

BENEFITS



Discover
Detect sensitive data upon ingest.



Secure
De-identify, route, filter and quarantine in stream.



Govern
Set central policies and audit compliance.

PROTECT DATA UPON INGESTION

- De-Identify
- Encrypt
- Generalize
- Hash
- Mask
- Randomize
- Scramble
- Tokenize

StreamSets Data Protector™

Discover, secure and govern sensitive data as it arrives.

For any company dealing with personal or sensitive data or struggling with regulatory compliance, StreamSets Data Protector provides software as a service to discover, secure and govern movement of sensitive data as it arrives from a source or moves between compute platforms. Unlike most solutions that analyze data at rest — after it has landed in a data store — Data Protector detects and deals with the data as it is ingested.

Handling sensitive data as it arrives helps companies comply with regulations such as GDPR or HIPAA, but can be applied to any data patterns the customer desires. Data Protector enables inspection of incoming batch and streaming data to discover sensitive data by comparing it to built-in patterns such as national ID, tax ID or driver license numbers, bank account or credit card numbers, or IP addresses. Customers can also add their own data identification rules and adjust the strictness of the pattern matching from exact match to varying levels of fuzzy matching.

Without the automation Data Protector provides, laborious hand-coding is required to continuously check each data source against dozens or hundreds of PII patterns. This approach becomes impossible as unstructured data is added to the mix, and data drift — unexpected changes to the structure and semantics of the incoming data — takes hold.

Data Protector also comes with central policy management to simplify the standardization of how sensitive data is handled across the enterprise. It includes support for Security Zones to segment policies as well as audit reporting to verify proper implementation. Lastly, Data Protector also integrates with systems for protecting data at rest to provide a complete solution.



StreamSets Data Protector includes transformations such as a Field Anonymizer that can obfuscate, mask or generalize sensitive data as it is ingested.

FEATURES

Discover

- Easily build PII monitoring into StreamSets pipelines.
- Continuously detect sensitive data in batch or streaming sources.
- Test against built-in PII patterns or create custom regular expressions.

Secure

- Obfuscate data using reversible and irreversible algorithms.
- Generalize and de-identify data to allow useful analysis.
- Integrate with leading metadata catalogs for data at rest.

Govern

- Centrally establish robust policies for consistent application of rules.
- Define Security Zones for varying levels of access.
- Utilize audit reports to monitor compliance.

The StreamSets Data Operations Platform is designed to simplify the entire dataflow lifecycle, including how to build, execute, operate and protect enterprise dataflows at scale. Developers can design batch and streaming pipelines with a minimum of code, while operators can aggregate dataflows into topologies for centralized provisioning and performance management.

ABOUT

StreamSets is headquartered in San Francisco. Our mission is to help enterprises harness their data in motion. StreamSets software is in use at hundreds of organizations and we're backed by top-tier Silicon Valley venture capital firms, including Accel Partners, Battery Ventures, Ignition Partners, and New Enterprise Associates.

LEARN MORE

Get up and running with StreamSets in minutes. Visit us at:

www.streamsets.com