



DATA SHEET

Virtusa's Data Policy Automation solution for Google Cloud Platform

virtusa

Using genAI technologies with a data governance tool to automate the governance of data policies

With the exponential growth of data, data governance has become critical for business success. Today's data-driven organizations are turning to automation and leveraging technologies such as Google Cloud Platform, Gemini Pro (a genAI LLM framework), and Collibra.

However, these same companies grapple with the following challenges:

- Data policies are often stored in PDF or Microsoft Word formats, making them less accessible and shareable.
- Data policies need to be implemented in data stores or pipelines.
- Determining which policy applies to what data and how to apply the policy.
- Policy enablement can be demanding and prone to errors.
- Data policies need to be analyzed regularly for any updates.

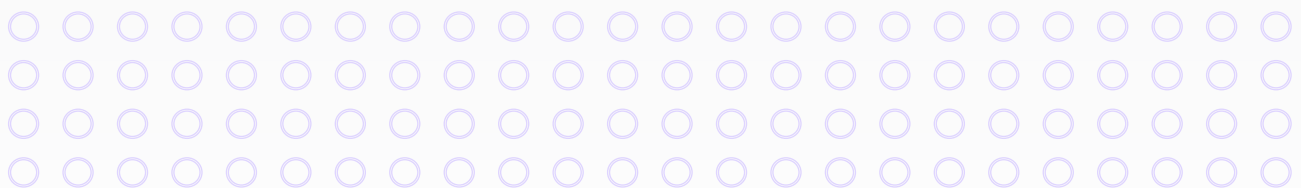
Virtusa's Data Policy Automation solution is perfect for enabling data policies and keeping them up to date where they matter most—in the data stores and pipelines.



Why clients are choosing Virtusa

Virtusa's Data Policy Automation solution is part of a broader intelligent data governance offering that will help clients make the most of their data assets and govern them well using a comprehensive, practical approach. This solution provides fast, automated, holistic data governance across the enterprise.

- Key benefits for our customers:
- Quicker time to deploy data governance initiatives
- Cost savings due to manual task automation
- Increased business and governance productivity
- Better usability and improved adoption of data governance tooling and initiatives





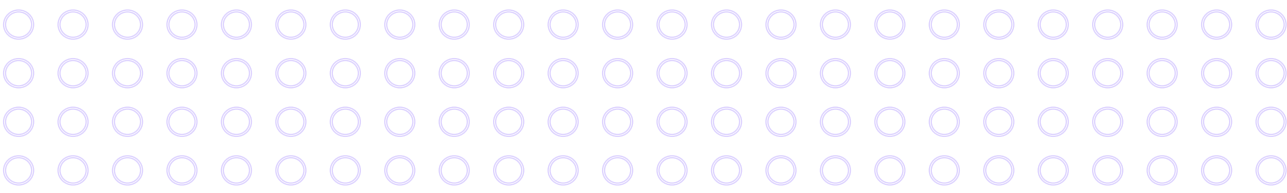
Capabilities

Modernize and automate your data governance platform with Virtusa's Data Policy Automation solution.

By leveraging the scalability of Collibra Data Governance Cloud and Google Cloud Platform with Gemini Pro, the Data Policy Automation solution can:

- Automatically catalog the data policies (i.e. retention, protection, privacy, etc.) into Collibra, extracting metadata from files using Gemini Pro's genAI
- Automatically link data policies to technical metadata based on algorithms
- Automatically enable the data policies across data stores or pipelines
- Automatically analyze data policies regularly for any updates
- Efficiently govern data policies from Collibra
- Trigger database events based on active metadata

The user has a web interface to upload files and content data policies and then ingest the extracted metadata into Collibra. After a governance process involving a few personas, the enablement can be triggered and applied to data and applications.



Any types of assets & metadata

Instead of policies, the process could generate other business assets such as business term, glossary, business process, business dimension, etc. It could scan any type of unstructured file format: word docs, chats, images, confluence's pages, etc.

Step 1

This process identifies and extracts from the pdf files the **policy's metadata**. The gen AI is must be trained to understand a wide variety of formats (iterative process)

Data policy generation process

Gemini

python

Adaptive config

Policy's metadata

- Name of the policy
- Description of the policy
- Lineage with other data assets

GenAI inputs/output

- Structured and unstructured files
- Text as instructions to extract the right text

Step 2

Via **Python** scripts using "Collibra API's"

Data policy harvest process (Collibra API's & Python)

CSV



JSON



python

Step 3

Data policies are curated - metadata and lineage is available as support for assessments, impact analysis etc.

The "Data Policy Owner" checks and approves the new policies

Data policy curation process (Collibra data catalog/metadata management)

Business Analyst

Data Steward

Data Engineer

Data Privacy Manager

Data Scientist



Collibra

Data Intelligence Cloud

Data catalog

Data lineage

Data quality & observability

Data privacy

Data governance

Collibra protect

Core services

Python layer steps

These steps are scheduled every day to check if there are new accepted policies and data to be delete. It also checks if relationships changed in the metadata model (i.e., new tables or new business dimensions have been added etc.)

Step 4

(Depending on type of policies)

1. Extract "Accepted" policies from Collibra (Via **Python** Scripts using "Collibra API's")
2. Call the **Data Policy Enabler Store** procedure to enable data policy

Data policy extract process (Collibra API's & Python)

Extract "Accepted" policy metadata query language

Call enablement store procedure

python

Step 5

(Depending on type of policies)

i.e. for retention: Data policy enablement via DB Store procedures - "deletion" of table's records subject to the retention policy

Data policy enablement (Database platform)

Google Cloud

Google Big Query

Database store procedures & DB tables

Collibra triggers the execution of data governance tasks when policies have been "Approved" as asset status.

We have seen data automation requests from various global clients, including LBG, AZ, Amex, FAB, Omantel, Odido, and KPN. Many clients are highly interested in such use cases. Therefore, having a more intelligent data governance offering—leveraging the power of Gemini Pro and Collibra—will propel Virtusa to the forefront of the industry.

Our partners



For solution
hosting



For LLM



For database



Streamlit

For Web UI



For framework



For data
governance



As programming
language

To learn more contact us today at marketing@virtusa.com

virtusa

Virtusa is a global product and platform engineering services company that makes experiences better with technology. We help organizations grow faster, more profitably, and more sustainably by reimagining enterprises through domain-driven solutions. We combine strategy, design, and engineering, backed by unmatched expertise at the intersection of industry, business, and technology to generate real-world business impact for clients.

Headquartered in Massachusetts with global delivery centers, Virtusa provides a broad range of services, solutions, and assets, including strategy and design, AI advisory and services, digital engineering, data and analytics, digital assurance, cloud and security, and managed services across industries such as financial services, healthcare, communications, media, entertainment, travel, manufacturing, and technology.

Virtusa is a registered trademark of Virtusa Corporation. All other company and brand names may be trademarks or service marks of their respective holders.

virtusa.com