PRIVITAR

Privitar Data Privacy Platform™

Technical Overview

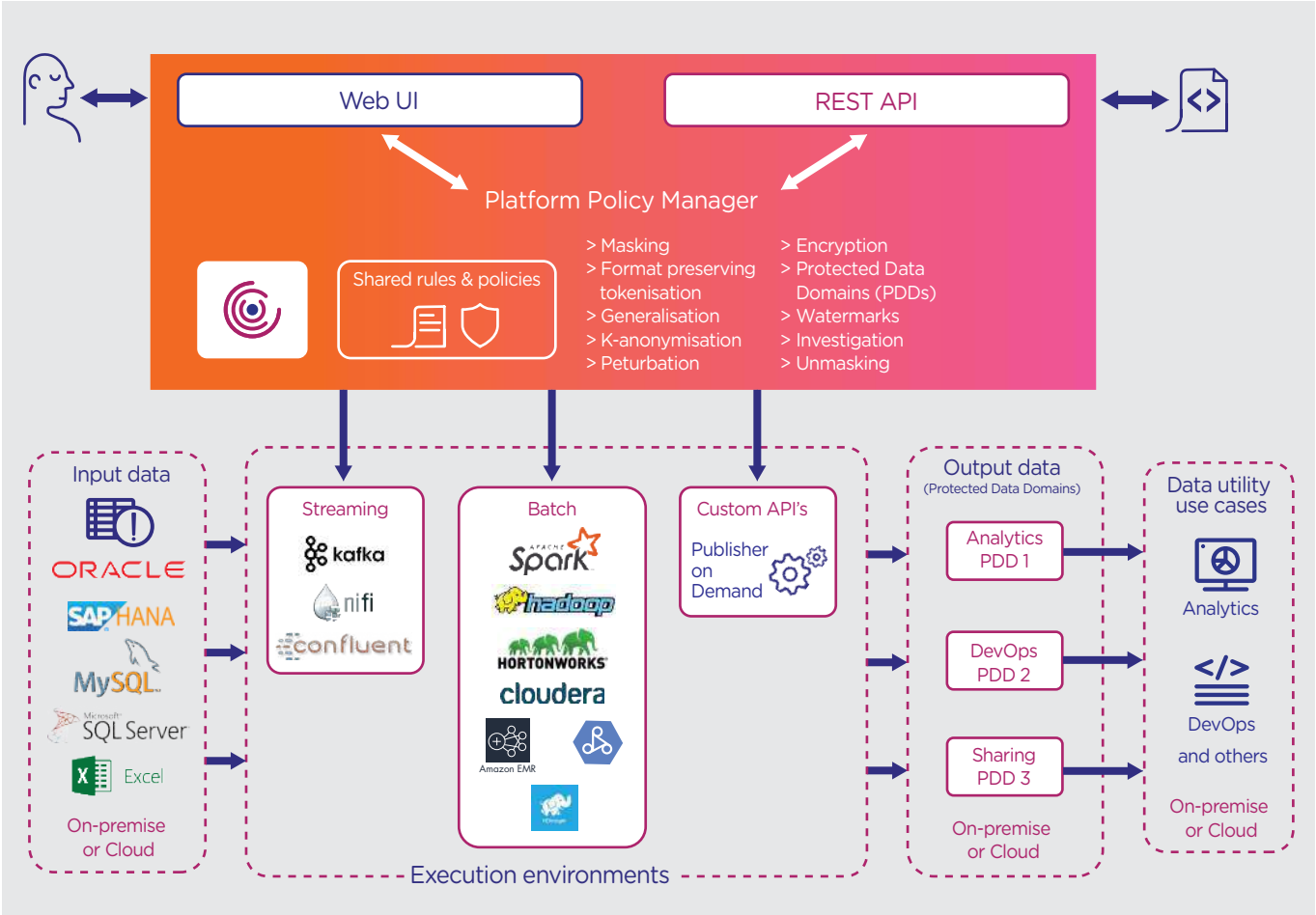
www.privitar.com

Introduction

Organizations increasingly recognize that data is their greatest asset, but an inability to access and use sensitive data due to associated privacy risks is a frequent barrier to innovation.

Technological controls are essential to ensure that the data economy can grow responsibly while sustaining the pace of innovation. Privacy engineering techniques can help companies to protect individual privacy by removing or obscuring the sensitive data through privacy enhancing techniques.

Privitar Platform Logical Architecture



The Privitar Platform is a comprehensive solution for data privacy protection and governance that helps organizations safely extract value from confidential data through a standardized privacy protection approach.

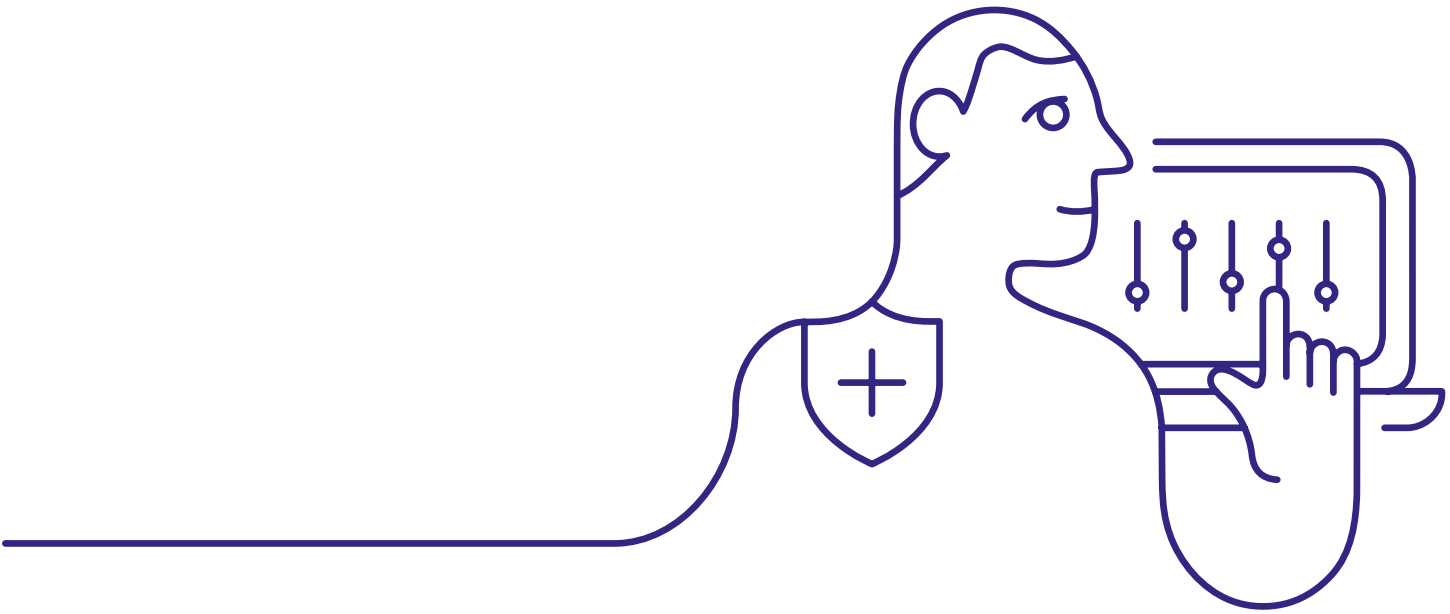
It allows organizations to create de-identified datasets with reduced privacy risk but preserved utility and suitability for analytics and other applications.

Benefits provided by the Privitar Platform

- Build trust with customers by enforcing clear and consistent privacy policies
- Limit exposure and reputational damage in the event of a breach
- Make sensitive data more broadly accessible and combine data assets to increase business value
- Preserve useful patterns in data, retaining structure, format and relationship
- Apply centralized, policy-driven governance via APIs and an easy-to-use Policy Manager interface
- Insert digital watermarks to control and track the distribution of sensitive datasets
- Manage and control the risk and linkability of data releases by organizing them in Protected Data Domains (PDDs)
- Execute consistent privacy policies on a variety of data architectures and processing patterns, such as cloud platforms, big data lakes, hybrid architectures, streaming and batch processing
- Perform advanced statistical transformation to maximize both privacy and utility

Data protected with Privitar is suitable for:

- Analytics, data science and machine learning
- Sharing data with third parties
- System development and testing
- Complying with regulatory standards
- Processing in cloud environments



The Privitar Platform Features In-Depth

Centralized Privacy Policy Management

Data privacy controls need to be highly contextual and dynamic. Privitar Publisher encapsulates privacy constraints as part of a policy that defines the privacy protections applied to the data. You can strike the right balance between privacy and utility by creating custom policies and implementing a risk-based approach to protecting your sensitive data.

Data releases are organized in Protected Data Domains (PDDs), specific to use cases and / or recipients. Data owners can apply data protection controls to PDDs, such as ensuring that datasets published inside the same PDD retain referential integrity and linkability, while not being directly linkable to data published into another PDD. Watermarks can be introduced in published datasets to trace them back to their original PDD in the event of a data breach.

Policies and Protected Data Domain are managed centrally using an intuitive and user friendly web interface or automation APIs. This provides transparency and traceability, ensuring the consistent, repeatable and auditable application of the policies.

Role-based access restricts which users may, for example, author policies, change them, or execute them. By defining custom Roles, each organization can configure Publisher to enforce internal governance requirements and reflect team structures. Different teams can retain independence and ownership of their policies, PDDs, de-identification processes and more, while sharing consistent de-identification rules across the organization. This way, the Privitar Platform allows organizations to conform to and exceed privacy regulations, and to demonstrate compliance.

Designed for Big Data Application

The Privitar Platform is highly scalable. The privacy engineering algorithms are implemented for parallel execution on distributed compute clusters and streaming data flows.

Privitar integrates directly with modern big data and streaming technologies such as Hadoop & Spark, Apache NiFi and Apache Kafka/Confluent. Furthermore, through Privitar On Demand, a proprietary web service, privacy policies and de-identification processing can be applied via a secure HTTPS API that can be integrated in custom applications. Across all these technologies, Privitar provides consistent privacy protection which is complementary to existing security controls.

The Privitar Platform runs on on-premise, cloud and hybrid deployments. The same policies can be executed on different technical infrastructures, enabling consistency and compliance across the organization.

The processing is performed close to the data, either at-rest on the cluster or in-motion in the data pipelines, greatly simplifying security architecture and processes.

Privacy policies, de-identification jobs and other configuration objects can be managed and executed via APIs, enabling the Privitar Platform to be integrated into automated workflows at scale.

Simplified Data Governance

The Privitar Platform records the data lineage of the protected datasets by integrating with the major enterprise metadata catalogs, such as Cloudera Navigator or Apache Atlas.

This allows for advanced and automated data governance, by using tools such as Apache Ranger to enforce access control policies based on Protected Data Domains metadata, or by controlling data lifecycle events.

Leading Privacy-Enhancing Techniques Included

Privitar employs a variety of techniques to protect data:

- Attribute-level suppression, tokenization and masking
- Attribute-level perturbation techniques
- Statistical anonymization (generalization)

Attribute-Level De-Identification Techniques

Privitar offers the following attribute level techniques that operate directly on attribute values:

- Suppression/Redaction: Fully or partially removing sensitive values from the dataset.
- Substitution: Replacing sensitive data with mapped values.
- Format Preserving Tokenization: Replacing sensitive data with a unique token. Tokens can be derived or randomly generated and can be constrained to a target format (e.g. email address or Luhn-valid credit card number). Privitar provides advanced controls over the referential integrity and consistency of the generated tokens, by publishing the data into Protected Data Domains.
- Encryption: Replacing a value with an encrypted version of that value using an encryption key.
- Perturbation: Adding random noise to data to obscure its original values.

Statistical Anonymization Through Generalization

Statistical anonymization techniques protect dimensions classified as secondary identifiers or quasi-identifiers.

Quasi-identifiers are variables which are not sufficient to identify an individual in isolation, but may do so when combined. Quasi-identifiers threaten privacy through linkage attacks, in which records are joined to another dataset, revealing identity information not present in the original records.

The Privitar Platform generalizes data to be resistant to linkage attacks while defending against the disclosure of an individual's sensitive attributes.

Generalization reduces the resolution or accuracy of the data to limit the knowledge that data consumers can obtain. Multiple quasi-identifiers are generalized to achieve k-anonymity, in a way that automatically minimizes distortion and so maximizes the utility of the data for analytics and secondary use.

Watermarks

Organizations need to ensure that their data throughout its lifecycle is protected in accordance with data governance standards. Companies not only face threats with regards to safe data usage and dissemination from external parties, but also internal ones (unauthorized behavior, mistakes).

Watermarks enables detection and attribution of unauthorized distribution or publishing of a sensitive dataset, and so acts as a deterrent against such unauthorized behavior.

Privitar embeds a digital watermark in anonymized data, which can be recovered should the data be leaked and traced back to the Protected Data Domain (PDD) the data came from. The PDD watermark gives access to its metadata, such as the file's intended purpose, who authorized the release of the data, and the Policy it was released under.

Enterprise-Level Security

The Privitar Platform integrates with enterprise security architectures including Active Directory and Kerberos for user and role management, authentication and authorization. It also integrates with Key Management Systems for encryption and supports key rotation.

Securely Collecting and Linking Data

Considerable insight can be gained from aggregating multi-party data and analyzing it centrally. However, trust barriers or other restrictions may prevent this from happening. To overcome this, data needs to be protected and de-identified in transit but still be linkable.

Privitar SecureLink enables this by allowing a central organization to securely collect, transfer and join fragmented data such that the data is de-identified and the central party cannot recover the raw identifiers. It protects identifiers by using sophisticated encryption schemes (homomorphic encryption) so that the identifiers are kept encrypted at all times, but can still be processed.

Rest APIs

The Privitar Platform offers a comprehensive set of HTTPS REST APIs to allow custom applications and scripts to interact with the Policy Manager and maintain Privacy Policies, Schemas, Jobs and Protected Data Domains.

The REST APIs can be used, for example, to:

- Manage complex and large numbers of data sets (schemas), privacy policies and de-identification jobs.
- Build custom integrations with metadata catalogues and schema repositories.
- Integrate Privitar in automated self-service data provisioning pipelines and wider data ecosystem in the organization.

We're Privitar

We help organizations engineer privacy-preserving data operations, using advanced data privacy techniques that protect sensitive information while retaining data utility. Our software accelerates and automates privacy-safe data provisioning, helping our customers get more business value from their data, generate data-driven insights, and drive innovation.

Contact us:

e: info@privitar.com
t: +44 203 282 7136
w: www.privitar.com

PRIVITAR

@PrivitarGlobal

www.privitar.com