Cloud Bioinformatics in a private cloud deployment

Victor Chang[1, 2]

[1] School of Computing and Creative Technologies, Leeds Metropolitan University, Headingley, Leeds LS6 3QS

[2]. School of Electronics and Computer Science, University of Southampton, Southampton SO 17 1BJ, UK.

V.I.Chang@leedsmet.ac.uk

Abstract. This paper describes service portability for a private cloud deployment, including a detailed case study about Cloud Bioinformatics services developed as part of the Cloud Computing Adoption Framework (CCAF). Our Cloud Bioinformatics design and deployment is based on Storage Area Network (SAN) technologies, details of which include functionalities, technical implementation, architecture and user support. Bioinformatics applications are written on the SAN-based private cloud, which can simulate complex biological science and present them in a way that anyone without prior knowledge can understand. Several bioinformatics results are discussed, particularly brain segmentation, which demonstrates different parts of the brain simulated by the private cloud. In addition, benefits of CCAF are illustrated using several bioinformatics examples such as tumour modelling, brain imaging, insulin molecules and simulations for medical training. Our Cloud Bioinformatics solution offers cost reduction, time-saving and user friendliness.

1   Introduction

Healthcare informatics has played a strategic role in the National Health Service (NHS) and has been influential to the way in the IT project development for different NHS Trusts. The ICT initiatives include Cloud Computing, which has investigations to understand how to process with Cloud adoption and the capacity to maximise the added value as a result of Cloud adoption. Cloud Computing offers a variety of benefits including cost-saving, agility, efficiency, resource consolidation, business opportunities and Green IT (Chang et al., 2010 a; 2010 b; 2011 a; 2011 b; 2011 c; 2012 a; 2012 b; 2012 c; 2013 a; 2013 b; Kangermann et al., 2011).

As more organisations adopt Cloud, the need for a standard, or a framework to manage both operation management and IT services is emerging. This framework needs to provide the structure necessary to ensure any Cloud implementation meets the business needs of industry and academia and include recommendations of best practices which can be adapted for different domains and platforms. Our framework is called the Cloud Computing Adoption Framework (CCAF). It helps organisations to achieve good Cloud design, implementation and services (Chang et al., 2011 a; 2011 b; 2011 c; 2011 d; 2011 e; 2012 a; 2012 b; 2012 c; 2013 a; 2013 b; Chang and Wills, 2013; Chang, 2013 a; 2013 b; 2013 c). CCAF may be used from service strategy to design, development, test and user support stages. The CCAF seeks to address two problems in particular:

- Calculating risk and return analysis of a large computer system adoption such as Cloud adoption systematically and coherently.

- Risk mitigation to migration of the Cloud

1.1 Service portability for Cloud deployment

This paper focuses on service portability which is the term we use to describe a recommended approach to Cloud adoption, that plays an important role in having a smooth transition to the Cloud environment. Service portability also influences the design and implementation of healthcare bioinformatics services. Beaty et al. (2009) and Chang et al. (2011 a; 2012 c; 2013) identify portability as an adoption challenge for organisational Cloud adoption. Although it is domain specific as there are different requirements for portability in each domain, communication between different types of clouds supplied by different vendors can be difficult to implement. Often work-arounds are needed which entail writing additional layers of APIs, or an interface or portal (Beaty et al., 2009; Armbrust et al., 2009).

Service portability (portability in short) is illustrated using examples from Cloud bioinformatics projects in the Healthcare industry where portability is influential in migrating the existing platforms and applications to the Cloud and later developing new applications and services. Bioinformatics is provided using in-house private clouds, initially to provide a working IaaS infrastructure for medical databases, images and analysis in a secure and collaborative environment. These Cloud projects have been successfully delivered and provide a

high level of user satisfaction and were followed up with further work to upgrade from IaaS to PaaS, which allows greater benefits, including better efficiency and better management of resources.

1.2   Two stages in the development of the private cloud

There are two phases in the development of the private cloud. The first phase is the design and deployment of the architecture to consolidate infrastructure, platform and resources. The objective is to provide a consolidated infrastructure before beginning the software development and offer of application services. The second phase is the application development built on top of the consolidated architecture. These applications offer both Platform as a Service (PaaS) and Software as a Service (SaaS). PaaS allows the developers to develop the code in the Cloud repository, which is a central platform for the developers to implement and test their prototypes. The internal cloud is used as a knowledge-based sharing resource so that any team members can be informed with the latest updates and lesson learned as a result of service delivery or troubleshooting experience. New knowledge and repository for best practices can be kept up-to-date. SaaS is the service offered to the internal users. From users' points of view, they do not need to know the complexity behind the scene but they can use the service at any way and any time. These SaaS services are easy to use and allow users to interact with simulations and obtain their experimental results, even without themselves involved with experiments.

The structure of this paper is as follows. Section 2 describes the first phase of the Cloud deployment and its architecture and Section 3 presents the bioinformatics services on offer. Section 4 explains one specialist area of the bioinformatics project, brain segmentation and its demonstrations. Section 5 presents three topics for discussions and Section 6 sums up Conclusion and future work.

2   Phase 1: Healthcare Cloud Bioinformatics Architecture and User Support

Supported by NHS UK, Guy's and St Thomas NHS Trust (GSTT) and King's College London (KCL) have worked together on projects to implement Cloud bioinformatics and deliver it as a service. The initial effort

was directed to an evaluation of the technology and developed a proof of concept service. CCAF is instrumental and influential in the way Cloud Bioinformatics has been developed:

- Healthcare Cloud Bioinformatics is a PaaS system, and needed careful planning and a thorough implementation. This required integrated adoption of multiple vendors' solutions.

- Healthcare Cloud Bioinformatics is an area to experience rapid growth in user requirements and big data management. Therefore, it had to be easy to use, and able to cope with increasing demand.

- Healthcare Cloud Bioinformatics is a new concept and implementation in the Health domain where private and in-house Clouds have been designed and deployed. Maintenance of data protection and security is a challenge.

Better performance in Healthcare Cloud Bioinformatics than previous storage service is regarded as a benchmark and measurement for success by executives. Recommendations, strategy and support from CCAF provided useful good services. Healthcare Cloud Bioinformatics has used trials during its design and implementation to ensure it meets its requirement to provide a robust service.

There are many thousands of data about patients (medical records) and tumours (detailed descriptions and images, and its relations to the patients). Data growth is rapid and the data needs to be carefully used and protected. The work involves integrating software and cloud technologies from commercial vendors including Oracle, VMWare, EMC, Iomega and HP. This is to ensure a solid infrastructure and platform is available. Researchers also use third party applications to access, view and edit tumour images from trusted locations. Security is enforced in terms of data encryption, SSL and firewalls. The Health Cloud Bioinformatics services provide scientific visualisation and modelling of genes, proteins, DNA, tumour and brain images. Users are very supportive in this project and some of them use it daily.

2.1 The Architecture: A Storage Area Network made up of different clusters of Network Attached Storage (NAS)

The Architecture design chosen uses two concurrent platforms. The first is based on Network Attached Storage (NAS), and the second is based on the Storage Area Network (SAN). The NAS platform provides great usability and accessibility for users. Each NAS may be allocated to a research group and operate independently. Then all the NAS can be joined up to establish a SAN. NAS supports individual backups with manual and automated options. One option is similar to the Dropbox pattern of backup enabling users to copy their files onto their allocated disk space without difficulty providing a backup facility which is easy to use and user-friendly. Such a manual service allows users to backup their resources onto a selected destination and can offer both compressed and uncompressed versions of backup as well as data encryption to enforce security.

The Storage Area Network (SAN) is a dedicated and extremely reliable backup solution offering a highly robust and stable platform. SAN can consolidate an organisational backup platform and can improve capabilities and performance of Cloud Bioinformatics. SAN allows data to be kept safe and archived for a long period of time, and is a chosen technology. A SAN can be made up of different NAS, so that each NAS can focus on a particular function.

The design of SAN focuses on SCSI, which offers dual controllers and dual networking gigabyte channels. Each SAN server is built on RAID system. RAID 10 is a good choice since it can boost the performance like RAID 0 but also has mirroring capability like RAID1. A SAN can be built to have 12TB of disk space, and a group of SAN can form a solid cluster, or a dedicated Wide Area of Network. There are written and upgraded applications in each SAN to achieve the following functions:

- Performance improvement and monitoring: This allows tracking the overall and specific performance of the SAN cluster, and also enhances group or individual performance if necessary.

- Disk management: When a pool of SAN is established, it is important to know which hard disks in the SAN serve for which servers or which user groups.

- Advanced backup: Similar functionalities to those described in the NAS, such as automation, data recovery and quality of services, are available here. The difference is more sophisticated techniques and mechanisms (use of enterprise software is optional) are required.

Some applications mainly based on PHP, MySQL and Apache have been written, to allow researchers to access the digital repository containing tumours. Users can access their Cloud Bioinformatics via browsers from trusted offices, and they need not worry about complexity, and work as if on their familiar systems. This Healthcare PaaS is a demonstration of enterprise portability. In addition, several upgrades have taken place to ensure the standard of Cloud Bioinformatics and quality of services. One example is the use of SSL certificates and the enforced authentication and authorisation of every user to improve on security. There is an automated service to backup important resources.

2.2 Selections of Technology Solutions

Selections of Technology Solutions are essential for Cloud Bioinformatics development as presented in Table 1.

Table 1: Selections of Technology Solutions.

Technology selections	What is it used	Vendors involved	Focus or rationale	Benefits or impacts
Network Attached Storage (NAS)	To store data and perform automated and manual/personal backup.	Iomega/EMC Lacie Western Digital HP	They have a different focus and set up. HP is more robust but more time-consuming to configure. The rest is distributed between RAID 0, 1 and 5.	Each specific function is assigned with each NAS. There are 5 NAS at GSTT/KCL site and 3 at Data Centre, including 2 for Archiving. Deployment Architecture is shown in Figure 4.
Infrastructure (networking and hosting solution)	Collaborator and in-house	University of London Data Centre	Some services need a more secure and reliable place. University of London Data Centre offers 24/7 services with around 500 servers in place, and is ideal for hosting solu-	Amount of work is reduced for maintenance of the entire infrastructure. It stores crucial data and used for archiving, which backup historical data and backup the most important data automatically and

			tion.	periodically.
Backup applications	Third party and in-house	Open Source Oracle HP Vmware Symantec In-house development	There is a mixture of in-house development and third party solution. HP software is used for high availability and reliability. The rest is to support backup in between NAS. Vmware is used for virtual storage and backup.	Some applications are good in a particular service, and it is important to identify the most suitable application for particular services.
Virtualisation	Third party	VMware VSphere and Citrix	It consolidates IaaS and PaaS in private cloud deployment.	Resources can be virtualised and saves effort such as replication.
Security	Third party and in-house	KCL/GSTT Macafee Symantec F5	Security is based on the in-house solution and vendor solution is focused on secure firewall and anti-virus.	Remote access is given to a list of approved users.

2.3   Deployment Architecture

There are two sites for hosting data, one is jointly at GSTT and KCL premises distributed in dedicated server rooms and the other is at University of London Data Centre to store and backup the most important data. Figure 1 shows the Deployment Architecture.

There are five NAS at GSTT and KCL premises and each NAS is provided for a specific function. Bioinformatics Group has the most demands. NAS 1 is used for their secure backup, and NAS 2 is used for their computational backup, which is then connected to Bioinformatics services. NAS 3 is used as an important gateway for backup and archiving and is an active service connecting with the rest. NAS 3 is shared and used by Cancer Epidemiology and BCBG Group. NAS 4 provides mirror services for different locations and offers an alternative in case of data loss. NAS 5 is initially used by Digital Cancer cluster, and helps to back up important files in NAS 3. There are two digital cancer clusters, which can back up between each other, and important data are backed up to NAS 8 for reliability and NAS 5 for local version. The reason for this is that a disaster recovery activity which took place in 2010 took two weeks full time to retrieve and recover data.

Multiple backups ensure if one dataset is lost, the most recent archive (done daily) can be replaced without much time spent.
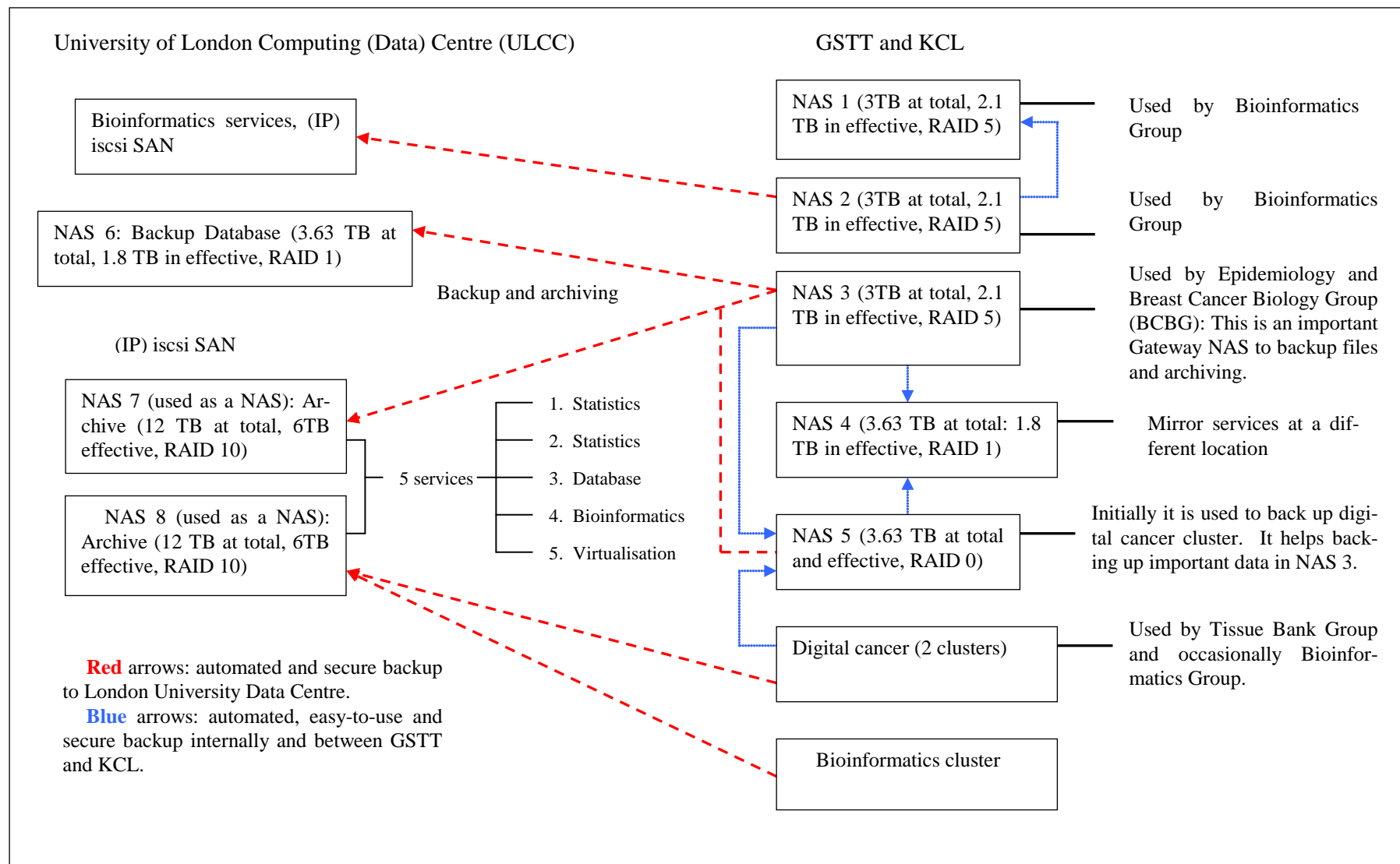
There are three NAS at the University of London Computing (Data) Centre (ULCC) where there are about 500 servers hosted for Cloud and HPC services. NAS 6 is used as a central backup database to store and archive experimental data and images. The other two advanced servers are customised to work as NAS 7 and 8 to store and archive valuable data. Performance for backup and archiving services is excellent and most data can be backed up in a short and acceptable time frame of less than one hour to back up data and images. This outcome is widely supported by users and executives. There are additional five high performance computing services based on Cloud technologies: Two are computational statistics to analyse complex data. The third one is a database to store confidential data and the fourth is on bioinformatics to help bioinformatics research. The last one is a virtualisation service that allows all data and backup to be in virtual storage format. These five services are not included in Cloud Bioinformatics for this paper.

2.4   User Support

The entire Cloud Bioinformatics Service has automated capability and is easy to use. This service has been in use without the presence of Chief Architect for six months, without major problems reported. Secondary level of user support at GSTT and KCL (such as login, networking and power restoration) has been excellent. There is a plan to obtain approval to measure user satisfaction.

Figure 1: Cloud Bioinformatics Deployment Architecture

University of London Computing (Data) Centre (ULCC)

Bioinformatics services, (IP) iscsi SAN

NAS 6: Backup Database (3.63 TB at total, 1.8 TB in effective, RAID 1)

Backup and archiving

(IP) iscsi SAN

NAS 7 (used as a NAS): Archive (12 TB at total, 6TB effective, RAID 10)

NAS 8 (used as a NAS): Archive (12 TB at total, 6TB effective, RAID 10)

5 services
1. Statistics
2. Statistics
3. Database
4. Bioinformatics
5. Virtualisation

Red arrows: automated and secure backup to London University Data Centre.
Blue arrows: automated, easy-to-use and secure backup internally and between GSTT and KCL.

GSTT and KCL

NAS 1 (3TB at total, 2.1 TB in effective, RAID 5)

NAS 2 (3TB at total, 2.1 TB in effective, RAID 5)

NAS 3 (3TB at total, 2.1 TB in effective, RAID 5)

NAS 4 (3.63 TB at total: 1.8 TB in effective, RAID 1)

NAS 5 (3.63 TB at total and effective, RAID 0)

Digital cancer (2 clusters)

Bioinformatics cluster

Used by Bioinformatics Group

Used by Bioinformatics Group

Used by Epidemiology and Breast Cancer Biology Group (BCBG): This is an important Gateway NAS to backup files and archiving.

Mirror services at a different location

Initially it is used to back up digital cancer cluster. It helps backing up important data in NAS 3.

Used by Tissue Bank Group and occasionally Bioinformatics Group.

3   Phase 2: Bioinformatics

Phase 2 focuses on the development of bioinformatics services activity started in September 2008 and was completed in February 2011. Applications are written in MATLAB, Mathematica and Visual C++. The objective is to archive all the code development to allow developers at different sites to work together on each project. Often the discussions prior the Cloud development was how to get different versions of codes together in a way that all developers could understand the progress and known issues in the development. Bioinformatics is an in-house solution focusing on scientific visualisation and modelling aiming to understand research analysis and improve existing services. The use of Cloud offers two distinct advantages:

(i) A PaaS for developers to simulate dynamic 3D modelling and visualisation for proteins, genes, molecules and medical imaging, where results can be instantaneous and data can be visualised, stored and shared securely.

(ii) Any complex modelling, such as growth of tumour and segmentation of brains, can be presented with the ease.

Each section is described as follows.

3.1   Tumour Modelling

Tumours develop as a result of abnormal and rapid growth of cells, and there are two types of tumours. The first type is benign tumours, which are harmless to human bodies. The second type is malignant tumours, which are malicious, should be removed and patients with them should be treated as soon as possible. Despite the fact that current technologies can take high-resolution pictures of tumours, it is extremely helpful for high performance Cloud resources to simulate the growth and formation of tumours, and this allows scientists and surgeons to diagnose possibilities of tumour growth and gain a better understanding about treatment (Grigoriadis et al., 2009). See Figure 2 for tumour modelling.

Figure 2: Selected figures in Tumour modelling

3.2   Medical Imaging

Medical imaging is widely adopted in Hospitals and medical institutes, and new ways to improve existing medical imaging services are regularly exploited. Bioinformatics Cloud platform allows computation and visualisation, and currently brain imaging can be used for demonstration. The aim is to study segmentation of brains, which divides the brain into ten major regions. The Cloud platform has these two functions: (i) it can highlight each region for ten different segments; and (ii) it can adjust intensity of segmentation to allow basic study of brain medicine. Figure 3 below shows selected brain imaging. Segmentation is an important aspect in brain study and it has two different functionalities. Firstly, it can highlight different areas in the cerebrum, where the different light intensity can highlight which particular areas. Secondly, segmentation can show different areas in the brain, including cerebellum, temporal lobe, mid-brain and so on. This allows medical students and instructors to understand the structure of human brain with the ease, but it also provides a platform to identity the right spot of the brain in a quick and efficient manner.
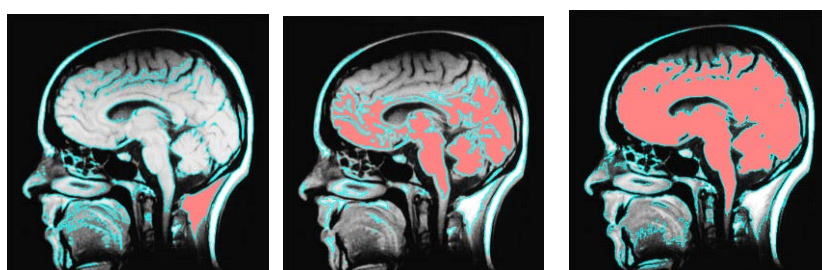

Figure 3: Selected brain imaging

3.3   Insulin Molecules

Insulin is a hormone central to regulating carbohydrate and fat metabolism in the body, and is important for type one diabetes treatment. Insulin has a molecular structure, and the study of its structure and for-

mation helps scientists to understand how to improve treatment. Cloud offers a platform for simulations and modelling enabling cutting-edge techniques to be used for Health Cloud for 3D Visualisation and modelling. This allows researchers to identify the areas in the molecule that they plan to study, and it allows 360 degrees rotation and zooming function, so that one particular area in the molecule can be magnified for different studies. Figure 4 shows the insulin molecule in original size and in zooms.
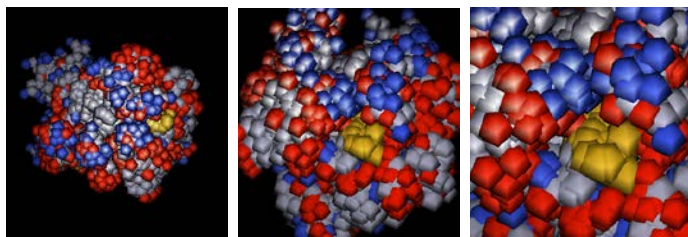

Figure 4: Investigation of insulin molecules on Cloud

3.4 Simulations for medical training

3D simulations on Cloud are very useful for medical education and workshop, since explanations can be made easier and participants can understand better with the aid of visualisation. 3D simulations such as DNA modelling, Poyllotaxis Spirals and cleavage of embryos have been used for training, and have positive feedback and support.

4 Brain segmentation: how human brains respond in details

Brain segmentation is an important specialist area for medical research to help scientists to understand how human brains work; to help trainee doctors to have better understanding with physiology and biological sciences related to brain; to help scientists to make more pioneering research for brain segmentation and the use of technologies in the aid of their analyses. Each part of the human brain has important functions to help our learning, memory, emotion and skills development. Simulations help scientists to understand how human learn new objects or skills; and also how scientists can discover better ways of identifying tumours. Often the detections of tumours require MRI scans to confirm, and any pioneering ways of discovery will help scientists to improve the quality of patients' life in the long term, and reduce the possibility of making errors in the process of treatment and operations (Baillard et al., 2001). Chang (2013 b; 2013 c) demonstrate the use of brain segmentation to investigate the relationship between learning and brain activities, as well as how to use

brain segmentation technology to improve on education. Chang (2013 c) also demonstrate the use of Cloud APIs to process thousands of datasets and produce results within seconds. The contribution is to present how Cloud Computing can work with other technologies and offer greater benefits to the medical community.

4.1 The core area of the brain segmentation – how the brain responds while recapturing a skill (eg. dancing)

Figure 5 shows the core area of the human brain, which covers the frontal lobe, parietal lobe, occipital lobe (these three are grey matters, the last one is situated at the back), corpus callosum, hippocampus (inner brain, the later one controls emotion) and brain stem (middle brain). The objective is to study how volunteers respond when they reuse a skill (which they learn some time ago but do not practice). Sophisticated equipment and expertise can be used but they are expensive. The use of simulations can reduce the costs in a way that results can be repeated. Figure 5 shows how human brain of volunteers (taking the average outcome) responds when reusing a skill such as dancing. The result shows that cells in the frontal lobe and occipital lobe are reacting actively, and the entire brain stem is responding very positively. This may mean volunteers require a high level of balancing. When they dance, their movement is fast, and must keep themselves balanced with swift but steady movements. The areas that represent the emotions are not so obvious but still can be seen with some activities.
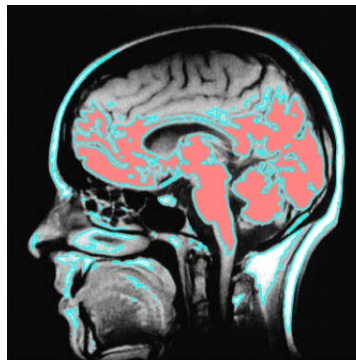

Figure 5: The core area of the brain segmentation (recapturing a skill, eg dancing, that was learned some time ago)

4.2 The core area of the brain segmentation – how the brain responds while learning a new skill (eg. dancing)

Figure 6 shows how the human brain responds for learning a new skill, dancing. Volunteers took very positive actions for their learning. This group of volunteers did not know the dancing moves but were extremely keen to take on dancing, particularly 'cool and impressive' dance moves, and the experimental session was full of fun, laughter and enjoyment. Although hippocampus controls human emotions, the entire section is in full pink colour. The lessons and dancing moves made volunteers very anticipated and enjoyable. As a result, all parts of the brain have hyper activities as indicated in the simulations. Figure 6 also shows that the volunteers have such high emotions that the entire brain is almost working at its own peak. In this case study, it confirms that positive learning can enlighten our learning capacities. If learners are involved in movement and are delighted in the process of learning, there is a strong possibility to link to positive learning outcomes. Other experiments and case studies will be conducted to consolidate this hypothesis. However, the role of simulations can help to ease the level of complexity while dealing with scientific research.
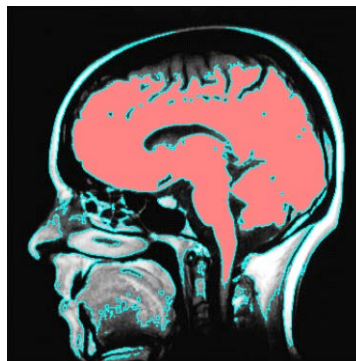

Figure 6: The core area of the brain segmentation (learning a new skill, eg dancing, for positive volunteers)

4.3 Other areas of the brain segmentation – work in progress

There are other work-in-progress that the use of simulations helps to identify any new projects on offer and also allows scientists to think more in-depth about what our bioinformatics services should provide. Figure 7 shows the outer area of the brain segmentation, which are closely related to brain analysis. Firstly, the outer area forms a protective layer to prevent the human brain from damage as a result of accidents. Second-

ly, the outer area can be helpful to calculate the ration of the size between the brain and the outer layer. This is helpful to some surgeons in terms of counting the number of tumours started or close to the outer region, particularly areas near spinal cord and pituitary gland, which may cause permanent damage and death. However, how these tumours develop is still not entirely known, and it is easier to find out from the outer layer rather than within the brain.
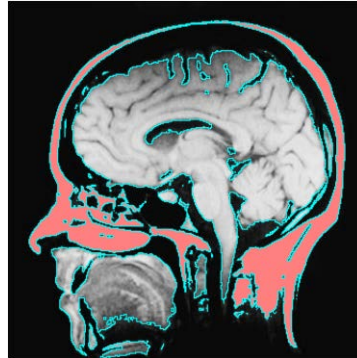

Figure 7: The outer area of the brain segmentation

Pituitary gland influences the way that hormone is used in the human body. Hormones cannot be controlled fully by other means and they play very influential development to human growth, health and development. There are projects-in-progress to determine how pituitary gland can be managed partially in terms of producing 'desired' hormones, such as insulin for diabetes patients instead of injecting insulin or taking medicine for treatments. Some scientists also discuss informally that the more effort they put in, there are more unknown territories for them to discover and work on. The role of simulations helps to identify the unexploited territory so that new results or improved methods can be known and adopted.
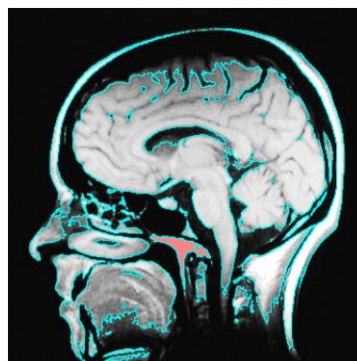

Figure 8: Pituitary gland of the brain segmentation

5   Discussions

There are three topics for discussions presented as follows.

5.1   User feedback on Cloud Bioinformatics

Currently Cloud Bioinformatics has provided users the following benefits:

- Cost reduction: The service is automated and saves costs in hiring and deploying staff and deployment of a larger and more expensive project that works the same. There is no need to hire a team to look after maintenance and daily services.

- Time-saving: Cloud Bioinformatics simplifies the complex backup process and saves time in performing backups. Users find that they need not spend significant time for back up.

- User friendliness: Cloud Bioinformatics offers easy to use features and users without prior knowledge can find it simple to use.

Healthcare community has a Data Protection Policy and not all types of services are able to release data. Services that do not use patients' data or confidential information are likely to be presented. Cloud Bioinformatics has another benefit for teaching. An internal survey was conducted that 95% of trainee doctors found that simulations were very helpful to their medical training. The focus for this service is no longer in technical implementation but rather user satisfaction.

5.2   Plug and Play Features in Cloud Bioinformatics – how this relates to e-commerce

There are papers explaining the importance and relevance of data intensive research, and why it is essential for Cloud development and services (Moretti et al., 2008; Hey, 2009). This Cloud Bioinformatics allows plugs and play, which means adding additional hard disks to existing NAS, or new NAS, can still provide services in place. This has been tested in 2010 where disk volume of NAS 7 and 8 were increased from 20 TB to 44 TB without interruptions of services. This Cloud Bioinformatics was also tested to store and protect data of up to 100 TB on another occasion. This allows any addition of hard disks and applications within 100 TB limit to provide user support and services. Additional applications can be added to the Cloud bioinfor-

matics without affecting the quality of the service, so that new features and services will be offered to the users.

Cloud Bioinformatics has been in used daily by medical researchers, and there are a few local administrators supporting a minimum level of services. Trainee doctors can find these simulations very easy to use so that all of them can understand the structure and functionality of the human brains in their lectures. In addition, simulations can help scientists to understand how our human brains respond when the volunteers are learning a new skill, or reusing a skill that volunteers have used some time ago, which help scientists to discover more about the power of human brain power.

All the development for the private cloud can be archived and presented in the form of a portal. This allows the internal and external users to be able to access the data and also analyse medical images and experimental results from their own laboratories. The portal acts as a one-stopped service to allow different roles of scientists to retrieve different types of data based on the access control policy, which is reinforced by the use of security tools and services described in Section 2. Scientists can view the simulations results visually at their own premises, and understand the outcome of simulations and experiments performed by other scientists. They can review those results and to judge whether they need to perform similar experiments at another occasion, or perform another experiments that other scientists have not performed as yet. This improves the co-ordinations and collaboration between different teams, so that each team focuses on one agenda at a time, without having the problems of duplicating each other's work, which is an issue in some of NHS trusts. The use of Cloud Bioinformatics can work as an internal and external e-commerce for scientists with the right of access control policy.

5.3 CCAF: The proposal for "Healthcare Platform as a Service" (HPaaS) for Research and Education

Cloud Computing offers contributions to research and development, as complex simulations can be computed and modelled with the on-demand capabilities, elasticity and scalability that Cloud can provide. Genes, molecules and medical imaging can be modelled at high speed and results can be computed and viewed in real-time. This is due to the establishment of PaaS to minimise the execution time so that 3D simulation can be running right after the code development on Cloud. The use of CCAF has been instrumental in the design,

deployment and support for Cloud services. The long-term strategic solution is to develop the CCAF as a consolidated and working framework. This allows different scientists from different research groups to work together to improve collaboration, efficiency, quality of research and funding opportunities, which explains why a more consolidated solution, Healthcare Platform as a Service (HPaaS), comes in place to aid the development.

Bioinformatics services also compare the performance improvements before and after introducing Cloud as an important ROI measurement. Chang et al. (2011 b) demonstrate that 1.2% - 7.2% time reduction for code development is achieved. Their objective is clearly met and project delivery is straightforward with progressive improvements. Different Health Cloud projects in Infrastructure, Bioinformatics, Statistics, HPC, Data Services and Security have worked together in an integrated environment to establish Health Platform as a Service (HPaas), which brings the following benefits:

- Different activities in private cloud can work together.

- The expertise in each area can be consolidated within the HPaaS.

- The outcome of one service can be the input to another.

Efficiency has improved as the Cloud saves time and resources to repeat the same processes, which can be automated. This is important in case the systems and/or services break that automated virtualised environments can quickly provision to the original setting. 3D Bioinformatics enhances the level of research and simulations can help surgeons and medical staff to make the right decisions. Chang et al (2011 d; 2012 b; 2013) and Chang (2013 a) also demonstrate Business Integration as a Service (BIaaS) that can further improve the process and integration of different activities in HPaaS.

6   Conclusion

This paper illustrates PaaS Portability in the form of Healthcare Cloud Bioinformatics, which is designed, deployed and serviced to GSTT and KCL under the recommendation of CCAF to ensure good Cloud design, deployment and services. Service Portability has been designed, implemented and serviced at participating organisations to provide added values such as efficiency improvement and time reduction in code development and execution time. User Groups for the system are divided into Bioinformatics Group, Databank and

Cancer Epidemiology Group, BCBG Group, Tissue Bank and Senior Clinicians. The CCAF was useful and helped the Health Community to achieve good private cloud design, deployment and services while following user requirements and challenges, and executives' feedback closely.

Healthcare Cloud Bioinformatics implements a data service as an easy-to-use, automated and collaborative platform which some users use every day. It is distributed between two physical locations: University of London Data Centre and GSTT/KCL and is designed and built to align with group and research requirements. It uses a private-cloud SAN architecture made up from different NAS services. The Deployment Architecture shows the connections between different NAS services and how they are related. These services include Bioinformatics (multiple services), joint Epidemiology and BCBG service, mirror services, two archiving services, digital cancer services and multiple backup services. Automated and secure backups take place between the two physical locations.

The first lesson from this activity is that recommendations from CCAF assist with achieving good Cloud design and deployment. A further lesson is that using experiments when designing and implementing a Cloud-based Storage Area Network (SAN) is helpful and execution time can be used as the benchmark to determine their success. The third lesson is that the private cloud allows developers to develop and test their code development in the repository, so that all known issues and new knowledge can be shared and effectively used in the Cloud project development. The efficiency and team collaboration have been hugely improved as a result of the Cloud adoption and new practice.

Bioinformatics simulations have been extremely useful for the healthcare scientists to make pioneering research. There are three added values as a result of Cloud adoption in the use of CCAF. Firstly, simulations can allow scientists to understand human organs (such as brains) and physiology (such as development of tumours and insulin molecules). This improves the efficiency for teaching, as trainee doctors can fully understand their expectations and medical lessons. Secondly, simulations help with pioneering research. Scientists can understand how volunteers learn and their reactions to the brain. Often human brain powers are not yet fully utilised, and the exploitations of human brain powers can be beneficial to some medical uses. Thirdly,

the use of bioinformatics can help different teams to focus on one agenda at a time without duplicating each other's work. This can reinforce the funding status and improves on each NHS trust's medical research outcomes and efficiency.

Healthcare platform (HPaaS) enables different activities to work together, so that expertise in one area can be consolidated. The use of 3D simulations allows developers to compute results in real-time and data can be stored, visualised and shared securely. 3D simulations of tumour, medical imaging and insulin have also helped to improve the quality of research analysis, as well as providing better understanding in the structure and formation of these analyses. All complex life science modelling can be presented with ease, so that it not only can promote greater awareness of health and disease issue, but also improves the quality of current research and development.

References

Armbrust, M., Fox, A., Griffith, R., Jseph, A. D., Katz, R. H., Kownwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., Above the Clouds: A Berkeley View of Cloud computing, Technical Report, No. UCB/EECS-2009-28, UC Berkeley, February 2009.

Baillard, C., Hellier, P. and Barillot, C., Segmentation of brain 3D MR images using level sets and sense registration, Medical Image Analysis (5), page 185-194, 2001.

Beaty, K., Kochut, A. and Shaikh, H., Desktop to Cloud Transformation Planning, 2009 IEEE International Symposium on Parallel and Distributed Processing, May 23-May 29 2009, Rome, Italy.

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J. and Brandic, I., Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, Journal of Future Generation Computer Systems, Volume 25, Issue 6, June 2009, Pages 559-616.

Buyya, R., Ranjan, R. and Calheiros, R. N., InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services, Algorithm and Architectures for Parallel Processing, Lecture Notes in Computer Science, 2010, Volume 6081/2010, 13-31 (Buyya et al., 2010 a).

Buyya, R., Beloglazov1, A., and Abawajy, J., Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges, PDPTA'10 - The International Conference on Parallel and Distributed Processing Techniques and Applications, 12-15 July 2010, Las Vegas, USA (Buyya et al., 2010 b).

Chang, V., Cloud Storage Framework – An Integrated Technical Approach and Prototype for Breast Cancer, Poster Paper and Technical Paper, UK All Hands Meeting, December, 2009.

Chang, V., Bacigalupo, D., Wills, G. and De Roure, D. (2010) A Categorisation of Cloud Computing Business Models. In: CCGrid 2010, The 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, May 17-20, Melbourne, Australia. pp. 509-512 (Chang et al., 2010 a).

Chang, V., Wills, G. and De Roure, D. (2010), A Review of Cloud Business Models and Sustainability. In: IEEE Cloud 2010, the third International Conference on Cloud Computing, 5-10 July, 2010, Miami, Florida, USA (Chang et al., 2010 b).

Chang, V., Li, C. S., De Roure, D., Wills, G., Walters, R. and Chee, C. (2011), The Financial Clouds Review, International Journal of Cloud Applications and Computing", 1 (2). pp. 41-63. ISSN 2156-1834, eISSN 2156-1826 (Chang et al., 2011 a)

Chang, V., De Roure, D., Wills, G. and Walters, R. (2011) Organisational Sustainability Modelling for Return on Investment: Case Studies presented by a National Health Service (NHS) Trust UK, Journal of Computing and Information Technology, 19 (3). ISSN Print ISSN 1330-1136 | Online ISSN 1846-3908 (In Press) (Chang et al., 2011 b)

Chang, V., De Roure, D., Wills, G. and Walters, R. (2011) Case Studies and Organisational Sustainability Modelling presented by Cloud Computing Business Framework, International Journal of Web Services Research. ISSN 1545-7362 (In Press) (Chang et al., 2011 c)

Chang, V., Wills, G. and Walters, R. (2011), Towards Business Integration as a Service 2.0 (BIaaS 2.0), In: IEEE International Conference on e-Business Engineering, The 3rd International Workshop on Cloud Services - Platform Accelerating e-Business, 19-21 October, 2011, Beijing, China. (Chang et al., 2011 d).

Chang, V., Wills, G. and Walters, R. (2011), The positive impacts offered by Healthcare Cloud and 3D Bioinformatics. In: 10th e-Science All Hands Meeting 2011, 26-29 September 2011, York (Chang et al., 2011 e).

Chang, V., Wills, G., Walters, R. and Currie, W., "Towards a structured Cloud ROI: The University of Southampton cost-saving and user satisfaction case studies", Sustainable Green Computing: Practices, Methodologies and Technologies (Chang et al., 2012 a).

Chang, V., Walters, R. and Wills, G. (2012), Business Integration as a Service, International Journal of Cloud Applications and Computing, 2 (1), ISSN 2156-1834, eISSN 2156-1826 (Chang et al., 2012 b)

Chang, V., Walters, R. J. and Wills, G., Cloud Storage in a private cloud deployment: Lessons for Data Intensive research (Best student paper), In the second international conference on Cloud Computing and Service Sciences (CLOSER 2012), Porto, Portugal (Chang et al., 2012 c).

Chang, V. and Wills, G. (2013) A University of Greenwich Case Study of Cloud Computing – Education as a Service. In, E-Logistics and E-Supply Chain Management: Applications for Evolving Business, IGI Global.

Chang, V., Walters, R. J. and Wills, G., "The development that leads to the Cloud Computing Business Framework", International Journal of Information Management, 2013, June, 2013 (Chang et al., 2013 a)

Chang, V., Walters, R. J. and Wills, G., Cloud Storage and Bioinformatics in a private cloud deployment: Lessons for Data Intensive research. In Cloud Computing and Service Science, Springer Lecture Notes Series, Springer Book, 2013 (Chang et al., 2013 b).

Chang, V. (2013), Business Integration as a Service: The SAP Case Study, International Journal of Next-Generation Computing, special issue on Cloud Computing, Volume 4, Number 3, 2013 (Chang, 2013 a).

Chang, V. (2013) Brain Segmentation – A Case study of Biomedical Cloud Computing for Education and Research, In, Learning Technologies Workshop, Higher Education Academy (HEA), University of Greenwich (Chang, 2013 b).

Chang, Victor (2013) Cloud Computing for brain segmentation technology. In, IEEE CloudCom 2013, 02 - 05 Dec 2013, Bristol, UK (Chang, 2013 c).

Grigoriadis A, Chang V, Schuitevoerder M, Gillet C, Tutt A and Holmberg L, "Cancer Cloud Computing - Towards an Integrated Technology Platform for Breast Cancer Research", Internal NHS Technical Paper, July 2009.

Hey, A. J. G., The fourth paradigm: data-intensive scientific discovery, Microsoft Publication, 2009, ISBN-10: 0982544200.

Kagermann, H., Österle, H., Jordan, J. M., IT-Driven Business Models: Global Case Studies in Transformation, John Wiley & Sons, 2011.

Latch, E. K., Dharmarajan, G., Glaubitz, J. C. and Rhodes. Jr., O. E., Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation, Conservation Genetics, 7:295–302, DOI 10.1007/s10592-005-9098-1, Springer 2006.

Moretti, C., Bulosan, J., Thain, D., and Flynn, P.J., All-Pairs: An Abstraction for Data-Intensive Cloud Computing, IEEE International Symposium on Parallel and Distributed Processing, 2008, IPDPS 2008, 14-18 April 2008, Miami, USA.