How Bol.com Uses RudderStack for Real-Time Event Data, Personalization, and Compliance

See how Bol.com replaced legacy pipelines with RudderStack to deliver real-time analytics, secure event data, and CI/CD-driven personalization.

Bol.com, the largest eCommerce platform in the Netherlands and Belgium, is replacing legacy DIY in-house infrastructure with RudderStack to drive advertising, personalization, and analytics - securely, scalably, and with full developer control.

About bol.com

Bol.com is the leading eCommerce platform in the Netherlands and Belgium, serving over 13 million active customers and offering more than 36 million products. Founded in 1999 as the Netherlands' first online bookstore, Bol.com has since evolved into a multi-category online marketplace and technology company with a mission to make daily life easier for consumers.

The company now operates with over 2,000 employees and supports 50,000+ retail partners through a sophisticated logistics and fulfillment network. Its robust platform powers everything from direct-to-consumer sales to partner-driven merchandising, advertising, and personalized shopping experiences.

As Bol.com expanded its product catalog and revenue streams, including the launch of its advertising platform and a proprietary fulfillment center, they faced new complexity in how it measures user behavior, attribution, and performance. Accurate, real-time event data became foundational not only for analytics but for driving revenue through personalization and media delivery.

To support this strategic shift, bol.com required a modern data infrastructure - one built with scalability, governance, and developer control at its core.

Industry
eCommerce and Retail Media

Region
Netherlands & Belgium (EU)

Monthly Event Volume
50+ billion

Use Cases
Behavioral analytics, personalization, advertising measurement, experimentation, recommendation systems

Key Stakeholders
Engineering, Data Science, Analytics, Product, Marketing, Security & Compliance

The Challenge

Bol.com was operating on an in-house data collection and measurement platform originally built over a decade ago. While it served the technical needs of advanced data scientists, the system had not scaled with the organization's business model or user base.

Key Pain Points

Siloed Ownership

The in-house platform was built and maintained primarily by backend engineering teams. Coordination between producers (frontend developers), platform engineers, and consumers (product analysts, marketers) was fragmented.

Overly Generic Schemas

Event tracking was largely freeform. Data producers had few controls, and consumers often had to reverse-engineer meaning from loosely structured payloads.

Low Usability for Analysts

While data scientists could work directly with BigQuery outputs, most analysts struggled to find or trust the data, resulting in duplicated work, shadow data pipelines, and friction.

Reactive Governance

There was limited real-time alerting or monitoring. Broken events often went unnoticed for weeks, and quality issues had real business consequences.

Security Gap

As Bol.com launched its retail media business, it faced mounting pressure to ensure data was collected securely, accurately, and with verifiable integrity - especially for monetized events like ad clicks and impressions.

The bol.com team realized that evolving their internal system would require a major investment in people, not just technology - and that a commercial-grade, cloud-native platform might offer faster time-to-value.

Why RudderStack

Bol.com launched a comprehensive evaluation process, exploring both internal rebuild scenarios and external customer data infrastructure platforms. RudderStack was selected based on its ability to meet enterprise requirements while offering engineering teams flexibility, transparency, and control.

Top Selection Criteria:

Performance at Scale

RudderStack was tested in Bol's infrastructure to handle 50B+ monthly events with high throughput and minimal latency.

Developer-Centric Workflow

Bol's engineering team adopted RudderStack's CLI, YAML-based tracking plan format, and GitLab CI/CD plugins to build data governance into their pipelines.

Security First Architecture

RudderStack introduced key features including transformation sandboxing, customer-controlled ingestion proxies, and automated alerting for anomalous traffic (e.g. bot detection, data exfiltration).

Collaborative Support Model

RudderStack embedded senior engineers and security leads directly into Bol's evaluation and onboarding process, creating a tight feedback loop across performance, compliance, and roadmap planning.

Customizability

RudderStack's open schema model and modular ingestion framework gave Bol the flexibility to tailor event specs, sessionization logic, and attribution methods to its unique business needs.

Implementation Highlights

Technical Evaluation

- RudderStack delivered a full-scale POC using Bol's real traffic (50B+ events/month) and demonstrated end-to-end delivery to BigQuery, downstream transformation, and error reporting.
- The platform was tested under simulated bot attacks and unexpected volume spikes, validating its autoscaling and backpressure resilience.

Security Engineering Collaboration

- RudderStack implemented alerting for anomalous ingestion patterns to support internal compliance requirements.
- Introduced container-isolated Python transformations after a pen test flagged a potential vulnerability.
- Jointly planned ingestion failover and replay strategies via S3 and PubSub integrations.

Developer Workflow Integration

- Engineers adopted RudderStack's CLI for tracking plan versioning and validation.
- GitLab CI pipelines automated schema validation and deployment into multiple RudderStack workspaces (e.g., staging, production).
- Planned enhancements include nested property support, conditional schemas, and transformation management via CLI.

"We use RudderStack as if it's a part of our own team. From load testing to transformation isolation, they've been all-in on building the right platform with us."

bol.

Koen Lijnkamp
Sr. Engineering Manager

Bol.com Results: Improved Governance, Real-Time Personalization, and Developer Velocity

With RudderStack, Bol.com has fundamentally upgraded its customer data infrastructure, delivering immediate benefits and laying the foundation for future innovation:

Improved Data Quality & Trust

- YAML-based tracking plans and CI validation workflows ensure only high-quality, schema-compliant events are accepted.
- Observability and alerting provide real-time visibility into event delivery and failures.

Stronger Security & Compliance

- Custom proxy ingestion flow ensures that event data is treated as first-party, which protects user privacy and preserving ad measurement in the face of browser restrictions.
- Python transformation containers and exfiltration monitoring reduce risk and meet internal InfoSec standards.

Empowered Engineering Teams

- Product teams now deploy instrumentation updates safely and independently using Git workflows.
- Platform engineers can version, test, and release changes to tracking logic with full automation.

Unblocked Strategic Use Cases

- Ad teams now have trusted, real-time performance metrics for sponsored products and campaign ROI.
- Product and data science teams are using clean, consistent behavioral data to feed recommendations and experimentation models.

What’s Next

The partnership between Bol.com and RudderStack continues to deepen, with shared roadmap alignment and co-development opportunities in the following areas:

CLI-based Transformation Management

Bring RudderStack Transformations into GitLab workflows for versioned, testable logic.

Nested Property & Custom Type Support

Enhance schema expressiveness for complex data models across events.

Disaster Recovery & Replay Optimization

Streamline ingestion failover with buffer/replay logic jointly managed between Bol.com’s proxy and RudderStack’s archive layers.

Advanced Attribution & Sessionization

Customize identity and session logic for advanced analytics use cases, including hybrid app/web flows.

Snapshot: Before & After

Capability	Before (In-House Platform)	After (With RudderStack)
Event Volume Support	Before: Unverified at scale	After: Proven at 50B+ events/month
Schema Governance	Unstructured, inconsistent	YAML tracking plans + CI validation
Real-Time Alerting	Manual detection	Automated ingestion + quality alerts
Transformation Management	Ad hoc, uncontrolled	CLI-based, sandboxed, testable
Data Consumer Experience	Fragmented, trust issues	Unified, versioned, and documented
Security Posture	Pen test vulnerabilities	Resolved via sandboxing and monitoring
Developer Workflow	UI-only, no version control	Git-integrated CLI with automated pipelines

Conclusion

The shift from Bol.com's legacy in-house event platform to RudderStack marks a key milestone in the company's data platform evolution. With real-time trust, strong security, and developer-first tooling, Bol.com has the infrastructure to scale innovation across every team - product, marketing, data science, and engineering alike. RudderStack now plays a central role in supporting bol.com's personalization, advertising, and measurement strategies.