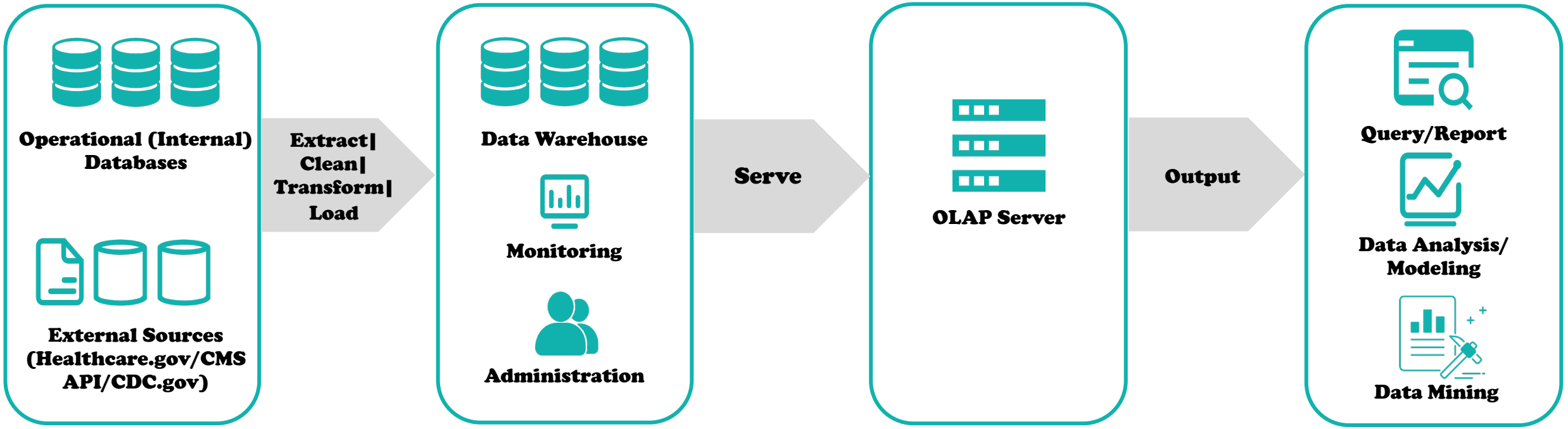
LifeBook Framework Proposal

LifeBook

Menghong & Jemma

September 2020

Big Picture


Operational (Internal) Databases
External Sources (Healthcare.gov/CMS API/CDC.gov)
Extract| Clean| Transform| Load
Data Warehouse
Monitoring
Administration
Serve
OLAP Server
Output
Query/Report
Data Analysis/ Modeling
Data Mining
Data
Bottom Tier
Middle Tier
Top Tier

Data & Manipulation

Operational Database (ODB): 

- Used to manage and store data in real time
- The source for a data warehouse
- Elements can be added and removed on the fly
- SQL/NoSQL based

External Sources: 

- Datasets from multiple sources
- API token is needed

Extraction: 

- The process of extracting data from various sources
- Fetch data from databases (SQL/NoSQL based)
- Read data from csv/xls/xml/json file

Transformation:

- Multistage Data Transformation
- In-Warehouse Data Transformation
- Data cleaning (Detection and removal of all major errors and inconsistencies etc.)

Loading:

- The transformed data is loaded into the target warehouse database
- Types of tables in data warehouse: fact tables and dimension tables

Bottom & Middle Tier

Data Warehouse:

1. Subject-oriented:

- Providing a simple and concise view around particular subjects by excluding data that are not useful in the decision support process

2. Integrated:

- Constructed by integrating multiple, heterogeneous data sources

3. Time-variant:

- The time horizon for data warehouse is significantly longer than that of operational systems

4. Non-volatile:

- A physically **separate** store of data transformed from the operational environment

OLAP (online analytical processing) Server:

- OLAP delivers warehouse applications such as performance reporting, marketing analysis etc. — applications that require historical, projected and derived data
- With OLAP servers' robust calculation engines, historical data is made vastly more useful by transforming it into derived and projected data
- Users could gain broader insights by combining standard access tools with a powerful analytic engine

Top Tier

Query/report:

- (Population level) Provide white paper analysis of health conditions of people in the States according to basic stats

Data analysis/modeling: (Technology/ Algorithm)

1. Risk Score Prediction using machine learning algorithms

- (Individual level) Predict risk scores of 10 chronic diseases for each patient based on both structured (lab values and demographics information) & unstructured data (patients' narration of his/her illness, the doctor's interrogation records and diagnosis)

2. Time series analysis of lab values

- Conduct time series analysis through checklists, discovering the trend and providing suggestions for further recommendation

Top Tier (cont.)

Platform Building: (Personalized service)

1. Checklist:

- Create checklists in the dashboard to record daily/monthly/yearly lab values and tests values

2. Patients Community:

- Segment patients according to different levels of their risk scores of certain disease; Daily checklist competition; Hold lectures for better health

3. Online Pharmacy:

- Sell medicines according to patient's risk scores/checklist/trend of lab values to enhance user experience and complete our platform's eco-system

4. Healthcare Provider/ Insurance Company Recommendation:

- Collect cooperative healthcare provider data, recommend providers/services to patients based on their health condition, provide online inquiry endorsement and options for further contact

5. Personalized Information Push:

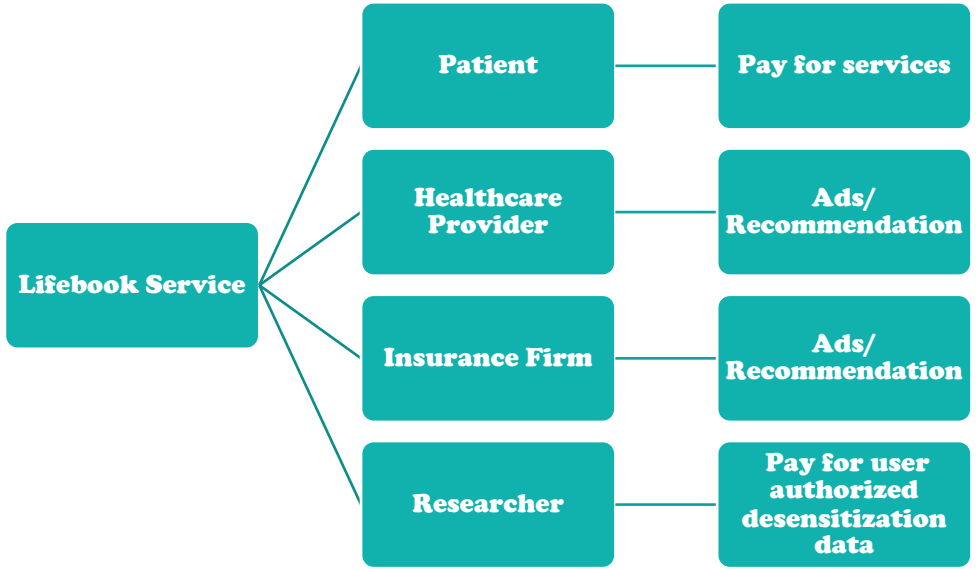
- Push personalized articles/news/tips/videos for patients

Profit Model - Ecosystem

The priority is that our platform can **help improve patients' care and reduce costs**

Services:

- **In the short term:** Free to patients and providers to gain more users
- **In the long term:** Four groups of entities are our targets



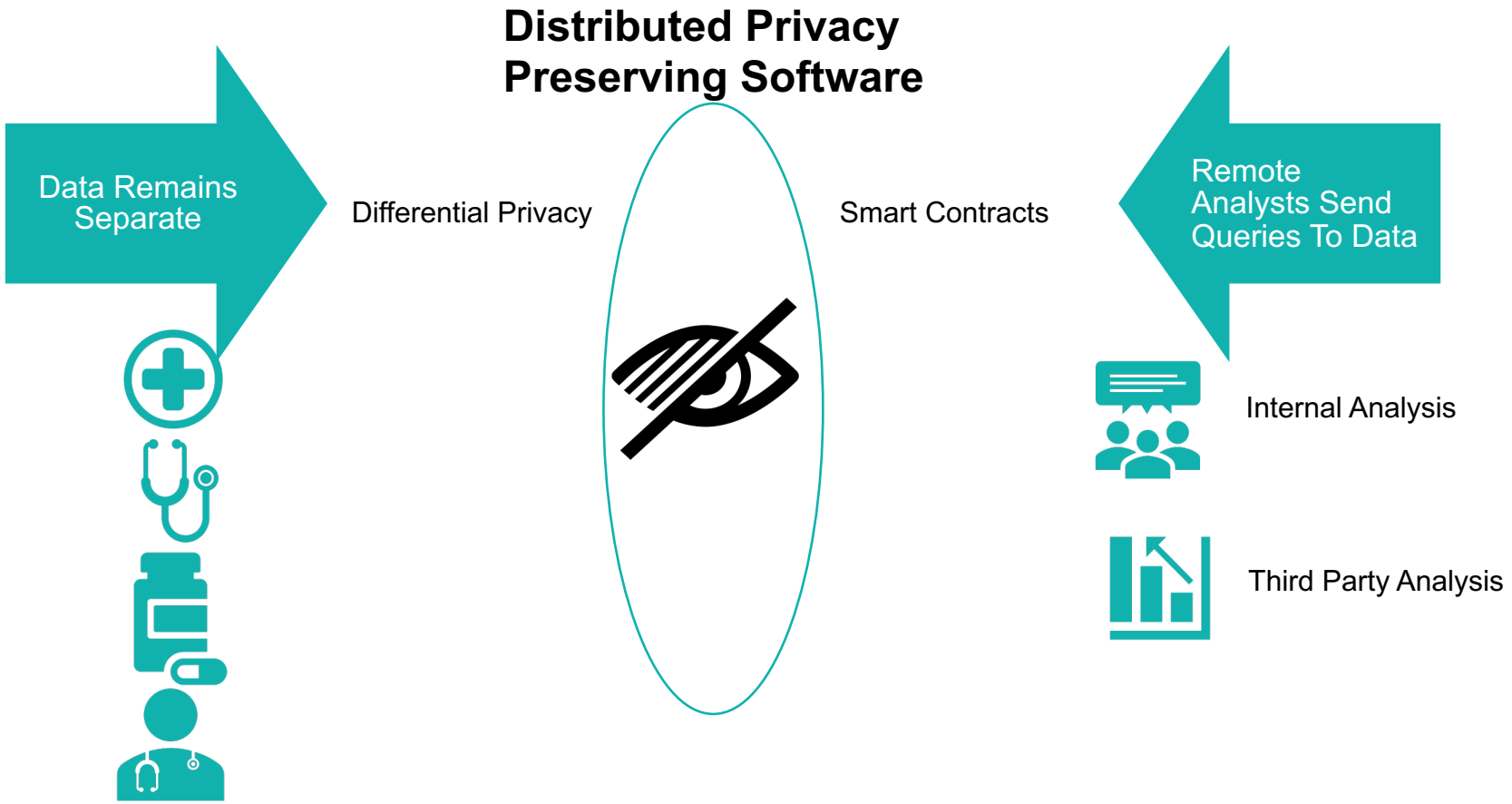
Community:

- Segmented by grouping patients according to different levels of their risk scores of certain disease and hold lectures/ tutorials/ other services
- Build a community patients can connect with each other anonymously, let them communicate to get company and fight again diseases together
- Make value of user-generated contents and encourage sharing disease control stories/ check-list result, etc. in and outside of community

Privacy Issues

Ideas about achieving "MILITARY-GRADE ENCRYPTED"

•Purchase/build **Distributed Privacy Preserving Software**, send analysts' queries without data transformation, uses smart contracts to verify queries.



References

1. Operational database: https://www.techopedia.com/definition/5711/operational-database-odb
2. Data Warehouse Architecture: https://panoply.io/data-warehouse-guide/data-warehouse-architecture-traditional-vs-cloud/
3. Data Cleansing: https://medium.com/hashmapinc/etl-understanding-it-and-effectively-using-it-f827a5b3e54d
4. OLAP server: https://www.oracle.com/technetwork/middleware/bi-foundation/olap-in-a-data-warehousing-solution-128690.pdf