VCF Tooling

Design and Recommendations for VCF Distributor Partners

Broadcom Limited
Web: www.broadcom.com
Corporate Headquarters: San Jose, CA

Revision History

Revision	Date	Change Description
1.0	May 17,2024	Creation of document draft

Contents

Revision History **1**
Contents **2**
Introduction **3**
Executive Summary 3
Requirements **4**
Assumptions **5**
Constraints **6**
Risks **7**
Services: **8**
Linux Environment 8
Recommended Minimum Footprint 8
Recommended Application Packages 9
Storage 9
Upload Destination 9
Support Case Data Storage 10
Support Case Data Analysis 11
Data Mover 15
Appendix I - Referenced Terminology **16**

Introduction

This document serves as an outline to describe how Broadcom Distributor Partners (Disti) can design and implement systems to allow support personnel to perform log analysis for VMware Cloud Foundation (VCF) products.

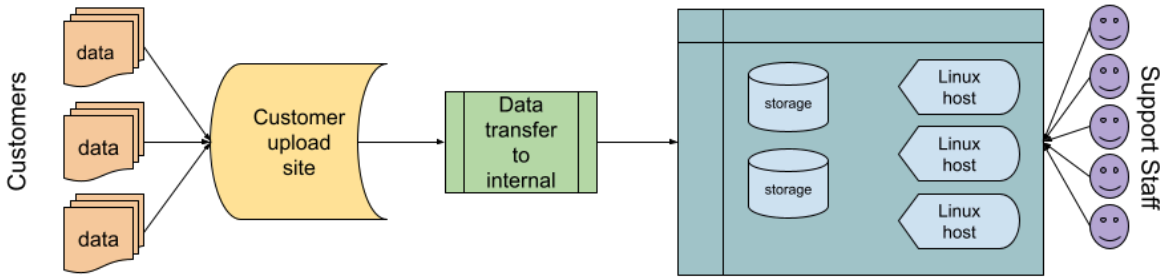
Distis require the ability to process VCF log file data. This log data consists of large datasets that must be managed in safe controlled environments that are collaborative and ensure proper security, governance and compliance. This recommended configuration draws upon the collective experience of VMware Global Support and provides design criteria found to contribute positively to the objective of providing efficient and accurate incident resolution.

Executive Summary

This document will outline a process for receipt and processing of log files required to address cases for the VCF product suite.

Files are uploaded to a standard repository accessible by customers. This repository ensures visibility is limited for individual customers, to prevent customers from seeing data outside their specific case. Once uploaded, this customer data is moved to an internal repository for review and processing by technical support. This internal repository consists of a shared storage infrastructure that is accessed via shared hosts with specific tools for log review and management.

The system controls the life cycle, maintenance and access to customer data. It ensures files are accessible to staff and removed(purged) from the system in accordance with appropriate regulatory and compliance requirements [RI.001].



This system is devised using the experience of the VCF Support organization's long history of providing technical support. It ensures accessibility to files by multiple technical support agents without requiring each technician to download the files individually. This not only increases efficiency by eliminating costly download times, but enables a more collaborative approach to support where the efforts of individual technicians are retained for use by any other technician brought into the case. Critically, it also minimizes the storage capacity needed for each technician's individual workstation, by preventing the need for downloading files locally.

Requirements

ID	Title	Description
Req.001	High Availability	All considerations for performance and redundancy should be implemented. This includes, but is not limited to, highly available hosting solutions, load balancing, shared storage, and other industry standard practices to eliminate single points of failure.
Req.002	Retention Policy	Data Retention must be configured according to all applicable policies. Different sectors such as Healthcare and Federal entities can have specific guidelines on how data must be stored and retained. [RI.001, RI.003]
Req.003	Role Based Access Control	Systems should be secured to meet all internal infosec requirements. Role based access controls are highly useful and beneficial.
Req.004	Isolation	To protect customer data, systems must only be accessible via secure corporate networks, and access limited to approved technicians. [RI.005]

Assumptions

ID	Title	Description
AS.001	Linux systems administration capabilities	Distis needs to have technical staff who are familiar with Linux systems administration. The details in this document assume an overall knowledge and understanding of Linux administration. Day to day tasks are not detailed and are assumed as part of normal business operation.
AS.002	All users require basic familiarity with Linux operating systems	Users of the system will need general experience and comfort operating within a Linux command line environment. Working knowledge of command syntax, execution and data parsing are required to successfully navigate.
AS.003	Average size of support bundle	Average support bundle file size is 2.8GB Max observed support bundle file size is 199GB
AS.004	File uploads	Customer file uploads will be processed through Wolken's natively provided S3 storage solution. This document assumes that the S3 storage is the target upload location.
AS.005	Average case data retention period	Case data is stored through the duration of the support interaction and then removed within "x" days of case closure. Data retention should be managed according to regulatory and compliance requirements. [RI.001, RI.003]
AS.006	Monitoring	There should be additional system(s) in place to monitor performance/state of the Linux environment and related Storage and Data Mover. [AS.001]
AS.007	Security	Systems should be secured, patched and monitored to maintain the integrity, availability and stability of the environment. [AS.001]

Constraints

ID	Title	Description
CO.001	Data Accessibility	Only those employees who require access to uploaded customer data to perform case-related investigation should have access to the customer data. Recommend internal review of data handling and access controls to ensure all regulations and compliance requirements are met. [RI.001, RI.003]
CO.002	System Accessibility	Employees granted access to the system should have the minimum set of rights/permissions in order to perform their work. In accordance with the principle of least privilege (here and here). [Req.004]
CO.003	Auditing	There must exist system logs such that user and automated actions can later be audited.
CO.004	Data Availability	Case data must be available to facilitate the speedy analysis and solution of customer issues. Data should be accessible through the life of a case.

Risks

ID	Title	Description
RI.001	Compliance	VCF customers upload support bundles that contain but are not limited to the following types of data. Log files, hostnames, Unique Hardware IDs, usernames, email address, and many other data types. Distis are responsible for ensuring any compliance requirements from either local, regional or international governing bodies are met. This document does not detail those requirements.
RI.002	Business Continuity and Disaster Recovery	These large datasets need to be available for continuous support, suitable Business Continuity and Disaster Recovery practices should be implemented to mitigate any risk factors to your business. The most common approach used for a data set of this size is array based replication. This document does not detail exact recommendations and the Disti should work with their internal teams to ensure customer data is backed up appropriately.
RI.003	US Federal Data Handling	This document does not intend to define any environmental details that would meet US Federal Requirement(s) and is not intended to handle federal customer data. This document does not define any ITAR or CMMC requirements that may be necessary to handle unclassified sensitive data from US Federal organizations and/or contractors.
RI.004	Performance	Compute, Networking and Shared Storage resources must be capable of retaining and serving the required data transfer rates and access requirements to provide an acceptable end user experience that can scale with load.
RI.005	Bad Actors	System security, permission minimization and auditing practices should be in place such that damage caused by Bad Actors is minimized and can be tracked.

Services:

Linux Environment

Any enterprise grade Linux distribution would be a suitable Operating System (OS) to host the environment. Linux is the best suited OS class for this system for a variety of reasons:

- The log source systems are Linux based
- Expansion of compressed files is simpler in Linux
- There are Linux specific tools that assist in log manipulation and analysis (e.g. grep, sed, awk)
- In Linux you do not need to download files to your local machine for analysis, which mitigates compliance and security risks for customer data [RI.001, RI.005]

To support the Linux systems, it is required to have sufficient staffing [AS.001] for regular maintenance activities such as distribution patching, package management, storage management, etc. Selection of an appropriately supported Linux distribution is critical to ensure availability.

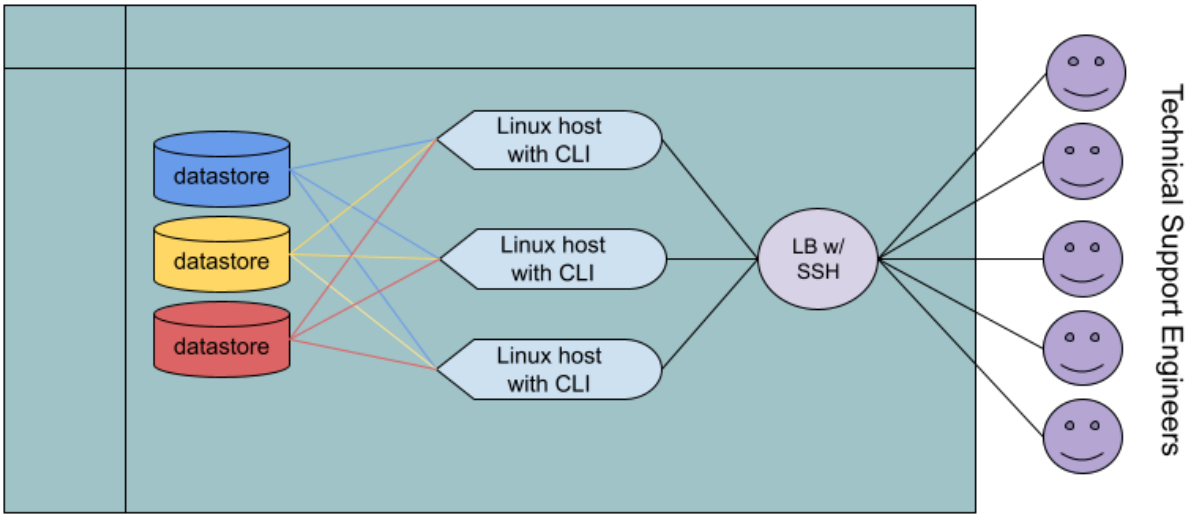
Best practices should account for these maintenance requirements and incorporate them into Disti architecture for high availability [RI.002]. Distis should adjust recommendations to meet their specific staffing and caseload requirements. [RI.004].

Recommended Minimum Footprint

- Multiple VMs (at least 3) to allow for OS patching
- Each VM (6 vCPUs, 12GB RAM) ** *This sizing can service approximately 20 technical support engineers accessing the system concurrently*
- Front-end Load Balancer (LB) for SSH connections
  - Session Persistence (required)
  - High Idle Timeout (recommended)
- Command line interface (no desktop distribution)

Resources can be scaled horizontally or vertically depending on the Disti size.

We recommend having additional environment(s) for testing/dev in addition to the production environment.



Recommended Application Packages

Application Page Name	Description
Linux Packages	neofetch, p7zip, p7zip-plugins, lynx, elinks, pv, unrar, glibc.i686, zlib.i686, libxml2.i686, openldap-clients, mysql, dialog, git, ruby, mysql-devel, redhat-rpm-config, gcc, python3.11-devel, the_silver_searcher, wireshark-cli, gcc-c++, sshpass, neovim, ripgrep, fzf, tmux, screen, sqlite, ccze, unar, emacs-nox, ack, zstd

Storage

Upload Destination

Wolken provides a built-in S3 storage solution on AWS. This document assumes the Wolken S3 storage solution is being used [AS.004]. This document does not define a File Transfer System, nor does it define the requirements that would need to be accounted for in any other file upload solution that might be used outside of the Wolken S3 storage solution.

Files uploaded to Distis can range from very small to very large. The upload system needs to be configured with sufficient capacity [AS.003]. The Disti should assume an average of between 2 and 5 TiB of customer uploads; with a maximum of 10 TiB of uploads in a 24 hour period. Customer uploads are primarily compressed archives (zip, tgz) which can expand by as much as a factor of 10.

These figures are based on the complete pool of customers. Distis should size their capacity appropriately for their specific customer base.

Support Case Data Storage

A shared storage system should be used to save and store customer data (support bundles). This shared storage should be accessible to the Linux system(s) that are used to perform log analysis. The storage system should allow for multi read / write IO. The storage system should allow for data encryption for the data at rest. Broadcom recommends an enterprise NFS storage system for this purpose, although not required.

Support case data is typically organized into directories, named by case number, one directory per case. Each top-level case directory contains any customer data uploaded to the case (images, compressed log bundles, etc). When a compressed log bundle is extracted, it is extracted into a sub-directory and then reassembled into the actual logs for analysis.

For added data security, each top-level case directory has its "sticky bit" turned on to avoid accidental deletion of customer data by users.

After the case is concluded (closed), the top level case directory and all of its contents are deleted (purged) in compliance with all business and legal requirements. [AS.005, Req.002]

Example directory structure / mount points

/srdata1	/srdata1/12345678/
	/srdata1/23456789/
/srdata2	/srdata2/34567890/
	/srdata2/45678901/
/srdata3	/srdata3/56789012/
	/srdata3/67890123/

Support Case Data Analysis

It is recommended to implement several command line tools to automate frequently repeated operations. These tools should be a part of the user's defined **PATH** to be accessible anywhere. Many linux based tools will allow for initial log analysis, but specialized tools are also recommended, some suggestions and pseudo-code examples are found below. NOTE: The following code examples are only for demonstration purposes and are non-functional. To acquire the functionality described, each Disti must develop them themselves

Log bundle extraction tool

Having one tool that can extract log bundles of all types will allow for support users to more easily maneuver in the system. Setting extracted folder permissions to allow for collaboration.

```
set_log_file_for_failures="__extraction_failure__.log"
permissions_dir=770
permissions_file=660

if [[ $file ]]; then
  # Identify file extension for decompression engine
  if [[ file extension -eq 1 && "extension1" -eq 1 ]]; then echo
"file_type"
  elif [[ file extension -eq 1 && "extention2" -eq 1]]; then echo
"file_type"
  elif [[file extension -eq 1 && "extension3" -eq 1]]; then echo
"file_type"
  return found_match
  else
  # Log and record no match for file
  return "no file extension match"
fi

# Perform decompression
while [[ $file_matched ]]; do
  file_type=file_type
  target=case_dir
  # Perform extraction in appropriate place
```

```
  extracted_dir=$target/ “$file_matched”_extracted
  mkdir $extracted_dir
  extraction_procedure $file >>
$extracted_dir/$set_log_file_for_failures
  exit_code=$?
  if [[ $exit_code -ne 0 ]]
    then error “$exit_code”; extract_dir=$file_extracted_failed;
mv $file_extracted $extract_dir
    else success; rm $extracted_dir/$set_log_file_for_failures
  # Change permissions on Files and Directories recursively
  find . type -d -exec chmod $permissions_dir $extracted_dir/
  find . type -f -exec chmod $permissions_file $extracted_dir/
```

NOTE: The above code is only for demonstration purposes and is non-functional. To acquire the functionality described, each Disti must develop it themselves.

Change to case directory tool

Navigates to the shared storage location of support ticket data easily by passing it the support ticket number.

```
symlink_path = path_to_symlinks
# Validate case_number
if isinstance(case_number, int):
  if len(case_number == 8 or case_number == 9):
    case_number = str(case_number)

# Check if symlink exists and return if true
if os.path.islink(symlink_path+case_number):
  return os.readlink(symlink_path+case_number)
else:
  return print(“Cannot find case number directory”)
```

NOTE: The above code is only for demonstration purposes and is non-functional. To acquire the functionality described, each Disti must develop it themselves.

Create case directory tool

Creates support ticket shared storage data folders manually by passing it the support ticket number.

```
symlink_path = path_to_symlinks
```

```
datastores = {
  datastore1name:datastore1path,
  datastore2name:datastore2path,
  datastore3name:datastore3path
}

# Validate case_number
if isinstance(case_number, int):
  if len(case_number == 8 or case_number == 9):
    case_number = str(case_number)

# Check if path exists
if os.path.islink(symlink_path+case_number):
  return os.readlink(symlink_path+case_number)
else:
  # Check space of data stores for best placement
  datastore_space = {}
  for k, v in datastores:
    datastore_space[k] = psutil.disk_usage(v).free
  most_space = max(datastore_space, key=lambda k:
datastore_space[k])
  # Create directory
  os.mkdir(most_space/case_number)
  source_path = most_space + “/” + case_number
  # Create symlink to datastore
  os.symlink(source_path, symlink_path)
  return source_path
```

NOTE: The above code is only for demonstration purposes and is non-functional. To acquire the functionality described, each Disti must develop it themselves.

Manual transfer of files from S3

Technicians can manually retrieve files from customer upload system (S3) to the Linux environment

```
ACCESS_KEY = “ABC”
SECRET_KEY = “123”
CASE_NUMBER = “12345678”

# Connect to S3 client
```

```
session = Session(aws_access_key_id=ACCESS_KEY,
aws_secret_access_key=SECRET_KEY)
s3 = session.resource(“s3”)
bucket = s3.Bucket(CASE_NUMBER)

# Check if case directory exists and create if needed
datastore_path = create_case_directory_tool(CASE_NUMBER)

# Retrieve data from bucket and download to datastore
if bucket is not None:
  for item in bucket:
    bucket.download_file(item, datastore_path)
```

NOTE: The above code is only for demonstration purposes and is non-functional. To acquire the functionality described, each Disti must develop it themselves.

Data Mover

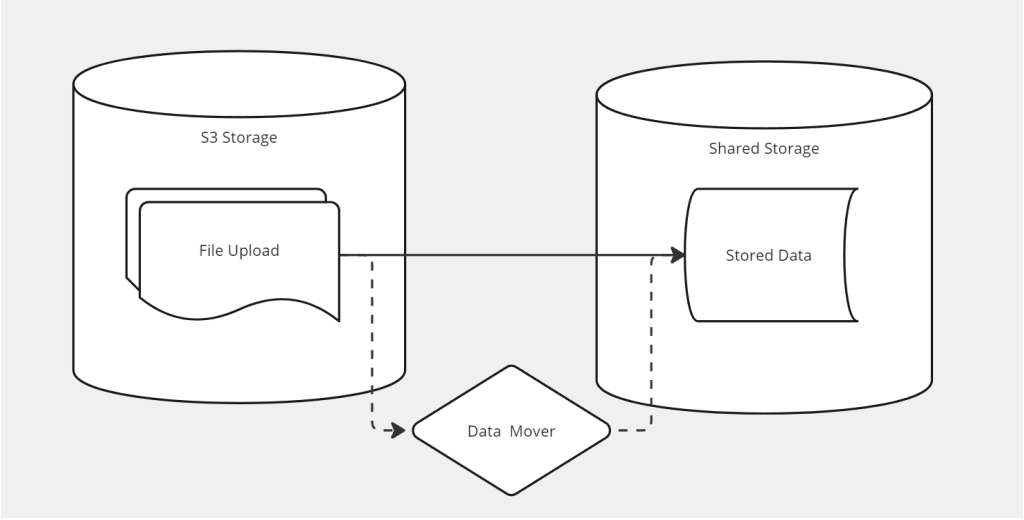
The Data Mover is not mandatory for system operation. However, in order to maintain separation between customer upload location and working directories without impacting Time to Resolution (TTR), there is a need for a mechanism to automatically transfer files. We refer to this as the Data Mover application.

Key components of this application should include the following capabilities:

- Technician notification of file upload (Wolken provides this function)
- Initiation (trigger) of Data Mover application when new customer data is present
- Secure connection between upload system and working directories
- Manual transfer of files from S3
- Recording/Logging of transfer information (e.g. beginning of transfer, file information, case number, etc)
- Ability to validate file is uncorrupted upon transfer (recommend sha256 hashing of data at every stage)
- Removal of files from upload location upon successful transfer to working directories

Ideally, the Data Mover application should be automated to ensure this process does not need human intervention. Manual processes should also be available to ensure timely operation when the Data Mover application is unavailable.

Diagram



Appendix I - Referenced Terminology

Terminology	Definition
Case	Case is the generic term for a ticket open in Wolken
Case ID	Case ID is the numerical ID used in Wolken to track technical issues reported by customers to the Disti.
CMMC	Cybersecurity Maturity Model Certification
Customer Data	Data uploaded from a customer environment. Including but not limited to: support bundles, diagrams, screen captures, etc.
Data Mover	Data Mover is an application that copies data from an upload location and saves it in an internal storage location
Disti	Broadcom Distribution Partners
ITAR	International Traffic in Arms Regulations
LB	Load balancer
NFS	Network File Storage
PII	Personally Identifiable Information - Includes but is not limited to: IP, MAC addresses, hostnames, Unique Hardware IDs, email address, usernames, etc.
Pseudo-code	Non-functional code to demonstrate the high level order of operations to accomplish a given task.
Sticky bit	A special permission on Linux directories that prevents file deletion within the designated directory.
Support Bundle	The generic term for a diagnostic collection of logs, command outputs and various priority data types that may be used in the pursuit of troubleshooting product issues
TTR	Time To Resolution - Amount of time from the incident report to resolved status
VCF	VMware Cloud Foundation
Wolken	Wolkensoft IT Service Management software

This page left intentionally blank