WHITE PAPER

BROADCOM & INFOSYS WHITEPAPER: REDEFINING INNOVATION AND AUTOMATION THROUGH PRIVACY FIRST DATA PIPELINES

Abstract

Enterprise view privacy controls as inhibitors of innovation.

For instance, Apple's move to adopt a privacy first consumer driven model enabling consumers to control data sharing met with tension from businesses. Enterprises feared giving up on data control will restrain innovation.

Enterprises need data without privacy risk for their AI and Research. This whitepaper explores how privacy first data pipelines powered by Broadcom and Infosys can be an enterprise enabler for innovation.

BROADCOM®

Infosys®
Navigate your next

Table of Contents

Table of Figures

Intended Audience

This whitepaper is intended for Enterprise Architects, Security and Privacy Professionals, Data administrators and Solution experts working on AI, ML, Analytics, application testing and other data centric services.

Finding the balance between innovation and privacy

The world as we know today is hyperconnected. The Global internet traffic was estimated to be more than 3 zettabytes, or 3,000,000,000,000 gigabytes in 2020. This is a 1000-fold increase in the last 20 years. In addition to transmitting valuable streams of information and ideas, data flows enable the movement of goods, services, finance, and people.

Enterprises battle to make sense of the constant stream of data from new sources, searches, communication, video, and transactions. Government regulation around data collection and data privacy continues to evolve and catch up to for governing enterprise data. For enterprises, the non-compliance with regulatory requirements across geographies includes and does not limit to reputational damage and hefty penalties. It can also have crippling effect to the brand value and customer trust.

To innovate and automate, enterprises must collect and analyze new forms of data, which can pose privacy risks if not responsibly managed. For example, the increasing use of artificial intelligence and machine learning algorithms can lead to the collection of enormous amounts of personal data, which can be used to find individuals and make decisions that affect their lives. This data can also be vulnerable to cyber-attacks and other forms of unauthorized access, which can result in harm to individuals and organizations.

Balance between innovation and privacy is a complex issue. On one hand, innovation drives progress and can lead to significant advancements in technology, healthcare, and other areas. On the other hand, privacy is a fundamental human right that must be protected to ensure that individuals are not subject to harm or discrimination.

Trends driving privacy standards in the data driven economy are:

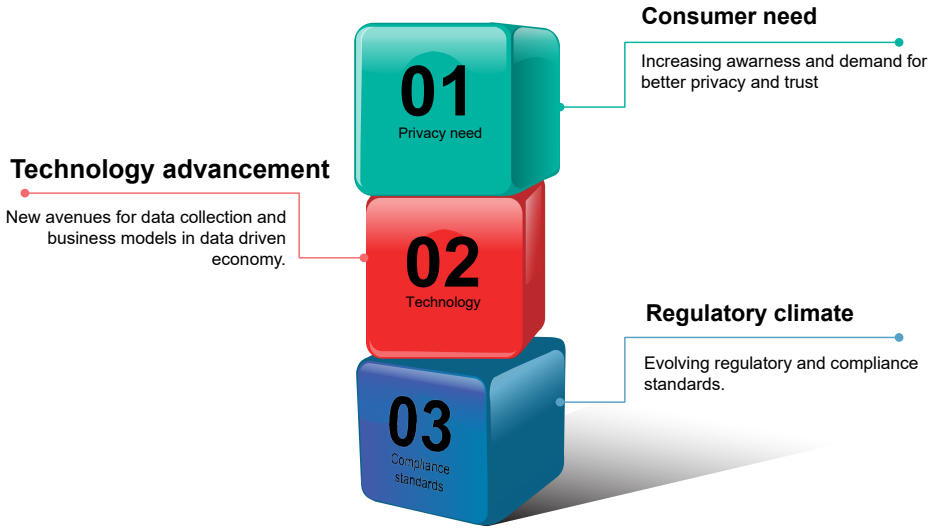


Figure 1 - Trends driving privacy which form barrier to innovation.

Control over data

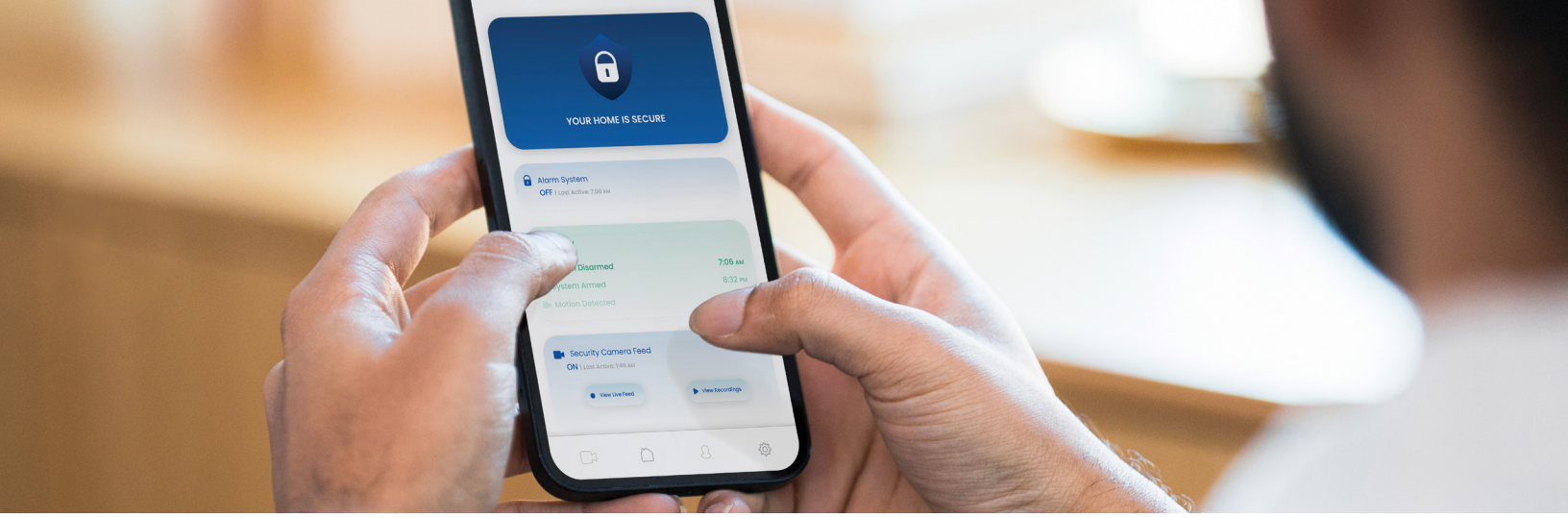
Users want more control over their data. Enterprises have built tools like web browsers with built-in cookie blockers, ad-blocking software, or incognito browsing modes to give consumers greater control of their personal information. As individuals become more aware of their privacy rights, they are willing to switch to competitors that offer more transparent and privacy-friendly services.

Advancements in technology

Advent of 5G, Metaverse and Artificial Intelligence powered by Internet of Things, are creating new opportunities for data collection and analysis. Data used widely has given rise to new privacy risks and biases. Further, COVID-19 pandemic has accelerated the shift to remote work and digital workplace, leading to an increase in online data collection and processing. Enterprises are working to balance the protection personal data with the pace to innovate and deliver digital services.

Changing regulatory climate

The approach to privacy and personal data protection varies across cultures, hence the difference across different countries. At one extreme, there is the absence of cross-border data flow regulation, and the other end there is a mandate to store data locally and take approvals by relevant authorities. 86% of the startups perform cross border data activity and have become micro multi-nations boot strapping on digital platforms like Amazon, E-Bay, Facebook, or Alibaba. By year-end 2024, Gartner predicts that 75% of the world's population will have its personal data covered under modern privacy regulations. Non-compliance with these rules when collecting data can lead to huge business costs. Enterprises are under increasing pressure to implement privacy-enhancing technologies and processes to follow these regulations and build trust with customers.



These trends have pushed the adoption of privacy enhancing technologies which can equip enterprises to protect personal data while delivering innovative products and services.

Innovation and privacy do not have to conflict with each other. Enterprises need to prioritize both innovation and privacy.

This whitepaper aims to achieve this balance through the adoption of privacy-by-design principles into the design of data pipelines for the right level of data protection and transparency for the choice to individuals on the collection and use of their data.

Enterprise privacy challenges and barriers to innovation

Innovation and privacy are both important for organizations, but they can sometimes have conflicting priorities. On the one hand, innovation often requires organizations to collect, process, and analyze substantial amounts of data, which can raise privacy concerns if not done responsibly. On the other hand, protecting privacy is a barrier to innovation if it prevents organizations from accessing and using data in ways that could drive new insights and opportunities.

For instance, a Europe-based medical manufactures who need access to patient data from US hospitals to build wellness and fitness monitoring devices. To design the right products, the manufacturers need right data on patient's health.

As per privacy regulations, sensitive patient health data cannot be for commercial product building and analytics. Unable to transfer data across the Atlantic, this medical manufacturer's product development and R&D fell behind by 6 months. The enterprise had started building data pipeline to move data, automate the flow of data, and ensure that the right data is available for application testing, R&D, Analytics and AI based decision making.

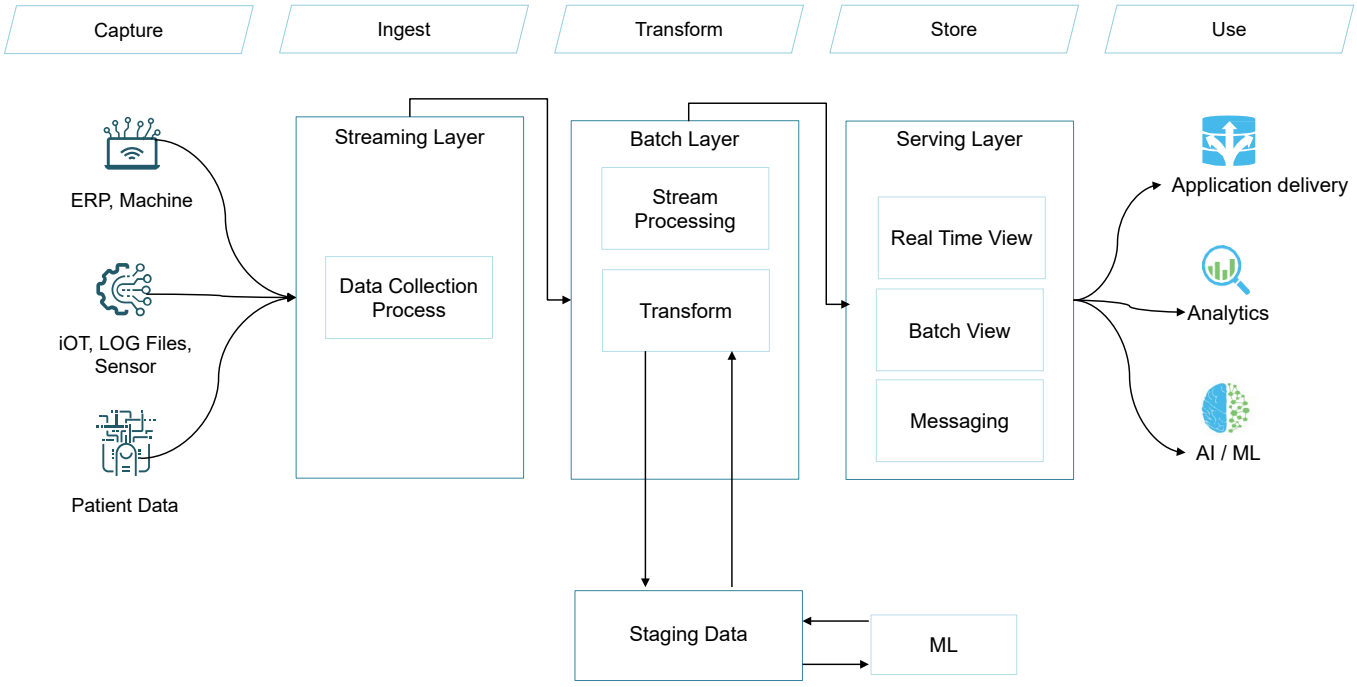


Figure 2 - Pipeline for analysis of patient data for device design

The enterprise data pipeline is built as a series of processing steps to prepare enterprise data for analysis, AI/ML and application delivery. A data pipeline drives efficiency across teams allowing the data from various sources, applications and user data to be shared and processed into one unified feed. The pipeline provides consistent and updated data but also presents key privacy risks across 5Cs.

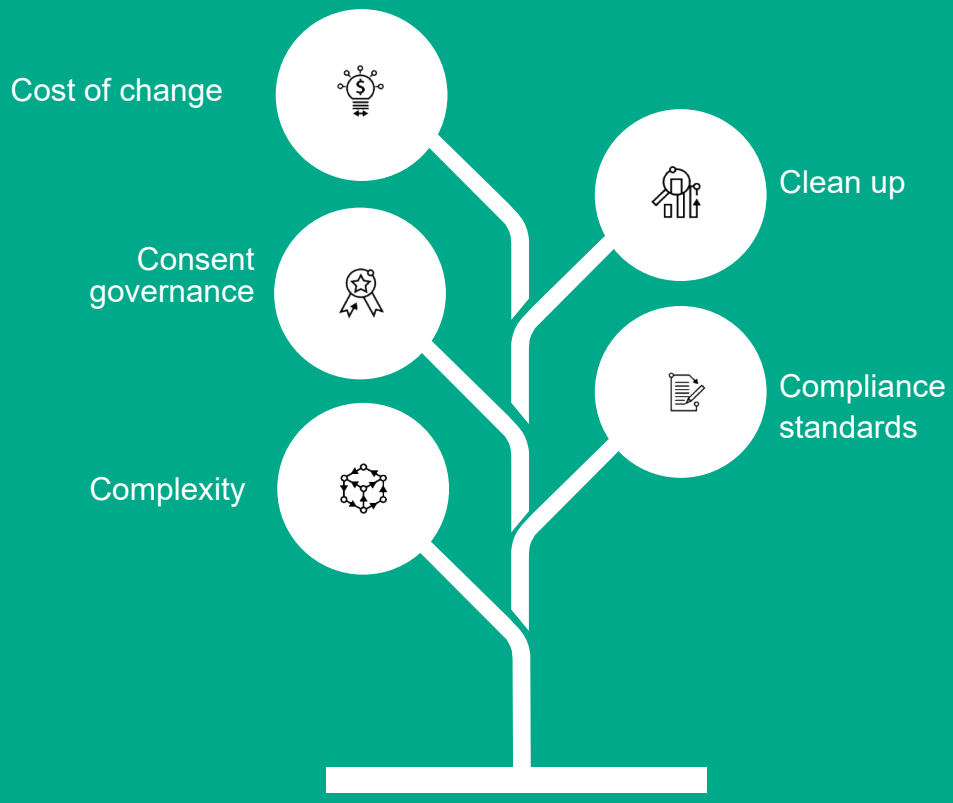


Figure 3 - 5C Privacy challenges which act as a barrier to Innovation.

Complexity

The data pipeline ingestion involves collecting data from various sources, such as databases, files, or APIs, and bringing it into a centralized repository for processing. There are no checks and balances in the collection of sensitive data. For instance, Meta's systems had grown increasingly complex and untraceable that engineers found sensitive data such as address details in six thousand data tables.

Compliance standards

Data processed for analytics and decision making need to be aligned with compliance standards of the land and business domain. Individuals have the right to access, correct, or delete their personal data, as well as the right to object to or restrict the processing of their data. Enterprises should implement mechanisms for individuals to exercise their data subject rights throughout the data pipeline.

Consent governance

Data driven economic models have made enterprises build data centric services and pipeline without building the right consent governance processes. Individuals can provide informed consent for the collection, processing, and use of their personal data. It is 40% more expensive to build consent management process after building a fully functional.

Cost of change

Data pipelines have become so complex and unwieldy that companies might not even know whether they are complying with regulations. As governments increasingly embrace Privacy by Design legislation, tech companies face a choice: either start from scratch or try to fix data pipelines that are old, extraordinarily complex, and already non-compliant.

Clean up

Enterprises must focus on collecting and processing only the data that is necessary for a specific business purpose and ensuring that it is deleted or anonymized when no longer needed. Without secure disposal, data minimization would be very difficult for organization to sift through large data volumes.

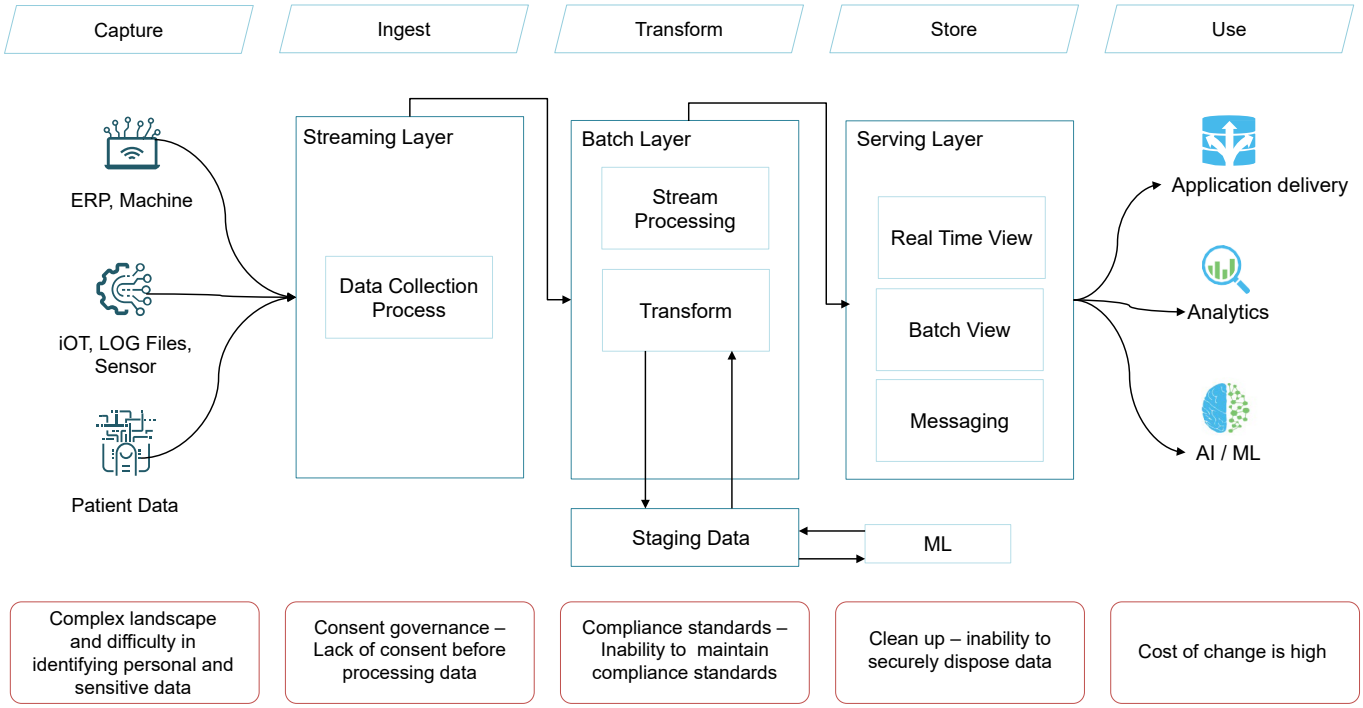
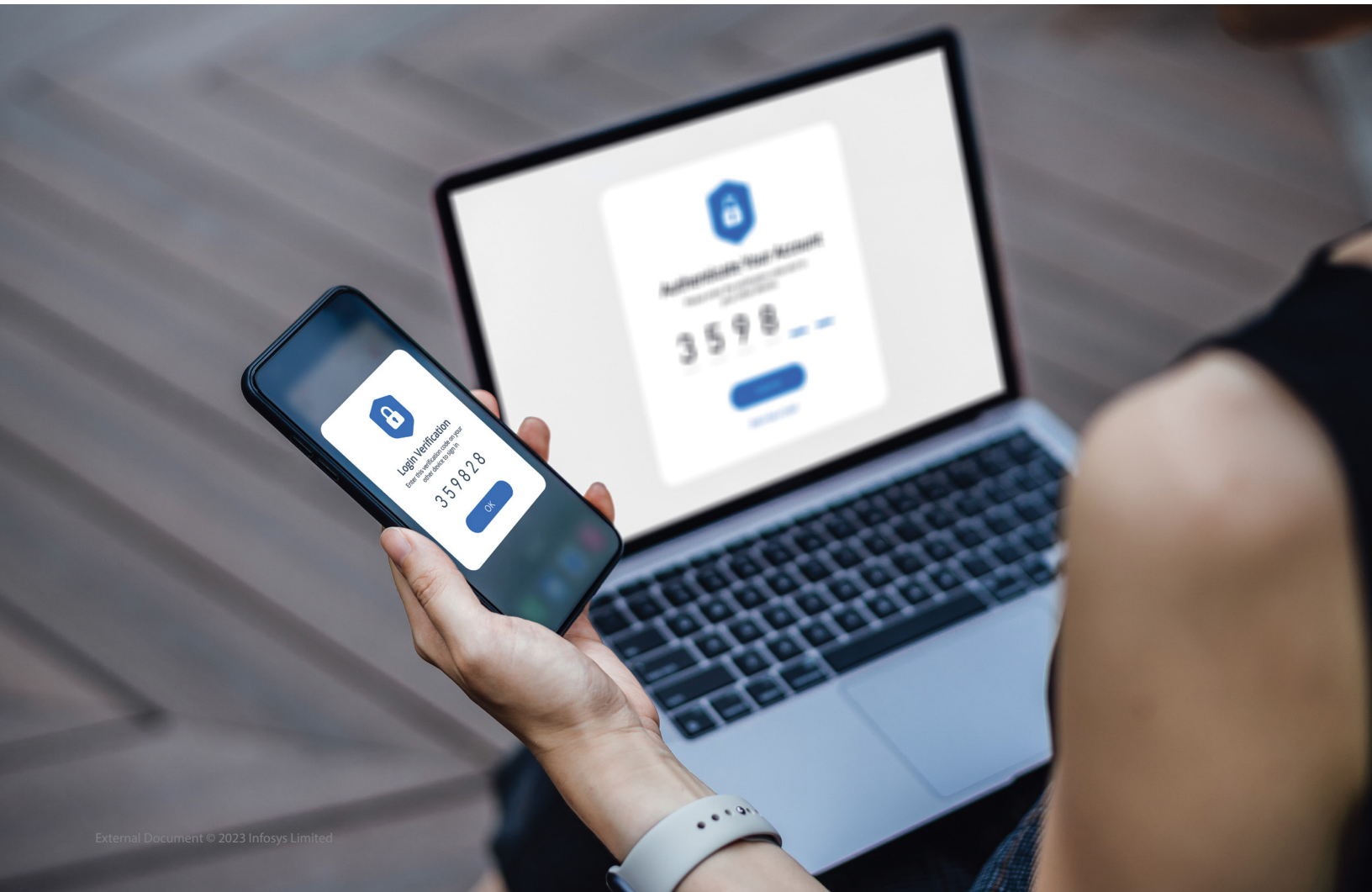


Figure 4 - Privacy challenges of a traditional data pipeline

Lack of privacy controls increase the risk of privacy breaches. Enterprises not managing personal data resulting in loss of trust and inability to create new business opportunities that rely on the responsible and transparent collection and use of personal data. Privacy controls should be designed and developed at the initial stages of the pipeline building rather than an afterthought



Implementation approach – Reference architecture for Enterprises

Privacy-first data pipelines are becoming increasingly important for enterprises across various industries, as they help to protect sensitive data and build trust with customers. By adopting privacy-by-design principles and implementing appropriate data protection measures, enterprises can prioritize privacy throughout the data lifecycle and ensure compliance with privacy regulations.

Apple has implemented a privacy-first data pipeline in their products and services. They collect the minimum amount of data necessary to provide their services and incorporate Privacy by Design principles into their product development.

A privacy-first data pipeline is designed to prioritize privacy throughout the data lifecycle, from data collection to analysis and distribution. This means that privacy considerations are built into every step of the pipeline, with a focus on minimizing the collection and use of personal data and ensuring that data is protected throughout the pipeline.

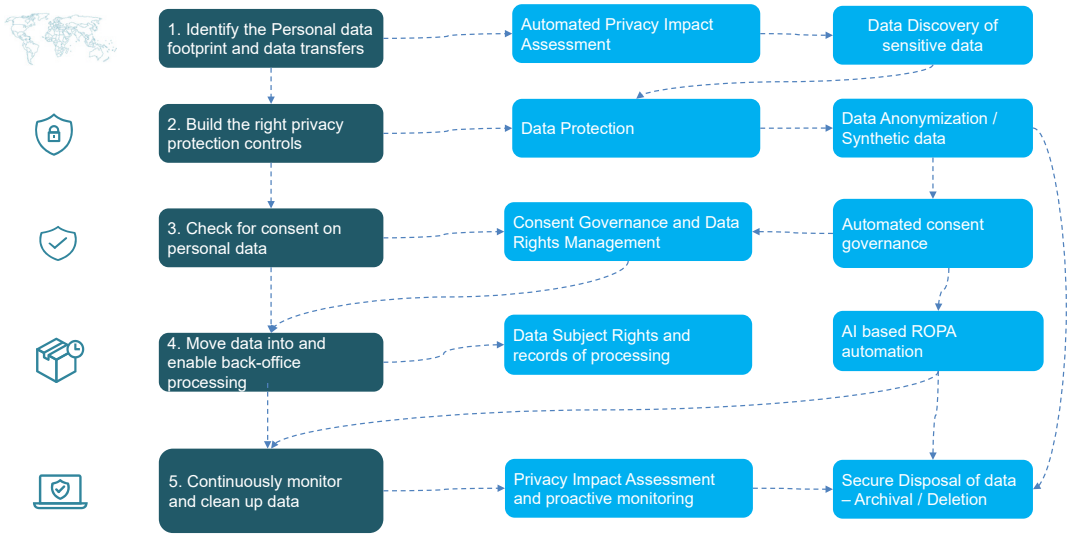


Figure 5 - Process steps to build a privacy first data pipeline.

Discovery of sensitive personal

This involves incorporating privacy considerations into the design of the data pipeline, from the data collection process to the data analysis and distribution. This can help to ensure that privacy is prioritized throughout the pipeline and that privacy risks are identified and addressed early in the process through a comprehensive data discovery process and automated privacy impact assessments.

Data protection

This involves implementing appropriate security and access controls to protect personal data from unauthorized access, use, or disclosure. This can include encryption, access controls, and data backup & recovery procedures. Many financial institutions, such as banks and insurance companies, have implemented privacy-first data pipelines to protect sensitive customer data and ensure compliance with privacy regulations. For example, JPMorgan Chase has implemented Privacy by Design principles in their data pipelines to protect customer data and build trust with their customers.

Transparency

This involves providing clear and concise information to data subjects about the collection, use, and sharing of their personal data. This can include implementing appropriate privacy notices and obtaining consent for data collection and use. A global retailer headquartered in Europe built a unified platform for collecting consent, processing, and managing access requests to personal data to build transparency into their data driven decisions.

Data subject rights

This involves providing data subjects with the ability to exercise their privacy rights, such as the right to access, correct, and delete their personal data. One of the largest telecom providers based out of the US used automated Data Subject Rights (DSR) process to enhance their data pipeline to ensure the end users could exercise their data rights on the data which is collected and processes in their application landscape.

Data minimization

This involves collecting and storing only the data that is necessary for the intended purpose. This can help to reduce the risk of unauthorized access or use of personal data.

Diagnose

The first step would be to take an inventory of all the personal data collected and processed by the organization, including data types, sources, and storage locations. Teams can analyze message definitions or algorithms identify sensitive data is the complex live data feeds to understand how variables flow relationships within data. Requirement designer and discovery engine of Broadcom TDM provides rule-learning algorithms, which identifies any missing data that data analysts can apply to reverse-engineer needs to be added for complete, rigorous testing. For a financial institute in Asia, data discovery was performed to identify the sensitive. A tool for assessing the privacy risks and implications of a particular project, program, or system, to identify and mitigate potential privacy risks.

Design

Post diagnoses, the platform should enable the design for effective privacy protection in a pipeline. Examples of critical processes include normalizing data before ingestion and verifying the degree of de-identification needed with the data. A process to assess risks in data clean rooms is also essential as it helps parties identify the specific privacy risks the analysis might encounter, and it guides them in identifying the right risk mitigation strategies.

Deliver

The focus of the pipeline is to deliver Data Protection at scale. This includes:

- Techniques for protecting the privacy of personal data by removing or replacing identifying information. Privacy first data pipelines must effectively enable promise of data protection, sharing, and analyzing enterprise data without privacy risks. These core data protections include strong identity and access management capabilities, encryption on data entering the pipe.
- Synthetic data - CA Test Data Manager capabilities to model and create synthetic data that covers functional variations. And with access to a comprehensive set of combinable data-generation functions, seed tables, and variables, test teams can create realistic data tailored to their specific testing and development needs.
- Advance solution should also include confidential computing and privacy-enhancing technologies — such as differential privacy. This can be enabled from Infosys IIN.

Defend

The pipeline should provide capabilities to dispose data which is no longer used keeping into data minimization considerations. Effective defense is built by capabilities to provision to detect, report and respond to data breaches. There should be capabilities available for incident response plans, breach notification requirements, and post-incident analysis. Further, there should be effective monitoring and reporting on compliance with privacy regulations, including data protection impact assessments, privacy audits, and regulatory reporting requirements.

By implementing a reference architecture for privacy, organizations can ensure that personal data is collected, processed, and shared in compliance with applicable privacy regulations, while also building trust with customers and other stakeholders.

Enabling data pipeline for the future

Privacy is a subjective concept, which is differentiating based on many factors. Ensuring privacy in data-driven organizations requires a comprehensive approach that incorporates Privacy by Design, data minimization, consent management, data security, privacy impact assessments, and transparency and accountability. By prioritizing privacy and data protection, organizations can build trust with their customers, comply with privacy laws and regulations, and create new business opportunities that rely on the collection and use of personal data in a responsible and transparent manner.

As Shamim quotes "Data privacy is a critical imperative in today's data driven world fueled by growth of AI/ML based applications. Data pipelines engineered with privacy-first principles help provide the framework for data protection, while allowing un-impeded access to data as and when needed. Broadcom provides industry leading tools for sensitive data discovery and identification, masking and desensitization and provisioning on demand that are especially well suited for the needs of agile application development and delivery that drives enterprise innovation.

We are proud to partner with Infosys to integrate our data privacy tools with their broader data privacy and governance solutions that enable enterprises provide equipoise between both innovation and privacy compliance needs."

Enterprises should balance privacy risk and need for high value data when selecting privacy first data pipeline. The pipeline can have varied level of privacy controls implemented based on the complexity of the data processed and the level of personal data consumed by the pipeline.

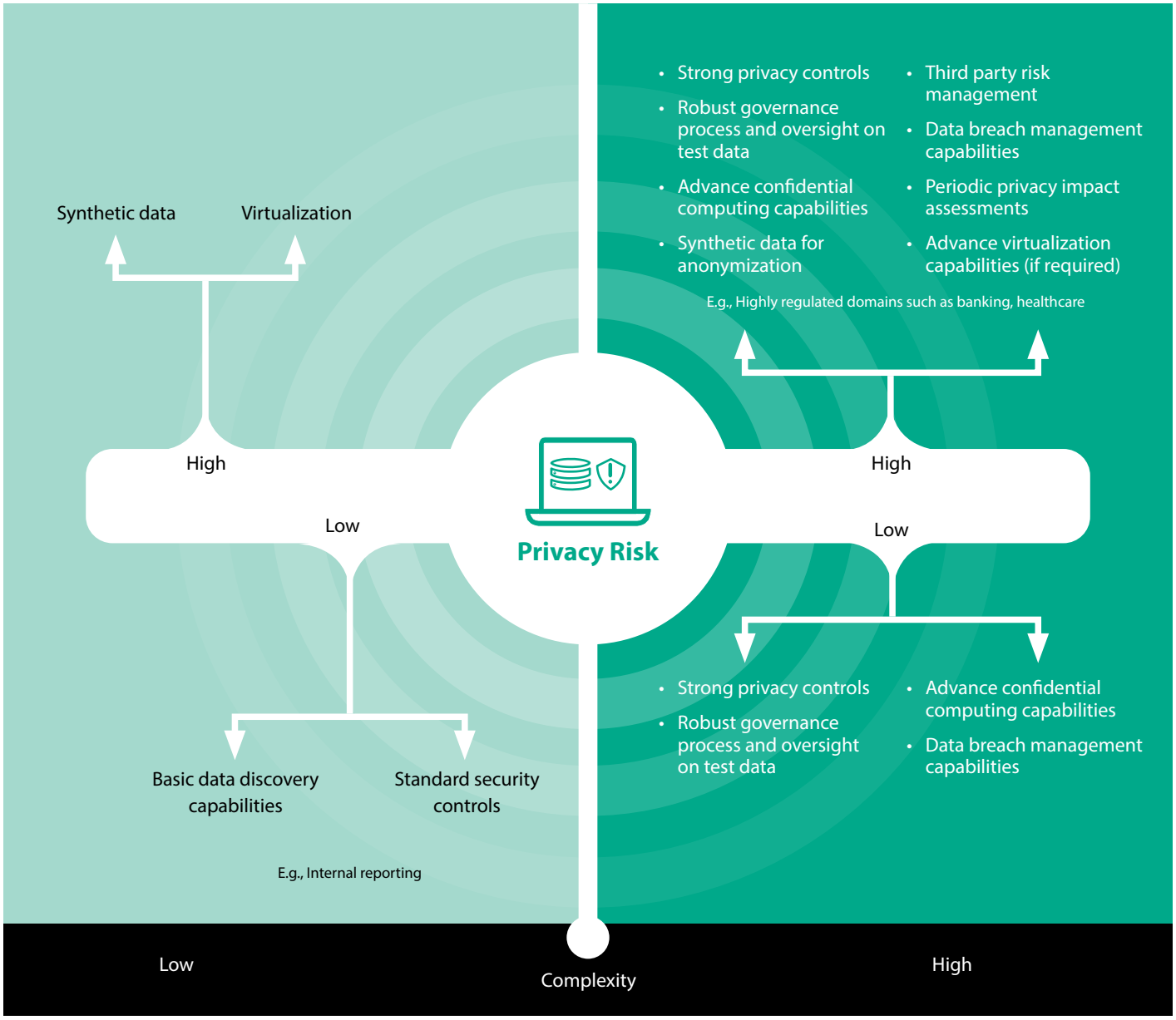


Figure 6 - Reference Architecture for a privacy first data pipeline

By taking a privacy-first approach to innovation, organizations can build trust with customers and other stakeholders, and create new business opportunities that rely on the responsible and transparent collection and use of personal data.

Operating Element	Traditional data pipeline	Privacy first data pipeline
Data Privacy Strategy	Functional and line of business based with Privacy as a After thought	Cross functional – Platform mind set strategy with Privacy built in every transaction
Experience	Focus on Efficiency, Usability, one size fit all for Privacy	Privacy First Inclusive design for the personalized content for each user with the right consent in place
Value Chain	Review and offline periodic reviews	Distributed agile process which are continuously monitored
IT Systems	Designed for known requirements and are feature driven with moderate automation	Design for evolvability, observable and platform based with Privacy First Design
Business Decisions and AI	Deterministic, What – if analysis basis Data Risk Assessment Reactive approach to Data breaches	Connected and curated data risk assessment Predictive and prescriptive stance on Data Breaches
Talent	Legal Advisors, Auditors, Lawyers	Multi - skilled Technical Functional teams with knowledge of Regulations

One is to safeguard the privacy of individuals and their personal data. Regulation differs based upon the varied approach to privacy and personal data protection across cultures. Countries may also restrict the flow of data, or mandate that data be stored locally, to meet other regulatory aims such as access to information for audit purposes. Restrictions to data flows might also arise for the protection of information deemed to be sensitive from a national security perspective, or to enable national security services to access and review data.

By taking a privacy-first approach to innovation, organizations can build trust with customers and other stakeholders, and create new business opportunities that rely on the responsible and transparent collection and use of personal data.

Reference

1. World Development Report, 2021
2. Gartner Identifies Top Five Trends in Privacy Through 2024, STAMFORD, Conn., May 31, 2022
3. Shamim A, Broadcom DevOps CTO - https://www.linkedin.com/in/shamim33

About the Authors

Karthik Nagarajan
Practice Manager and Senior Industry Principal

17+ years of Data privacy and Cx (Salesforce and CRM) focused on Practice building, Solution Architecture and Business Development.

Deepak Dinasi
Enterprise Solutions Architect at Broadcom

Over 20+ years of experience in DevOps focused on Data. Management building solutions involving complex cross- functional technologies and providing technical thought. leadership through the application delivery life cycle.

For more information, contact askus@infosys.com