INTEGRIS SOFTWARE

Integris Data Privacy Automation | Data Handling Solution Brief

Hadoop Data Lake Integrity. **Enabled.**

Unlock your data lake in a way that protects customer privacy and fuels Innovation.

Regulations like GDPR and the California Consumer Privacy Act (CCPA) are forcing companies to lock down their data lake for fear of missuese. With Integris Software, you can finally use your data without fear.

Integris Software will discover and calissify sensitive data across your data lake, map it back to data handling obligations, indentify violations and automate actions. Integris` unique set od capabilities enable regulatory hydiene, while maintaining the productivity of your data lake.

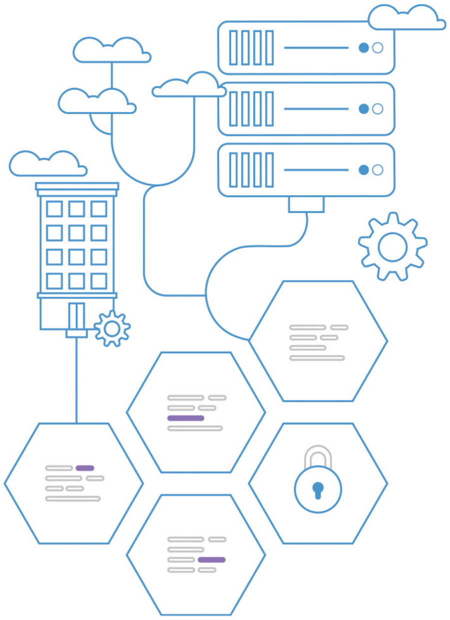
- Meets all the challenges of volume, variety and velocity for data lake hygiene.
- Multiple data scanning options such as complete scans, reservoir sampling, and random sampling.
- Continuosly classifies, labels, and maps your sensitive data to your retention and encryption policies.
- Multi-Label Classification to identity combinations of data that alone are benign but together become highly sensitive.
- Data in-motion discovery and classification as data streams in and out of your data lake.

CUSTOMER SPOTLIGHT

> “Data privacy automation is needed to visualize where personal information is located across the organization`s geographical footprint, prove adherence to regulatory standards, and empower strategic decision making.”

Xavier Quintuna
Principal Big Data Architect

orange™



Data Lake Success Factors, How Inegris Software Helps, and Customer Highlights

Adhering to the Data Handling Best Practice of De-identification

Importance to Data Lakes

Data lakes ingest disparate pieces of cutomer data from a variety of sources. When combined, this data has the potential to reveal customer identities along with highly sensitive personal information.

De-indentification prevents data analysts from connecting an individual to their personal information. This enables the data analysts to access useful data without compromising customer privacy.

How Integris Helps

Integris` deeper inspection down to the data element level informs you exactly what`s in your data lake, not just what the metadata implies.

Integris discovers and classifies your data and surfaces which sensitive data can be tied back to individuals.

Integris Multi Label Classification identifies highly sensitive combinations of data across your data lake.

Integris Customer Highlight

An online retailer wanted to retain and perfom analytics on customer buying history, geography, and gender.

Using Integris, the retailer was able to validate that the dataset couldn`t be used to identify and individual.

87% of the US population can be identified using only their Zip Code, Gender, and Birthdate.*

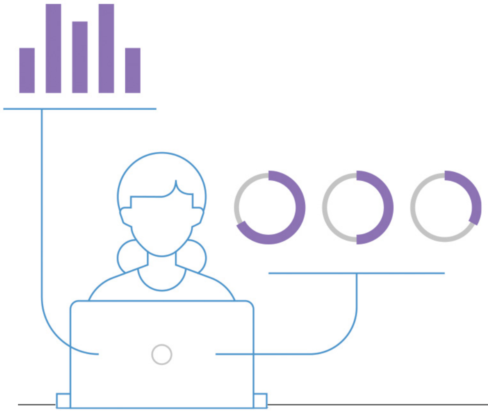
*Source: https://dataprivacylab.org/projects/identifiability/paper1.pdf

Validating Encryption Policies

Importance to Data Lakes

Regulations and internal policies require encryption to preserve the confidentiality and integrity of sensitive data.

The massive volume of data in your data lake render traditional encryption validation tools obsolete.



How Integris Helps

Integris continuously classifies, labels, and maps your sensitive data to encryption requirements emanating from regulations or your internal use policies.

During the mapping process, Integris will flag encryption violations and automate actions to remediate issues. For examples, Integris can kickoff workflows with your existing ticketing system.

Integris Customer Highlight

A financial servicies company separates their data in their data lake into “Open” and “Restricted” tiers in order to control and limit access to sensitive data. The financial services company uses Integris to identify any unencrypted sensitive data in the Restricted Tier.

As Integris discovers unencrypted sensitive data, it triggers an event in the company`s encryption software to encrypt the data.

Integris` ability to label at the data element level enables Integris to pass detailed metadata such as the data`s location which helps automate a process in which the encryption software encrypts the data.

Once this process is complete, and Integris can`t identify any more encryption violations, the data is then moved to the Open Tier for the wider use.

Validating Encryption Policies

Importance to Data Lakes

Regulations and internal policies specify how long datasets should be retained.

The massive volume of data in your data lake makes it impossible to manually validate controls around data retention.

Since data is flowing into your data lake from disparate sources there`s a good chance you`re calculating retention in different ways (e.g. creation data, date of last transaction, or other metric).

How Integris Helps

Integris alerts you to retention issues across your data lake.
You can view sensitive data by geography, system, and policy, as well as snapshots of the amount, type, and recency of data.

Integris can adapt to any method you`re using to calculate retention, then flag retention issues for human intervention or kickoff workflows with your existing ticketing system.

Integris Customer Highlight

A retailer has a five-year retention policy on sensitive data contained in customer service chat record.

They use Integris to discover and classify the unstructured chats in their large data lake. Integris automatically maps this data to the five-year retention obligation and reports any issues.

Inspecting Data in-Motion at Points of Ingress and Egress

Importance to Data Lakes

Most organizations don`t have an effective way to identify and monitor sata streams to ensure they are adhering to all data handling policies.

When the data that streams in and out of your data lake is a blind spot, it can cascade into multiple problems.

For example, if you have data sharing agreements, you`ll want to know that inbound data is clean and limited to what`s in your contracts. You`ll also want to monitor outbound data, so it doesn`t fall out of scope, and leak private information into a data sharing pipeline.

It`s good hygiene to monitor the data being ingested into your data lake. Better to catch inappropriate or problematic data early so you can manage it, and before contamination sets in. Classifying data as it enters the data lake can ensure continued compliance with business obligations.

How Integris Helps

Integris taps directly into the message queue that is processing data as it moves in and out of the data lake to provide discovery and classification services on data in-motion.

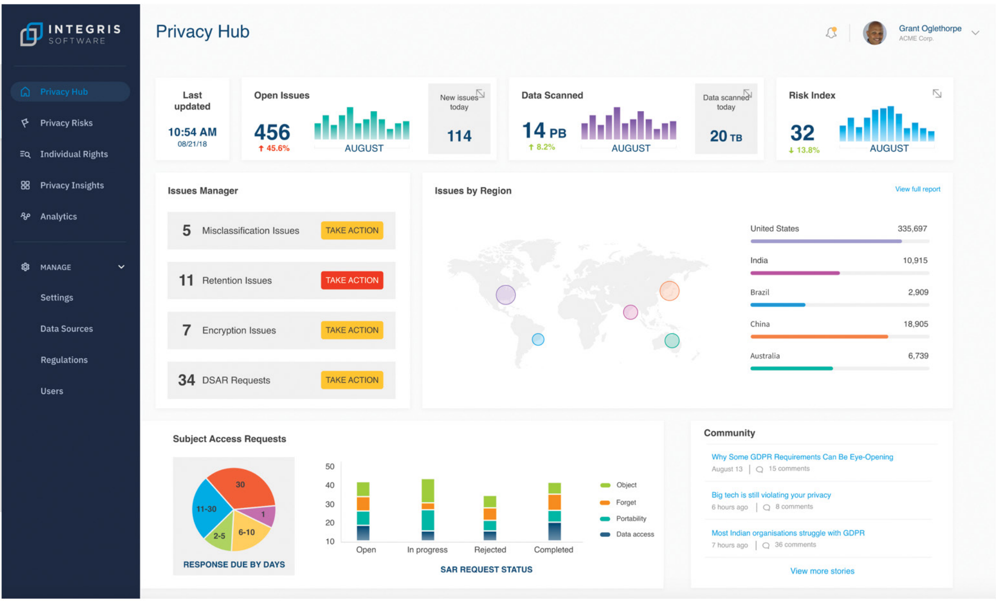
Integris` ability to process data in-motion is key to helping you understand which data is entering or leaving your organization via data sharing agreements, and the streams and feeds your company relies on for continuous innovation.

Integris integrates into popular streaming analytics tool such as Apache Kafka, AWS Kinesis, and more.

Integris Customer Highlight

A technology company has an inbound data sharing agreement with a logistics company for shipping and return processing.

The data is transmitted via Kafka, and they use Integris to scan the data in-motion at the Ingress point to validate that it only includes customer names and addresses, and that no other sensitive data can be tied to the customer.



Get your CTO, CISCO, CDO, and CPO on the same page with your Integris Data Privacy Hub.

Empower your security, privacy, and data governance leaders to make fact-based decisions about the use and transfer of customer data.

Assess Data Risk Prior to M&A Transactions

Importance to Data Lakes

As data is acquired through the M&A process, data lakes cand become contaminated with unexpected, inappropriate, or problematic data.

Due diligence should include the inspection of the data being acquired. This allows organizations to properly evaluate the rick prior to merging large datasets.

How Integris Helps

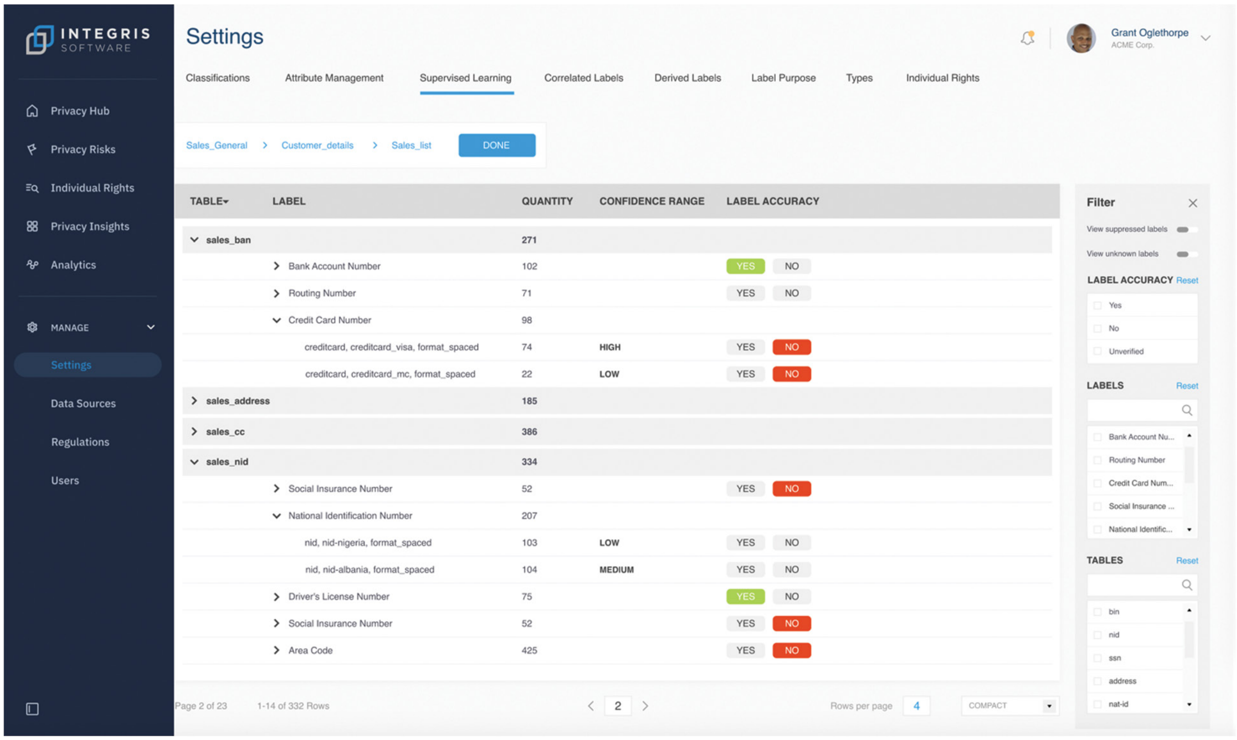
Integris helps maintain data lake hygiene by scanning the data lake to ensure that both the data in the lake and the data acquired by M&A activity is consistent with what is expected.

An effective method is Integris`s ability to sample data already in the data lake, sample data at the Ingress point as data is merged into the data lake, and even an ongoing scanning schedule to insure continued compliance.

Integris Customer Highlight

A Travel & Expense company was acquiring a smaller company. The target company attested that there was no sensitive data in their data lake. The Travel & Expense company used Integris to sample the target company`s data lake to identify sensitive data.

The Travel & Expense company was surprised to learn that the data lake included highly sensitive behavioral information about customers.



TABLE▾	LABEL	QUANTITY	CONFIDENCE RANGE	LABEL ACCURACY
sales_ban		271		
	Bank Account Number	102		YES NO
	Routing Number	71		YES NO
	Credit Card Number	98		
	creditcard, creditcard_visa, format_spaced	74	HIGH	YES NO
	creditcard, creditcard_mc, format_spaced	22	LOW	YES NO
sales_address		185		
sales_cc		386		
sales_nid		334		
	Social Insurance Number	52		YES NO
	National Identification Number	207		
	nid, nid-nigeria, format_spaced	103	LOW	YES NO
	nid, nid-albania, format_spaced	104	MEDIUM	YES NO
	Driver's License Number	75		YES NO
	Social Insurance Number	52		YES NO
	Area Code	425		YES NO

Page 2 of 23 1-14 of 332 Rows

2

Rows per page 4 COMPACT

Filter

View suppressed labels

View unknown labels

LABEL ACCURACY Reset

□ Yes
□ No
□ Unverified

LABELS Reset

□ Bank Account Nu...
□ Routing Number
□ Credit Card Num...
□ Social Insurance ...
□ National Identific...

TABLES Reset

□ bin
□ nid
□ ssn
□ address
□ nat-id

Machine learning is tuned to personal information, making it more accurate than broas-based data mapping tools.
You can even provide feedback on data labeling to further tune machine learning accuracy.

Integris Software Hadoop Technical Advantages

Integris Software`s secure, scalable, microservices architecture meets the demands of petabyte-scale processing. Flexible, hybrid deployments go where your data resides, minimizing costs and deployment friction, while maximizing processing power efficiencies.

- Multiple data scanning options including complete scans, reservoir sampling, and random sampling of large tablets with the option to set maximum counts (e.g. up to 30% capped at 2,000,000 rows).
- Integris hadoop Connection Engine sits as an adjacent edge node inside your Hadoop cluster and utilizes existing computer power.
- Integris Multi-Network Zone Installation means the data residing in HDFS will be processed on the same Hadoop cluster.
- Ability to work between network zones. For example, even if your Hadoop cluster resides in a VPN/secure area, Integris containers can still be deployed and communication across different network zones via 80/443.
- Cataloging process uses Hive/HCatalogue information for automatic data discovery.
- Jobs can be scheduled on existing YARN queues or choose intelligent job scheduling between queues to maximize YARN scheduling capacity.
- Post-process, Integris only stores aggregated results (in MBs) in HDFS.
- Integration into popular streaming analytics tools such as Apache Kafka, AWS Kinesis, and more.
- Support for all popular file formats including Parquet, OCR, Avro, RC, Sequence file, Delimited, and more.

ABOUT INTEGRIS SOFTWARE

Integris Software, the global leader in data privacy automation, helps enterprises discover and control the use of sensitive data in a way that protects privacy and fuels innovation.

Privacy is now critical to an effective data protection strategy. By sitting upstream from security, Integris tells you what data is important and why so you can be precise in your InfoSec controls.

Integris works securely, at scale, no matter where sensitive data resides. You get a live map of your sensitive data where you can apply policies, surface issues, and automate remediations via your broader ticketing and InfoSec ecosystem.

Regulations like GDPR and the California Consumer Privacy Act (CCPA) are triggering knee-jerk reactions as companies lock down their data for fear of misuse. With Integris, there is finally a way to use your data without fear.