

CASE STUDY

# 96% Reduction in Data Quality Issue Resolution Time and Improved Data Operations Result in Millions of Dollars in Savings for Global Data Provider



## Customer Profile

- Global data provider and world's largest source of commercial information
- 6,000 employees, \$2B annual revenue
- Delivers reliable data products to 90% of Fortune 500 companies
- Delivers commercial data and analytics to empower businesses in evaluating creditworthiness, mitigating risks, and making well-informed choices regarding prospective business associates, clients, and vendors.
- Decades of expertise used to cleanse, transform, and deliver data products for global clients

## Massive scale of data

- Collects and analyzes data on 480 million+ global organizations
- Data sourced from tens of thousands of global data suppliers
- Diverse data types collected daily (addresses, revenue, legal records, etc.)
- Over 3,000 attributes per company
- Operates massive dataset: 2.5 - 3 PB; 500MM+ records on GCP
- Reduced data quality issue resolution time to 45 days
- Multi-petabyte database

## Challenges Around Ensuring Data Quality

- Legacy data quality solutions couldn't manage petabyte-scale data or complexity
- Millions spent on support costs for legacy data quality tools
- Lack of a singular, global view or control over data quality
- Legacy vendors required separate data quality checks for each application
- No real-time processing or automated monitoring with legacy solutions
- Data incidents found by consumers had disastrous effects on business, revenue, and brand reputation
- Legacy data quality tools were incapable of tracing bad data back to the source
- No recommendations for resolving data issues





“

Don't let dirty data come inside your home!

Acceldata's shift-left data observability catches data quality issues early - at the source - before it causes exponential damage downstream. That's what we love about Acceldata.

Head of Data,  
Global Data Provider

## How Data Observability Improves the Company's Data Operations

-  Run at scale
-  Ability to explain insights (root causes)
-  Works with disparate sources
-  Properties/tags/labels

## Technical Challenges

- **Unacceptable resolution times:**
  - Troubleshooting data issues from suppliers took nearly 30 days
  - Legacy data quality tools required 14 - 27 days for data processing and quality measurement
  - Follow-ups with suppliers extended resolution time to 45 days
- **Inefficient use of data engineering team:** 80% of team time was consumed managing infrastructure for data quality support
- **Legacy data quality tool limitations:**
  - Required specific data formats, necessitating data reformatting
  - Couldn't handle complex business rules without multiple steps
  - Depended on expensive tools and infrastructure, resulting in a large compute footprint

## Consequences and Risks of Legacy Data Quality Solutions

The above challenges led to:

- Millions of \$\$ of unnecessary spend
- Countless engineering hours required to ensure data quality
- Customer dissatisfaction and churn due to bad quality data
- Lower renewal rates
- Risked legal action and FTC compliance issues leading to expensive fines
- Reputational damage

## Company Business Goals

The company has set the following rigorous business goals to ensure continued reliability of data products that are delivered to end consumers:

- Increased quality of data
- Reduced tech debt
- Improve quality of data product development and deliver products to the end-consumer faster
- Reduce dependency on unscalable legacy data quality tools
- Increase scalability to align with company's growing and increasingly complex PB-scale data environment



## The 1×10×100 rule of data quality:

- For every dollar spent to fix and prevent data quality issues in the landing zone, it costs \$10 to fix it in the transformation zone and \$100 to fix it after it has gone into consumption.
- Acceldata's "shift-left" data reliability capabilities detect issues and fix data as it comes in saving customers money

## Improvements Provided by Acceldata

-  Ability to work on data in motion and data at rest
-  Data profiling
-  Reusable rules and policies
-  High performance (processing speed)
-  Alerts, notifications & recommendations

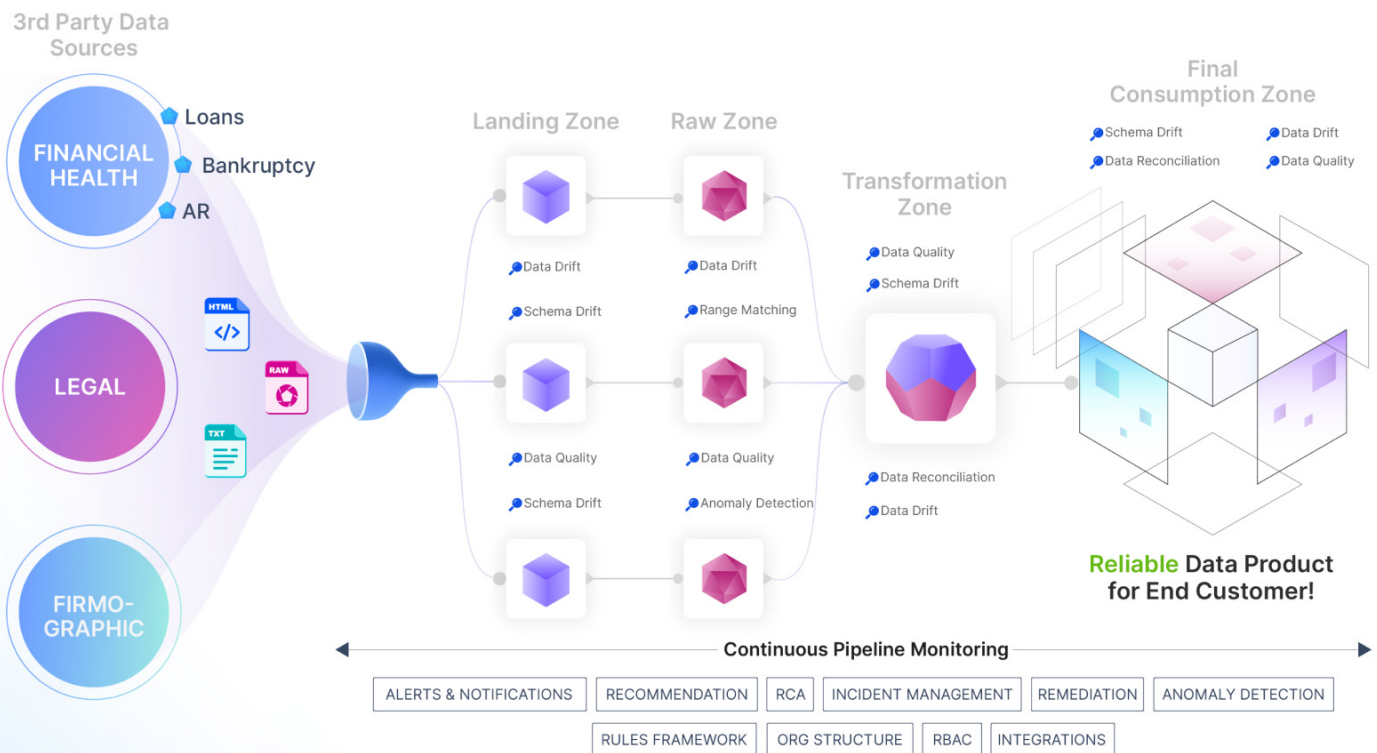
# The Solution: Acceldata

- Optimized data operations by ensuring end-to-end data quality and pipeline reliability in company's data supply chain
- Acceldata's shift-left data observability identifies and rectifies quality and pipeline issues at the source, preventing expensive downstream contamination
- Continuous, automated checks applied across the data supply chain, covering data at rest, in motion, and in consumption, enhancing data quality and reliability



**Acceldata transformed** and optimized the customer's internal data supply chain. All data is validated and cleansed by Acceldata before it enters the data supply chain.

## Data Observability For Data Providers



# How the Customer is Using Acceldata

- Cataloging and creating metadata for all data sources and files
- Deployed 600+ custom data policies on 500+ billion rows in just six months using Acceldata's rules framework
- Reusing policies across multiple data assets, including rapidly bringing new data sources into production
- Ability to write data quality and validation rules using SQL, Python, and Spark
- Running data quality and data validation rules on both static and streaming data
- Detecting schema, and attribute-based changes
- Can handle large data volumes, including data files with 100+ million records
- Pipeline observability enables quality checks on data in pipelines, including at landing zone, and in data assets
- Can now identify and alert data owners to data issues within hours, rather than weeks
- Benefiting from automated recommendations for resolving data quality, pipeline, and cost issues.

## Results Observed with Acceldata

- Significantly **reduced customer churn**
- Enhanced **compliance** with **FTC guidelines** for data quality reporting
- **Eliminated** SEV1 data issues
- Increased **customer satisfaction** due to greater data consistency and higher data reliability
- Dramatic reduction in processing time - from **22 days to 7 hours**
- **30PB** quality checks performed monthly
- **20x faster rule creation led to** improved speed and accuracy
- Continuous data checks with 200+ rules for 500 billion rows
- Improved **engineering outcomes** with data quality automation
- Accurate **ROI** assessment for each data source
- MTTR reduced from **15 days to 2-4 hours** - reduction by **96%**
- Noticeable **morale boost** in data engineering teams using Acceldata as they received **formal kudos** from business teams

## Our Customers



Get started today for free.

Request Demo