An Architecture For Healthcare Big Data Management And Analysis

Presented By: **Lina Ben Salem**

Plan

I. Introduction

Introduction

- Computers, Internet, big data, artificial intelligence, numerous emerging technologies are changing the world and almost every industry.

- In the healthcare field, more and more people are eager to conveniently obtain more intelligent healthcare services without seeing a doctor. For instance, a chronic disease patient is allowed to rest at home and he would like to know his body state which can be monitored by wearable medical devices. Such devices or systems are able to monitor patient's vital signs and send them to a remote data center.

- Both historical and real time data can be utilized for further analysis and treatment suggestion. However, the amount of such data would be as large as we never saw before and we call it as healthcare big data with following "5V" features

II. Goal

Introduction

- Therefore our goal is to build an architecture for healthcare big data management and analysis to help gather and predict healthcare data such as monitoring blood pressure in a real-time manner and showing warning information by comprehensively analyzing both historical and current blood pressure collected from patients (this architecture can be used for any other medical data that is available in the system)

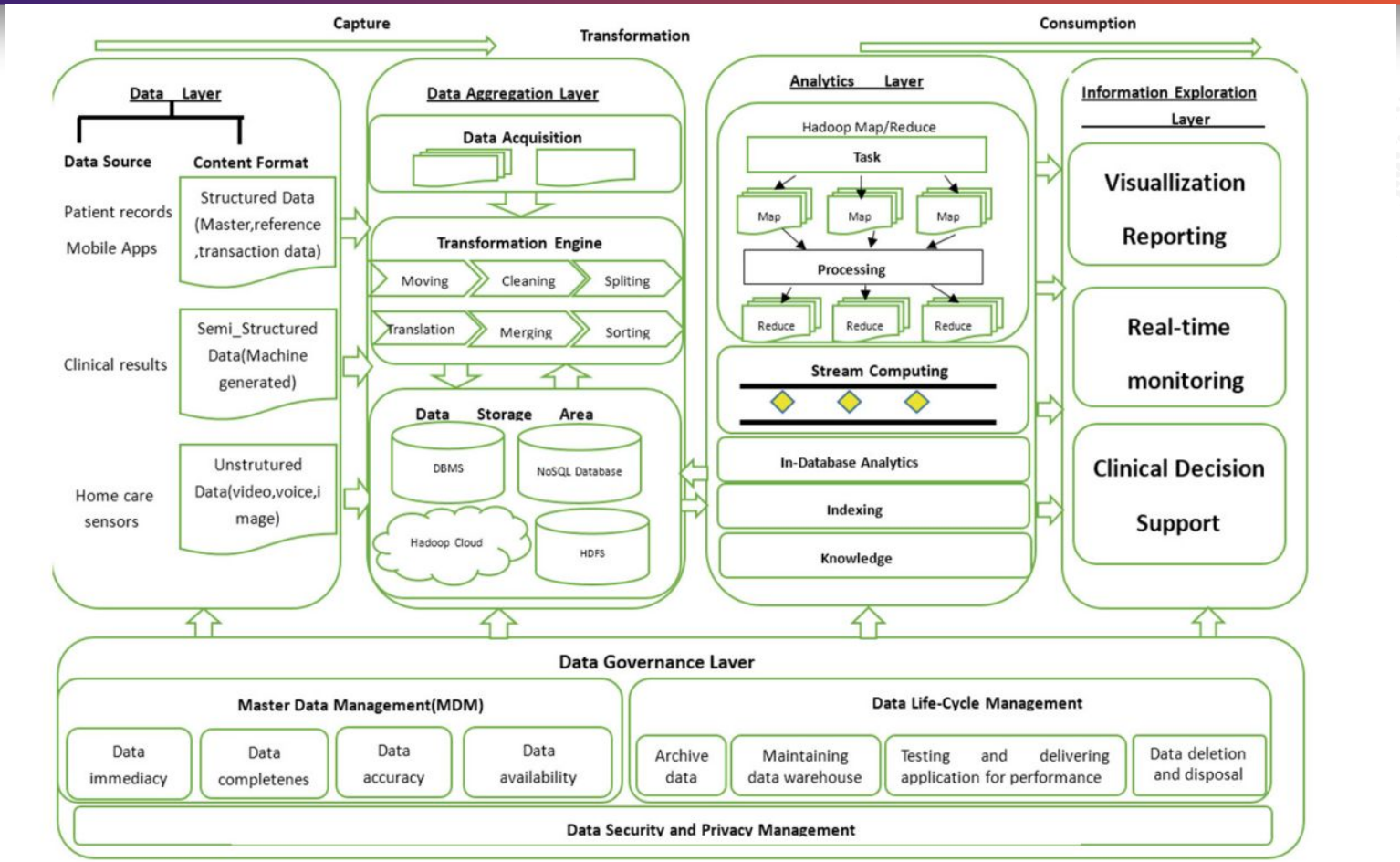
III. Architecture

Architecture

The architecture proposed to tackle healthcare big data consists of five layers including:

1. Data Layer
2. Data Aggregation Layer
3. Analytics Layer
4. Information Exploration Layer
5. Data Governance Layer

Architecture

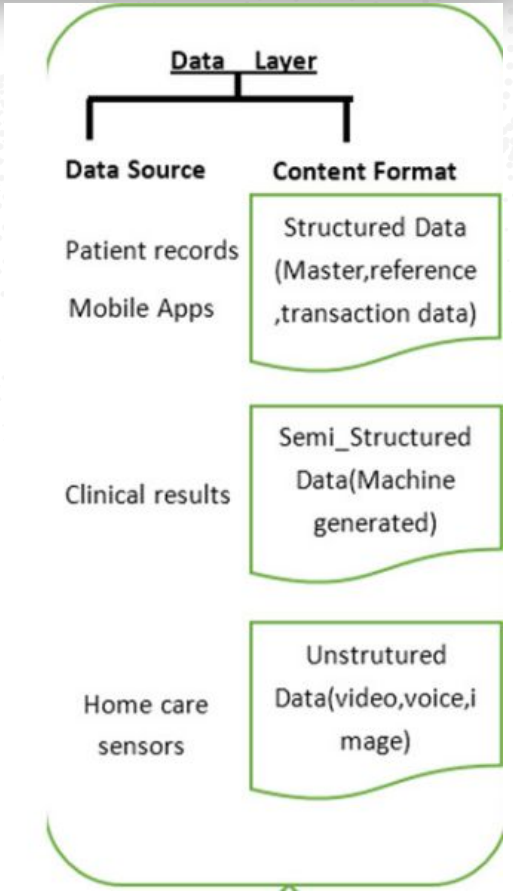


IV. The Layers

The Layers

1. Data Layer

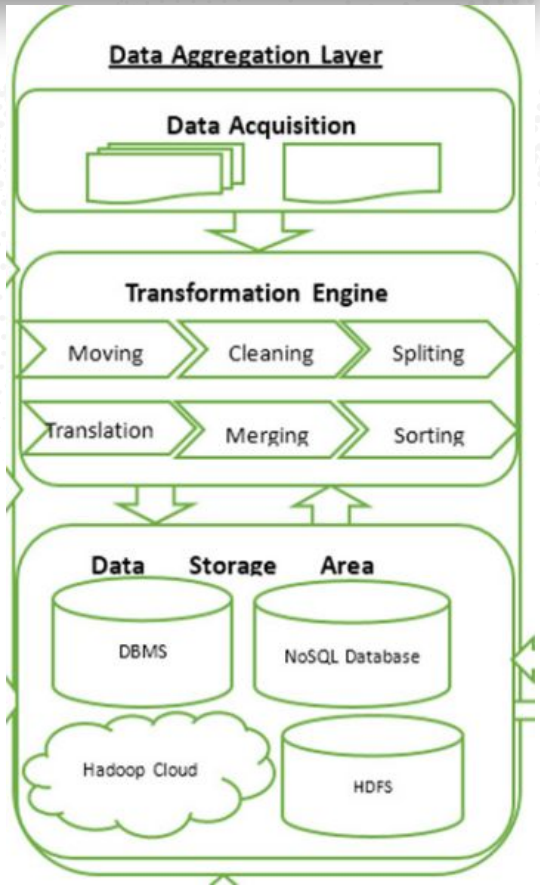
Healthcare data has multiple sources such as EHR (Electronic Health Record) and different types of medical devices.
At the same time, the healthcare data is collected with different formats such as structured data, semi-structured data and unstructured data, which lead to the challenges for data collection and preprocessing.
The Data Layer aims to provide services facilitating data collection and pre-processing for health records, back-end data of mobile healthcare Apps and streaming data generated by wearable medical sensors.



The Layers

2. Data Aggregation Layer

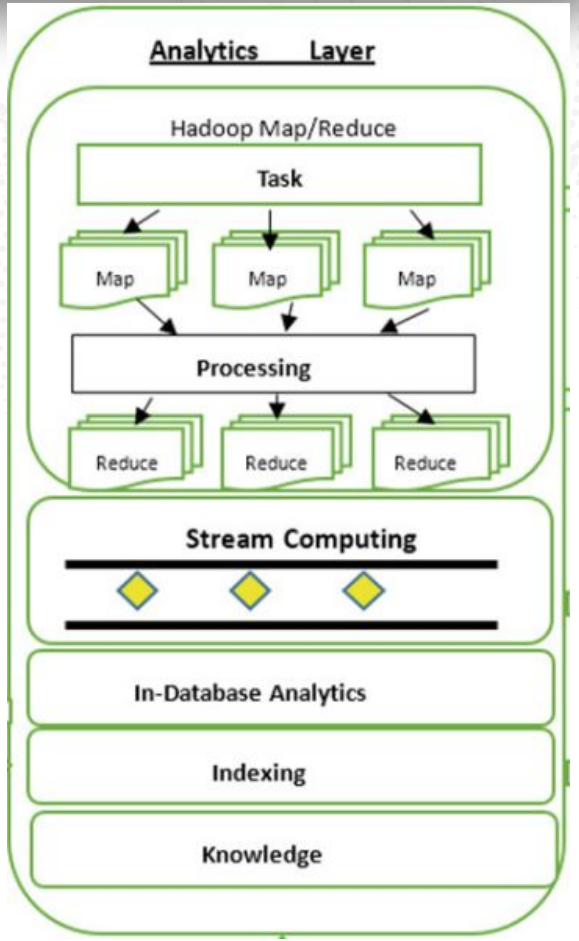
The main tasks of the Data Aggregation Layer include data extraction, transformation and loading into storage system. With the support from Data Layer, necessary operations including data moving, cleaning, splitting, translation, merging, and sorting can be performed. Afterward, the healthcare big data with standard format can be loaded into a storage system which may be relational databases, NO-SQL databases, distributed file systems and etc.



The Layers

3. Analytics Layer

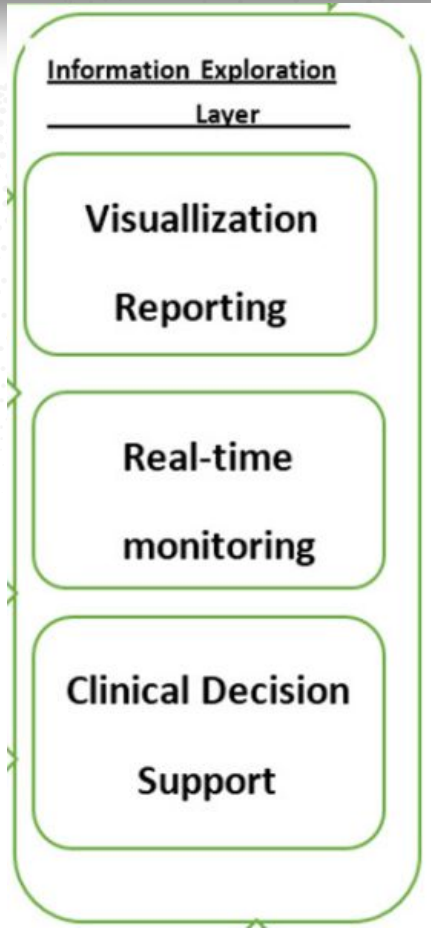
With the support from the Data Aggregation Layer, the Analytic Layer focuses on basic statistical analysis work.
Usually, work on this layer includes On-Line massive healthcare data analytical processing, streaming data processing, database construction & optimization, indexing, etc.



The Layers

4. Information Exploration Layer

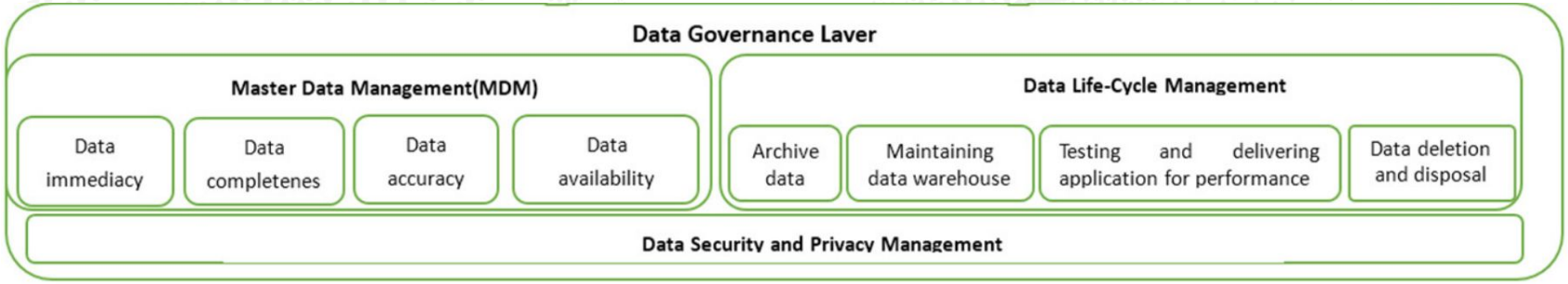
It consists of visualization/reporting, real-time monitoring and clinical decision support.
As we know, the healthcare big data could be massive and complex, which makes it difficult to understand and observe. Therefore, powerful techniques for efficiently visualizing and summarizing the healthcare big data become vital. And for patients, we care about analysis results on not only the historical data but also the current vital signs.
For this purpose, real-time monitoring based on transient vital signs of patients is needed. Besides, according to the further investigation on historical clinical data, it is feasible to provide better clinical decision support for doctors. So far, some artificial intelligent algorithms such as Byesian model, logistic regression model, decision tree, support vector machine, random forest and others can be integrated with domain knowledge for clinical decision purposes



The Layers

5. Data Governance Layer

Data Governance Layer, which is integrated with all the other four layers, is responsible for meta-data management, data life-cycle management and security/privacy management.



V. Next Step

V. Next Step

Under the guidance of the proposed architecture, a prototype system can be developed. The main function of the prototype system is to monitor blood pressure in a real-time manner and show warning information by comprehensively analyzing both historical and current blood pressure collected from patients.

The prototype system can be based on HBase, Hive, Spark MLLib and Spark Streaming to introduce personal health problem detection and real-time vital sign monitoring