



Augmented Real-time Clinical DataMart

Phani S Srinivasan Ponnappalli, Syneos Health
Subrahmanyam Rayaprolu, Syneos Health

Agenda

- Introduction
- Traditional Clinical Data warehouse vs Digital Data
- Modern Data warehousing architecture for Digital Data called “Data Lake”
- Bridge connection between Data Lake and Traditional Clinical Data by “Augmented Real-time Clinical DataMart”
- Apply Cognitive analytics on integrated information
- Conclusion

Introduction

Digital capabilities are disrupting multiple aspects of business and providing unparalleled opportunities across industries. In recent years, the Pharmaceutical industry has recognized this tremendous opportunity and started embracing digital information to optimize the time from “Lab to Life”.

The rapid consumer adoption of wearable devices for the collection of health data is laying the foundation for the next revolution in clinical trial operations.

Clinical trial models based on the integration of digital capabilities demonstrate compelling benefits, including:

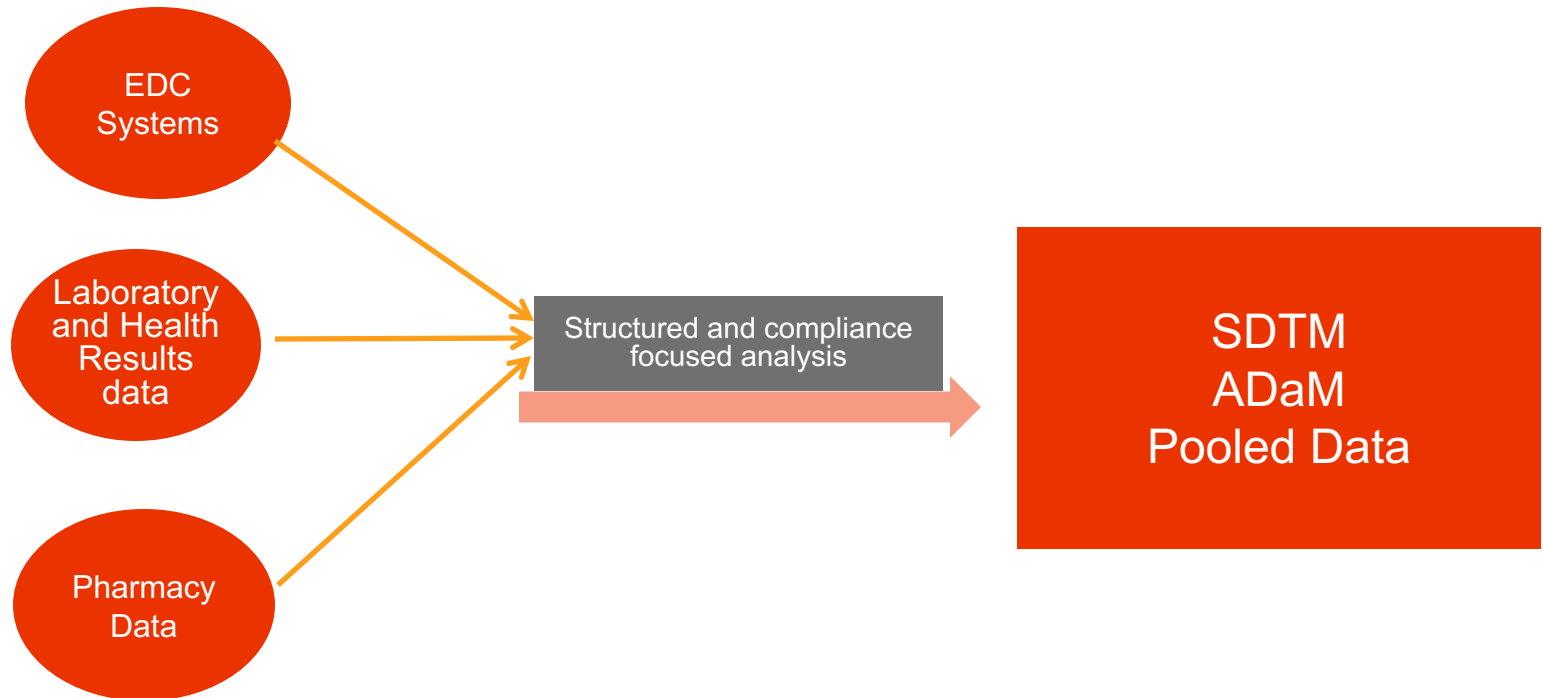
- Real-time patient monitoring and adherence
- Improved patient retention and compliance through convenient e-alerts and information sharing
- Cost reduction by reducing clinical visits
- Discovery of hidden patterns and actionable insights

Applying cognitive analytics on integrated Digital Data and Traditional Clinical Data will provide a new dimension of possibilities, including hidden patterns and actionable insights.

“Augmented Real-time Clinical DataMart” (ARCDM) is a term we are coining for this approach.

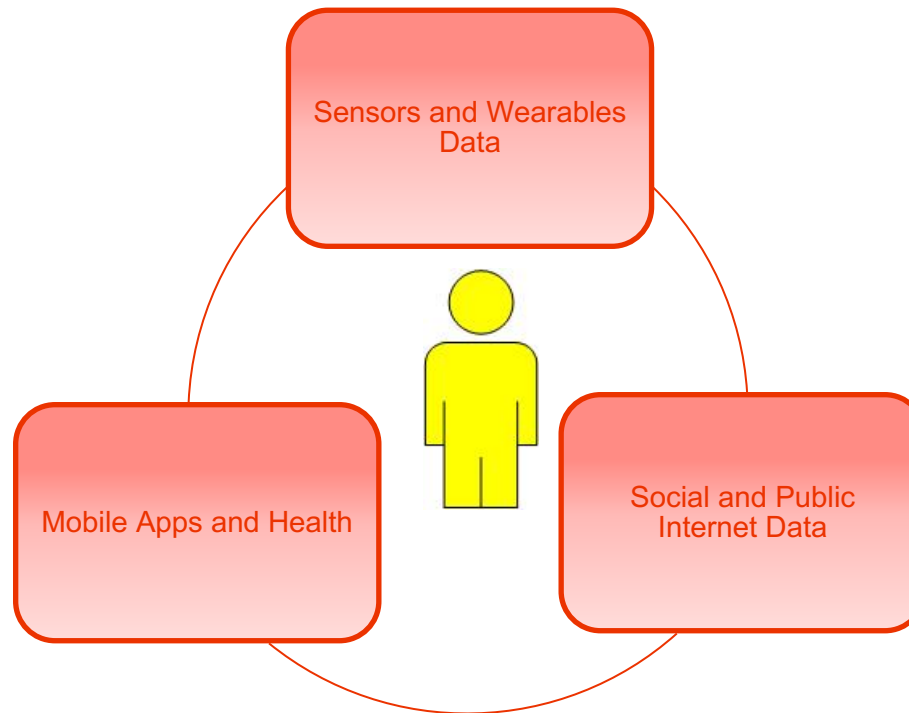
Traditional Clinical Data Warehouse

- Standards and Compliance focused.
- Relational Data Model.



Digital Data

- Patient and Physician Centric.
- Engagement and Adherence.
- Health Monitoring and Alerts



Traditional Clinical Data vs Digital Data

• HOW DIFFERENT IS DIGITAL DATA FROM TRADITIONAL DATA?

Digital / Big Data is characterized by the 3 Vs:

- Volume – number of transactions / complexity
- Velocity - number of transactions / time
- Variety – disparate data types and formats

Example:

In traditional clinical data there will be one record per visit which is taken at site. Medical Devices are available as patch to a patient's body that can monitor ECG every minute.

• CAN WE HANDLE DIGITAL BIG DATA WITH TRADITIONAL DATA MODEL?

Digital Data being mostly unstructured / semi-structured do not fit the enterprise relational data warehouse model. A new operational process has to be defined with the following focus areas to handle Big Data:

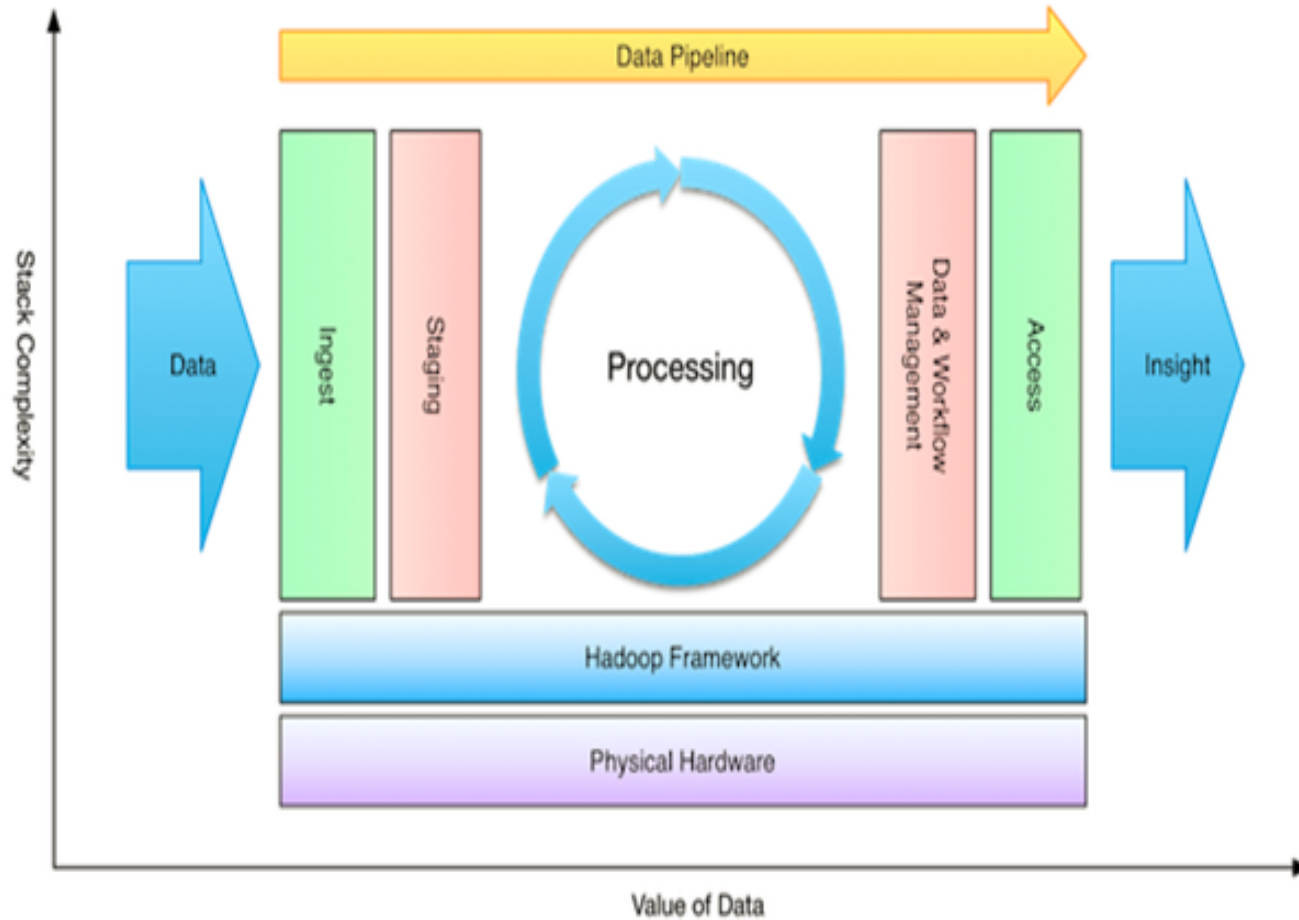
- Identify and use tools that easily manage unstructured data. Tools need to provide efficient text parsing and analytics, taxonomy and metadata management.
- Integrate unstructured data with existing information systems.
- Store and process data in volumes too large for a traditional database.

Big Data

Big Data architecture is designed to handle the ingestion, processing, and analysis of data that is too large or complex for traditional database systems. Big data solutions work on a fundamentally different principle to handle device data and streaming data. Big Data architecture has the following characteristics.

- **Distributed Data and parallel processing:** Big Data solutions store huge data in a distributed manner in a file system. Process the data in parallel on a cluster of nodes. In simple words, Data is broken in to small pieces of information and stored in multiple blocks called HDFS. Tasks are executed in parallel by processing these blocks in parallel and results are merged back, is called map reduce.
- **Fault Tolerance:** Big Data solutions work on failure to tolerance by redundancy. The same information is stored in multiple places called racks. There are multiple machines available as cluster for processing. If any machine in the cluster goes down still system works due to data redundancy and multiple machines.
- **Scalability:** Big Data systems are very flexible in scaling storage space and computing power on fly whenever required.
- **Cost effectiveness:** Big Data systems like Hadoop are open source and uses commodity hardware. They do not require a very high-end server with large memory and processing power. This makes the system very cost effective.

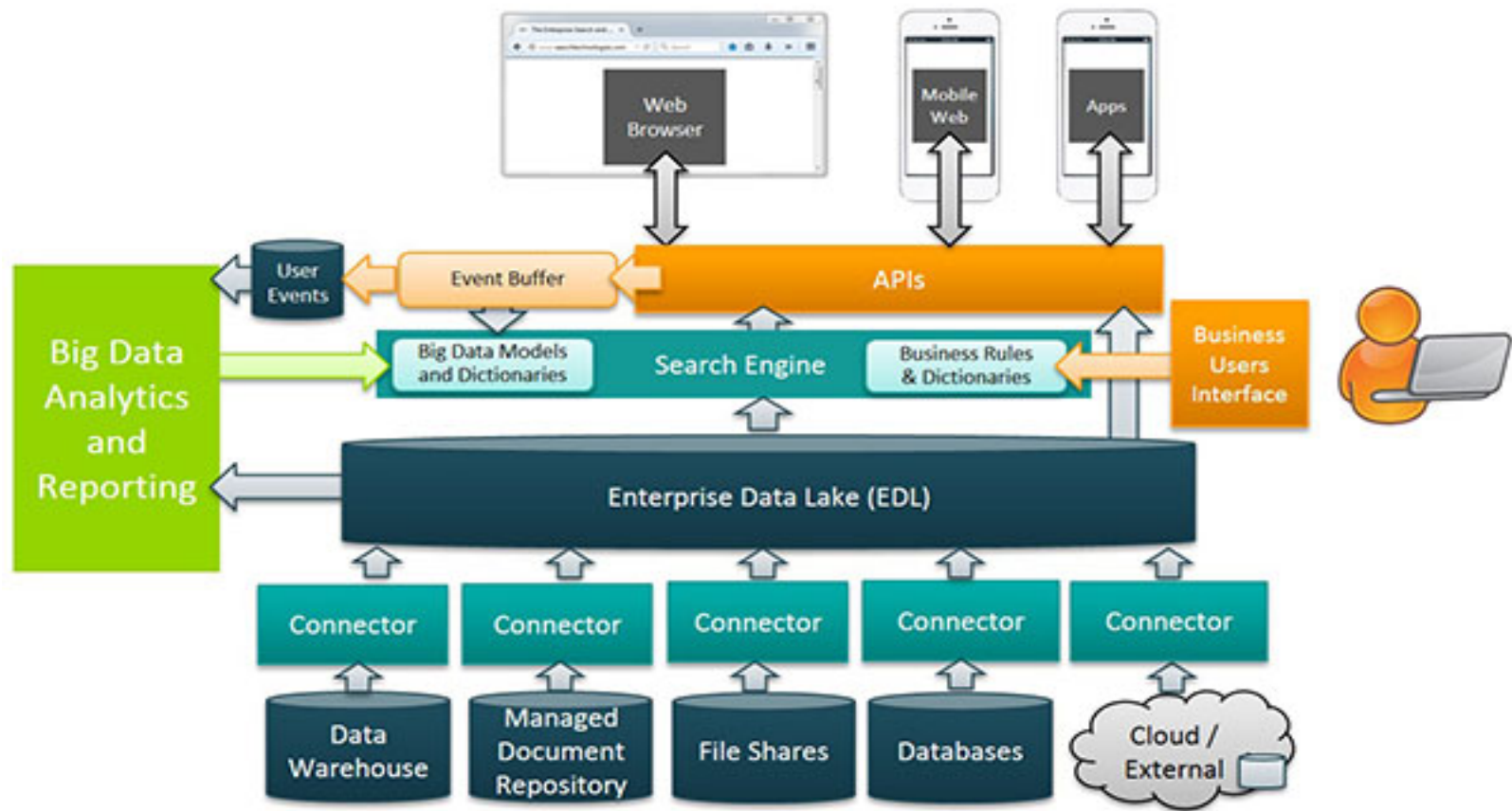
Big Data Architecture



Data Lake

- A data lake is a large storage repository that holds a vast amount of raw data in its native format until it is needed. An “enterprise data lake” (EDL) is simply a data lake for enterprise-wide information storage and sharing.
- The advent of Big Data has also given rise to the concept of a “Data Lake” which is a storage repository that can hold vast amounts of data in raw data format including structured, semi-structured and unstructured data. A data lake accommodates data of varying structures, which are able to resolve at the application delivery level, i.e. when the data structuring requirements are known.
- Data lakes are considered as a re-creation of data warehouses in the Big Data era. Data lakes are significantly different from warehouses, in terms of the ways they structure and manage data. The main difference is that data warehouses deal with structured data only, while data lakes store raw data and transform it to some structure when it is time to use the data (e.g. as part of an application). Also data lakes are closely affiliated to Big Data technologies as in most cases they leverage tools and techniques from the Hadoop/Big Data ecosystem. Finally, data lakes provide agility in terms of their processing. Data schemas can flexibly change, while data warehouses adhere to given schemas in order to benefit from well-structured data.

Data Lake Architecture



Bridge Connection between Data Lake and Traditional Clinical Data

- Cognitive analytics offer a way to bridge the gap between Big Data and the reality of practical decision making.
- Cognitive analytics applies human like intelligence to certain tasks, such as understanding not only the words in a text, but the full context of what is being written or spoken, or recognizing objects in an image within large amounts of information. It is a data forward approach that starts and ends with what's contained in information. This unique way of approaching the entirety of information (all types and at any scale) reveals connections, patterns and collocations that enable unprecedented, even unexpected insight.
- Applying this Cognitive Analytics framework, a hybrid information layer is created from the data lake that could be integrated with a traditional data warehouse for discovering hidden patterns and produce actionable insights, thus the “Augmented Real-time Clinical DataMart” (ARCDM).

ARCDM ,Three Step Approach

1. IDENTIFY THE QUESTIONS TO BE ANSWERED

Analytics is about asking—and answering—smarter questions.

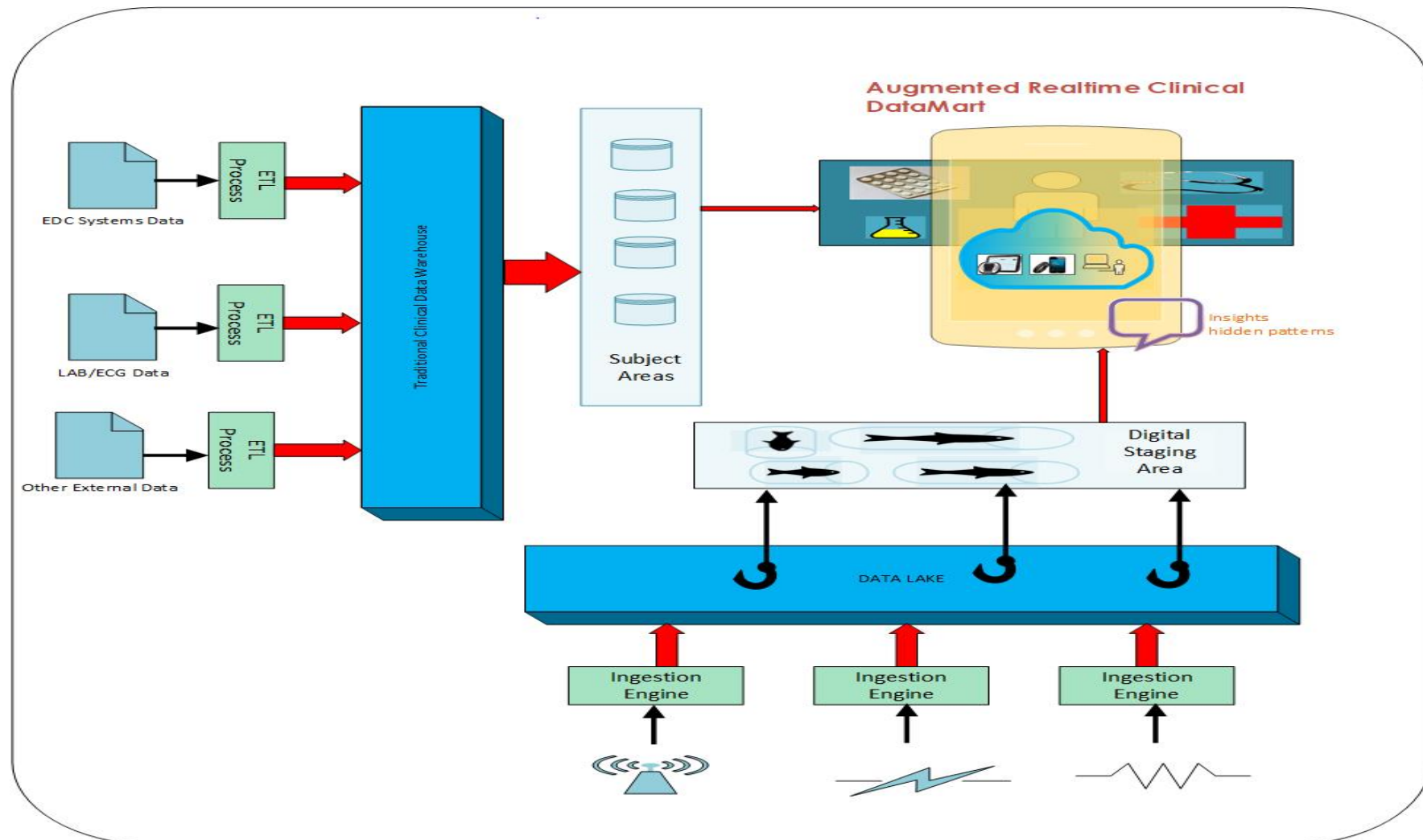
2. FISHING THE DATA LAKE

Data Lake and Big Data Ecosystems contain tons of data in raw format. We need to extract the relevant information with which we will be able to answer the questions which are not available in traditional data warehouses.

3. AUGMENTED REALTIME CLINICAL DATAMART

“Augmented Realtime Clinical DataMart” (ARCDM) is a term we are coining to define a data abstraction layer which combines the traditional clinical data and the data fish we extracted. Selectively move data that are deemed to have the highest business value in a complementary warehouse.

ARCDM Architecture



ARCDM , Features & Benefits

• FEATURES OF ARCDM

- Centralized data repositories streamline translational clinical research
- Provide relevant data to researchers for research
- Facilitate storage and maintenance of datasets
- Promote collaboration

• BENEFITS OF ARCDM

- Discovering hidden patterns
- Producing actionable insights
- Augmenting traditional data with real time digital data
- Answering the open questions

Conclusion

- The vast volume of digital data generated by health devices and sensors will be transformative across the entire health care spectrum, from wellness and prevention to treatment and research.
- During the next five years, Big Data and AI technologies will mature to enable advanced research models, including cloud-based health databases of continuously uploaded patient data and Internet-based trials conducted remotely.
- ARCDM is one of the possible strategies to combine traditional clinical data and this new form of enormous data.
- Cognitive analytics framework on digital data augmenting the traditional data will open doors for a new era of clinical research.

References

- <http://blog.cloudera.com/blog/2014/09/getting-started-with-big-data-architecture/>
- <https://www.searchtechnologies.com/blog/search-data-lake-with-big-data>
- <https://blogs.informatica.com/2016/03/30/fishing-for-big-data-insights-with-an-intelligent-data-lake-2/#fbid=4g3xvtj52zd>
- <http://www.itexchangeweb.com/blog/fishing-in-the-big-data-lake/>
- <https://itknowledgeexchange.techtarget.com/iot-agenda/gone-fishing-in-iot-data-lakes/>
- <http://support.sas.com/resources/papers/proceedings10/183-2010.pdf>
- <http://www.oracle.com/us/industries/life-sciences/guide-clinical-data-warehouse-ar-1563726.pdf>
- <https://www.cognizant.com/perspectives/information-architecture-trolling-in-the-data-lake-get-yourself-a-fish-finder>
- <https://www.analyticsvidhya.com/blog/2013/07/big-data/>

Thank you!

