Wearable Computing

Ioan RAICU
Illinois Institute of Technology
iraicu@cs.iit.edu

Matthieu MUNHOVEN
ESIGELEC | Graduate School of Engineering
m.munhoven.12@groupe-esigelec.fr

Armand LEOPOLD
ESIGELEC | Graduate School of Engineering
a.leopold.12@groupe-esigelec.fr

September 2, 2016

Abstract

The number of wearable devices (smartphones, smartwatches ...) is increasing along the years. All of these devices are becoming more powerful, and more efficient sensors are added to them. Unfortunately, the enormous amount of data produced by these devices is only used for local applications, games or fitness for example. Nowadays, a lot of health laboratories and medical centers need this kind of data and are ready to pay expensive surveys to collect them. Collecting these data directly from wearable devices with the user agreement would be easier and less expensive for health research. This amount of data may also be useful for any commercial use, smart homes for example. This project focuses on developing an end to end wearable device data management architecture, including app, database, and data processing algorithms for research purposes.

I. Background information

There are two major OS (operating system) for smartphones and three for smartwatches: iOS (iPhone), Android (Android Phones), watchOS (iWatch), Android Wear (majority of the smartwatches) and Tizen (Samsung Gears). On these five OS, there are three different programming languages: iOS and watchOS are using Objective-C or Swift, Android and Android Wear are using Java, and Tizen use C/C++. We can see, after looking at the charts 1, the iWatch is twice more sold than the others smartwatches. Also there are five time more Android smartphones than iOS ones 2 . The problem for Android smartphones is that there are a lot of different smartphones from different manufacturer with different sensors. Some of them have light sensors and others don't. Or maybe the sensors have different sampling frequency or unit of measurement.

Furthermore it is not the only difficulty, the

Figure 1: *Smartwatches Operating System Marketshare Source*

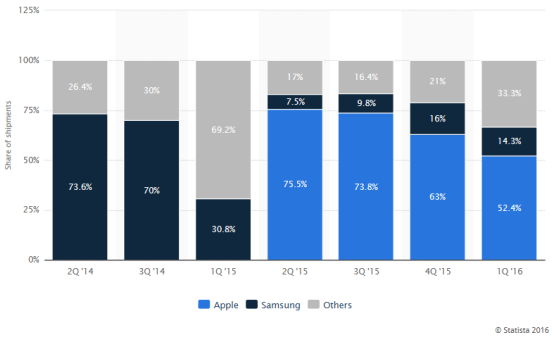
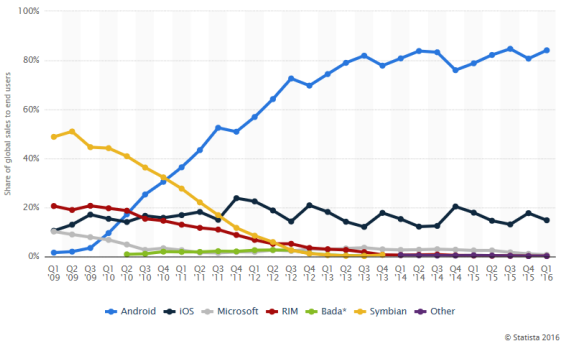


Figure 2: *Smartphones Operating System Marketshare Source*



way of collecting data is different on each OS. With iOS and watchOS, the use of the HealthKit API is required to extract the data, however these data have their own pattern. In this case, it's impossible to access directly to the sensors value. Unlike Tizen, Android and Android Wear, where it's possible to access to the sensors value.

As a consequence, finding the best database management system (DBMS) for the application is essential for working with a big amount of data, for example accelerometers produce around one million lines in one hour. Moreover, the amount of data will increase rapidly with a lot of users.

There is two different types of database system, RDBMS (Relational Database Management System) and DDBMS (Distributed Database Management System).

A relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model. RDBMSs are a common choice for the storage of information in new databases used for financial records, manufacturing and logistical information, personnel data, and other applications. Relational databases have often replaced legacy hierarchical databases and network databases because they are easier to understand and use.

A distributed database is a database in which storage devices are not all attached to a common processing unit such as the CPU, and which is controlled by a distributed database management system (together sometimes called a distributed database system). It may be stored in multiple computers, located in the same physical location; or may be dispersed over a network of interconnected computers.

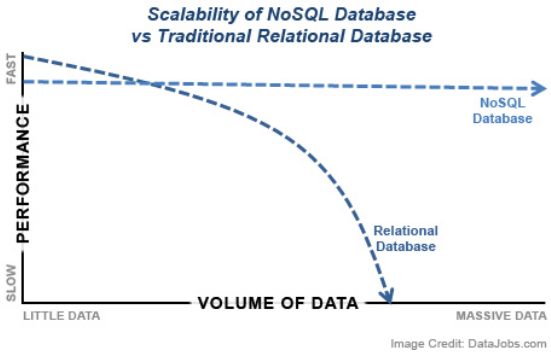
II. Problem statement

The first problem is to ensure a low energy consumption on the different devices while doing measurements with the sensors. We have to take enough samples without disturbing the everyday user experience of the devices. Managing and standardizing big dataets are some major factors that should be concerned in designing the database architecture. Moreover, it is necessary to have a constant data processing workload to prevent overload on the database due to real-time input. Finally, labelling the different activities based on raw data without nearly explicit supervised learning is a big problem who will need artificial intelligence and deep-learning algorithms.

III. Related work

There are already a lot of applications specialized in health and fitness management as most of the actual wearables are made for them. Nevertheless, data and algorithms are made for individual use only and can't be merged all together to form a more reliable and consistent dataset of multiple data source. AlgoSnap recently released such dataset [1] (February 2016) which involve a variety of wearable devices (smartwatches , smartphones, etc ...) and multiple people doing different activities. This database was used for our preliminary work for designing the database architecture. DDBMS is a consistent and well established database architecture part of the NoSQL paradigm [2]. One great example is Cassandra [3] who's reliable due to it's developement by Facebook wich became over the year the best compagny in the field of Big-Data; Or BigTable by Google [4]. AI and deep learning algorithms are often used for improving accuracy labelling

Figure 3: *Scalability of NoSQL Database Source*



to a large amount of data, though there were used for many years, no real application were made using them. Labelling user data activities with a very high rate accuracy will ensure reliable ground data pattern detection for future research project. [8] [7] [6] [5]

IV. Proposed solution

The first step of the work would be to develop an application for each device we want to integrate in our wearable device environment. Learn the different languages related to each operating system and standardize the data from the different sensors. We have decided to use the Android platform because Apple's platform need a specific developing environment, which we don't have. We will develop an application on an Android smartphone which records some sensors' data and input them in our database.

The second step would be to determine the best DBMS and configure it. The greatest challenge of this step would be to benchmark the performances of the workflow between the application and the database to ensure that the architecture is scalable and optimized for a high loading rate 3.

We have chosen Apache Cassandra thanks to her scalable capability and great robustness against failure with which we can manage a big workload, it's also a fast optimized DBMS for big amount of data. We also have chosen to associate Apache Spark with Cassandra as we want to do some work with these data, it will be more appropriate and faster.

The final step, would be to labelize the user activities using an algorithm (stand, sit, walking, running, ...) based on his wearable sensor records.

V. Evaluation

We are planning to evaluate our database arcitecture of our project with a distributed database on a 10 machine cluster on Amazon AWS. We are going to evaluate the overall input time of one day dataset records and extrapolate to a large population to determine if the system architecture sustains it.

We will also evaluate the accuracy of user activities detection with our labelling algorithm. All along the evaluation process we will record data about energy consumption and compare it to a normal use.

VI. Time line

The schedule for the previous and next weeks of the project is shown in 1.

VII. Deliverables

The final results of this project are a fully working end to end wearable data collection system with android app, server with Spark/Cassandra, and labelling algorithm.

VIII. Results

For our Benchmark we decided to compare the input time for different file size on a single node local cluster. So we choose file of 1000 4, 10000 5, 100000 6, 1000000 7 rows. As we can see with these graphs, the file size doesn't matter for the input time because it depends on the input time rate for one row which is a constant. This is why the final time is still the same no matter the file size

Table 1: *Work Timeline and deadlines*

Goals	Week 1	2	3	4	5	6	7	8	9	10	11	12
Study State-of-the-art Wearable devices	X	X										
Study State-of-the-art Sensors		X	X									
Study State-of-the-art Database Systems				X	X							
Android application								X				
Cassandra implementation						X	X	X				
Spark implementation							X	X				
AWS deployement									X			
Labeling algorithm									X			
Evaluation										X	X	
Final Report											X	X

Figure 4

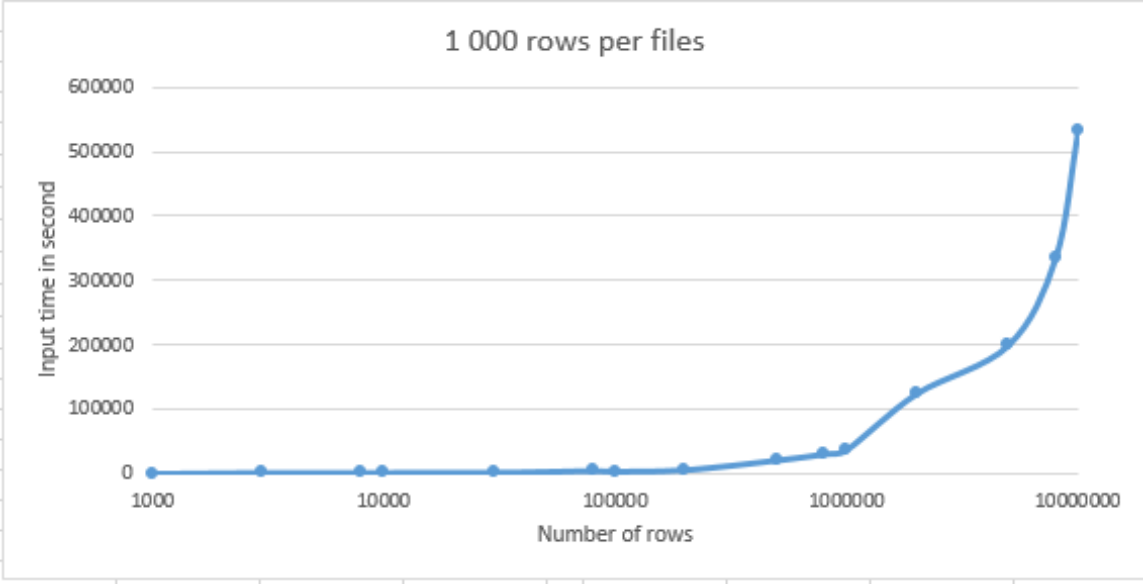


Figure 5

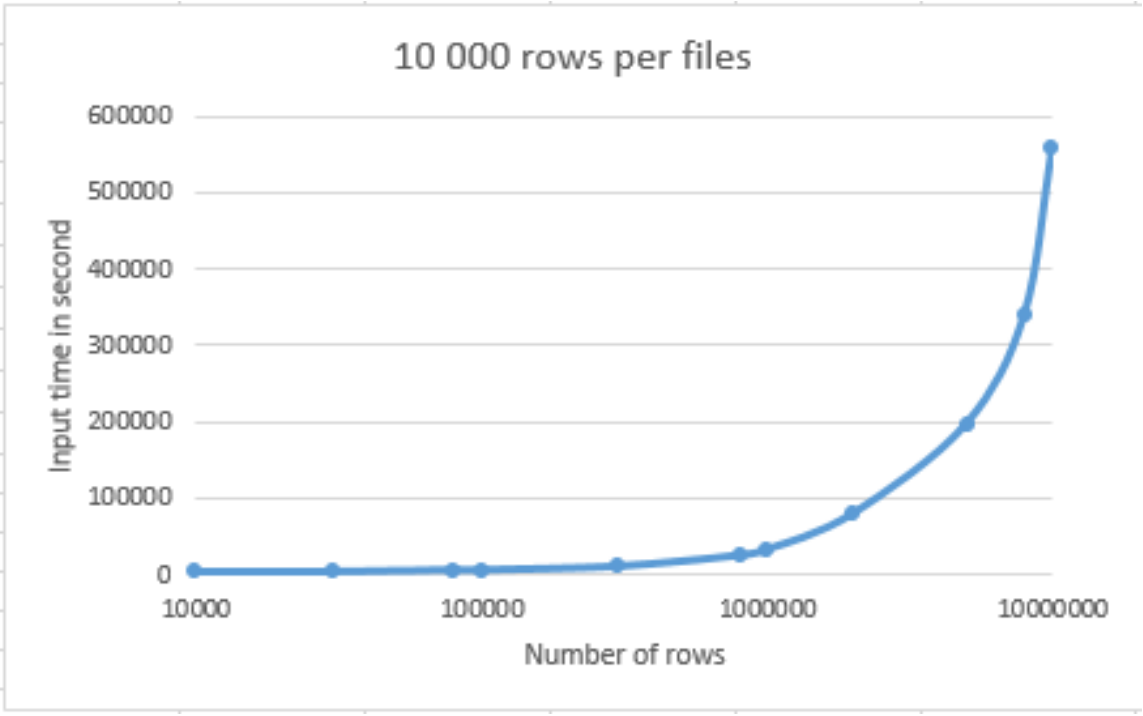


Figure 6

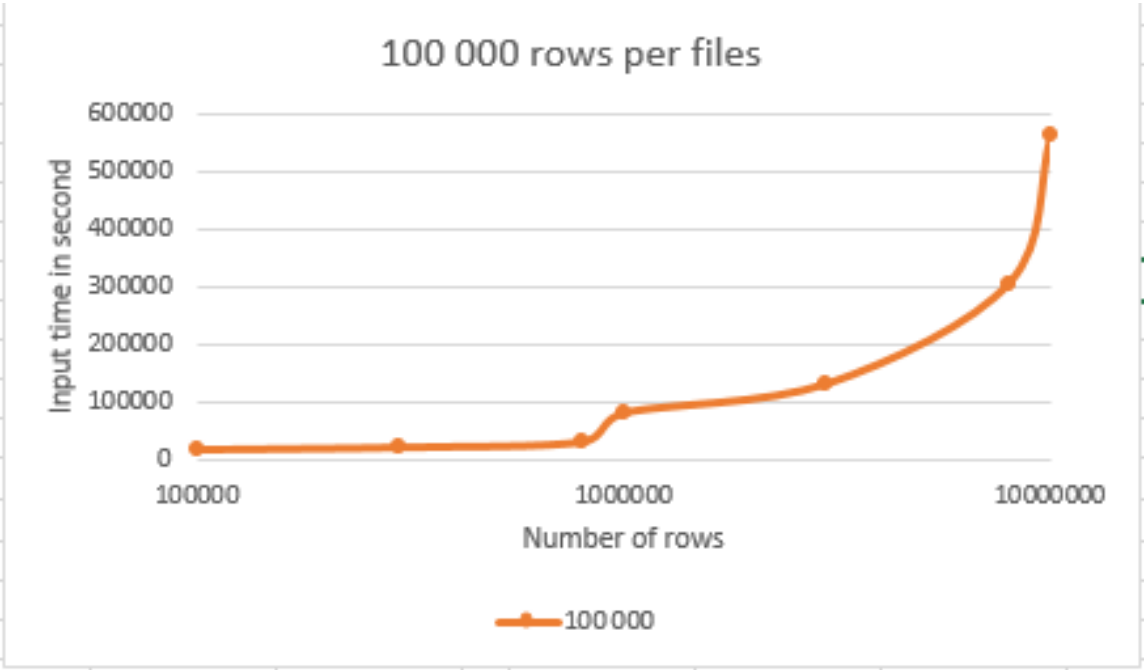
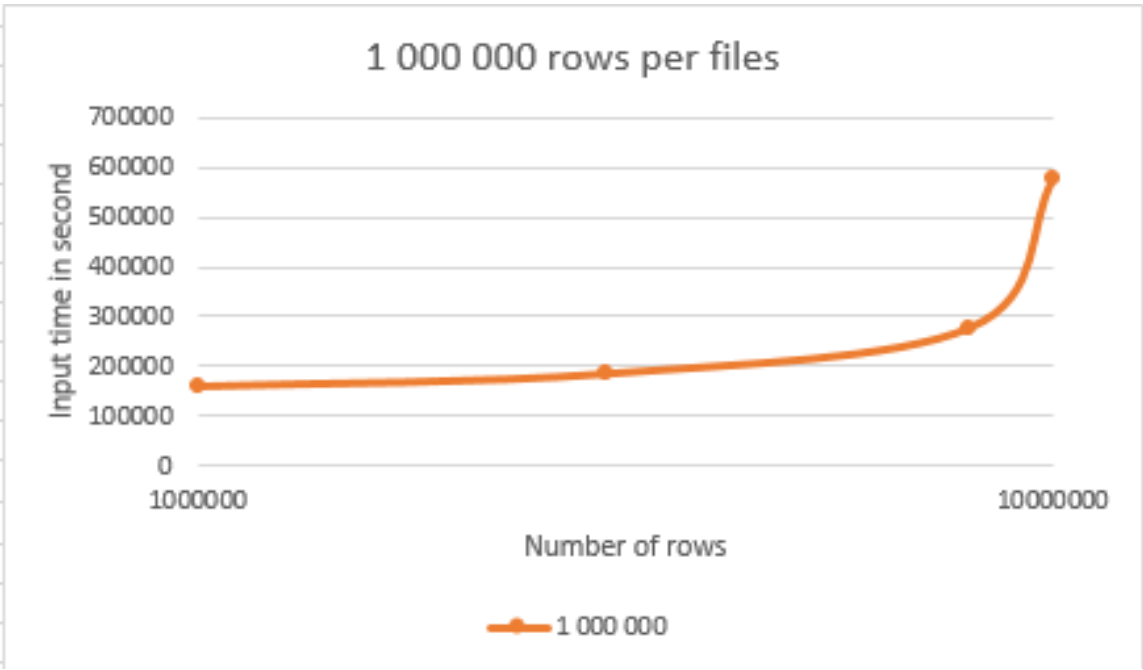


Figure 7



IX. Future Work

Deep learning AI for labelling accuracy, integrate more device type, open source the system architechture, benchmark on a large multi-node cluster.

X. Work related GitHub repository

Wearable-Computing-Labelling-Data-Algorithm

This is the labelling algorithm project for labelling user activities based on his sensors data collected previously related to the 2016 summer "Wearable computing" research program at iit.edu

Wearable-Computing-Input-Data-Bench

This is the spark/cassandra project to input the sensors data from smartphones related to the 2016 summer "Wearable computing" research program at iit.edu

Wearable-Computing-Android-App

This is the android 6.0 project for collecting the sensors data from smartphone related to the 2016 summer "Wearable computing" research program at iit.edu

XI. Conclusions

The quantity of unused data produced by wearable computing is a good opportunity for building a big database for health companies. Developing an environment to collect data and processing them with a user activities detection algorithm add a true value for possible work application.

References

[1] CrowdSignals.io, A crowdfunding campaign to fund the largest ethically collected set of high-quality, labeled mobile and sensor data for use by the research community. CrowSignals website - 2015

[2] Survey on NoSQL Database, Jing Han, Haihong E, Guan Le ,Jian Du. Survey on NoSQL Database - 2011

[3] Cassandra - A Decentralized Structured Storage System. Avinash Lakshman , Prashant Malik. Cassandra - 2010

[4] Cloud Bigtable - A high performance NoSQL database service for large analytical and operational workloads BigTable

[5] Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. Machine Learning - February 2010

[6] Human Activity Recognition from Accelerometer Data Using a Wearable Device Human activity recognition - 2011

[7] Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields gpsTraces - September 2013

[8] Robust Active Learning Using Crowdsourced Annotations for Activity Recognition Acive Learning - 2011