Arkivum
Every bit archived

ArkivumLimited
R21 Langley Park Way
Chippenham
Wiltshire
SN15 1GE
UK

+44 1249 405060

info@arkivum.com
@Arkivum
arkivum.com

Making the Cloud a Safe and Secure Place for Data

Whitepaper

Introduction – Cloud Services

There seems to be an endless number of cloud types: public, private, virtual, hybrid, multi- tenant, smart, club and community to name a few. There's plenty of hype and confusion on what cloud really is, sometimes it's traditional models dressed up with a new cloudy name, for example hosting kit you own in someone else's data centre or renting servers in a rack from an ISP. Sometimes it's a fully managed service that you can use as little or as much of on- demand and then only pay for what you use, Amazon is one of the pioneers in this space. The concepts of flexible capacity and online access to a measured service tend to crop up most often in definitions of the cloud[1]. The idea of a 'measured service' then leads to service level agreements and billing models based on metering by use[2], with pay-as-you-go being the archetypal model for 'public cloud' services.

It's these attributes of cloud services that allow organisations who use them to achieve a host of benefits. The most obvious are the ability to save cost, especially capital expenditure, and reduce or free-up in-house resources that are not at the core of the business. But these aren't always the main benefits, a survey by Richardson Eyres[3], showed that over 60% of the 3000+ respondents surveyed said that cloud for risk reduction/resilience was an imperative or significant driver. It was the number one driver at the strategic decision making level.

Using the cloud for risk reduction makes huge sense when it comes to storage of data, especially for the long term. Almost all enterprises have to retain data for a variety of reasons: for knowledge discovery, for reuse and monetisation, and to adhere to regulatory and compliance requirements. But few organisations have the expertise to do this effectively in house – long term data archiving and digital preservation needs specialist skills, a particular mind-set, and the use of specific and dedicated infrastructure. No surprise then that the Taneja Group[4] forecast cloud storage as a $10 billion business by 2014 – and a big part of that will be archive with a study by Forrester Research[5] in 2011 suggesting that cloud archiving will grow between 28% and 36% a year.

Considerations – Data Safety and Security

So, if you want to realise all those benefits of using a cloud service, then how do you know your data is in good hands, that it is safe and secure, and that whatever happens you won't be left high and dry? NIST[6] in the US, and ENISA[7] in Europe, have both published extensive recommendations on what to do and what to look for when considering the use of cloud services. A risk assessment and risk management approach is the key, and this is exactly what we suggest at Arkivum when considering the cloud for data archiving, or any other data storage for that matter where data safety and security are paramount.

There are seven questions which need to be answered when making decisions on data archiving:

Q1. Can you be sure of the integrity of your data in the cloud?

Are you guaranteed that every single bit of data you pull back from your cloud service is identical to what you put in - and can you prove it?

Q2. Is your data always confidential in the cloud?

Can you be sure that there is no way for anyone else to access and understand your data when it is 'in the cloud' - and that includes any service providers used to hold the data?

Q3. Do you know the authenticity of your data at all times?

For your business, do you know who has put data into the cloud service, who has then accessed that data from the cloud, and then for the cloud provider do you know who has been handling your data inside the cloud service in order to keep it safe and secure?

Q4. Is your data available from the cloud where and when you need it?

If you need data quickly then can it be retrieved in a matter of minutes? Can this be done for every single file you want to store?

Q5. Do you know who is in possession of your data at all times?

Do you know exactly where your data is stored in the cloud and who runs the facilities? Do you know who controls how your data is stored and who is responsible for its safekeeping?

Q6. Can you always use your data no matter what happens to the cloud service?

What happens if your cloud service provider goes out of business or stops support? Are you locked into proprietary data formats or tools? How do you get your data back?

Q7. Can you always assert the provenance of your data when it's in the cloud?

Faced with a compliance audit, the need to understand whether you're litigation ready, or the need to prove the authenticity of your data, can you provide an audit trail that says what the data is, where it's come from, and everyone and everything that's happened to it since?

The questions about confidentiality, integrity and availability are core to information security and are known as the CIA triad, which should be a priority in assessing a cloud service.

The questions about authenticity, possession and utility are further essential facets of risk management for information assets and apply in the cloud as much as anywhere else – probably more so. The final question on provenance is being able to support auditability: asserting a chain of custody for data assets.

Arkivum's the Answer

Questions are easy to ask, but it is the answers that count. At Arkivum, we ensure that we have all the answers. The files in our service are stored in two fixed separate data centres that are known to our customers, in online tape libraries for resilience and high availability, and if we ever need to move customer data to another location we always gain permission first. Each customer's data is on dedicated tape, which can be viewed in person if necessary. To give further peace of mind a third copy of the data is stored offline with an independent third-party escrow site, assuring that your data is accessible to you no matter what happens to our business or service. If the worst does happen each escrow tape comes complete with a full manifest and set of open source tools that can be used for easy recovery of the data.

All tapes are only ever handled by trained Arkivum staff who are responsible for their safekeeping through use of our highly automated data management systems. The use of tape libraries means that access to files is quick – less than five minutes to reach the start of any file.

To ensure data integrity we generate checksums on all data as it enters our systems, which are agreed with the customer, these are then regularly checked to detect any problems and then fix them using the replicas we store for each and every file. Whenever a customer asks for a file from the archive we will use the checksums to confirm that it is bit-identical to the original – no files are ever returned without passing this test. We have a full information security management process in place that covers all aspects of security including people, process, facilities, equipment and software.

All customer data is encrypted on-site with AES encryption before it leaves the customer network. The customer owns the key-pair and Arkivum does not have a copy of the keys at our data centres or offices, so we have no way to decrypt customer data. For even greater security, data is transferred to our data centres, which all have extensive physical security measures, over a two-way authenticated VPN connection.

We store a complete record of where every file in our service has come from, what its ownership and permissions are, when it has been retrieved by the customer, when we've migrated it to new media, when we've checked its integrity, when it went into escrow – as well as any unauthorised attempts to access the file.

Together with our information security management system and tightly defined policies, we provide a remarkably complete and rigorous record of everything and everyone who has interacted with your data in any way, be it access you have made or actions we have taken to ensure integrity and authenticity.

Summary

Cloud Computing is here to stay and storing data in the cloud is certain to form a major part of any enterprise's IT strategy moving forward, not least because it presents businesses with very tangible benefits in terms of increased resilience, reduced costs, and release of in-house resources. However, the decision to opt for the cloud forms just a small part of a much larger underlying process; it is the steps that you take before putting your data in the cloud which are crucial. Care must be taken in choosing your provider. Maintaining a corporate archive is much more than merely warehousing data. Your choice needs to reflect not just excellence in data management but advanced policies on bit preservation, security, and access. The business providing the archive must be secure and well-provisioned, with 'fail-safe' backup plan to ensure that data can be recovered without undue complexity should the company fail. The seven questions outlined above should provide a great starting point for your journey and if you get the answers you're looking for from all of them then putting your data in the cloud should be smooth and seamless.

About Arkivum

The author, Matthew Addis, is part of the team that developed an innovative new approach to data archiving, which is now commercialised in the company Arkivum.

The company was launched in 2011 following a four year research project at the University of Southampton IT Innovation Centre. Partners in the project, which benefitted from UK government funding, were BBC R&D, Ovation Data Services, University of Edinburgh, Xyratex and IT Innovation. The aim was to develop a new approach to planning and managing large- scale, sustainable and integrated digital archive solutions.

Arkivum now has established data centres in the UK and is working with clients from a number of industry sectors on archiving solutions which provide a 100% guarantee that all data can be recovered at any time.

[1] http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf
[2] http://www.gartner.com/it/page.jsp?id=1035013
[3] http://www.richardsoneyres.co.uk/pdf/survey_report_4pp_mr.pdf
[4] http://tanejagroup.com/profiles-reports/request/abstract-taneja-group-emerging-market-forecast-emf-on-cloud-storage-product
[5] http://www.forrester.com/Your+Enterprise+Data+Archiving+Strategy/fulltext/-/E-RES58169
[6] http://www.nist.gov/manuscript-publication-search.cfm?pub_id=909494
[7] http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment