



TRANSFORMING TO A DIGITAL ENTERPRISE PROVIDES DATA IN REAL-TIME WITH 99.9% ACCURACY

IMPACT

- Clean and consolidated data available for everyone, with 99.9% accuracy
- Data constantly governed to maintain quality and security
- Reduced 10-12 person job that took 1.5-2 months down to 1-2 person job that takes one week
- Cost savings of up to \$2-3 million per quarter

All across the globe retailers are needing to transform how they do business - becoming more data driven to create long lasting relationships with customers. This multinational retailer's digital transformation began with the goal of building a seamless shopping experience for customers. The accuracy and accessibility of data was paramount to achieve this goal. Before starting on this journey, data was inaccurate or unavailable to the users who needed it.

Inaccurate reporting impacting the business

Access to data was a big challenge that the various lines of business organizations faced, specifically in the finance department. Data was brought in from various sources, and users were often unclear of which source was the correct one to use - especially for the merchandising and marketing teams. Processing of data was unreliable and inconsistent, both in cost of goods (COGS) and profitability measures. Business leaders began noticing huge differences with the two sets of reports they received and no way to track which source was used to generate the reports. Understanding accurate COGS and profitability on an intraday basis when taking into account promotional and pricing discounts at the unique item level were a challenge.

The accuracy, reliability, stability, and credibility of the data used were of concern. There were no common definitions of hierarchies, and no standard naming conventions. This created a convoluted structure that the finance department had to sift through, to do heavy lifting, and remedy the reports. This created process inefficiencies taking up to 30-70% of their time on manual data gathering and reporting to equip business leaders with accurate information to power decision-making.

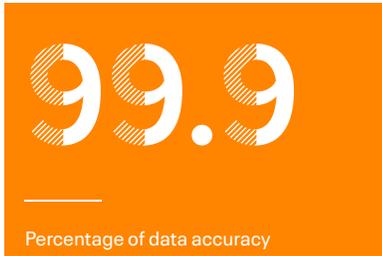
A need for real-time data to meet SLA's

There was a need to not only radically improve these processes, but also access real-time data. Business leaders wanted to slice and dice data, comparing the current year with last year to better understand performance with all causal (temporary price reduction, ad, display) factors incorporated. The finance department encountered the challenge of meeting the business' service-level agreement of supplying updated data through 6:00 AM. Most business users were in the Central Time Zone, but there was late arriving data from the stores located in Alaska and Hawaii. To meet the SLA's, the team had to institute a 4:00 AM cut-off time to accommodate late arriving data. Retrieving data, handling vast volumes, and measuring it on time were all not possible with the legacy landscape of systems and applications.

To try and address these challenges, the company was still investing in legacy systems, serving as a quick fix, for a limited number of users and did not address the complete problem. Furthermore, the fragmented approach meant license fees and operational costs were spiraling out of control. The company needed an enterprise platform solution that could address all these challenges.

Enabling data governance on the data lake - ensured quality and accuracy

This sparked a data lake initiative using Cludera's enterprise data platform as the foundational solution. It became the single source of truth for the entire organization, across all source systems, with a central and economically scalable repository that enabled users to access accurate data efficiently.



“Working with Cloudera Professional Services has been instrumental for us in terms of our time efficiencies, upgrades, and architectural design. It’s valuable to have them available whenever we need.”

Member of the engineering organization

Having a robust solution, with proper guidelines and data governance on top of the data lake, ensured high quality and accuracy. The retailer built an in-house ETL framework, and paired it with a hybrid data architecture consisting of Cloudera’s HDP and a rich and robust private cloud using Google Cloud Platform (GCP). There are three zones for data brought into the data lake: RAW (data ‘as is’), catalogued (published data with proper lineage), and non-catalogued (allowing users to bring their own data).

With the Cloudera enterprise data platform, the business and analytic teams now have trust and confidence in the data: data and metadata lineage allows them to verify the information’s origin as well as see where it is being used. Ranger policies were defined to give the right access to the teams, making sure they only have access to the data they need.

“An enterprise needs to be adaptable and flexible. We’re providing solutions and tools on top of the data lake, where users of different capabilities can use the same data for their own unique queries. We’re utilizing Druid and Spark on top of the data lake for high speed performance, so that all queries and results are returned in sub-seconds,” said a member of the engineering organization.

This multinational retailer realized tremendous cost efficiencies through having a single source of truth. By migrating workloads from older legacy systems that were subsequently decommissioned, this retailer was able to save money on licensing costs. Built on Cloudera’s platform, the team created an enterprise data analytics solution, with multi-tenancy, that all teams and organizations across the business could access through a single source. All data is now brought into the data lake, so business users always have the most current and up to date information at their fingertips.

The team also saw significant value in partnering with Professional Services.

“Working with Cloudera Professional Services has been instrumental for us in terms of our time efficiencies, upgrades, and architectural design. It’s valuable to have them available whenever we need,” said a member of the engineering organization.

Clean and consolidated data available for everyone, with 99.9% accuracy

Finance was the first department that was on-boarded onto the data lake and now the team has the ability to track multiple KPIs for customer data, supply chain, and merchandising. This included improved insights into the impact of markups, markdowns, and inventory receipts. The retailer produces hundreds of gigabytes of data weekly, across 200 million global items. Now, whenever business users are looking at particular numbers, they are all coming from the same source since all reporting is done on top of the data lake. If they see an issue or a negative trend, they can drill down quickly to a specific item or department at the store level to fix issues immediately.

Building an enterprise solution has helped remediate most of the initial challenges that the company was facing. Now they have integrated multiple data sources into a single source of truth. This helps avoid a lot of manual steps, which leads to more efficient processing.

“Previously, when people were getting in many different sources and trying to combine them into a single source, you’d see a lot of data inconsistencies. This resulted in an accuracy of 90%. We then had to do some reconciliation, manual steps, to bring the accuracy up to 99%. With our data lake, whenever you query it, you’ll always get an accuracy of 99.9% because all the reconciliations are done automatically before being sent to the user,” said a member of the engineering organization.

These results have been transformative in terms of the business impact. For example, previously the tax team used to be 10-12 people working for 1.5-2 months on extracting, cleaning, and reconciling data before numbers could be reported. A substantial amount of time and resources were spent on generating those reports. Now with the data lake in place, it has reduced this to a 1-2 person job that takes a week at most.

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises. Learn more at cloudera.com

This results in significant time and cost savings, up to \$2-3 million per quarter. Now these resources can be reallocated to other areas.

The finance organization went through an evolution, and some additional results it has seen include:

- Clean and consolidated data available for everyone, with 99.9% accuracy
- Data constantly governed to maintain quality and security with Ranger policies
- All users have access to clean, high quality, consistent finance data through different layers like ThroughSpot, Tableau, Druid
- The business keeps changing, so the finance team is constantly improving on the existing model, always optimizing, automating, and gathering more data

The data science team is also users of the data lake, and they're using it to build faster, more accurate sales and demand forecasting models. They've built a global data portal that has automated dashboards. Where reports were static before, they are now updated with data from the most recent 30 minutes utilizing the Cloudera enterprise data platform.