**Beren Clinical Data Mart**



A picture containing graphical user interface

Description automatically generated

Contents

**[1. PURPOSE 4](#_Toc85558736)**

**[2. SCOPE 4](#_Toc85558737)**

[3 DEFINITIONS & ABBREVIATIONS 5](#_Toc85558739)

**[4 SYSTEM DESIGN 8](#_Toc85558750)**

**[5 SAS Computing Environment Schematic 8](#_Toc85558751)**

**[6. SYSTEM ARCHITECTURE DIAGRAM 13](#_Toc85558755)**

**[7. DATA FLOW DIAGRAM 14](#_Toc85558756)**

**[8. SAS DIRECTORY STRUCTURE DIAGRAM: 15](#_Toc85558757)**

**[9. REFERENCES 15](#_Toc85558758)**

**APPROVAL**

|  |  |  |  |
| --- | --- | --- | --- |
| **Document Approval / Review** | | | |
| **Roles** | **Name & Title** | **Signature** | **Date** |
| Business Owner |  |  |  |
| System Owner |  |  |  |
| Quality |  |  |  |

**Version History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date** | **Changes** | **Author** | **Reviewer** |
| 1.0 | 17-OCT-2021 | Final Copy | Sushanth Shetty | Satyoki Chatterjee |

## PURPOSE

This document describes the configurations of Beren Clinical Data Mart (BCDM) with other infrastructure components installed in it

## SCOPE

The requirements specified in this document are for Beren Clinical Data Mart (BCDM) Environment

## In Scope:

The following software and integration are in scope of this environment:

* + AWS Infrastructure
* Data Bricks
* DynamoDB
  + sFTP
  + NetApp ONTAP
  + Data Mart - (Redshift)
  + Spotfire
  + SAS Office Analytics
  + SAS system, which consists of Base SAS (hereafter referred to as SAS)

**Out of Scope**: The SAS Programs and SAS Macros for Beren studies are out of scope of this document.

# DEFINITIONS & ABBREVIATIONS

|  |  |
| --- | --- |
| **Terms** | **Definition** |
| **AD** | Active Directory |
| **Amazon Elastic Compute Cloud (Amazon EC2)** | Amazon EC2 provides scalable computing capacity in the Amazon Web Services (AWS) cloud. |
| **AWS** | Amazon Web Services |
| **CloudTrail** | AWS CloudTrail is a service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With CloudTrail, one can log, continuously monitor, and  retain account activity related to actions across your AWS infrastructure. |
| **CloudWatch** | Amazon CloudWatch enables one to set alarms and automate actions based on either predefined thresholds, or on machine learning algorithms that identify anomalous behavior in your  metrics. |
| **Datacentre** | A storage area or a dedicated building or group of buildings used to house large computer systems and its associated components  such as telecommunications and storage systems. |
| **Directory Services** | Map the names of network resources to their respective network addresses. |
| **EBS** | Amazon Elastic Block Store provided raw block-level storage that can be attached to any Amazon EC2 instances and is used by Amazon relational database services. |
| **EG** | Enterprise Guide |
| **Instance** | Instances in AWS are basically virtual environments. These virtual environments are isolated from the underlying base OS. |
| **BSC** | Beren SAS Computing |
| **Beren Cloud AWS (BAWS)** | Cloud environment specifically qualified for hosting Beren SAS Computing system |
| **Private Subnet** | Private subnets are back-end servers that do not accept incoming traffic from the internet and therefore do not have  public IP addresses. |
| **R Shiny** | R Shiny is a package of R language and is used to build interactive web apps directly from R. |
| **RAM** | Memory or device used to store information for immediate use in a computer related hardware and digital electronic devices. |
| **SAS** | Statistical Analysis System |
| **SAS Compute Server** | Enables clients to submit data set options in SAS programs and  stored procedures for processing by using the SAS language. |

|  |  |
| --- | --- |
| **Terms** | **Definition** |
| **SAS Dataset** | A SAS dataset is a SAS file stored in a SAS library that SAS  creates and processes. |
| **SAS Enterprise Guide** | SAS Enterprise Guide is a functionality of SAS from a point and click windows interface. It’s an easy-to-use menu and wizard  driven tool for analyzing data and sharing results. |
| **SAS**  **Metadata/Spawner** | Program to run SAS Workspace servers and SAS stored  process servers. |
| **SAS Studio** | SAS studio is a web browser-based interface for SAS programmers that also suits the needs of novice users by  providing an assistive framework |
| **SAS Universal Viewer** | The SAS Universal Viewer is a replacement for the SAS System Viewer and is an application for windows environment to view  SAS data sets and other simple text-based files. |
| **SSD (Solid State Drive)** | A solid-state drive is a solid-state storage device that uses  integrated circuit assemblies to store data persistently, using flash memory and functioning as a secondary storage. |
| **Subnet** | A part of the larger network |
| **Topology** | It is the arrangement of various elements (link, nodes, etc.), including the device location and code installation of a computer  network. |
| **VCPU** | Virtual Central Processing Unit |
| **VPC** | Virtual Private Cloud |
| **Win RAR** | Win RAR is a powerful archiver extractor tool and can open all  file formats. |
| **Beren Cloud AWS (BAWS)** | Cloud environment specifically qualified for hosting Beren SAS Computing system |
| **Private Subnet** | Private subnets are back-end servers that do not accept incoming traffic from the internet and therefore do not have  public IP addresses. |
| **R Shiny** | R Shiny is a package of R language and is used to build interactive web apps directly from R. |
| **RAM** | Memory or device used to store information for immediate use in a computer related hardware and digital electronic devices. |
| **SAS** | Statistical Analysis System |
| **SAS Compute Server** | Enables clients to submit data set options in SAS programs and  stored procedures for processing by using the SAS language. |
| **SAS Dataset** | A SAS dataset is a SAS file stored in a SAS library that SAS  creates and processes. |
| **SAS Enterprise Guide** | SAS Enterprise Guide is a functionality of SAS from a point and click windows interface. It’s an easy-to-use menu and wizard  driven tool for analyzing data and sharing results. |

|  |  |
| --- | --- |
| **Terms** | **Definition** |
| **SAS**  **Metadata/Spawner** | Program to run SAS Workspace servers and SAS stored  process servers. |
| **SAS Studio** | SAS studio is a web browser-based interface for SAS programmers that also suits the needs of novice users by  providing an assistive framework |
| **SAS Universal Viewer** | The SAS Universal Viewer is a replacement for the SAS System Viewer and is an application for windows environment to view  SAS data sets and other simple text-based files. |
| **SSD (Solid State Drive)** | A solid-state drive is a solid-state storage device that uses  integrated circuit assemblies to store data persistently, using flash memory and functioning as a secondary storage. |
| **Subnet** | A part of the larger network |
| **Topology** | It is the arrangement of various elements (link, nodes, etc.), including the device location and code installation of a computer  network. |
| **VCPU** | Virtual Central Processing Unit |
| **VPC** | Virtual Private Cloud |
| **Win RAR** | Win RAR is a powerful archiver extractor tool and can open all  file formats. |

## SYSTEM DESIGN

The diagram 1 in this section illustrates the SAS Computing Environment schematic. End users will utilize Amazon Workspaces or Beren Datacenter’s local systems to access the applications such as Enterprise Guide and file share server within AWS. The diagram 2 in section 5 illustrates the SAS Topology for test and production environments. The system will be maintained under formal Change Control.

## Diagram 1: SAS Computing Environment Schematic

Diagram

Description automatically generated



















1. **SAS SERVER AND WORKSPACE CONFIGURATION:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Configuration**  **Specification No.** | **Instance Role** | **Qty** | **Server Name** | **Configuration** |
| CS 1 | SAS Test | 1 |  |  |
| CS 2 | SAS Prod | 1 |  |  |
| CS 3 | Workspace | 4 |  |  |
| CS 3 | Spotfire |  |  | EC2 instance type – M5. large |
| CS 5 | NetApp ONTAP |  |  |  |
| CS 6 | Dynamo DB |  |  |  |
| CS 7 | AWS Redshift |  |  |  |
| CS 8 | AWS lambda |  |  |  |

**Diagram 2: SAS Topology**

Diagram

Description automatically generated

Diagram

Description automatically generated

# Number of Environments

There will be two environments:

* + 1. Production Environment
    2. Test Environment

# Backup and Recovery

Snapshot of EC2 instance would be taken daily when the backup happens at the scheduled time of the day.

# Network Security

All ports are open for internal network only and nothing is opened for external network. The port numbers are 3389, 8561,8581, 8591, 8801 and 135 – 139. These ports are specified during the SAS Computing Environment installation and will be documented in CIR-005, NSC IQ, PQ and OQ Protocol Execution Report. No inbound ports are opened for the Amazon Workspaces.

## SYSTEM ARCHITECTURE DIAGRAM

**Diagram

Description automatically generated**

The system architecture shows how the users can access the SAS EG via the Amazon Workspace. The directory services and the servers are hosted in the VPC in AWS cloud. They can login using Beren domain credentials to the Workspaces and access the Directory services (only authorized users). The SAS server has inbuilt functionalities of SAS Dataset, SAS Studio, SAS EG and SAS Spawner/Metadata. They are hosted on a private subnet and can be monitored via the CloudWatch and the CloudTrail.

## DATA FLOW DIAGRAM

**Diagram

Description automatically generated**

This diagram illustrates how files will be received and can be uploaded/downloaded by SAS programmers, data managers and other authorized users by logging into SAS EG through their local machines or via the Amazon workspace account.

Files are received by Data Managers / SAS Programmers via emails or any other source to their local machine. The authorized users can connect to SAS server via SAS EG on their local systems or via the Amazon Workspaces. They can upload the files to the SAS server in the secured Active Directory structure which has been created. Similarly, if any authorized users want to download the files, they can do so from the SAS directory (from the server) onto their local machines via Amazon Workspace or via SAS EG. The SAS server is mapped as a drive onto the local system or the workspace for the users to access the directory structure.

## SAS DIRECTORY STRUCTURE DIAGRAM:

Primarily SAS home directory will have program folder i.e., BerenA01 which contains subfolder for single study and pooled study. Subsequently the Single and Pooled folder will also contain subfolders as described in the figure below. In addition to this the Utility folder will contain information for SAS programs, macros, templates for test and prod.

## REFERENCES

* 1. SOP IT101 Computer System Risk Assessment (CSRA)
  2. CIR-002 BCDM System Validation Plan
  3. CIR-003 BCDM System Requirements Specification