



# Data Ops/Data Integration: How Retailers Can Win with Real-Time Data Integration

When you interact with IBM, this serves as your authorization to Dataversity or its vendor to provide your contact information to IBM in order for IBM to follow-up on your interaction. IBM's use of your contact information is governed by the IBM Privacy Policy.



71%

of organizations say decisions must be made faster and more frequently.<sup>1</sup>

**Gartner**

80%

of organizations rely on stale data for decision-making<sup>2</sup>

 **businesswire**  
A BERKSHIRE HATHAWAY COMPANY

85%

of data leaders cost their company money with stale data<sup>3</sup>

 **Agility**  
PR Solutions

Let's discuss **Retail** –  
and the cost of not using  
**real-time data**

1

**\$1.75 trillion** lost annually due to overstocking,  
stockouts, and customer dissatisfaction.<sup>1</sup>

---

2

**66%** switch retailers if items are unavailable

- **71%** expect personalization
- **76%** feel frustrated when expectations aren't met.<sup>4</sup>

---

3

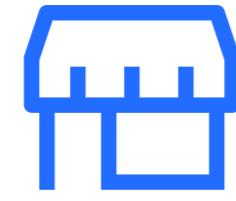
**\$121.6 billion** lost to theft

- **\$461.86** average theft per incident
- **2%** of shoplifters are caught.<sup>5</sup>

# Use Cases for Retail



# Built for Retail Industry Use Cases



## Real-Time Inventory Visibility

Stream inventory data in real time to ensure accurate, channel-wide stock visibility.

Gain accurate, real-time insights across all stores and distribution centers to:

- Ensure reliable fulfillment and on-time delivery
- Meet customer expectations for product availability
- Maximize revenue by preventing stockouts and delays



## Personalized Promotions

Trigger instant, tailored offers using real-time behavioral data.

StreamSets enables real-time analysis of user interactions – fueling hyper-personalized, timely promotions that

- boost revenue
- streamline operations
- elevate customer satisfaction.



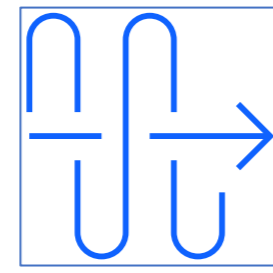
## Real-Time Loss Prevention

Real-Time Loss Prevention  
Detect and act on suspicious activity with live data streams.

IBM StreamSets unifies real-time data from transactions and sensors to instantly detect anomalies and trigger automated alerts – enabling faster, proactive loss prevention

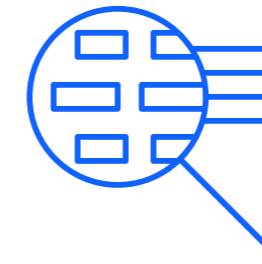
# How does IBM StreamSets provide value?

Build intelligent, reliable, [streaming data pipelines](#) across hybrid-cloud environments, at scale. Improve real-time decision making and reduce the risks associated with data flow across an organization



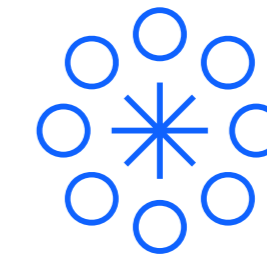
## Simplified pipeline creation

Design and deploy streaming and CDC pipelines in minutes with an extensible low/no-code UI – leveraging 200+ prebuilt stages to accelerate development



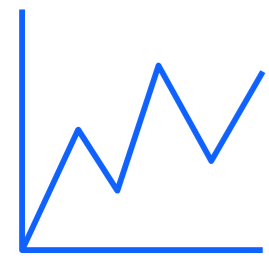
## In-flight data transformations

Transform, enrich, and cleanse data in motion using prebuilt quality operators or custom code—ensuring clean, reliable data during pipeline execution.



## Best-in-class developer experience

Features like preview, validation, version control, and clear error messages—backed by robust documentation—simplify workflows and accelerate productivity.



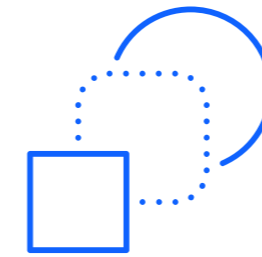
## Programmatic approach to pipeline creation

Build and deploy pipelines programmatically with a powerful Python SDK—fully integrated with the UI to ensure bulk changes sync effortlessly across teams.



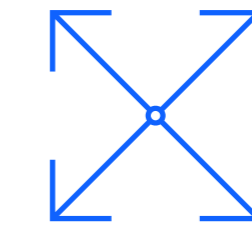
## Pipeline observability

Monitor and alert on StreamSets jobs with the integration of IBM Databand—detect data issues early, resolve them faster, and ensure high-quality, reliable data.



## Data Drift Protection

Shield pipelines from unexpected changes in data structure, meaning, or infrastructure—ensuring stability while enabling agility and innovation across your organization.

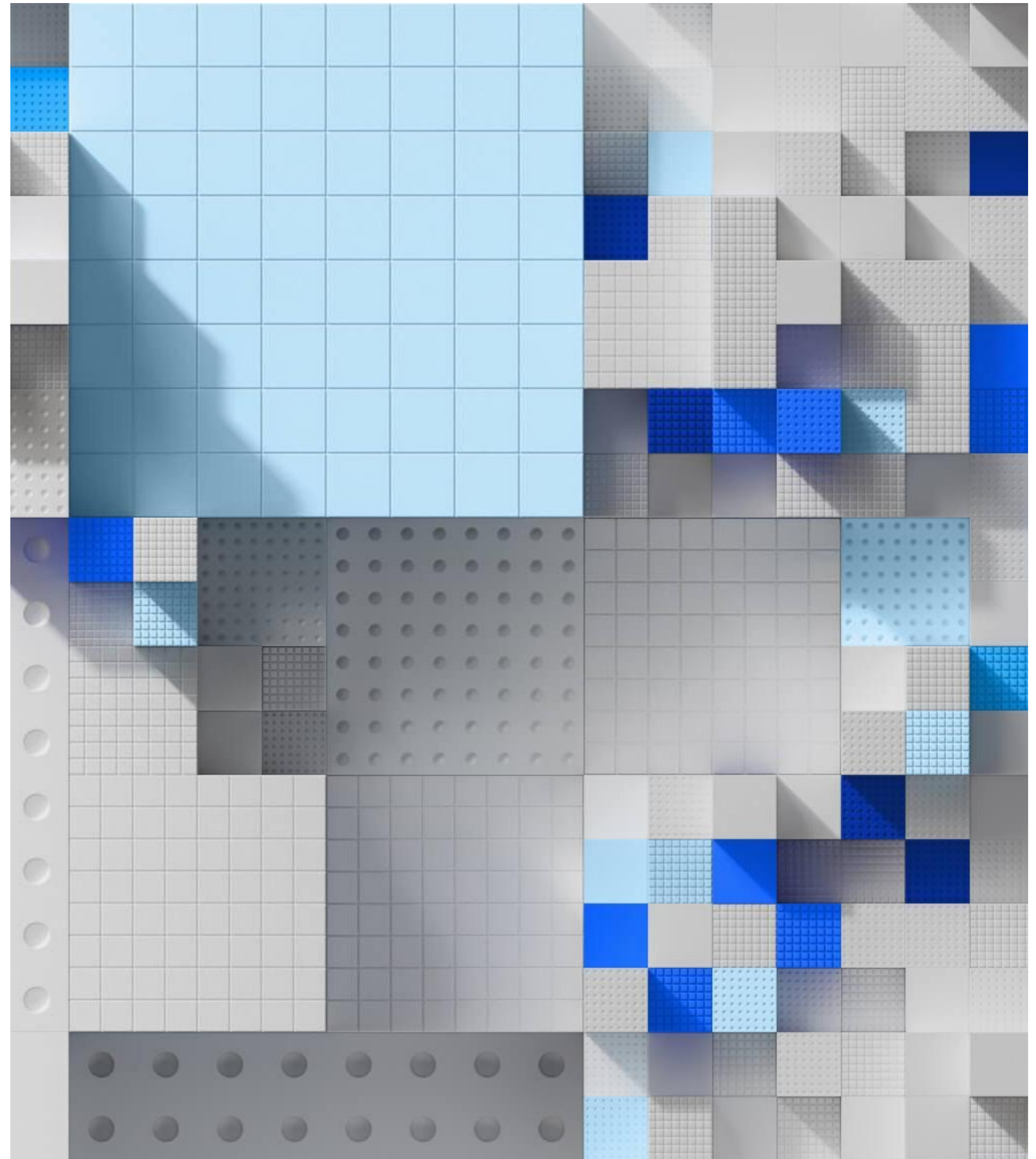


## Flexible deployment

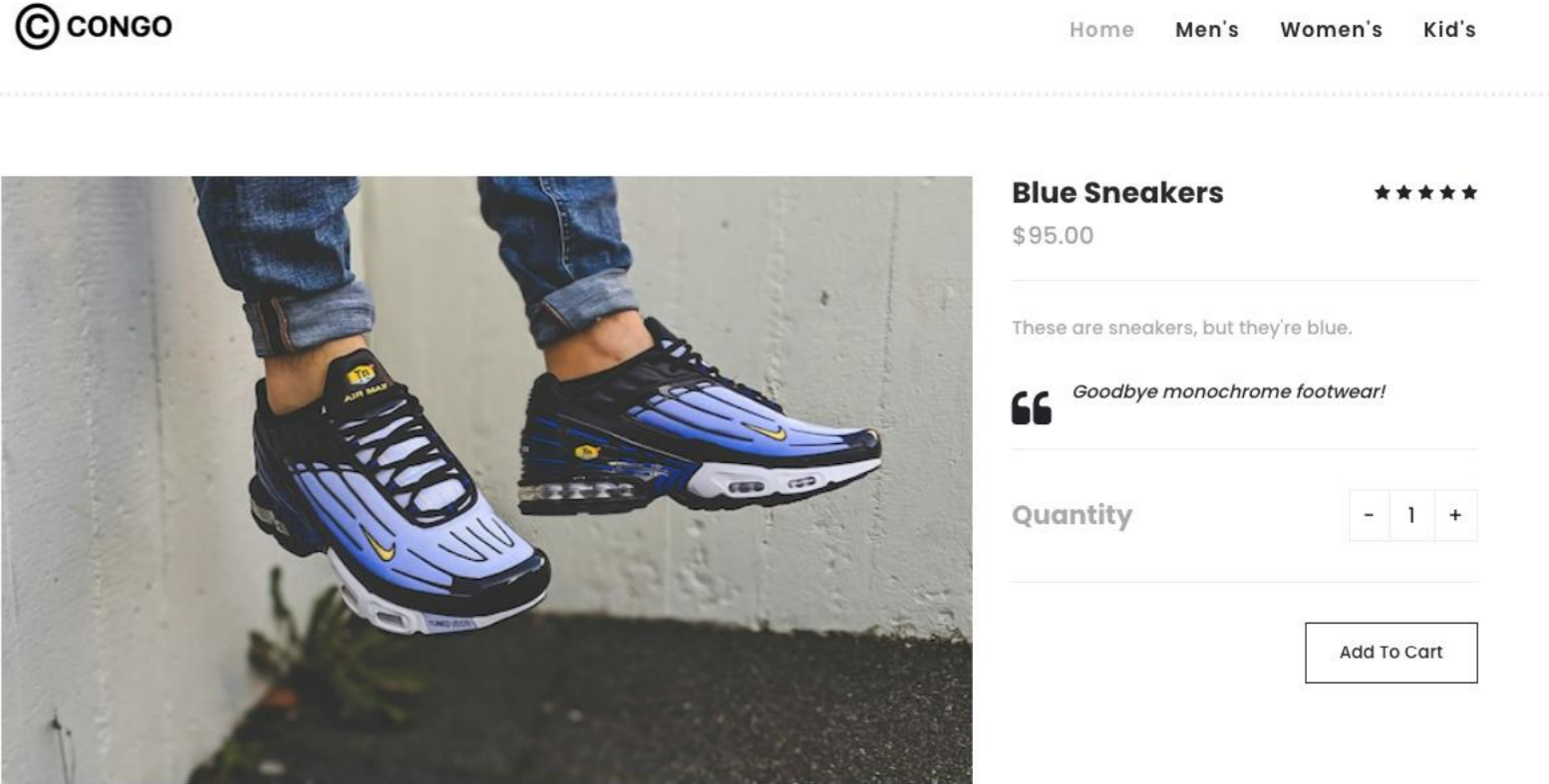
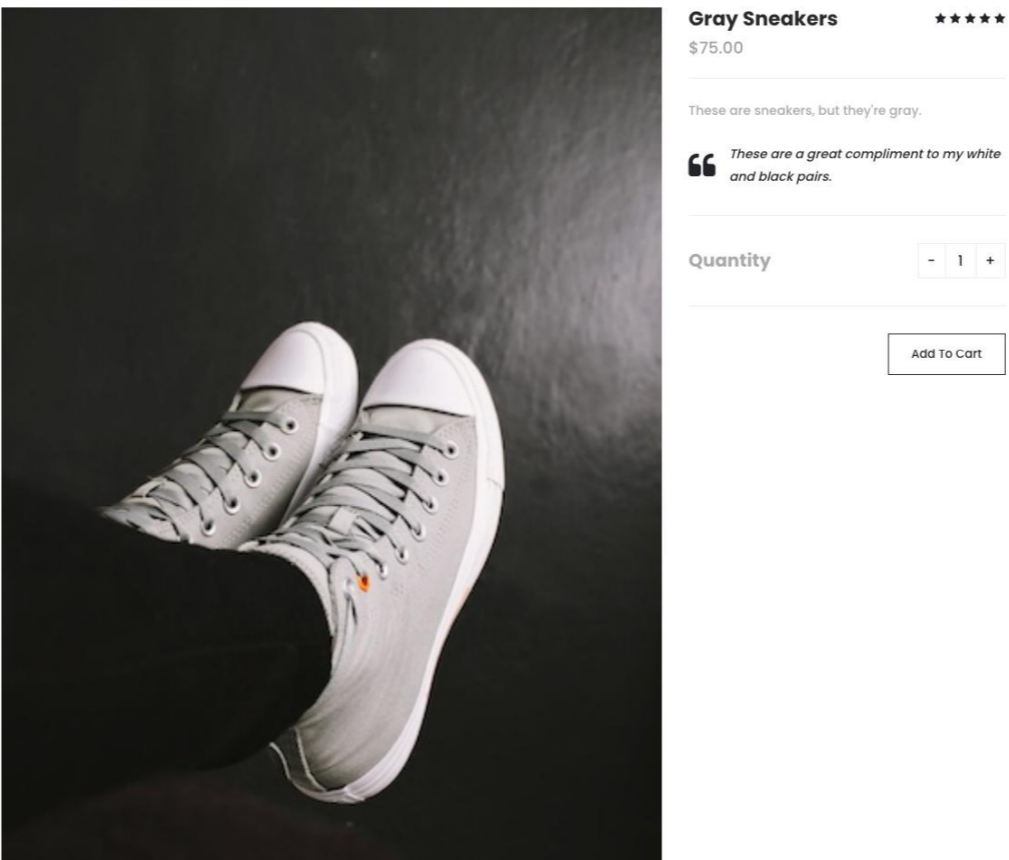
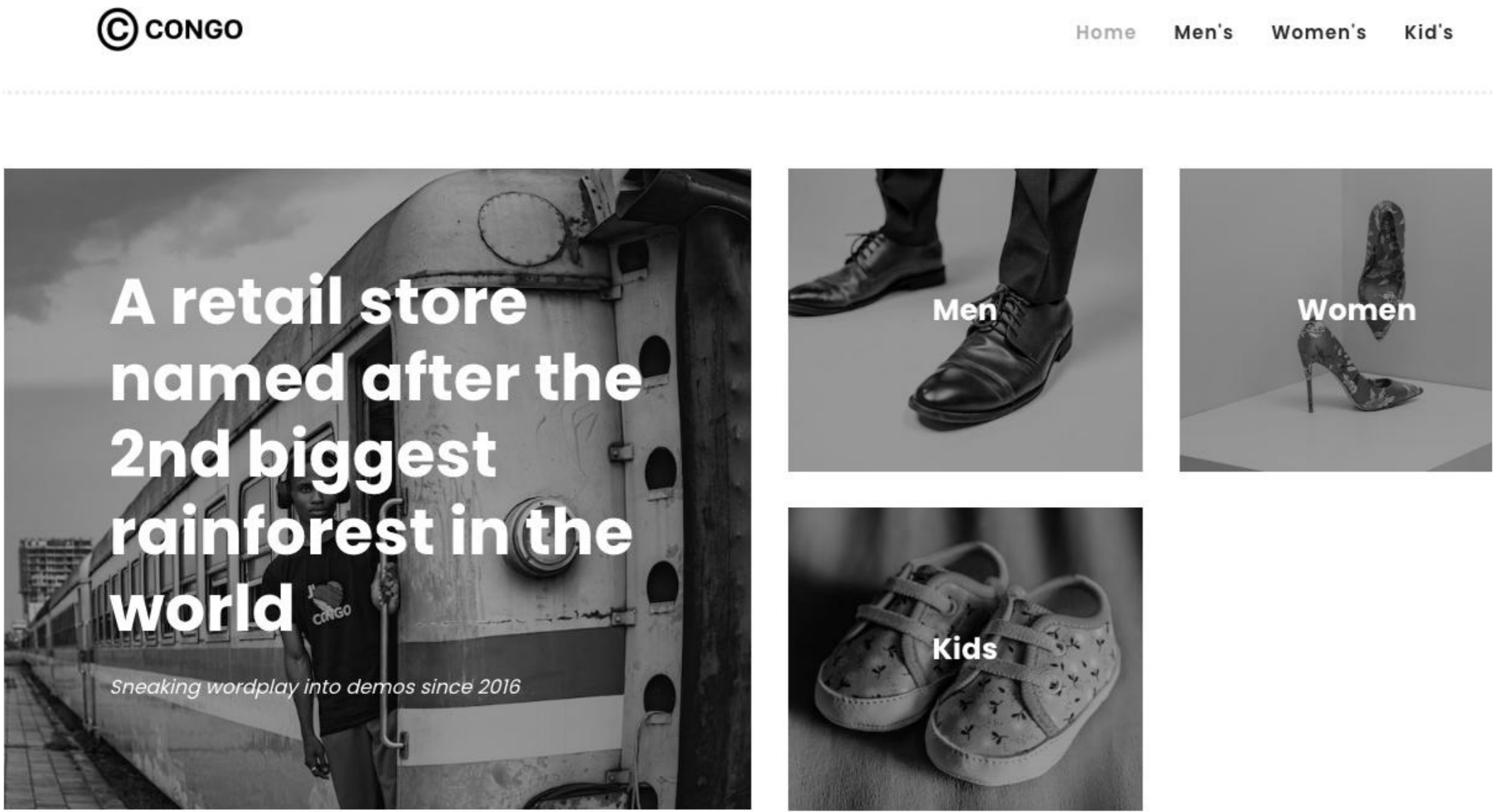
Execute data pipelines wherever they reside, in any cloud, geo, or on-prem - minimizing egress/ingress, latency, and security risks.

# Real-Time Customer Offers

with IBM StreamSets



# Example: Raj's Shopping



- Customer Interacts with Retail Website
- If the customer looks at the same item more than once in 30 seconds, offer them a discount

# Pipeline



Transaction Stream



Lookup and Apply Business Rules



Interaction Store

**Gray Sneakers** ★★★★★  
\$75.00

nonce: {STRING} "7795bdfd33566637cdd51d95096edafa"   
username: {STRING} "chaucer"   
url: {STRING} "https://127.0.0.1:8443/item-3"   
timestamp: {STRING} "2025-07-07 17:50:03.403215"

**Gray Sneakers** ★★★★★  
**\$69.99**

# The Algorithm

1. For each website interaction, query the nonce, username and URL of the product page the customer is interacting
2. Make sure that the timestamp of customer interaction is greater than 30 seconds
3. Move Customer in Repeat Access Discount program
4. Add Expiration Date of the real time off as 30 seconds from now

```
SELECT COUNT(*) + 1 as access_count_30s
FROM clickstream
WHERE nonce != "${record:value('/nonce')}}"
AND username = "${record:value('/username')}}"
AND url = "${record:value('/url')}}" AND timestamp > UTC_TIMESTAMP() - INTERVAL 30
SECOND;
```

Output Field * ⓘ	Field Expression * ⓘ
/program	repeat_access_discount
/expiration	\${time:dateAddition('SECOND', 30, record:value('/clickstream_utc_timestamp')) }

# Summary

# Summary

- Built a real-time customer offer pipeline using IBM StreamSets
  - Ingested transaction stream from Kafka in real-time
  - Enriched records using queries to SingleStore
  - Flagged `program field` records using business rule logic
  - Wrote transactions to SingleStore
  - Which allowed the store website to offer real-time discounts

