SDataWaha Guidelines and Procedures

Table of Contents

1. Introduction

1.1 Purpose

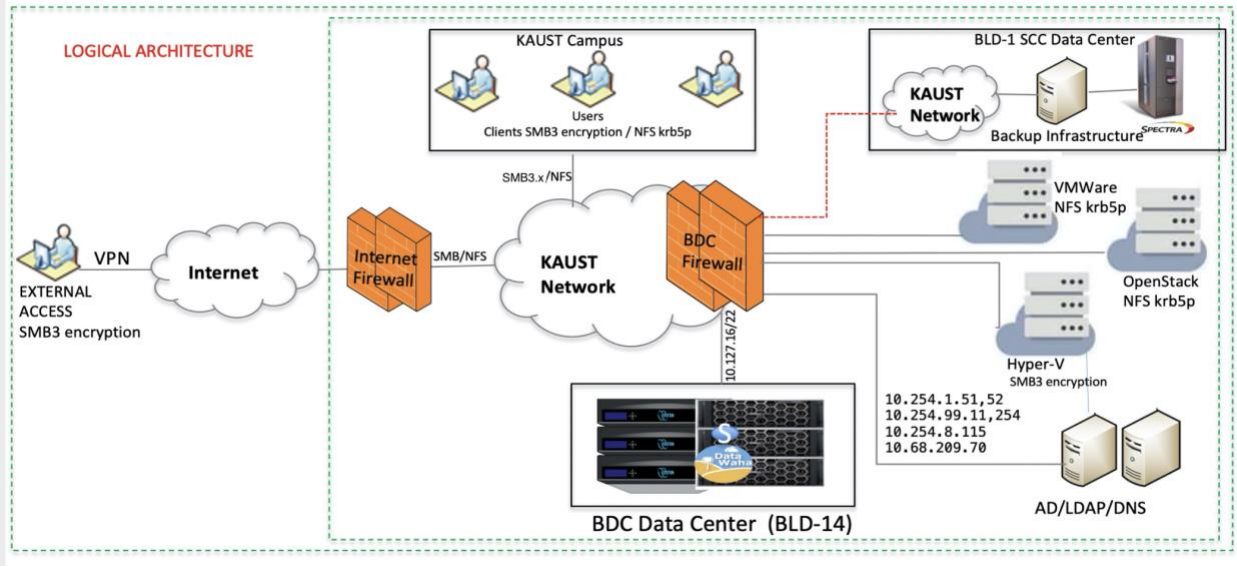
This document consists of standard operation policies procedure for of SDataWaha.

1.2 Overview

SDataWaha provide data storage and archive system for sensitive research data such as Human Genome Data. Which allows users to store and archive in a highly secure manner. This data service aims to address end-to-end data (E2EE) encryption, data privacy, security, and compliance requirements, including: BEC, NIH data sharing guidelines, ISO 27001 standard, KAUST Information Security policies and standards, and the Dell EMC Isilon security best practices. SDataWaha provides data encryption at rest, data backup encryption, and encryption-in-transit in conjunction with KAUST Active Directory Authentication and Access Control.

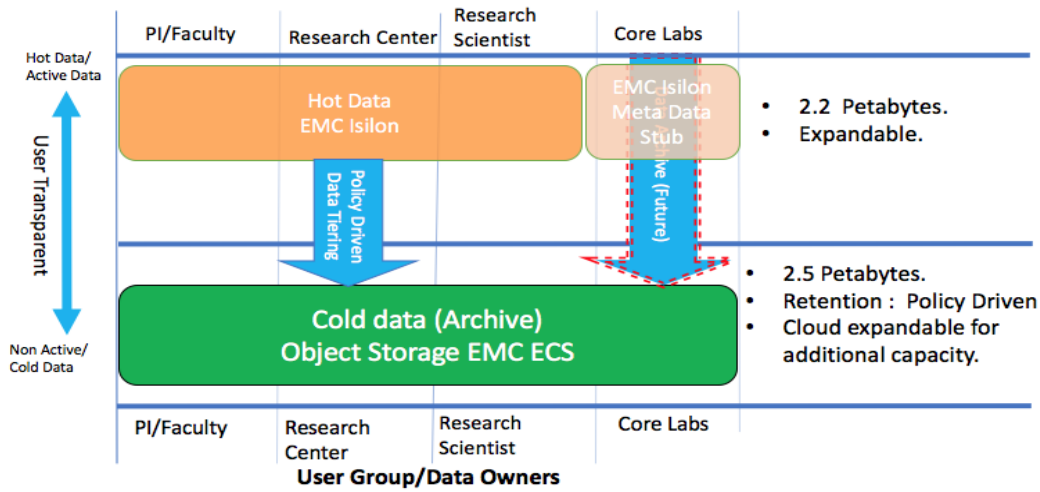
The consumer of SDataWaha should review the "Assumption section", and adopt the requirements for the data accessibility and data sharing. the client reasonability is to gain an understanding of any assumption made in the guidelines.

1.3 High Level Logical Diagram



1.4 Data Tiering and Data Archive

Active data will be on tier1 EMC PowerScale Isilon, cold/archive data user must move it to tier-1 object storage. For Core Labs all the data will be moved to Tier2 cold storage only meta data will be maintain on Isilon Tier1.



2. Guidelines

2.1 Assumption

SDataWaha should consider assumptions about usage and environmental settings as requirements for the SDataWaha installation and its operating environment.

This will ensure the proper and secure operation of the SDataWaha

The following assumptions are made regarding the use and deployment of the SDataWaha

- SDataWaha Data path users are identified, authorized and authenticated prior to gaining access to the data.
- SDataWaha will be located within controlled access facilities BLD-14 Data Center, which will prevent unauthorized physical access.
- SDataWaha will be accessible from IT Managed Clients for Mac/Windows and Linux
- Client machines will adhere to KAUST IT Security Standards
- Owner of the SDataWaha data folder is responsible to maintain and authorize user access.
- NFS4 export is only available for Linux Client which are integrated with KAUST Active directory, adhere to KAUST IT Security Standards and Kerberos is implemented.
- Mac and Windows client must be integrated with KAUST AD and SMB3.x with encryption supported.

2.2 How to request DataWaha storage

- Open a ticket with VITA requesting SDataWaha Disk Storage or Data Encrption".

2.3 Default quota

- **20 terabytes tier-1 and 80 terabytes tier-2(archive) for each PI/RC/Lab folder**.
  Note: The Quota will be accounted against the usage of **DataWaha** + **SDataWaha**

2.4 How to request additional capacity

- Business case justification must be required for additional disk storage beyond 20 Terabytes **max up to 100 Terabytes**.
- Funding must be required for requirements beyond 100 Terabytes.
- SDataWaha is also cloud enabled to meet growing demand of data storage growth. cost of Cloud storage usage is subject to discussion.

2.5 Data archive policy

- Every folder must have data management policy to migrate non-active (Cold data) to cost effective object storage system for data archive.
- Default data archive policy is 180 days, files older than 180 days will be migrated to teir2 storage automatically and data access is user transparent.

2.6 Data Directory/Data Location

- Each folder or directory under DataWaha will be unique.
  For e.g. /sdatawaha/<PI Folder>, /sdatawaha/<RC>, /sdatawaha/Corelab.

2.7 Authentication

- KAUST Active Directory (KAUST Portal id)

2.8 Access protocol

- SMB3.X with encryption from any devices within KAUST network.
- NFS4 with krbp5 and it is limited to data center devices (including dm, remote workstations, OpenStack Virtual machine and HPC front end node)

2.9 Limitation

- SDataWaha cannot be use to run HPC compute jobs.
- SDataWaha will not be accessible directly from HPC Compute Nodes including Shaheen and IBEX.
- SDataWaha can be mounted on Linux devices using NFS4 Kerberos with privacy for in-transit encryption or SMB3.x with encryption.

- Client machines must be integrated or using KAUST Active Directory authentication.

2.10 How to access from outside the KAUST

- KAUST VPN is required

3. Operation procedures

3.1 How to access SDataWaha from Linux.

- Individual NFS exported "PI folder" is accessible from identified Linux clients which is integrated with KAUST Active Directory and Kerberos is configured See NFS Client Configuration Ref_Guide.pdf
- Add below mount point in (/etc/fstab) on Linux client and mount "PI Folder" or you can configure automount with the help of IT Storage team.

sdatawaha.kaust.edu.sa:/ifs/sdatawaha/<PI Folder Name> /sdatawaha/<PI Folder Name> nfs4 rw,proto=tcp,nfsvers=4.0,sec= krb5p,noatime,acl,intr 0 0

- Default Directory permissions
  - PI AD Group: Full control
  - User: Full Access
  - Others: No Access

Note :Each <PI Folder> will be exported to selected Linux client using their IP addresses or subnet.

3.2 How to access From Windows.

- \\sdatawaha.kaust.edu.sa\<pi folder>

Note: only SMB3.x with encryption is supported check your OS release for SMB3 support.

3.3 How to access From MAC

- smb://sdatawaha.kaust.edu.sa/<pi folder>

Note: only SMB3.x with encryption is supported check your OS release for SMB3 support.

3.4 Data transfer to and from DataWaha

- From Windows and MAC
  - Directly mount the SDataWaha and use windows copy tools.
- From Ubuntu
  - Use rsync from any Ubuntun client to dm.kaust.edu.sa
    - *For .e.g. rsync –av localdata dm.kaust.edu.sa:/sdatawaha/<pi folder>*
- From HPC Cluster
  - From login nodes aloing, ilogin, Shaheen use rsync.

- *For e.g rsync –av dm.kaust.edu.edu.sa:/sdatawaha/<pi folder>*

3.5 DataWaha Access and Permissions

- SDataWaha is accessible within the KAUST. VPN is required to access from outside the KAUST using portal id
- Access is controlled by AD group permission, each parent folder will be owned by Active Directory "group", with full control.
- individual user access is controlled by adding or removing the members in the group.
- Additional access control and permission can be configured from windows machine.
- Default Directory permissions
  - PI AD Group: Full control
  - User: Full Access
  - Others: No Access

********** End of the Document**********