SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT

Virtual Appliance installation manual

Document history

Version	Date	Status	Author	Comments
1	2017-06-14	Draft	Marcello Galli	First version of the manual
1.1	2017-06-27	Draft	Flavian Gheorghe	New structure and corrections
1.2	2017-06-29	Draft	Dick M.A. Schaap	Added introduction
1.3	2017-07-03	Draft	Marcello Galli Flavian Gheorghe	Updated structure of document and text
1.4	2017-07-27	Draft	Flavian Gheorghe	Updated chapter 4.5
1.5	2017-10-12	Draft	Flavian Gheorghe	Updated manual following testing by partners
1.6	2018-01-24	Draft	Marcello Galli Flavian Gheorghe	Updates for DM-1.4.7
1.6.1	2018-02-28	Draft	Marcello Galli	Added note for importing in vmware
1.6.2	2018-03-01	Draft	Flavian Gheorghe	Added note for importing in KVM. Updated chapter 5.
1.6.3	2018-03-01	Final	Marcello Galli Flavian Gheorghe	Added notes on security. Formatting and grammar corrections

Contents

1. Introduction

SeaDataNet provides online unified access to the vast resources of marine and ocean datasets, managed by the distributed data centres, via its portal website. The Common Data Index (CDI) Data Discovery and Delivery service gives users a highly detailed insight in the geographical coverage, and other metadata features of data across the different data centres. Users can also request access to identified datasets in a harmonised way, using a shopping basket. This requires a one-time registration to the Marine-ID AAA services and acceptance of the SeaDataNet data policy to become a SeaDataNet user. Thereafter users can follow the processing of requests via an online transaction register (Request Status Manager (RSM)) and can download datasets in the SeaDataNet standard formats.

Data centres can connect to the CDI Data Discovery and Access service to support automatic processing of data set requests, as far as possible. Therefore a data centre has to locally install a Java component, '**Download Manager (DM)**', that handles all communication between the data centre system and the CDI Request Status Manager (RSM) service. This ensures that the requested files are made ready for downloading by users (if not subject to access restrictions) via personal download pages at the data centre. The DM software delivers SeaDataNet NetCDF formats and/or ODV ASCII data files for profiles, time series and trajectories.

Through its cooperation with many EU projects, and its active role in the development of EMODnet, the number of connected data centres has steadily risen to 103 at present. Thus the CDI service provides metadata and access to more than 1.9 million data sets, originating from more than 580 organisations in Europe, covering physical, geological, chemical, biological and geophysical data, acquired in European waters and global oceans.

However the installation and configuration of the Download Manager software can be challenging due to different configurations, firewalls etc, which in practice results in having different versions installed among the connected data centres which results in possible service differences.

ENEA has explored ways to make the Download Manager easier to install and configure. One way forward is by providing the Download Manager as a Virtual Appliance that can be delivered and deployed including operating and other supporting software that otherwise have to be installed and configured one by one.

This manual provides guidance how to install and configure the first release of the Virtual Appliance as developed and tested by ENEA.

2. General description of the virtual machine

The starting point for the virtual appliance (in the following VM), is a “Tomcat on Apache” Turnkey virtual machine (version 14.02). TurnkeyLinux is an Israeli firm (see www.turnkeylinux.org) which offers pre-configured Linux virtual machines suited for specific tasks.

All the included software is open-source, and freely available: it consists of a minimal Debian Linux distribution (the stable version: Debian 8, codename Jessie), coupled with some procedure for initialization and management. In addition Java 8 (Openjdk-8) and Tomcat 8 have been installed. Also Oracle Java is installed; the user can change the used Java implementation by the “alternative” mechanism of Linux Debian. The VM is extensible: all the software packages in the Debian distribution can be easily installed. Security updates are automatically applied.

Instruction on operating and usage of a Turnkey VM are available at the mentioned site. Instructions specific to the “Tomcat on Apache” VM are: https://www.turnkeylinux.org/tomcat-apache

In order to share data, the data producer has only to deploy the VM in the right environment, load the data, prepare the related coupling files, input the correct center_id (EDMO code) and the public IP of the VM in some configuration files. However, a minimal knowledge of the management of a Linux system is required.

The ‘postfix’ mailer installed is configured for local delivery only. It is used by the system to send warning mails to the ‘root’ user in case of problems.

In the directory: /home/Maintenence there are some procedures which can be used as guidelines for some tests, moving the VM from test to production operating mode and for installing the Data manager on a Debian-like Linux machine.

Note: Different versions of the virtual machine are available for the dominant software virtualization platforms. If you are using the KVM, please read the instruction in section 5.2 on how to deploy the machine.

2.1. Users present in the Virtual Appliance

The prepared VM of the Virtual Appliance has only 3 users:

- root (the Linux privileged user); password: SeaDataCloud2017
- tomcat8;
- seadataclient;

tomcat8 was created without password and login access, since it is used only to run the batches and the servlet.

The seadataclient user was created without password, but it has login access. It is used for the tasks related to the data and vocabularies. The seadataclient user should be used instead of root, due to security reasons.

To easy tomcat administration and the installation and testing phases, the tomcat administrative web interface is active on the appliance. The user should **change user and password**, which are defined in the file "/etc/tomcat8/tomcat-users.xml, or deactivate the service, if not used. The default access credential are below:

- user: admin
- password: SeaDataTom2017.

A MySQL database is installed, but it is not used. The MySQL management credentials are:

- user: root
- password: SeaDataSQL2017.

It is strongly recommended that all the passwords present in this manual to be changed during the installation, since this document is publicly available. Section 3.3 of this manual contains a description on how to change the default passwords.

Note: The sudo application is not installed on this machine.

2.2. Menu of the TurnKey solution

At each start of the virtual machine a configuration panel is shown. It is used for a basic setup of the machine: setting or changing the host name, the IP number, the date, time and time zone. This menu will appear every time that the machine starts (after a reboot or after a shutdown).

2.3 Accessing the virtual machine

The VM can be accessed through the console (terminal), but also through the network.

The following services are available, mainly through a simple web interface and the https protocol (the certificate is self-signed):

- Port 80: to access to the Download Manager, via the http or the https protocol and a browser. A menu for management is reachable via the URL: http://host_IP/cp/ (where host_IP is the IP number or DNS name of the appliance); from there the Tomcat manager pages can be also reached.
- Port 12320: for a secure shell access via https and a browser. The "shellinabox" application is used.
- Port 12321: for the VM management via https and a browser. The "webmin" application is available, for a full and graphical management of the VM.
- Port 22: for ssh access; the sftp transfer protocol is enabled.

3. Installation procedure of the virtual machine

3.1. Install virtual machine monitor

If you organization is not using a virtualization machine monitor (hypervisor software), you can install VirtualBox. It is a virtual machine software solution supported by Oracle and it is completely free. It is available here for download: https://www.virtualbox.org/wiki/Downloads

If you use this software for the first time, the manual has to be read before installing the software. The manual for the software is available here: https://www.seadatanet.org/Software/Download-Manager

3.2. Deploy the virtual appliance

Download the OVA file containing the virtual appliance from .

Import virtual appliance by double clicking on the OVA file. Or using "File" → "Import appliance" from the VirtualBox Manager window. A dialog window will appear describing the machine presets. Click on the “Import” button.

After the import is finished, select the machine from the panel on the left. Go to Menu Devices->Network->Network Settings. You will provide the network connection type. This action has to be made considering the organisation computer network’s specification.

You can do one of the following:

1) You can select in the Adapter menu => NAT mode. Skip the form window for selecting the network configuration of the Virtual Appliance. You will forward the port used by the machine through, so that the machine is accessible from the outside.
2) Set Adapter => Bridged Adapter. If the organisation network is set to use the DHCP protocol to assign the IP addresses for new machines, then you will select in the start form of the Virtual Appliance for IP the DHCP option. Note that in the case of usage dynamic IP address you will have update the configuration of the Download Manager’ DM_Servlet, once the IP address is changed.

Start the machine and complete the TurnKey configuration form with information on the machine.

Note: that every time you restart the machine you will have fill the form above.

3.3. Secure the machine

Please, note that, to easy installation and testing, in the Virtual Appliance a strong security policy is not implemented. After installation and testing the appliance should be secured. It is up to each organization to secure the machine, the security measures depend on the specific policy adopted by each organization and the environment in which the appliance is installed.

In particular one or more of the following action should be considered:

ssh access: ssh access is, by default granted only to root. A good idea is to define one or more users, without special privileges, to be used to access the machine. The root user should be used only for system maintenance, with direct ssh access denied (editing the file: /etc/ssd/sshd_confif and setting "PermitRootLogin no")

webmin access: the webmin interface allow for an easy, graphical management of the appliance. This application listen on port: 12321. This port should be protected and accessible only to trusted machines. If not needed, the service can be deactivated.

Shell access: the "shellinabox" application is used to allow to login to the VM via a browser interface, this appliacation listen on port 12320. Also this port should be protected or the application deactivated, if not needed.

Tomcat webmin access: available on port 80, at the url: http://VM_IP/manager/, the username and password, defined in the file /etc/tomcat8/tomcat-users.xml, must be changed, the service can also be deactivated when not needed, considered that port 80 must be accessible from internet.

Turnkey web management: available also on port 80, at the url: http://VM_IP/cp , this is an interface for easy access on all the other services available on the VM. It is a Java appliance. This also should be hidden from the outside, or deactivated, if not needed. Java appliances are activated in the configuration file: /etc/tomcat8/mod_jk.conf.

Default web site: as a default, a simple web page: /var/www/index.html redirects to the Turnkey web management page. This page should be changed to avoid random visitors to be redirected to a management interface.

A firewall is not installed on the VM, but the "iptables" tool is available and can be used to limit the access to the VM or built a custom firewall. For iptables usage see:
https://www.frozentux.net/documents/iptables-tutorial/

The following guides can also be useful:

https://www.debian.org/doc/manuals/securing-debian-howto/ for securing Debian operating system

https://www.owasp.org/index.php/Securing_tomcat provides a great list of the actions which are required to be taken to secure the Tomcat installation.

3.4. Change of the existing users passwords

Log into the machine using the user root with the assigned password (SeaDataCloud2017). The recommended action is to change this user password. The command for this action is:

passwd root

Type twice the new password.

Before login of the seadataclient user, a password of the must be given. This will be done by logging as the root user and typing the following command:

passwd seadataclient

and similarly to change for the root user type twice the new password.

To change the password for the Tomcat webserver, follow the instructions open the /var/lib/tomcat8/conf/tomcat-users.xml, look under the tomcat-users tag for the tomcat8 user. After you have found it, change the existing password in the password attribute. The existing tag:

<user username="tomcat8" password=" SeaDataTom2017" roles="manager-gui"/>

For changing the root user password of the MySQL RDBMS use one of the methods presented in this article: https://www.techrepublic.com/article/how-to-set-change-and-recover-a-mysql-root-password/.

4. Configuration of the Download Manager

4.1 Download Manager's installation folder structure

The Download Manager (abbreviated as DM) version 1.4.7 is installed. The Download Manager is already configured for "test mode" configuration. The present installation was done following the instructions given in the Download Manager installation manual, which can be downloaded from: Software/Download-Manager/

The seadataclient user is used for for the management of the Download Manager: configuration files, data and other files for the Download Manager software are present under /home/seadataclient. This directory is writable also by the tomcat8 user. The structure of this directory is:

```
/home/seadataclient/
        tmpdir/            : temporary folder, used for installation
        dmtest-install/  : for the installation running in "testing mode"
                DM_batches/
                        DM_Batch.jar        : prepares data
                        DM_Checker.jar     : checks the coupling table
                        DM_ToolsBatch.jar : cleans the old zip files containing the client requested data
```

config/ : contains the configurations files for the data manager
config/BODC_vocabs/ : local version for the required BODC vocabularies
config/tmp/ : contains the temporary files created by the batches
commontest/
download/ : the zip files with the data files prepared by the DM
mapping/ : mapping files
logs/ : the log files created by the Download Manager components

dm-install/ : for the installation once it is running in "production mode"
DM_batches/
DM_Batch.jar : prepares data
DM_Checker.jar : data checks
DM_ToolsBatch.jar : cleans old files
config/ : contains the configurations files for this component
config/BODC_vocabs/ : local version for the required BODC vocabularies
config/ tmp/ : contains the temporary files created by the batches
common/ : data prepared by the DM in production
download/ : the zip files with the data files prepared by the DM
mapping/ : contains the mapping XML files
logs/ : the log files created by the Download Manager components
data/ : original user data read by the DM

These directories are accessible for reading and writing to both the seadataclient and the tomcat8 users.

The subdirectory config contains the configuration files:

coupling.txt : coupling table
config.properties : Download Manager configuration

The structure and syntax of the coupling table file and that of mapping files are described in the Download Manager user manual. It is available from: Software/Download-Manager/

The following sections describe the actions required for the configuration of the Download Manager and how to add the required files in the machine. *The manual will not go into in detail over the Download Manager, because it will then duplicate the manual of the Download Manager. That is why you are advised to consult that manual during the installation.*

4.2. Modify the Download Manager configuration files

The Download Manager software is already setup on the machine to run in "test mode".

You must only perform the following actions:

1) change the EDMO code present in the file config.properties of the DM_Batch (/home/seadataclient/dmtest-install/DM_batches/config/) from 0000 to your organization's EDMO code;
2) if not present, set the IP address of the machine in the configuration files web.xml of the Download Manager servlet found under /var/lib/tomcat8/webapps/dm-test/WEB-INF/;
3) if your organization is using NAT or a local machine to re-route the requests to the servlets or a proxy for the machine, then you need to update the IP address list for the rsm_server_name configuration parameter with the respective machine IP address. The RSM IP address is 77.87.163.227;
4) DO NOT ADD the local IP address in the web.xml serverName tag value. Only the external IP address or the FQDN of machine can be added.

For action 2 and 3 you will need to restart the tomcat8 service for the changes to take effect. You can do this using the Tomcat Manager interface which will be accessed via the browser.
Or you can run as root the command:
service tomcat8 restart

In the case your organization is storing the coupling table in a database, you will update the config.properties of the DM_Batch. You will modify the coupling_table_dbms parameter, filling the required parameters for connecting to the database. Don't forget to remove the "#" characters.

The configuration file provides examples of connection string for different relation database management systems.

4.3. Copy into the Virtual Appliance data files, coupling table file and mapping files

If you are using a Windows OS based environment as the host of the data files or mapping files:

You any data transfer program, which supports SFTP (such as FileZilla), to add the needed files. This data transfer will be done via port 22.

You can also install a CIFS (or SMB) client, make and mount a virtual folder, then copy the data to the Virtual Appliance.

Example for the CIFS utilization:

apt-get install cifs-utils

mount -t cifs -o username=name_user,password=password_user //IP_address/folder /mnt

cp /mnt/dm-data/* /home/seadataclient/data

For Linux based environment:

You can use scp utility to copy the needed files.

Example:

scp username@theserver:/folder_1/folder_2/filename.extension /home/seadataclient/data

or

scp username@theserver:/folder_1/folder_2/* /home/seadataclient/data

4.4. Test the Download Manager installation

The cronjobs for the batches of the Download Manager are already present in /etc/cron.d/. You do not need to create them.

After the configuration of the Download Manager is finished, you will test:

- if the log files of the DM_Batch contain entries every five minutes and there are no error message after you had finished configuration;
- run the following command in the terminal of the Virtual Apliance wget “” –O /home/seadataclient/ dmtest-install/DM_batches/tmp/web.xml to see if the servlets are accessible locally and run correctly.
- To test the access to the special HTML page from the outside via:

If the tests are passed, notify the MARIS team for the installation. You will send the email to . You will also send the IP address(es) of the machine, the URL for the controller and the userPage. The information to be sent to MARIS is presented in the Download Manager installation manual.

4.5. Move the Download Manager installation to production mode

The steps to move the Download Manager software to production mode are outlined below. They are in accordance with chapter 5 of the Download Manager installation manual.

Before proceeding stop the tomcat8 and apache2 services: (with: ‘service tomcat8 stop; service apache2 stop’). The steps below are executed as the root user.

1. Copy all the contents of the folder dmtest-install/ to the directory dm-install/
2. Write the following commands
   cd /home/seadataclient/dm-install/
   cp –r /home/seadataclient/dmtest-install/* .
3. Go to /home/seadataclient/dm-install/DM_batches/, open the file config.properties
   a. Change the value for the test_mode to 0

b. Change all the reference to the path: /home/seadataclient/commontest/ into: /home/seadataclient/common/ for the data files location
c. Check the rsm_server_name parameter. You can add here the IP of the machines you want to have access to the servlets.

4. All the subdirectories present in the dmtest-install should be also now in the directory dm-install. Check the directory structure and create missing directories, if any.
5. All the content of /home/seadataclient/dm-install/ must be accessible both to tomcat8 and seadatanet,
   a. this can be done by using the unix file permissions, which can be set with the following commands:

```
cd /home/
chown -Rv tomcat8:seadataclient seadataclient
chmod -Rv g+rw seadataclient
```

   b. As an alternative way the 'setfact' command can be used; i.e.: to set the correct user rights for the tmp directory: setfacl -m u:tomcat8:rwx /home/seadataclient/dm-install/DM_batches/config/tmp/

6. Change the existing cronjob to run the DM_Batch, DM_ToolsBatch in production mode by modifying in the /etc/cron.d/datamanager the path /home/seadataclient/dmtest-install/DM_batches/ to /home/seadataclient/dm-install/DM_batches/, also the 'HOME' parameter must be changed
7. Copy the DM_Servlet from dm-test to dm using the command

```
cp -r /var/lib/tomcat8/webapps/dm-test/* /var/lib/tomcat8/webapps/dm/
```

8. Set the permissions for the folder to the tomcat8 user:

```
cd /var/lib/tomcat8/webapps/
chown -Rv tomcat8:tomcat8 dm
```

9. In the file: /var/lib/tomcat8/webapps/dm/WEB-INF/classes/log4j2.xml you have to change /home/seadataclient/commontest/logs/dm_servlet.log into: /home/seadataclient/common/logs/dm_servlet.log . The full path of the log file is needed for some versions of the download manager.

10. Check that in the file: /var/lib/tomcat8/webapps/dm/WEB-INF/web.xml, you have the public IP of your VM

11. Check the file: /etc/tomcat8/mod_jk.conf change 'dm-test' into: 'dm'. You must have the lines:

```
JkMount /dm      ajp13_worker
JkMount /dm/*    ajp13_worker
```

or you will redirected to the test environment appliance instead of the production one.

12. Restart apache and tomcat (with : service apache2 start | service tomcat8 start)
13. Check the log files to verify that all is running as expected

5. Troubleshooting common problems

5.1. The RSM cannot access the machine?

The reason could be that the RSM IP address is not present in the configuration file of the Download Manager DM_Servlet. Check if the correct IP address 77.87.163.227 is present. If the machine is behind a proxy or firewall the correct settings have to be made.
The IP addresses and ports for the firewall rules that must be set for the Download Manager machine are described below:
1) Port 80 must be open to communication in order to allow client to download their data files
2) The RSM addresses are
77.87.163.227 for the requests coming the RSM to your machine,
seadatanet.maris2.nl (77.87.163.211) for the requests going from your machine to the RSM through port 80;
3) The addition to 80, the port 443 must be open the authentication, authorization and accounting service IFREMER CAS - users.marine-id.org (or the IP address 134.246.142.22)
in order for the clients to log into the personal page on the Download Manager;
4) Allow access through port 80 to the BODC vocab webservice served by the server with hostname vocab.nerc.ac.uk (or IP address 192.171.196.70);
5) The IP address of HCMR\ HNODC NAGIOS - 195.251.37.48 has to be kept in the DM_Servlet configuration file (also port 80);
6) OGS Nagios monitoring server has the address tritone.ogs.trieste.it (IP address 140.105.70.47).
To prevent unauthorized persons tampering with the DM, please, set up a rule in your proxy similar to these:

```
<LocationMatch "/dm/(controller|status)" >
    Order Deny,Allow
    Deny from all
    Allow from 77.87.163.227
</LocationMatch>
<LocationMatch "/dm/(status)" >
     Order Deny,Allow
     Deny from all
     Allow from 195.251.37.48
</LocationMatch>
<LocationMatch "/dm/(index.html)" >
    Order Deny,Allow
    Allow from all
</LocationMatch>
```

5.2. How do I run the DM_CheckerBatch, the DM_ToolsBatch?

The Download Manager provides information on the commands to run these batches, you only need to indicate the correct folders.

```
java –Xmx512m –jar /home/seadataclient/dmtest-install/DM_batches/DM_batches/DM_ToolsBatch.jar
-all -config /home/seadataclient/ dmtest-install/DM_batches/config/config.properties
```

5.3. Importing the virtual machine in VMWare VSphere 6.5

Problems have been reported when deploying the VMWare version of the Virtual Appliance on VSphere 6.5. The reason is the use of the SHA1 hash in the manifest file contained in the ova file (which is a tar archive). The problem can be solved by deleting the manifest file, or rebuilding the file with sha256 hashes, on a Linux box the manifest file can be deleted from the ova archive by the command:
"tar --delete DM147V02bVbox.mf -f DM147V02b-vmware.ova".
The other option is to rebuild the manifest file, putting in the file sha256 hashes; this can be done by using the ovftool program to manage the ova file and the sha256sum program to compute the hashes.

5.4. Importing the Virtual Appliance in KVM

The colleagues at HNODC\ HCMR have been able to import the OVA image of the Virtual Appliance using the steps below:

1) Untar the ova file

$ tar -xvf DM147V02b-vmware.ova

2) After the succesfull "untaring" locate the created .vmdk file(e.g. DM147V02Vbox-disk001.vmdk).

   a. For QCOW2 file

$ qemu-img convert -O qcow2 DM147V02Vbox-disk001.vmdk DM147V02Vbox-disk001.vmdk.qcow2

   b. For RAW file

$ qemu-img convert -O raw DM147V02Vbox-disk001.vmdk DM147V02Vbox-disk001.vmdk.raw

5.5. Back-up the machine

You can use the snapshot functionality of Virtual Box to backup\ restore the machine. A tutorial on this functionality is here: https://blog.en.uptodown.com/snapshots-virtualbox-tutorial/

A similar functionality is offered by VMWare.

It is highly recommended that you back- up the data files or the database, the metadata XML files in two different locations, situated in different places.