IMPROVING CLINICAL OUTCOMES:
UNLEASHING THE POTENTIAL OF HEALTHCARE DATA LAKE

A Thought Paper by Scalable Health

SCALABLE HEALTH

TABLE OF CONTENT

EXECUTIVE SUMMARY

Data is growing in volume, velocity and variety. The density of healthcare delivery system can sometimes, contribute complications in accessing right care at right time. Though navigating through health system can be confusing, caregivers need to make healthcare more accessible with less appropriate settings. In response, healthcare systems need to adopt data storage facility to fully customize and integrate the huge amount of data to be used for their specific requirements. Healthcare and life sciences organizations are increasingly, turning towards data lake platforms to get more actionable insights and flexibility to drive scalability and innovation to leverage the already overflowing unstructured and semi-structured medical data. Moreover, if data is not managed efficiently through data governance, data integration and data security method, data lakes can quickly become land of unusable data with higher risk of data disorder and non-compliance among healthcare organizations. Therefore, it is necessary for organizations to use a health data lake where they can holistically identify, integrate, cleanse, and prepare structured data while collectively governing and managing all the data assets.

Built on an intelligent data analytics platform, healthcare data lake can easily support unstructured, semi- structured, and big data. It also helps clinicians to use advanced analytics for a centralized access to data, which were previously inaccessible from EHR/EMR applications. Using machine learning assisted automation, they can also get faster insights from medical device sensors, EMR/EHR applications, social media, doctor's notes, and HL7 and claims data through self-service data preparation, optimized data processing and rapid intelligent data discovery.

The data lake offers the capability to maintain the fluid data requirements of current healthcare organizations as they try to analyze huge volumes of clinical data in real-time from different data sources in a variety of formats.

Healthcare organizations like data lake concept as it helps clinicians to carry out research & development, in-depth analysis of patient outcomes, clinical trials, fraud, and waste management. In data lake, predictive and prescriptive analytics can be used to support healthcare use cases and initiatives. With unparalleled levels of patient data figures at their fingertips, thanks to the data lake, caregivers can make informed, and data-driven decisions at the point-of-care.

INTRODUCTION

The advancements of digital transformation have arrived at such a time when healthcare organizations are working their way around to improve the efficiency of their electronic health record (EHR). Thus, healthcare organizations need to discover new analytical model to detect at-risk patients, reduce adverse events, and use evidence-based medicine.

Huge amount of important unstructured data remains yet to be explored, all due to difficulty in accessing unstructured data. This makes it tough to migrate towards fast evolving expectations for preventive care and speedy diagnosis and treatment. With the huge amount of medical data available in the majority of healthcare organizations, such as payers, providers, pharmaceuticals and third-party vendors; there is exclusive prospects to harness these sources of information for analytical insights. Using these insights, they can improve quality of care, reduce costs, and prevent resource waste. However, the complexity of emerging data sources somehow presents new challenges in attaining strategic goals, and within the subsequent applications in the clinical environment.

This is where Data Lake comes into the picture. Instead of trying to pull all these data from disparate sources and integrating it manually, data is put directly into a single data lake which is capable of getting through multiple systems. Here, the data is greatly organized, and any type of external analysis technique can be used to integrate it more efficiently and gain actionable insights from information for providers and patients. The main trait of this approach is the security levels, which is individually maintained as data passes through each separate database. This is done to ensure that patient data is encrypted and healthcare organizations adhere to strict privacy and compliance laws. All the patient data are handled with complete control and right data is shared with the right person. Additionally, old and inactive data is routinely archived, thereby reducing higher medical costs, minimizing possible problems, and inefficiencies of resource overrun.

As the emerging data sources are in rise, healthcare organizations need a smarter way to collect, manage and harness the data that can add value to healthcare. Organizations that leverage the healthcare data lake will be entrusted to reduce costs, maximize patient data access, and ultimately get more benefit from data. In the end, this will lead to higher satisfaction among patients and providers and will bring effective outcomes in patient health.

THE HEALTHCARE TRANSITION AND THE NEED FOR DATA LAKE

As the healthcare industry is transforming from a volume-based model to value-based model, it is unsurprising to see the need of more meaningful data is increasing rapidly from all corners of the industry.

Today, the information are stored in traditional EDWs. And, the need to use relevant patterns and trends in data had escalated an upsurge in using "Data Lake" approach among healthcare organizations to integrate that data in their work. The data lake platform manages raw data in original form, thus, creating an information pool that can be dipped into and queried in an ad hoc manner at any time period. This action helps healthcare organizations to gain insights from data and can assists them in making quick decisions. As new processes gets implemented, the effect of each intervention can be rapidly calculated.

Some of the key benefits of data lake approach are:

- Comprehensive view of patient care: Using data lake approach, healthcare organizations can collect and standardize a wide range of data, such as claims and Rx data, clinical information, health survey, administrative data, patient registries, data from EHRs, and EMRs,. All these data can be combined together to create a comprehensive view of patient, assisting in a variety of use cases comprising better outcomes, cost reduction programs, medical decision-making, and quality improvement initiatives, etc.

- Processing huge data volumes at once: A data lake is a merger of huge amount of data. It can process different varieties of relevant medical data, ranging from unstructured, to

semi-structured, and/or structured while delivering the agility to configure underlying schema- a feature favored by data lake. Moreover, the raw data can be stored in original format and is never lost.

- Enhanced query processing: As data governance is used in the data lake, it reduces the effort of clinical researchers to know about the amount of data ingested. This leads to higher patient information efficiency, better concurrency, enhanced query processing, and produces quick results. This helps researchers to take rapid decisions and helps to speed up the care delivery process.

- Cost effective: As Hadoop is used in data lake to store huge data, it performs in an efficient manner. Data lake minimizes the costs while maximizing the return on data investment. Using a data lake platform puts minimum effort on clinical researchers for integrating the data, thus reducing the cost.

- Faster time to insight: Earlier, clinicians used to store information on patients and various test on a manual form. This decreased the data accessibility and visibility and created a gap in the care delivery network. Now, by using optimized data processing, machine learning assisted automation, self-service data preparation and intelligent data discovery, a culture can be created where the data collected is valued, where insights is generated from the collected data piece and is used during appropriate care interventions.

- Improving and handling treatments in real-time: Creating an empowered network between patients, care providers and specialist by using data integration method will offer a complete analysis of patient's claims and clinical data and help in identifying early at-risk patients. This will help in providing right care to right patient at right time.

- Faster time-to-value: Using pre-built parsers and connectors, clinicians can accumulate, process, and deliver a range of patient related data from medical device sensors, EMR/EHR applications, images, doctor's notes, and claims data. With data lake, they can get insights from large data sets in weeks rather than years.

- Create care management model: Technology, care coordination, analytics and patient engagement forms the structure of an effective care management model. Using integrated data lake platform, healthcare providers can make decisions which will lead to reduced in-hospital complications and preventable readmissions, promote use of personalized medicine, detect genetic markers, develop clinical trial safety, and much more.

- Get centralized data access: Medical data is collected from various set of public and private data collection systems, such as administrative enrollment and billing records, medical records, and health surveys, used by various entities (CHCs, physicians, hospitals, and health plans). Data collected on ethnicity, race, and language are collected, to some extent, is used by all these entities, which reflects the potential of each to contribute information on patients or enrollees. Thus, a data lake helps in providing a centralized data access to all the entities involved in the data sharing process, thus reducing the redundancies.

- Reduction of information silos: A data lake offers a potent data architecture with an integrated location to help reducing information silos across the healthcare system. This helps to collect data from large number of trusted sources including genomic research centers, payers, public health databases, bio banks, and social media feeds.

- Easy analysis of medical data: The data lake lets clinicians, and data science teams to effectively analyze cross data and incorporate from trusted external and internal data sources for mining and analysis. With such insights on future-healthcare, providers can advance accountable care initiatives, which creates a new realm of data science for detecting patterns, trends, correlations, and discoveries that can have great impact on integrated patient care.

- In-time patient monitoring: Healthcare organizations are looking to provide more integrated care to their patients by regularly monitoring patient vital signs. The data generated from various monitors can be examined in real-time and alerts can be sent to care providers so they know directly about the changes occurring in a patient's condition. With machine learning algorithms, physicians' can process real-time events to get insights that will help them in making a lifesaving decision and allow for effective interventions.

NAVIGATE A HEALTHCARE DATA LAKE

With the shift towards clinical, financial, and operational advancements using advanced data analytics approach, healthcare organizations first of all need to develop a long term roadmap and strategy with a multidisciplinary team to prioritize tasks.

Designing a data lake that meets both healthcare and technology goal is important. With requirement to process huge amount of data and the need to support wide range of interfaces, data structures and processing methods, sometimes, data lakes can turn into a messy data swamp, and fail to provide a promised analytical value. Data lakes require data analytics experts to develop a logical or physical separation of data acquisition, insight development, optimization and

governance, and analytics consumption.

On the next move, healthcare organizations need to collect data scattered across data warehouses, data marts, operational systems, transactional systems, and external data sources. The data lake approach in healthcare industry can provide power to share data and support for rapid exploration and discovery processes. Data science team can use these tools to discover variables and metrics that better predict clinical performance and support decision making. By align clinical requirements, it can also enable predictive and prescriptive analytics necessary to support healthcare use cases and initiatives to get meaningful insights for better outcomes.

Navigate a Healthcare Data Lake

Develop Big Data Analytics Vision and Roadmap

Collect Data from Internal and Trusted External Data Sources

Align Clinical Requirements to Optimize Data Lake Infrastructure

1

2

3

4

5

Work with Multidisciplinary Team to Prioritize Cases

Execute Proof-of-value Project

HEALTHCARE DATA LAKE USE CASES FOR PAYERS

1

Improving Outcomes by Supporting Health Initiatives

With the latest health initiatives, healthcare organizations are developing and deploying mobile applications that help patients manage their care, locate providers, and improve their health. By collecting data from these mobile interactions along with the data from wearables and analyzing the resulting data, payers can monitor adherence to drug and treatment regimens and detect trends that will lead to individual and population wellness benefits.

2

360^0 View of Member

Using healthcare data lake platform offers healthcare payer organizations deliver a highly personalized experience by offering a complete, single and consistent view of member across all touch points, while meeting tactical cost of care reduction goal and strategic health wellness goal. This will bring a greater opportunity for payers to make decisions, improve profitability for both the provider and the health plan, reduce preventable and inefficient processes and procedures.

3

Fraud Detection

Using integrated data mining and statistical techniques in massive data sets from internal and external, structured and unstructured sources in real-time; payers can facilitate more accurate detection, prevention of fraud and management across the functional areas of healthcare industry. Data lake allows healthcare payers to extract, analyze, interpret and transform the area of proactive fraud prevention

4

Billing Opportunities in Unstructured Text

The analysis of huge amount of unstructured data can offer massive opportunity for a complete billing practices. Payers in order to get quick meaningful insights on unbilled texts can use this information from different sources such as transcripts of physician's notes, labs, and related minor performed procedures. Based on data generated, payers can apply risk adjustments to segments of their patient population for compensating on per- patient basis.

HEALTHCARE DATA LAKE USE CASES FOR PROVIDERS

1 Creating Population Health Management Model

A population health management model includes care coordination, technology, data analytics, and patient engagement. An integrated data lake in healthcare industry helps in making providers better equipped with disparate data sources both structured and unstructured data, using which healthcare providers can make accurate decisions. This will lead to delivery of comprehensive care, reduce readmission rate, create better patient outcomes at an efficient cost.

2 Psychographic Prescriptive Modeling for Analysis

Using psychographic data from patient health records (PHR), one can get considerable information into additional disease risk factors. According to a study of over 1000 published medical papers it was found that heart disease is effected not only by measurable factors (such as blood pressure, and cholesterol) but also demographic (age, gender, and race) and psychographic factors (values, attitudes, and lifestyles) as well. Using data lake, healthcare provider can analyze data from multiple sources to arrive at a more accurate risk stratification.

3 Streaming from the Bedside for Care Optimization

Using information collected from bio-monitors, bedside sensors, and other IoT enabled devices, healthcare providers can find valuable patterns using which researchers, clinicians and physicians can bring changes in patient care delivery and assist in care coordination.

4 Implementing 'Precision Medicine'

With healthcare data lake to sort and analyze huge disparate data to gain valuable, quantitative information is key to improving the quality, delivery and efficiency to healthcare services. Precision medicine being a coveted areas of healthcare is rapidly developing. It helps to predict illnesses, and prescribe medicines based on individual's genetic.

HEALTHCARE DATA LAKE USE CASES FOR PHARMA

1 Finding more qualified candidates, faster

Big data analytics can help clinical trial recruiters to identify patients accurately, drawing insights from data stored in healthcare data lake. With such specified patient information, healthcare organizations can pre-establish nuanced enrollment criteria and set up automated screening processes, thus, improving the speed and efficiency of enrollment drives.

2 Improved Clinical Trials

Using pre-built parsers and connectors, healthcare organizations can collect, process and deliver patient related data from claims, doctor's note, Electronic Medical Record (EMR), digital images, Rx data and medical device sensors etc. This will help clinicians in getting actionable insights and make clinical decisions faster.

3 Real-Time Information

Healthcare data lake supports source-system agnostic data ingest, obviating cumbersome programming for each eClinical systems. Sponsors and CROs can get the benefits of accessing all clinical data regardless of its format, including Clinical Trial Management Systems, lab, imaging, and safety data in near real-time.

4 Better Drug Combinations

The design and analysis of clinical trials can bring better drug combinations with remarkable improvements on overall toxicity and survival. Using statistical models and data available from various sources, healthcare organizations can develop an optimization models that will choose treatment regimens which can be tested in clinical trial process, thus, improving its overall effectiveness.

DATA LAKE USE CASES FOR GLOBAL HEALTHCARE

1

Epidemic tracking

Combining data from clinic reports, media updates, transactional data from pharmacies, social media, helpline data, and evaluating them collectively can help to detect trends of disaster such as Ebola crisis. Integration of huge volume of data and analyzing it in its raw format as soon as it is produced across the globe. Combining such vast quantities of public healthcare information already available, healthcare data lake can help to ensure those working in hazardous conditions are able to stay on top of ever-changing situations.

2

Quality and safe delivery of vaccine

Quality and safe delivery of vaccine is the foremost priority of global health bodies. By using healthcare data lake, global health organizations can collect, process and report data to monitor and identify unusual and potentially dangerous trends, temperature fluctuation and sends alerts for corrective action in real time.
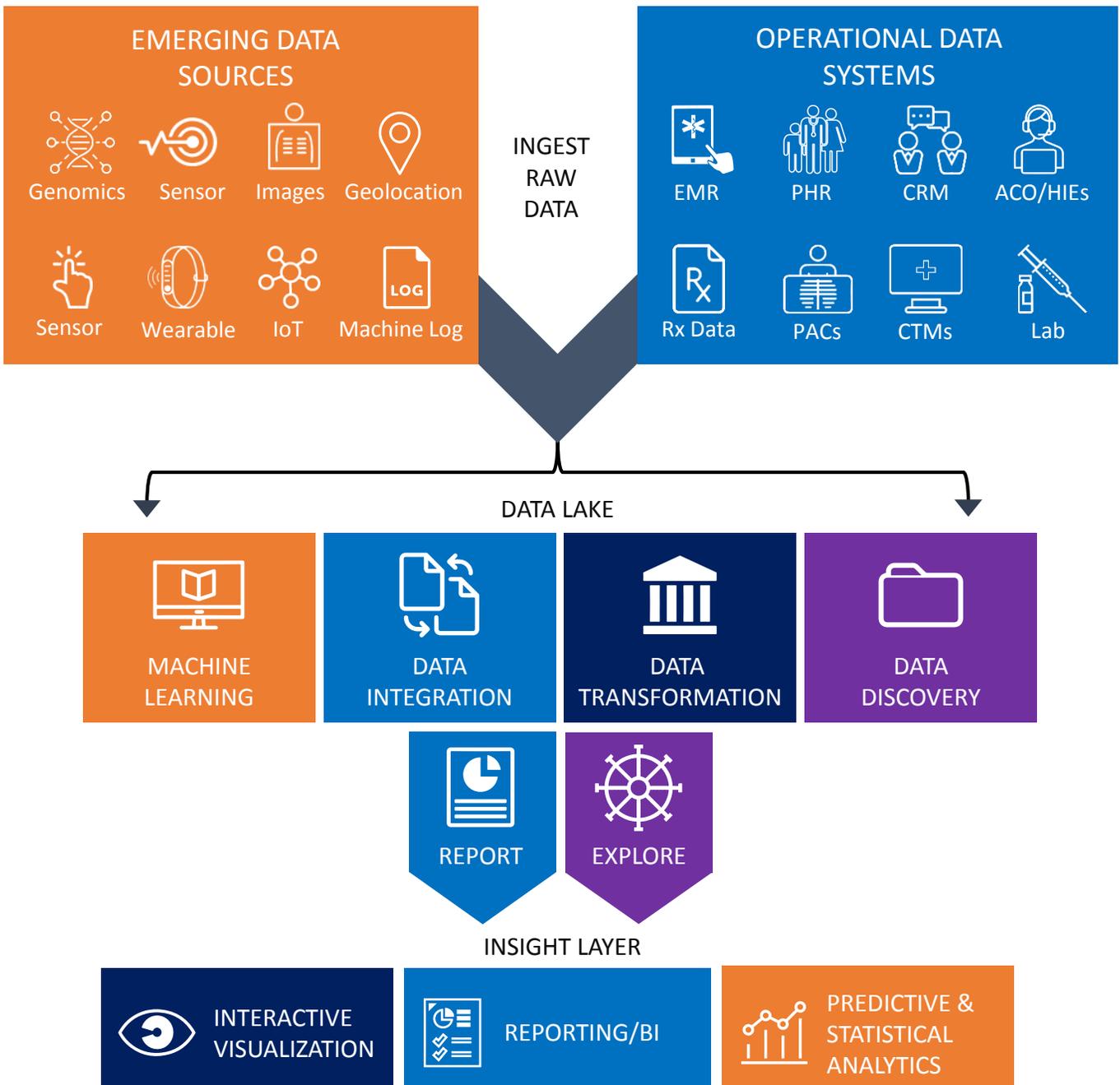
4

Health data for informed strategic planning

Big data can help to gain insights into care provider's motives. Organization can study check-up results of population in different demographic group and classify the factors that discourage people from taking up treatment. By using free public health, census, weather, social network data and Google Map to create heat maps for targeting multiple issues such as chronic diseases and population growth, healthcare administrator can match this data with the medical services to gain a better insight. This will help global health organizations to evaluate their healthcare delivery strategy and add more care units to the most problematic area.

3

Mapping high-risk areas

Using remotely-sensing and other geographic data about environmental, human and animal factors stored in data lake on top of it machine learning techniques can use high-resolution, global maps to identify possibility of epidemics rise in certain area and estimate number of people living in highest risk places. For example, such type of analysis can help to detect likely location for diseases virus to thrive, and even identify areas virus is expected to establish itself.

MODERN ARCHITECTURE OF DATA LAKE

The modern architecture provides healthcare organizations with complete control over the medical data collected from disparate data sources in unstructured and unorganized form. Using data exploration and integration method, these data are reconfigured into structured format, which enables clinicians to make better clinical decisions.

Having a data lake architecture can help to process query in real time while accommodating manual entry and imaging solutions. Using a modern data lake architecture can help clinicians to replace manual or batch review processes with rules and in-time workflow and eliminate or reduce duplication errors while avoiding costly clinical review and prior authorization processes.

EMERGING DATA SOURCES

Genomics Sensor Images Geolocation

Sensor Wearable IoT Machine Log

INGEST RAW DATA

OPERATIONAL DATA SYSTEMS

EMR PHR CRM ACO/HIEs

Rx Data PACs CTMs Lab

DATA LAKE

MACHINE LEARNING

DATA INTEGRATION

DATA TRANSFORMATION

DATA DISCOVERY

REPORT

EXPLORE

INSIGHT LAYER

INTERACTIVE VISUALIZATION

REPORTING/BI

PREDICTIVE & STATISTICAL ANALYTICS

Enabling Real-time Healthcare Insights with Data Lake

CONCLUSION

Driven by the enormous growth of current data and use of innovative technology to manage it, healthcare organizations are now consuming data and analytics based like never before. The rise of data driven decision-making is also real. And it is spectacular. Data lakes has become an excellent solution for extracting and putting all the relevant big and small data in a single repository.

But, this is just the tip of the ice berg. We are just scratching a thin layer of data lake application in healthcare industry. There is more to come as data lake gradually gets used for more use case including- addressing data accessibility and integration concern, creating scalable data lake structure, function of data lake and how a lake matures as it is used widely by the healthcare organizations.

However, there is more to gain for the healthcare industry by using a data lake platform to get valuable cost saving insights through massive available data. Massive data siloed problem is real and Healthcare Data Lake can help.

REFERENCES

1. http://tech.newstatesman.com/big-data/how-healthcare-big-data-lakes-aid-population-health-patient-care

2. http://searchhealthit.techtarget.com/feature/Semantic-graph-database-underpins-healthcare-data-lake

3. https://www.healthdatamanagement.com/opinion/how-healthcare-can-play-safe-in-the-data-lake

4. https://healthitanalytics.com/news/partners-data-lake-offers-healthcare-analytics-as-a-service

5. http://www.axisimagingnews.com/2015/08/data-lakes-will-healthcare-trailblazer/

6. https://insidebigdata.com/2016/11/16/case-studies-big-data-and-healthcare-life-sciences/

7. https://healthtechmagazine.net/article/2017/08/providers-tap-data-lakes-boost-patient-care

8. http://www.emrandhipaa.com/tag/healthcare-data-lake/

9. http://analyticsindiamag.com/biggest-big-data-trends-healthcare-2017/

10. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/using-agile-to-accelerate-your-data-transformation

11. http://blog.cambridgesemantics.com/diving-deeper-into-hl7-standards-with-the-smart-data-lake

12. https://tdwi.org/articles/2017/01/23/data-lake-management-innovations.aspx

About Scalable Health

Scalable Health specializes in providing next generation healthcare data, analytics and digital transformation solutions enabling healthcare organizations to optimize healthcare.

Scalable Health

Email: info@scalablehealth.com
Web: www.scalablehealth.com