Certified Solution Architect – Professional

Contents

[**Section 3: Designing Solutions for High Availability and Business Continuity** 1](#_Toc29029185)

**12/31/2019**

# **Section 3: Designing Solutions for High Availability and Business Continuity**

* Business continuity and Disaster Recovery
* Resiliency and Fault Tolerant
* Redundancy and high availability
* Cost optimization
* Performance
* Security
* Monitoring
* Saleable and Elasticity
* Ease of deployment
* Migration and Hybrid architectures

Fault Tolerant: Ability of a system/application/infrastructure to be operational even if one or more system fails is call fault tolerant. A fault tolerant system should be the one which can recover/failover with minimal or no human interventions on a even of any failure.

AWS recommended to create a library of own AMI, with own standard & best practice. Ensure the AMI is up to date and stable for production usages.

Floating IP (elastic IP) = the IP that can be shift from the primary instance to secondary on an event of any failure.

Create regular EBS snapshots – EBS are region specific, to be used in another region we need to copy the snapshot to the desired region before creating the EBS volume from that snapshot. EBS are region specific and EBS volume are AZ specific.

Multiple site architecture: In this approach the goal is to have two or more independent copy of each application stack into two or more availability zone (site). Every application tier need to have redundant copy , in case of failure the traffic can be route to the alternative availability zone (site) – this can further be made more effective by introducing elastic load balance which can automatically route traffic between AZ, OR having a route53 health check created to route traffic between different region, alternatively one can also use elastic ip for routing traffic, so that when the instance fails the elastic ip can be remapped to another running instance in same/different AZ. Implement Autoscaling to address flatulating production load.

**01/01/2020**

Disaster Recover – Recovery plan from a system failure to resume business continuity without impacting business operation.

|  |  |
| --- | --- |
| **Benefit of having AWS as alternative site** | **Counter Argument** |
| Don’t need to negotiate contract with other provide in another region. | Avoiding vendor locking – have a multi-vendor/multi-cloud landscape. |
| One can use the same underline AWS technology across region | Need to have trained staff in more than one cloud technology – will be cost intensive to maintain multi talent pool. |
| One can use same tool/AWS build artifacts/API across region | One need to maintain different tool/build artifacts/APIs for different cloud. |

**Recovery Time Objective (RTO) –** Time taken after a disruption to restore a business process to its service level agreement as define by the operational level argument (OLA).

**Recovery Point Objective (RPO)** – Acceptable amount of data loss measure in time. For example, If the disruption occurs at 12 Noon and RPO is 15 minutes, the system should be successfully able to restore all its transactions which have occur till 11:45AM.

Note: RPO and RTO is defined by the financial impact to the business when systems are unavailable.

* Disaster Recovery Plan for Commute from AMI backup aspect:
* AMIs can help in restoring failed instance quickly.
* AMIs can be used within the context of the autoscaling to scale out DR site on an event of primary site failure.
* AWS recommended to maintain a library of preconfigure AMIs, including the application stack, so that on an event of a disaster these can be used to create new (EC2) instances to replace the old/failed once.
* Keep the AMIs available on DR sites.
* Disaster Recovery Plan for Storage from S3 Fault tolerance aspect [FOR OBJECT STORAGE]:
* Can be used as primary object storage.
* Highly redundant, with 99.99999999999 (11 9’s) redundancy rate. Objects stored within S3 buckets are automatically backed up in multiple facilities within the region.
* Additional security can be provided within S3 bucket for ensure high degree of data retention through MFA for delete operation, versioning, cross region replication – for auto backup in DR site, bucket level policies & object level policies.
* One can also use S3 Glacier storage class for archiving/storing data at very LOW COST, however RTO needs to be set to meet the recovery time as object archive in S3 Glacier takes longer time (usually 3-5 hours).
* Disaster Recovery Plan for Storage from EBS Fault tolerance aspect [FOR BLOCK STORAGE]:
* Point-in-Time snapshot can be created to backed up EBS content into S3 bucket.
* Snapshot created can be used to create a new EBS volume & then connect it the EC2 instance to replace failed EBS volume.
* EBS volume data are stored in different system within the Availability Zone (AZ), this provide protection from single system failure within AZ.
* Disaster Recovery Plan for Storage from Storage Gateway prospective
* Storage gateway provides an easy means to backup/store on-premises data on to AWS cloud. Customer can download VMware image from AWS console, and install a VM instance from the same on premises which will provide NFS/iSCSI interface for backup/storage.
* Storage Gateway – File Gateway (NFS interfaces) [OBJECT STORE]: files are asynchronously backed up into S3 bucket, there is a one-to-one mapping between on premises files and S3 objects. User can ALSO access the S3 objects(files) from the S3 bucket. S3 sub resources like lifecycle policy, versioning etc. can be implemented on the uploaded file(objects).
* Storage Gateway – Volume Gateway - Cached | Storage Mode (iSCSI interface) [BLOCK STORE]: Unlike File gateway, in case of the volume gateway, on premises files are stored into S3 bucket as BLOCK store. There are two possible operating modes – *Cached Mode* where frequently use files are cached on-premises while its asynchronously backed up on S3 bucket as volume store, in *Storage Mode* there is NO local file store everything is backed up on S3 as EBS snapshot. One can’t access the files stored in S3 directly – to access the files one need to create EBS volume from form the snapshot (*in case of cache mode, first snapshot needs to create from the volume store prior to creating EBS volume*).
* Storage Gateway – Tape Gateway: it provides an iSCSI VTL (Virtual Tape Library) interface to backed up on premises files into virtual tape data store within S3 bucket or can be achieve into S3 Glacier. It consists of virtual media changer, virtual tape drive, and virtual tapes.
* Disaster Recovery Plan for Storage from AWS Import/Export prospective:
* If large amount of data needs to migrate (backed up) quickly into AWS one can use AWS import/export feature – solution like snowball, snowball edge and snowmobile can be leverage for the same.
* Rule of thumb – “*if the data takes more than week to transfer over the available connection (VPN/Direct Connection/Open internet), then better use AWS import export instead*”.
* Disaster Recovery Plan for Storage from VM Import/Export prospective:
* Ease means to transfer VM image (hypervisor base virtual image) from/to AWS.
* This can be helpful in creating DR site at AWS or on premises.
* NOTE: VM Export is ONLY AVAILABLE to those instances which are initially brought in using VM import feature. AMI based EC2 instance cannot be exported using AWS Export feature.
* Disaster Recovery Plan for Storage from RDS Fault Tolerance aspect:
* RDS Multi-AZ: Primary-to-Secondary synchronous data replication within a region. On an event of any failure the primary DB instance DNS entry will be automatically swapped with secondary DB instance. Help in archiving zero downtime during system patching/upgrade.
* Read Replicas: For read intensive applications, a read-only DB instance can be provide with asynchronous data replication from the primary instance. On an event of any failure read replicas can be promoted
* Automated Backup: Automated backup in conjunctions with transaction logs to help recover failed DB instance with RPO up to 5 min.
* Manual Snapshot: this are necessary for backing up a DB instance OR restoring a DB instance in another region.
* Can be synced from on premises database to RDS and vise-versa – failover from the on-premises to AWS OR AWS to on-premises is possible.
* RDS instance size can be upgraded to a bigger size BUT can’t be downgraded to a smaller size.
* For DynamoDB one can copy the data into a S3 bucket and using data-pipeline it can be copied to another region DynamoDB. Also, one can use cross-region replication of the DynamoDB to replicate the data into another region DynamoDB table. For Sync operation (preparation phase), the dynamo DB can be created with lower read-write capacity, during recovery phase the capacity can be altered to meet the production demand.
* For redshift database: During the preparation phase one can create snapshot of the database in to S3 within same/different region, during the recovery phase redshift cluster can be (re)created from the copied snapshot.

**01/04/2020**

* Disaster Recovery Plan for Networking, on an event of a disaster the ability to quickly shift the network settings from the production site to the DR site is very much required.
* Route53 AWS managed service helps in quickly change and restore network connectivity (routing) on an event of a failure.
  + - Route53 is an AWS managed, highly available and highly scalable global service.
    - Route53 automatic Health-check configuration can route incoming traffic to healthy site.
* Using Elastic IP (floating IP): these are the static IPs that can be configured to the system, on an event of any disaster these IPs can be swap with the DR instances to quickly resume the production workload.
* Elastic Load balancing:
  + - It can load balance the traffic within a given region
    - Elastic IP addresses can be allocated to the ELBs, which make the transition simple on an event of any disaster. (DNS can configure to pass the load to the elastic IPs, which can load balance the traffic among healthy instances).
* Amazon VPC: On an event of a disaster Amazon VPC can help in extending the network topology to the cloud. This is mostly applicable in extending the on-premises application to the cloud during disaster.
* Amazon Direct Connect: This helps in connecting the AWS cloud network to the on-premises network by providing a low latency, high bandwidth dedicated physical connection.
* Disaster preparation phase: Before disaster strikes, the stage where the disaster/backup plans are made are called disaster preparation phase.
* Disaster recovery phase: After disaster strikes, the steps that needs to be taken to recover from the disaster is call disaster recovery phase.
* Different Strategies for Disaster Recovery are: These strategies can be mix and match to find a right solution that meets the RPO and RTO expectations.
* **Backup & Restore** 
  + - **Backup solution:** Snapshot | Object (file copy) | Volume copy.
    - **Storage solution:**  S3|Glacier | Tiring solution (according to the usages storage class will be selected).
    - **Data transfer solution:** Over internet using custom solution| AWS File Gateway | AWS Volume Gateway – Stored/Cached mode | AWS Tape Gateway | AWS import/Export | AWS VPN Connect |AWS Direct Connect
    - **Key Consideration:** 
      * Select appropriate means: tools/services for backup/storage/data transfer solutions.
      * Define retention policy – till when a backed-up objects/files/snapshots needs to be stored? How soon they need to be made available? Can they be archived?
      * Ensure adequate security measures are applied to secure backed up contents.
      * Plan regular game-day/dry run to ensure the data are correctly backed up and can be restore correctly when needed.
* **Pilot Light:** DR strategy where a minimal version of the cloud is always running on the DR site. During preparation phase: The core component will be identified and a minimal version of the solution will be running on the DR site and will be in Sync with the production. During restore phase the other components on the solution can be quickly provisioned and the production traffic will be routed to the DR site. RDS site may be upgraded to meet the production traffic.
  + - **Key Consideration** 
      * Backed up AMI can be use to provision the remaining component.
      * Running Instance size should be upgraded to meet the production demand.
      * Add resilience to the DR Site: Plan for fallout instance failure.
* Warm Standby:
* Multisite
* Data Replication and Self-Healing