strategy for product innovation with privacy-preserving data analytics

Table of contents

Context

This paper describes how protects and utilizes customer's real-world production data. Further it describes how we continually improve our product line while keeping personal and business data private and secure.

Executive summary

As a smart home experience company focused on data management and business intelligence, nsiders it mission critical to ensure data is always secure. We use the data we collect to benefit the user, such as improving online protection and digital well-being. Our privacy preserving strategies are in compliance with industry frameworks.

In response to market needs for an improved smart home management platform, we added cloud-based experience and data-insight services. New strategies directly correlate with the amount of time we store—and analyze—real-world and de-identified data. The time we spend developing new solutions with protected data is informed by the:

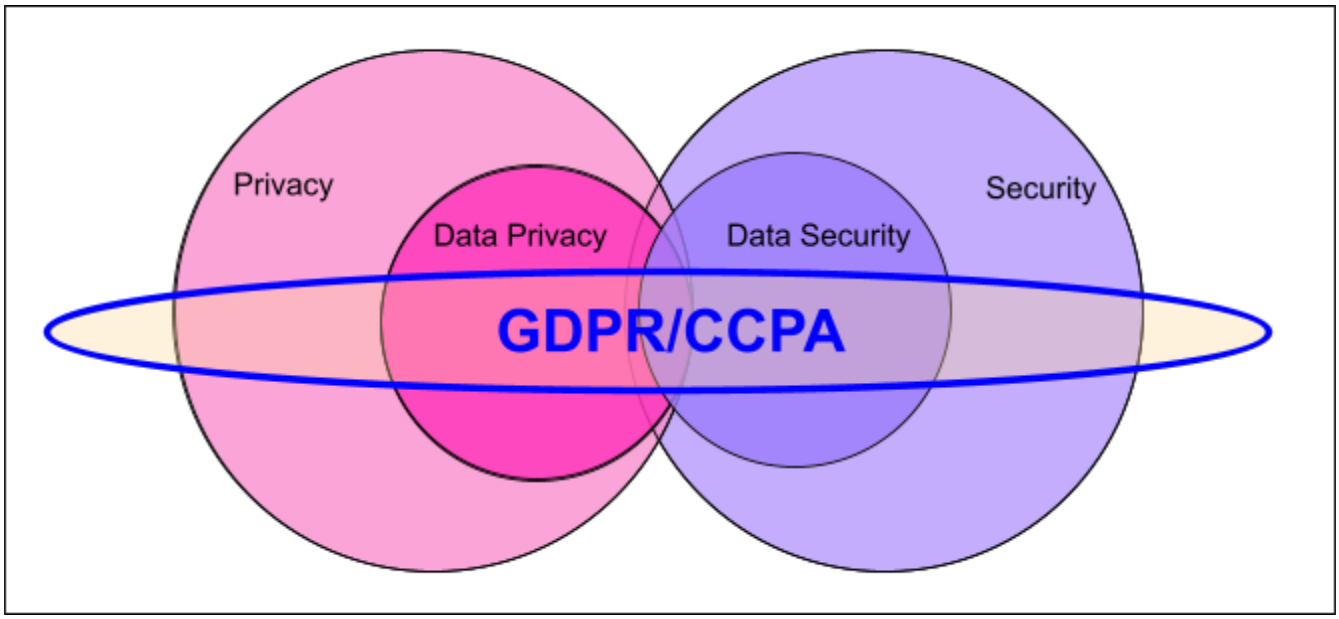
- business need for which it was collected,
- contractually required duration, and
- legally required retention for certain transactional records.

Our commitment to protecting personal and business data extends to all individuals who interact with us: Communications Service Providers (CSPs), business partners, vendors, and end-customers.

CSP Viewpoint

created the first SaaS experience platform dedicated to CSPs and their subscribers (users). As a SaaS provider, our top priority has always been to protect data with security tools and individuals with data privacy (privacy) tools. We baked in strong encryption, so that data is always secure. This applies to data at rest on a device or in the cloud, and data traffic from the internet.

The guidance on security and privacy are complementary but come from distinct fields, with different goals. Successfully protecting data and privacy in the cloud means that they have to be integrated with each other.



security and privacy within regulation guidelines

Our data security and privacy practices are in compliance with the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).

Walking the tightrope

Rules and regulations can keep us up at night as we walk between what is right for our customers and what next new thing needs technical investments. The quandary is that the research required to develop that next great solution/feature depends on working with real-world data for analytics and machine learning algorithms. To ease the tension of working with

real-world data collected by doing business, we protect the data with de-identification and anonymization solutions so that personal attributes in real-world data are not traceable to any one person.

data privacy and security solutions

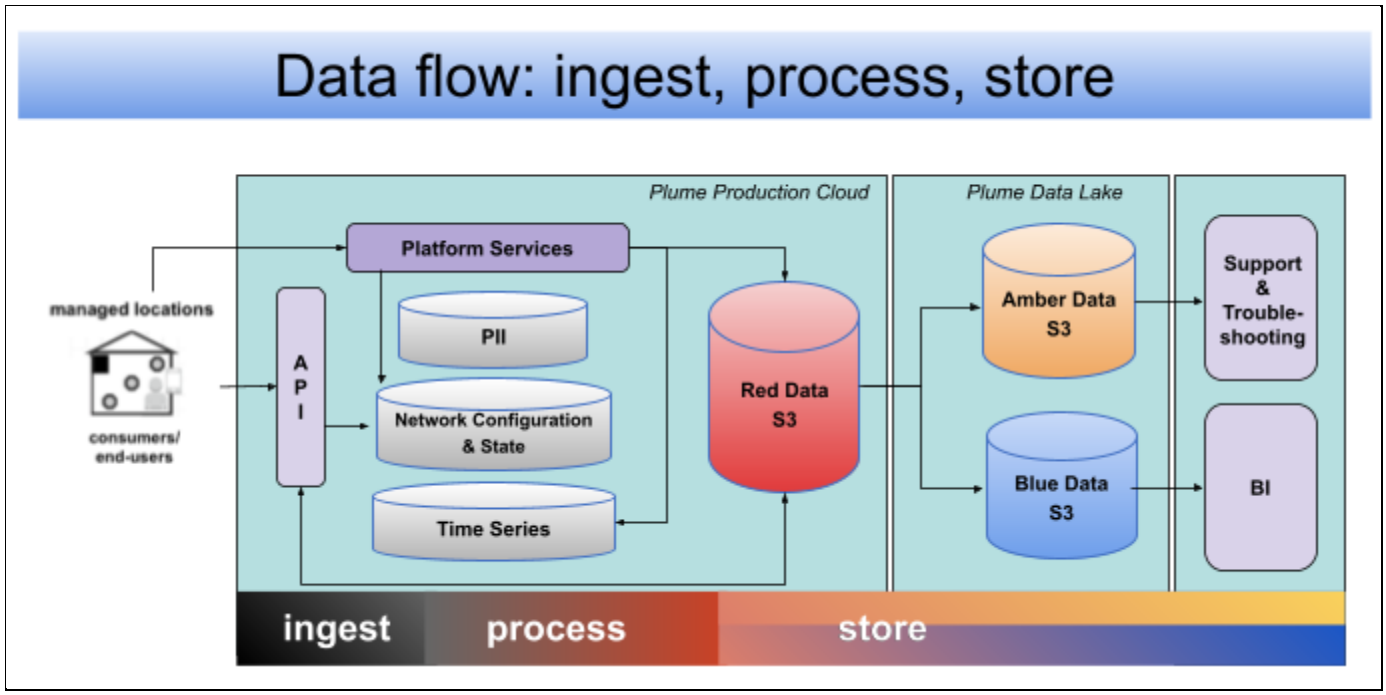
strategy for bringing personal data protection up to contractual and regulatory requirements is to fold in advanced processes on top of tried and true methodologies. The following table shows how these requirements are addressed.

Privacy and security requirements and solutions table

Data privacy and security requirements	Plume solutions
Ensure appropriate controls based on sensitivity of information.	Treat all personal data with the same level of security protection.
All personal data must be encrypted at rest and in transit.	• A blanket encryption at rest is in all data stores using server side encryption and Plume managed keys. • Encrypt data in transit for internet traffic.
All access to personal data must be authorized, audited and limited to the purposes of customers obligations.	• Strictly define user-access roles with audit capability when reidentifying an individual customer. • Perform periodic data-access audits at a minimum of twice a year.
Appropriate technical and organizational measures to ensure a level of security for personal data appropriate to the risk.	• Host hardening to protect data in memory while processing data. • Implement vendor third-party security risk assessments to understand data transfer risks. All Plume vendors are required to comply with data classification, privacy and de-identification practices. • Implement Plume Data Processing Agreements (DPAs) for customers and vendors to handle data transfer obligations.
No personal identifiers should be retained in raw data format when stored for long-term analytics or machine learning.	Tokenize personal identifiers in data stored for use in analytics and machine learning.
Customer personal data should not be used in any development or testing environments.	Retain only de-identified data for development or testing environments.
Plume may collect the minimal personal data required for business and technical purposes, and use it only for purposes disclosed to and consented by the CSP customer or consumer.	Adopt Privacy by Design practices in secure development lifecycle to align processing activities with business purpose and secure appropriate consent.
Personal data gained from a customer account must be deleted when that account is deleted.	Remove all customer account data when an account is deleted.

Strategy for preserving the privacy of data analytics and product innovation

Our privacy-preserving solutions keep personal attributes in real-world data untraceable to any one person. In this way we can continue delivering reliable and high quality insights and visualizations.



Data flow to ingest, process, and store

How I collects data

OpenSync-enabled CSP Gateways and (network nodes) create a managed WiFi network in home or small business locations. When end-users connect to the WiFi network, the network nodes report the account, network, and internet connectivity data (with personal identifiers). This information is required for monitoring and optimizing WiFi performance, providing online protection and motion features, as well as enhancing services for CSPs.

Platform services ingest and store production data from managed locations in a Cloud deployment, while API services ingest and store account-profile and WiFi configuration data

from the customer mobile app at managed locations in a l Cloud deployment. The ingested data is stored within the PII, Time Series, Network Configuration, and State databases.

How uses data

A data ingestion pipeline periodically copies select production data sets; these sets inform product innovation algorithms in the Plume data lake. This data is further prepared with data enrichment and aggregation pipelines to create dimensional models and reporting tables that:

- we deliver to consumers through authenticated mobile and Web Apps, like HomePass or WorkPass for online protection and digital well-being,
- deliver personalized trends with email or push notifications,
- are available through business intelligence dashboards, such as Panorama, CorpDash, etc.,
- we publish as aggregated and anonymized global trends on our plume.com resources page, and
- CSP customers can receive daily data exports.

How l classifies information for data management

manages all information in accordance with Plume's information classification and retention policy. This means that information must be classified and handled based on its value and sensitivity. The classification levels determine what baseline data protection safeguards are appropriate when handling information. We color-code assignments of red, amber, blue, and green to simplify associating data access conditions and rules.

The personal data categories are:

- RED raw data

  Confidential— personal data collected from customer locations.

- AMBER pseudonymized data

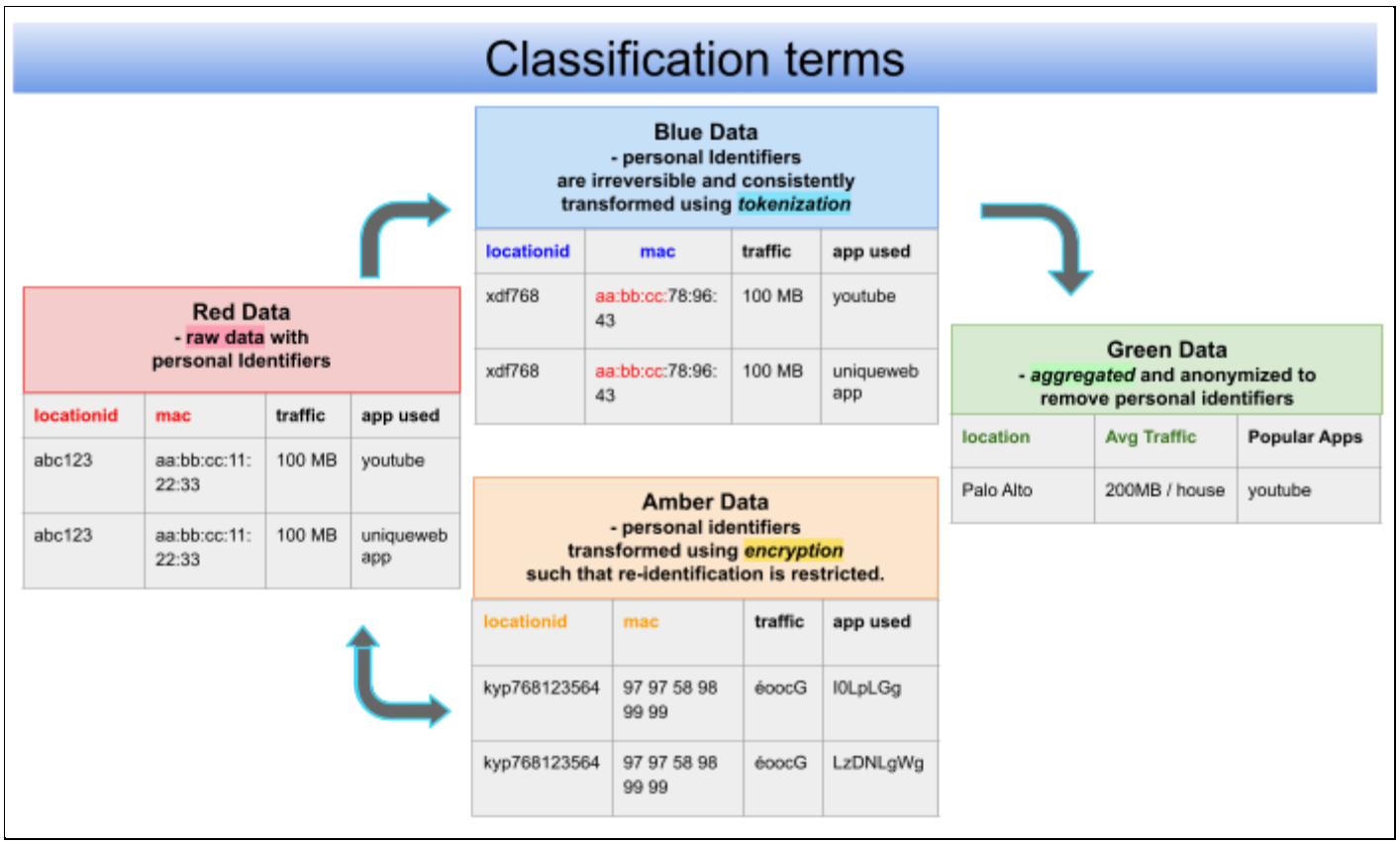
  Internal—encrypted personal data using symmetric encryption algorithms; auditable records of customer re-identification.

- BLUE pseudonymized data

  Internal—tokenized data that preserves statistical qualities of the original but cannot be re-identified. Data is protected and reasonably de-identified.

- GREEN anonymized data

  Public—aggregated data that can be released to the public. Data is anonymized.



Classification terms and examples

How refines data-use cases for product improvements

We examine our business use cases to find analytics and transactional data that can potentially be used to enhance customer experience.

Specifically we use:

- tokenized and de-identified (Blue-zone) data for data analysis in service insights, product improvement and machine learning, and
- encrypted and re-identifiable (Amber-zone) data for troubleshooting or transactional scenarios.

All analytics data is protected and de-identified using a third-party solution, Tonic.ai. All transactional data are protected using the AWS platform's encryption capabilities with Plume-managed encryption keys.

How implements personal data de-identification

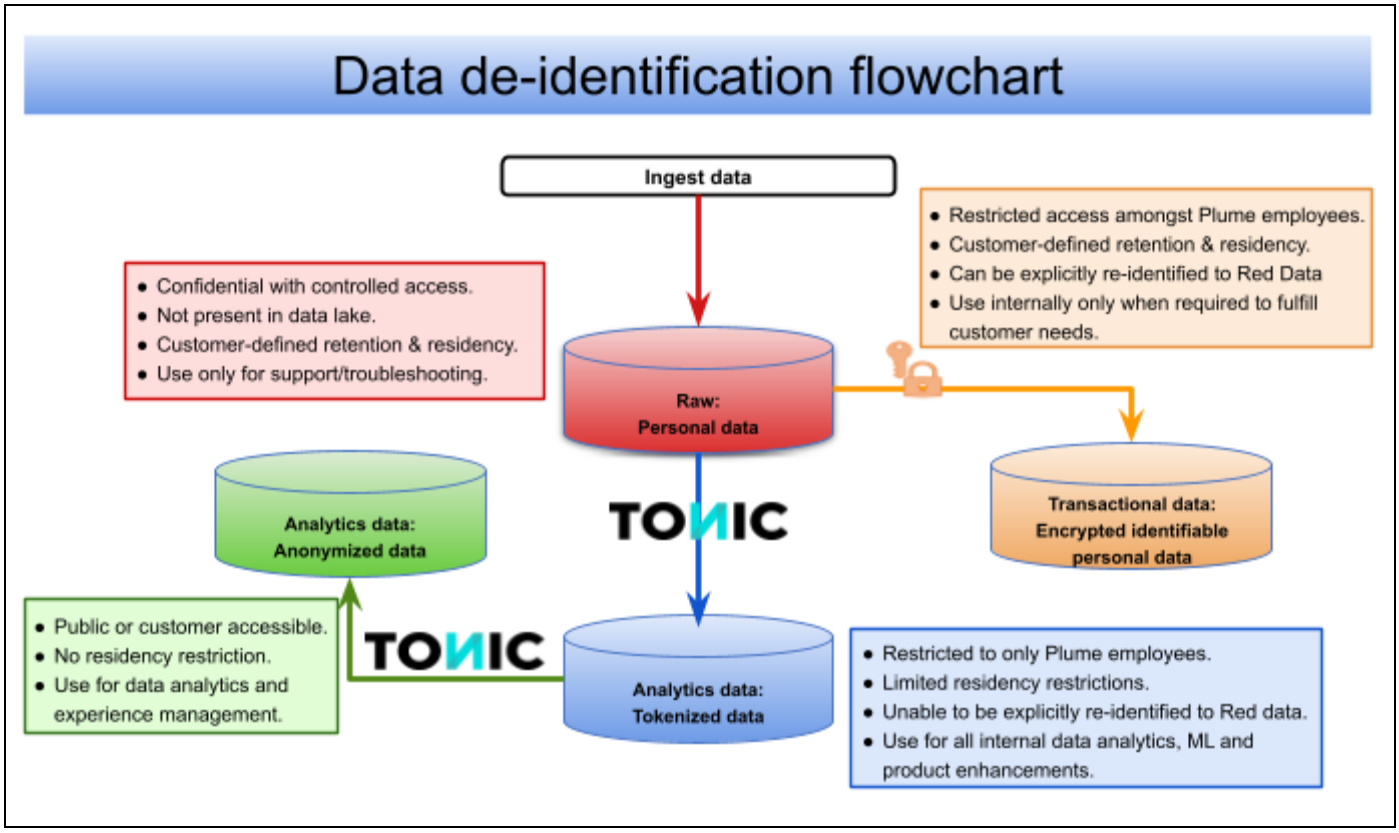
Periodically running dual ingestion pipelines from production environments into our data lake are critical to implementing the Red-to-Blue and Red-to-Amber classification criteria. In one pipeline, data transformation jobs handle the following:

- Transform personal identifiers in service and usage data using Plume Information Security team defined data transformation policies.
- Write the tokenized data with consistently preserved statistical qualities to the "Blue" data zone, such as tokenizing the mac address while preserving the first 3 octets. (*See Classification terms and examples figure.*)

Access within the Blue-zone is authorized and granted to machine users and dev/test "human" access roles based on auditable access records. In the other pipeline, data transformation jobs handle the following:

- Write service and usage data into buckets within the "Amber" data zone wherein personal data is encrypted using a symmetric encryption algorithm with a Plume-managed key stored in the AWS key-management service.
- Ensure logical isolation of data within AWS regions using region-specific keys.

Access within the Amber-zone is restricted to machine users and restricted, privileged “human” roles based on auditable customer support requests. Following data transformation, the Blue-zone data undergoes automated data-quality-regression tests before it’s available for wider-consumption.



Data de-identification flowchart

How anonymizes data

So far we have discussed personal data protection solutions and our data classification with methods to protect and de-identify raw data. In this section we discuss ta aggregation and anonymization.

Aggregation is a statistical pipeline process where data becomes protected and anonymized, making it safe to view publicly. It can be viewable in groups or as part of a summary, but is not viewable at an individual level. The statistical function, aggregation criteria, includes average, sum, cohort, etc.