Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

Project PEACH at UCLH: Student Projects in Healthcare Computing

Navin RAMACHANDRAN[a], Dean MOHAMEDALLY[b] and Paul TAYLOR[b1]

[a] *University College London Hospital, London*
[b] *University College London, London*

Abstract. A collaboration between clinicians at UCLH and the Dept of Computer Science at UCL is giving students of computer science the opportunity to undertake real healthcare computing projects as part of their education. This is enabling the creation of a significant research computing platform within the Trust, based on open source components and hosted in the cloud, while providing a large group of students with experience of the specific challenges of health IT.

Keywords. Capacity building, Informatics education. Open source

1. Introduction

A key challenge for the health informatics community is to train a workforce capable of delivering the potential of new technology. While attention has focused on initiatives to give clinicians training in informatics, we believe that we must also train computer scientists to have a strong understanding of the special requirements of healthcare: the additional responsibilities for patient safety and confidentiality that healthcare IT involves, and the complexity of the environment in which systems have to work.

To put this in a UK context, the NHS workforce now numbers around 1.3 million staff. A 2008 estimate (there isn't a more recent figure) suggested that informatics staff represent 3% of total and that vacancy rates were as high as 16%. [1] Although written evidence is hard to find, other employers tell us that they too struggle to recruit technical staff with specialist knowledge of healthcare.

Five years ago the Department of Computer Science at UCL began to move away from the use of invented or artificial problems in teaching. This initiative led to the development of the Industry Exchange Network (http://ixnet.org.uk/) to train students on real-world problems through term-based projects with real clients. This is now the largest network of client engagement projects for term-based CS students in the country. In 2016 the authors of this paper began what we hope will be a long-running initiative to host a large number of these projects within a hospital. The students are supervised by staff from the university's computer science department, with input from a health informatics specialist. The projects are proposed by a clinical team who act as

[1] Corresponding author, Paul Taylor, Institute of Health Informatics, UCL, 222 Euston Rd, London, NW1 2DA, United Kingdom; E-mail: p.taylor@ucl.ac.uk

clients for the duration of the projects, have regular meetings with the students, and provide guidance and advice throughout.

The projects, collectively known as PEACH, are all based on the use of open source components and, it is hoped, will provide a proof-of-concept for open source software development in the NHS as well as a realistic exposure to the special demands of health informatics.

2. Methods and Participants

Twenty UCL MSc students, five from the specialist Software Systems Engineering MSc and fifteen from the generalist MSc in computer science (designed for students who did not take computer science as a first degree) volunteered to work on a suite of summer projects based at UCLH.

The projects were proposed by the clinical clients: Dr Navin Ramachandran (the first author of this paper and a consultant radiologist) and Dr. Wai Keong Wong (consultant haematologist at UCLH). Projects were identified and students assigned roles within each project. Students worked in five teams and each project was assigned an architect from the students in the SSE MSc.

All students had a supervisor from UCL Computer Science who met with them at the outset (in June 2016), at the mid-project review, at the final assessment (in Sept 2016) and as required in between. The clients met with the students weekly, or more often if required. All students completed online training in information governance and data security for healthcare researchers.

Additional support was supplied by Microsoft UK who attended early meetings, provided a mentor for students and donated Azure credits to support the applications.

Test data for the projects was supplied by UCLH and consisted of 5 years' worth of anonymized data from the hospital RIS system and a set of test data from the Intensive Care Unit provided through the Health Informatics Collaborative.

2.1. Core Platform

A core platform was engineered to support the deployment of multiple healthcare applications. The basis of the platform was an integration and analytics engine composed of the following elements. Apache NiFi, a dataflow tool, was used to import the data from the files supplied by the Trust. Apache Kafka was used both as a messaging hub and a source of truth (in a Kappa architecture).[2] This works on a publish -subscribe model, whereby sources publish data to Kafka and applications subscribe to the data feeds. Druid was used for online analytical processing and Apache Spark for machine learning analytics. The whole platform was deployed on DCOS (data centre OS), hosted on Microsoft Azure, which also provided the required role based security.

For novel healthcare data the client selected EtherCIS as the main datastore. [3] EtherCIS is an open source system, compliant with the openEHR information modeling specification.[4] This system provides an interoperable backend service to applications, allowing them to use RESTful APIs to persist and retrieve healthcare data.

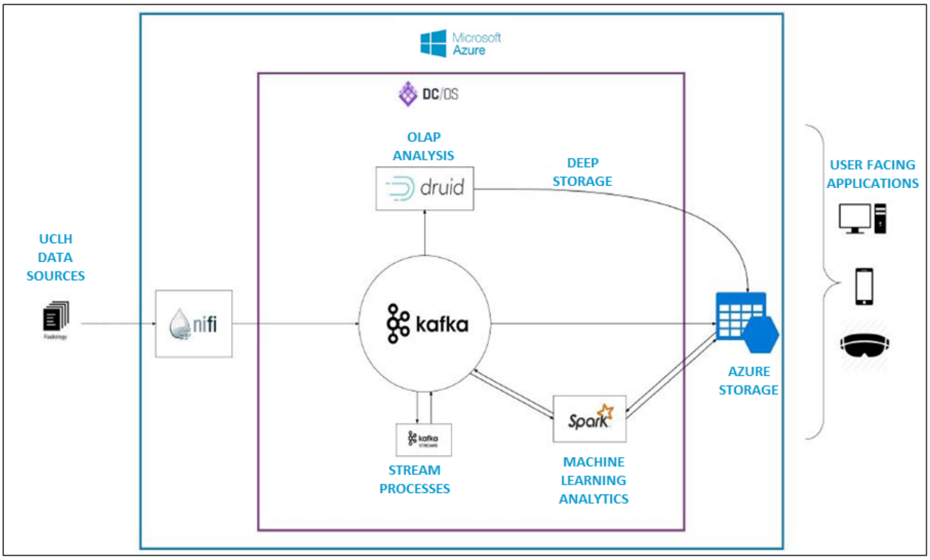

Figure 1. Architecture diagram of the PEACH data platform

2.2. Applications

- The "Multidisciplinary Team (MDT)" application is an innovative application to manage information on patients reviewed during cancer MDT meetings.
- The "Dashboards" application is a customizable reporting tool for multiple users, supporting interactive analysis on event data to provide real time visualization of key statistics, with radiology data as the first use case. Users have access to a variety of services based on roles.
- The "Chatbot" team developed an interactive conversational chatbot using natural language processing and machine learning to perform an electronic Holistic Needs Assessment (eHNA) for cancer patients, in order replace a very time-heavy process.
- The "Reality" application allows for surgeons to interact with 3D models of patients pre-operatively using virtual reality, in order to prepare for surgery. The first iteration of the project considered preoperative planning for renal cancer surgery.
- The "Clinical Trials" application provides a searchable online database of clinical trials in the region, for administrators, clinicians and patients.
- The "Visual Report" application is a novel approach to healthcare data recording, using tablet / pen input to record diagrams of prostate cancers on a template, with analysis of the diagram to provide automated risk scoring of prostate regions.

3. Results

The PEACH project consists of multiple projects, which were assessed separately. Performance of the core platform was measured by its processing time. In Table 1 we provide a sample measurement in the pre-processing layer based on 10,000 records of one of the csv data files (DATAEXAM file). Complexity of each stage of testing was measured by counting the number of processors involved at that stage. The metrics demonstrated that a CSV file could be pre-processed to singular JSON objects in less than 4 milliseconds per record in Stage 2. In Stage 3, the time required to finish a 10,000 record transfer to the message hub cluster was 2 hours and 47 minutes, or less than 1.001 second per record. These figures demonstrate that using a single process to deal with both batch data (large volume) and streaming data (low volume but realtime) in a Kappa architecture was a viable approach.

Table 1. Sample result of measurement in pre-processing layer

Stage	Input files	Input MB	Output files	Output MB	Time	Complexity
1	1 (CSV)	494.31	299 (CSV)	494.31	00:00:21.227	13
2	1 (CSV)	1.65	10,000 (JSON)	7.17	00:00:37.419	9
3	10,000 (JSON)	7.17	10,000 (key,value)	7.17	02:46:50.722	5
2-3	1 (CSV)	1.65	10,000 (key,value	7.17	00:00:50.467	12

The MDT team created a ‘Dockerised’ version of EtherCIS, a significant achievement. [5] However there were problems with the reliability of solutions for role-based access. The team was unable to demonstrate a working front-end.

The Clinical Trials application team was assisted by experts in clinical terminologies and was able to solve many of the backend challenges in representing trial inclusion and exclusion criteria. One problematic aspect of this project was that the team chose to implement the app using a different and suboptimal framework (sails.js). Future projects will use a single consistent architecture.

The Visual Report team created a working intuitive product which it is hoped could be published on the Windows App store with further development,

The radiology Dashboard team created a real-time batch pipeline using the core platform. Unfortunately the dashboard interface was only partially completed. A major problematic requirement was for a UK data-centre, in order to meet information governance rules - these were all in Technical Preview phase with limited functionality. Microsoft have since provided further assistance to resolve this issue as well as 20,000 USD Azure credits for the next phase of the work.

The Chatbot was less successful due to suboptimal integration between components developed by different team members. Future projects will be stored on a common GitLab repository, enforcing continuous integration, to mitigate this risk.

The Reality project worked well, allowing 3D anatomical reconstructions to be viewed on the HTC Vive, but did not have a fully-developed pipeline for producing the required 3D models. The virtual reality visualization also proved slightly disorientating for users. The next phase will use Hololens Augmented Reality (which should be less disorientating), and also develop a model construction pipeline and design language.

One student wrote of the experience of leading a PEACH project: “It was very useful especially for us because we haven't had any previous experience on patient data before PEACH. It was really challenging to keep it secure especially when the UCL-UCLH sharing agreement process demanded a long time/effort before the data could actually be securely transferred to our server. Our everyday communication with the doctors and healthcare specialists gave us a better understanding of how a real life problem looks like and the kind of influence our project (with its weak and strong points) could have in people's lives. For me, the aspects of security, efficiency and consistency of data across a real time infrastructure were really fascinating. It's like a dream for me to help people through technology and try to solve a real need.”

4. Conclusions

The following lessons were learned and are being applied in the next phase:

- The number of core technologies has been reduced to ensure interoperation and reusability. A team will build a reusable component library.
- Many students had not been sufficiently encouraged to think about teamwork. Some supervisors instructed students to concentrate on their own components and not on the reusability of the application. The guidance has been changed.
- There were a number of tensions between the supervisor’s and the client's requirements. For example the client might ask for a simpler interface where the supervisor may want the student to use demonstrate more complex skills. We will improve communication between supervisors and clients offset this.
- We have improved the guidance at the outset of the project so that students have a clearer statement of requirements and principles of continuous integration and continuous deployment are enforced.
- The cloud services are available from the start.
- We will generate a realistic synthetic dataset to use in testing.

Acknowledgements

Dr Wai Keong, Dr Jay Kola and Dr Ian McNicoll provided input and guidance to the student projects. Efthymia Kazakou and the other PEACH students contributed to the success of the work reported here.

References

[1] Tribal Consulting, “NHS Informatics Workforce Survey 2007/08,” 2008.
[2] J. Kreps, N. Narkhede, and J. Rao, “Kafka: a distributed messaging system for log processing,” in *Proceedings of the NetDB*, 2011.
[3] “EtherCIS by ethercis.” [Online]. Available: http://ethercis.github.io/. [Accessed: 04-Nov-2016].
[4] A. Moreno-Conde *et al.*, “Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis,” *J. Am. Med. Inform. Assoc.*, vol. 22, no. 4, pp. 925–934, Jul. 2015.
[5] D. Merkel, “Docker: Lightweight Linux Containers for Consistent Development and Deployment,” *Linux J*, vol. 2014, no. 239, Mar. 2014.