Copernicus Maintenance Team

DHS Suite Easy Deploy Installation Manual

Role/Title	Name	Signature	Date
Authors	Simone Garofalo		19/07/2024
Verified/Approved	Elisabetta Giuliani, Giulia Carosi		19/07/2024

Change Register

Version/Rev.	Date	Description
1.0	19/07/2024	First version of the document

Table of Contents

Index of Tables

1. Introduction

1.1 Scope

This document aims to detail step-by-step instructions to install, configure and run all software being part of the DHS suite through a DHS Suite Easy Deploy tool.

1.2 Document applicability

This document is referring to DHS Suite Easy Deploy tool supporting the DHS#6 delivery.

Document version	Component	DHS Release
1.0	DHS Suite Easy Deploy tool	DHS#6

1.3 Document structure

The document is structured as follows:

- Section 1 (this section) contains scope and purpose, providing document structure, reference documents and definitions/acronyms.
- Section 2 contains an overview of the DHS Suite Easy Deploy tool.
- Section 3 contains the installation procedure of the DHS Suite Easy Deploy tool, troubleshooting issues and utilities scripts.

1.4 Reference documents

Table 1 - Reference Documents

Ref.	Title	Reference and Version
RD-1.	GSS COTS installation	COPE-SERCO-TN-23-1461 - GSS COTS installation v1.3
RD-2.	GSS Administration Manual	GAEL-P311-GSS - Collaborative Data Hub Software GSS Administration Manual_1.6.5
RD-3.	Transformation Framework Installation and Configuration Manual	COPE-SERCO-TN-21-1218 - Collaborative Data Hub Software Transformation Framework Installation Configuration Manual v4.0
RD-4.	COPSI Installation and Configuration Manual	ALIA-COPSI-ICM-22-0001_Installation_and_Configuration_Manual_3.0.0

RD-5.	Semantic Framework Installation and Configuration Manual	COPE-SERCO-TN-22-1329 - Collaborative Data Hub Software - Semantic Framework Installation Configuration Manual v3.5
RD-6.	DAFNE Installation and Configuration Manual	ALIA-DAFNE-ICM-2021-0001_Installation_and_Configuration_Manual_1.5

1.5 Acronysm and Abbreviations

Table 2 - Acronyms and Abbreviations

Acronym	Definition
COPSI	Copernicus Space Interface
COTS	Commercial-Off-The-Shelf
DAFNE	Data Flow Network Environment
DHS	Data Hub Software
ESA	European Space Agency
GSS	GAEL Store Service
ICD	Interface Control Document
NFS	Network File System
RD	Reference Document
SF	Semantic Framework
TCP	Transmission Control Protocol
TF	Transformation Framework
UDP	User Datagram Protocol

2. DHS Suite Easy Deploy tool.

2.1 Overview

The DHS Suite Easy Deploy tool is a comprehensive package designed to simplify the installation process of various DHS suite components. It includes all the necessary software and scripts to ensure a seamless setup.

For a technological point of view, the DHS Suite Easy Deploy tool is based on Docker Swarm which guarantees the possibility to perform both a deployment on a single node and on a distributed environment.

From now on, for convenience, we will refer to the "DHS Suite Easy Deploy tool" as "Installation Tool". These terms will be used interchangeably.

2.2 Supported DHS components versions

This version of the DHS Suite Easy Deploy tool is applicable to the following DHS components.

- **COPSI v3.0.1**:
  - copsi 3.0.1
- **DAFNE v3.0.2**:
  - dafne frontend 3.0.2
  - dafne backend 3.0.1
  - postgres 12.5 (release version: 3.0.2)
- **Transformation Framework v1.5.3-osf**:
  - esa_tf_restapi latest
  - esa_tf_worker latest
  - scheduler dask 2023.3.2-py3.10
  - nginx 1.21.6 (release version: 1.5.3-osf)
- **Semantic Framework v3.6**:
  - virtuoso-opensource-7 latest version
  - sf-datareceiver v3.6
  - semantic_framework v3.6
  - sf-natural-language v1.4
- **KEYCLOAK**:
  - custom version ciam-swarm-keycloak 1.0
- **GSS v1.4.3**:

- cdh-admin 1.4.3
- cdh-catalogue 1.4.3
- cdh-ingest 1.4.3
- cdh-notification 1.4.3

2.3 Pre-requirements

2.3.1 Hardware requirements

Here follows the hardware requirements for the DHS Suite Easy Deploy tool installation:

- **CPU**: Minimum 16 cores, with additional cores required for Transformation Framework (TF) depending on the scale of deployment.
- **RAM**: Minimum 32 GB, with an additional 6 GB per TF core.
- **DISK**: Minimum 350 GB of storage to accommodate the installation and operation of all components.

These values can be different in case of distributed deployment, in that case, adjust the resources of each node of the Docker Swarm cluster according to the hardware specifications defined in the documentation of the software to be installed on that specific node.

2.3.2 Software pre-requirements

Here follows the software requirements for the DHS Suite Easy Deploy tool installation:

- Docker engine>=v20.10.21 (on each node in case of distributed deployment)
- Docker compose>=v2.12.2 (on each node in case of distributed deployment)
- SOLR>=8.0
- Postgres>=10.12
- Kafka>=3.3.1
- Zookeeper=3.8
- Linux NFS package

It is important that the TCP/UDP ports 2376/2377/7946/4789 are opened on the machine hosting the Installation Tool.

Note: Before launching the Installation Tool, the GSS COTS installation procedure [RD-1] must be executed.

2.3.3 Additional pre-requirements

- The machine hosting the suite must have the "/shared" folder configured in the "/etc/exports" file and exported as an NFS volume to the subnet with the other nodes of the Docker Swarm cluster.
- In case of installation in a distributed environment, the following steps must be followed to create the Docker Swarm cluster:
  - To enable the docker swarm cluster execute the command: "docker swarm init --advertise-addr ip" (the ip shall be reachable from the other node of the cluster)

- From the leader node exec only one of the following commands:
  - "docker swarm join-token manager"
  - "docker swarm join-token worker"
- The output of these commands (i.e. docker swarm join --token TOKEN <server_ip>:2377) is a command to be launched on workers/managers node of the swarm cluster. Running this command on each node will create the docker swarm cluster.
- Running the command "docker node ls" from the leader node (or master nodes) will show the list of the nodes in the cluster.

3. Installation procedure

3.1 Suite download

The suite package is downloadable at the following link:

- https://github.com/DHS-Components/dhs-suite-easy-deploy/archive/refs/tags/1.1.0.zip

This package must be downloaded under /home folder by 'root' user.

Since Semantic Framework and IAM software must be installed in the same machine hosting the Installation tool, it must be ensured that the Installation tool must be unzipped in the machine dedicated to the Semantic Framework and IAM software.

After the download it is needed to unzip the package and go inside it where is the installer script:

- *unzip /home/dhs-suite-easy-deploy-1.1.0.zip*
- *cd /home/dhs-suite-easy-deploy-1.1.0*

The version of the package for this installation is 1.1.0.

3.2 Suite configuration

3.2.1 Suite configuration

The only file to configure is the "config.cfg" stored in the DHS Suite Easy Deploy tool installation folder.

The file "config.cfg" is composed by:

- The field "nfs_server_ip" that must contain the IP of the leader node hosting the Installation Tool.
- A list of tags to be set. Each tag's value represents the destination host for the distributed deployment of each DHS software component. The host value assigned to each tag must match the HOSTNAME column shown by the `docker node ls` command (executed from the leader node).

  If the DHS software components need to be deployed on a single machine, the value of each tag will correspond to the hostname of the machine running the DHS Suite Easy Deploy tool.

Table 3 - Tags definition

Tag name	DHS Software component
copsi_tag	COPSI
dafne_tag	DAFNE
gss_admin_tag	GSS Admin
gss_catalogue_tag	GSS Catalogue
gss_ingest_tag	GSS Ingest
gss_notification_tag	GSS Notification
iam_tag	KEYCLOAK
tf_tag	Transformation Framework

Please note that the Semantic Framework software must be deployed only on the node hosting the Installation Tool.

Check that each `docker-compose.yml` file of each DHS component is configured according to the applicable environment and to user specific needs. The NFS IP configuration of each `docker-compose.yml` file will be automatically substituted according to the parameter "nfs_server_ip" contained in the suite "config.cfg" file.

If it is needed to change the exposed port of a DHS component, modify the value of "published:" setting in each `docker-compose.yml` to be modified.

Before deploying the software components, please ensure that there is no conflict within the ports used on the same host.

3.2.2 DHS components configuration

Each DHS component requires specific configurations, which must be set according to the deployment environment. These configurations include database settings, network parameters, and service-specific options. It must be ensured that all configurations are thoroughly reviewed and adjusted to match the infrastructure setup.

All software packages can be configured changing the content of the following files directly in the DHS Suite Easy Deploy tool directory (relative path starting from the root installation folder of DHS Suite Easy Deploy tool):

- **COPSI:**

    - ./copsi_install_pkg/data/copsi/config/config.json
- **DAFNE:**
    - ./dafne_install_pkg/data/dafne/front-end/config/config.json
    - ./dafne_install_pkg/data/dafne/back-end/config/config.json
    - ./dafne_install_pkg/data/dafne/back-end/config/db_credentials.env
- **GSS Admin:**
    - ./gss_install_pkg/config/admin/application.properties
- **GSS Catalogue:**
    - ./gss_install_pkg/config/catalogue/application.properties
- **GSS Ingest:**
    - Xml configuration
        - ./gss_install_pkg/config/ingest/gss-consumer.xml
        - ./gss_install_pkg/config/ingest/gss-producer.xml
    - Admin API Configuration
        - ./gss_install_pkg/config/ingest/database-configuration-for-ingestion.properties
- **GSS Notification:**
    - ./gss_install_pkg/config/notification/consumer-for-notification.properties
- **KEYCLOAK:**
    - ./keycloak/docker-compose.yml
- **Semantic Framework:**
    - ./sf_install_pkg/config_nl/keycloak_configuration.json
    - ./sf_install_pkg/config_sf/configuration.json
    - ./sf_install_pkg/config_sf/keycloak_configuration.json
- **Transformation Framework:**
    - ./esa_tf_install_pkg/.env
    - ./esa_tf_install_pkg/config/esa_tf.config
    - ./esa_tf_install_pkg/config/hubs_credentials.yaml
    - ./esa_tf_install_pkg/config/traceability_config.yaml

Please refer to documentation of each software regarding how to configure the files above and only for GSS software please refer also to [COPE-SERCO-TN-23-1461] - GSS COTS installation v1.3 [RD-1].

3.3 Suite execution

The script "2-click-installer.sh" is the core of the DHS Easy Deploy tool and it is in the "/home/dhs-suite-easy-deploy-1.1.0/" path.

The installer script "2-click-installer.sh" has only one parameter, the list of software installable comma separated.

To install all software of the DHS suite, the command to be executed from 'root' user is:

- ./2-click-installer.sh "copsi,dafne,tf,sf,iam,gss-admin,gss-ingest,gss-catalogue,gss-notification"

It is also possible to install one or more DHS components using the 2-click-installer.sh command by passing the name of the components to be installed as comma-separated arguments e.g.:

- ./2-click-installer.sh "copsi,gss-notification"
- ./2-click-installer.sh "sf"

Table 4 - 2-click input parameters

2-click input parameter	DHS Software component
copsi	COPSI
dafne	DAFNE
gss-admin	GSS Admin
gss-catalogue	GSS Catalogue
gss-ingest	GSS Ingest
gss-notification	GSS Notification
iam	KEYCLOAK
sf	Semantic Framework
tf	Transformation Framework

Every command in the script above is executed as 'colluser' user, pre-created by the script itself before the execution of installation commands.

During the script execution the folder "/home/dhs-suite-easy-deploy-1.1.0" will be moved to the folder "/home/colluser/dhs-suite-easy-deploy-1.1.0". As consequence in the case the script must be re-lunched it can be executed from the folder "/home/colluser/dhs-suite-easy-deploy-1.1.0".

After running the 2-click installation, verify that the services are running with the command `docker ps` or check the logs with `docker logs <container_name>` to ensure the software is running properly. All the software,

excluding SF, will be deployed as a Docker stack. Monitor the Docker stack using the command `docker stack ls` to verify the correct initialization of the stack.

In case of a GSS installation from scratch, the first service to be installed is the "gss-admin". This is necessary to configure the datastores, producers and consumers via API.

3.3.1 Troubleshooting Common Issues

- **Network Issues:** Ensure all nodes can communicate with each other. Check firewall settings and network configurations.
- **Port Conflicts:** If services fail to start, check for port conflicts. Adjust port settings in configuration files.
- **Service Logs:** Review service logs for errors. Common commands include:
  - docker logs <service_name>
- **Resource Limitations**: Ensure the host machine meets the hardware requirements. Check CPU, RAM, and disk usage.
- **Proxy/NoProxy**: Ensure to configure the proxy according to the software configuration or in the case that a proxy is not used, ensure that the lines in the COPSI and DAFNE docker-compose file are uncommented.

3.4 Suite utilities

Under the folder "utils" several scripts are available to perform cleanup operations, such us clean docker volumes or removing docker tags assigned to the cluster nodes.

Here below a brief description of each script:

- *remove_all_tags.sh*
  - it removes all the tags from every node inside the cluster.
- *remove_tag.sh*
  - it removes a specific tag passed as argument from each node of the cluster.
- *remove_all_stacks.sh*
  - it deletes all the stacks of the cluster (it must be used carefully)
- *remove_single_stack.sh*
  - it deletes the stack of the software indicated as argument defined in [Table 4 - 2-click input parameters (i.e. copsi)
  - Since Semantic Framework is not deployed in stack, to stop all the running containers perform the following steps from the Installation tool folder:
    - cd sf_install_pkg
    - docker compose down

    in few seconds the container will be stopped and removed.
- *remove_volume.sh*

- it deletes the volumes of the software indicated as argument defined in [Table 4 - 2-click input parameters (i.e. copsi)

  In case of distributed deployment, it must be run on each node of the cluster.