcare smarter. They work with major health providers to take data -- ranging from electronic medical records, claims, lab work, physician notes, and more -- and transform it into specific recommendations to improve patient outcomes and reduce delivery costs, known as Lumiata Insights. Their business is built on delivering this advanced data product

Given the foundational importance of data, Lumiata built a world-class team of highly sophisticated data engineers and data scientists. These teams leveraged modern technologies to build the data pipelines that would fuel their Insights product, including Airflow, Scala, and Spark. These tools, however, also came with significant engineering overhead and the team's productivity was being eroded as they battled with long iteration cycles and difficult maintenance. In the pursuit of simplicity and speed, the team had successfully abstracted away much of these tools' complexities, but the resulting system introduced its own maintenance burden on Lumiata's team. These teams needed to find a new path forward and began prototyping their pipelines on Ascend.

In just three weeks, Lumiata reduced the lines of code powering these pipelines by **98%**. New pipeline creation was not only **7x** faster, but Ascend also enabled the data science team to self-serve and create their own iterations. Overall, Ascend helped Lumiata simplify their data systems, focus on the data and resulting Insights, and scale the impact of their data teams.

98% REDUCTION IN
LINES OF CODE

7x FASTER PIPELINE
CREATION

Pipeline Challenges

For Lumiata, pipelines are the backbone of everything they do. Hundreds of millions of data records, of all types and structures, arrive from a variety of sources and go through multiple stages of complex transforms to:

- Scale down to only the most critical columns and timeframes
- Reformat to Parquet to support big data systems
- Cleanse and handle data or schema errors
- Model and enrich
- and ultimately join everything together into a Curated Table

This Curated Table is the basis for the Data Science team to develop the Lumiata Insights. However, with higher volumes of client data and faster SLA requirements, both sides of this process began to strain.



On one side, the pipelines to create the Curated Table were manually stitched together using a mix of Apache Airflow, Apache Spark, Python and over 100,000 lines of custom code. Onboarding each new client required bespoke development and the over-extended data engineering team was responsible not only for this development, but also for maintaining and monitoring the pipelines, as well as the health and performance of the underlying Apache Spark jobs.

Meanwhile, the data science team required a certain amount of experimentation and iteration to develop the Lumiata Insights, but were completely dependent on data engineering to provide necessary adjustments to the Curated Table. Each request would take a week or more; the data engineering team would need to go back to the unruly codebase, find where to make the update, hand-code it, test the change, deploy back into production, and wait for the Curated Table to finish updating. Since the data science team was only working from a partial, cleansed view of the data, and weren't able to translate back to the source codebase, these adjustments and back-and-forths were constant to get the right data prepared in the right way. End-to-end, this whole process would take six weeks or more, and induced a heavy maintenance burden to keep everything running. As the company looked to scale to take on more clients with their existing team, they needed a new approach.

“With the raw data we get from clients, there’s always something unexpected. The schema changes or the data has a new delimiter. It’s a manual, time-consuming process to readjust for these at each stage of the pipelines without disrupting the downstream Model.”

Miguel Alvarado, CTO

“Ascend is like autopilot for my data science team. It’s so easy to create new transformations and adapt to any data changes. And the ability to build end-to-end pipelines all in SQL is game changing for not only how our teams work together but also how we can share Insights with clients.”

Rohun Kshirsagar
Senior Director of Data Science

Intelligent Pipeline Orchestration & Automation with Ascend

Lumiata opted to give Ascend's Autonomous Dataflow Service (running on Amazon Web Services) a try. Within three weeks, they were able to migrate all their existing pipelines to the Ascend Service. Creating the Curated Table now involved only 2,000 lines of reusable code, reducing the total codebase to maintain by **98%**. Ascend's declarative programming model ensures the resulting code is focused solely on the data and logic development (not on executing tasks or infrastructure management), so both data engineering and data science teams can collaborate and iterate together clearly and quickly, without risk.

Additionally, the automation that the Ascend Service provides fundamentally changed the scale with which they could work with data. When onboarding new datasets, the data engineering team can now ingest and work with all of the source columns with no extra code or complexity. And, since Ascend automatically converts all data to Parquet upon ingestion, they've been able to fully eliminate that time-consuming processing stage.

The biggest impact, however, was on the speed of iteration. The data science team now has full visibility into the context of the data and resulting Curated Table. They can clearly trace operations done on data fields, and even pull in new fields or adjust the logic directly. At any stage of the pipelines, they're able to experiment and iterate directly using SQL, Python, or custom functions interchangeably -- all in a matter of minutes, not weeks. Ultimately, the process to go from raw data to Lumiata Insight is now **7x** faster.

"As a growing company, we needed to maximize the impact of our small team wherever possible. Ascend has enabled us to automate many of the manual development steps involved with pipeline creation, so more people can build sophisticated pipelines to an extent I'd never have believed was possible. Our data engineers can finally stop plumbing and truly focus on engineering, enabling them to take on a far broader scope of development work."

Miguel Alvarado
CTO

Ascend Benefits Across Teams

Data Engineering Team

Working from the concise, declarative codebase has made it faster to build new pipelines and relieves much of the maintenance burden for existing pipelines. Additionally, since Ascend's Control Plane automates the underlying Spark job generation and execution, this team no longer needs to manually schedule or monitor the jobs themselves. However, full visibility into the underlying runtime stats, resourcing, and cloud spend is always available.

Data Science Team

The ability to self-serve updates and experiment rapidly has allowed this team to become Citizen Data Engineers -- giving them more accurate models more quickly and relieving pressure from the central data engineering team. All changes are fully traceable and the supporting automation ensures these changes can be made safely, without impacting other pipelines or result sets.

Executive Team

Delivering accurate Lumiata Insights as quickly as possible, without exponentially scaling the teams is what matters most. With Ascend, data engineering and data science can collaborate together for faster, more agile development cycles. This has resulted in an **83%** reduction in time-to-insight and has enabled them to take on more projects with their existing team.