staffcircle

DISASTER RECOVERY POLICY

VERSION 1.6 LAST UPDATED: 14.11.2020

STAFFCIRCLE LTD. LCB DEPOT. 31 RUTLAND STREET. LEICESTER. LE1 1RE. UNITIED KINGDOM

Disaster Recovery & Resilience

Goals

- Ensure the StaffCircle Platform can recover from large scale zone level critical disaster scenarios within 24 hours
- Ensure the StaffCircle Platform is resilient and can recover from minor and medium level failures within 3 hours

Scope

This document covers the potential disaster scenarios that may affect the usage of the StaffCircle Platform as of 20 Oct 2020. This focuses solely on the technical aspects of the StaffCircle Platform including infrastructure and applications only.

Background and strategic fit

The SLA for the StaffCircle Platform is set at 99.9% uptime average across a calendar month as measured at the start of each following month.

StaffCircle platform is built natively on Microsoft Azure Cloud infrastructure and able to utilise global scaling and resilience options provided for by Microsoft.

Assumptions

Azure functionality should be focused on to ensure both disaster recovery and resilience.

Measurement

All external facing APIs and UX are measured in real-time by our customer facing status engine which is available at https://status.staffcircle.com

Production Resiliency State

Resource	Region(s)	Automatic Failover	Manual Failover	Notes
App Services	UK-S	☐	☒	manual failover/redundancy
SQL Databases	UK-S, UK-W	☐	☒	Replication enabled, but failover needs DNS changes to work properly
Virtual Machines	UK-S	☐	☐	No failover/redundancy
Redis Cache	UK-S	☐	☐	No failover/redundancy
Prod storage accounts	UK-S Primary, RA-GRS UK-W	☐	☒	
Shared storage accounts	UK-W Primary, RA-GRS UK-S	☐	☒	
Key vaults	UK-S + RA-GRS paired region	☒	☐	Built in failover to paired region (read only)
Service bus	UK-S	☐	☒	manual failover/redundancy
App insights	N-EU	☐	☐	No failover/redundancy, some telemetry saved locally until endpoint available again
Search service	UK-S	☐	☒	manual failover/redundancy
SignalR	W-EU	☐	☐	No failover/redundancy
Text Analysis	N-EU	☐	☐	No failover/redundancy
DNS	Global	☒	☐	100% SLA
SQL Backups	RA-GRS Paired region	☒	☐	7 Day PTR, 2 Week LTR
Storage Backups	None	☐	☐	None, no native backup provider, soft delete enabled for 7d
VM Backups (Recovery services vault)	GRS Paired region	☒	☐	Daily, 30d retention

Resources

Azure Availability Checklist

Service Bus

Search Api

Key Vault

DNS

Potential Failures

The following potential failures have been identified as key service affecting risks that we can mitigate against.

Failure	Result	Severity
Azure Region Outage	UK South down, complete lose of platform services in this region.	CRITICAL
Azure Multiple Region Outage	UK South and alternative region down, complete loss of platform services.	MINOR
Azure Storage Outage	UK South blob storage failure, documents, avatars and images unavailable.	CRITICAL
Azure SignalR Service Failure	In app notification services down	MINOR
Azure Service Bus Failure	All events across the platform are down	CRITICAL
Twilio SMS Services Failure	SMS notification services are down	MINOR
SendGrid Email Services Failure	Email notification services are down	MINOR
Azure Search Api Outage	Information feed unavailable	CRITICAL

Failure	Result	Severity
Azure Redis Cache Outage	Users unable to performance account administration including: 1. Activating new users 2. Reset password 3. Change username	CRITICAL
Azure Key vault Outage	Api services unable to reload configuration. Any restart of a service will cause a outage of the Api.	
Azure DNS Outage	Api or UI services unable to some or all tenants.	CRITICAL
Azure Cognitive Services	Article keywords unavailable for new or updated articles being published to the search index.	MINOR
Azure Availability Test Outage	Monitoring within the Azure environment unavailable	MINOR
Azure Application Insights Outage	Monitoring and diagnostics unavailable to the team	MINOR
Azure App Service Outage	One or more Api's and UI unavailable resulting in a loss of functionality or availability of the platform.	CRITICAL
Azure SQL Outage	One or more Api's data is unavailable resulting in a loss of functionality or availability of the platform with risk of permanent data loss.	CRITICAL
SCP Security Api Failure	User unable to login or authenticate against the platform. All functionality unavailable.	CRITICAL
SCP Platform Api Failure	Users unable to login or authenticate against the platform. All core functionality unavailable.	CRITICAL
SCP Directory Api Failure	Users unable to login or authenticate against the platform. All directory functionality unavailable.	CRITICAL
SCP Gateway Api Failure	Users unable to access all Api's, all functionality unavailable.	CRITICAL
SCP Platform UI Failure	User unable to load application outside of using cached version (PWA/SW). All new users will not be able to use the platform.	CRITICAL
SCP Events Api Failure	All events raised during Api operations would fail to be queued resulting in data inconsistencies across distributed data sets and no notifications.	MINOR

Failure	Result	Severity
SCP Events Workers Failure	All events raised would not be actioned but would be queued until service restored. Data inconsistencies would exist across distributed data sets and no notifications during failure.	CRITICAL
SCP Notification Integration Workers Failure	All inbound content and status updates from sms and email would fail, resulting lose of inbound emails and sms as well as outbound status loses.	CRITICAL
SCP Notifications Api Failure	Users unable to send or receive in app notifications until service is restored.	MINOR
SCP Documents Api Failure	Users unable to upload or retrieve documents, images, and avatars from storage until restored.	CRITICAL
SCP Feed Api Failure	Users unable to publish, or update content, access comments or article details. Tags unavailable across the platform.	MINOR
SCP History Api Failure	Users unable to access timeline functionality or event update timeline and audit records.	CRITICAL
SCP Reviews Api Failure	Users unable to access reviews and objectives functionality.	CRITICAL
Azure Portal Failure	StaffCircle Team unable to access portal to provide infrastructure support.	CRITICAL
Azure Active Directory Failure	StaffCircle Team unable to access portal or teams to provide support, or deploy new tenants.	MINOR
Jira Failure	StaffCircle Development Team unable to access ticketing system.	MINOR
Zoho CRM Failure	Users unable to access account and invoice information, StaffCircle Team unable to manage existing accounts and tickets.	MINOR
Azure DevOps Outage	Development team unable to access code or deploy new code	CRITICAL

Recovery and Resilience Options

Azure Region Outage

Fail over to a secondary region - Expected Recovery Time - **4 - 24 hours**

In this scenario all StaffCircle Platform Services would require to be failed over to a secondary instance in a alternate region. To mitigate this we will need a replica of our infrastructure in a cold but up to date state with data readily available.

Recommendation

At this stage due to the overhead of maintaining and cost of a complete cold environment in a secondary region, we would be best covering core components include App Services, SQL Databases and Service Bus in the secondary region, with other services added and configured during the fail over. It is expected that this process would take between 4 - 24 hours.

Azure Multiple Region Outage

Fail over to a third party - Expected Recovery Time - **Infeasible (Weeks/Months +)**

Due to the dependencies on Azure specific technology deploying to a third party provider would require rearchitecting large parts of the platform to move away from Azure specific technologies, as such this would be an unsuitable path for recovering within a quick enough timeframe.

For a long term Azure outage (weeks/months+) in this scenario all StaffCircle Platform Service would require to be moved to a alternative provider. To achieve this level of recover we require the following:

1. Backups of all data sets outside of Azure prior to outage
2. Commission the required infrastructure in the new provider
3. Rearchitect all code to remove dependencies on Azure specific technology
4. Update deployment pipelines to deploy and configure services
5. Update DNS outside of Azure to point to new servers

Azure Storage Outage

Fail over to alternative storage - Expected Recovery Time - **Immediate with reduced functionality to 24 hours + for full recovery**

Storage accounts are currently set up with Read-access geo-redundant storage (RA-GRS) - In the event of a failure in the primary region the storage account can be manually failed over to the secondary region in read only mode. This will immediately restore functionality to access existing data, but any changes cannot be made. A new storage account would need to be set up and data copied over in the event of a prolonged storage outage to enable write access.

In the event of a global Azure Storage outage data would be need to restored to a third party provider from any available backup, and changes to the platform made to read/write data to the new storage provider, with considerably longer recovery times expected

Recommendation

Ensure read-access geo-redundant storage on all production storage to allow for users documents to remain available and only provide a degraded service during the recovery period. Blob storage backups should be saved to a third party provider if possible.

Azure SignalR Service Failure

In this scenario all InApp notifications, some live data refreshing and tenant provisioning progress from the StaffCircle Platform would fail. We have the following options to mitigate this:

1. Allow Azure to resolve issue - Expected Recovery Time - **8.76 hours (Azure SLA)**
2. Commission and configuration a new SignalR instance for the StaffCircle Platform. This would be dependent on the SignalR service being available in our primary region or a alternative region - Expected Recovery Time - **4 hours**
3. Upgrade our Notifications Api to host the SignalR service - Expected Recovery Time - **24 hours**
4. Have redundant SignalR services running, with primary and secondary roles configured on consumers

Recommendation

Given the relatively low impact on platform services we should allow Azure to resolve any SignalR issues rather than take action ourselves, failing back to commissioning a new SignalR service if outside of the Azure SLA.

Azure Service Bus Failure

At present Azure Service Bus only supports Geo-Replication and Availability Zones in Premium Tiers. Based on the pricing model this would at base cost ~£500 a month per unit

We have alternative options of:

1. Commission a secondary Service Bus Namespace in alternate region and send copies of all messages to each namespace tagged with a unique identifier, and deduplicate messages on the receiving end
2. Move away from Service Bus to Azure Queue Storage, but this comes with downsides such as losing deadletter support
3. Upgrade to Premium Tiers to support Availability Zones, which ensure data is synchronised across three copies within a single zone with automatic failover, and geo-redundancy can be enabled for cross-region resiliency but this does **not** support syncing existing data, only metadata, so any existing messages (including scheduled) would be lost either until the outage is resolved, or permanently.
4. Build resilience into service bus clients to write failed messages to storage, and fail over manually to a secondary namespace in the event of a outage, losing any scheduled messages in the original namespace.

Recommendation

At present we should choose option 4 with the premium tier, an optimal solution would be a combination of options 3 & 4

Twilio SMS Services Failure

Allow Twilio to resolve the issue - Expected Recovery Time - **8.76 hours** (Twilio SLA)

In this scenario all SMS notifications from and to the platform would fail until this issue was resolved. Manual intervention would need to occur to re-queue sms notifications and retrieve inbound messages.

SendGrid Email Services Failure

Allow SendGrid to resolve the issue - Expected Recovery Time - **8.76 hours** (Sendgrid SLA)

In this scenario all Email notifications from and to the platform would fail until this issue was resolved. Manual intervention would need to occur to re-queue email notifications and retrieve inbound messages.

Azure Search Api Outage

As per https://docs.microsoft.com/en-us/azure/search/search-capacity-planning there is no built in mechanism for disaster recovery.

In a outage scenario the feed would be unavailable, and all new articles would not be published. Azure provides no built in mechanism to handle this and advises this to be covered by out code. In this scenario we would have the following options:

1. Allow Azure to restore the service and rebuild indexes for all tenants.

2. Commission a new Azure Search Service, configure the platform to use this service and rebuild indexes for all tenants.
3. Commission a secondary Azure Search Service in a alternative region and have a instance of the article worker sync data to this.

Azure Redis Cache Outage

In a outage scenario the platform would lose the ability to send or valid any tokens sent to users resulting in the lose of activation, forgot password, change username and access to secure data functionality. To mitigate this scenario we would have the following options:

1. Upgrade to Premium Redis Cache allowing for Geo-replicated cache - **2 - 4 hours**
2. Remove the Redis Cache and replace with Table Storage which can be replicated easily - Excepted Recover Time - **Unknown**
3. Simply create a new Redis Cache in a secondary region and reissue and tokens manually - Expected Recover Time - **2 - 4 hours with lose of issued tokens**
4. Upgrade to Standard to recieve 99.9% SLA - Expected Recovery Time - **8.76 hours**

Recommendation

We should look to move away from Azure Redis Cache for token generation and expiration. At present we are utilising very little functionality that it provides, and this is a major point of failure for the platform. We should look to move to Table Storage for tokens with a Azure function to maintain expired tokens. In the interim, we should simply create a new Redis Cache in a secondary region.

Azure Key vault Outage

In this scenario all platform Api's would lose configuration after restarting resulting in failure to serve clients data. This would cause failures across the platform impacting our SLA. To mitigate this Azure Key Vaults have multi-region fail-over built in, with no intervention required from the StaffCircle Team, however while failed over only read operations are permitted so new operations such as inserting connection strings for new tenants during tenant provisioning will fail. This process is documented here https://docs.microsoft.com/en-us/azure/key-vault/key-vault-disaster-recovery-guidance.

In the scenario where this was unavailable we would need to replace the key vault configuration mechanism across our services and reissue SQL credentials for all tenants, as well as any other credentials stored in keyvault - Expected Recover Time - **24 hours**

Azure DNS Outage

Azure offer a 100% SLA for DNS services. In the event of a DNS failure outside of this we would have the following options to mitigate this:

1. Update DNS directly at domain source to use alternative name servers and reroute directly to Azure resources or traffic managers - Expected Recovery Time - **48+ hours**

Recommendation

Backup all DNS zones at least daily to a third party provider to ensure nameservers can be changed and DNS entries can be restored to another provider in the event of a global outage

Azure Cognitive Services

Allow Azure to resolve this issue - Expected Recovery Time - **8.76 hours** (Azure SLA)

In this scenario, article keyword would be lost during the outage period. We would not aim to mitigate this outage due to the low impact this will have on the platform usability.

Azure Availability Test Outage

Allow Azure to resolve this issue - Expected Recovery Time - **8.76 hours** (Azure SLA)

In this scenario, we would lose internal monitoring within the Azure platform. To mitigate this we will:

1. Fail over to UpTimeRobot only
2. Disable internal alerts during outage if required

Azure Application Insights Outage

Allow Azure to resolve this issue - Expected Recovery Time - **8.76 hours** (Azure SLA)

In this scenario, we would lose internal monitoring within the Azure platform. Some logs should be cached locally and send to app insights once connectivity is restored (see this documentation). To mitigate this we will:

1. Fail over to UpTimeRobot only
2. Disable internal alerts during outage if required

We also have the option to investigate a third party logging platform to hold logs externally to aid diagnosis outside of Azure Application Insights.

App Service Outage

In this scenario, one or more Api's and UI unavailable resulting in loss of functionality or availability of the platform, directly affecting our SLA. To mitigate this we have the following options:

1. For single instance failures not affecting a region - implement a minimum of 2 instances of each app service running in the Production Environment primary region - Expected Recovery Time - **30 minutes**

2. Implement a cold instance in a secondary region with traffic managers to fail over in outage scenarios - Expected Recovery Time - **2+ hours**
3. Implement and deploy code to alternative provider with traffic managers to fail over in outage scenarios - Expected Recovery Time - **8+ hours**

Recommendation

Options 2 is the preferred route for disaster recovery with option 1 for additional resilience in a single region. Additional to this we will need to consider:

1. Maintaining the cold instance of the app in the secondary region (options of cloning, PS scripts or adding to deployment pipeline are available)
2. Managing configuration in the secondary region (this should ideally continue to point to primary region)
3. Maintaining multiple region instances during deployments
4. Certificate management with traffic managers

Azure SQL Outage

In this scenario, one or more Api's data is unavailable resulting in a loss of functionality or availability of the platform with risk of permanent data loss. To mitigate this we have the following options:

1. Periodic built in backups so we can always recover the data with point in time recovery & long term retention policies
2. Enable Geo-replication of Azure SQL Databases through a fail over group to cold secondary region and enable fail over group, this option risks data in transit being lost - Expected Recovery Time - **15 minutes**
3. Periodic snapshot replication to alternative region, this option risks data from the last snapshot until the outage being lost and will require manual intervention - Expected Recovery Time - **4 hours**

Recommendation

Based on this options 1 & 2 should be configured.

SCP API/UI Failure

In this scenario, one or more modules within the system would become unavailable degrading functionality or causing a outage. In the case of the following Api's this would result in a outage across the platform due to their dependencies:

1. Security Api
2. Platform Api
3. Directory Api
4. Gateway Api
5. Platform UI

Recommendation

For all services, option 1 & 2 of Azure App Service Outage should be implemented to ensure maximum up time. Alongside this we should also implement the following:

1. Auto-healing through web.config
2. Auto-scaling to ensure capacity of each service is adequate

SCP Workers

In this scenario, all events raised would not be actioned but would be queued until service restored. Data inconsistencies would exist across distributed data sets and no notifications during failure. To mitigate this we have the following options:

1. Use a deployment slot to store the last known good version as part of the deployment pipeline (staging slot). This can be re-enabled manually to ensure services stay functioning until the issue can be resolved - Expected Recovery Time - **30 minutes**
2. Manually fail over to a cold version in a secondary region - Expected Recovery Time - **30 minutes**

Recommendation

Based on this both options 1 & 2 should be configured.

DevOps Unavailable

Allow Devops to resolve this issue - Expected Recovery Time - **8.76 hours** (DevOps SLA)

Azure Portal Failure

Allow Azure to resolve this issue - Expected Recovery Time - **8.76 hours** (Azure SLA)

Azure Active Directory Failure

Allow Azure to resolve this issue - Expected Recovery Time - **8.76 hours** (Azure SLA)

Jira Failure

Allow Atlassian to resolve this issue - Customer Agreement

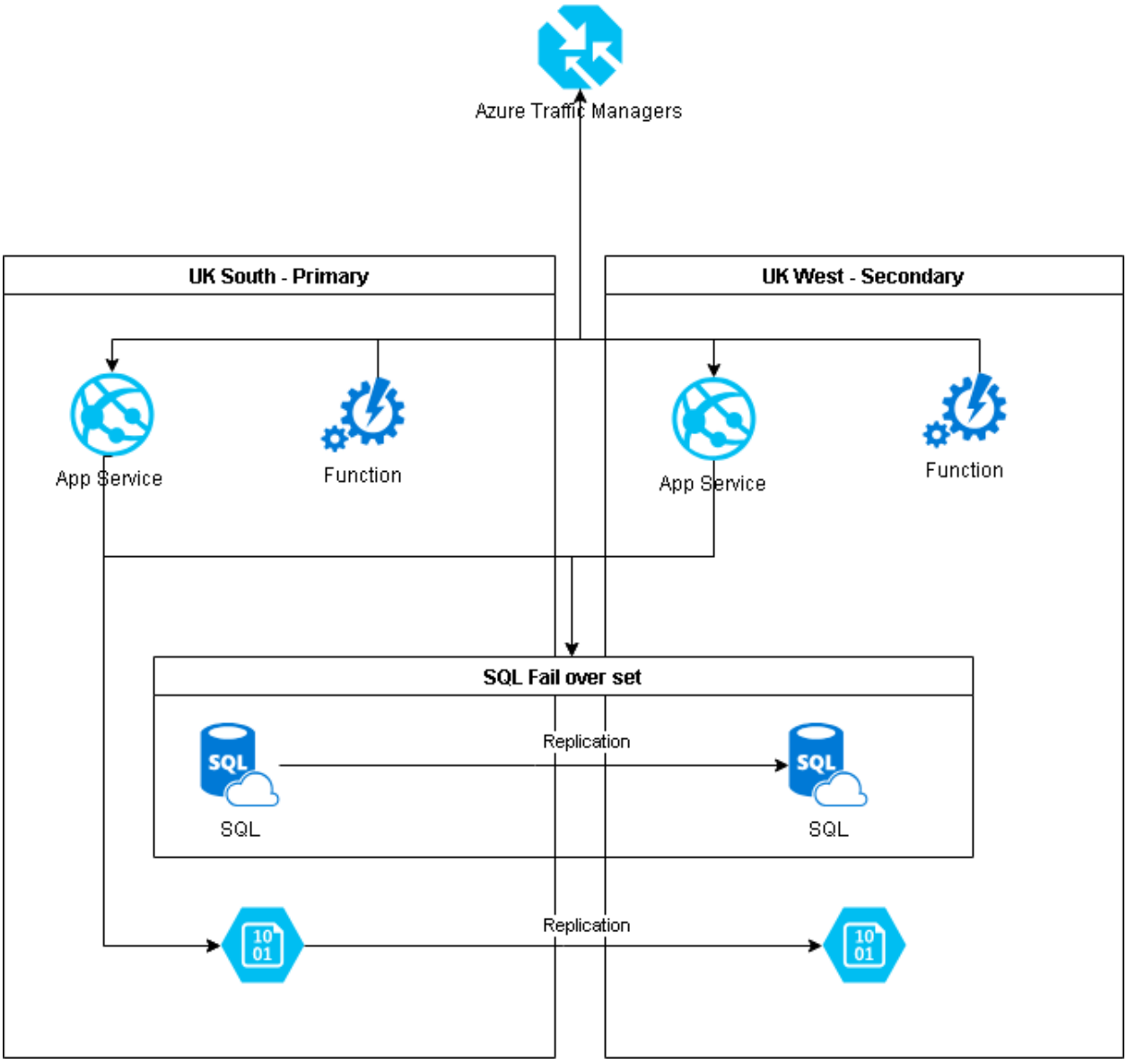
Zoho Subscription Engine Failure

Allow Zoho to resolve this issue - Terms of Service

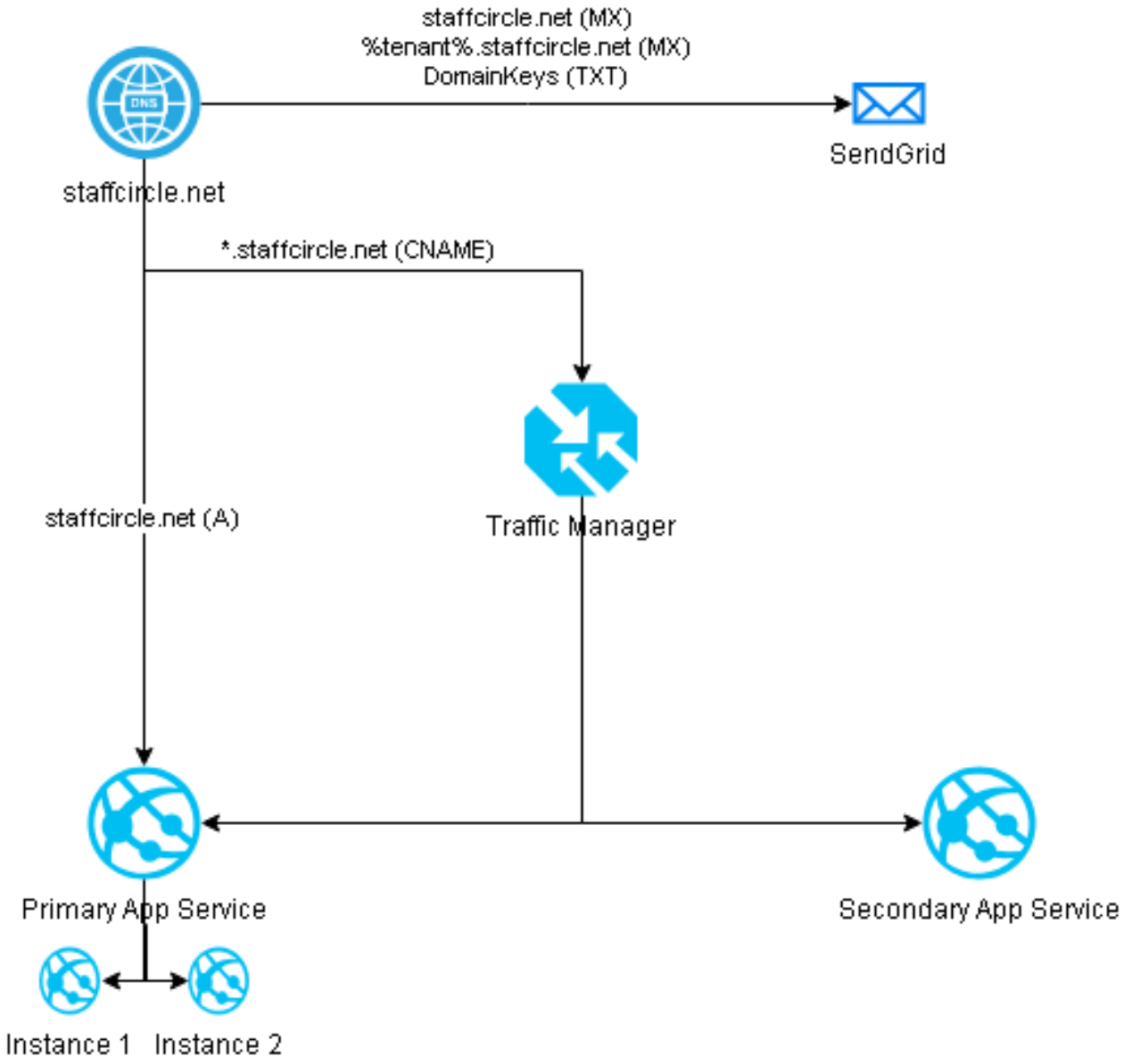
Requirements

#	Title	Task	Importance	Dev	UAT	Prod
1	Azure Storage	Configure Read-access geo-redundant storage across all production storage containers	MUST	DONE	DONE	DONE
2	Azure Service Bus	Configure Events Api to handle fail over	MUST	DONE	DONE	DONE
3	Azure Service Bus	Migrate SCP Events to use Azure Queue Storage	SHOULD	DONE	DONE	DONE
4	Azure Search Api	Migrate Activity Feed search and index functionality to the Feed Api	SHOULD	DONE	DONE	DONE
5	Azure Search Api	Migrate Activity Feed search and index functionality to the Feed Api using Lucene	SHOULD	DONE	DONE x	DONE
6	Azure Redis Cache	Migrate token functionality to Azure Table Storage	SHOULD	DONE	DONE	DONE
7	Azure DNS	Configure Azure Traffic Managers for all DNS records routing to Azure Resources	MUST	DONE	DONE	DONE
8	Azure App Insights	Investigate a third party logging platform	COULD	DONE	DONE	DONE
9	Azure App Services	Configure multiple app instances within primary region	MUST	DONE	DONE	DONE
10	Azure App Services	Commission cold instances of app services in secondary region	MUST	DONE	DONE	DONE
11	Azure App Services	Configure auto-healing across all app services	MUST	DONE	DONE	TBC
12	Azure App Services	Configure auto-scaling across all app services	COULD	DONE	DONE	DONE
13	Azure Functions	Configure deployment slot for event workers	MUST	DONE	DONE	DONE
14	Azure Functions	Configure cold instance in secondary region	MUST	DONE	DONE	DONE
15	Azure SQL	Configure nightly backups of all SQL databases to storage	MUST	DONE	DONE	DONE
16	Azure SQL	Configure Geo-replication on all SQL databases to secondary region	MUST	DONE	DONE	DONE
17	Azure App Services	Enable app caching on all app services	MUST	DONE	DONE	DONE
18	Development	Tear down additional resilience on development environment	MUST	DONE	DONE	DONE

User interaction and design



DNS and Traffic Manager Configuration for Platform UI



Updates Required

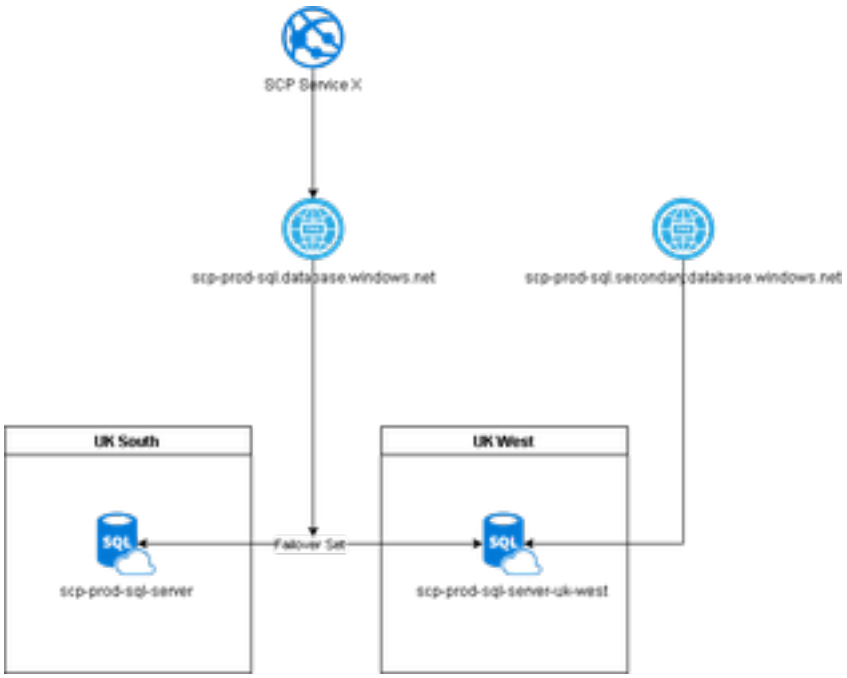
Area	Azure Resource	Update	Notes
DNS	staffcircle.net	Add new *.staffcircle.net record	This should point to app then be moved over to traffic manager
App Service	scp-prod-app-platform-ui	Add new *.staffcircle.net domain and assign SSL	
DNS	staffcircle.net	Move tenant MX recorders to staffcircle.net DNS set	

SendGrid	scpprodsendgrid	Verify staffcircle.net against production SendGrid account	
Traffic Managers	scp-{env}-platform-ui scp-{env}-gateway-api scp-{env}-security-api scp-{env}-directory-api scp-{env}-documents-api scp-{env}-events-api scp-{env}-feed-api scp-{env}-history-api scp-{env}-notifications-api scp-{env}-platform-api scp-{env}-reviews-api	Add new traffic managers	
DNS	scp.staffcircle.net	Update DNS to point to traffic managers	
DNS	staffcircle.net	Update DNS to point to traffic managers	
DNS	staffcircle.net	Remove procomm name server record Remove us name server record Remove pitacs name server record	
DNS	pitacs.staffcircle.net procomm.staffcircle.net us.staffcircle.net	Remove resource	

SQL

Fail over group

Using a Azure SQL Fail over Group will allow all databases to be managed in a single fail over group. This will allow for replication to be configured at Azure SQL Server level essentially pairing two server instances. During a failover we can initiate a failover of all SQL Databases without changing configuration of the application. This has the disadvantage of a fail over scenario affecting all databases rather than independent databases.



Backup

Azure offers rich back up facilitates allowing for retention of weekly backups for a specified period of time alongside the point in time backup for the past 7 days.

Updates Required

Area	Azure Resource	Update	Notes
Azure SQL Server	scp-prod-sql-server-uk-west	Create new SQL Server instance	
Azure SQL Server	scp-prod-sql-elastic-pool-uk-west	Create new Elastic SQL pool	

Azure SQL Server	scp-prod-sql-server	Configure fail over group to include all databases	• We should ensure we set to manual fail over at this stage
Keyvault Configuration	scp-prod-keys-*	Update all keyvault configurations to point to new fail over DNS record	
SCP Services Config	scp-prod-app-platform-api scp-prod-app-security-api	Update configuration in DevOps to use new fail over DNS record	
SCP Services	scp-prod-app-*	Restart all services	
Azure SQL Server	scp-prod-sql-server	Configure weekly retention of backups	

Azure Service Bus