# Healthcare Aware Distributed Tracing™

## Using Machine Learning to Solve the Mystery of Tracing PHI Across Microservices

**ClearDATA's Chief Privacy and Security Officer and Founder Chris Bowen and the VP of Architecture Adam Greenfield share insights on:**

- The move to microservices and containers in healthcare
- Using machine learning with distributed tracing
- The benefits of Healthcare Aware Distributed Tracing™

*This whitepaper is designed for healthcare execs looking for a foundational understanding of what distributed tracing means to healthcare in the public cloud, and how Healthcare Aware Distributed Tracing™ can be the game changer needed to allow healthcare organizations to trace PHI across Kubernetes environments – a necessity for PHI inventories and audits as well as to comply with complex regulatory frameworks including GDPR.*

ClearDATA®
SECURE · HEALTHCARE · CLOUD

The rapid evolution of healthcare IT systems is creating an increasingly agile, scalable and dynamic environment that is also increasingly complex, regulated and attractive to hackers. Emerging technologies are undeniably providing new opportunities to innovate and radically improve patient care, but the obligation to 'do no harm' and protect patient data looms in IT environments that make tracking and tracing patient data difficult.

## Microservices and Containers in Healthcare

Modern design principles in IT environments are turning to microservices and containers. Containers allow software to rapidly deploy and run with reliability when moving from one compute environment to another, as when moving from development, to staging to production. Microservices provide more efficient consumption of compute services with the ability to quickly scale up or down, paying only for what you need in the public cloud. Additionally, containers and microservices can provide protection to PHI because rather than PHI living and moving throughout one singular monolithic environment, packets are moving in much smaller batches that in essence, limit the blast zone if a compromise occurs. If a container is compromised only a small packet of information can be accessed rather than the entire database of patients, for example. But with this advantage of placing only portions and parcels of data in each container comes the huge challenge of knowing which data is where, especially when in transit across microservices in complex IT environments. The need to know where PHI is – at rest and in transit – has never been greater.

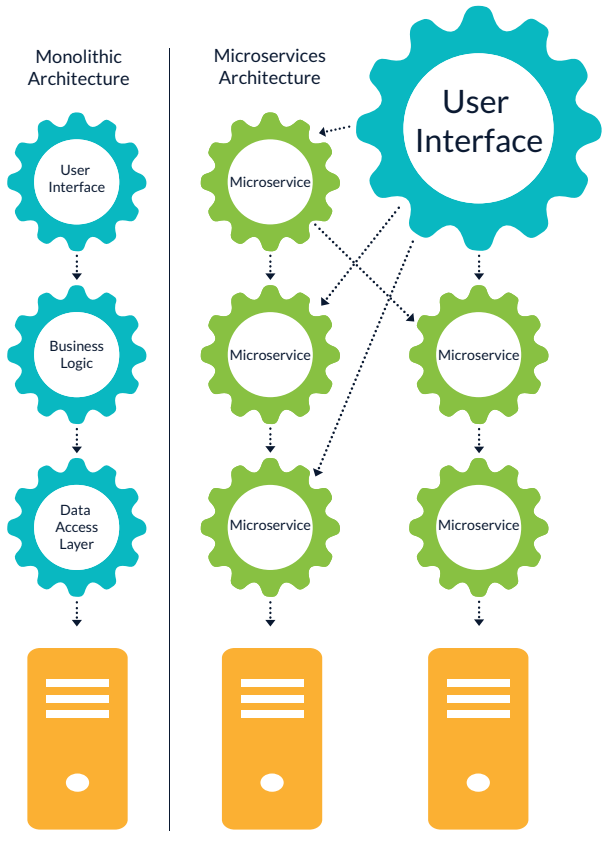


*Image 1: This diagram shows the changes from monolithic architectures to microservices and containers. Microservices can provide more efficient compute resources, scale with better agility and provide better protection to PHI.*

While PHI inventories are not only a good practice, but also a requirement, new regulations that protect citizens rights such as the EU's General Data Protection Regulation (GDPR) provide EU citizens with many rights, including the right to be forgotten. To be able to erase a person's records, the administrator must know where their PII or PHI is.

## Distributed Tracing

Distributed Tracing systems have a rich history in the technology world, and they are seeing renewed interest recently. Distributed tracing is a concept for microservices that can trace data requests as they are processed by microservice architectures. Patterns involving containers and microservices are being utilized in many organizations as they look to leverage the power of the modern cloud computing to meet their business objectives for scaling, agility and security. Distributed tracing creates a record of a request as it passes through each component of

a complex system. The level of detail and the volume of information provided are dramatically higher than traditional simple logging-based systems. The difference in terms of available information could be as stark as the difference between a clinician performing an external physical examination versus having access to real-time medical imaging of a patient.

Currently, network traffic inspection tools are often challenging to leverage because of the best practice encryption of data in transit which is often required to operate in certain environments, especially healthcare. Additionally, modern application developers are implementing secure overlay networks and service meshes to ensure data in transit is protected consistently without requiring these developers to consciously integrate encryption into all of their network connections.

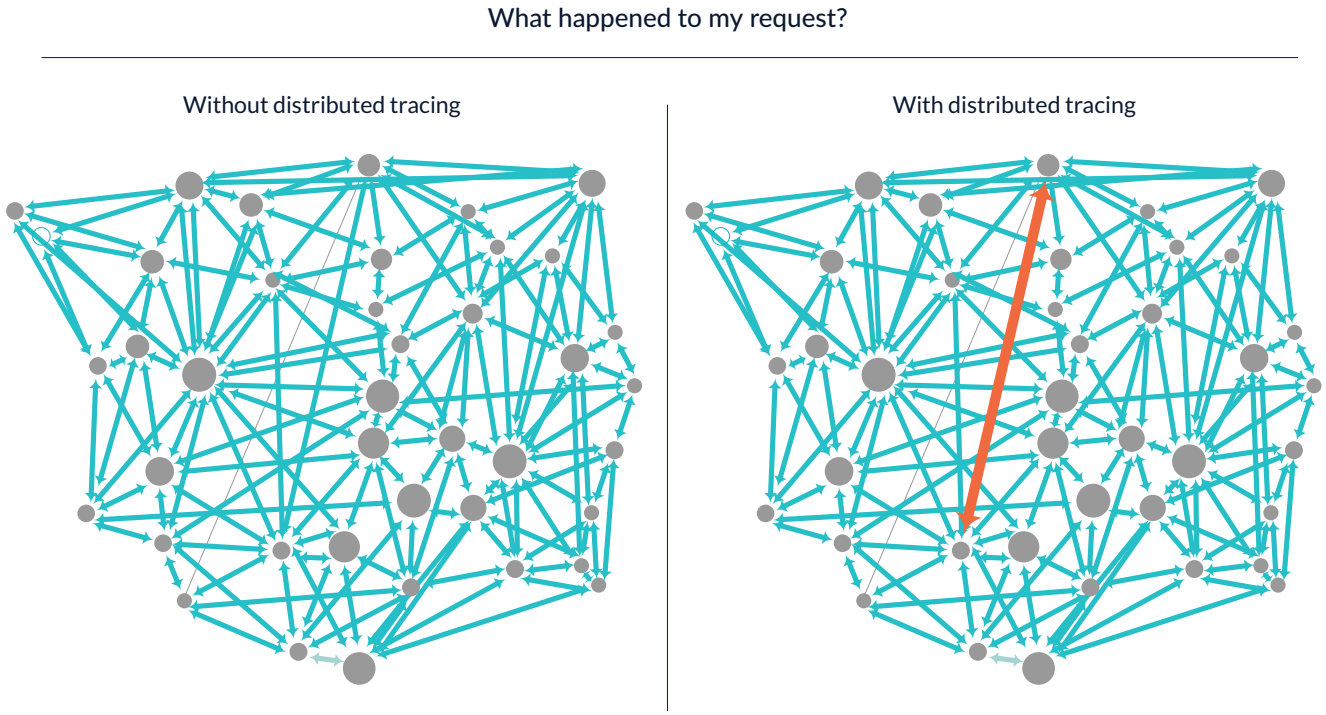


*Image 2: Left: Without tracing, it is difficult to track requests and determine potential bottlenecks or errors. Right: With tracing, a request's path can be easily determined to help pinpoint potential issues.*

Organizations can either create an audit record of all data that flows through the application, or they can build a PHI audit record into each microservice. Both of these routes take significant development investment, which takes time away from what the organization should be building.

Distributed Tracing APIs have been used to identify performance issues or diagnose operational trouble within an application, but there has not been a tracing tool that can determine if the traffic is PHI. As microservice architecture is widely adopted across an organization and scales, it can be difficult for the PHI audit requirements to be maintained. This can cause delays in software development or a gap in the risk posture of an organization.

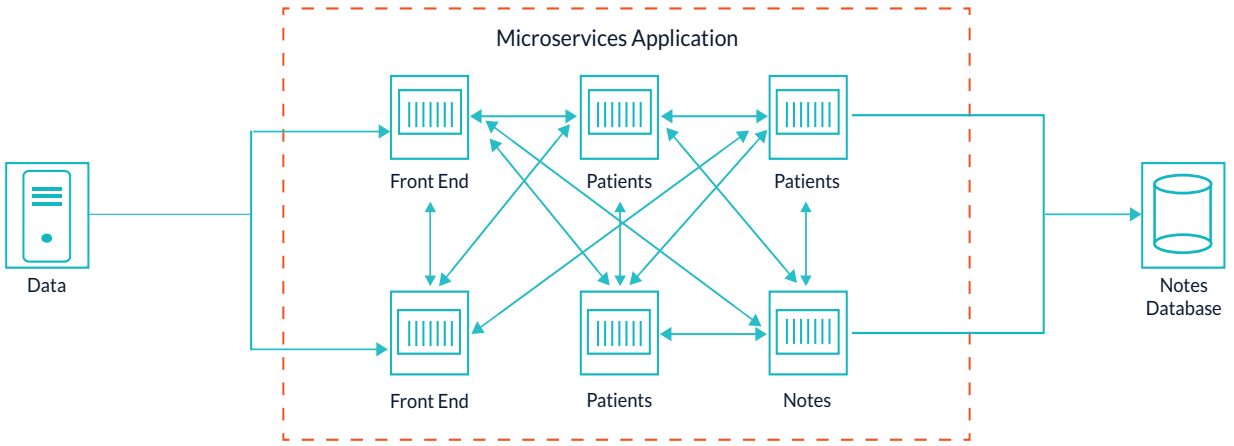


*Image 3: Diagram depicting a small microservice application.*

# The Geography of Data in the Public Cloud

In many ways, cloud innovation has increased resiliency, introduced pay-as-you-go models, and unharnessed the potential for eliminating data silos globally. Much of this is a positive development. However, when international regulations promulgate data locality requirements, an organization can quickly get sideways with regulators when a dataset that should be located only in Germany, for example, finds its way to Japan with the click of a few buttons in a cloud console. To further complicate the situation, consider that many third-party integrations also make it easy to transmit massive amounts of data everywhere.

# Kubernetes and Healthcare Aware Distributed Tracing™

Advancements in machine learning, including natural language processing, and on-going protocol standardization are making it possible to know which request paths involve PHI, thereby creating auditing and compliance feedback about data in transit within an application. Simply put, this has the potential to change everything when it comes to complying with PHI security and privacy guidelines and regulatory frameworks.

ClearDATA's Healthcare Aware Distributed Tracing™ software solution is built on top of industry standard open source tools as well as cutting edge machine learning models geared specifically to healthcare. With this solution, healthcare organizations can plug this into an existing Kubernetes-based application and immediately start seeing tracing reports that are focused on healthcare data. Requests that flow through a Kubernetes pod will be evaluated against healthcare aware machine learning models that can identify if the data contained in the request are PHI or not. This tracing can be implemented in Kubernetes with ease, allowing healthcare organizations to start tracing PHI without any development effort.

Imagine looking at a dynamic map on your big screen, laptop, or mobile device, that geographically displays your cloud systems throughout the world. You see bi-directional lines of network traffic containing sensitive identifiers from one location to another. Clicking on one of those lines or markers reveals the specific

identifiers, the systems in which they transit, the times and dates of that transit, as well as the type of identifier. An additional click on the data icon reveals the state of compliance of the specific packet. The flows are logged and become a part of your record of evidence to help you create and maintain your record of evidence for auditors, patient requests for an accounting of disclosures, and if necessary, regulatory investigations.

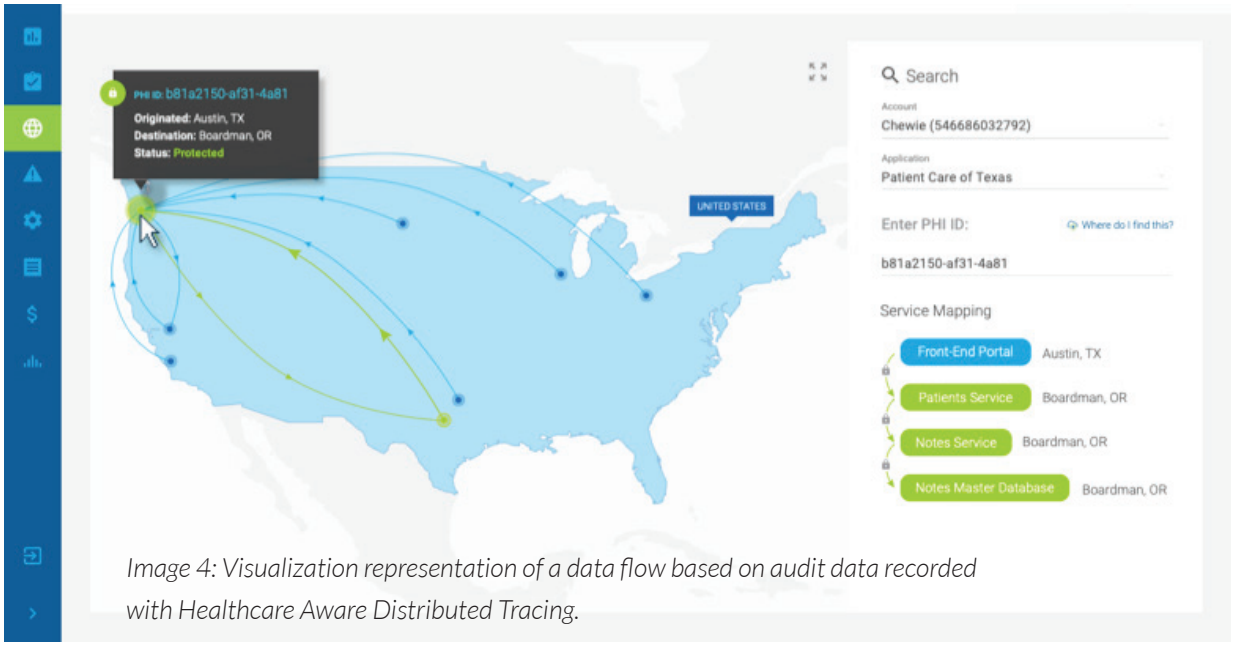


*Image 4: Visualization representation of a data flow based on audit data recorded with Healthcare Aware Distributed Tracing.*

This solution will enable healthcare organizations to build a strong risk posture in their modern, microservice-based applications and environments. This automated tagging and auditing allows the development team to focus on their core application, rather than building auditing into each microservice to satisfy various compliance requirements.

ClearDATA's Healthcare Aware Distributed Tracing provides a solution that allows the developers to maintain their velocity, while the security and compliance team is comfortable knowing that all data is being audited.

# Who Can Benefit from Healthcare Aware Distributed Tracing™

Within a healthcare environment, several vital roles will benefit from this promising new technology as it provides visibility of patient data flowing through complex systems.

## *The Application Developer*

The brains behind innovation in healthcare often belong to development teams charged with bringing innovation to the organization. These teams are bringing innovations to healthcare that are making possible increased patient engagement, precision medicine, deep data insights, home monitoring and sequencing of genomes, to name but a few. This effort to innovate often meets the buzz saw of strict security, privacy, and regulatory compliance requirements and ultimately slows the pace of innovation.

## *The Compliance Professional*

Part of the requirements for protecting sensitive information is inventorying the data, and applying safeguards to that data throughout its flow. While it is relatively easy to document a patient data inventory in a working session, this activity is mostly academic. When compliance officials walk the halls of the organization and examines the data in every component of an on-premises data center, or peers into a cloud environment to gain insights to the data lifecycle, what they find is that data sprawl is alive and out of control. Without Healthcare Aware Distributed Tracing™, compliance professionals may never achieve the real awareness necessary to create and maintain an accurate inventory of their data.

## *The Privacy Official*

Privacy officials must employ new approaches and sophisticated technologies like machine learning and data science to classify data because personally identifiable data is created, used, distributed, maintained, stored and destroyed in volumes that have scaled to a scope never imagined in human existence. To further complicate matters, data that should be classified as sensitive is likely encrypted (or should be), and may also flow across borders, across data centers, or even through different continents. The responsibility remains, however, to create a living data inventory and ensure that safeguards are applied to ensure data protection, privacy and compliance including data locality regulations like GDPR.

## *The Security Officer*

Similar to compliance officials, this overburdened role is responsible for applying the security controls to the sensitive data system. They must provide insights to their stakeholders for authorized and unauthorized access, security events, control failures, log flows, and data integrity, all within the world of data sprawl.

### *The Patient*

At the beginning and end of all of this, patients are the ultimate stakeholders. There are more than 194 million active patients in the U.S. healthcare system – a startling 59.8 percent of the country's population. While more and more of their sensitive health information is being gathered, stored, and shared, patients have very little control or visibility into their data, yet have tremendous vulnerability to having that data compromised or stolen. If patients wish to monitor their data; or know where it has been and with whom it has been shared, they have had very little opportunity to have their questions answered. And, if they are among the many unfortunate who have had their data compromised or stolen for nefarious purposes, monitoring it to protect their rights, their safety and their credit can be a near impossible task. Healthcare Aware Distributed Tracing™ gives these patients insight, voice and power into owning their own healthcare data.

### *Policy as Code*

The static data inventory approach used across healthcare organizations only provide a point-in-time, limited glimpse of the general location of sensitive data, but ClearDATA's Healthcare Aware Distributed Tracing software service changes the game. This software service allows organizations to implement policies using code that direct microservice systems to enforce data locality requirements or provide insights as to which partners are using which data elements - and whether they are doing so appropriately. The software service could even be used to provide records of data processing activities required by GDPR, or power an organization's ability to enable the HIPAA requirement for fulfilling requests for an accounting of disclosures.

## Closing the Gap in Audit Processes

Microservices can create gaps in an audit picture. This software solution builds a foundation of secure, reliable PHI tracking through modern microservices environments, enabling healthcare organizations to build a better audit posture and comply with current and increasingly complex regulatory frameworks. By accurately tracking PHI, a security incident has more limited damage because an organization can definitively convey to the Office for Civil Rights (OCR) the actual records compromised rather than having to work under the OCR's current policy of assuming all records were compromised if the organization is unable to document which ones, because of which organizations now often face fines in the millions of dollars. In fact, if an organization is able to clearly identify just those records that were compromised in a microservice environment they may find that they are now de-escalated from a breach notification to a much less damaging security incident. Imagine being able to prove in a breach investigation that 90% of a dataset was safe and that only a fraction of the records were compromised. The cost of processing and storing sensitive data could plummet, and the patient state of mind could be protected.

Another use case for Healthcare Aware Distributed Tracing improving your culture of compliance and audit stature is with GDPR's "Right to be Forgotten." This is not only a concern for those dealing with EU citizen records as states within the U.S. including California are now beginning to implement regulations that require organizations to find all data related to a person and delete it upon that person's request. Without Healthcare Aware Distributing Tracing, finding and deleting a patient's PHI could be impossible to comply with.

# Innovating within Compliance Frameworks

Using the Healthcare Aware Distributed Tracing software, ClearDATA builds upon its strong culture of Privacy by Design and Defense in Depth by enabling:

- ✓ The adoption of more innovative technology to build new apps while managing multinational regulatory frameworks, thus modernizing the healthcare industry.
- ✓ Assurance for PHI inventory requirements with insight on patient records during each stage of the application while creating a reliable and continually updated inventory; and
- ✓ Allowing deeper analytics to gain insights that combine to better detect fraud, discover areas for cost optimization or improve customer service.

ClearDATA's Healthcare Aware Distributed Tracing™ solution is implementing with beta customers in Q1 of 2019 and will be generally available in late Q2 of 2019.

## About ClearDATA

Healthcare professionals across the globe trust the ClearDATA HITRUST-certified cloud to safeguard their sensitive data and power their critical applications available across the major public cloud platforms. For healthcare organizations, customers receive one of the most comprehensive Business Associate Agreements (BAA) in the industry, combined with market-leading healthcare-exclusive security and compliance solutions, and multi-cloud expertise. ClearDATA's innovative solutions protect customers from data privacy risks, improve their data management, and scale their healthcare IT infrastructure, enabling the industry to focus on making healthcare better by improving healthcare, every single day. To learn more, please visit https://www.cleardata.com