Morningstar Empowers Investor Success
with a Unified Data Lake

MORNINGSTAR®

As challenges arose due to evolving data needs, Morningstar decided to bring its datasets into a single location and leverage them for a more intuitive and powerful software experience for investors. To configure its numerous and complex pipelines, Morningstar needed an Amazon Web Services (AWS) - native Extract, Transform and Load (ETL) solution that could be deployed quickly, and Etleap met those needs.

"We've moved from on-prem to AWS because we don't want to be in the business of building data centers," explains Alex Golbin, Morningstar's Chief Data Officer. "Similarly, we're leveraging Etleap because we don't want to be in the business of building data pipelines."

By teaming with Etleap, Morningstar saved many months of engineering work and gained new abilities to pursue excellence and efficiency on an accelerated timeline.

An independent voice in the industry

During its nearly four-decade history, Morningstar has remained a pioneer of financial insight, built on comprehensive data and granular analysis.

> "
> Just as we've moved from on-prem to AWS because we don't want to be in the business of building data centers, we're leveraging Etleap because we don't want to be in the business of building data pipelines.
> "
>
> **Alex Golbin, Chief Data Officer, Morningstar**

"Data is at the heart of everything we do," explains Jeffrey Hirsch, Head of Technology for Data and Analytics at Morningstar. "Our researchers perform daily quantitative and qualitative analysis and need accurate data to make long-term, value-based assessments."

Morningstar's goal is two-fold: maintain a rich trove of historical and real-time equity data and manage it in a way that benefits their clients as they scale. Over the last 25 years the quantity and speed of available data have increased exponentially. To maintain its competitive advantage in this increasingly complex environment, Morningstar has evolved in tandem.

Morningstar relies on key datasets to deliver valuable outcomes

Morningstar's software products leverage financial data to create intuitive workflows and generate insights for customers. Financial advisors, asset managers, and other financial professionals depend on these workflows to guide their day-to-day investment decisions.

As Morningstar succeeded, matured, and acquired new companies, it also faced the common technical challenges that accompany growth. Over time, Morningstar's internal data sources were peppered across the organization, so extra effort was required for separated parts of the organization to collect and share data with each other.

"Acquiring a company is just the beginning," explains Hirsch. "Integrating that newly acquired company into the parent company [from a data perspective] is extremely tedious and can take years."

By finding a solution that would streamline their processes, Morningstar's teams would be able to collaborate more effectively.

Unifying disparate datasets into one location enables new abilities

Three examples of Morningstar's strategic datasets include:

- Environmental, social, and corporate governance (ESG) data from Morningstar's acquisition of Sustainalytics
- Equity and fixed-income data used to power Morningstar Indexes
- Venture capital, private equity, and M&A data from its PitchBook platform

While powerful individually, these datasets yield more sophisticated insight when joined in a central location. Centralization removes the potential for data inconsistencies that result from keeping multiple systems in sync over time. In addition to giving teams more robust and holistic insights - thus enhancing the overall client experience - a centralized data repository allows for easier scaling, simplicity, and cost savings.

Building out an S3 data lake in AWS

The architecture to build out Morningstar's data lake on AWS was complex. Data would move in sequence from source to landing zone, to quarantine zone, curated zone, ledger, then finally to structured form within the S3 data lake. Building around AWS's core services, Morningstar created APIs, workflows, and pipelines to provide a consistent way for team members to ingest data into the lake.

Although Morningstar sought to generalize this process, they encountered scalability issues as their codebase underwent enormous expansions as new use case permutations entered the mix.

Initial analyses suggested that Morningstar's vision would require more than 100 on-site technical specialists building out thousands of data pipelines over five years – all at a significant expense. To streamline the process, they evaluated several outside ETL vendors but most failed to fulfill their unique business and operational needs.

To build or to buy?

Considering the heavy workload required of both mainstream vendors and building in-house, Morningstar sought out better options. As part of their data lake roadmap, their engineers were pursuing workflows to facilitate data ingestion, operations, pipeline monitoring, and scaling.

Etleap's product already had built-in solutions for many of these, including:

- Incremental updates to data
- A workflow for triaging data movement to support multiple big data file formats
- Optimizations of storage for performant queries
- Maintenance-free scaling to large data volumes
- Incremental schema updates
- Deployment flexibility to enable strict security requirements

Morningstar made a comparison between the investment required for building out on their own versus the time and cost savings of buying from Etleap. Etleap cut their time-to-build by more than two years, and did so for less than half of the initial estimated expense.

"Etleap specializes in doing ETL at scale," says Hirsch. "So collaborating made perfect sense."

Etleap cut their time-to-build
by more than two years,
for less than half of the
initial estimated expense.

Bridging the gap with Etleap

In Etleap, Morningstar found a next-generation solution built from the ground up on AWS technology.

By employing Etleap's Data Wrangler Morningstar can gather data from producers, then use Etleap's modeling function to quickly and easily transform the data in ways that are intuitive to data consumers, and that models how the data relates to other similar data in the lake. Now, Morningstar has the power to build models that leverage core financial data such as entities, securities, and portfolios in a consumer-friendly way – true to the nature of business relationships. These models relate data in such a way that consumers can access relationships quickly and generate new insights.

"That's where the real power and true value of Morningstar come to fruition," says Hirsch. "It's all of our data used collectively to generate insights across different lenses."

"

That's where the real power and true value of Morningstar come to fruition. It's all of our data used collectively to generate insights across different lenses.

Jeffrey Hirsch, Head of Technology for Data and Analytics, Morningstar

"

Data lake and ETL solution unlocks innovation potential

Morningstar and Etleap team up to deliver client success

After acquiring Sustainalytics, Morningstar leveraged Etleap to help integrate their ESG investment decisions and recommendations into the data lake within six months. This success paved the way for Morningstar to rapidly build pipelines for other financially pertinent datasets. Now, Morningstar is integrating valuable datasets in a matter of days.

In today's era of climate action and corporate sustainability, ESG data is increasingly part of the investing conversation. With this dataset, portfolio managers can apply an ESG lens on their investment recommendations - whether it's investment portfolios or indexing. Having that data available and accessible in a consistent way, and clearable at scale from the data lake is the next step to building and sustaining new customer capabilities.

With Etleap's support, Morningstar can also build new product offerings like Analytics Lab, which give clients a multifaceted view of data so they can make decisions with less friction and leave behind time-consuming processes.

For this innovation to function properly, data must be rapidly deposited into the lake so that it can be exposed through the Data Explorer, where it immediately becomes available to researchers and data scientists to query that data, notebook it, join it with other data, generate insights, and share that notebook with others.

Previously, Morningstar's quantitative research team would have to search through all Morningstar databases to retrieve the pertinent data, gather it into a single place and cleanse it, then build a model over top of it to generate results. If the results proved fruitful, a developer would take over to productize the idea – a process that end-to-end took months. The result was that their research team had to be very deliberate and judicious on what ideas they thought were good because the barrier to entry to developing an idea was extremely high.

With Morningstar's data lake and ETL infrastructure, Morningstar researchers can focus on their core mission without worrying about the prerequisites and hurdles that they had to overcome previously to manage the complexity associated with increasingly larger datasets.

"We're at the very beginning of a new level of innovation from our research team at Morningstar," concludes Hirsch. "We now have a solution for data management and modeling that allows for the easy creation of more innovative, interesting, and exciting products and methodologies. More so than ever before, all the data and the ability to manipulate it is right at our fingertips."

Morningstar, Inc. is not affiliated with Etleap.

ABOUT ETLEAP

Etleap's mission is to make data analytics teams more productive. Etleap's ETL solution lets analysts build data warehouses without internal IT resources or knowledge of complex scripting languages. This reduces the time of typical ETL projects from weeks to hours, and takes out the pain of maintaining data pipelines over time. Etleap was founded by Christian Romming in 2013. Before founding Etleap, Romming was the CTO of an ad-tech company, where he recognized that the available solutions for building data pipelines required monumental engineering resources to implement, maintain, and scale. Etleap is backed by world-class investment firms First Round Capital, SV Angel, BoxGroup, and Y Combinator.