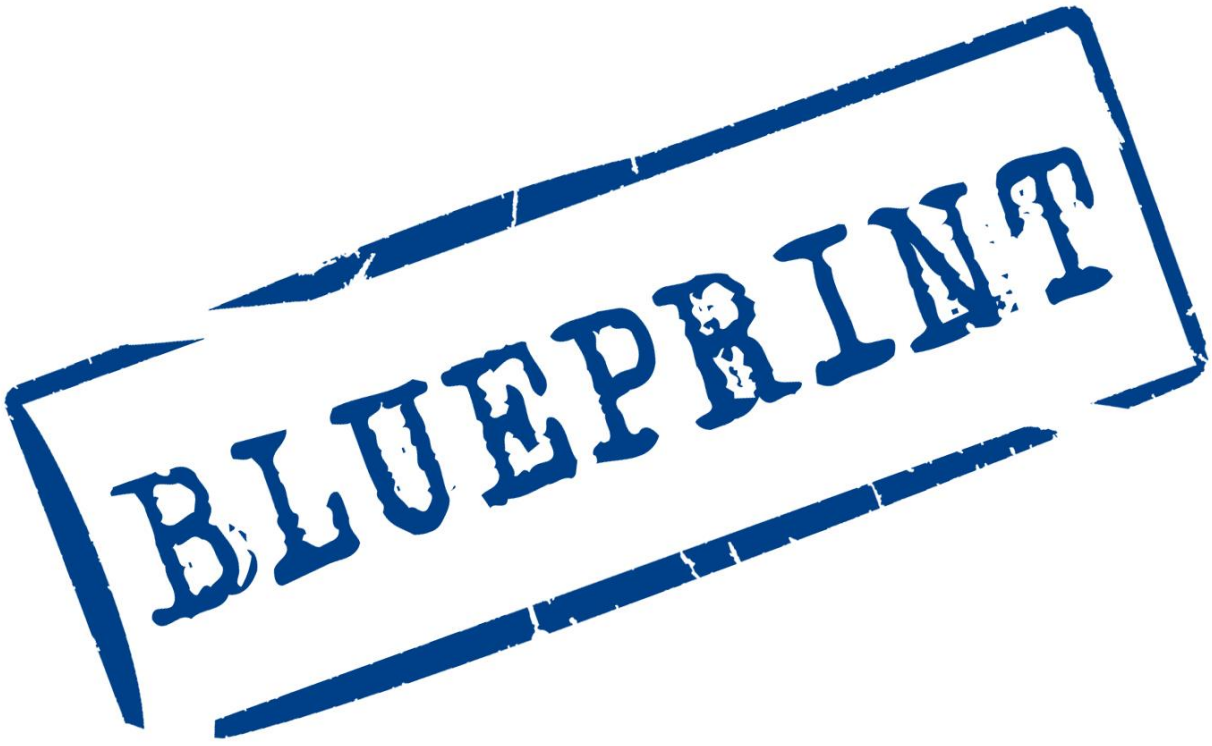
ukcloud

Pure commitment.

High Availability and Disaster Recovery Options



UKC-GEN-101

INTRODUCTION

Ensuring critical IT systems can be quickly and easily recovered or continue to function after a disaster is a vital element of any organisation's business continuity plans.

Effective planning to deliver high availability and ensure disaster recovery is more cost-effective in the long run by reducing response and recovery costs as well as minimising any disruption.

Cloud computing can remove the need to manage and pay for redundant remote infrastructure, releasing resources to focus on core business activities. Organisations are able to leverage the active/active nature of the cloud platform to create disaster recovery solutions which closely meet their needs.

Your organisation is obviously unique in terms of the applications it runs and the priorities assigned to each, as well as associated services and data. So your disaster recovery plans will also be unique.

This Blueprint aims to inform and assist you as you specify your disaster recovery plans. It details measures put in place by UKCloud to protect against major failures, how our platform has been built with service continuity in mind and how we offer extensive capabilities that can be used to protect your business against a disaster. We have detailed a number of scenarios and explained different options for how our platform can be utilised to enable effective disaster recovery solutions.

IN THIS BLUEPRINT

WHAT COULD POSSIBLY GO WRONG?

When designing for high availability and disaster recovery for mission critical applications, it is useful to understand how our cloud platform protects against the many sources of failure that are possible in the physical infrastructure underpinning an Infrastructure as a Service (IaaS) cloud.

So what could possibly go wrong in a cloud computing environment, and how does the UKCloud architecture mitigate against these failures? Many of these features are taken for granted by large enterprise customers, but many smaller organisations may not have the budget or capability to implement these enhancements.

Component	Mitigation
Power	Dual power supplies into data centres from different national grid providers (N+1) UPS equipment (N+1) Diesel generators
Physical network infrastructure	Fully redundant routers and switches Diversely routed cabling/fibre/ducting
Server/device infrastructure	Redundant power supplies, fans Redundant internal disks using RAID, hot-swappable Device/Host clustering Virtualised contexts
Storage infrastructure	Redundant power supplies, fans Redundant dual storage fabrics Redundant disk configuration (RAID), hot swappable

Over and above the physical infrastructure, what benefits does virtualisation bring? Our cloud platform also provides the following additional mitigation:

Component	Mitigation
Physical host failure due to memory error, CPU failure or something equally substantial	VMs on failed host are detected as being offline and are restarted automatically on other nodes in the cluster
Physical host becomes overloaded	Less busy VMs on constrained hosts are migrated away to other hosts in the cluster with no loss of service to rebalance cluster
Physical host needs to be taken down for maintenance	VMs on target hosts are migrated away to other hosts in the cluster with no loss of service
VM operating system crashes (BSOD)	VM can be restarted a number of times, then raise an alert if required
Major datacentre failure	VMs restart instantly in the other site (if using the ENHANCED service level) retaining VM configuration such as hostname and IP address. VMs are rebooted in a crash-consistent mode.

DESIGNING FOR DISASTER RECOVERY

A quick search of the Internet will reveal large quantities of information on the subject of High Availability compared to Disaster Recovery. For the purpose of this Blueprint, we will focus on how UKCloud offers capabilities to protect against the catastrophic failure of an entire data centre.

Ideally, cloud applications should be architected so that they are 'Designed for Resilience'. As described in our Blueprint entitled 'Characteristics of Cloud Applications', the architecture should reflect loose coupling between components, through solutions such as Message Queuing. The architecture should also prefer multi-master shared-nothing clusters rather than traditional active/passive shared storage clusters (where the storage remains a single point of failure).

Applications should be designed so that if a component has failed, it automatically tries a secondary component (even in another data centre or on a cloud platform). In that way, there is no 'recovery', rather applications continue to run as designed regardless of the failure. That said, in our Blueprint "Legacy & Enterprise Applications in the Cloud", UKCloud recognises that there are many applications which pre-date cloud and hence are not 'Designed for Resilience'. These applications expect to use traditional technologies such as clustering, replication, failover and backup, and hence need to be 'recovered' when a failure occurs.

In either case, you should remember that the infrastructure delivered by UKCloud comprises separate & independent elements available in both sites (or availability zones):

- Compute Storage – block storage associated with virtual machines via VCE Vblock and optional EMC VPLEX Storage Replication
- Cloud Storage – scalable object storage via EMC ATMOS cloud storage platform
- Automated VM Backups via EMC Avamar Purpose Built Backup Appliance

All these capabilities are located in separate technologies in different locations in both data centres.

If the Compute as a Service platform were to fail, the VM backups would still be safe in another location using a different technology. Similarly, if you had backups or longer term storage in Storage as a Service, it is physically separate to the storage on the Compute as a Service platform.

For most use cases, disaster recovery is used to protect against the worst type of outage – the loss of an entire data centre. An example of this was the Buncefield Old Depot explosion in Hemel Hempstead in 2005 where several IT companies lost data centres.

UKCLOUD BUILDING BLOCKS

UKCloud offers a number of capabilities via individual cloud products and flexible Service Levels to enable High Availability and Disaster Recovery solutions.

The UKCloud Compute as a Service cloud platform has been built using best of breed technology from VMware, Cisco and EMC, and from the amalgam of these, VCE, in the form of Vblocks – factory-integrated solutions containing Cisco UCS blades servers, EMC VNX storage arrays and Cisco Nexus converged networking fabric.

For automated VM backups, UKCloud makes use of EMC Avamar storage, and use the VMware APIs for Data Protection to enable efficient crash-consistent snapshot backups of entire VMs on a nightly basis – which have no impact on your workloads. These backups are stored on independent purpose built backup appliances located in a remote site.

For Cloud Storage, UKCloud provides scalable object storage via EMC Atmos which is natively designed for cloud environments with its support for Petabyte scale, multi-tenancy and API accessibility. The service is addressed via the native API which is also compatible with the Amazon S3 API and many applications which support S3. Various solutions exist to make the object storage appear as a network file share to VMs.

The Compute as a Service cloud platform is powered by ubiquitous VMware technologies including VMware vSphere with vCloud Director. Several physical clusters have been implemented across both sites and security domains to provide you (via the UKCloud Portal powered by vCloud Director) with an entirely self-service environment that provides complete control over networks, firewalls, catalogues, virtual machines and related resources.

UKCLOUD VM REPLICATION

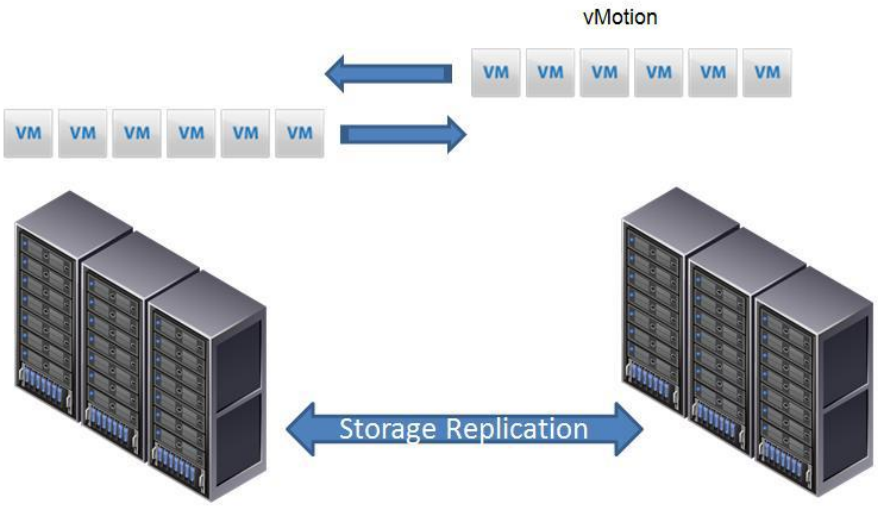
A key benefit of our cloud platform is that it is available independently in both of our UK data centres and so provides an active/active infrastructure to support your disaster recovery solutions.

For the majority of our service levels, each virtual machine runs on a VMware vSphere cluster which is hosted in a single site. At these service levels, you are advised to have independent virtual machines running at each site and to configure application level replication and failover as appropriate. This is equivalent to the concept of running your application across multiple Availability Zones.

For our VMs using the geo-resilient option, UKCloud has uniquely built a VMware vSphere cluster that is stretched across both sites. This has been achieved by taking advantage of technologies including EMC VPLEX (storage replication), Cisco OTV (network virtualisation), VMware Metro Cluster Services and the substantial Data Centre Interconnect (DCI) bandwidth that UKCloud has provisioned.

spans both sites. At this service level, your VMs can be proactively migrated whilst still running (and with no downtime) between sites (unlike most cloud platforms today). However, in most scenarios, the migration between sites will be a reactive event responding to an unplanned outage at the primary data centre. In these scenarios, your VMs will simply be restarted at the other data centre in a crash consistent state. In all scenarios, your VMs retain their original configuration (such as hostname and IP configuration).

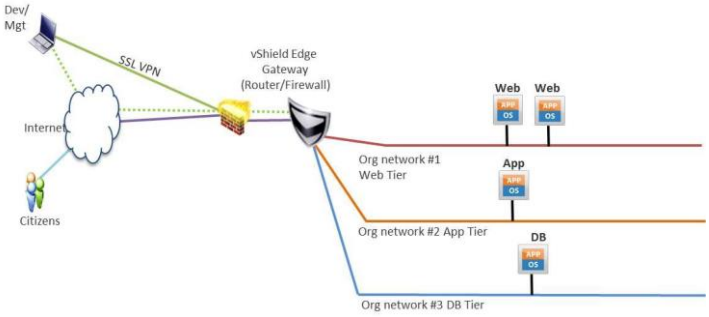
The synchronous replication provided by the geo-resilient option is enabled by secure data centre interconnects which have a latency less than 5ms round-trip-time. As with any synchronous implementation, the solution has an impact on data-write performance (i.e. fewer IOPS relative to asynchronous implementations) but delivers a recovery point (RPO) of near zero data loss. The recovery time (RTO) varies from a few minutes to a few hours depending on the precise failure condition and the efficiency of your application to recover when restarted in crash consistent mode.



This enables VMs using the geo-resilient option to be delivered in a Virtual Data Centre which logically

USE CASES

Consider the simple example of a three tier application below. It is comprised of three zones, one for each tier. The vShield Edge Gateway is providing load balancing across the two web servers.



How can you make this more highly available, for example to cater for disaster recovery?

The following questions can help you specify your requirements:

- As well as catastrophic disaster (affecting system availability), how do you protect against data loss or data corruption (affecting data integrity)?
- Do you require point-in-time system backups or continuous replication?
- How much data can you afford to lose in the event of a disaster (RPO)?
- How quickly must your application become available again following a disaster (RTO)?
- How long does it take your application to recover from a crash consistent state?
- Can your application be configured for application level replication or multi-site deployment?
- How much of your environment is stateless and can be re-provisioned when required?
- Can you deploy a solution to redirect users on failover or is it essential that the original configuration is retained?

We'll now consider a number of different ways of architecting a disaster recovery solution to address the different requirements you may have.

Use Case 1: Transparent replication and failover using the geo-resilient option

This option provides the easiest access to a disaster recovery solution. It is ideal for solutions where there isn't enough time, budget or capability to engineer the solution across two data centres.

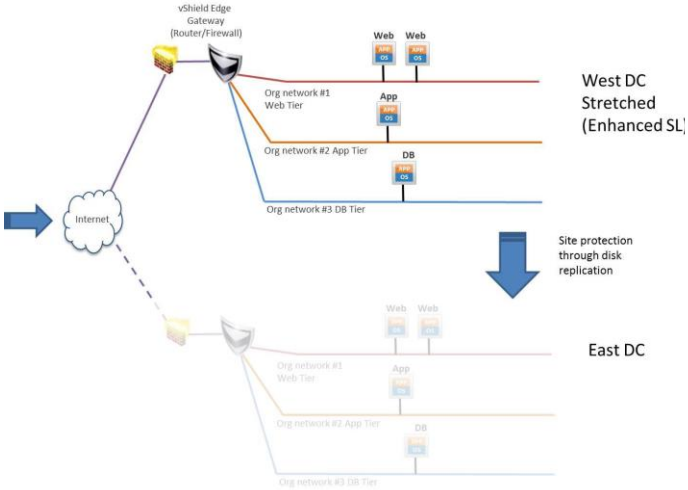
In this example, the entire application stack is deployed at the geo-resilient level. Each VM is automatically and continuously replicated to the second data centre by our cloud platform. In the event of a major failure at the primary site, our cloud platform will automatically restart the whole application (including the vShield Edge Gateway and associated networking and security configuration) at the secondary site. The unique design of the UKCloud platform ensures that your VMs do not require any reconfiguration before, during or after failover – the VMs are simply restarted rather than recovered.

The synchronous replication provided by the geo-resilient level (sub 5ms round-trip-time), provides a RPO of very close to zero. The RTO will vary from a few minutes (for the VM to start rebooting at the other site) to some hours (dependant on how long it takes your application/database to run integrity checks to bring it into a state of consistency; if required).

Please note, as data is replicated synchronously, any malicious or accidental changes that occur at the primary site will be immediately replicated to the DR site. Hence the recovery time may be extended if data needs to be restored from backup or if the application needs to roll-back through its logfiles.

Advantages	Disadvantages
Simple and convenient	Poor control and visibility
Near zero data loss	Difficult disaster recovery testing
Automatic failover	Lower storage performance
Application agnostic	Indiscriminate replication

The diagram below shows that the application ordinarily runs in a single site with continuous replication ensuring that the application and associated configuration data is always available to be restarted at the second site. UKCloud automatically configures virtual machines with 'Site Affinity' such that all virtual machines in the geo-resilient level Virtual Data Centre (VDC) run in the same site to reduce latency and inter-site traffic. This ensures that the entire application exists in one site and automatically fails over to the second site.

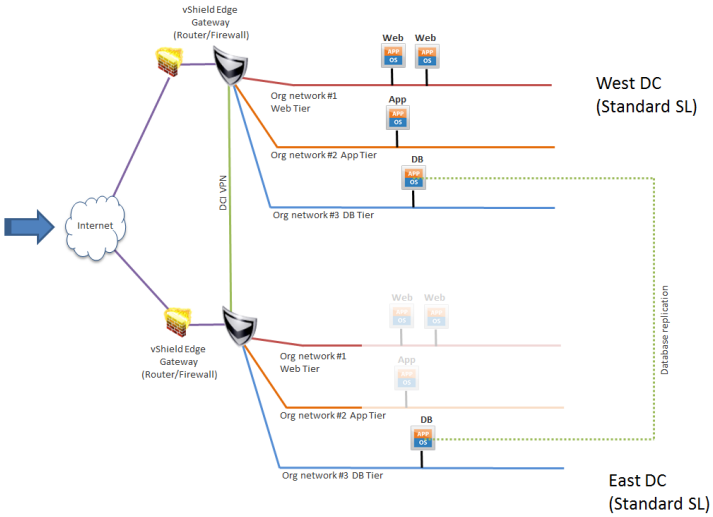


In the event of a significant failure affecting the primary site, our cloud platform automatically restarts the affected virtual machines at the second site in a crash-consistent mode. This means the application retains the same configuration and data as existed in the primary site up-to the point of failure. Some applications are able to restart and recover automatically from a system crash whereas others will require manual intervention. In all scenario's, the customer (or partner) is responsible for bringing their applications online.

Use Case 2: Self-managed application recovery for ESSENTIAL, POWER or PRIORITY VMs

This option is ideal for advanced solutions which can be engineered to run across multiple sites. Using this option, solutions can leverage UKCloud's active/active cloud platform to provide near zero downtime as the application is in full control of the replication, failover and failback.

In the example below, a Virtual Data Centre (VDC) has been deployed in each UKCloud site. The VDCs in each site can be connected together using a Site-to-Site VPN between the vShield Edge Gateways over the UKCloud data centre interconnects. Virtual networks in each VDC have been configured to reflect the different tiers/zones for the application. VMs in the Web Tier and App Tier are provisioned and then powered off in the second site to reduce costs. Only the database VM is configured to remain running in the second site so that database level replication can be used to enable continuous data replication.



The Web Tier and App Tier in the second site (East DC) are shown greyed out because:

1. The VMs are pre-deployed but powered off to reduce running costs.
2. The VMs will be automatically provisioned during failover (for example, using orchestration solutions like Puppet and CFEngine).
3. The VMs will be deployed from a catalogue instance/template.
4. The VMs will be restored from the UKCloud 'automated VM backup' optional feature via a manual service request.

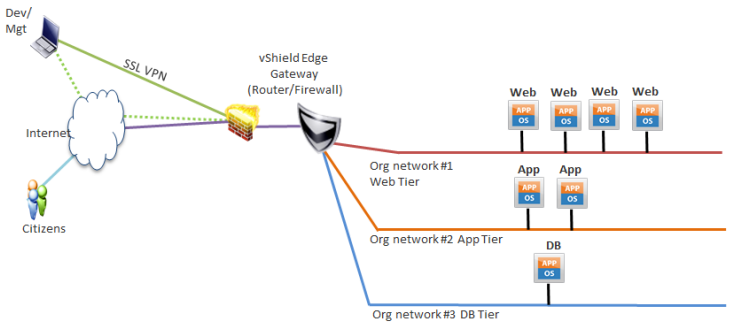
In this scenario, both the RPO and RTO are completely controlled by your solution as UKCloud operates independent platforms in each site in an active/active configuration. Unlike Use Case 1, on failover the solution will be accessed by a different set of IP addresses and so may require some reconfiguration (such as DNS hostnames) or solutions such as Global Load Balancers, Content

Delivery Networks, or similar. This also provides the ability to perform DR testing or to take elements of the solution down for maintenance without impacting end-users. Further, as the application controls data replication it can be implemented to minimise the effect of malicious or accidental changes through techniques such as checkpoints and asynchronous replication.

Advantages	Disadvantages
Familiar traditional architecture	Complex failover and feedback
Provides most control and flexibility	More virtual machines to manage
Can deliver near zero RPO and RTO	Requires application layer functionality
Application consistent failover	Slower RTO unless redundant VMs used
Easier DR testing and validation	Potentially higher running costs
Potentially better storage performance	

Use Case 3: Restoring from backup in the other site — considerations

This use case addresses a common scenario suggested by UKCloud customers. In this scenario there is an application running in one UKCloud site withincorporated automated VM backups by default. These backups are configured by default to be stored on purpose built backup appliances located in the remote site which ensures that the data backups remain available even if the primary data becomes unavailable.



In the event of a major failure to the primary site, you might want to restore the offsite backups of VMs to the Compute as a Service cloud platform running at the second site.

Although this is technically possible, it should be remembered that you will not have a Virtual Data Centre (VDC) in the other UKCloud site to host these restored VMs. Hence, for this option to be viable it would be necessary for:

5. You to formally request a new VDC(s) in the other UKCloud site via a Service Request. A delay will be incurred whilst the request is queued to be actioned.
6. UKCloud to create the VDC(s). The ability to do this is subject to available capacity [1]
7. You would need to configure the VDC(s), provision the virtual networks and apply appropriate vShield Edge configuration (such as NAT rules, firewall rules and VPN configuration) to mirror the original VDC.
8. You would need to update external DNS entries with the new Public IP addresses assigned to the VDC(s)
9. You would need to raise a further Service Request for the VMs to be recovered from 'automated VM backup' system to the newly created VDC(s) [1].

[1] In a disaster recovery scenario, due to the likely volume of requests for new VDC(s), UKCloud could not guarantee that there would be sufficient capacity to restore the VMs. This is because spare capacity will be allocated firstly to VMs using geo-resilience and to existing VDCs already running on the cloud platform. Hence, UKCloud strongly recommend that customers consider Use Case 1 or Use Case 2 in preference to this option.

Advantages	Disadvantages
Inexpensive	Unreliable
Simple to understand	Indeterminate RTO
	Difficult to perform DR test
	Poor RPO (last successful backup)
	Complex recovery process

Use Case 4: Using cloud as a DR platform for your current data centre

Whereas the previous use cases focus on how you can architect cloud-based applications for disaster recovery, this use case shows you how you can use

our cloud as a DR platform for your current in-house hosted applications.

There are many third party solutions such as Neverfail, Zerto, PHD Virtual which can further facilitate this use case but for this example we will focus on vCloud Connector (VCC) which is provided free by VMware.

vCloud Connector is designed to connect your in-house vSphere solution with vCloud powered platforms such as UKCloud Compute as a Service. The solution enables you to:

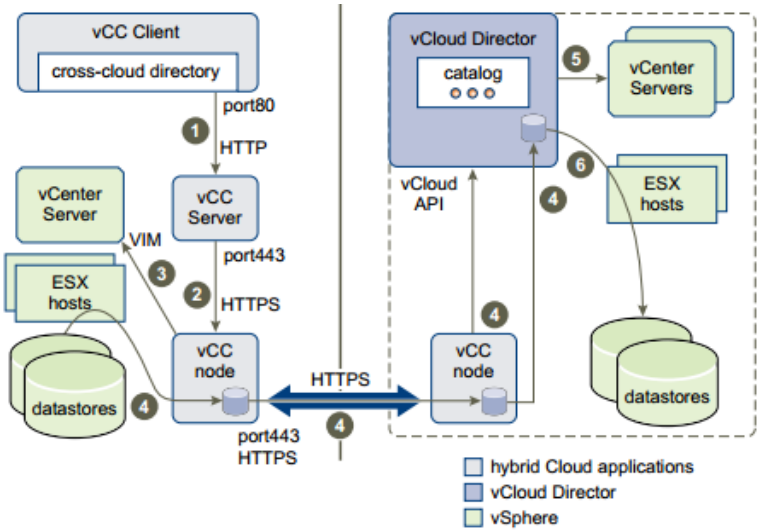
1. Extend your in-house vSphere platform into the cloud

2. Maintain replicas of your templates synchronised across your in-house platform and the cloud platform

3. Copy VMs or vApps to and from our cloud service

The solution might be viable for point-in-time DR solutions rather than those that require continuous replication — although the latter can be supported by using vCloud Connector to seed the VMs into the cloud and then enable application-level replication between the in-house source VMs and the cloud based destination VMs (much like Use Case 2).

This solution requires you to establish a connection between your vSphere infrastructure (the left half of the diagram below) and the UKCloud Compute as a Service platform (the right half) using VMware vCloud Connector.



Once the connection has been established:

1. You can transfer selected VMs or vApps to the UKCloud platform via the vCloud Connector application.

2. vCloud Connector automatically exports VMs from your local platform, transfers them to our cloud platform and then imports them into your Private Catalogue in your VDC.

3. As required, instantiate your vApp templates from your Private Catalogue to create running VMs.

GENERAL DISASTER RECOVERY CONSIDERATIONS

UKCloud operates a dual site multi-tenanted cloud and through the use of self-service controls, pooled resources and capacity planning processes, is able to offer customers seemingly limitless Compute as a Service. However, although we maintain a buffer of unused capacity at both sites to allow for failover of VMs using the geo-resilient option, on-boarding of new customers and the expansion of existing customer solutions, it would be wrong to think that UKCloud maintain capacity to cater for force majeure events.

It should therefore be considered that in the event of a major failure, resulting in failover to the remaining datacentre that:

- VMs with the geo-resilient option will have reserved capacity and will begin to be restarted within minutes of a site failure. (Please note that UKCloud may stagger the rate at which VMs are restarted)
- Any remaining capacity will be used to enable customers to order additional services. Customers who already have VDCs in the remaining data centre will be able to scale their environment autonomously
- Customers who do not have a VDC in the remaining data centre will have to formally ask UKCloud to provision them. It must be noted that UKCloud may not be able to provide uncontended capacity for all these customers as preference will go to existing customers in the remaining data centre.

ABOUT UKCLOUD

UKCloud has developed a range of cloud services designed specifically for the UK public sector, to help increase efficiencies, reduce costs, significantly improve procurement times and increase transparency. Our services are *easy to adopt, easy to use and easy to leave* to ensure that our customers remain in complete control with minimum risk. We were one of the first G-Cloud providers to achieve Pan Government Accreditation (PGA) up to Elevated OFFICIAL, and our services continue to achieve formal UK Government accreditations which make them suitable for all data at OFFICIAL (including OFFICIAL-SENSITIVE).

UKCloud's full offering consists of:

1. Infrastructure as a Service (IaaS) – seven offerings around Compute and Storage on demand

2. Software as a Service (SaaS) – offerings for email and collaboration

3. Platform as a Service (PaaS) – based upon Open Source Digital Application Platform and Hadoop which provides organisations the benefits of using a commodity cloud platform without the added management overheads

All of UKCloud's UK sovereign cloud computing services are hosted in one (or both) of our highly resilient Tier 3 UK data centres in Farnborough and Corsham. UKCloud services are delivered with leading technologies from UKCloud Alliance Partners: QinetiQ, VMware, Cisco, EMC and Ark Data Centres. The Cloud Alliance also provides a collaborative resource which drives innovation and technical product development, helping to continually improve UKCloud's offering to meet the needs of the UK public sector.

UKCloud is focused on providing cloud services in a more agile, secure and cost effective manner. We strive to deliver solutions that harness technology as a way to facilitate the changes that are needed to streamline processes and reduce costs to support the UK public sector and, ultimately, UK citizens and taxpayers.

MORE INFORMATION ▶

For further information about UKCloud and how we can help you, please send an email to info@ukcloud.com

UKCloud Ltd

A8 Cody Technology Park
Ively Road
Farnborough
Hampshire
GU14 0LX

+44 (0)1252 303300

info@ukcloud.com

www.ukcloud.com

UKC-GEN-101 • 07/2016