

# Manage Your Sensitive Data With an AI-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.

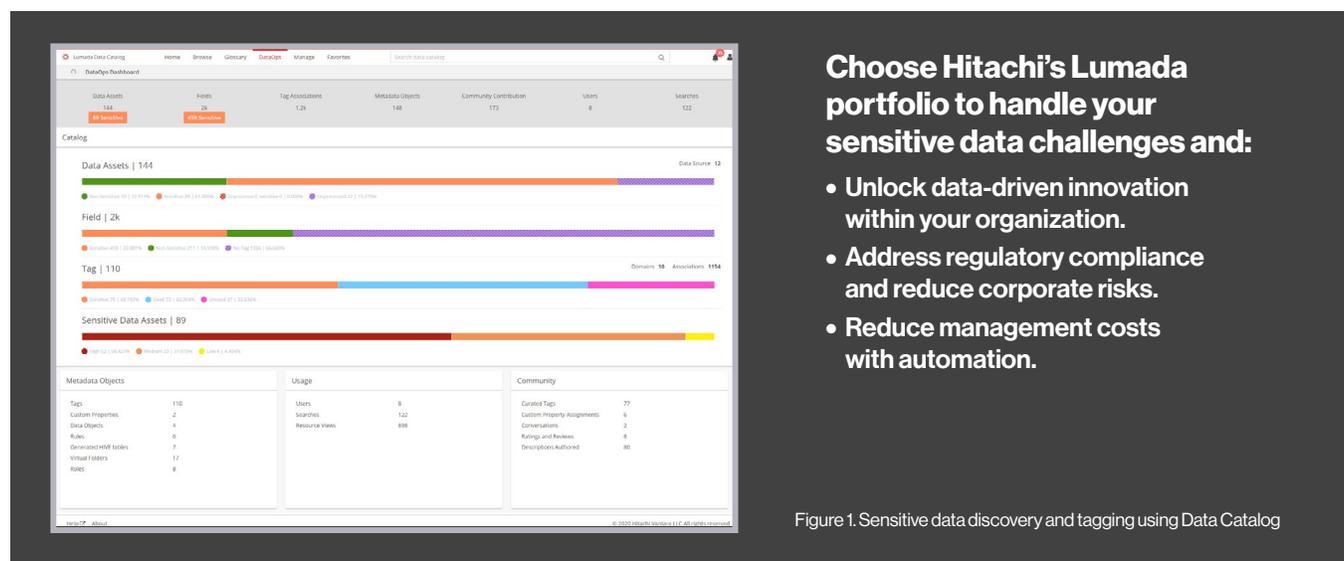
**Data Catalog** is an AI-driven data catalog that automatically detects, tags and secures sensitive data across the enterprise, lowering the cost of regulatory compliance and reducing the risk of financial penalties. It also supports rapid response to evolving regulations, such as GDPR and CCPA.

- 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 analytics-ready 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

### 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:

#### 01

##### Automatic

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

#### 02

##### 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.

#### 03

##### High Accuracy

The solution is 99% accurate with AI-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.

#### 04

##### 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.

### Hitachi Vantara

**Corporate Headquarters**  
 2535 Augustine Drive  
 Santa Clara, CA 95054 USA  
[hitachivantara.com](http://hitachivantara.com) | [community.hitachivantara.com](http://community.hitachivantara.com)

**Contact Information**  
 USA: 1-800-446-0744  
 Global: 1-858-547-4526  
[hitachivantara.com/contact](http://hitachivantara.com/contact)

