



Dataworks System Services Guide

UNAVCO initially established the GNSS data management service Dataworks as a full stack independent server running on Dell Hardware operating CentOS as its operating system. Beyond that initial development UNAVCO has pursued development of Dataworks for both a server and a cloud-based Virtual Machine environment. This document covers the preparation needed to take a new server and get all of the prerequisites in place for the Dataworks software installation.

In either the server or cloud environment it is possible to use other operating systems, but the tested and currently recommended operating system for Dataworks (considering safety, security, cost and ease of setup and administration) remains GNU/Linux CentOS version 6 and higher from the Centos Project at www.centos.org.

The Centos project provides a complete set of processing, memory, and storage space requirements, with installation instructions and manuals to establish a baseline server environment in both hardware and virtual machines, to establish a platform server for Dataworks GNSS data management. Additionally the installer should see Section 2 of this operations guide for specific Dataworks recommendations for user accounts.

After the initial setup of the server, Dataworks users will need to add the required Dataworks support software to the server in the appropriate order. The first package we recommend you install is the extended Java support environment using the latest Oracle JDK. The software and instructions for the full Java Development Kit needed to support Dataworks GSAC software can be found on the Oracle web site in the Downloads section.

After the JDK support is established the next recommended package to add is to setup and configure the database software. At its core Dataworks depends on the MySQL database service. Installing the latest version of MySQL is recommended and the software and instructions are also available from the download area of the Oracle web site.

After establishing the database installers should add and configure several services starting with VSFTP to support File Transfer services for data downloads. This product is available from security.appspot.com. Webpage access for GSAC then requires the user to install the Tomcat services from www.apache.org. We also strongly recommend the installer also add the latest Apache web server and the Webalizer web log analytics packages from www.webalizer.org.

The most effective way to install all these services in CentOS is as a root user or a user who has administrative privileges and can run the command `sudo`, and utilizing



the native software manager YUM or DNF. Section 3 of this manual has detailed Dataworks configuration files for the setup of the specific software packages.

For the automation of Dataworks file transfers the server depends on customized scripting that requires the Python version 2.7 or higher version 2 branch for the script execution language. Depending upon your version of CentOS (Python 2.7 is not available in CentOS 6), you may have to install Python 2.7 following the instructions on the Python support page at www.python.org. Besides the standard Python 2.7 language the downloader and mirroring scripts depend on two extended Python libraries to interact with the MySQL database.

Installers should acquire the mysqldb support library from the Python support website pypi.python.org by searching for "MySQL-python" and following the instructions located there for installing the latest version of the package in Python 2.7.

For the MySQL connector Python library, users should follow the Oracle supported instructions located at dev.mysql.com by selecting the download tab at the top of the page then select the MySQL Connectors tab on the left side of the page.

Other Recommended Resources:

Server Manuals for your selected hardware
VMWare or AWS online documentation for Virtual Machine setup
UNAVCO Dataworks Web site

<http://www.unavco.org/software/data-management/dataworks/dataworks.html>

Section 2

User Accounts: Recommended Practices.

UNAVCO's experience in data management operations suggested creating multiple user accounts for security and audit purposes. A 'developer' account is required to have sudo privileges to manage services and deploy applications. A second 'operator' account is used to manage the daily operations for Dataworks, and a 'dbadmin' account supports database management.

The installer may also wish to consider creating a web account to support the web services and the web statistics functions.

The recommended OS for Dataworks is the latest version of CentOS Linux with a default Bash shell for the user accounts for ops, dev, dbadmin, and root users.



The standard CentOS installs Java OpenJDK by default but as mentioned earlier, Dataworks GSAC requires the use of the latest version of Oracle's JDK.

The CentOS user accounts should be modified to use the Oracle JDK by editing the `/home/<user>/.bash_profile` file and editing the `PATH` variable extending the path with the destination of the JDK file:

```
# User specific environment and startup programs
```

```
PATH=/usr/java/jdk1.7.0_XX/bin:$PATH:$HOME/bin
```

```
export PATH
```

```
JAVA_HOME=/usr/java/jdk1.7.0_XX/
```

```
export JAVA_HOME
```

It is also recommended the installer update the user account for ops, by editing `.bashrc`, to add:

```
umask 0002
```

Section 3

Support Software Settings:

MySQL Database Server Install Notes

The startup script for the MySQL server is here:
`/etc/init.d/mysqld`

This script calls the `/usr/bin/mysqld_safe` script to start the database services

The database configuration file is here:
`/etc/my.cnf`

System Administrators may need to modify these scripts if the default values for port, logging, etc. need to be adapted for the work environment.

By default, the logs are written here:
`/var/log/mysqld.log`

By default the data space used by the server is here:
`/var/lib/mysql`



The installation directory for language and helper files is here:
`/usr/share/mysql`

Additional notes:

Installers should follow the instructions at dev.mysql.com/docs section for adding users.

§ After following instructions for installing and starting the services you will be required to create a password. For security reasons do not use that password for daily database operations.

§ Only use root to create a DB instance; then GRANT users permissions;

§ There is a default user named mysql, w/o password or privileges; try to avoid creating another conflicting user.

MySQL client tools include both the command line access and the Workbench can be found at:

<http://www.mysql.com/products/workbench/>

Tomcat Servlet Container

Docs - <http://tomcat.apache.org/>

Tomcat is Apache's Java based HTTP Web applications container for managing multi-user sessions to backend/server resources, e.g., databases, file systems, web services, and other servers, etc.

Dataworks GSAC is also a Java application that uses the Tomcat service to connect the Dataworks data collection to the web.

The startup script for the Tomcat service is here where ? represents a version number:

`/etc/init.d/tomcat?`

the configuration file is here:

`/etc/tomcat6/tomcat?.conf`

Users should start, stop or restart Tomcat with this command:

`$>sudo /etc/init.d/tomcat? start|stop|restart`

These specific changes need to be added to the configuration file:

`#1 set where the java installation resides`

`JAVA_HOME="/usr/java/jdk1.?.?_XX/"`



```
#2 set catalina_opts memory allocation
CATALINA_OPTS="$CATALINA_OPTS -server -Xms1024m -Xmx2048m"
```

Specific workspaces for Tomcat (installation directories) are here:

```
/usr/share/tomcat6/webapps
/usr/share/tomcat6/logs
/usr/share/tomcat6/conf
```

Tomcat should be installed with the Web manager application for controlling deployed applications including Dataworks GSAC. The manager applications is found at the url: <https://hostname:8080/manager/html>

Users should have their network administrators ensure this port is only accessible inside a firewall to keep the server secure.

VSFTPD – Secure ftp service

For Dataworks a secure FTP service runs to provide only anonymous downloads for data files. The FTP service is configured to block uploads and maintain local file ownership to preserve data integrity (can't overwrite or upload unauthorized files).

The config file is located at : /etc/vsftpd/vsftpd.conf

To start the service the following command should be run with a user that has administrative privileges:

```
$> sudo service vsftpd start|restart|stop
```

These are recommended changes to vsftpd.conf:

```
# Allow anonymous FTP? (Beware - allowed by default if you comment this out).
anonymous_enable=YES
#
# Uncomment this to allow local users to log in.
local_enable=YES
#
# Uncomment this to enable any form of FTP write command.
write_enable=NO
#
# The target log file can be vsftpd_log_file or xferlog_file.
# This depends on setting xferlog_std_format parameter
xferlog_enable=YES
....
#
# You may fully customise the login banner string:
ftpd_banner>Welcome to Dataworks FTP service.
```



```
#  
#  
pam_service_name=vsftpd  
userlist_enable=YES  
tcp_wrappers=YES  
# add root drop-in directories  
use_localtime=YES  
local_root=/data/pub  
anon_root=/data/pub  
---- end file ----
```

Apache HTTP Server

A Web server is an optional capability but it is highly recommended for supporting Dataworks clients and developers with a convenient access interface. The Apache web server front end to GSAC and other web applications allows greater configuration and control of public URLs, data access applications and web documents.

For the core Apache web server, the startup script is here:

```
/usr/sbin/apachectl
```

and the configuration file is here:

```
/etc/httpd/conf/httpd.conf
```

Full configuration and installation instructions are provided by the additional Apache documentation found at:

```
httpd.apache.org/docs/
```

To start the Apache web services, use the following command:

```
$>sudo /usr/sbin/apachectl -f /etc/httpd/conf/httpd.conf
```

Using the standard Apache configurations the html files for the primary web documents, are placed in the root directory:

```
/var/www/html/
```

Apache logs are here:

```
/var/log/httpd/
```

These are the recommended changes to httpd.conf:

Section1.

```
#  
# Dynamic Shared Object (DSO) Support  
#  
....
```

```
LoadModule jk_module modules/mod_jk.so
```

Section 2.

```
#  
# ServerAdmin: Your address, where problems with the server should be  
# e-mailed. This address appears on some server-generated pages, such  
# as error documents. e.g. admin@your-domain.com  
#  
ServerAdmin your-web-admin@ your-domain.com  
#  
# ServerName gives the name and port that the server uses to identify itself.  
#  
.....
```

When using Apache with Tomcat in order to seamlessly connect the web server with Tomcat there are three files to edit and one to review. Completing this integration allows administrators to by-pass Tomcat's built-in web server and control the urls to the Dataworks applications through the Apache web server configuration.

First, the Apache httpd.conf file must be edited in Section 2, add the following configuration:

```
# Tomcat jkmount  
Include /etc/httpd/conf/mod_jk.conf  
JkLogFile logs/mod_jk.log  
JkLogLevel info  
JkLogStampFormat "[%a %b %d %H:%M:%S %Y] "  
JkWorkersFile /etc/httpd/conf/workers.properties  
JkShmFile /etc/httpd/logs/mod_jk.shm  
# for each application, set the worker  
JkMount /dataworksgsac* tomcatd
```

Next add the /etc/httpd/conf/mod_jk.conf file to define the location of the shared library. The mod_jk.conf contents:

```
<IfModule !mod_jk.c>  
LoadModule jk_module "/etc/httpd/modules/mod_jk.so"
```

</IfModule>

Finally, add the `/etc/httpd/conf/workers.properties` file to define workers and how they connect to Tomcat. The `dataworks/workers.properties` file is:

```
# The workers that jk should create and work with
#
worker.list=tomcatd
#
# Defining a worker named ajp13w and of type ajp13
# Note that the name and the type do not have to match.
#
worker.tomcatd.type=ajp13
worker.tomcatd.host=hostname
# Tomcat6 is setup as 8009 for ajp
worker.tomcatd.port=8009
```

Note1: using the default AJP port defined in `/usr/share/tomcat6/conf/server.xml`

Note2: the 'tomcatd' worker matches the one defined in `httpd.conf`

Make sure to add a default `index.html` file to the `/var/www` directory. This should provide a standard web page interface for your server, should a user enter only a host name in the browser. If you need additional information on the use of the Apache server with Tomcat be sure to refer to both sets of documentation on the Apache.org web site. This is especially useful for understanding the worker relationship between Apache Web services and Tomcat.

Webalizer

Webalizer is another optional package for Dataworks that proves extremely useful for collecting web metrics for Dataworks (file downloads, users, busiest times, best times to do maintenance, etc.)

Verify that inside Apache's `httpd.conf`, you have established a Document root setting. The default value is `/var/www/` and you will use that to link the Webalizer usage directory to the Apache server instance.

Using the system command 'ln', add symlink in `/var/www/html` connecting `/var/www/html/usage` to `/var/www/usage`

Next stream line the webalizer configuration. Note there are two `webalizer.conf` files in `/etc/webalizer` and `/etc/httpd/conf.d`. Installers should delete `/etc/httpd/conf.d/webalizer.conf` and use other.

The following are changes to `webalizer.conf`



LogFile defines the web server log file to use. If not specified

.....

LogFile /var/log/xferlog

To start the software an administrator should run

```
$>/usr/bin/webalizer -c /etc /webalizer.conf
```

If you need further assistance with customizing the webalizer configuration be sure to consult the website for extended documentation.

Section 4 Additional tips

If you are having trouble establishing any of the services for Dataworks support be sure to consult the log files for that service. They may be able to provide you error messages and recommended solutions. Also be sure to check the log file errors against the online documentation for additional trouble shooting information.

MySQL Logs: /var/log/mysqld.log

Tomcat Logs (GSAC): /usr/share/tomcat6/logs

VSFTPD Logs: /var/log/xferlog

Apache Logs: /var/log/httpd/error_log

Tomcat Manager (GSAC): hostname:8080/manager/html

User/password:networkname/hostnamebackwards; These are set via /usr/share/tomcat6/conf/tomcat-users.xml

The remaining documentation establishing procedures to install the DB Schema, GSAC services, and Dataworks Downloader software can be found online at:

http://www.unavco.org/software/data-management/dataworks/lib/docs/Dataworks_for_GNSS.pdf

Installers can request additional assistance from the Dataworks email group at dataworks@unavco.org

Copyright UNAVCO 2016