The health data ingestion stack

Ali Anil Sinaci, Gökce Banu Laleci Erturkmen, and Tuncay NAMLI, *SRDC Software Research, Development and ConsultationLtd, Ankara, Turkey*

Introduction In order to address the challenges introduced by the increasing prevalence of chronic diseases with the aging society, EHRs are widely being adopted, and varieties of health-relevant parameters are captured via wearable devices and smartphone apps. There is also a growing movement called 'Quantified Self' to capture person's state and daily behaviours to establish a basis for preventive medicine interventions. The amount of healthcare and wellbeing data collected becomes unmanageable through conventional IT systems. This creates a big opportunity of big data analytics applications for healthcare data, where the requirements of high volume, velocity (real-time data from medical sensors), and variety (heterogeneity) of healthcare data sources can be addressed adequately.

Method In this study, we introduce the underlying architecture of the health data analytics platform that has been designed as a highly scalable ingestion stack respecting to the principles of Big Data Lambda architecture. The stack starts with the Inbound Adaptors acting as an interface to various health data sources such as EHRs, medical/tracking devices and mobile/Web/IoT applications. Several adaptors have been implemented for standards based communication: (i) HL7 CCD based clinical documents, (ii) ISO/IEEE 11073 compliant medical devices and (iii) Bluetooth LE enabled tracking devices. For device integration, an Android application has been implemented to collect data from medical devices and trackers, which also captures the measurements from the sensors on the phone and from Google Fit. According to the nature of the analytics services, data received through the Inbound Adaptors follow either the batch layer or the speed layer of the ingestion stack. If real-time analytics is required, the spark streaming based speed layer processes the data. Otherwise, data is processed by the batch layer through Apache Spark using Cassandra for data extraction, mediation, medical terminology mapping, summarization and complex rule processing. The results of the ingestion processing are received by the outbound adaptors to be exploited by the consumers such as data visualization services or machine learning algorithms.

Results ISO/IEEE 11073 compliant blood pressure and blood glucose measurement devices, a wristband that provides skin temperature, heart rate and measures for movement through Bluetooth LE and an Android application collecting data from phone sensors and Google Fit have been implemented and integrated to the ingestion stack, and tested with *in vivo* data collection mechanisms. In the next phases, the project will focus on EHR integration.

Discussion Collected data will be used through a use case where the outbound results will be smart, adaptive and personalized interventions for diabetic patients so that the interventions resulting from the extensive analytics can lead to behavioural changes in those patients for reducing risks and improving quality of life.

Conclusion The health data ingestion stack is being developed in the scope of ITEA3 Medolution Project. With this stack, the aim is to be able to process data from various health data sources, perform predictive and prescriptive analytics for both medical professionals and patients themselves and lead to better-informed decisions and interventions with the patients.