DATA MINIMIZATION

CHECKLIST, STRATEGIES, AND STEPS

PROTECTO

www.protecto.ai

Minimize Data,

Reduce privacy and security risks!

DATA MINIMIZATION

Privacy principles and privacy laws prescribe companies to collect, process, and store the data that is essential for the purposes for which an organization needs the data. For instance, GDPR requires data collection must be limited to what is adequate, relevant and limited to what is necessary for relation to the purposes for which they are processed (data minimization).

Collecting and storing data increases privacy and security surface area. Hence, instead of saving all available data, data officers must limit data collection to what is relevant, minimizing data and attack surface.

STRATEGIES

Narrow Data Collection: For successful data minimization, you must also narrow down the techniques used to collect the data.

Deidentify Personal Data: Mask personal data or create test/mock data that limit processing and sharing of personal data.

Establish Effective Data Governance: Frequently update metadata information and processes to avoid storing stale and unusable data.

NARROW DATA COLLECTION

Companies need to decide and specify beforehand what data will be adequate, necessary, and relevant to achieve their goal. It is essential to be clear about the kind of data you need and what you don't need at the initial planning stage.

□ What is the intended purpose of the data to be collected?

□ Are there other alternative ways of achieving that purpose without collecting the data?

□ What's the duration it will take to achieve the purpose before discarding the data?

DE-IDENTIFY PERSONAL DATA

Understanding the context of how data will be used is critical in determining the minimization options. For example, in development and testing environments, mock data could satisfy the business need.

□ Understand the environment. Why do we need the data?

□ Can we omit sensitive data for development/testing? Limiting sensitive data from the data storage is the best and easy fix.

□ If not, can mock test datasets that can enable the full range of testing scenarios?

□ If can't be mocked, can we de-identify PII and obfuscate other personal data?

EFFECTIVE DATA GOVERNANCE

To achieve data minimization, companies must get rid of data that has outgrown its lifespan and usefulness. Outdated information should be deleted as they could pose a security threat.

□ Do we have express permission from the source of the data that I am collecting?

□ Do we track the specific purpose for which we collect personal data? Do we know the lifespan of the data assets?

□ Do we periodically review the data that we hold? Do we know who is the data owner/steward?

□ Do we track what data assets are actively used? who uses data assets?

□ Are we regularly deleting the data that we don’t use?

Sign up for a free demo and consultation at www.protecto.ai

BENEFITS OF DATA MINIMIZATION

1 **Compliance with European Union Data Protection Act**
This Act requires that businesses hold information about EU citizens to apply data minimization policies to protect such citizens.

2 **Reduce Attack Surface**
Effort, resources, and time needed to secure the data significantly reduces. Privacy violation risks and possibilities of potential leaks decreases.

3 **Reduced Cost of Data Storage**
Data storage costs money. The less it is, the less it costs to the benefit of the organization. With this in mind, it is advisable that companies only collect relevant data and store it for the duration of its usefulness.

4 **Efficient Data Management**
Data storage and retrieval are more manageable when there is less of it. Builds confidence that the retrieved data is current and appropriate. When requests are sent out, data managers are sure to respond quickly.

5 **Improved Customer Participation**
Analysts can improve their analysis by discovering other interesting data assets that are related to a data asset. You can discover the most used data assets, top use cases and popular data joins.

WE CAN SIMPLIFY AND AUTOMATE

Protecto can help improve your privacy and security posture by simplifying and automating your data minimization strategy.

Automatic
To determine what data is necessary and relevant, our solution analyzes data use and activities.

Supports Variety of Data Sources
Our solution scans across multiple data repositories including big data cloud storage to identify sensitive data. Creates a detailed map of your data. Supports all major structured databases, and big data sources (Hive, HBase, Presto).

AI/ML and Privacy Engineering
Applies AI/ML models and privacy engineering to identify the potential list of data assets that can be deleted/archived.

Designed for Enterprise
Integrates with existing enterprise services such as Active Directory. Our solution can be easily deployed on the cloud (AWS, Azure, GCP) or on-premises.

Sign up for a free demo and consultation at www.protecto.ai