|  |
| --- |
| Final Sotera Logo w ds 300dpi (2) |
| Datawake Dev Environment Installation Guide 1.0 |
| DARPA-Memex |
| Datawake-Dev-InstallationGuide.docx |
| Brodie McDougald, Michael Frame |
| 2/11/2015 |

Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Revision | Reviser | Description |
| 02/11/2015 | 1.0 | Brodie McDougald | Initial draft – Windows Deployment |
|  |  |  |  |
|  |  |  |  |

Table of Contents

[1 Setup for Windows 4](#_Toc411417479)

[1.1 Prerequisites 4](#_Toc411417480)

[1.2 Installation 4](#_Toc411417481)

[1.2.1 Install & Prepare Docker 4](#_Toc411417482)

[1.2.2 Install Datawake 4](#_Toc411417483)

[1.2.3 Check To See That Everything is Up and Running & Save State 5](#_Toc411417484)

[2 Setup for Linux 6](#_Toc411417485)

[2.1 Prerequisites 6](#_Toc411417486)

[2.2 Installation 6](#_Toc411417487)

[2.2.1 Install & Prepare Docker 6](#_Toc411417488)

[2.2.2 Install Datawake 6](#_Toc411417489)

[2.2.3 Check To See That Everything is Up and Running 7](#_Toc411417490)

# Setup for Windows

This section describes the setup of the Datawake dev environment under Windows. The environment consists of various system components (Kafka, Datawake, MySQL, Tangelo, etc) configured and running in containers utilizing Docker and Virtual Box.

## Prerequisites

There are a few prerequisites required to run Datawake in a Windows dev environment. These include the following:

* Microsoft Windows
* Boot2Docker (software and instructions available @ <https://docs.docker.com/installation/windows/> )
* A connection to the internet / Sourceforge

## Installation

The following steps need to be performed to get the Windows dev environment installed, configured, and working.

### Install & Prepare Docker

* + Install Boot2Docker via the downloadable version found at <https://docs.docker.com/installation/windows/>
  + Launch the Boot2docker application from the icon on your desktop.
  + When it finishes loading it will be at a ‘*docker@boot2docker:~$*’ prompt
  + Type ‘*git config --global core.autocrlf false*’
  + Create a directory for the GiT repository: ‘*sudo mkdir <whatever you want to call it>*’
  + cd to that directory

### Install Datawake

* + Type ‘*sudo git clone* [*https://github.com/Sotera/Datawake*](https://github.com/Sotera/Datawake)’ (this should clone the Datawake repo to your folder)
  + From the directory you just cloned Datawake into, type ‘*cd Datawake/dev-env*’
  + Create a fig.yml from the fig.yml template file
    - Type ‘*sudo cp fig.yml.template fig.yml* ‘
  + Next, edit the newly created ‘*fig.yml*’ file by typing ‘*sudo vi fig.yml*’.
    - Change the IP addresses listed as 192.168.59.103 to whatever the IP is of your Docker (this is shown when you first start Docker, or you can go to a Windows Command Prompt and type ‘*boot2docker ip’*). To enter edit mode, use the insert key.
    - Locate the “*volumes:*” section of **dwstream** and **datawake** in the fig.yml file. These start with *‘- /Users*’. These lines need to be changed to reflect the correct path of the GiT repo we cloned earlier. Following the steps above, this will likely be *‘- ~/<whatever you called the folder>/Datawake/…*’
    - Save your changes, and exit vi (esc key to exit insert mode, then *‘:wq’*)
  + Install Installation Support Files
    - Type ‘*tce-load -wi python.tcz*’
    - Type ‘*curl https://bootstrap.pypa.io/get-pip.py -o - | sudo python*’
    - Type ‘*curl https://bootstrap.pypa.io/ez\_setup.py -o - | sudo python*’
    - Type ‘*sudo pip install fig*’
  + Bring up the mysql container by typing ‘fig up –d mysql’
  + Insert the tables and mock test data into mysql by typing ‘*./init\_db.sh*’
  + Bring up all of the containers for Datawake by typing ‘*fig up –d*’

### Check To See That Everything is Up and Running & Save State

* + Open your browser and navigate to ‘*http://<your docker ip>/domain/loader*’
    - You should get the domain upload page
    - For the domain name fill in “memex.program”, for description type “*Memex Default domain*”
    - Browse to the default domain file from your local computer’s Datawake/etc directory and click the Submit button
    - Your new domain entry should appear in the table at the bottom of the page.
  + Open Oracle VM VirtualBox Manager (notice your boot2docker-vm is running)
    - Click on Snapshots on the upper right
    - Right Click on the ‘*Current State*’ listed in the snapshot tree, and choose ‘*Take Snapshot*’
    - Right Click on the ‘*boot2docker-vm*’ listed in the left-hand pane, and choose ‘*Close > Save State*’. These steps should save your docker vm in its current state with everything set up. It will also provide a clean starting point if you ever need to revert to a “clean slate”.

# Setup for Linux

## Prerequisites

There are a few prerequisites required to run Datawake in a Linux dev environment. These include the following:

* A Linux Distribution (Ubuntu, CentOS, etc). This guide is from a Ubuntu 14.04 perspective.
* Docker (software and instructions available @ https://docs.docker.com/installation/ )
* A connection to the internet / Sourceforge

## Installation

This installation assumes that you have already installed a base install of Ubuntu or other Linux distribution, and know the credentials that are required (including those of root).

### Install & Prepare Docker

* + Type ‘*ifconfig*’ at the Linux prompt to get your IP address (take note of it for later).
  + Type ‘*sudo apt-get update*’ to update apt-get if necessary.
  + Install the new version of Docker (not Docker.io)
    - Type ‘*curl –sSL* [*https://get.docker.com/ubuntu/*](https://get.docker.com/ubuntu/) *| sudo sh*’
  + Install apparmor (needed by docker to run properly on Ubuntu)
    - Type ‘*sudo apt-get install apparmor apparmor-utils -y*’
  + Install python pip
    - Type ‘*sudo apt-get install python-pip -y*’
  + Create a source folder for the Datawake project
    - From your user home directory (usually ‘~’), type ‘*sudo mkdir src*’
    - Type ‘*cd src*’

### Install Datawake

* + Pull the latest Datawake code to the newly created src directory
    - Type ‘*git clone* [*https://github.com/Sotera/Datawake*](https://github.com/Sotera/Datawake)’
  + Change directory to the Datawake/dev-env folder
    - Type ‘*cd Datawake/dev-env*’
  + Create a fig.yml from the fig.yml template file
    - Type ‘*sudo cp fig.yml.template fig.yml* ‘
  + Next, edit the newly created ‘*fig.yml*’ file by typing ‘*sudo vi fig.yml*’.
    - Change the IP addresses listed as 192.168.59.103 to whatever your Linux IP is (the one you got when you ran ‘*ifconfig*’ earlier). To enter edit mode, use the insert key.
    - Locate the “*volumes:*” section of **dwstream** and **datawake** in the fig.yml file. These start with *‘- /Users*’. These lines need to be changed to reflect the correct path of the GiT repo we cloned earlier. Following the steps above, this will likely be *‘- ~/src/Datawake/…*’
    - Save your changes, and exit vi (esc key to exit insert mode, then *‘:wq’*)

### Check To See That Everything is Up and Running

* + Start Docker and it is up and running
    - Type ‘*sudo service docker start*’
    - When you are returned to the prompt, type ‘*sudo docker ps*’ (you should see the docker containers listed)
    - if Docker failed to start, check to see if it still has a ‘*docker.pid*’ in /var/run
      * Type *‘- cat /var/run/docker.pid’*
      * If it exists, delete it by typing ‘*sudo rm /var/run/docker.pid*’
  + Install Fig
    - Type ‘*sudo pip install fig*’
  + Setup your MySQL container in docker
    - Type ‘*sudo fig up -d mysql*’
  + Setup your MySQL database, and create the test user
    - Type ‘*sudo ./init\_db.sh*’
  + Startup Docker and all its containers (using the settings in fig.yml that you created)
    - Type ‘*sudo fig up -d*’
  + Open your browser and navigate to ‘*http://<your ip address>/domain/loader*’
    - You should get the domain upload page
    - For the domain name fill in “*memex.program*”, for description type “*Memex Default domain*”
    - Browse to the default domain file from your local computer’s Datawake/etc directory and click the Submit button
    - Your new domain entry should appear in the table at the bottom of the page.