## HITACHI Inspire the Next

## DATASHEET

# Manage Your Sensitive Data With an Al-Driven Data Catalog and Integration

Not complying with government regulations can be cost prohibitive. The European Union's General Data Protection Regulation (GDPR) fines can be up to 4% of global revenue for the corporate group (including subsidiaries and parents). In the U.S., the California Consumer Privacy Act (CCPA) fines are \$7,500 per violation for every person whose data you store. Violations can quickly add up to millions in penalties, not to mention both personal risks and corporate risks to company brand sentiment, business and intellectual property.

Enterprises constantly face pressure to provide self-service data access to drive business agility and outcomes. They are often forced to balance this with governance and compliance. Are you being asked to make this forced tradeoff?

Data Services from Hitachi can help you manage your sensitive data challenges and fuel data-driven innovation, as you reduce risk and cost.

**Data Catalog** is an Al-driven data catalog that automatically detects, tags and secures sensitive data across the enterprise, lowering the cost of regulatory compliance and reducing the risk of financial penalties. It also supports rapid response to evolving regulations, such as GDPR and CCPA.

• Automatically identify and secure sensitive data as it flows into or is created within your enterprise data lake.

- Track metadata required for regulatory compliance, such as intended use, content type and retention period for each dataset, to eliminate violation risk.
- Integrate your existing data security and privacy solutions for access control, masking and encryption.
- Gain insight into how much data has been cataloged, some discoveries we've made, and how your teams are using it. (see Figure 1).

Pentaho Data Integration, part of the Lumada portfolio, is a modern data integration and orchestration platform to access, prepare and blend data from all sources. It delivers analyticsready data to end users faster with visual tools that reduce time and complexity. Without writing SQL or coding in Java or Python, organizations immediately gain real value from their data from sources like files, relational databases, Hadoop and more, which are in the cloud or on premises.

- Simplify integration and blend new with existing data.
- Reduce time and complexity of data integration across big data sources.
- Realize broad connectivity to a variety of diverse sources: structured, unstructured and semi-structured.
- Employ rich data profiling capabilities and data quality operators.

# Choose Hitachi's Lumada portfolio to handle your sensitive data challenges and:

- Unlock data-driven innovation within your organization.
- Address regulatory compliance and reduce corporate risks.
- Reduce management costs with automation.

Figure 1. Sensitive data discovery and tagging using Data Catalog

O DataOps Dashboard	Home Browse Gloss					
Data Assets	Fields	Tag Associations	Metadata Objects	Community Contribution	Users	Searches
144 80 Secultive	28 459 Sensitive	1.23	148	173	1	122
atalog						
Data Assets   144						Data Soc
Non-densities 32   22.9176	😝 Secolabor 201   54.20005 🛛 🔕 Sarpe	races and servicine 6   0.000% 🛛 🔕 the service of 22	15.278%		8	
Field   2k						
	_					
Sensitive 459   22.881%	Non-Sensitive 211   10.510% 🔕 No					
Tee   110						Domains 10 Association
Tag   110					_	Domains 10 Association
	1 Used 73   66.364% 🛛 🛑 Unused 37   :	22.636%			_	Domains 10 Association
		23.436%.)				Domains 10 Association
Sensitive Data As	sets   89					Domains 10 Association
Sensitive Data As						Domains 10 Association
Sensitive Data As	sets   89			Community		Dorsains 10 Association
Sensitive 76   68.1824 Sensitive Data As	sets   89		8	Community		Dorsains 10 Association
Sensitive 75.1 64.100 Sensitive Data As ights2156.427k Metadata Objects	sets   89 dom.31   37.0776 🥚 Low 4   4-664	Usage	8 122	Curated Tag		
Environe 75   63 125% Sensitive Data As Sensitive Data As Sensitive Data As Metadata Objects Taps	sets   89 dum 20   37.0075 • av 4   4.600 110	Usage Users		Curated Tag	erty Assignments	77
Sensitive Data As Sensitive Data As Herica (Sectors) & M Metadata Objects	sets   89 down 32   37507% • Low 4   4-694 110 2	Usage Uters Searces	122	Curated Tag Custom Prop	erty Assignments 6	77 6
Exercise 75   63.100% Sensitive Data As Sensitive Data As ters2   56.02% Metadata Objects Tags Custom Properties Diak Objects	sets   89 down 22   37.077% • Love 4   4.4041 110 2 4	Usage Uters Searces	122	Curated Tag Dustern Prop Conversation	erty Assignments 6 teviews	77 6 2
Sensitive Data As Sensitive Data As • tento ( ) 54.02% • M Metadata Objects Tap: Custom Properties Data Objects Raies	sets   89 dom.22 370796	Usage Uters Searces	122	Curated Tag Custern Prop Conversation Ratings and	erty Assignments 6 teviews	77 6 2 8
Sensitive Data As Sensitive Data As Metadata Objects Tap: Cuton Interestis Data Objects Rule: Generated HM tables	sets   89 sum 21 (2007). • Lon 4 ( 4.60) 110 2 4 6 7	Usage Uters Searces	122	Curated Tag Custern Prop Conversation Ratings and	erty Assignments 6 teviews	77 6 2 8

## **Answer Sensitive Data Concerns**

The solution brings the power of the two products to help users address their sensitive data problems, end to end.

Organizations can now shape all data based on their needs and permission requirements.

- Understand the permitted use of all data.
- Tune exposure to sensitive information. For example, set visibility of the entire social security number, or show only last four digits, or even completely mask it (see Figure 2).
- Show the granular data to some users while summarizing it for others.
- Provide 360-degree views by blending in additional information from other data sources.

#### Figure 2. Mask and blend sensitive data using Pentaho data integration

## WE ARE HITACHI VANTARA

Hitachi Vantara solves digital challenges by guiding you from what's now to what's next. Our unmatched industrial and digital capabilities benefit both business and society.

## **Solution Benefits**

Key benefits of the Lumada solution for sensitive data include:



## Automatic

Al-driven automatic identification and enforcement across petabytes of hundreds of millions of columns.



#### Comprehensive

Go beyond standard PII to more firm-specific elements. For instance, Uber driver ID is sensitive to Uber only: no other companies. No other catalog does this.



### **High Accuracy**

The solution is 99% accurate with Al-based tagging that reduces the need for analyst help to fix the false positives. An entry that looks like SSN but isn't being tagged as sensitive means more work afterwards.



#### Trustworthy

It all comes down to trusting the data. Crowdsourcing with community-based tagging and ratings leads to higher trust.

# Learn More

Use Lumada to address your sensitive data management needs.

 $\rightarrow$ 

## **Hitachi Vantara**

Corporate Headquarters 2535 Augustine Drive Santa Clara, CA 95054 USA hitachivantara.com | community.hitachivantara.com Contact Information USA: 1-800-446-0744 Global: 1-858-547-4526 hitachivantara.com/contact

© Hitachi Vantara LLC 2022. All Rights Reserved. HITACHI and Lumada are trademarks or registered trademarks of Hitachi, Ltd. Pentaho is a trademark or registered trademark of Hitachi Vantara LLC. All other trademarks, service marks, and company names are properties of their respective owners.

HV-CBE-DS-Manage-Data-Cataloging-Intergration-8Jun22-B