

This report presents a Health solution proposed by our group and discusses in detail its Value Proposition, Design, Architecture and a Bird-eye view of Prototype

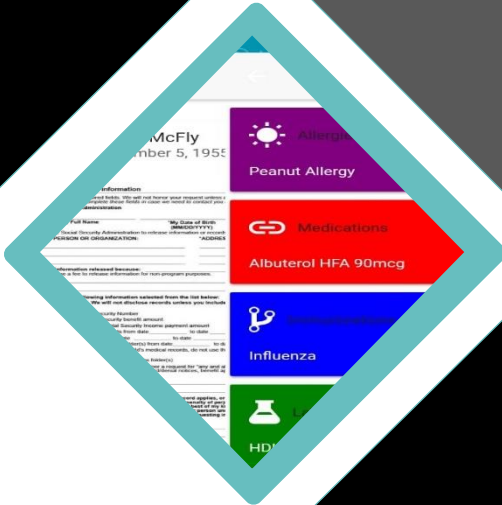
HEALTH YOU

2018

HEALTHYOU

**A Data Driven Solution for tomorrow's
Health Industry**

Project Report



Names: Ahsan Aftab, Camillo Baratta, Jorge Camino, Sharon Marin, Theo Tortorici,
Nikita Khudov and Michellé Greaves.

Master of Science in Business Analytics and Big Data

Course: Big Data and Health

Submitted to: **PROFESSOR DAVID GONZALEZ PISANO**

Date: 21-06-2018



Team D, Section O-2



IE Business School, Spain

TABLE OF CONTENTS

Preface	4
Value Proposition and Business Opportunity	5
The Problem	5
The Solution	5
Value Proposition	6
Our Customer	7
Product and Service Design	7
Product Usage AND FLOW Diagram	7
HealthYOU Lifecycle and Journey	10
Technical Architecture and Features of HealthYOU Platform	11
Mobile APP and Personalized Profile for Patients	12
Med-Care View	13
Smart-Health View	13
HealthYOU Technology stack and Architecture	16
Data Privacy and Governance	18
Evaluation	19
Benefits	19
Risks	20
Scalability and Growth	20
Final Words	20
Appendix	21
Appendix A-1 Technologies Landscape	21
Appendix A-2 Interfaces of HealthYOU	22
Appendix A-3 ARCHITECTURE of HealthYOU	24

Appendix A-4 DICOM DATA Example with HealthYou API.....	25
Appendix A-5 Machine Learning Example with HealthYOU	26
Appendix A-6 Network Analysis Example	27
Appendix A-7 HealthYOU Platform Components Explained.....	28

PREFACE

This report is submitted as a part of final project in Big Data Analytics and Health Course and presents a solution developed by team D to enhance the customer experience in the field of health and medical science. The product offers different interfaces for different parties (patients and doctors) to interact. Although the user interfaces of the product are highly user-friendly and simple, the backend product proposed in this report is equipped with highly advanced technologies of today like Big Data backend, real time data analytics, machine learning and predictive analytics. The product is currently being offered as B2C, with a partial B2B incorporation, however our team sees a much broader aspect and perspective of B2B business growth in the future. Keeping in view the fast paced technologies, the product has been designed to allow for the maximum horizontal, vertical scalability in technological stack as well as scalability in business model.

The report first discusses the problem and identifies the business opportunity in the field of health services followed by our proposition and solution. The report discusses some of the major benefits of the product and the evolutionary changes it brings in the life cycle of a patient in particular and health services in general. This is followed by description of the technologies being used in the product development and the key benefits they bring along with them. Also discussed in detail is the architecture of the product and the description of key features that were introduced by it into HEALTHYOU platform.

The report also covers an appendix with a few important examples and MVPs that have been developed in the process addressing some of the key issues being faced by the health industry. The report also briefly highlights how HEALTHYOU team plans to scale the product in the future and incorporate government health bodies and other businesses as a key stakeholder in the process. In a broader view, we plan to be a part of Smart City, Smart People and Smart Governance concepts in the future by effectively contributing to the health services and health industry. And in the entire process, the privacy of end users and governance of underlying data at HEALTHYOU platform is of the utmost and prime concern to us. That's why the platform has been designed with specialized data security, governance and data management layers.

The ultimate goal of HealthYou team is to promote practices and solutions that facilitate and ease access to the medical and health services and enhance the quality of experience and quality of service for the general public. In the process we also intend to augment the operational efficiency and quality of operational management in the operations related to health services and while working in collaboration with health and medical service providers. And in the end we would like to thank our team for being a part of this wonderful journey and dedicating themselves for a great cause.

VALUE PROPOSITION AND BUSINESS OPPORTUNITY

The first section of this report discusses the basics of the problem, its impact and some of the key issues identified in the today's health industry, followed by the business opportunity it creates for us in this digital era of big data analytics and AI. The HEALTHYOU solution is particularly focused on Health Care and Medical Industry and discusses the various prospects and features of the product hence forth in the coming sections and how these features help the customer to overcome this challenge.

THE PROBLEM

If you ever have had to change a doctor, in case of an emergency, or had to retrieve historical medical records you know that it is a hassle to access this information. Although most industries have undergone digital transformations, healthcare falls behind in this area. However, with the creation of a centralised system where all medical data is held, a doctor or patient will easily be able to access information.

THE SOLUTION

With the creation of a centralised data system it entails that all data is stored in one central location, moving data from silos into a collaborated system. A doctor or patient can easily access the information whenever or wherever it's required and moreover it allows a doctor or a patient to easily access the relevant information from any internet connected device.

The solution provided by HealthYOU is simple to understand, easily accessible, efficient and user friendly. In terms of technology, the solution or platform offered is highly scalable and flexible incorporating connectivity and seamless integration and fully equipped with today's almost all the popular technologies of Big Data Analytics and advanced machine learning. (Please refer to the Appendix-A1 to see the list of technologies being covered by our product).

This flexibility and scalability was desired as some of the in-place systems and technologies being used in the health industry today, greatly vary across the regions around the globe. This flexibility was also desired to be able to integrate the product with various in-place and running systems today in the health industry.

Our proposed solution gathers all medical records for patients through this application, data is processed and stored at a central point or hub. This means that all medical records can be easily accessed by doctors, specialists, during emergencies, by patients and hospitals upon a valid request.

The solution offers:

- Instant Access to your medical profile and your entire medical history
- Highly flexible and connected system, brings the care and efficient response time making health services available to you and your family at your fingertips.
- This is a way to give users and medical practitioners access to health information. Holds details like allergies, medical condition and treatments, medicine details and test scan results in one place.
- The solution offers additional integration with a central database or repository of governments. This allows the government to better visualize the trends and key metrics related to health industry in the general population.

The end user of the app you can personalize the Healthcare Manager App by adding:

- His or her personal notes to the personalized profile.
- A personal health summary that he can access from anywhere in the world
- The user can add allergies or adverse reactions if any.
- A Personal health note: with a date, title and short description

The doctor or healthcare professional may review and add:

- Symptoms found in the patient during visit.
- Remedy and tests proposed.
- Any health services provided to the customer and the feedback.
- The doctor can see the entire patient history, the services he had previously acquired for health issues and the outcomes of tests, treatment or surgeries can also be seen in real time on previous medical and test reports.

The government section in general allows to review the numbers in the population in a summary format and may not be able to access the personalized content of patients or individual users. However, some of the very important figures the government bodies may obtain and review:

- Different health related issues encountered in a day at all the health facilities who have registered formally through our platform.
- Trend of health issues or certain diseases over extensive periods of time in the general population
- The population can be segmented based on demographics in a country or a topology that can be pre-defined by the government services.
- Analyse the data on big data platform and run machine learning algorithms to classify and predict the diseases within the population.

Hence, the solution brings the three main pillars of health industry the people, the healthcare providers and the government under the same roof of a highly advanced medical and health care system.

VALUE PROPOSITION

Taking healthcare into the future, through and accessible data. We are going to help you and your family have better control over your medical records.

Our mission is to be the number one companion for patients and doctors before, during, and after treatment. We as a team plan to be a part of the entire life cycle of patient and facilitate our customers in better managing health and treatments.

Our value proposition is to help patients and doctors to diagnose and complete the treatment by sharing critical information and supporting the patient during and after treatment. This would not only make the health services more efficient and affordable but enhance the quality of experience for both patient and health service provider.

OUR CUSTOMER

HealthYOU as a responsible company assures its services as a company to the following customers and sectors

- Customers: Patients and Doctors (easily transferrable information)
- Pains: Current difficulty and issue to address is that all current or past medical records are kept in silos and in certain cases still paper documents and not electronic copies. Further to that the files are in different formats and in different locations
- At times a new doctor or medical practitioner may need to contact a previous medical doctor that saw you. The contact details for previous medical practitioners will be available if needed, or in case of an emergency.

PRODUCT AND SERVICE DESIGN

As seen in many of the industries today like Telecom and social media, Velocity and Volume of the data poses a real challenge. However interestingly, Medical and Health industry faces a different challenge, the most important challenge being faced by the health industry today is the Variety and Access of data. And in order to address this issue, our platform addresses it using some of the most popular big data technologies which are reviewed in the coming sections. The platform ensures that the end user has the most simple and enjoyable interface, allowing them to easily browse through the services and get to the target information in almost no time.

The customer or patient may book an appointment, review and receive his test reports, consult a doctor in case of emergency through a personal message and carry his entire medical history with him or her in the pocket entire team. The patient may access the online readily accessible profile on mobile and web and review and print in desired formats in almost no time.

The doctor or medical services provider on the other hand can use our service to enhance their operational efficiency, management and better understand their patients. The interface is different from the one offered to patients and has many more features to improve drastically the performance and response time of medical services. Both the patients and the doctors may set reminders through the application interface to make sure they never miss an urgent appointment. The past history of patient enables a doctor to better understand the medical issue at hand and come up with an enhanced diagnosis for the patient.

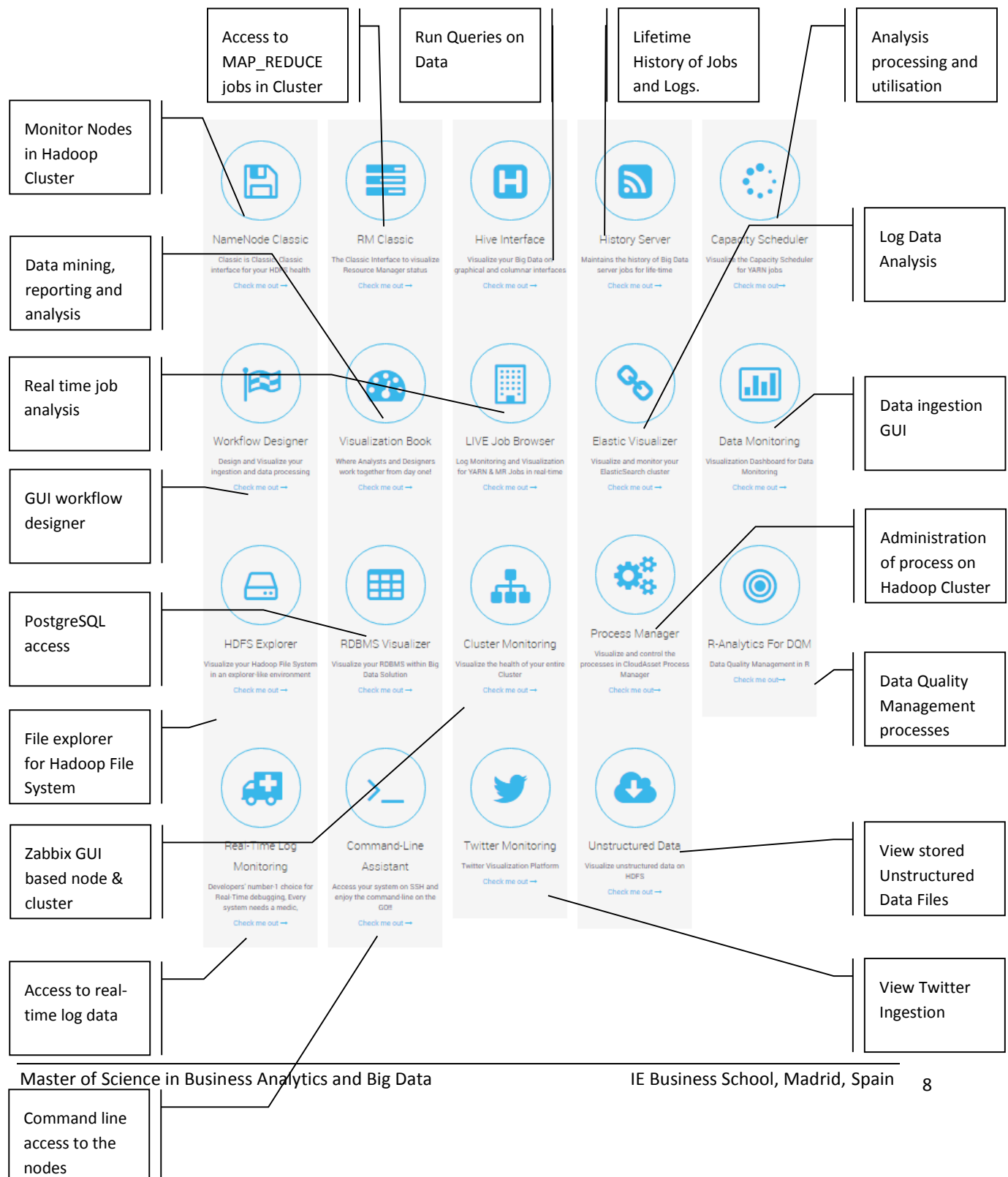
PRODUCT USAGE AND FLOW DIAGRAM

The product offered by HealthYou team has been designed by a team of data scientists and solution architects in consultation with various health facilities and experts. It's a complete platform that gives access to three key stakeholders in the journey:

- The Patient
- The Doctor
- The Government and other businesses (in near future)

The general architecture of HealthYou platform looks like: (Please refer to Appendix A-7 for description of each component below)

HEALTHYOU SCHEMATIC

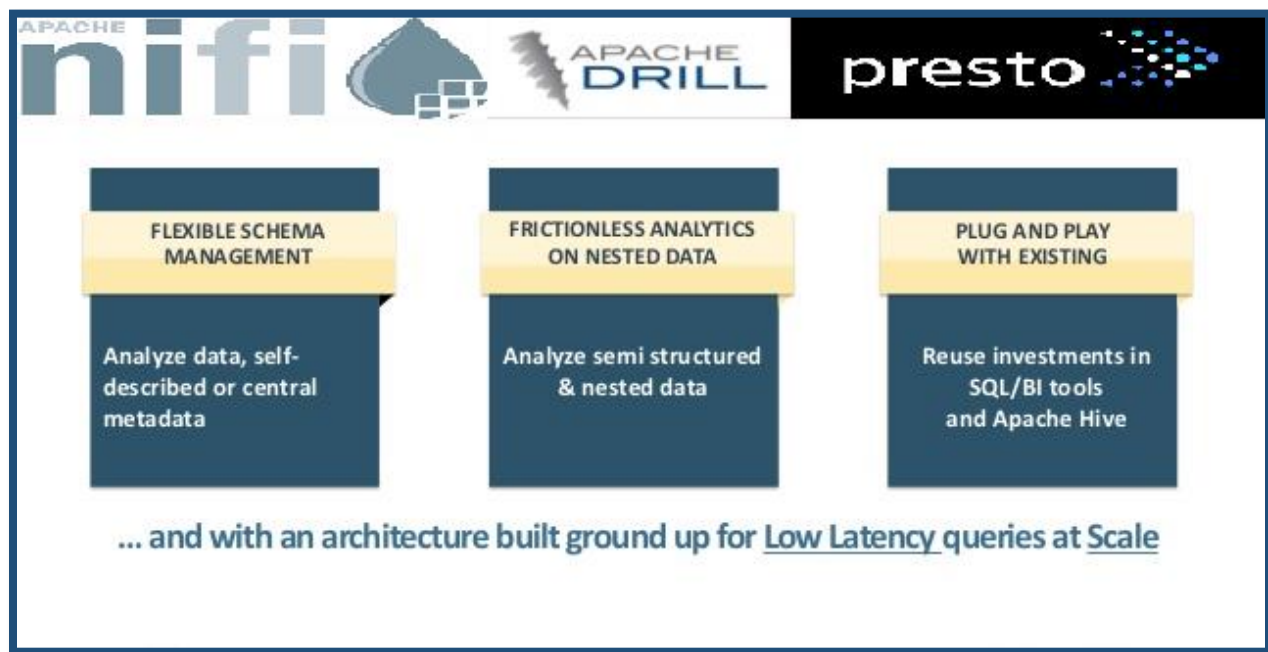


As seen in the above figure the HealthYou System sits in the centre of all key technologies when it comes to health and the medical history of patients. This allows us to provide services to all the key stake holders in a patient's health journey without any hassle. Our design experts are fully aware during the entire product development cycle that the following key challenges are being faced by medical industry when it comes to centralization of data:

- Since the medical issues arise or occur to patient with a sufficient in-between interval, the information can get easily lost in the process. Therefore making this information **readily accessible** is the key.
- In case of an emergency the information must be readily accessible to both patient and the doctor. This is where **system stability** is a key for our business.
- Some health care facilities have some kind of digital solution in place, results in **variety of data**.
- Some health care facilities might have to start from scratch, results in efficient solution deployment strategy
- Governments are unable to understand health data due to variety and veracity of data in medical field
- Data can be both structured or unstructured(image, scans, prescriptions etc.) posing yet another challenge

In order to address these issues, our team of technology experts wanted to design a robust system that can integrate seamlessly with any popular database and data acquisition technology today. An example of this kind of integration with medical data is provided in **Appendix A-4** where image data could be easily extracted from DICOM files using our technology stack in almost no time. The data provided by DICOM format is generally considered as unstructured, containing within it both image and text data. Our solution could easily read this data using the popular DICOM Java library included as one of the core libraries of the platform and made it possible in almost no time. Please refer to appendix to analyse in depth the case of DICOM data.

Additionally in order to address the above key challenge, our team designed an API inspired by the popular big data open source technologies like **Apache Drill**, **Apache Nifi** and **Apache Presto**. These technologies were the key of inspiration for our developers as we developed a centralized and configurable key API application that gets integrated with any sort of database that exists today. The output of the API endpoint is a simple readable JSON formatted text, which in turn can be used by various other applications for analytics, modelling and predictive data mining.



The three technologies are at the core of our built in interface to integrate seamlessly with any health care database and technology. The output of the platform is just an API endpoint that allows the governments to connect to the central platform and obtain the key health stats and trends of the population through a centralized platform.

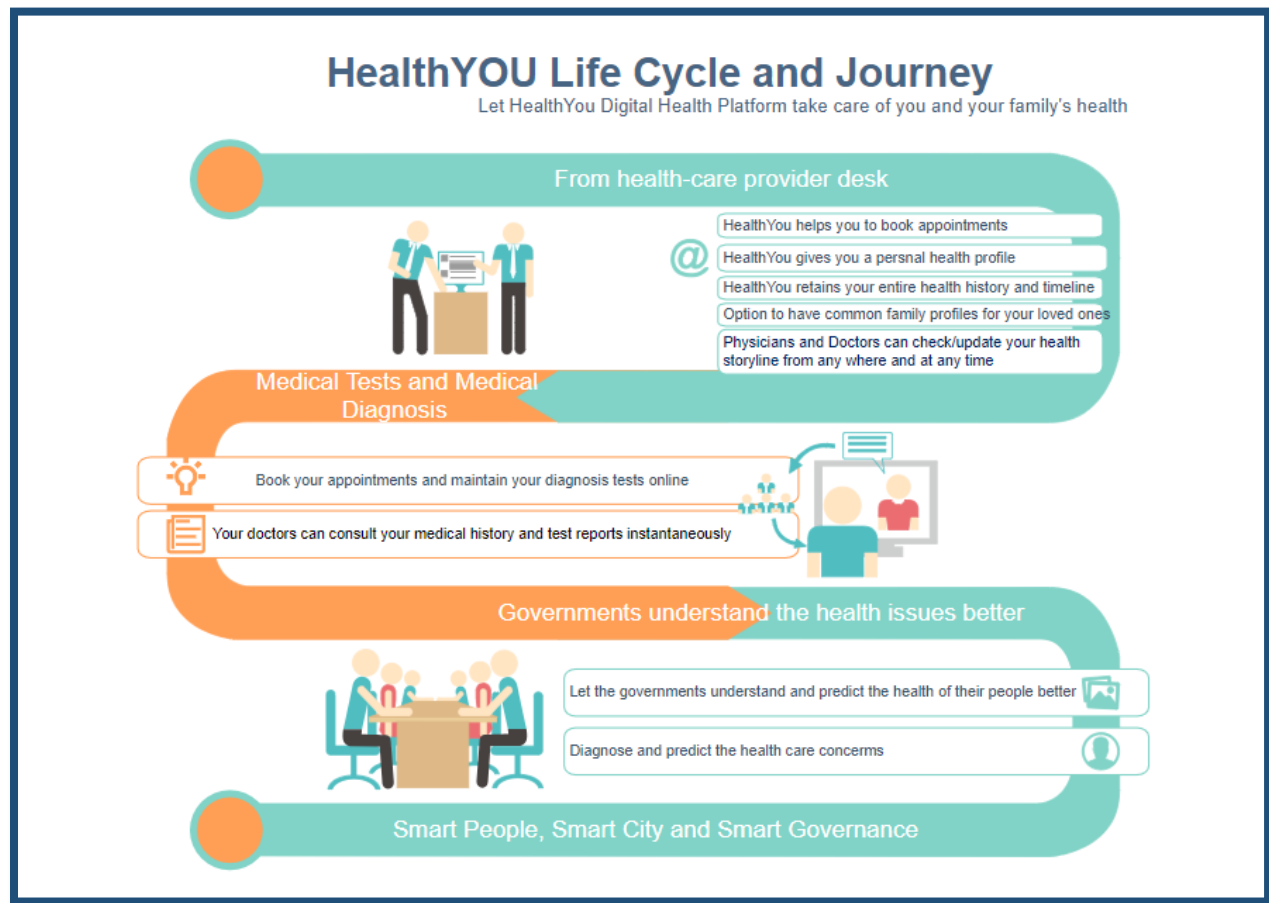
HEALTHYOU LIFECYCLE AND JOURNEY

The HealthYou solution offers three different interfaces that allows it to be accessed by patients (general public), health service providers (doctors, physicians, surgeons and health care facilities) as well as government health care institutions.

The patients and doctors may access the platform on mobile or notebook devices from anywhere in the world using their personal credentials.

The interface for government is only accessible through a special Java Application and web interface. Since Java is a universal language and applications designed on Java can almost run on any platform, therefore the HealthYOU platform provides a seamless integration with different operating systems like Linux, MacOS, and Windows.

A common customer cycle for HealthYou platform usage might look like:



TECHNICAL ARCHITECTURE AND FEATURES OF HEALTHYOU PLATFORM

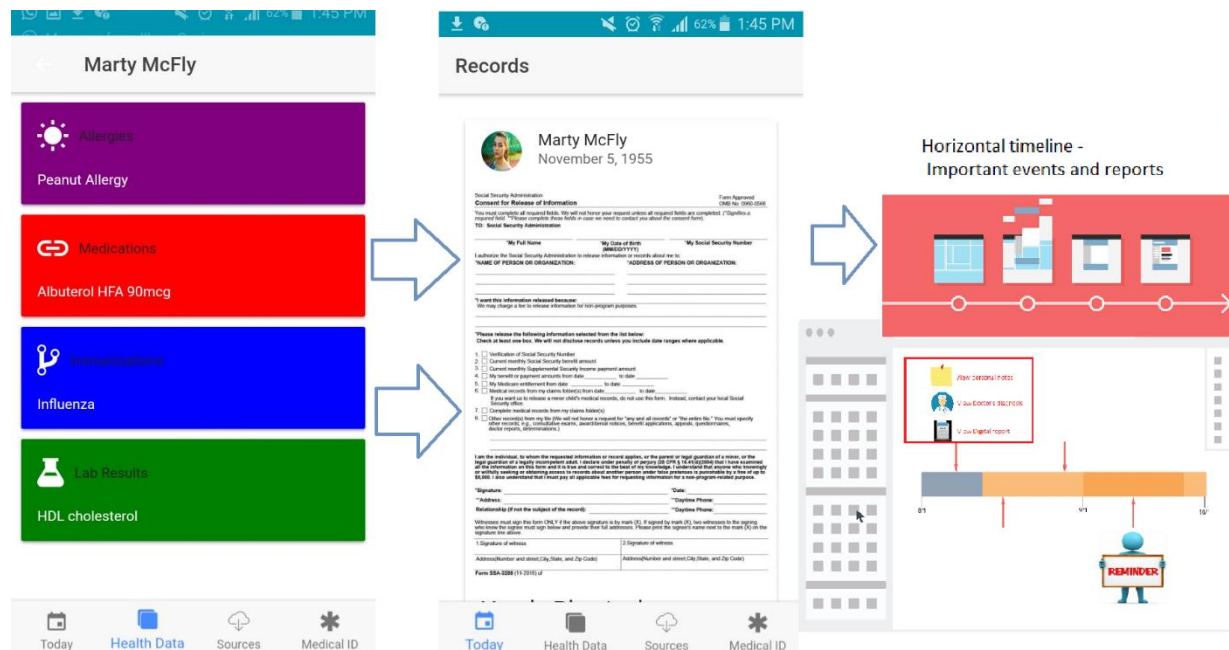
The data is centrally stored on Big Data repositories of HealthYOU that is highly fault tolerant and is highly scalable. The system is designed with the leading technologies of Big Data and Machine Learning era that allow us to both horizontally and vertically span the system. By horizontal integration we intend to make the system grow in terms of additional capacity to manage the volume and velocity of the data. By adding vertical scalability we allow the systems to be replaced by more advanced technologies both in terms of hardware and software and primarily address the issue of variety in the data.

The HealthYOU platform currently offers three different interfaces for its usage and productivity:

- Personalized View – Designed for the patients and allows them access on mobile and web
- Med-Care View – Designed for health services providers
- Smart-Health View – Designed for data discovery, HEALTHYOU R&D and governments

MOBILE APP AND PERSONALIZED PROFILE FOR PATIENTS

The simplicity, user-friendly design and ease of usage was the main concern for our developers while developing this interface. This interface is available on both mobile and web in the form of an easy to use application. The application can be accessed online from anywhere in the world. Our data governance ensures and complies strictly with the data usage and privacy rights of our customers. The following figure shows how the flow of actions might look like to a general patient or subscriber. The services offered by us are freemium and therefore in the reach of every person. The interface offers a main browsing section followed by detailed profiles for each section, and further seamlessly integrates to a relevant history section. This flow allows the patient to reach to the desired information in almost no time.

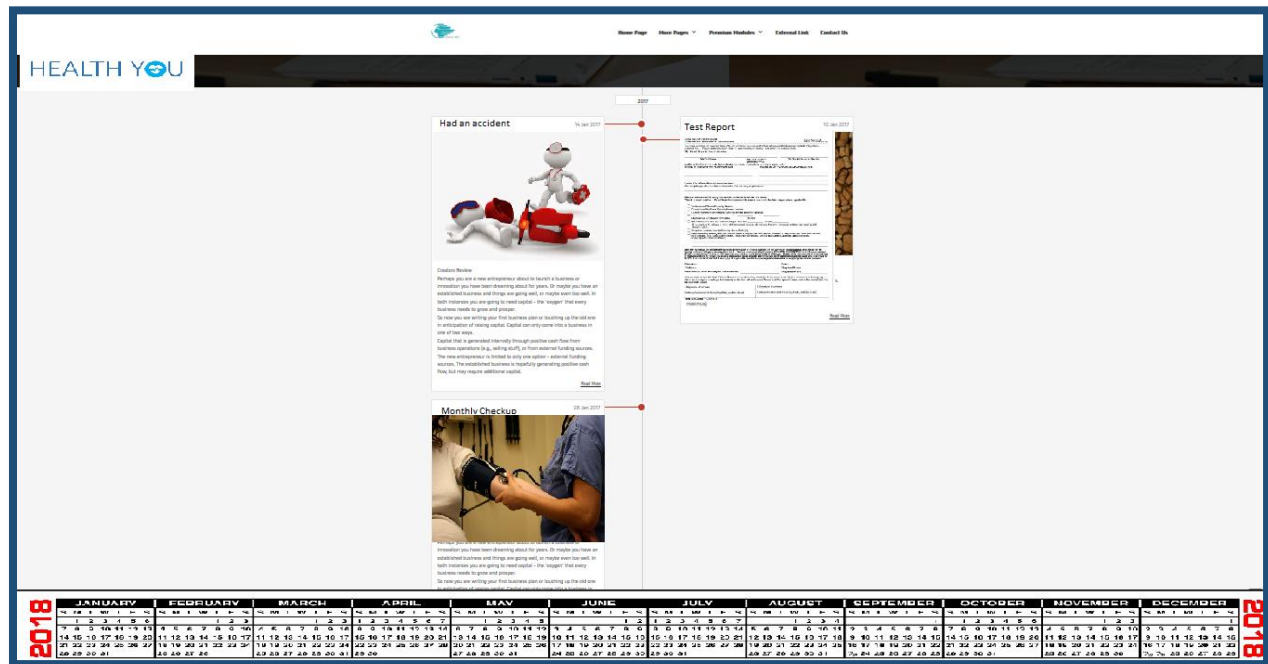


The patient may share certain reports or areas of their personalized profile with their medical doctor or advisor. This requires a user confirmation from both parties and once received, this information is readily available to both parties.

Certain actions or prescripts by a doctor will also appear in the patient's personal medical history and the application allows the user to set reminders or book the next appointment right from the app. The schedule and availability of doctors is also available to the customers in a specified section.

MED-CARE VIEW

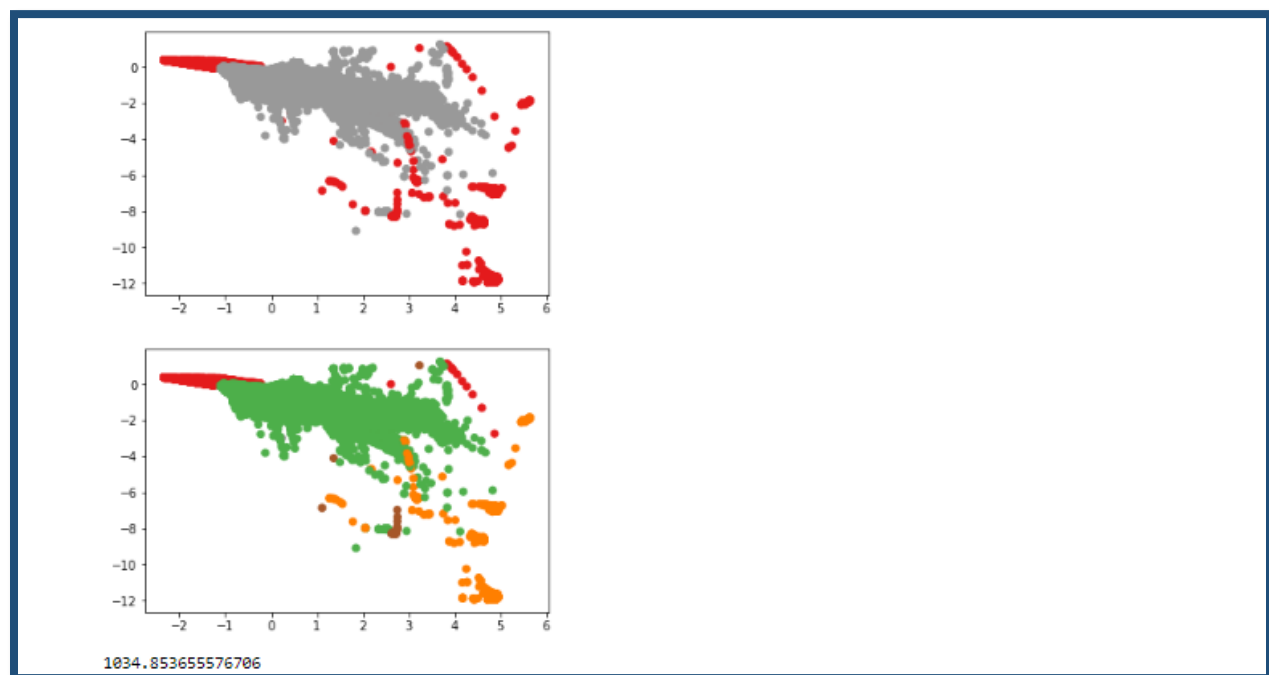
The med-care review is designed for medical specialists who can see the medical history of the patient along with all medical diagnosis in the past. It also incorporates different test reports and diseases the patient has previously suffered from. A general view of this interface look like:



The prescripts or messages by the doctor readily appear on the patient's profile and the patient may also book appointment and review the changed schedules of doctor or appointments if any. This avoids the hassle of calling each patient separately and helps the medical team to efficiently run their services.

SMART-HEALTH VIEW

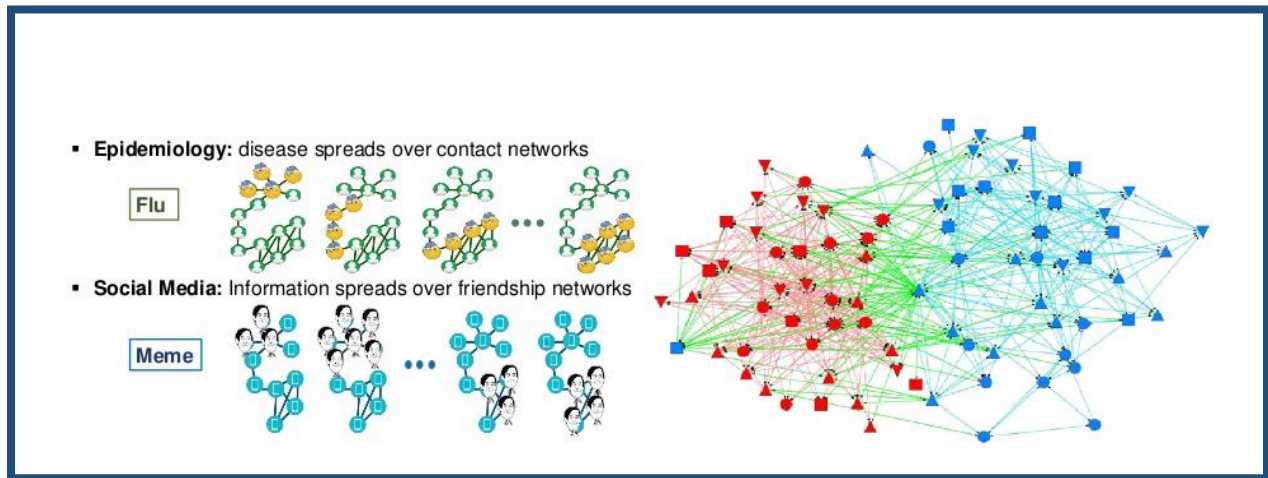
The smart health view allows the government health institutes and our experts to analyse **anonymised data** from a central platform. The platform can be accessed through a Java Application or Web interface and interacts with the central HealthYou platform through a secured API endpoint. This allows the government bodies to statistically analyse the health data of population and perform some advanced machine learning and data mining on the given data. The interface also provides advanced features for predictive analytics and cutting-edge machine learning capabilities. An example of such classification applied through SVM is provided in the **Appendix A-5**



The above model generated a classification model of the population based on demographics and as has been observed may provide us the key insight why people living in the certain areas of city are more affected by certain type of diseases. This information can be very insightful for the government and international health organizations

and data scientists studying specific health issues and outbreak of diseases. Certain patterns may reveal in this study which eventually might lead to the cure of the problem.

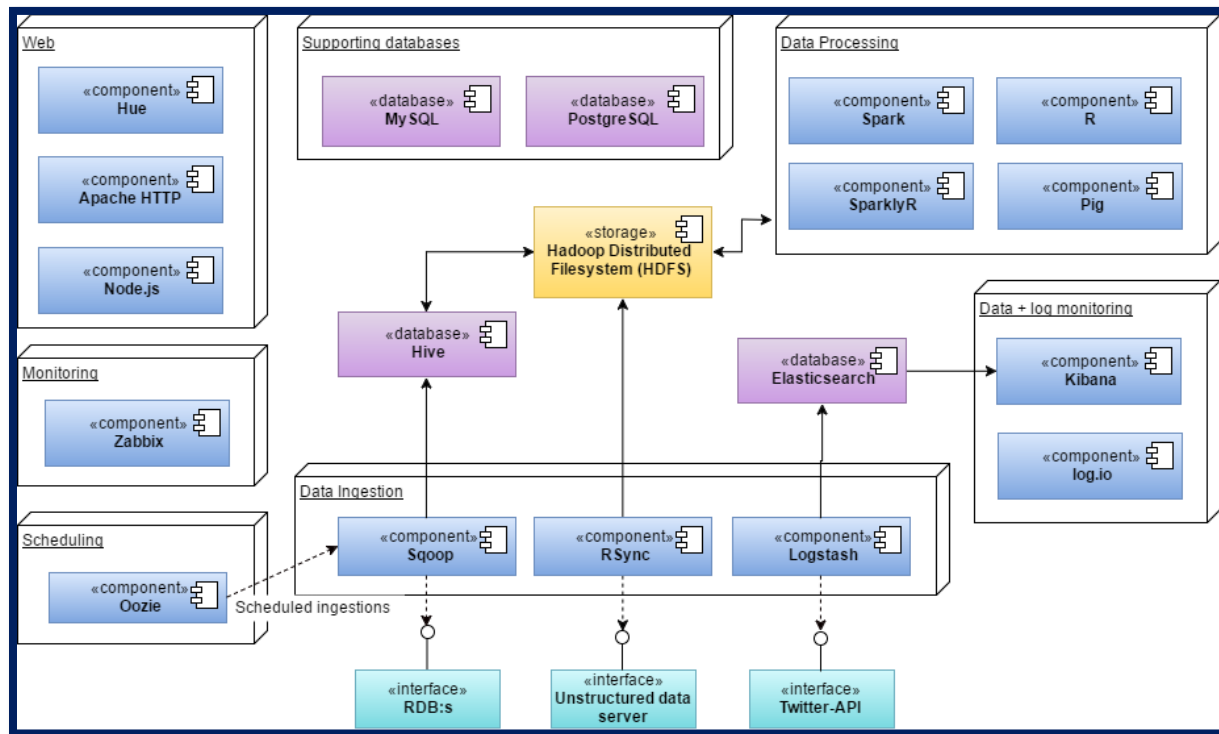
Similarly the following view and feature in the application provides and builds a graph network for the patients and helps the government to ascertain the reasons of certain disease outbreak or seasonal infections.



The graph output when analysed over dynamic timeframes allows us to see the origin and spread of certain diseases over time. It has helped to figure out the root causes and mysteries of various diseases that health industry had been struggling over years to figure out.

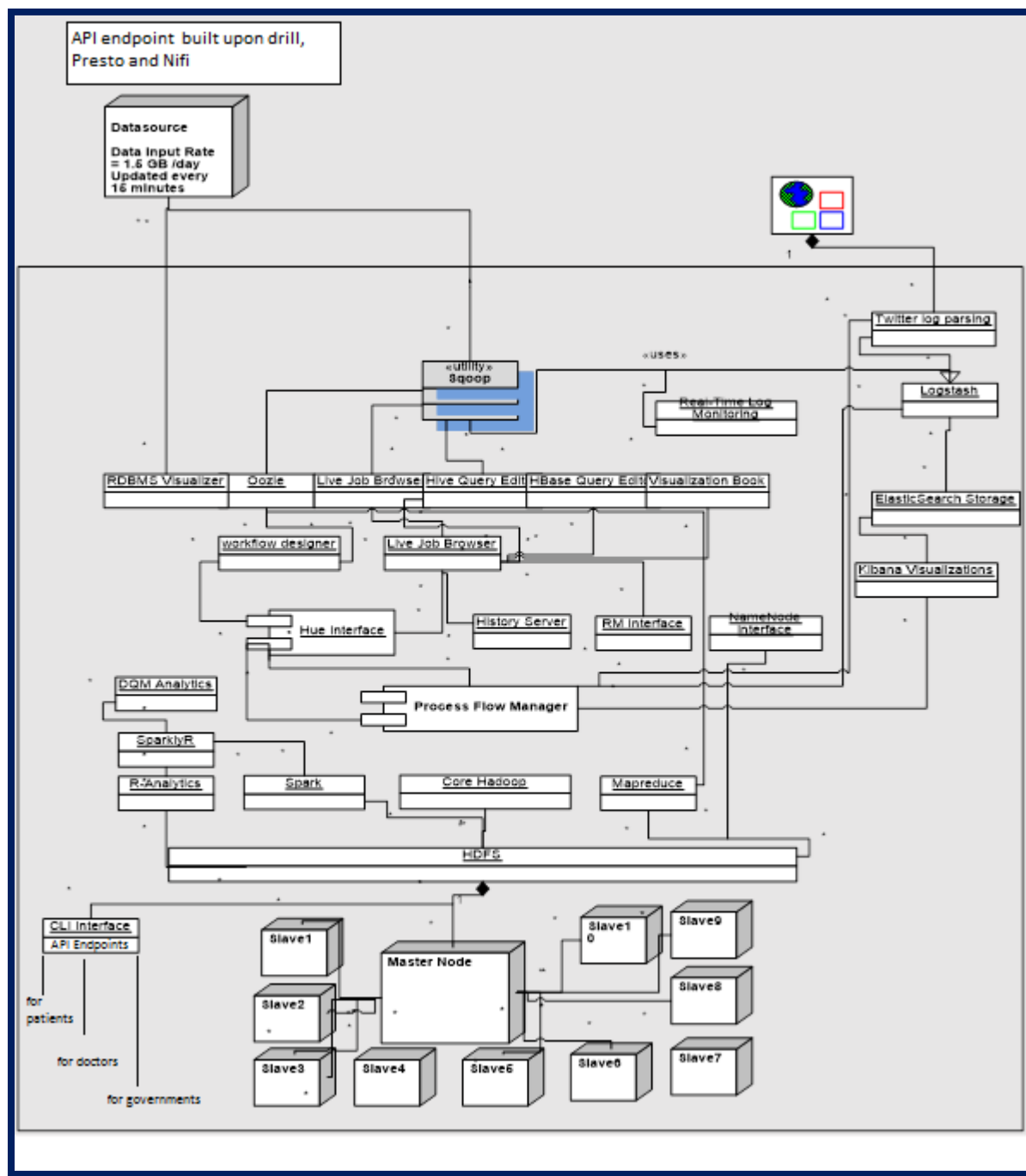
HEALTHYOU TECHNOLOGY STACK AND ARCHITECTURE

The technology stack of HealthYou platform consists of following layers of data management and as can be seen below:



The stack is built upon the secure, robust and powerful components of latest big data technologies. The stacks allows our product to grow both horizontally and vertically when it comes to development of platform. These technologies provide key tools to our data specialist to both analyse and predict. The stack is highly flexible and customizable allowing any future big data technology like Cassandra, Redis, Carto and others to instantly integrated with our product. The technology stack of HEALTHYOU also allows us to ready integrate our core platform with any traditional databases and technologies like RDBMS, SQL, and Plain text file formats. The platform has the capability to process unstructured data from various sources like IP streams, images, P2P services etc.

Please refer to the following diagram for the complete architecture of the core platform. The centre of this platform is occupied by today's most powerful technologies and offers high scalability in terms of future demand and requirements. The components are readily replaceable with the updated versions through a version control mechanism and that ensure no downtime or slowdown of the platform in the entire process.

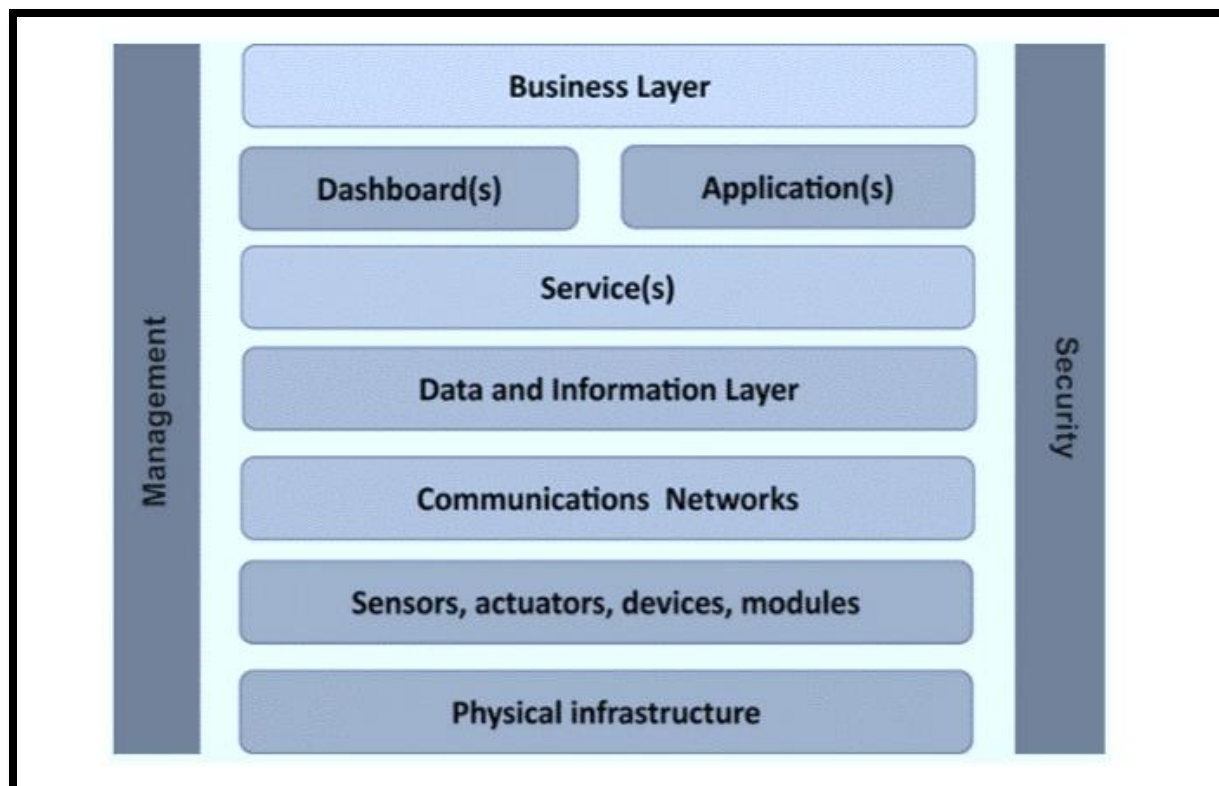


DATA PRIVACY AND GOVERNANCE

Our big data platform for health management strictly follows the standards of industry for data governance and privacy. The different layers of the platform have been secured by Security and Management layers and this allows us to implement the best practices in the industry for the data security and privacy.

Our company strictly complies with the terms and conditions of GDPR and other international laws when it comes to data security, privacy, data governance and data management.

The solution retains the logs of any changes or modifications observed during the entire process or cycle of the product. Any data breach observed is addressed with these logs with high confidentiality and while maintaining the international standards of data governance.



EVALUATION

BENEFITS

During the evaluation phase of the product our team came up with the list of following highly beneficial outcomes:

- Previously patients often had files of hard copies of consultations now with this centralized database and app it allows that after a consultation with a doctor or medical practitioner the file can be added to the database easily. Also patients can have the records/files emailed to them and added to the app to keep track of previous records.
- When travelling it is often a concern that you do not have access to medical record, however with this centralized system/platform a doctor or patient can easily access the medical information required from an internet device.
- Even if you move locally or internally a user will have access to their medical records and doctor would be able to have access to this information.
- For both clients/patients and hospitals or clinics it entails that historic data is easily accessed for consultations
- The data will be stacked in a manner that reflects the latest files first
- With the health manager the benefit is that every time a user sees a new doctor or are in an emergency (conscious) you don't have to repeat your past medical history and worry about remembering the details of everything.
- Medicines, chronic condition details and the dates of recent tests are available when you see a new doctor or healthcare provider.
- For a new doctor or in case of an emergency an overview of your health history can be accessed. This can be viewed as a useful source from where to gather historical timeline of medical occurrences.
- Beneficial information can be accessed such as
 - If needed hospital discharge summaries are accessible
 - Medications prescribed by doctors
 - Organ donation decisions can be added
 - Indigenous decisions can be added
 - Medical/aid plan details can be added
- If a patient needs to get a record or requires customer assistance a chat-bot can be created to offer another level of support to the app
- A benefit of the App would be that it could translate the information into the desired language
- In the health manager app the childhood development area will hold information about early health, growth & development. For instance one could add vaccinations that took place:
 - Personal measurements
 - Immunizations & health checks
 - Knowledge information about growth and development
 - Information for parents
 - Emergency contact details/information
 - Parental observation
 - Track children's development
- At times a new doctor or medical practitioner may need to contact a previous medical doctor that saw you. The contact details for previous medical practitioners will be available if needed, or in case of an emergency.

RISKS

As mentioned above, our team identified and understand the following risks of utmost importance in the product growth and development stage:

- Identity theft
- Hackers
- Ensuring the right data is captured under the right patient
- Data sharing between patients & doctors
- Data privacy and cyber security

In order to address these issues, we have set a foolproof mechanism and auto-check features that ensures the data governance is at its best when it comes to HealthYOU. Please refer to Data Security and Governance section above for the complete details.

SCALABILITY AND GROWTH

In terms of growth and scalability, we see the following areas where our team and company intends to span their efforts over coming years

1. Treatment recommendation engine based on symptoms for app users
2. Computer vision & NLP to translate written prescriptions into knowledge
3. Cloud computing to make data accessible from different devices around the world
4. Personalized medicine - employ data science to custom-tailor a regimen
5. Integration and Scaling of the business by having key partnerships with other health related business and government bodies.

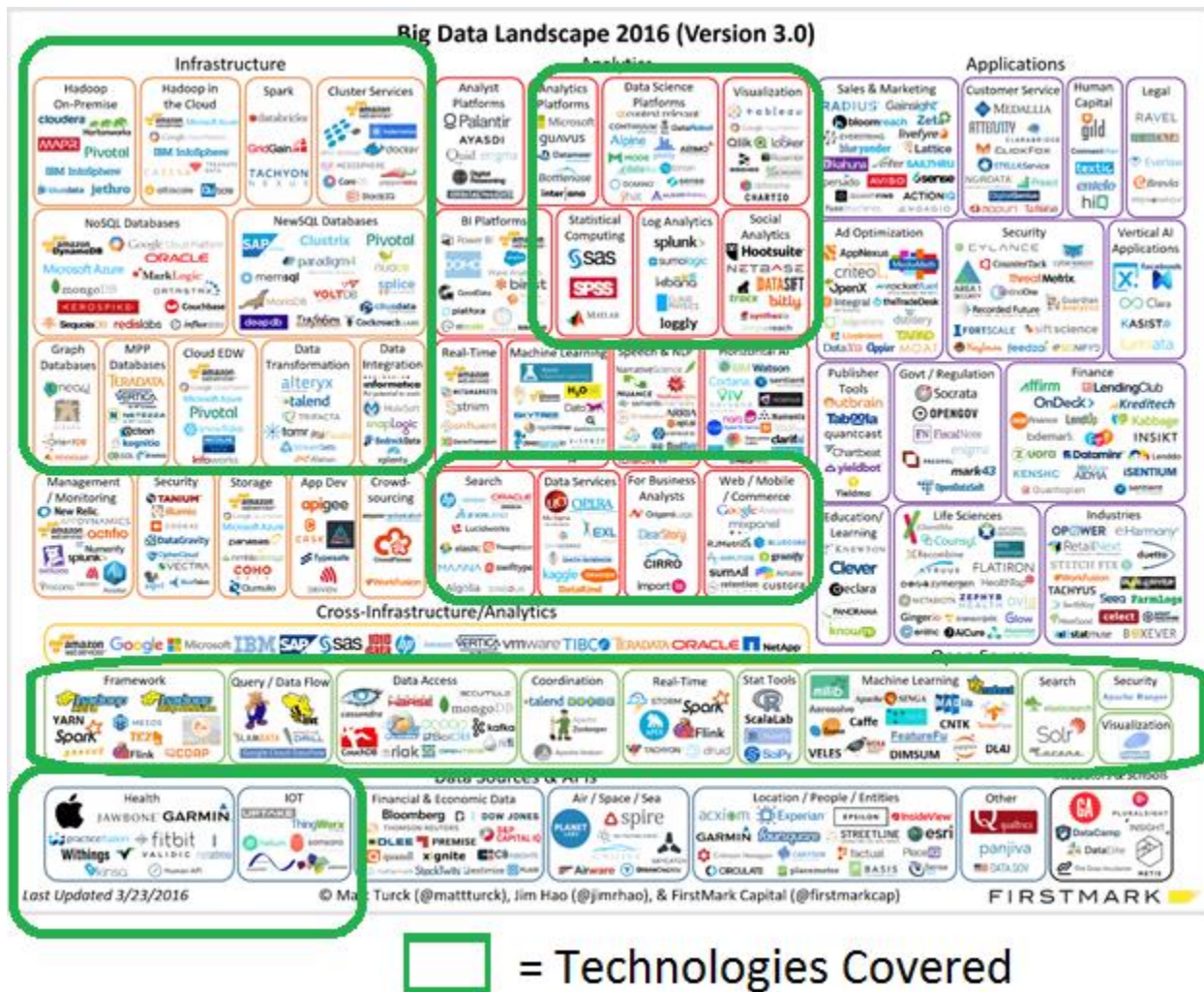
In order to review the scalability from technical perspective, please review the technical architecture section above.

FINAL WORDS

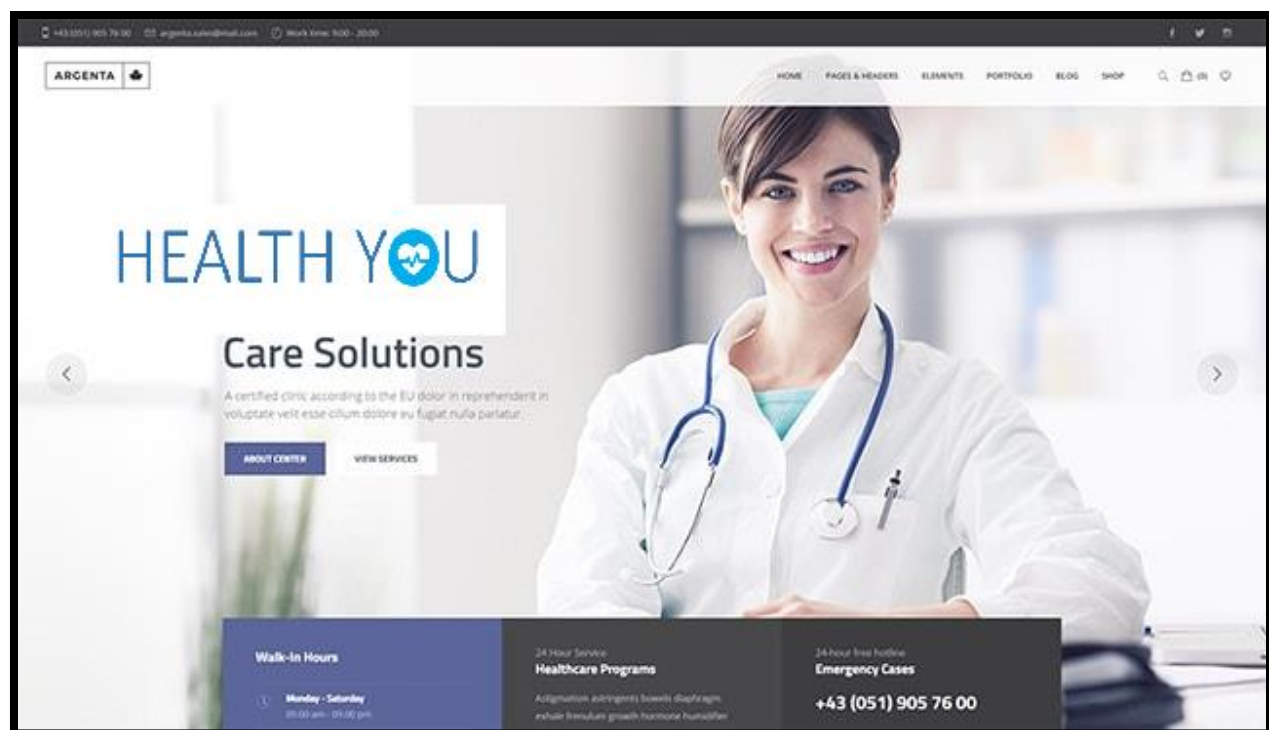
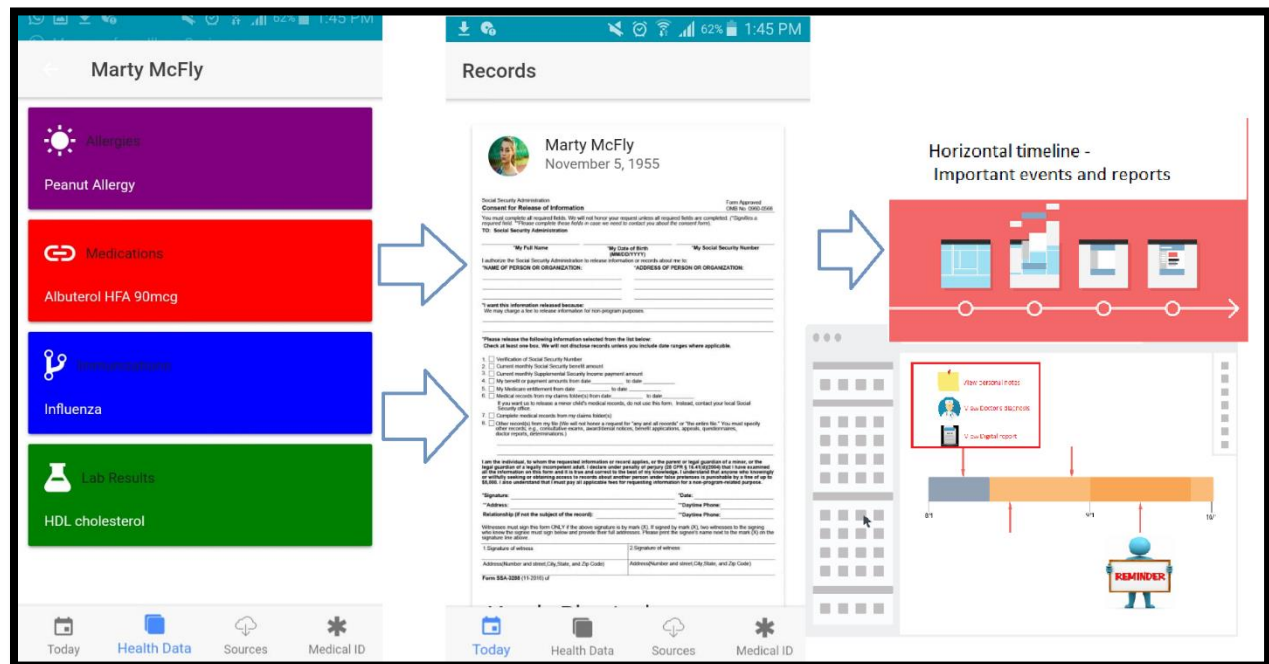
HealthYou as a team and service provider, is highly motivated and dedicated to bring this solution to reality and mainly due to the positive impact it intends to bring to an industry vital as health. The services have defined business and technological objectives in the business and operations domain. We as a team intend to make the health services available to everyone everywhere in the world. And our sole commitment is to the amelioration of customer experience, relief and satisfaction. We aim to bring new findings through better data discovery and engagement. Our sole mission is to bring new and innovative ideas to the industry with a category of products that are in-line with the requirements of patients, health savvy people, doctors and medical service providers.

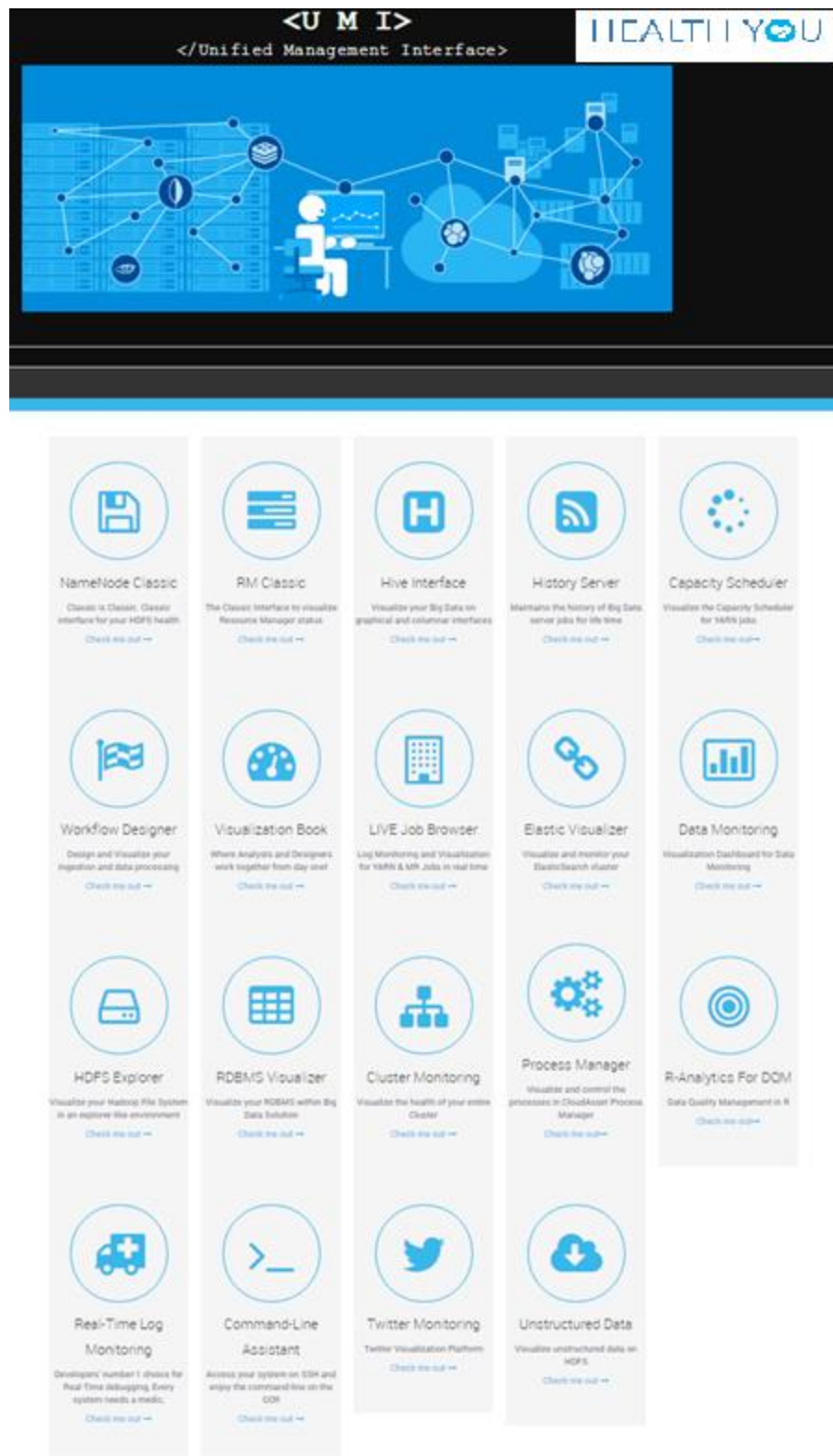
APPENDIX

APPENDIX A-1 TECHNOLOGIES LANDSCAPE

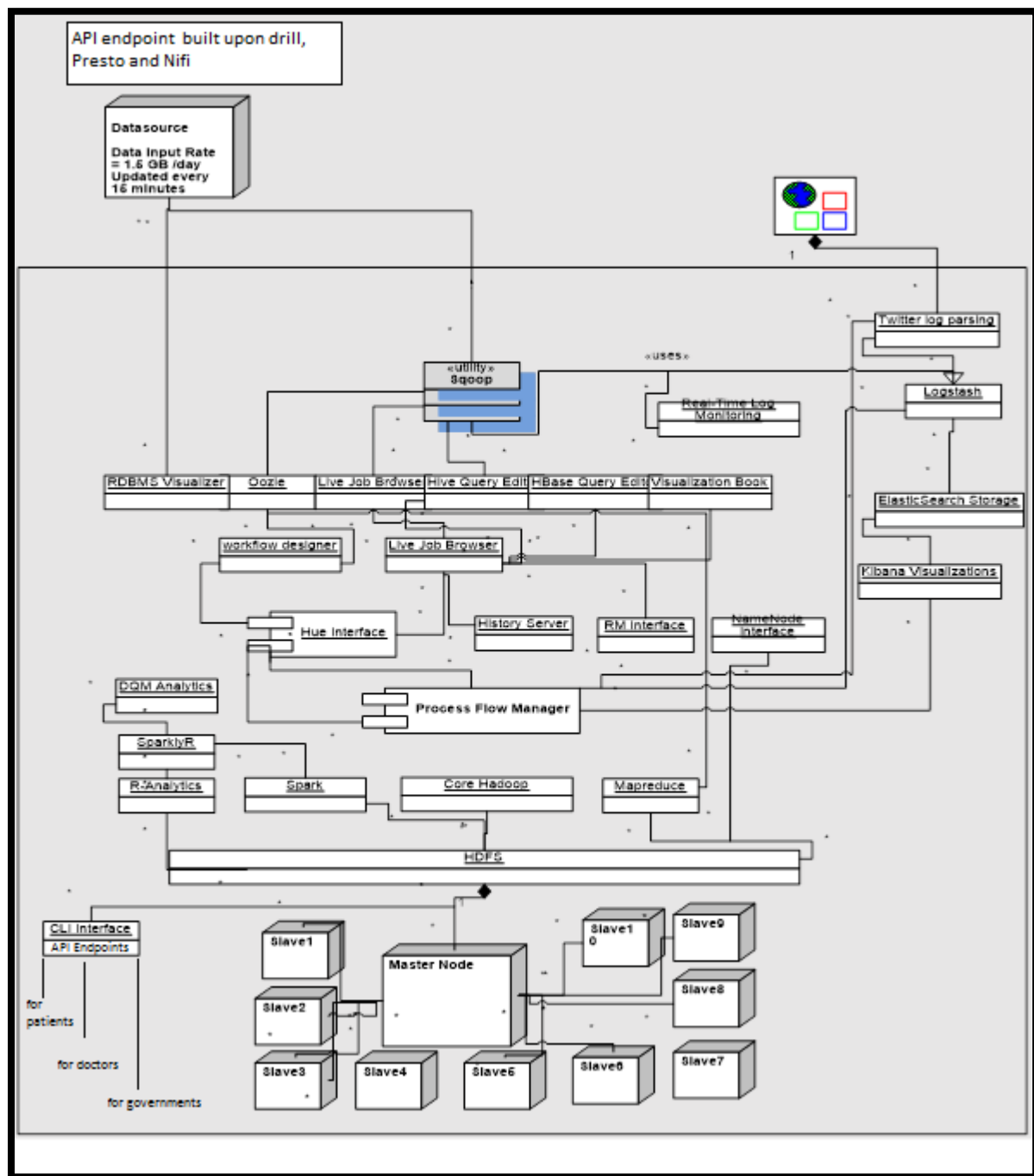


APPENDIX A-2 INTERFACES OF HEALTHYOU





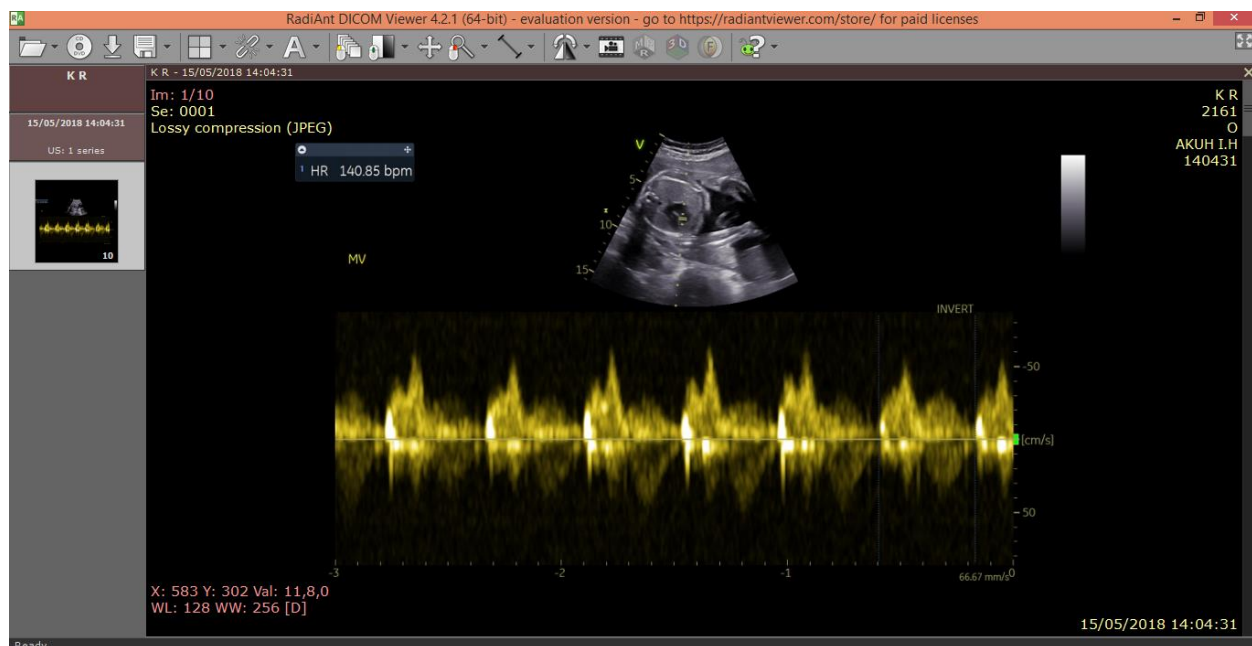
APPENDIX A-3 ARCHITECTURE OF HEALTHYOU



APPENDIX A-4 DICOM DATA EXAMPLE WITH HEALTHYOU API

Digital Imaging and Communications in Medicine (DICOM) is the standard for the communication and management of medical imaging information and related data. DICOM is most commonly used for storing and transmitting medical images enabling the integration of medical imaging devices such as scanners, servers, workstations, printers, network hardware, and picture archiving and communication systems (PACS) from multiple manufacturers. It has been widely adopted by hospitals, and is making inroads into smaller applications like dentists' and doctors' offices.

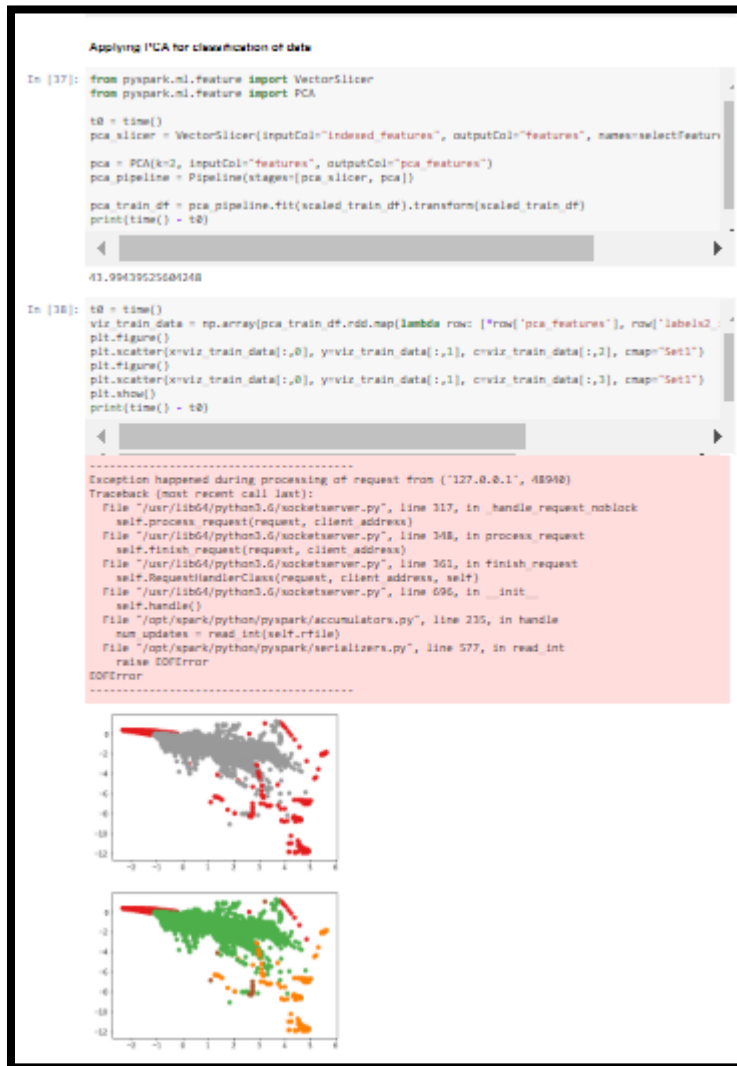
DICOM files can be exchanged between two entities that are capable of receiving image and patient data in DICOM format. The different devices come with DICOM Conformance Statements which clearly state which DICOM classes they support, and the standard includes a file format definition and a network communications protocol that uses TCP/IP to communicate between systems. The following image is a case where HealthYOU team was consulted by a medical team for transforming the following blood flow scans at the facility into the information that can be analysed efficiently and fed into machine learning algorithms.



HealthYou Core Platform Core libraries have OrdDicom library as inherent part of its API design. This allowed us to transform the above medical image into structural data in almost no time and this data can be then stored in standard database along with reference to the standard image file. A brief description of OrdDicom library and class is given below:

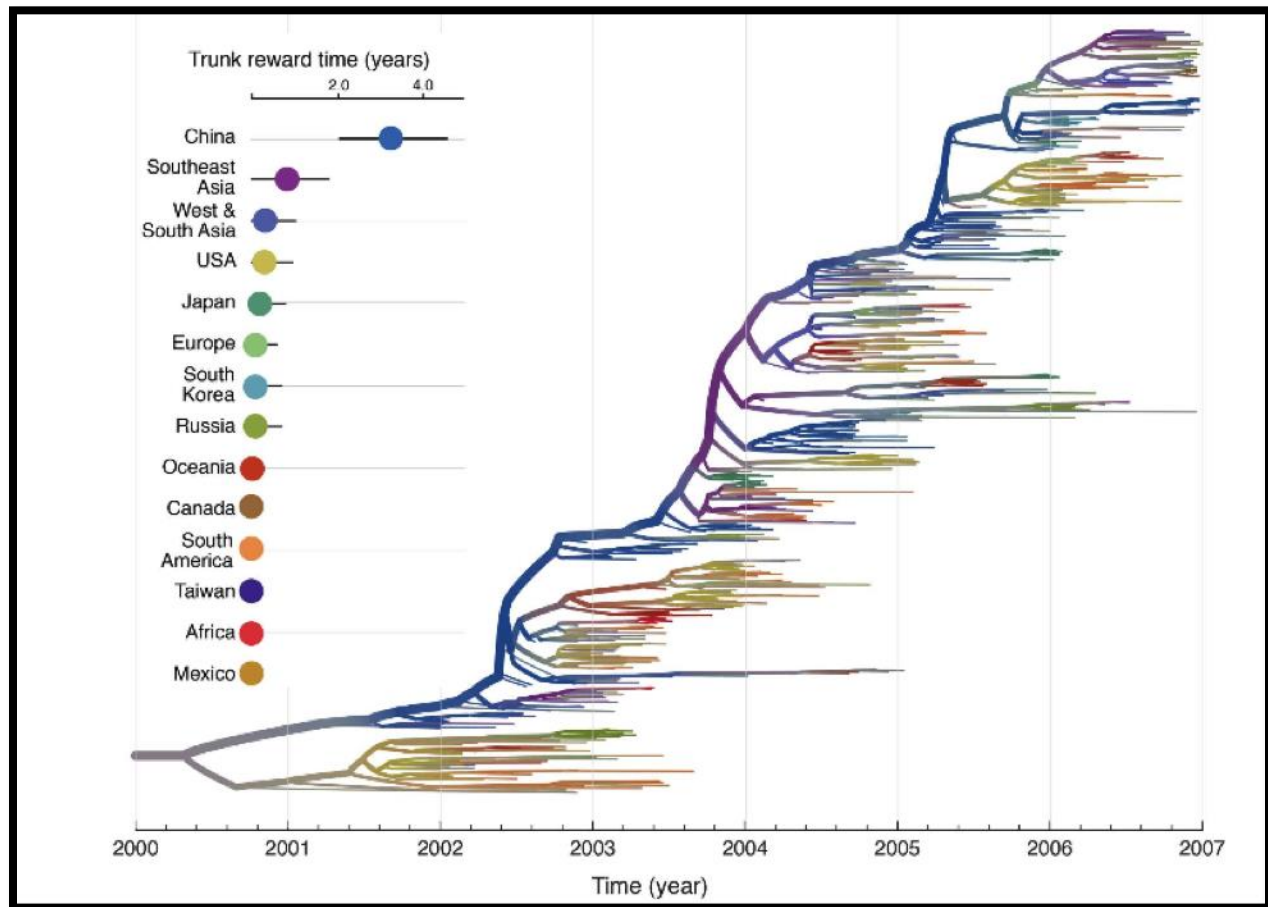
The OrdDicom class is used to represent an instance of the ORDSYS.ORDDicom database object type in a Java application. The OrdDicom class includes a set of methods to get object attributes, as well as a set of methods to perform various operations on an OrdDicom Java object. Users of this class should be familiar with Oracle Multimedia DICOM Developer's Guide. format in a structural database.

APPENDIX A-5 MACHINE LEARNING EXAMPLE WITH HEALTHYOU



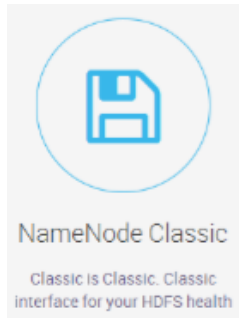
APPENDIX A-6 NETWORK ANALYSIS EXAMPLE

The importance of graph or network analysis can be well understood by a recent studies that showed the spread of Influenza in different regions of the world. Therefore including graph analysis capabilities is impertinent to our health platform



The graph features are currently in the stage of development and our team is highly impressed by the value they bring within big data analytics and health domain.

APPENDIX A-7 HEALTHYOU PLATFORM COMPONENTS EXPLAINED

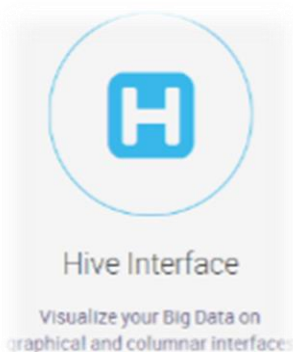


The Namenode classic UI or the Namenode web interface is used to monitor the status of the Namenode or master node in a Hadoop Cluster. It's a very useful tool to monitor and observe the basic health stats of the cluster.

Overview section provides Block Pool Id, Cluster ID parameter, Configured Capacity, DFS used, NON-DFS used, DFS remaining, Live and Dead Nodes status, Safe-mode Indicator and Namenode Journal. Namenode Classic also features HDFS browser and Logs browser.

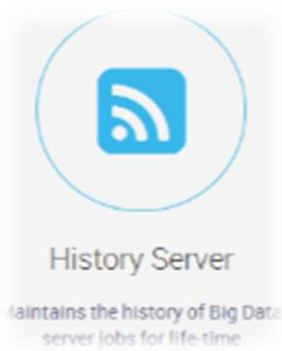


RM or Resource Manager Classic Interface provides real time access to running MAP-REDUCE jobs on Hadoop Cluster. The interface is a very helpful tool to visualize the on-going jobs, their status, cluster resource usage in terms of CPU and RAM, live and dead nodes status, and classification of running jobs as Successful, Suspended, Running or Killed. The Resource Manager can be a primary interface to explore processing related issues or performance tests. The jobs history is retained in Resource Manager for limited periods or unless the cluster is restarted. The retired jobs are archived and moved to History Server.



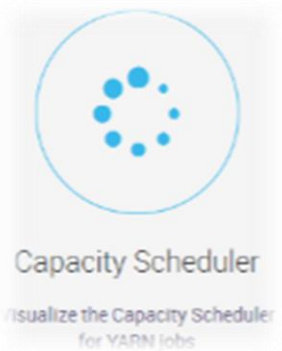
The Hive interface, powered by Hue, is a primary tool to run user queries onto Hive data. The interface is very powerful in that the user queries can be designed, saved and tagged in this interface. Once the designed queries are finalized they can be added to notebooks for enhanced dashboards.

Hive interface also provides analytics for teams to analyse data on graphical interfaces. Thereby ensuring that the developers and analysts can start working on the available data from the first day. Hive interface also indicates queries status, runtime and retains the results for a few hours in temporary storage on HDFS.

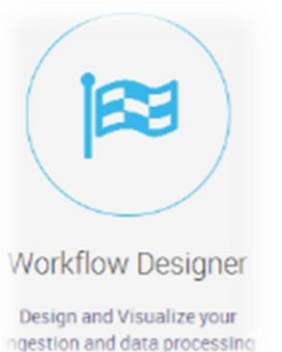


History server is the archive repository for Resource manager. Whenever a job is finished (successfully or otherwise), it is considered as retired and pushed to the History Server. The history server maintains the logs and jobs history for the lifetime of the cluster on HDFS. Thereby ensuring that historic checks and past job analysis may be conducted on any cluster whenever required. The history server also provides advanced counters and logs for analysing in depth the status of every job that ran on the cluster.

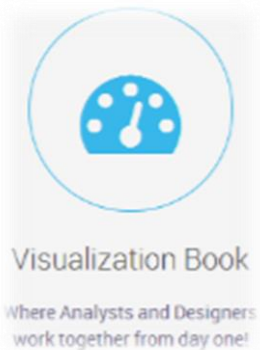
Can be a very useful tool for administrators, system analysts and Hadoop designers. The teams also use it during various performance tests on the cluster.



Capacity Scheduler, the built in component within Resource Manager is a perfect tool to analyse the job queues with processing load and utilization % shown in real time for all running jobs on cluster. Capacity scheduler can be used to visualize the remaining processing capacity within the system at peak hours, investigate the priority queues, and analyse the running processes in depth. The display can be used to see the status of Available and Remaining Cores within the cluster, RAM/CPU utilization per node, per job and per queue. This tool is the perfect choice while performing upgrade tasks on the cluster.

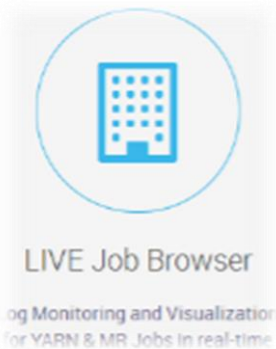


Workflow designer, powered by Hue and built over the top of Oozie is the most powerful and the most versatile tool in the stack. It allows users to create almost any workflow in a graphical user-interface, keep track of the workflow status, analyse logs in real time and schedule the workflows in an advanced job coordinator. Assisted with bundled feature of Oozie, it allows to define the groups or batches of workflows, their priority and their organization. We believe any solution can be integrated with Workflow designer thus allowing its users to visualize the ingestion, administration and mediation functions in a modern GUI.

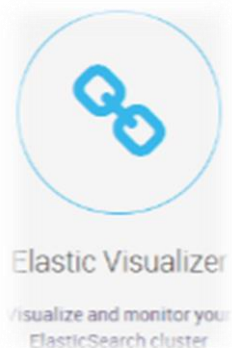


The Visualization book, powered by Hue, is a pre-defined visualization dashboard editor. It basically allows users to create and define their own dashboards in no time.

The Visualization Book or the Notebook can be integrated with some of the very advanced visualization tools in the market and at the same time can act as a stand-alone product. The ease of integration of queries into dashboards makes it an ideal tool for real time analysis, report generation and data mining in absolutely no time!

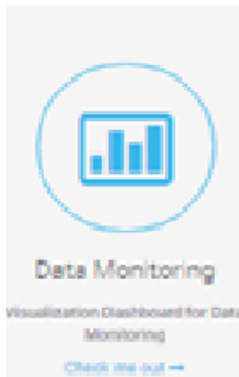


Live Job Browser, the tool provides an advanced interface for Resource Manager Visualization. Based on the concept of classic RM UI, the Job Browser can be used to analyse the jobs status in real time, as well as analyse the logs and counters of each job in a very advanced interface. Search filters can be applied on various criteria, thereby allowing administrators and system designers to have a complete insight of their system running on Hadoop cluster. Live Job Browser is a detachable component within Hue, and can be integrated very easily with multiple clusters.



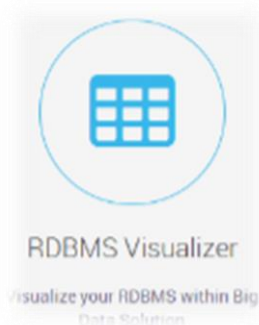
A web administration tool for Elasticsearch.

Offers a minimalistic GUI for easy performing of common management tasks of the Elasticsearch cluster. Enables features like cluster storage management, and querying for specific data retrieved by Logstash. This data includes data from Hadoop logs and Tweets from Twitter-API in JSON format.

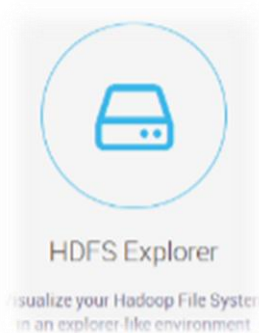


A visualizer / dashboard tool for the Elastic stack

Kibana offers graph and dashboard creation for visualizing the data parsed from logs created by individual components of the platform. One example is to find ingestion problems quickly without the need to go through logs manually.

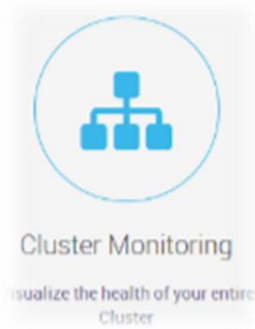


RDBMS visualizer, is a tool that has been integrated into Hue and can be used to analyse and access traditional databases like PostgreSQL, MySQL and oracle. The RDBMS visualizer enables us to visualize the ancestor databases from our Big Data Platform. It has also been used to deploy metastores for various subsystems of Hadoop. Metastores for subsystems are a key recommendation of Hadoop eco system. The metastores ensure that records are locked and shared amongst the resources in the most effective and efficient manner. This section can also be used to analyse source databases of HAIL on PostgreSQL servers.



HDFS explorer, is an advanced HDFS browser that allows its users to explore the HDFS and its directories like any of the most advanced file explorers. It can be used to upload, download or modify data on the HDFS. It can also be used to control the permissions on various directories in the file structure of HDFS.

HDFS file explorer is the number one choice of the developers who are analysing in depth the data that is residing onto their Hadoop cluster's HDFS. HDFS cannot be accessed by users without using Hadoop's URI, and therefore this tool eliminates the need of a file explorer for Hadoop.



Zabbix, a solution for monitoring the health of the cluster nodes

Zabbix provides graphs and alerts for monitoring the status and use of physical resources on each node. Different trigger values can be set, for example to give e-mail alerts to the administrators from disk space running low, CPU usage too high or any similar event.



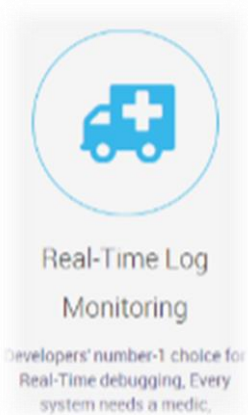
CloudAsset Process Manager, allows its users to run and stop processes on entire Hadoop cluster from the same window. It's a very advanced interface, built for cluster environments, thereby to give control to its users over every process within the cluster. The Process Flow Manager also supports advanced debugging messages, alerts, and Hadoop cluster control in almost no time.

It offers a great interface to perform a general health check on the process of Hadoop cluster. Every service within our ecosystem will be controlled by this Process Flow Manager.

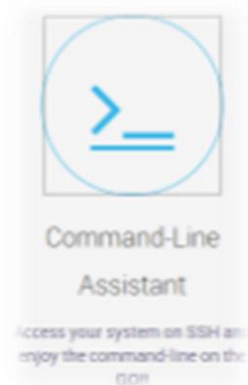


Data Quality Management is a tool for monitoring and analysing data from our various tables to detect sensor errors like Flat Values, Missing Gaps, Out of range values, Outlier, Inhomogeneity and Missing Patterns.

Using some of the cutting-edge technologies that are R, Spark and hdfs for analytics, and Shiny for visualization. Its user friendly graphical user interface helps intuitively observing the error values and help them to filter data and dig down into the issue and work out for its solution.



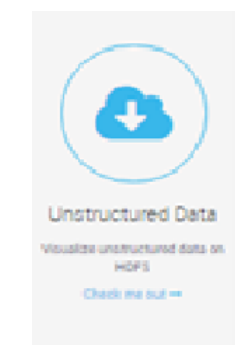
Another lightweight cloud application that offers real-time monitoring functionalities for Hadoop developers. Also allows filtering and searching from the logs in real-time. Can track logs from multiple nodes (Master and slaves alike) and updates to logs are shown as it happens. Also can help visualize logs from multiple services on the cloud simultaneously. This is a number one choice for analyzing multiple services or systems on cluster during debugging activities. Can also be used very effectively in the performance tuning of Hadoop cluster.



Command-Line Assistant, a web command-line SSH tool for accessing the nodes. Useful for doing any kind of maintenance that requires the command-line, without the need to install a separate ssh tool. The developers can now access their hadoop cluster from laptops, mobiles, workstations or any platform supported with a web browser. Thereby allowing administrators and system devops to work on the system from any location without hassle. The interface can be used to interact with the core of shell, hadoop and any service on the cluster.



Twitter Monitoring is constantly reading online tweets for any keywords, previously suggested by designers of process, and then it forwards those tweets to our big data platform. Those tweets are further segmented and filtered and made available as key statistical measure and key indicators on which analysis can be performed using our platform's capabilities. The solutions is highly flexible and can be easily integrated with the workflow designer.



The unstructured data interface allows to access unstructured data directories on HDFS on a visual interface. Files can be copied, viewed, modified and analyzed in a general file explorer like application or structure. The data is synced with HAIL's original unstructured data node, and it keeps populating data from that node in almost real time. The unstructured data can be further analyzed in the future solutions.