



*Eagle Genomics Ltd. White Paper*

# *Ten Steps to Successful Cloud Migration*

*Richard Holland, Eagle Genomics Ltd., Cambridge, UK*



## Executive Summary

Cloud computing has attracted a lot of hyperbole since it became a trendy topic for IT managers to talk about. Companies frequently trumpet their cloud enabled services but rarely give up details on precisely how they achieved this or how much of their infrastructure has been fully migrated. Security and reliability of cloud services are often raised as concerns. By understanding the basics of cloud computing and knowing how to assess important factors such as security and the identification of systems that are suitable for migration, it becomes much easier to design and implement a cloud strategy. This white paper will provide readers with essential facts about the cloud and will outline a checklist to facilitate successful cloud migration.

### What is the Cloud?

Cloud is a term that has a very loose definition in computing, generally meaning any situation where data storage and processing takes place without the user being able to pinpoint the specific physical computer carrying out the work. This may sound risky at first but in practical terms it really is no different to the traditional model of renting a server in a data centre. When renting a physical server it is unlikely that its precise location within the data centre will be known to the customer. The cloud is no different in this respect except that the physical server is replaced by a simulated (virtual) one.

### A brief history of Cloud

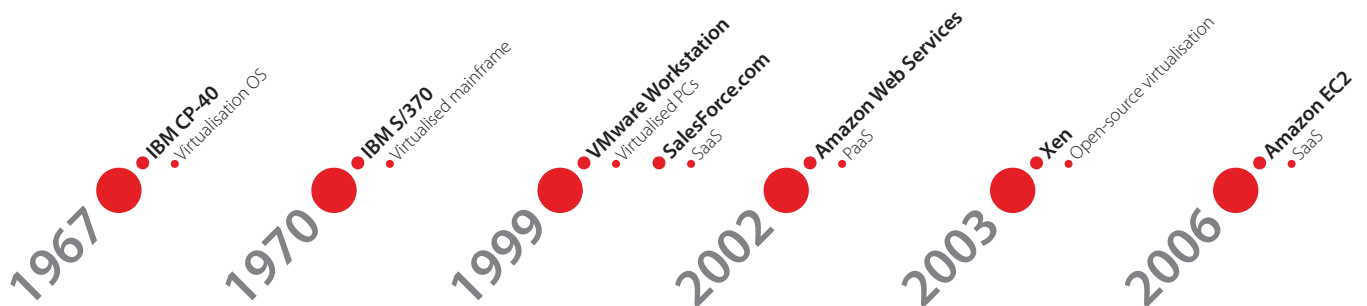


Figure 1 *Firsts in cloud computing.*

Traditional mainframes of the 1960s and 70s were the forerunner of cloud computing, although it wasn't called that at the time. Machines such as the IBM System/370 provided users with a number of simulated servers (virtual machines) running simultaneously on a single piece of powerful hardware. Each virtual machine ran its own operating system (for example IBM CP/CMS) and appeared to the user to be a completely independent entity functioning without reference to its host. The user did not need to know precisely which corner of the mainframe's hardware their virtual machine was running in – as long as it worked, that was all that was required.

Modern technology allows this same process of virtualisation to take place on commodity server hardware and even desktop computers. It is very common for companies to use out-of-the-box virtualisation technology such as VMware or Xen to be able to run several virtual machines on one physical server in the same way as mainframes used to. This allows them to increase the number of servers they have available without necessarily purchasing any additional hardware.

Cloud computing is merely the application of virtualisation technology on a larger scale. Instead of a single physical server hosting multiple virtual machines, clouds consist of groups of one or more physical servers sitting behind an access point that distributes requests for virtual machines between them. Clouds can be built in-house on top of existing hardware, or access to external clouds can be rented on demand from third-party vendors such as Amazon, Microsoft, or Google.

## Three types of Cloud

1. **Software-as-a-Service (SaaS).** Third-party software tools are made available across the web which allow storage of data remotely and access to it on demand. Examples include webmail providers such as Google Mail, or other web-based tools such as Salesforce. Each of these organisations leverages an internal cloud in their own data centre to distribute data across a number of machines and locations to facilitate its efficient update and retrieval.
2. **Platform-as-a-Service (PaaS).** Services such as Google App Engine and Microsoft Azure allow users to upload their own programs that conform to a programming language and interface (API) that the service provider has specified. The API allows the program to interact with the service provider's cloud infrastructure to store and retrieve data in the same efficient way as predefined SaaS services like Google Mail. The service provider then deploys the program's web interface on their internal cloud in order to enable it to scale on demand to cope with any number of users.
3. **Infrastructure-as-a-Service (IaaS).** This is precisely the same service as renting a physical server in a data centre or deploying a virtual machine on a physical server that has been enabled with virtualisation technology. Users are able to gain access to what appears to be a dedicated machine and are given complete control of it (often known as 'root access'). Amazon Web Services are market leaders in this area and it is often this particular definition of cloud that IT managers are referring to. SaaS clouds are usually built on top of PaaS clouds which in turn are often based upon IaaS clouds.

## Who is using it?

This is a good question, but a better question is: who isn't?

60% of business have moved or are planning to move into the cloud according to a February 2011 article in InfoSecurity magazine<sup>1</sup>, confirming a recent Gartner report cited in the same source that 40% of firms would have no appreciable local IT infrastructure within the next 4 years. (As an aside, it goes on to say that despite this widespread takeup of cloud computing 72% of respondents are concerned or very concerned about security in the cloud, but this white paper will attempt to explain these security concerns in the next section.)

An interesting side-effect of the cloud is that it encourages use of open-source software to run cloud-based services<sup>2</sup>. This is due to the lack of licencing restrictions – in a dynamically scaled and constantly changing computing environment, having to continually count and limit instances of licenced software is next to impossible. Many proprietary software vendors have yet to come up with licencing models that suit the cloud paradigm.

This combination of difficulty of using commercially licenced software with concerns over security is often the only excuse IT managers need not to use the cloud, but there are good arguments and techniques for overcoming both that mean that they do not necessarily need to be stumbling blocks.

*“As with any other outsourced hardware, externally-provisioned clouds are maintained and physically secured by professionals.”*

## Security check

With any outsourced service, security is paramount. We shall ignore application security for now as this is no different on an outsourced service than if it were to be hosted in-house – an in-house piece of software is just as hackable as an external one, only externally there are more people looking and more chances that someone will find the backdoor. If security testing has been done properly at the application level then the difference in the chance of it being hacked as a result of bad programming whether it be hosted on the cloud or in-house is negligible.

Application security aside, there are three key areas which require attention when deploying a service on any outsourced provider's hardware, cloud or otherwise.

1. **Encryption and authentication of communication.** The only way to avoid data being snooped on and intercepted whilst travelling between the corporate network and the external application is to encrypt it. Various well-established and easy-to-implement technologies exist to do this, such as the HTTPS protocol for encrypting websites. At the same time, steps must be taken to ensure that only authorised users are able to access the external service, which by default would otherwise be visible to the whole internet. Again many established technologies already exist to do this, such as OpenAM.
2. **Virtual firewalls, virtual separation, and encryption at-rest.**
  - a. Steps must be taken to ensure that the virtual server is only accepting connections to the application of interest. Allowing access to unrelated software increases the chances of hackers

1. <http://www.infosecurity-magazine.com/view/7539/as-cloud-takeup-and-budgets-rise-it-security-firms-are-ripe-for-takeover/>

2. <http://www.computerweekly.com/blogs/inside-outsourcing/2010/08/cloud-computing-will-increase-open-source-take-up-in-big-business.html>

gaining access through routes other than the application, and increases the maintenance overhead of ensuring these other access points are all suitably patched with the latest security fixes.

- b. In a virtualised environment, true separation of virtual machines within a physical server must be guaranteed so that two virtual machines on the same physical server cannot accidentally access each other's resources.
  - c. In the unlikely event of a successful hacking attempt, make it a point to encrypt any sensitive data whilst it is stored on disk so that intruders are unable to read it.
3. **Physical firewalls.** These are also known as doors, locks, keys, walls, and security guards. The easiest way to take over any machine is by walking up to it and taking physical control of it, so the most effective defence against the most determined intruder is a good physical security system at the data centre. Any data centre worth its salt will at the very least be certified to international standards of physical security.

### *Using commercial software*

Many commercial software packages are licenced on a per-CPU or per-seat basis. Per-seat licences can sometimes be suitable for cloud use depending on the mechanism they use for checking the licence validity. Every package and licence is different however, so it is always advisable to ask the software vendor whether it is permitted to use their software on the cloud under the existing licence or if an alternative licence would need to be negotiated.

The per-CPU licences present a different challenge. It is a simple task to migrate to the cloud a piece of software licenced in this way if it will be installed into a fixed-size environment as it would in the physical world, but if auto-scaling of that instance is required then the licence will more than likely preclude the dynamic creation and addition of servers. In principal a bulk licence could be purchased to allow up to a set limit of machines, but the user is then paying for licences for this additional capacity even when it is not being used. The financial overhead of doing this negates the cost benefit of being able to use the cloud to scale resources to meet demand.

Therefore when migrating per-CPU licenced software it is necessary to consider whether or not the software is required to scale on demand. If not, then it is not an issue. If scaling is required then special licences may need to be negotiated with the software vendor, or it may simply be easier to migrate to a less restrictive offering.

If a decision is made to swap the licenced software for something less restrictive, the most suitable alternatives at present are largely open-source as many proprietary software companies are still lagging behind in their provision of suitably cloud-proof licencing models.

### *Why do they use it?*

Finally, an easy question to answer!

#### *Bigger*

Cloud is scalable. The instant deployment of additional resources at times of peak demand, and the equivalent removal of these resources when demand has subsided, enables applications to continue responding to users even when under heavy load. The vast scale of most cloud installations means that it is unlikely that any single application will ever hit the upper limit of available resources, giving a perceived infinite level of scalability and availability.

#### *Better*

As with any other outsourced hardware, externally-provisioned clouds are maintained and physically secured by professionals. Chances are high that the security and maintenance regimes at cloud data centres are much better than many in-house corporate data centres, because the entire business reputation of the cloud vendor is at risk if they fail. Not only that, but the hardware itself will most likely be newer and more frequently upgraded to keep pace with the demands placed on the cloud vendor by its customers.

#### *Slow Is Fast*

Cloud is not necessarily faster (in terms of processor speed), but it is much bigger and allows more to happen at the same time. This means bigger jobs can be broken up into smaller ones with each small one run in parallel and results combined at the end – all in less time (by the clockface) than running the original big job. This benefit is shared with all existing cluster and grid installations, but with one key difference.

Unlike clusters and grids, cloud deployments do not require fixed-size resources. A cluster or grid will contain several physical nodes and will not be able to run even a simple short task without having most of those nodes configured and available. Therefore even if a task can be completed by a single machine, a traditional cluster or grid would have any number of machines sitting doing nothing, but still depreciating, consuming electricity and otherwise costing money for nothing, whilst the single machine does its work. Only the very busiest of grids and clusters are able to make the most efficient use of this fixed-size resource.

### *Cost-effective.*

In an externally-provisioned cloud such as Amazon, users only create virtual machines for as long as necessary, and only as many as necessary. If the job in hand requires only a single machine for 2 hours, then that is all that they need to fire up and pay for. If it requires 50 machines, then that can be done too. Only the machine-hours used are paid for.

In an internal cloud, there is still a fixed number of machines up and running, but instead of requiring expensive system administration time to manage the resources available on each node they are now capable of running virtually any task through the deployment of virtual machines through the cloud infrastructure. This allows servers to be repurposed almost instantly according to user demands and also allows the risk of letting users install their own software to be removed from the system administrator and placed firmly on the shoulders of the users themselves. The system administrator only needs to ensure that the virtual machines are fairly distributed and the physical resources made available to support an optimal service.

Simply put, internal clouds reduce the need to buy more servers, and especially so for short-term projects or those with uneven spikes of resource demand.

The pricing models of external cloud vendors reward sporadic use. It may sometimes still be cheaper to rent a physical server than to have a cloud-based virtual machine left running 24/7/365. For short-term or fluctuating use however the external cloud is hard to beat in terms of cost.

### *To migrate, or not to migrate?*

The information outlined above is a lot to digest, and readers would be right to interpret some of it as self-contradictory. The arguments change substantially when considering in-house clouds versus hiring time on an externally-provisioned cloud and what is true for one may not necessarily be true for the other. However the majority of discussions at present focus on the latter so this section of the white paper will focus solely on that aspect – i.e. what to look out for when migrating to an externally-provisioned IaaS cloud.

Care and attention to detail is vital when making the decision to migrate a project to an external cloud. Here are ten key points to consider when looking to migrate:

- 1** *Look for an established vendor with a track record.*  
A cloud vendor that is well established will have a wider breadth of knowledge and deeper insights into potential pitfalls than a smaller less-established vendor. They are also more likely to have higher security standards, a better range of services, more resources available to meet peak demand, and a better quality of support and training available for their users.  
The support and training provisions alone are likely to make the biggest difference to a customer that is new to the cloud – the vendor must be able to answer questions in great detail and within an appropriate timescale.
- 2** *Does the project really need to be migrated?*  
It may sound obvious, but not every project is suited to migration to the cloud. If management and customers are happy with the current hosting arrangements, and particularly so if they are cheaper than the cloud option, then there is no reason to move. Cloud migrations are usually only necessary when considering large-scale hardware purchases in order to sustain or scale existing in-house projects or enable new projects to take place. Such migrations are only value-for-money if the perceived benefits of the migration outweigh the costs of performing it.
- 3** *Consider data security.*  
It goes without saying that when putting applications and data onto a system outside an in-house data centre, customers will want to be sure that only the right people can access it and that its contents remain secure. Take a long hard look at the applications that will be migrated and consider getting an ethical hacker to attempt to break into

them so that developers can close any loopholes before the move takes place. Use firewalls liberally to ensure no accidental backdoors are opened through routes other than the application itself. Encrypt all communications with the external application and lock it away behind a proven authentication system that will guarantee that the only people who can access it are those who are permitted to.

## **4** *Data transfer.*

IaaS clouds are internet-based, therefore there is usually only one method of getting data into them: uploading files across the internet. Many internet connections used by smaller businesses offer far slower upload speeds than download speeds and it can take an eternity to upload even a gigabyte of data. Even the fastest corporate networks struggle to upload a few tens of gigabytes within a reasonable timeframe, e.g. a dataset from a DNA sequencing machine.

If the transfer of large datasets to and from the migrated software is a requirement then careful consideration needs to be made as to how this could be achieved, reduced, or avoided. Some IaaS providers offer the option of shipping hard drives of data to them to avoid these upload bottlenecks, but be aware that shipping delays and workload at the vendor's data centre may mean the process is far from instant. The best approach is to take a very close look at the data transfer requirements and see if they can be minimised, e.g. by pre-processing large datasets locally to produce a smaller summary dataset for upload.

Conversely, the cloud can be a good way of improving download speeds for customers using an application. Cloud vendors can deploy an application on whichever of their physical servers are closest to a customer's location. This can make a big difference to the response times customers get from the application.

## **5** *Data storage and location.*

How much data does the application really need? Cloud data storage can be expensive, particularly for very large quantities, so consideration should be given to data retention policies. Should old data be archived off-cloud to a tape library or other external resource? Is raw data needed at all or are summaries sufficient?

Whilst not hugely expensive, movement of data within the cloud does cost money in terms of internal bandwidth charging, depending on the vendor, so applications should avoid moving it around unnecessarily. Shared data resources such as shared drives or central relational databases can be more effective than directly copying data between virtual machines, particularly for data sources that are usually accessed randomly or partially rather than sequentially or in their entirety.

## **6** *Scaling.*

The scalability of cloud applications is not something that magically happens upon deployment (at least, not in IaaS – although PaaS deployments of single applications do inherit a certain amount of scalability from the host environment). Applications have to be placed behind load-balancers/auto-scalers within the cloud in order to be scaled up on demand. Some cloud vendors offer these as part of the service, others require the installation of third-party tools, however most scaling and balancing solutions incur some additional expense.

Once the application is behind a load-balancer or auto-scaler, the application itself needs to be aware that it could be scaled. If migrating an in-house application that already sits behind an in-house load-balancer then chances are that very little will have to be changed to support scaling in the cloud. For applications that have not yet been load-balanced in house, developers will need to assess the code to ensure that it can cope with a changing environment. How will user sessions be persisted? How will they co-ordinate access to any central data resources in order to avoid conflict?

## **7** *Service level guarantees.*

The first question to ask any cloud vendor is what their availability guarantee is, and what recompense there might be if they fail to live up to their claims. A cloud vendor failing to provide the agreed service is the worst possible situation for any cloud application to be in. Particular attention should be paid to the processes in place in case of vendor collapse or takeover.

Once confident of the vendor's service guarantees, the next check is to look at vendor backup plans. Do they take on- or off-site backups? What is their disaster recovery plan in case of loss of a data centre? Do they guarantee to recover the

contents of the virtual servers, i.e. the applications and data, or will they only recover the base operating systems? Independent off-site backup plans should be built with these answers in mind.

## 8 *Upgrade and maintenance schedules.*

Cloud vendors will need to update their systems from time-to-time to install the latest security patches and new features. Applications built on top of the cloud will need to be aware that these patches take place and have plans in place to ensure that they won't be adversely affected after the upgrade. Vendors often give an option to decide when the upgrade will take place, subject to a fixed final deadline when it will happen anyway, so application developers should carry out testing well in advance to ensure service continuity.

Likewise, if a vendor schedules planned downtime of cloud services to perform maintenance, try to schedule application maintenance windows to coincide with this in order to prevent excessive numbers of outages for customers using the application.

## 9 *Software architecture.*

Traditional applications are designed for traditional hardware configurations. Cloud applications are designed for cloud infrastructure and features. The two are not necessarily equivalent.

Whilst it is entirely feasible to take a traditional application and simply copy it to a cloud-based replica of its original environment, this is not always the most effective use of cloud functionality. Questions need to be asked regarding the choice of infrastructure – does it need a grid/cluster equivalent or can cloud alternatives such as Hadoop or self-instantiating instances provide a better service? Does it need an integrated load-balancer or can the cloud's default load-balancer suffice? Does it need database replication to distribute requests or can a single larger virtual database server handle all the traffic on its own?

## 10 *Check with the lawyers.*

The final hurdle when migrating to the cloud is almost certainly going to be a legal one. Data protection or other acts of law may prevent the placement of data in certain locations (e.g. French law prevents clinical trial data from being transferred to locations in other countries, even within the EU). The contract with the cloud provider must also provide suitable protection for data transmitted to it.

Checks must also be made to establish which jurisdiction's laws will apply in case of a dispute – the application owner's, or the vendor's head office, or the vendor's data centre locations where the application and data is being kept?

Some lawyers express concerns regarding intellectual property (IP) of any data that is outsourced to an external location, cloud or otherwise. The opinions and rules vary widely depending on local custom and precedent so seek legal advice before putting anything on the cloud that could construe potential IP.

Using licenced software on the cloud may in some cases contravene the terms of the licence, or may invoke special clauses that would not apply elsewhere. Check the licence text carefully and if necessary consult with the software vendor to gain appropriate permissions or renegotiate the licence.

### *The Eagle approach*

Eagle Genomics have extensive experience in designing and implementing cloud-based solutions (and other projects where the cloud is not the best choice). Eagle's cloud customers have included Big Pharma as well as small companies and academic institutes. To reassure customers that their data is safe, Eagle has engaged a team of ethical hackers to prove the effectiveness of the security framework used in Eagle solutions.

### *Track record*

Eagle Genomics' track record of delivering outsourced projects of all sizes on time and on budget to a number of the world's largest pharmaceutical and agri-biotech organisations stands testament to our ability to understand and clearly define client projects from the outset and get it right first time.

Eagle Genomics' main business is genomic bioinformatics, but our cloud knowledge is applicable across the board and our consultants are available to advise companies of all kinds on the best approach to migrating their systems to the cloud.

[www.eaglegenomics.com](http://www.eaglegenomics.com)



For more information about Eagle Genomics cloud consultancy services and about Eagle Genomics Ltd. in general, please visit our website or contact Richard Holland on:

Email: [holland@eaglegenomics.com](mailto:holland@eaglegenomics.com)  
Tel: +44 (0)1223 654481 ext. 3

**ElasticEagle** is a registered trademark of Eagle Genomics Ltd.

© Eagle Genomics Ltd 2011. All rights reserved.

Eagle Genomics Ltd. is a company registered in England and Wales, Company no. 6587071.