# Design Document-RStudio-for-SWB

**Design Document-RStudio-for-SWB** **1**
1.0 Introduction 3
1.1 Customer Impact 3
2.0 Requirements 3
2.1 Additional Document Reference 9
2.2 Out of Scope 9
3.0 Solution 9
3.1 Security Certificate Lifecycle 11
3.2 Secure (SSL) connection End-to-End 12
3.3 Client Role IAM Changes 12
3.4 Application Load Balancer 13
3.5 Route53 Changes 14
3.6 RStudio AMI 14
3.7 RStudio CFT 15
3.8 SWB Hosting Account Configurations 15
3.9 RStudio Active Counter Tracking (per Account) 15
3.10 SWB DynamoDB Change details 16
3.11 SWB Provision Environment Workflow 16
3.12 SWB Terminate Environment Workflow 18
3.13 Secure “Connection” Action for User for RStudio 19
3.14 Disallow Access of Researchers to connect to other RStudio Instances 19
3.15 ALB Routing Rules Changes with RStudio Lifecycle 19
API used: 20
3.16 CIDR Security Considerations 20
3.17 Cost Tracking for Provisioned Assets 21
3.18 Packaging and Delivering RStudio in a separate RL Hosted Repo 21
3.19 Packaging and Delivering RStudio as AWS Marketplace option 21
3.20 Ongoing Support, upgrades and enhancements considerations 21
Security Impact 23
Story Breakdown 24
Assumptions 24

**Appendix** **25**

## 1.0 Introduction

RStudio is a very popular software used by the Scientific Research Community and supported by AWS Service Workbench (SWB). The existing model of packaging and deploying RStudio for end customers had usability challenges hence an effort was put to evaluate the friction points and come up with an enhanced design to tackle the same.

### 1.1 Customer Impact

Researchers use RStudio very commonly in their day to day efforts. While RStudio is a popular product, the process of installing RStudio securely on AWS Cloud and using it in a cost effective manner is a non-trivial task specially for Researchers. With AWS SWB the goal is to make this process very simple, secure and cost effective for Researchers so that they can focus on “Science” and not “Servers” thereby increasing their productivity.

## 2.0 Requirements

See RStudio Requirements document.

The primary assumptions and guidelines as per Requirements Review is following:

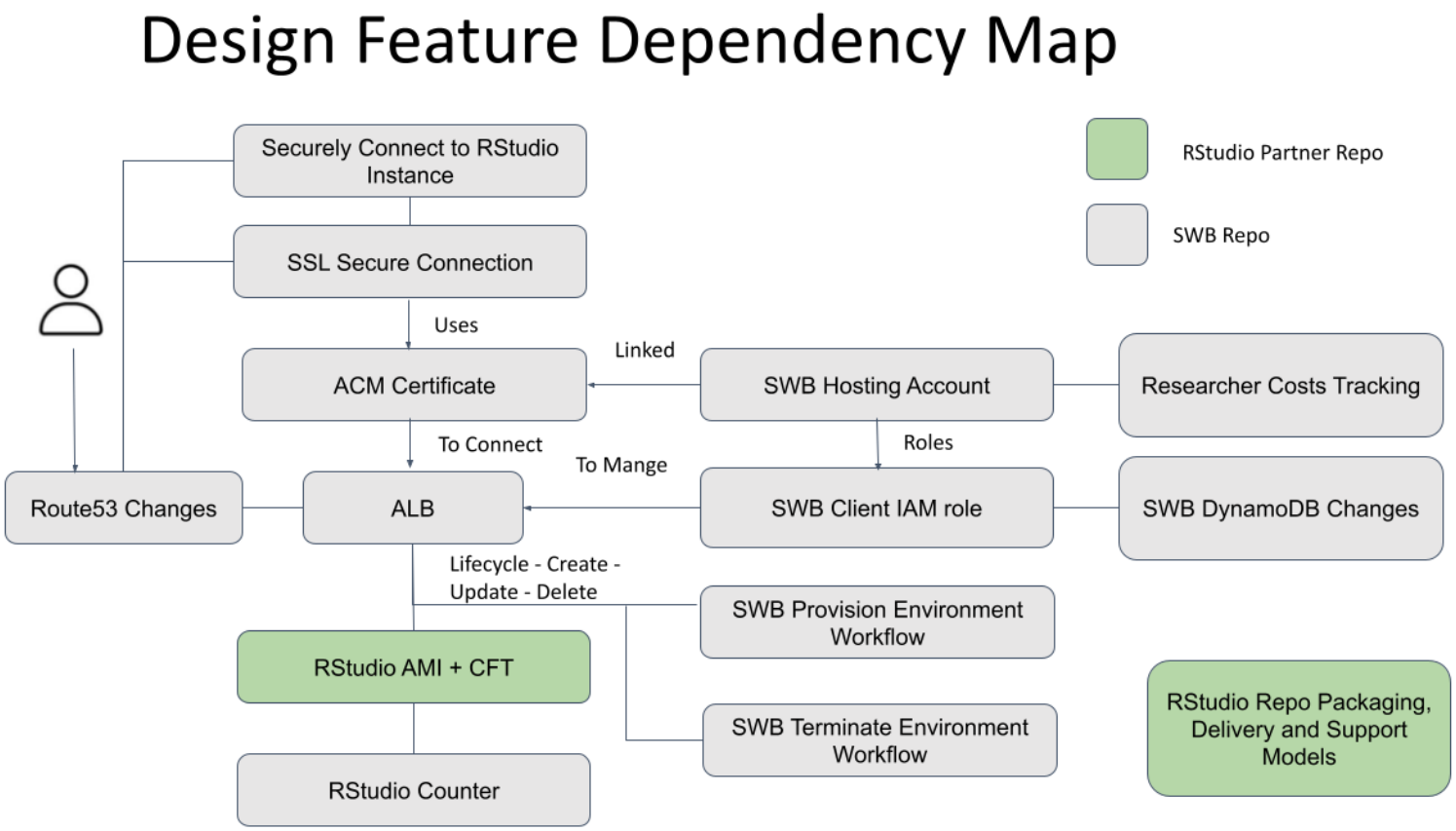
1. Provide a better end customer experience on RStudio usage
2. Replace current RStudio version from SWB with this new implementation
3. Deliver the new RStudio from a separate partner hosted Repo (RL) and not in the base SWB Repo hosted by AWS, which will link to new Partner Repo for RStudio
4. The new RStudio changes will still have dependency between SWB Changes and RStudio features hence both Repo’s and feature releases will have to be synchronized for first release and with all subsequent upgrades.
5. Current implementation still assumes customers will continue with a custom DNS for their SWB installation and that is used for ALB + RStudio configurations on routing

Currently Supported SWB Product Versions:

- To be added

**Snap-shot of High Level "Design Feature" enhancements being addressed in this Design Document for RStudio use in SWB.**

**SWB Base Repo:** All existing SWB Code and existing templates
**RStudio Repo:** All RStudio Related files (Templates and scripts) that will be separately hosted.
**Customer Setup:** All changes related to customer specific environments where SWB is installed



| Design Feature No. | RStudio Design Feature | Original Design Approach | New Design Approach | Impacted Component (RStudio Repo, SWB Base Repo, Customer Setup?) |
|---|---|---|---|---|
| 1 | User Generated Security Certificate for SSL Secure Connections to RStudio. Move from External certificates to AWS ACM | Users have to create a certificate (like LetsEncrypt) and use it with RStudio EC2 Instance with NGINX server. This creates complexity in the Certificate lifecycle. Complex for end users to create, maintain and renew. The RStudio AMI also needs to manage the Certificate lifecycle | Bring in a shared AWS ALB (Application Load Balancer) and used AWS ACM certificates for each Hosting Account to simplify the Certificate Management Lifecycle | **Customer Setup**<br>**SWB Base Repo**: This process has to be automated while the Hosting Account is Configured. Needs discussion, **currently manual process**.<br>**RStudio Repo**: Expects this as dependent parameter for ALB provisioning |
| 2 | SSL Secure Connection. | Create an SSL connection with Nginx Server on RStudio EC2. Related to custom certificate management. | Replaced with ALB at an Account level and shared by all RStudio Instances in an account. User Portal to ALB connection secured by ACM. For ALB to RStudio EC2 secure connection, use unique self-signed Certificates to encrypt connection per RStudio. | **RStudio Repo** : AMI<br>**SWB Base Repo:** Bootstrap script |
| 3 | Client Role (IAM) changes in SWB | Client role is provided necessary permissions for setup purposes. | **CrossAccountEnvMgmtRoleArn**<br>**CrossAccountExecutionRoleArn**<br>Current roles have most elevated ec2-access levels capable of creating ALB and its components. | **SWB Base Repo**: Associated CFTs |
| 4 | ALB Design | Not existing in original design. | Shared ALB design per Hosting Account to be shared between Projects. Details in Design document on creation, deletion | **RStudio Repo** Code & Packaging<br>**SWB Base Repo** |

| | | | | |
|---|---|---|---|---|
| | | | lifecycle. Each ALB is expected to cost about $20-50 monthly in shared mode with average use. API model used to create/delete ALB. | **See sections 3.4, 3.11, 3.12, 3.15 and 3.16** |
| 5 | Route 53 Changes on Main account | A CNAME record gets created with the EC2 DNS name | A CNAME record gets created with the ALB DNS name | **Customer Setup:** Main account |
| 6 | RStudio AMI | Embedded with Certificate details. Related to custom certificate management. | Independent of user provided Certificate details. Also AMI has been enhanced to include the following:<br>Self signed SSL and additional packages (as commonly requested by researchers) are baked into the AMI. | **RStudio Repo** |
| 7 | RStudio CFT | Original one to be removed from SWB. | Added a new output to indicate the "Need ALB" flag. Also create the new target group to which the ALB can route requests | **RStudio Repo**<br>**SWB Base Repo**<br>**See section 3.18** |
| 8 | SWB Hosting Account Configuration | Did not have to provision certificate AWS ACM | Manual process to set up a certificate in a new hosting account. | **SWB Base Repo** |
| 9 | Provisioned RStudio per Hosting Account Active Count Tracking | None | Needed to ensure ALB is created first time when RStudio is provisioned and deleted after last RStudio is deleted to optimize cost overheads of ALB. | **SWB Base Repo** - DynamoDB changes and workflow changes. |
| 10 | SWB DynamoDB Table Changes | DynamoDB used for all Tables by SWB. | Modifications needed to support the new design. New Table added to store additional parameters related to RStudio and ALB change. Explore using the existing DeploymentStore table in SWB design. More details in the document below. | **SWB Base Repo** - DynamoDB changes |

| | | | | |
|---|---|---|---|---|
| 11 | SWB Provision Environment Workflow | Standard design | Additional Step added to check if “Workspace Type” needs ALB and if it does, when check for ALB and either create or pass reference to existing one. | **SWB Base Repo** - Workflow changes |
| 12 | SWB Terminate Environment Workflow | Standard design | Additional Step added to check if last Active RStudio being deleted and if so, also delete ALB to reduce idle costs. | **SWB Base Repo** - Workflow changes |
| 13 | Secure “Connection” Action from SWB Portal to RStudio instance | To ensure each RStudio has a secure connection for each user a unique connection URL is generated during the user session that is valid for a limited period. | The same design of the original implementation is preserved. Internally the routing is managed through ALB but the concept remains the same. This ensures users do not have to remember userid/password for RStudio and a secure connection is always made available. | No changes done. |
| 14 | Secure “Connection” from SWB Portal disallowing other users from accessing RStudio resources given shared ALB | NA | Using the design feature (Step-13) ensures that even post ALB the connection for a User (Researcher and PI) is still restricted to their provisioned RStudio only and they cannot access other Researchers Instances. The unique connection is system generated using User to RStudio mapping uniquely. | Evaluated with Source IP and Host Header Forwarded to specific Target group. ( access is restricted to only allowed IP ) |
| 15 | ALB Routing Rules for RStudio secure connections given shared nature. | NA | Every time an RStudio is created or deleted, changes are made to ALB rules to allow a secure connection between the User session and the linked RStudio. Same rules are cleaned up during RStudio delete lifecycle. These changes to ALB routing rules are managed from SWB code under Workflow customizations. (Step-11 and 12) | **SWB Base Repo** |

| | | | | |
|---|---|---|---|---|
| | | | using APIs. | |
| 16 | RStudio Configuration parameters related to CIDR | Original design allows only whitelisted IP addresses to connect to associated RStudio instances - this can be modified also from configurations. | RStudio CFT should take CIDR as Input Parameter and pass it through as an Output Parameter for the SWB to take it and create the ALB Listener Rule.<br><br>SWB code will take CIDR from RStudio CFT output, subsequently update the ALB Listener Rule with respective Target Group. | |
| 17 | Researcher costs tracking | Original design had RStudio costs tracked for Researchers. Custom certificate costs were not tracked, if any. | In the new design RStudio costs are tagged and tracked per researcher. ALB costs are treated as shared costs for the Hosting account. | **SWB Base Repo**<br>**RStudio Repo** |
| | | | | |
| | | | | |
| | Packaging and Delivery Models | | | |
| 18 | RStudio Packaging and Delivery for new customer - Repository Model | Bundled with standard AWS SWB repo and installed | New model for RL to create a separate Repo and host RStudio with associated documentation and templates for customers to use. | |
| 19 | RStudio Packaging and Delivery for new customer - AWS | None | RL to provide RStudio on AWS Marketplace for SWB customers to add to standard Service Catalog and import (Phase-2 priority) | |

| | | | | |
|---|---|---|---|---|
| | Marketplace model | | | |
| 20 | Upgrade and Support Models for RStudio | | To be managed by RL teams | |

### 2.1 Additional Document Reference

RStudio Architecture Diagrams document link
RStudio LLD Working Document

### 2.2 Out of Scope

Any major callouts of requirements/functionality that is out of scope for this design.

- Deployment of SWB without a custom DNS setup is out of scope. Current design assumes a custom domain is made available when SWB is set up as per existing architecture.

## 3.0 Solution

***Note:*** *You should always suggest a single preferred design instead of describing multiple alternatives without picking one to move forward with. If you haven't picked a solution you haven't finished designing! You should include an "Alternative approaches considered" section at the end explaining why it was rejected.*

As part of looking for solutions to make use of RStudio simpler in SWB following alternatives were evaluated.

**Primary Architecture Considerations**

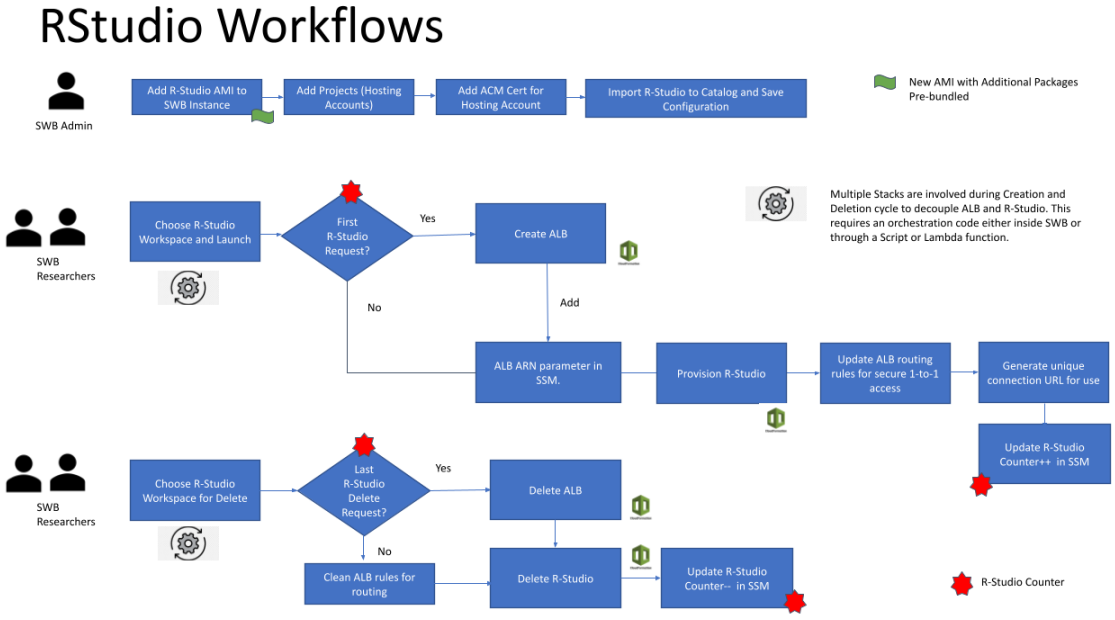
| Options | Solution Approaches | Evaluation |
|---|---|---|
| Option-1 | Automating LetsEncrypt Cert lifecycle for the provisioned Server (and renewal) | Possible but will only partially solve the problem and not a big enhancement to current model |
| Option-2 | Routing from an AWS ALB that frontends the secure connection and connects to RStudio server | Increases cost of ALB but most elegant solution and can be used to share ALB costs across more applications requiring SSL access. Best aligned with AWS Best practices and scalable in future for supporting private network access resources (for high trust environments) |
| Option-3 | Using a Secure Workspace to front-end for Researchers and access research tools and applications (like RStudio) inside private network | More costly option with dedicated workspaces and needs more extensive changes to SWB that currently supports public network assets |

| | | |
|---|---|---|
| Option-4 | Using Bastion Host (or equivalent) to reach out to the RStudio (if deployed in a private subnet with a App connection shortcut link) | Not suitable. This works for SSH but in this case we want access to a Web app. |

Based on the review of options, it was decided to select Option-2 with use of ALB for the current needs. However further aspects of this design were optimized as part of a second level review keeping in mind customer experience, costs and security.

**Further ALB Architecture Optimizations**

| Optimization Considerations | Solution Approaches | Selected option |
|---|---|---|
| 1 | Which Load Balancer to choose - classic LB, Application Load Balancer, Network Load Balancer? | Based on RStudio application routing needs and flexibility classic LB was not suitable and ALB was the optimal option. RStudio support recommends ALB. |
| 2 | How many ALBs to provision:<br>- Per RStudio<br>- Per Project Account<br>- Per Hosting Account<br>- Per SWB Deployment? | Most optimal consideration is one ALB per Hosting Account shared by Projects. Each ALB is expected to be able to scale to 100 RStudio instances concurrent connections and sizing wise + cost wise seems most optimal. Dedicated ALB would have cost about $30-40 per RStudio that was not optimal while shared costs will be much lower and usage based. Refer section 3.4. |
| 3 | When should ALB be provisioned and deleted?<br>- During SWB Creation<br>- During Hosting Account Setup<br>- During first RStudio Provisioning and Deleted with last RStudio deletion in Account | Most optimal option was to create when the first RStudio is requested in a Hosting account and deleted when the last RStudio is deleted from the account hence keeping the costs linked to active usage lifecycle. |




Notes: New RStudio AMI will be deployed into the main account during deployment. Refer section 3.17

## 3.1 Security Certificate Lifecycle

Original design of RStudio expected a User Provided Certificate (e.g LetsEncrypt) to be baked into AMI and then used by NGINX Server on RStudio EC2 box for creating a secure (SSL) connection. The use of an external certificate was needed since AWS ACM certificates could not be used with the above design. However the lifecycle of Certificate management was heavy and not suitable for researchers leading to complexity with initial setup and on-going Certificate renewals etc.

To solve the above problem and enable use of AWS ACM required a change in design to introduce ALB to front-end the RStudio instances and these can use AWS ACM seamlessly.

One ACM Certificate is needed per SWB Hosting Account setup that is used by the associated ALB linked to that account. The renewal of certificates is automatic so users do not have to worry about the lifecycle after initial setup.

- **<u>One Certificate linked to AWS Account + Region as a unique combination</u>**

**Alternate options explored:**

- Manual Certificate creation (this option is fine currently)
- Automated Certificate creation (this is more optimized and can be considered as a future enhancement.). Since the number of configured Hosting accounts are not very high per installation, this feature may not be needed on Day-1.

Design Review comments and clarifications:

1.) Does the design use ACM SSL Public certs?
    a.) YES. This should be fine but should be highlighted.

### 3.2 Secure (SSL) connection End-to-End

Original design used nginx in EC2 instances for SSL termination using the user provided SSL certificate. This approach is now being replaced with ALB using the ACM certificates based on section "Security Certificate Lifecycle".

In the current approach the ALB terminates SSL and to secure the connection between ALB and EC2 instances, a self signed certificate is being used.

**Self-signed Certificate design:**

A self signed certificate is created per Rstudio instance when the bootstrap script runs on provisioning of the instance.
A self signed certificate is baked inside the AMI during the AMI creation process to act as a fallback when the bootstrap script fails or is not triggered. This will be explained in the following sections.

**Code changes:**

| | |
|---|---|
| **File to be changed** | bootstrap.sh (script that runs on instance provisioning) |
| **Changes** | Create self signed certificate for type **rstudio** and move the certificate to nginx folder |

**Alternate options explored:**

- Not using the self signed certificate to encrypt traffic between ALB and EC2 to reduce complexity of creating an additional certificate. But, this will raise compliance issues.
- Common Self-signed certificate for all RStudio Instances vs unique Self-signed certificates. Unique ones used for higher security.

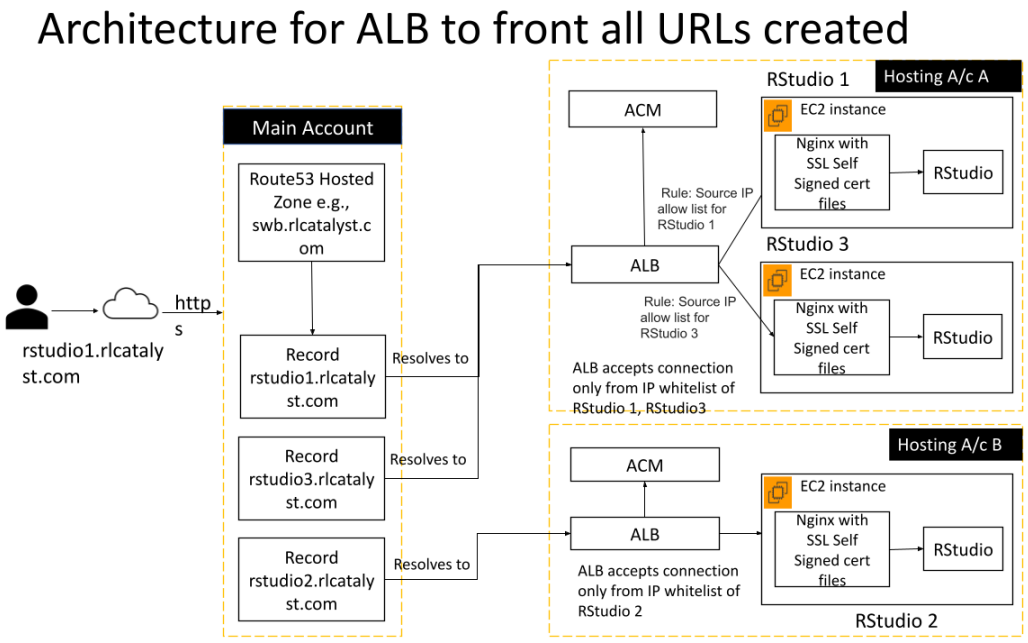
### 3.3 Client Role IAM Changes

Client role of SWB root account provides privileges to create resources for Hosting accounts. With the new resources of ALB getting introduced additional privileges need to be provided for:

- ALB Create and Delete
- Listener and Target group changes for ALB

### 3.4 Application Load Balancer

Based on the Optimized architecture consideration explained in Solutions section above, the following architecture was selected for ALB sharing.




**ALB Cost Considerations**

A cost of $0.0225 per Application Load Balancer-hour (or partial hour) is expected as per current pricing. An additional cost of $0.008 per LCU-hour (or partial hour) is expected based on the following:

LCU Details
An LCU measures the dimensions on which the Application Load Balancer processes your traffic (averaged over an hour). The four dimensions measured are:
**New connections:** Number of newly established connections per second. Typically, many requests are sent per connection.
**Active connections:** Number of active connections per minute.
**Processed bytes:** The number of bytes processed by the load balancer in Gigabytes (GB) for HTTP(S) requests and responses.
**Rule evaluations:** It is the product of the number of rules processed by your load balancer and the request rate. The first 10 processed rules are free (Rule evaluations = Request rate * (Number of rules processed - 10 free rules)

**Refer AWS pricing sheet for more information:**
**https://aws.amazon.com/elasticloadbalancing/pricing/**

**AWS Pricing Calculator:**
https://calculator.aws/#/estimate?id=b219c77289f75206d370e0814d897a367f3d879e

**ALB Capacity Considerations**

**Note: Based on the above, a maximum of 100 RStudio instances can be supported per ALB. This assumes a limitation of 5 CIDRs per RStudio instance. This limitation of 100 is based on the ALB limit of 100 rules per ALB instance.**

### 3.5 Route53 Changes

Current approach uses EC2 DNS name to create CNAME record in route53.
The new design will use the ALB DNS name instead of EC2 DNS because the traffic is now routed via ALB.

**AWS SDK API changes:**

| | |
|---|---|
| **API Name** | route53Client.changeResourceRecordSets() |
| **Parameter to be changed** | ResourceRecords |
| **action** | CREATE |
| **Existing Value** | [{ Value: EC2DNSName }] |
| **New Value** | [{ Value: ALBDNSName }] |

### 3.6 RStudio AMI

The current Rstudio AMI is embedded with a user provided certificate and key for the custom domain. The new design will eliminate the need for those and bake the AMI with self signed certificates. The self signed certificates are used to encrypt only the traffic between ALB and EC2.
The AMI is also packed with additional R packages that are commonly used by the researchers.

**Code changes:**

| | |
|---|---|
| **File to be changed** | provision-rstudio.sh (script that runs on AMI creation) |
| **Changes** | Create self signed certificate and install additional R packages |
| **Library used(self signed certificate)** | openssl |
| **Additional packages installed** | • tidyverse<br>• devtools<br>• kableExtra<br>• survival<br>• survminer<br>• MASS<br>• quantreg<br>• DescTools |

### 3.7 RStudio CFT

The CFT for the new RStudio shall incorporate two changes.

1. An output that indicates "NeedsALB". This will be used by the SWB code while importing the catalog item. See section "SWB Provision Environment Workflow"
2. A target group shall be created and the RStudio instance shall be added as one of the targets.

### 3.8 SWB Hosting Account Configurations

The ACM certificate which will be used by the ALB should ideally be requested when the hosting account is onboarded but that will be a manual step for now.

### 3.9 RStudio Active Counter Tracking (per Account)

Active Workspaces that need ALB are tracked by maintaining a counter in the DB.
The counter gets incremented and decremented with the lifecycle of workspaces that has the flag **NeedsALB.**

**Alternate options explored:**
Active Rstudio instances count can be tracked using the DB query to get the items in the table **EnvironmentsSc** with the following filters.

- projectId - ID of the projects which are using the AWS account
- status - COMPLETED or STOPPED

- output - Contains key NeedsALB

This approach of getting all the environments, filtering it with all the above filters by looping through the environment increases the complexity as the number of environments increases.

### 3.10 SWB DynamoDB Change details

The existing table **DeploymentStore** is being used to store the details of the resources created in the hosting account like ALB, Listener and rules. Each row will have the type **account-workspace-details** and value as the resource details stored as JSON string. **id** will be the AWS account id from **AwsAccounts** table since the details are per hosting account and not per project.

**Table details:**

| **Table name** | DeploymentStore |
|---|---|
| **Columns** | • type - account-workspace-details<br>• id - awsAccountId<br>• value - {<br>id<br>albStackName<br>albArn<br>listenerArn<br>albDnsName<br>albDependentWorkspacesCount<br>} |

### 3.11 SWB Provision Environment Workflow

Added a new step in the standard provisioning workflow that checks for a flag **NeedsALB** in the CFT output. When there is such a flag,

1. Read the output **MAX_COUNT_ALB_DEPENDENT_WORKSPACES**
    a. Get the count of ALB dependent workspaces as per section "RStudio Active Count Tracking"
    b. If count >= MAX_COUNT_ALB_DEPENDENT_WORKSPACES, raise an error saying the maximum number of workspaces allowed reached and stop the provisioning
2. Check if there is an ALB existing for the hosting account
3. If there is no ALB present, a new ALB and a listener is created in the hosting account

    a. Listener on port 80 redirects to listener on port 443
    b. Listener on port 443 uses ACM certificate for SSL
    c. Listener on port 443 has rules that enforce IP white-listing, with the default rule returning HTTP code 403 (Forbidden)
4. ALB is provisioned using a CFT
    a. The CFT will be registered using the **cft-templates-plugin** in SWB code

The details of the products created are being stored in the table created in section "SWB DynamoDB changes".

Once ALB is created, The **launch-product** step is called and the Rstudio is launched. On successful completion of the provisioning,

- Create a Route53 record. Explained in section "Route53 changes"

**AWS SDK API used:**

**API 1: Get CFT template URL**

| **API Name** | describeProvisioningArtifact |
|---|---|
| **Parameters** | ProductId<br>ProvisioningArtifactId |
| **Response** | TemplateUrl |

**API 2: Read CFT from the URL**

The Bucket and Key are parsed from the url and passed as a parameter

| **API Name** | getObject |
|---|---|
| **Parameters** | Bucket<br>Key |
| **Response** | CFT YAML content |

The YAML content will then be parsed using the **js-yaml** library

**API 3: Deploy ALB create stack**

| **API Name** | createStack |
|---|---|
| **Parameters** | StackName |

| | |
|---|---|
| | TemplateBody<br>Parameters (Parameters mentioned below) |
| **Response** | CFT YAML content |

**Provision ALB CFT details**

| | |
|---|---|
| **CFT Name** | provision-alb |
| **Parameters** | Namespace<br>Subnet1 (Subnet created during account onboard)<br>CertificateArn<br>VPC (VPC created during account onboard)<br>PublicSubnet2Cidr - 10.0.32.0/19<br>● Since Subnet1 uses the CIDR 10.0.0.0/19, we are using the above CIDR |
| **Resources** | AWS::EC2::SecurityGroup<br>AWS::EC2::Subnet<br>AWS::ElasticLoadBalancingV2::LoadBalancer<br>AWS::ElasticLoadBalancingV2::Listener |
| **Response** | LoadBalancerArn<br>ListenerArn<br>ALBDNSName |

**IAM Roles Added**

Added the below policies for the role RoleWorkflowLoopRunner

- Added **DeploymentStore** table in the policy **db-access**
- Created a new policy **service-catalog-cft-s3-access**
    - **s3:GetObject** permission to buckets **arn:aws:s3:::cf-templates-*/*** (Needed to read the template in **API 2**)

**Alternate options explored:**

- Explored alternate option of using APIs to create all the resources instead of a CFT. But, since there are 4 resources to create and are dependent for ALB to be created, CFT approach will be better and deleting the resources on failure is also handled by the CFT.

### 3.12 SWB Terminate Environment Workflow

Added new functions to delete the resources which got created as part of the Rstudio provisioning.

1. Added a function in the **terminate-product** workflow step that checks the provisioned product output **MetaConnection1Type**.
    a. If the connection type is Rstudio, remove the listener rule before triggering the terminateProvisionedProduct function.
        i. The Targetgroup needs the rules to be deleted before termination.
2. On successful termination of the product, perform the below operations. Add the code in the existing function **rstudioCleanup**.
    a. Delete the Route53 CNAME record created with the ALB DNS.
3. Added an additional step in the **terminate-environment-sc workflow.** Check if the product has output **NeedsALB.**
    a. Get the count of ALB dependent workspaces as per section "RStudio Active Count Tracking"
    b. If Count is 0, delete the **provision-alb** stack so that the ALB and Listeners will be deleted

Deleting listener rule will be explained in section "ALB Routing Rules Changes with RStudio Lifecycle"
Note: RStudio CFT deletes the target group created for that instance. When the ALB is being deleted all target groups are already deleted.
**AWS SDK API used:**

**API 1: Delete Route53 record**

| | |
|---|---|
| **API Name** | route53Client.changeResourceRecordSets() |
| **Parameter to be changed** | ResourceRecords |
| **action** | DELETE |
| **Existing Value** | [{ Value: EC2DNSName }] |
| **New Value** | [{ Value: ALBDNSName }] |

**API 2: Delete ALB**

| | |
|---|---|
| **API Name** | deleteStack |
| **Parameters** | StackName |

### 3.13 Secure “Connection” Action for User for RStudio

To ensure a secure connection for the user to access an Rstudio instance, a secure signed URL is being generated in the current architecture. The original approach is preserved since changes in the routing mechanism does not affect this implementation.

### 3.14 Disallow Access of Researchers to connect to other RStudio Instances

See section “CIDR Security Considerations”.

### 3.15 ALB Routing Rules Changes with RStudio Lifecycle

After the product provisioning is complete and the connection type is **Rstudio**, a listener rule will be created that redirects the request to the provisioned Rstudio instance. The same rule will be deleted when the Rstudio instance gets terminated and the count of Rstudio instances decreases to zero..
The RuleARN will be stored in the product output and the same will be used while deleting the rule

**AWS SDK API used:**

**API 1: Create Listener Rule**

| **API Name** | createRule |
|---|---|
| **Parameters** | ListenerArn<br>Priority - Count of Rstudio instances +1<br>Conditions<br>● Field - host-header<br>● HostHeaderConfig.Values - HostName<br>Actions<br>● Type- forward<br>● TargetGroupArn (Read from CFT output) |
| **Response** | RuleArn |

**API 2: Delete Listener Rule**

| **API Name** | deleteRule |
|---|---|
| **Parameters** | RuleArn |

**Alternate options explored:**

- Storing the Rule ARN as an entity in the environment table. But this will add another column in the table for an entity that is very specific to Rstudio. So we fixed the approach of storing the ARN in the output of the environment.
- Provide environment id as an input parameter and SWB resolve it before provisioning of the product. CFT uses this parameter to create rules for different rstudio hosts. But this adds a dependency from the user to choose ${envId} as the parameter value and we have to validate raise error before provisioning when user gives wrong envId.

### 3.16 CIDR Security Considerations

Only the CIDR blocks “allowed” by the user should be allowed to connect to the RStudio instance. Since the RStudio instances are behind the ALB in this new design, the client IP address is not visible to them. Hence the security group of the RStudio EC2 instance cannot be used to enforce IP white-listing. It will be enforced using the ALB rules using the source-ip in the condition.
e.g. if (source-ip = CIDR1 OR source-ip = CIDR2 OR source-ip = CIDR3) AND (host-header = rstudio1.swbdomain.com) forward to TargetGroup1

A user may provide multiple CIDRs in the IP white-list. The ALB rules can OR the conditions to check for multiple CIDRs. However there is a limitation on the number of clauses that can be used (5). One work-around for this limitation could be to use multiple rules for the same RStudio instance.

e.g. if (source-ip = CIDR4 OR source-ip = CIDR5 OR source-ip = CIDR6) AND (host-header = rstudio1.swbdomain.com) forward to TargetGroup1

The default rule shall return http code 403 (Forbidden).

The combination of source-ip check and host-header should allow the same level of security as is provided by the security group in the current design.

The same rule shall be edited during the EditCIDR operation.

Additional security can be provided by:
1. Opening the security group of the RStudio instance to traffic only from the ALB. This will require that the ALB details are passed to the RStudio as part of the parameters.
2. Add the CIDRs in the allow-list of all RStudio instances to the SG of the ALB.

### 3.17 Cost Tracking for Provisioned Assets

Standard existing SWB design will be used to tag RStudio instances for cost tracking. The ALB costs that are shared to be tracked at Hosting Account level and not specific to researchers.

### 3.18 Packaging and Delivering RStudio in a separate RL Hosted Repo

We are in the process of building a new landing page in the RL repo for RStudio template.

This would be a top-level folder under the main branch "https://github.com/RLOpenCatalyst/Service_Workbench_Templates" . The README.md in this folder provides a landing page that will have background information, diagrams, etc.

Additionally, we will include a one-click button which is the quick-create link, pointing to the built template residing in a public S3 bucket of RL.

| | |
|---|---|
| **Pre-Requisites** | • Create Custom Domain with the required SSL certificate for the Domain.<br>• Create RStudio AMI using Packer scripts. |
| **Implementation Steps** | • Download RStudio template from Git Repo, add it to Service Catalog in Main account.<br>• Import the RStudio ALB template from the main account into the target account under 'Workspaces Types'.<br>• In the Configuration, Input Parameters provide details – VPC, Subnets, IAM Policy, AMI Id etc.,<br>• Publish the Workspace for provisioning. |

### 3.19 Packaging and Delivering RStudio as AWS Marketplace option

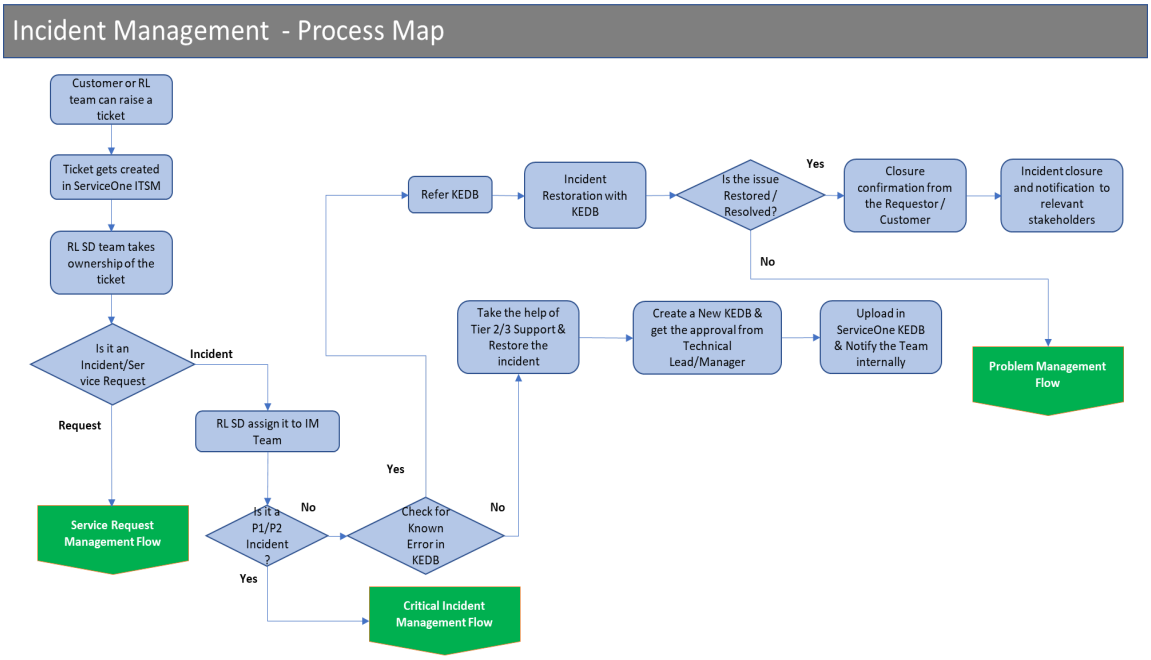
In future the RStudio will be delivered as other standard Marketplace listed products. (Sample - https://aws.amazon.com/marketplace/pp/B06W2G9PRY)

## 3.20 Ongoing Support, upgrades and enhancements considerations

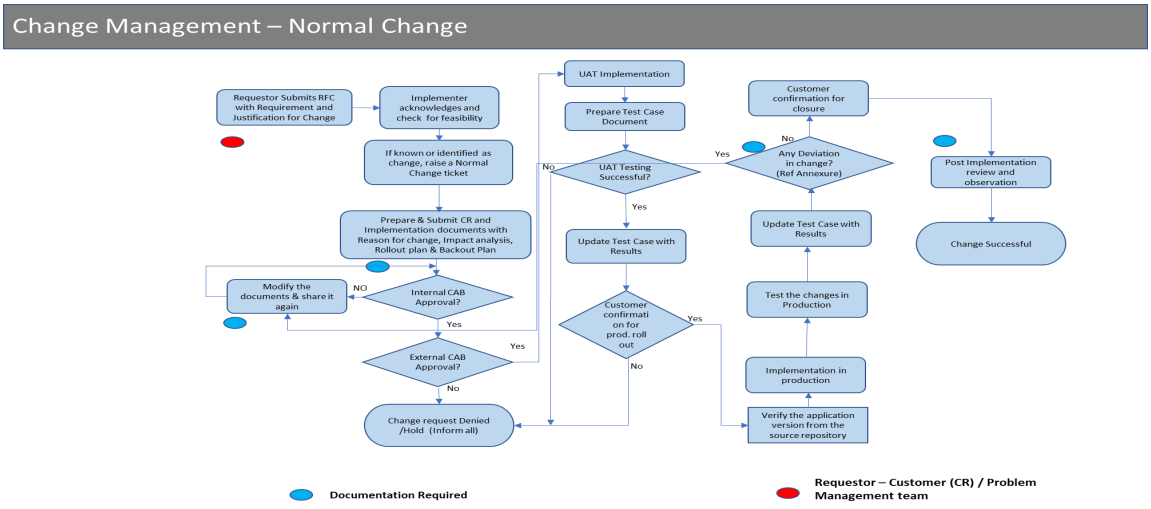
Customers can report issues with our RLServiceOne platform. They can do this by sending an email to rlserviceone@relevancelab.com.

This would generate a ticket and an acknowledgement with the ticket number would be sent back to the requestor.

Please find the Incident handling process for the ServiceOne Framework




Also for the enhancements, and upgrades, we will follow the standard change management process.



Additional Guidelines

You **MUST** include the following items:

1. High-Level Architecture Diagram with entities involved.
    a. Using https://drawio.corp.amazon.com is recommended. It is good practice to keep a link to the source diagram in the doc.
2. Full end to end user experience walkthrough of all customer use cases.
    a. Balsamiq is recommended for UI wireframes

After that, the rest of the solution design is up to you and the format varies depending on the project. However, here are a few guidelines:

- For new APIs, fully specify the inputs/outputs and error cases.
- For updates to existing APIs, specify the proposed changes and note how backwards compatibility is maintained. Breaking API changes should be ideally avoided, but if they occur they must have a strong design justification.
- If your solution involves some sort of state machine, formally specify all the states and transitions. (This applies not only to explicit state machines like Step Functions, but also to entities that go through different "states" i.e. the lifecycle of a SWB instance)
- For new DDB tables, specify the key schema and any additional indexes. Enumerate all the the access patterns and the queries that fulfill them (See example at the bottom of ddb docs page)

Here are a few more yes/no questions to think about. Answering "yes" to them triggers additional details being provided.

- Are you Introducing dependencies on AWS Services that we weren't consuming before that might not be available in all regions?
- Are there documentation changes?
- Are there significant changes to our cost structure / cost model?
- Any changes to our installation?
- Is your design constrained by limits in AWS services?

## Security Impact

1. The IP white-listing feature of SWB for RStudio was implemented using the security group of the EC2 instance to control which CIDRs are allowed to connect. With the introduction of the ALB, it is no longer possible to do this via the Security Group of the EC2. The IP white-listing now has to be enforced via the rules on the ALB. See section CIDR Security Considerations.
2.

***Note:*** *This section is a required section. If you do not believe there is a security impact to your design change, you must explicitly state why there is no security impact.*

1. **List of high level security concerns** - How could attackers benefit from this change? How could they try to attack it? How is your design mitigating security risks?
2. Update the threat model if necessary

Anything that changes/modifies/touches the following items, MUST go through an AppSec Review:

- Authentication or authorization logic
- Networking infrastructure changes (VPC, Security Groups, etc.)

## Story Breakdown

1. Organized into milestones
2. Include milestone dependency graph
3. Include point estimates for stories, extrapolate to estimated number of sprints to deliver project
4. Add the max parallelism for each milestone

See example

## Assumptions

1. We expect the product to be deployed in a VPC that has both public and private subnets. Creation of such a VPC is outside the scope of the current work.

**Recommended reading:**

- Great compilation of design questions to think about when designing distributed systems: https://w.amazon.com/index.php/User:Drr/DesignQuestions

References:

[1] https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-connect-set-up.html

# Appendix

**Trail of Review comments and MOM**

**Key points from Weekly review meeting 29th Apr 2021**

- Since the changes to support new workspaces like RStudio have dependent changes on SWB Base Repo need to find a way to build a “Contract Model” between SWB and RStudio (also future similar components) that creates a clean separation of responsibilities between the base platform (SWB) and additional component (Optional) RStudio. This design will help create an architecture that is extendable and easy to maintain (upgrade etc.)
    a. Addressed in the section below.
- Provide additional Meta-Data on Partner provided Workspaces so that additional information is visible in SWB UI while importing Catalog items like
    a. Version Number
    b. Provided by which partner
    c. URL to additional details about the Catalog product

  These details are provided on a template of the Catalog item and imported by SWB.
    1. Addressed in the section below.

- Have a mechanism for Partner Provided products to track which end customers are using the same and what versions to be able to provide them on-going support and proactive notifications for upgrades/patches etc.
    a. Repo based model
    b. Marketplace based model (Registration form)
- If a Researcher copies the Unique URL provided as part of “Secure Connection” and shares with another user, can the same be accessed? This should not be possible due to CIDR restrictions but needs to be validated.
    a. Test without CIDR restriction
    b. Test with CIDR restriction
- Provide details about the maximum number of RStudio that can be provisioned
    a. This depends on number of EC2 instances, Target groups and Listener rules that can be created in an account
    b. There is a chance that the customer may have increased their limit of any of the above resources
- Detail out the cost of ALB (both pay as you use and the fixed monthly cost) - DONE
- Put all the capacity limitations on the usage of ALB. (DONE)

**Contract model between SWB Platform and Partner Provided Service Workbench Template/Marketplace Product**

- standard contract should be defined
- standard contract should be used by all partner templates
- Lifecycle management should be streamlined - creation, updation, deletion etc.

| | | SWB Responsibility | Partner Responsibility |
|---|---|---|---|
| 1. | Linking of the Repos | | Partner shall host the new templates and provide an installer to load them into the main account. |
| 2. | Clean-up of any existing template being replaced (RStudio) | SWB shall clean-up any existing product being replaced. | |
| 3. | Use of Common Components managed by SWB Platform | SWB code shall provide a mechanism to check which 3rd party templates require common components like ALBs and create the same. | Partner templates shall provide the appropriate meta-data to indicate the need for common components. |
| 4 | Metadata of third party products | | Partner templates shall provide the appropriate meta-data to indicate the need for common components. |
| 5 | Installer | | Partner templates shall be |

| | | | |
|---|---|---|---|
| | | | accompanied by installers to install the templates. |
| 6 | Registration and Support Model | | Partner to provide links for registration and support. |
| 7 | Documentation linkage | | Partner to provide documentation links. |
| 8 | SWB Components being updated for the support of new template<br>1.) Workflows<br>2.) Account Setup Script (if relevant like ACM)<br>3.) Review Role Privileges<br>4.) UI Changes to handle additional Meta-data<br>5. Dynamo-DB Changes | | |
| | | | |

## Provide additional Meta-Data on Partner provided Workspaces

- Installer provided by Partner to provide this data
    - Version
    - Partner Name
    - “Know More” URL
    - Support URL
- SWB code has to be able to understand and provide UI changes
- **It has been decided to implement this in the current release (Action item from 6th May call)**

### SWB CFT Code Changes details:

1. Incorporate the partner product label as a tag in the tag options of the product CFT.
2. Create the tags as parameters with the key pair values such as.
    a. Partner Name = Provided by Relevance Lab
    b. Version Name = V1

    c. Know More = The Know more link will lead to a page on the partners website which has details of that offering.
        i. Eg : The link which leads to the RL Repository.
    d. Support = Support link will lead to a partner website with a support form.
        i. Eg : The link which lead to https://serviceone.relevancelab.com/support/login web page
3. Insert the key pair values into the tags section of the aws service catalog product.

**SWB Backend Code Changes details**

Create the module which would retrieve the values from the service catalog product and store it into DynamoDB.

**SWB Frontend Code Changes details**

Design and implement the partner product label in the product provisioning card as we import in the SWB portal.

1. Create the following the parameter in the product provisioning card
    a. Partner Name
    b. Version Name
    c. Know More
    d. Support
2. Retrieves the values from the backend api's and binds them into the product provisioning card UI.

## Customer registration model for use of Partner provided template

- A registration form will be provided on RL ServiceOne Portal and referred in documentation
- In a marketplace model this will be automated
- Customer can also drop a mail to rlserviceone@relevancelab.com for registration
- Registered customers will get notifications from ServiceOne on important changes

## Provide details about the maximum number of RStudio that can be provisioned

- MAX_COUNT_RSTUDIO = 100 will be captured in RStudio CFT Output parameter similar to “NEEDS_ALB”
- Provision Environment Workflow will do the needful check and notify user