

# Real-time Big Data Analytics in Health Care Using Tools From IBM

The goal of this article is to provide the readers a basic technical understanding for big data applications in the health care system as well as provide information to develop or integrate various technologies with definite possible positive effects as return.

This article, **Real-time Big Data Analytics in Health Care Using Tools From IBM**, does not cover the other areas of application of Big Data Analysis, like predicting trend of communicable diseases such as flu, influenza or genetic analysis. There are already known existing ready to solutions.

Central aspect here is the objective presentation of the possible applications and, in particular, the opportunities that are currently used and in near future paths which has opened up. In order to illustrate the theoretical concepts and to illustrate the practical usage, selected application scenarios are described which could benefit from Big Data technologies in the future. In this context, IBM and IBM Big Data Analysis platform is chosen as tool as they :

- 1 Uses Open Source/Free Software softwares mainly

which are developed & maintained by Apache Foundation

- 2 Has free to test developmental environment ([we demonstrated usage before with Hadoop, Pig](#))
- 3 Has pay as you go Bluemix cloud platform
- 4 Has enterprise grade alternative to Bluemix cloud platform for the larger players
- 5 Has existing works in healthcare data analytics segment

## Table of Contents

- [1 Introduction](#)
- [2 Data Sources](#)
- [3 Challenges](#)
- [4 Alternatives to IBM Bluemix](#)

# Real-time Big Data Analytics in Health : Data Sources

Data sources can be stationary data sources in which, both the device and the person under question are not moving. These data sources include as example, patient monitors, and devices which measure air pressure, temperature, humidity to sleep trackers. Mobile data sources are from the wider range of devices such as wearable computers and smartphones. Sometimes, this segment is erroneously identified as “Medical Device

Data". Data can be generated via Image Processing or Signal Processing. In this context, data is generated via signal processing.

Wearables are small portable computers which should allow new communication capabilities, complementary to the smartphones and existing devices. Wearables are becoming increasingly important in our society's day to day life as well as in healthcare industry. The areas can easily be guessed – streaming real-time data usually captured by the devices like patient monitors which are designed to collect data like cardiac rhythm for analysis from sports field for optimization of the training performance to the field of ambulatory or inpatient medicine as an aid for diagnosis of heart diseases such as, for example, hypertension, heart attacks or cardiac insufficiencies, or in research and development, for example artificial pacemakers. Other similar fields are real-time data collected by blood glucose monitors such as glucometers, sleep evaluation tools etc. For medical studies wearables are becoming more and more interesting and opening up completely new possibilities. The collected and prepared data are used to predict the future decision. Predictive Analytics describes a concept to make predictions about a situation. In particular, intention is to correct decisions and to avoid false decisions, such is helpful in many areas for future planning to better estimate how things are likely to become. The technology and technical basis for the predictions are the evaluations of existing data using statistical models and empirical methods to predict events which could be expected with a certain probability. By means of statistics, the available data are

first processed, there are two types:

- 1 Descriptive statistics:** Structuring existing data and determining key figures such as average, variance or extreme values for better assessment of the data.
- 2 Exploratory statistics:** Systematic analysis and production of causal connections from the available data.

Also, analogous to webserver log data, collected real-time data in healthcare can be stored and obviously analyzed later.

## Challenges To Big Data Analytics in Health Care

Privacy and healthcare standards always a big hurdle for this segment. Healthcare industry always generate large amounts by record keeping, compliance & regulatory requirements, and patient care. Most of these data can easily be stored in hard copy form, although the current trend is toward rapid digitization. Oddly, the real-time monitoring data by modern devices such as in-OT, in-ICU streaming data against blood pressure readings to EKGs are mostly ignored in fields such as trauma patient monitoring, operating room monitors, anaesthesia, operating theatres possibly due to the nature of being less known to the non-healthcare individuals.

Currently available medical devices widely vary in the context of IEEE standards related to Healthcare and Electronics. Also, many of them lack the initiative to

embrace usage of open source hardware technologies. Some of these embedded devices has no way of data retrieval in any format. Also a remote possibility of collective social phobia associated with the phrase Big Data Analytics in healthcare may be liable for such lesser penetration. Also, many healthcare providers are not aware of the possibilities of usage of the already available tools for Big Data Analytics in Health Care and Real-time Big Data Analytics although, the nature of data source does not hugely vary compared to a webserver and later is widely known to many type of users.

Embedded devices like patient monitors in current usage is largely dominated by the hardware industry with specialised microprocessors like from Intel for healthcare. However, several open source hardware already demonstrated by the developers and makers on standard embedded project sharing websites with common computing hardware, open source single board computers. At present, streaming data from the proprietary devices involves ownership of higher models. We can brief up the challenges in to :

- 1 Privacy and data security issues
- 2 Issues with healthcare standards
- 3 Lack of experts in cross-domain knowledge
- 4 Lesser penetration of open source hardware technologies and monopoly of certain hardware companies
- 5 Ignorance

## Current Solutions

As can be revealed via web search, the current offered solutions by vendors like Bluemix less explored by the developers. Streaming analytics on Bluemix actually exists to capture data and enable application of other common tools compatible with Open Source (H/W and S/W) development :

- 1 [https://www.ibm.com/developerworks/websphere/library/techarticles/1603\\_chowdhury-bluemix-trs/1603\\_chowdhury.html](https://www.ibm.com/developerworks/websphere/library/techarticles/1603_chowdhury-bluemix-trs/1603_chowdhury.html)
- 2 <https://developer.ibm.com/streamsdev/2016/12/09/introducing-streams-healthcare-analytics-platform/>
- 3 <https://www.siam.org/meetings/sdm13/sun.pdf>
- 4 <https://assets.sourcemia.com/31/a6/cb1b019c4d6cb338fab539eea360/ims14428usen.pdf>

As for the common proprietary patient monitor models like from LG or Siemens Healthcare which are used in ICU, hybrid OT, ICU – the models which has standard ports for networking options/remote monitoring, streaming data can be retrieved from them. That topic of integration with cloud analytics is model specific matter and really off-topic to the this discussion. IBM Healthcare actually a separate dedicated division. There are lot of services from IBM which actually in current use. We have attached a PDF below for that part :

- [IBM Big Data Healthcare](#)

## Alternatives to Big Data Solutions From IBM Bluemix

In any case, one may want to install all the softwares on own on-premise servers which are actually hosted, managed by Bluemix for streaming data ingestion. The methodology essentially is streaming data ingestion with the Kafka for real-time analytics. Batch processing is for processing data gathered over a period when fast response time is not critical. Streaming data ingestion with the Kafka is faster. Apache Drill enables self-service data exploration on big data with a schema-free SQL query engine. Most of our existing guides to install the required components to build on-premise own real-time big data analytics in context of healthcare is possible but a costly alternative. At present, most of the standard big data analytics vendors offers managed Free Software/Open Source based solutions with minimal proprietary components.