



Healthcare **Big Data** Warehouse in the Cloud

Accelerating Population Health, Precision Medicine and Interoperability Initiatives



Healthcare **Big Data** Warehouse in the Cloud

Accelerating Population Health, Precision Medicine and Interoperability Initiatives

Table of Contents

Page 2

Introduction

Page 3

The Big Data Challenge

Page 5

Big Data Warehouse in the
Cloud

Page 6

Balancing Privacy and Data
Utility

Page 8

Cloud Benefits

- Cost
- Flexibility
- Get Started Quickly
- Proof

Page 10

Cloud-Based Big Data Ware-
house Checklist

Data is an increasingly critical part of how healthcare organizations operate. The patient data that resides in disparate clinical, diagnostic and financial systems will play an important role in how providers document quality and outcomes, manage new reimbursement models, address population health trends, and deploy precision medicine and other clinical approaches to improve clinical outcomes.

The aggregation of this data is also important for research purposes, as more healthcare organizations undertake clinical research projects and also search for ways to more quickly access research results and implement best practices based on the findings. The challenge here is capturing complex data types, both structured and unstructured, that exists in a variety of formats (images, files, EMR data, etc.).

Data also plays a key role in precision medicine, which is dependent on the aggregation of genomics data as well as environmental and lifestyle information, medical history, and other data. Collecting, storing, extracting value from that data is a task that requires a more advanced approach than most healthcare organizations have available in-house.

Big Data and analytics tools help companies use these large volumes of data to create actionable intelligence – to identify population health trends, spot emerging outbreaks, measure effectiveness of treatment, and develop better clinical responses more quickly based on empirical evidence. Creating a Big Data warehouse that can help your organization effectively collect, catalog, and analyze healthcare data can be expensive and complex.

If want to help your researchers and analysts get better insights faster, while also keeping protected health information (PHI) secure and private, you should consider a cloud-based Big Data warehouse.

The Big Data Challenge

Data is everywhere; making good use of that data is challenging because it resides in disparate locations and systems. Healthcare organizations will not only need to use this data in order to provide performance and outcomes information to payers, they will also need this data for internal analysis and research. Using population data to analyze health trends and effectiveness of care, and to identify performance gaps will help health providers improve patient health and the overall patient experience of care, as well as reduce the per-capita cost of providing that care. Hospitals have even been able to combine patient data with geospatial information and census information to identify localized causes of very specific disease and injury trends, and then work to address them.

Analytics is proving to be a huge potential in helping healthcare organizations achieve what the Institute for Healthcare Improvement calls Triple Aim: improving patient care

quality and satisfaction, improving population health, and reducing the per capita cost of care. But setting up a Big Data warehouse for analytics, gathering complex data into a central repository, and then on top of all that, assuring your privacy officer, your researchers, your board, and your patients that their data is secure can be challenging. If you take a do-it-yourself approach, it can be costly, time-consuming, and ultimately ineffective because of the lack of in-house expertise and technical knowledge at most healthcare organizations.

According to Gartner's 2015 Hadoop Adoption Study, skills gaps continue to be a major obstacle for 57 percent of respondents, while figuring out how to get value from Hadoop was cited by 49 percent of respondents. "While tools are improving, they primarily support highly skilled users rather than elevate the skills already available in most enterprises," the report notes. "Hadoop vendors are responding to this challenge by



offering a variety of training options. However, Gartner estimates it will take two to three years for the skills challenge to be addressed. Beyond skills, demonstrating the value of Hadoop is the second-highest challenge.”

Big Data capabilities will be more important as the need for these analytics capabilities expands beyond traditional research teams at academic medical centers. More clinicians are engaging in research and demanding faster access to research results in order to enable precision medicine and other initiatives that are highly data-dependent. Researchers and clinicians want to get to insights faster so that evidence-based medical practices can be more quickly implemented into actual care plans. As academic medical universities lead the way in applying their research, other institutions will follow.

In the past, researchers have developed ad hoc approaches, including consolidating data on their own laptops or desktop systems, which doesn't scale, especially in the case of omics type of research.

This is not only costly, time-consuming and potentially duplicates data sets, it also exposes the institution to the risk of data theft or a HIPAA violation – ultimately leading to costly

finances and potentially irreparable damage to the provider's reputation among patients.

Large, on-premise Big Data solutions offered by major enterprise software vendors can crunch this data, but they require months of planning and a significant up-front capital expenditure, in addition to the need for additional server hardware to manage and maintain. That approach makes it difficult for researchers to quickly launch projects in order to respond to more immediate data needs. DIY solutions may look attractive, but staff may find themselves downloading a Hadoop toolkit and realizing they've done the equivalent of purchasing a massive Lego set with no instructions.

The right cloud-based Big Data solution can solve this problem for you. Such a solution can accelerate the consolidation of data across the organization, while providing the right level of privacy, security, and governance to comply with internal and regulatory requirements.

Don't be deterred by concerns over complexity or privacy – technology exists today that not only can get you up and running quickly, but can also keep your data private and secure.

DIY solutions may look attractive, but staff may find themselves downloading a Hadoop toolkit and realizing they've done the equivalent of purchasing a massive Lego set with no instructions.

Big Data Warehouse in the Cloud



Implementing a Big Data warehouse using cloud infrastructure can accelerate research and analytics activities, and help get data into your researchers' hands faster. You can enjoy all of the scalability and economics of a data warehouse without having to build out the features and functionality, and without the cost of implementing the server infrastructure required to house and analyze that data.

Centrally located data isn't the only advantage that this type of data warehouse solution can provide. A cloud solution that provides indexing and cataloging features can make it easier for your team to create and publish data sets, search for information, and easily select the data required using self-service functionality. Data can be indexed, cataloged, and tagged with metadata at the point of entry, and then automatically standardized, encrypted, scrubbed, or processed so that it is available in the format required by your analytics applications. Further, data locked in documents in PDF format, such as X-rays, echocardiograms, and diagnostic reports, can be pulled out and made part of the data record for analysis. The data sets can be centrally curated, allowing easy access across the organization and multiple facilities.

Metadata and governance policies can guide the use of those data sets by authorized users (based on organizational policies) and track who used the data and how. A data warehouse built to easily integrate with third-party applications and analytics tools can help providers and research organizations leverage their existing investment in technology, while greatly expanding their ability to access and analyze data.

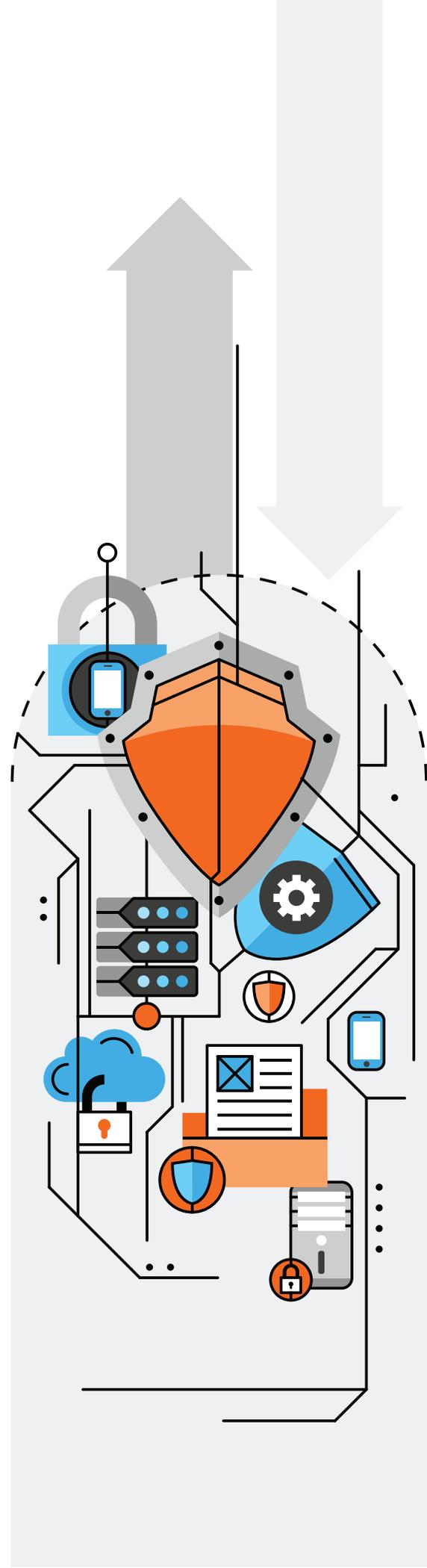
Balancing Privacy and Data Utility

Data security and privacy are top concerns for healthcare organization considering any type of cloud-based solution, let alone creating a massive offsite database of patient and treatment data. Both HIPAA and HiTECH call for stringent protection of PHI; as a result, privacy officers at many healthcare organizations have some times been reluctant to approve cloud-based projects.

According to the report "Big Data: Seizing Opportunities, Preserving Values," from the Executive Office of the President: "Though medicine is changing, information about our health remains a very private part of our lives. As Big Data enables ever more powerful discoveries, it will be important to re-visit how privacy is protected as information circulates among all the partners involved in care. Healthcare leaders have voiced the need for a broader trust framework to grant all health information, regardless of its source, some level of privacy protection."

Balancing the needs of the organization to protect data integrity with the ability of clinicians and researchers to use this data to improve care is one of the key struggles the healthcare industry will grapple with over the next several years.

"Though medicine is changing, information about our health remains a very private part of our lives." - Big Data: Seizing Opportunities, Preserving Values - Report from The Executive Office of the President



While the privacy team may have a tendency to over-state the risk of putting PHI in the cloud, clinical and research teams may understate it. There has been a perception in the past that healthcare data is not an attractive target for hackers, since most well-known data breaches have involved payment data. That's no longer the case. According to a research brief from Forrester Research ("Healthcare Organizations Face Mounting Pressure to Protect PHI"): "While this [perception] may have been true in the past, now the value of healthcare records far surpasses that of credit card information sold on the black market, fetching anywhere from \$20 for a single health record to over \$500 for a complete dossier on a patient with driver's license, health insurance information, and other sensitive information."

Security budgets for many healthcare providers are limited. Selecting a data

warehouse tool that complies with HIPAA and HiTECH requirements and that takes a data-centric approach to security can ensure the privacy and integrity of PHI without a substantial investment in internal resources.

The cloud solution should also leverage detailed metadata in order to automate and track data security. In this model, the metadata includes the data type and provenance, as well as full policy and governance rules, indexing, rights management, retention periods, and privacy agreements. Using the metadata to enforce access rights, the data warehouse solution can redact, de-identify, or even conceal data based on pre-defined user authorizations. Data use should also be trackable and auditable so that the organization can determine who accessed the data, when they did so, and what their authorizations were.

Attribute-Based Access Control

Attribute-based access control is an access paradigm that grants rights to users via the use of policies that combine various attributes together.

Metadata and attributes can be used to create rules and policy management that provide the basis for granular, attribute-driven, policy-based control over who can see and do what. This ensures that access control is implemented automatically and uniformly.

In a Zero Trust Data strategy, users are decoupled from data. This decoupling requires a strategy for describing data (metadata) that is extensible and modifiable as policies evolve. It also requires a way to describe users (people and applications). These attributes can be delivered in sets: this user, who has this role, accessing from this location, using this device, etc.

The solution should also provide the ability to de-identify PHI automatically and as efficiently and securely as possible. Look for systems that follow the HITRUST De-Identification Framework, which provides a managed methodology for de-identification of data and sharing of compliance and risk information among entities and stakeholders. This is critical for research teams who will rely on access to data sets that have been stripped of patient identifiers. Automated, attribute-based access makes it possible for teams to access that data much faster. Organizations can even securely provide data access to third parties.

The data warehouse and cloud infrastructure provider should be willing to sign a strong business associate agreement (BAA) that assumes compliance risk and outlines specific consequences for violation of the terms. This creates a single, compliant point of service.

Controlled, auditable access also makes it easier to minimize human error in the security process – by far, the largest source of HIPAA violations. According to the Department of Health and Human Services, the majority

of HIPAA violations have been due to theft or loss of paper documents, hard drives, or mobile devices/laptops. Wrapping security around the data and keeping it in the cloud (as opposed to, say, on a laptop) mitigates those risks.

Medical and support staff sometimes consider security policies a hindrance to doing their jobs and efficiently sharing information. A cloud-based data warehouse provides an opportunity to encrypt and secure data centrally, while expanding access across usage scenarios and devices. According to the Forrester report “Healthcare Technology Innovation Continues to Outpace Security,” this data-centric strategy “is especially critical as your firm increasingly shares patient data with other hospitals, information exchanges, and patients themselves. To mitigate this risk without inhibiting use, inventory and classify your sensitive data, document how both medical staff and data scientists are using the data, then implement the appropriate controls (such as cloud encryption, careful scrutiny of HIPAA partner agreements, and privacy policy enforcement).”

Cloud Benefits

Once you have established that your data can be securely housed and accessed in the cloud, you also have to build a business case for investing in this type of Big Data warehouse solution. Compared to an on-premise solution, the cloud offers two valuable and quantifiable benefits: a reduction in infrastructure capital costs, and a much faster time to insight.

Cost

The cost benefits of the cloud are driven by two key benefit areas. First, transitioning to the cloud shifts the data warehouse from a capital expense (CapEx) model to an operating expense (OpEx) model. Second, the costs for maintaining and upgrading the infrastructure fall largely on the cloud services providers.

CapEx IT investments (in servers, software, and peripherals) require ongoing maintenance and updates to meet changing needs. The amount of data you collect and manage is only going to expand, and storage capacity (and compute power) has to expand with it. OpEx items are easier to integrate into a budget.

The cloud also eliminates ongoing data center, server, and OS maintenance and support. IT resources can be redeployed to manage projects that improve clinical outcomes or analytics capabilities instead of tending to servers.

Flexibility

Cloud solutions also provide a more flexible expense model. You only pay for what you use, so there is no need to invest in capacity for future growth or to run workloads that require short bursts of high compute resources and storage. You can start small and scale up (or down) as demand and storage needs change.

A flexible solution also enables burstability – so researchers can tap into significantly larger amounts of capacity on short notice, and respond quickly to emerging clinical needs.

Get Started Quickly

Even large, advanced organizations with well-established research infrastructures may feel as if they are falling behind. That's because data gathering capacity has increased much faster than data analytics capabilities at many institutions. A cloud-based approach allows these teams to quickly launch projects and incorporate clinical, genomics, geospatial, and other data sources in their analysis. These platforms also make it easier to centrally disseminate findings to multiple departments or other organizations so that managers or clinicians can act upon that data.

The pay-on-demand model makes it easier for even smaller organizations to access massive compute power as they need it. For providers that may only need those capabilities for brief periods of time or for specific projects, this is a much more affordable approach, and further enables the ability to scale in the future.

Proof

You can also more rapidly demonstrate real results to executive managers. By demonstrating short-term success using affordable and effective resources, you will be in a better position to expand and grow your analytics operation. This type of quick, demonstrable value can also help get buy-in from clinical staff and other team members because they can see the benefits of collecting and analyzing the data.

Cloud-Based Big Data Warehouse Checklist

Partnering with a qualified, cloud-based data warehouse provider can help you improve your research capabilities, and strengthen population health, precision medicine, and other initiatives. Not all platforms are created equal. Below are some key considerations for selecting a provider:



Designed for Healthcare Data

Find a solution that is designed with healthcare data in mind, and includes libraries that can handle the importation of different types of healthcare data, as well as the ability to process and transform that data into usable intelligence. That should include the ability to use metadata to describe the various types of healthcare files and associated data. That makes it easier to search through the data without having to manually search through files for specific pieces of information.



Attribute-Based Control Features

The platform should provide strong attribute-based control features. You should be able to provide different levels of access to the same data based on user-defined parameters. The solution should also provide an auditable record of who accessed the data, when they did so, and what they did with the information.



Robust De-Identification

De-Identification should be robust and reliable. The system should follow the HITRUST De-Identification Framework or other accepted standard such as the HIPAA Safe Harbour Guidelines and HIPAA Expert Determination Method.



Advanced Catalog Functionality

The solution should offer advanced cataloging functionality. Data isn't going to do anyone any good if they can't find it. The cloud solution should provide advanced cataloging and indexing to enable rapid access to data and fast creation of data sets. Using manual methods, this can take weeks or months. With a centralized view of available data, researchers can quickly determine if there is enough data available for a given project (for example, certain types of patients with a particular disease) to conduct the research they want without actually having to request the data.



Ability to Handle Complex Data Types

The solution should be able to handle complex data types. Healthcare data frequently comes in unstructured or semi-structured forms, including large genomics files, PDFs, Excel spreadsheets, XML documents, images, and data generated by EMR systems. An advanced, cloud-based Big Data solution can extract information from those documents, and then catalog and index that information quickly. This removes a significant barrier to end-user adoption of the analytics tools.



Plug-and-Play Solution

Plug-and-play out of the box solution will accelerate the time-to-benefit. Research and clinical teams don't want to wait any longer to apply Big Data analytics within the organization. Using a cloud-based tool specifically designed for healthcare applications allows end users to immediately begin using the tool without having to gain any Big Data expertise, or developing proprietary privacy or cataloging features. This also allows users to continue with their favorite analytics and visual tools and solutions like ODBC, R, Matlab/SAS/SPSS, Genospace, 5AM Solutions or other home grown applications.



Should Support Your Existing Environment

The data warehouse should support your existing analytics environment. A Big Data warehouse addresses the storage, management, and access to your data. It should support your existing ecosystem of analytics and data science tools. This allows end users to access much larger and more organized data sets, while doing so using tools they are already familiar and comfortable with.



Must Be Scalable

The solution should scale. Data collection and analysis requirements are only going to expand for healthcare organizations in the future. The data warehouse should scale with those demands. Cloud-based Big Data solutions are uniquely positioned to provide that scalability more affordably and flexibly than an on-premise solution.

Both researchers and clinicians are anxious to expand their abilities to analyze healthcare data and share their findings quickly to have a greater impact on clinical operations and quality of care. A Big Data warehouse can provide rapid and secure access to valuable data sets. By leveraging a cloud-based Big Data warehouse, healthcare organizations can quickly and affordably launch research, precision medicine, quality improvement and other initiatives without compromising patient privacy.

About PHEMI

PHEMI delivers privacy, security, governance and enterprise-grade management for big data. Its flagship product, PHEMI Central, applies the Zero Trust Data approach, which embeds and enforces consent, data sharing agreements and privacy policies at the data level, removing a critical roadblock standing in the way of enterprises that aim to become data-driven.

PHEMI Central lets organizations easily access and mine any variety of data at any volume to drive insights that lower costs, improve outcomes, and allow better decisions faster. PHEMI brings both privacy and performance to big data. For more information visit <https://phemi.com/> and follow us on Twitter.

About ClearDATA

The Premier Healthcare Managed Cloud Company

ClearDATA is the nation's only healthcare-exclusive cloud computing company. As the premier healthcare managed cloud company, our solutions are positioned to solve the three fundamental challenges facing HIT:

- Modernize the Infrastructure
- Secure & Protect Patient Data
- Improve Data Interoperability

Our intimate knowledge of healthcare data workflow, security and compliance is a key differentiator. Our "just right" solutions for enterprise healthcare as well as individual private practice provide flexible options that can fit any size healthcare organization.



About Us

ClearDATA is the nation's fastest growing healthcare cloud computing company. More than 310,000 healthcare professionals rely on ClearDATA's HIPAA compliant cloud computing HealthDATA platform and infrastructure to store, manage, protect and share their patient data and critical applications.

For more information

1600 W. Broadway Road, Tempe AZ



(800) 804-6052



www.cleardata.com

