



SaaS company for data management, BI, and POS applications

**80% performance improvements  
in data pipelines**

## **Business Challenge**

The client captures, collates, and analyzes POS and sales data to generate insightful reports for CPG manufacturers and retailers. They were unable to onboard data from bigger retailers as it was difficult to scale up their existing platform. The data feed that they received from different retailers had to be cleansed, enriched, transformed and stored. The client switched to an Azure-based cloud platform to accommodate and process multiple feeds in parallel and on time. The new platform had its share of challenges with regards to performance, scalability, stability, and high cost that had to be resolved.

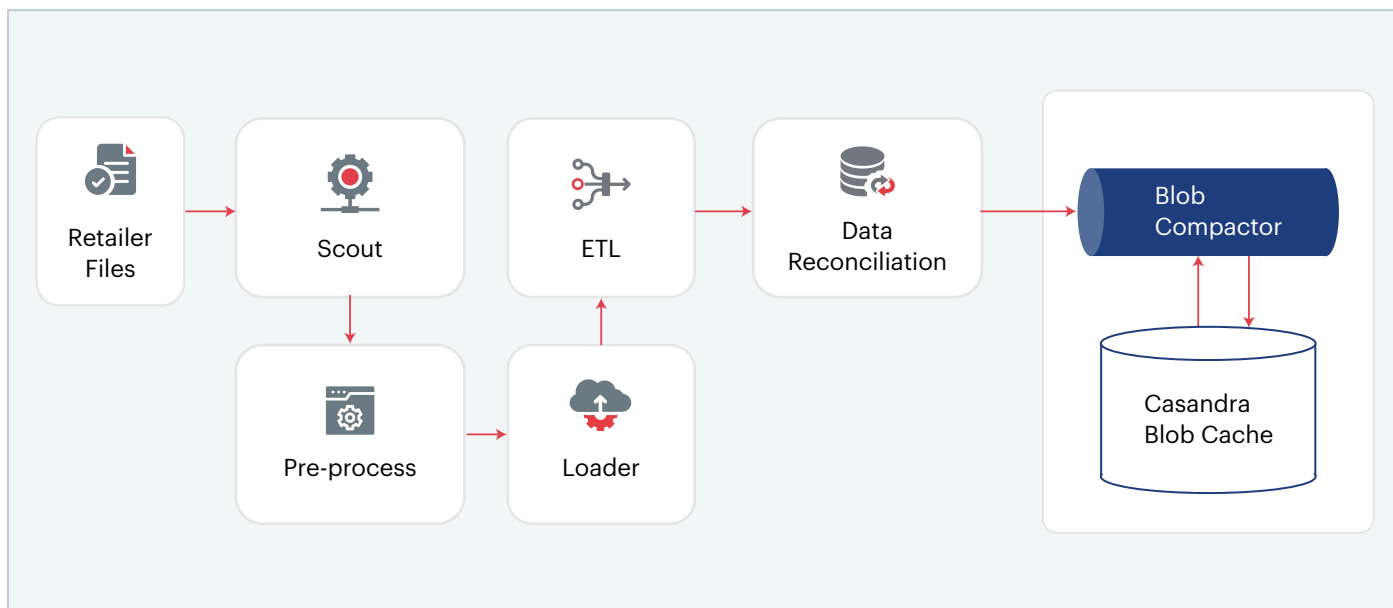
**Sigmoid's ETL optimization and architecture improvements made the data pipelines highly scalable and performant in the client's Azure-based cloud platform and helped ingest high-volume data from the large retailers like Walmart, Walgreens, and others.**

## Sigmoid's Solution

After reviewing the existing ETL data architecture, Sigmoid performed a comprehensive analysis and benchmarking of different components and business rules. The existing system wasn't horizontally scalable and it was crucial to resolve the biases of the legacy platform. The ETL pipeline and data reconciliation codebase were rewritten and incremental changes were made to quickly test the impact.

The Spark transformations code tuning was done based on the data size, cluster size and memory footprint, apart from implementing the autoscaling features in the Spark cluster. As part of the entire architecture overhaul, a cache was built to make Parquet read faster. Sigmoid's strategic approach helped the client teams test the data from large retailers like Walmart, and addressed different performance challenges. Additionally, the DataOps automation brought more agility to deploy and test the changes faster.

### Tech Stack



## Business Outcomes



Sigmoid's solution resulted in more stable data pipelines and faster data processing, reducing the execution time of Spark transformations from 90+ mins to under 15 mins. This led to enhanced reporting capabilities and improved performance, apart from significant cost savings due to a reduction in computing resources.

## About Sigmoid

Sigmoid enables business transformation using data and analytics, leveraging real-time decisions through insights, by building modern data architectures using cloud and open source. Some of the world's largest data producers engage with Sigmoid to solve complex business problems. Sigmoid brings deep expertise in data engineering, marketing analytics, AI, and DataOps. Visit [www.sigmoid.com](http://www.sigmoid.com) to know more.

### Let's Get Started!

To know more about how we can help optimize your existing data pipelines and architecture, contact [marketing@sigmoid.com](mailto:marketing@sigmoid.com).