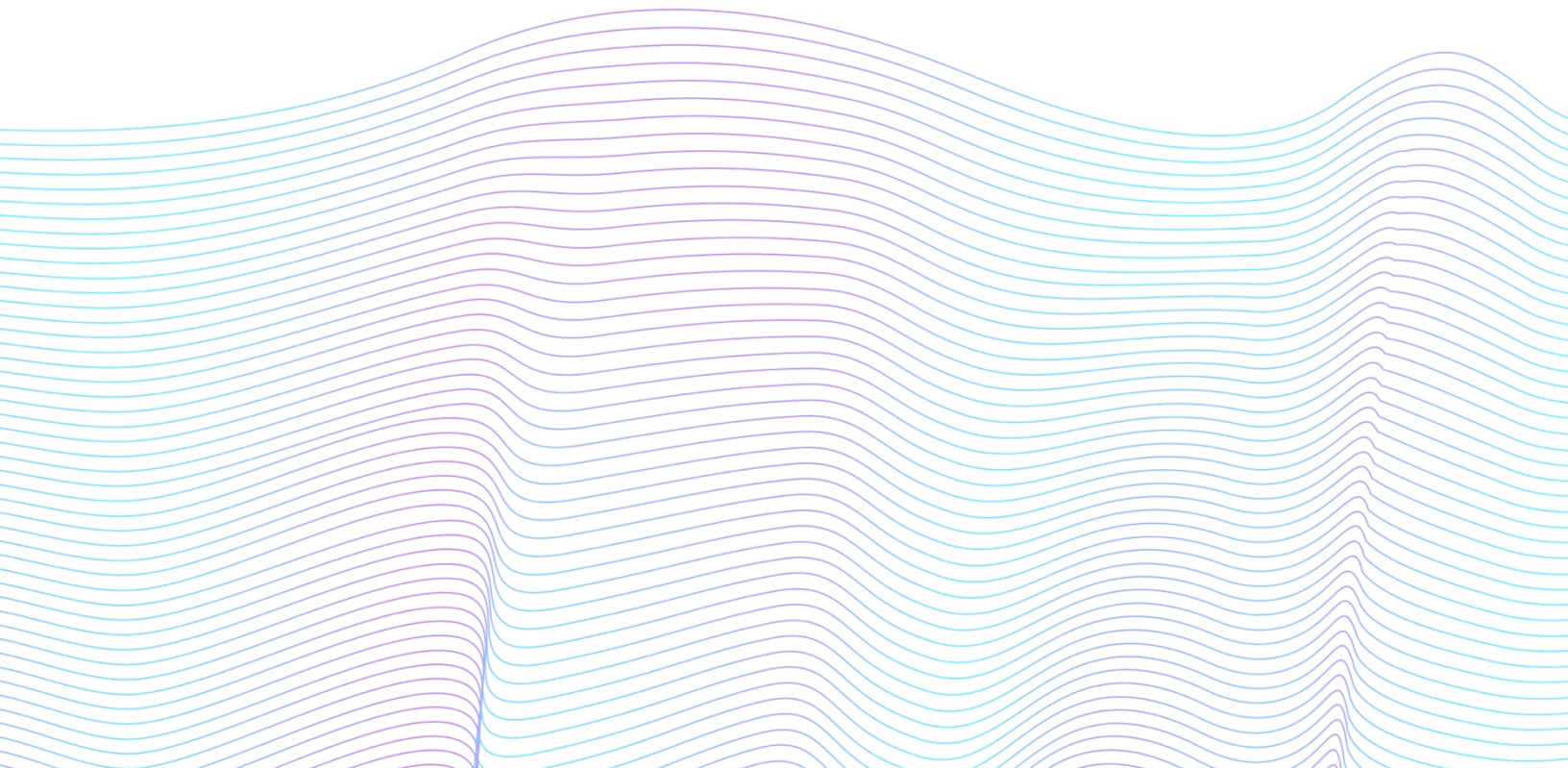
Plume®

Plume's strategy for product innovation with privacy-preserving data analytics

A Plume® whitepaper

September 2022



Disclaimer:

This whitepaper (this "White Paper") was prepared by Plume Design, Inc. ("Plume") for informational purposes only and not for any other purpose. Nothing contained in this document is, or should be construed as, legal or any other advice nor as any form of recommendation, promise or representation by the presenter or Plume or any officer, director, employee, agent, or advisor of Plume. This White Paper is Plume's proprietary information and nothing herein grants any rights or licenses to any of the contents of this White Paper or any other intellectual property of Plume to any third party.

Plume employees have prepared this document in good faith and based on their beliefs and assumptions in light of currently available information. While such Plume employees have endeavored to ensure that the assumptions, assessments, statements and information contained in this document are accurate, the information contained in this White Paper may not be complete, comprehensive, accurate, adequate or correct. The contents of this White Paper are presented "as is" and without any representation or warranty of any kind, either express or implied, regarding the accuracy or completeness or other quality whatsoever. In no event shall Plume or any of its representatives have any liability to any third party (or any of its affiliates or its or their representatives or agents) relating to or arising out of any use of the contents of this White Paper. This White Paper contains information as of its date of publication (or such other date as may be referenced) and Plume is not responsible for updating any information contained in this White Paper following its publication.

This White Paper may contain certain preliminary, unaudited financial information which is subject to revision as well as forward-looking statements. Forward-looking statements are, by their nature, subject to significant risks and uncertainties. These forward-looking statements may include, without limitation, statements relating to Plume's business prospects, future developments, trends and conditions in the industry and geographical markets in which Plume operates, its strategies, plans, objectives and goals, its ability to control costs, statements relating to prices, volumes, operations, margins, overall market trends, risk management and other factors. Actual results and events may differ materially from information contained in the forward-looking statements as a result of a number of factors, including any changes in the laws, rules and regulations relating to any aspects of Plume's business operations, general economic, market and business conditions, including changes or volatility in interest rates, foreign exchange rates, equity prices or other rates or prices, the actions and developments of Plume's competitors and the effects of competition in the insurance industry on the demand for, and price of, Plume's products and services, various business opportunities that Plume may or may not pursue, changes in population growth and other demographic trends, Plume's ability to identify, measure, monitor and control risks in Plume's business, including its ability to manage and adapt its overall risk profile, its ability to properly price its products and services, market demand for its products and services, and factors beyond Plume's control.

Nothing in this White Paper obligates Plume to negotiate or enter into any agreement of any kind with any party. Plume has no obligation of any kind whatsoever with respect to a possible business transaction, whether by virtue of this White Paper or otherwise.

Table of contents

Context

This paper describes how Plume Design Inc. (Plume) protects and utilizes customers' 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, Plume considers it mission critical to ensure data is always secure. We use the data we collect to benefit the user, by, for example, improving online protection and digital wellbeing. 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. The period of time during which we store—and analyze—real-world and de-identified data correlates with our strategies for privacy-preserving analytics and is informed by the:

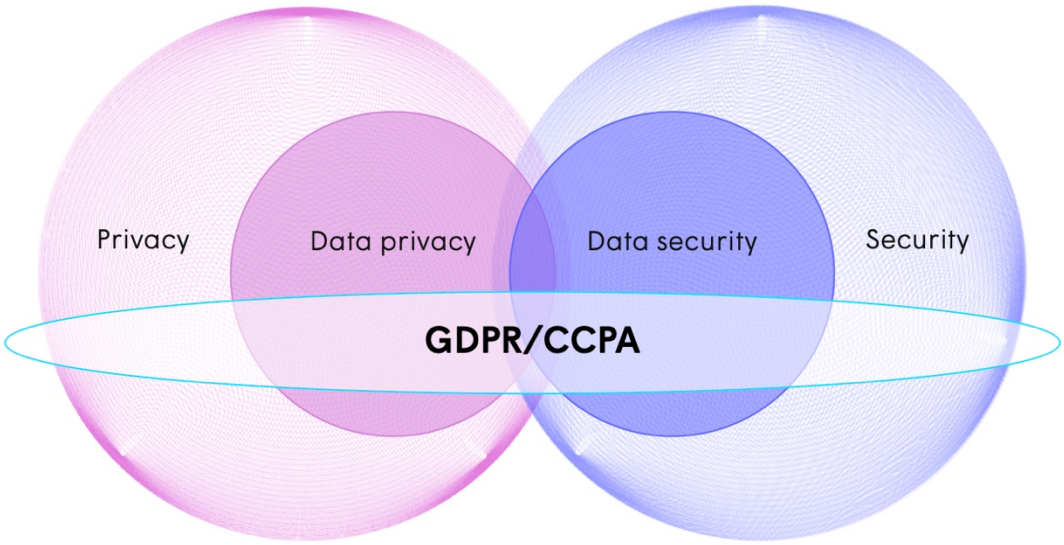
- business need for which it was collected.
- contractually required duration.
- legally required retention for certain transactional records.

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

Plume 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 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.



Plume 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

We regularly find ourselves walking a tightrope 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.

Plume data privacy and security solutions

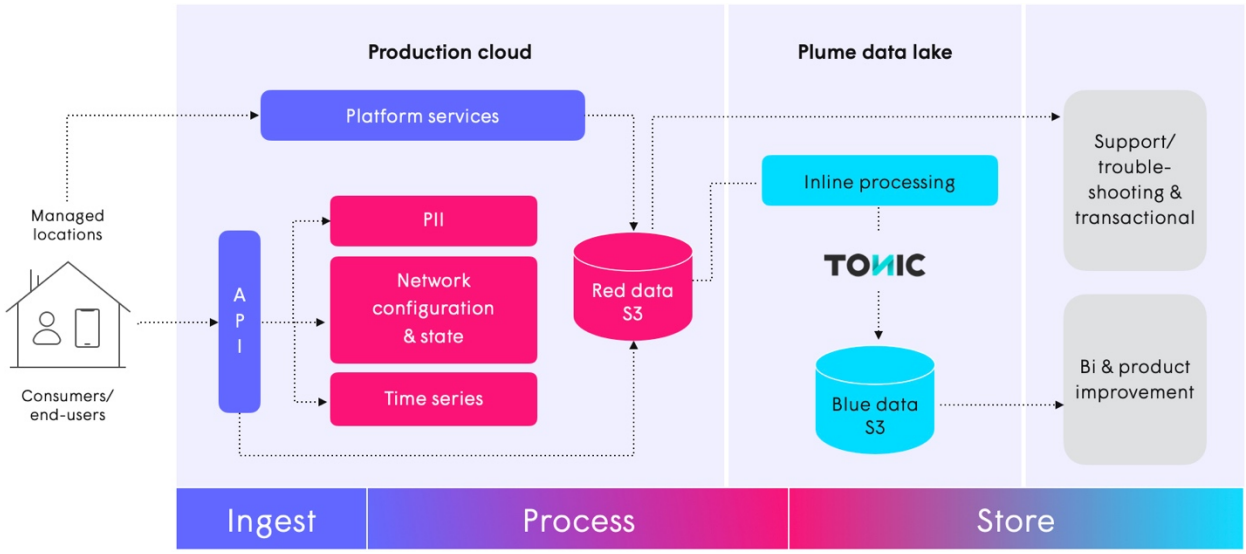
Plume's strategy for bringing personal data protection up to contractual and regulatory requirements is to fold in advanced processes on top of tried and true methodologies. The following table shows how these requirements are addressed.

Data privacy and security requirements	Plume solutions
Ensure appropriate controls based on the 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 once 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 the 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.

Privacy and security requirements and solutions table

Strategy for preserving privacy in 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, high-quality insights and visualizations.



Planned data flow to ingest, process, and store

How Plume collects data

OpenSync-enabled CSP gateways and Plume pods (network nodes) create a Plume-managed WiFi network in residential 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 Plume Cloud deployment, while API services ingest and store account-profile and WiFi-configuration data from the customer mobile app at managed locations in a Plume Cloud deployment. The ingested data is stored within the PII, Time Series, Network Configuration, and State databases.

How Plume uses data

A data ingestion pipeline periodically copies select production data sets; these data 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:

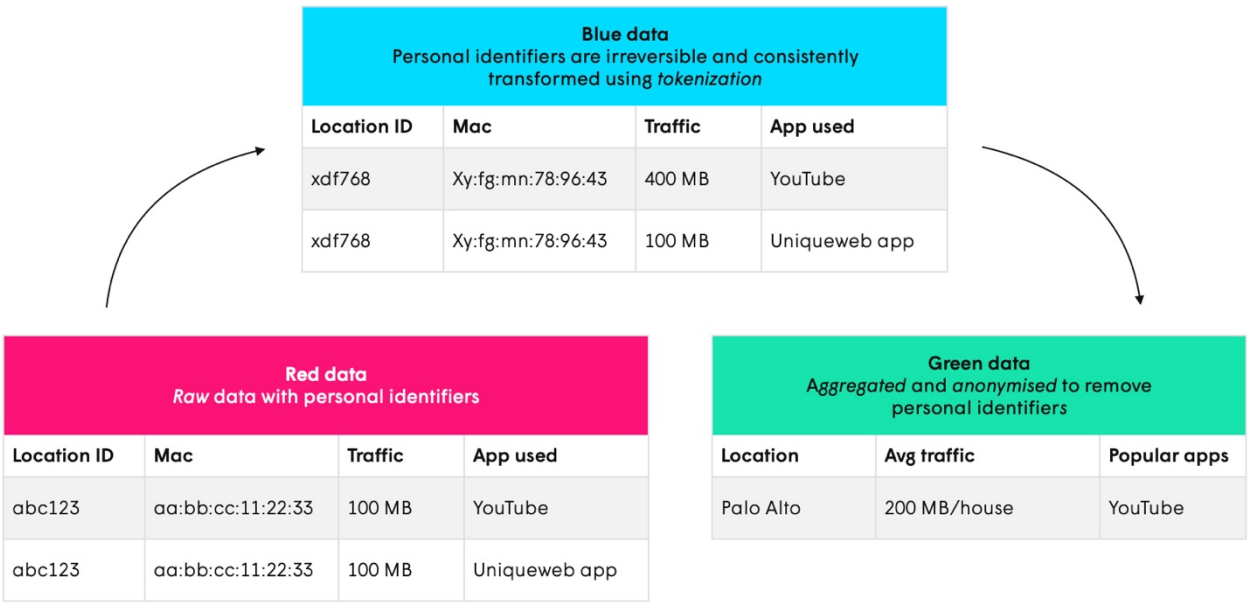
- we deliver to consumers through authenticated mobile and web apps, like HomePass® or WorkPass® for online protection and digital well-being.
- deliver personalized trends with email or push notifications.
- are available through business intelligence dashboards, such as Panorama™, CorpDash, etc.
-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

How Plume classifies information for data management

Plume 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, blue, and green to simplify associating data-access conditions and rules.

The Plume personal data categories are:

- RED raw data
  - Confidential—personal data collected from customer locations and encrypted at rest.
- BLUE pseudonymised 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.



Blue data
Personal identifiers are irreversible and consistently transformed using *tokenization*

Location ID	Mac	Traffic	App used
xdf768	Xy:fg:mn:78:96:43	400 MB	YouTube
xdf768	Xy:fg:mn:78:96:43	100 MB	Uniqueweb app

Red data
Raw data with personal identifiers

Location ID	Mac	Traffic	App used
abc123	aa:bb:cc:11:22:33	100 MB	YouTube
abc123	aa:bb:cc:11:22:33	100 MB	Uniqueweb app

Green data
Aggregated and *anonymised* to remove personal identifiers

Location	Avg traffic	Popular apps
Palo Alto	200 MB/house	YouTube

Classification terms and examples

How Plume 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.

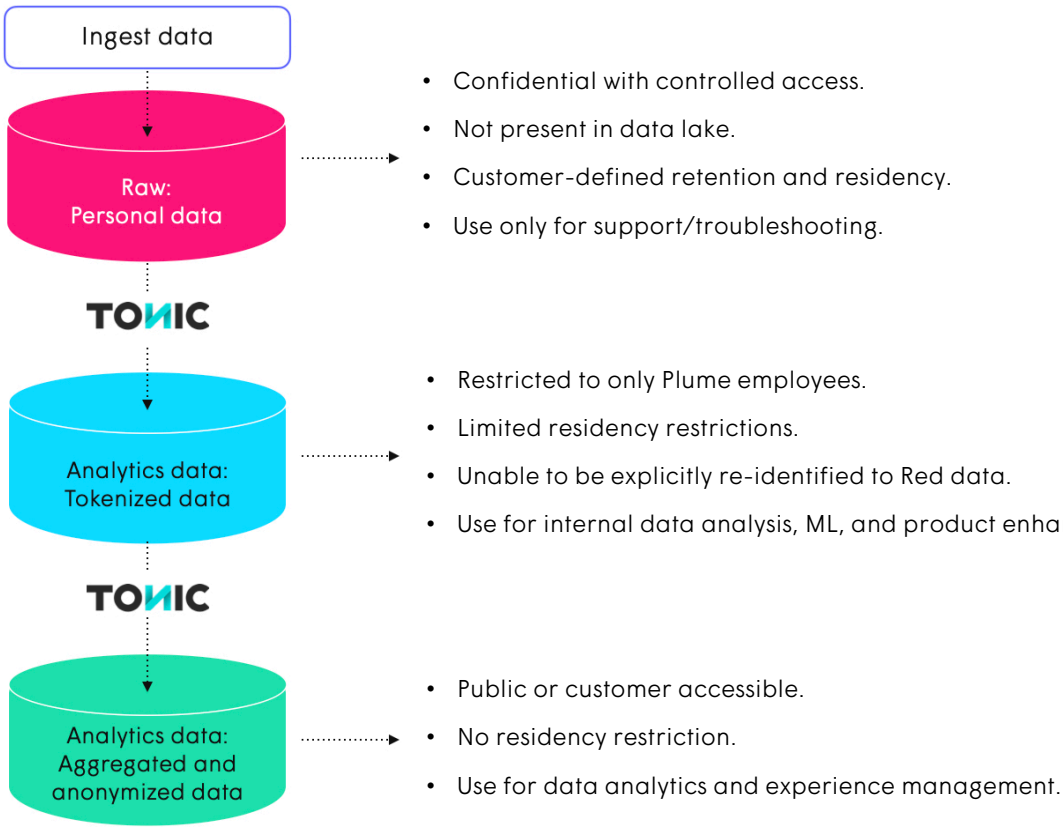
All analytics data is protected and de-identified using a third-party solution, Tonic.ai. All transactional red data is protected using the AWS platform's encryption capabilities with Plume-managed encryption keys.

How Plume implements personal data de-identification

Periodically running data-ingestion pipelines pull raw data from production environments into our data lake to implement the Red-to-Blue classification criteria. In the pipeline, inline data transformation jobs handle the following:

- Transform personal identifiers in service and usage data using Plume Information Security team-defined data transformation policies configured with Tonic.ai generators.
- Write the tokenized data with consistently preserved statistical qualities to the "Blue" data zone, such as tokenizing the mac address. (*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. 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 Plume 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 Plume data 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, and aggregation criteria, include average, sum, cohort, etc.

We use tokenized data with two different data aggregation solutions:

- Generate dimensional models by location, such that we use dimension and fact tables aggregated by location-pseudo-identifiers at a daily or monthly frequency.
- Add dimensional data to generate reporting database tables that can be easily used in dashboards.

Certain location characteristics are unique to our data sets. When these characteristics are correlated with raw production data sets, it creates direct identifiers and can re-identify one or more end-users. To address this risk, we use data access and retention controls to restrict red data access.

How long Plume plans to retain data for analytics and product innovation

The length of time we keep personal data depends on the following criteria:

- Business needs for which it was collected.
- Contractually required durations.
- Legally required retention for certain transactional records.

Every 30 days, or when a customer account is terminated, raw personal data will be deleted from our production clouds. Anonymized and de-identified data will be retained for long-term analytics in our development and test environments.

Our personal data categories and the associated retention approach we have taken for the Plume data lake is:

- BLUE pseudonymized data (**Internal**) that is protected and reasonably de-identified. It is retained in the Plume data lake for longer durations (currently up to 5 years), commensurate to Plume-acceptable risk levels.
- GREEN anonymized data (**Public**) that is protected and anonymized. It is retained in the Plume data lake and on public websites for longer durations than Blue data, commensurate to Plume-acceptable risk levels.

Conclusion

In this paper we have described how Plume, in a high-growth industry, has constructed a careful approach to ensuring our product line continues along the path of innovation, problem-solving, and ultimate ease of use for clients and end-users. By designating and color-coding categories of personal data protection solutions and safety, a clear course has been set with internal and external expectations and requirements.

Our goal is to continue on this path of privacy-preserving strategies for data analytics and product improvements.

Top

Appendix A

Glossary

Aggregate criteria of regression analysis is to investigate the common influence of several potential influence factors on the target parameter. For example, Cox regression or Poisson regression can be used for the data analysis of cohort studies, depending on the target parameter.

AWS key-management service (KMS) is a secure and resilient service that uses hardware security modules that have been validated under FIPS 140-2, to protect Plume keys. AWS KMS is integrated with AWS CloudTrail to provide logs of all key usage to help meet any regulatory and compliance needs.

Anonymize is the complete and permanent removal of personal identifiers from data, such as converting personal, identifiable information into aggregated data. Anonymized data is data that can no longer be associated with an individual in any manner. Once this data is stripped of persona-identifying elements, those elements can never be re-associated with the data or the underlying individual.

Data handling occurs after the conclusion of a business function or project ensuring data is stored, used, archived, or disposed of in a secure and private manner. This includes policy development and procedures to manage data handled electronically or by non-electronic means.

Data ingestion is a short-hand way of saying that data is being prepared for analysis. This usually includes steps to extract (taking the data from its current location), transform (cleansing and normalizing the data), and load (placing the data in a database where it can be analyzed).

Data lake is a storage repository that can store a large amount of structured, semi structured, and unstructured data.

Data privacy is a subset of privacy and refers to the rules we apply to handling personal data. Data privacy defines the policies that data protection tools and processes employ and is concerned with the proper handling of data (e.g. collection, consent, use, transfer to third parties, etc.), particularly under regulatory obligation.

For protection, it is up to the companies handling data to ensure that it remains private. Data privacy is a legal issue and data protection is essentially a technical issue. Because this paper is on “data privacy” we refer only to privacy to imply data privacy.

Data protection is concerned with the unauthorized use, corruption, loss, and availability of personal data.

De-identify is removing personal, identifying information in order to protect personal privacy. In some definitions, de-identified data may not necessarily be anonymized data (as defined in this document). This may mean that the personal identifying information may be able to be re-associated with the data at a later time.

In such cases, anonymized data is a subset of de-identified data. Data is considered de-identified under the Privacy Rule when a number of specified data elements are removed. (45 C.F.R. §§ 164.502(d)(2), 164.514(a) and (b).) De-identified data is not regulated by HIPAA and may be shared without restriction.

Limited data sets are stripped of many categories of identifying information but retain information often needed for public health and research (such as birth dates, dates of treatment, and some geographic data). (45 C.F.R. § 164.514(e).).

NOTE: The de-identify definition excludes references to health information in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

Dimensional database is a relational database that uses a dimensional data model to organize data. This model uses fact tables and dimension tables in a star or snowflake schema. A dimensional database is the optimal type of database for data warehousing.

Encryption at rest and in **transit** are both data protection concepts. Data can be exposed to risks both in **transit** and at **rest** and requires protection in both states.

- For protecting data in **transit**, you can mount a file system so that all NFS traffic is encrypted in transit using Transport Layer Security 1.2 (TLS) with an industry-standard AES-256 cipher.
- For protecting data at **rest**, enterprises can **encrypt** sensitive files prior to storing them and/or choose to **encrypt** the storage drive itself.

Direct personal identifiers (*commonly known as direct personal data*) include any pieces of information whereby an individual is directly identifiable using nothing but the information one possesses.

Host hardening has several meanings in computer security such as limiting network access to a system by turning off unnecessary network services, by firewalling, or by enforcing authentication to use a service.

Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information that, when collected together, can lead to the identification of a particular person, also constitute personal data.

Indirect personal identifiers (commonly known as indirect personal data) includes any pieces of information whereby an individual is not directly identifiable using the information alone, but one may be able to identify the individual by using other information obtained from a reasonably accessible source.

PII database is secured such that links between it and the rest of our cloud data can only be accessed by our API software.

Privacy is a state in which one is not observed or disturbed by other people.

S3 or Amazon Simple Storage Service is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network.

S3 data lake—see S3 and data lake definitions.

Security refers to preventing unauthorized access to personal information, through technologies like network security, firewalls, encryption, etc.

State databases hold the last known committed value for any given key. They are populated when each peer validates and commits a transaction. The state database can always be rebuilt by re-processing the ledger.

Time series databases store data as a sequence of data points collected over time intervals, giving the ability to track changes over time. By storing data in this way, it makes it easy to analyze time series, or a sequence of points recorded in order over time, efficiently and continuously add, process, and track massive quantities of real-time data with speed and precision.

Tokenized data is undecipherable and irreversible and therefore has no meaningful value. It is generated by a process whereby personal identifiers are replaced with randomly-generated values (i.e. iterate through each character in the string and randomly replace it with a different character in the same 'character class', for additional reference see: Tonic Generators Common Usage). When consistency is enabled we'll ensure that a given cell value always gets mapped to the same output cell value. Additionally, if the character is whitespace, punctuation, or a Unicode Math Symbol we leave the character as-is and do not modify it. In the case of a mac address, for instance, the generated octets, however, are unique in the sense that we will only ever generate that specific set of 6 octets for the original mac address), known as tokens, without having any mathematical relationship with the original identifiers. Data in the "Blue" zone criteria has preserved statistical qualities in addition to being tokenized data. In some cases, the statistical quality to join datasets across multiple tables is preserved by generating consistent tokens for the same input across an entire database or multiple databases.

Top

Appendix B

References

1. Aggregating over Anonymized Data, 2019, International Association of Privacy Professionals, Lea Kissner
2. California Consumer Privacy Act (CCPA), 2018, State of California Department of Justice, Attorney General office
3. Design considerations for building privacy-protecting analytics services, 2019, The International Association of Privacy Professionals, Rafae Bhatti, CIPP/US, CIPM, IAPP Member Contributor
4. General Data Protection Regulation (GDPR) and Art. 5, 2018, official PDF of the Regulation (EU) General Data Protection Regulation, in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018 as reported by InterSoft Consulting, IT security and IT forensics
5. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII), 2010, National Institute of Standards and Technology (NIST) and Information Technology Laboratory, Computer Security Resource Center, Erika McCallister (NIST), Tim Grance (NIST), Karen Scarfone (NIST)
6. Guidelines for Data De-Identification or Anonymization, 2015, EDUCAUSE, a nonprofit association
7. Opinion 05/2014 on Anonymisation Techniques, 2014, Data Protection Working Party, an independent European advisory body on data protection and privacy
8. Privacy-preserving data analysis: How can we enable personal data to be shared without compromising our privacy?, ongoing, Turing Research Group, interestgroups@turing.ac.uk
9. Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics, 2015, European Union Agency for Network and Information Security, Giuseppe D' Acquisto (Garante per la protezione dei dati personali), Josep Domingo-Ferrer (Universitat Rovira i Virgili), Panayiotis Kikiras (AGT), Vicenç Torra (University of Skövde), Yves-Alexandre de Montjoye (MIT), Athena Bourka (ENISA)
10. Salesforce Security White Paper for Salesforce Government Cloud, 2015, SalesForce, as part of a FedRAMP annual assessment
11. Salesforce Shield Enhance protection, monitoring, and retention of critical Salesforce data, 2017, Salesforce State of IT Report
12. The consumer-data opportunity and the privacy imperative, 2020, McKinsey & Company, Venky Anant, Lisa Donchak, James Kaplan, and Henning Soller

Top

Appendix C

List of figures