1 Anaconda Enterprise Notebook Runbook

- Version: 4.0-3 | October 12, 2016

Anaconda Enterprise Notebook (AEN) is a Python data analysis environment from Continuum Analytics. Accessed through a browser, Anaconda Enterprise Notebooks is a ready-to-use, powerful, fully-configured Python analytics environment. We believe that programmers, scientists, and analysts should spend their time analyzing data, not working to set up a system. Data should be shareable, and analysis should be repeatable. Reproducibility should extend beyond just code to include the runtime environment, configuration, and input data.

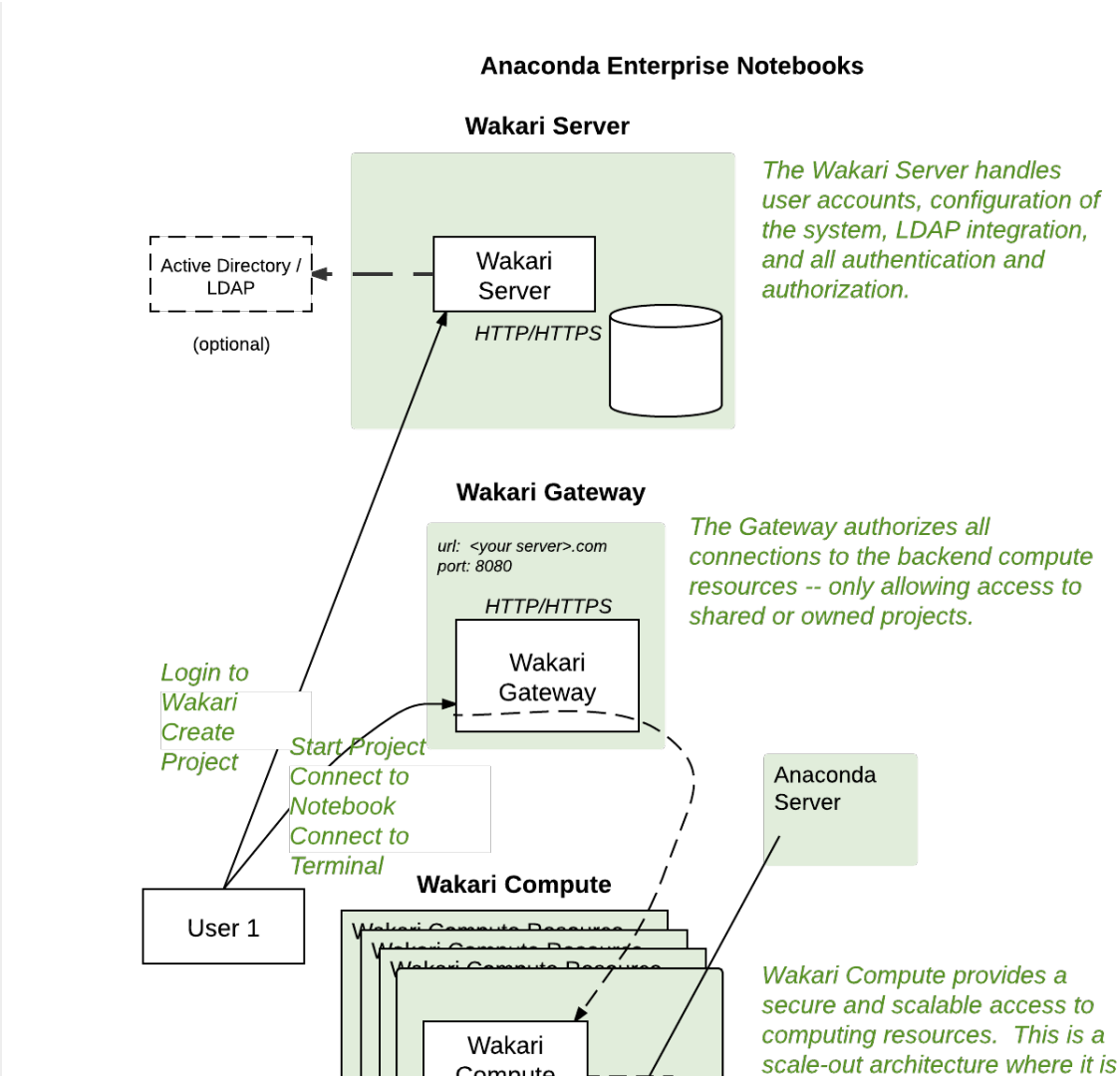
Anaconda Enterprise Notebooks makes it easy to start your analysis immediately.

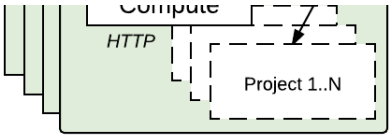
This runbook walks through the steps needed to install a basic Anaconda Enterprise Notebook system comprised of the front-end server, gateway, and two compute machines. The runbook is designed for two audiences: those who have direct access to the internet for installation and those where such access is not available or restricted for security reasons. For these restricted a.k.a. “Air Gap” environments, Continuum ships the entire Anaconda product suite on portable storage medium or as a downloadable TAR archive. Where necessary, additional instructions for Air Gap environments are noted. If you have any questions about the instructions, please contact your sales representative or Priority Support team, if applicable, for additional assistance.

AEN Server: The administrative front-end to the system. This is where users login to the system, where user accounts are stored, and where admins can manage the system.

AEN Gateway: The gateway is a reverse proxy that authenticates users and automatically directs them to the proper AEN Compute machine for their project. Users will not notice this component as it automatically routes them. One could put a gateway in each datacenter in a tiered scale-out fashion.

AEN Compute nodes: This is where projects are stored and run. AEN Compute machines only need to be reachable by the AEN Gateway, so they can be completely isolated by a firewall.





easy to add new compute resources.

Network Ports in Use

Wakari Server: 8080/80, 8443/443
Wakari Gateway: 8080
Wakari Compute: 5002
Anaconda Server: 8080/80
LDAP:389,636,3268

1.1 Requirements

1.1.1 Hardware Recommendations

AEN Server

- 2+GB RAM
- 2+CPU cores
- 20GB storage

AEN Gateway

- 2 GB RAM
- 2 CPU cores

AEN Compute (N-machines)

Configure to meet the needs of the projects. At least:

- 2GB RAM
- 2 CPU cores
- At least 20 GB

1.1.2 Software Requirements

- RHEL/CentOS 6.7 on all nodes (Other operating systems are supported, however this document assumes RHEL or CentOS 6.7)
- **/opt/wakari:** Ability to install here and at least 5GB of storage.
- **/projects:** Size depends on number and size of projects. At least 20GB of storage.
- **ACL:** This directory needs the filesystem mounted with Posix ACL support (Posix.1e). Check with `mount` and `tune2fs -l /path/to/filesystem | grep options`

1.1.3 Linux System Accounts Required

Some Linux system accounts (UIDs) are added to the system during installation. If your organization requires special actions, here is the list of UIDs:

- mongod (RHEL) or mongodb (Ubuntu/Debian): Created by the RPM or deb package
- elasticsearch: created by RPM or deb package
- nginx: created by RPM or deb package
- wakari: Created during installation of Anaconda Enterprise Notebooks

1.1.4 Software Prerequisites

AEN Server

- Mongo Version: >= 2.6.8 and < 3.0
- Nginx version: >= 1.4.0
- ElasticSearch
- Oracle JRE 8

AEN Compute

- git

1.1.5 Security Requirements

- root or sudo access
- SELinux in Permissive or Disabled mode - check with `getenforce`

1.1.6 Network Requirements

- TCP Ports

direction	type	port	protocol	optional	configurable	comments
inbound	TCP	80	HTTP	No	No	Server
in/out	TCP	8089		No	No	Gateway
in/out	TCP	5002		No	No	Compute

1.1.7 Other Requirements

Assuming the above requirements are met, there are no additional dependencies necessary for AEN.

1.2 Air Gap vs. Regular Installation

As stated previously, this document contains installation instructions for two audiences: those with internet access on the destination server(s) and those who have no access to internet resources. Many of the steps below have two sections: **Air Gap Installation** and **Regular Installation**. Those without internet access should follow the **Air Gap Installation** instructions and those with internet access should follow **Regular Installation** instructions.

1.2.1 Air Gap

Docs assume the air-gap data is available on target server at *$INSTALLER_PATH*. Refer to 1.2.1 Air Gap for instructions on obtaining full AirGap archive which contains AE-N.

If the full archive was not downloaded during AE-Repo install, the smaller AE-N archive called: *aen-`date +%Y-%m-%d`.tar* is about 1.5GB. Here's an example using *oct-2016* archive:

```
nohup curl -O https://s3.amazonaws.com/continuum-airgap/2016-10/aen-2016-09-30.tar
tar xf aen-2016-09-30.tar -C $INSTALLER_PATH
```

AE-N archive contents:

```
$INSTALLER_PATH
mongodb-org-tools-2.6.8-1.x86_64.rpm
mongodb-org-shell-2.6.8-1.x86_64.rpm
mongodb-org-server-2.6.8-1.x86_64.rpm
mongodb-org-mongos-2.6.8-1.x86_64.rpm
mongodb-org-2.6.8-1.x86_64.rpm
aen-compute-4.0.0-Linux-x86_64.sh
aen-server-4.0.0-Linux-x86_64.sh
aen-gateway-4.0.0-Linux-x86_64.sh
nginx-1.6.2-1.el6.ngx.x86_64.rpm
elasticsearch-1.7.2.noarch.rpm
jre-8u65-linux-x64.rpm
```

1.3 Download the Installers

Download the installers and copy them to the corresponding servers.

- **Air Gap Installation:** Copy installers from the Air Gap media/download
- **Regular Installation:**

```
RPM_CDN="https://820451f3d8380952ce65-4cc6343b423784e82fd202bb87cf87cf.ssl.cf1.rackcdn.com"
curl -O $RPM_CDN/aen-server-4.0.0-Linux-x86_64.sh
curl -O $RPM_CDN/aen-gateway-4.0.0-Linux-x86_64.sh
curl -O $RPM_CDN/aen-compute-4.0.0-Linux-x86_64.sh
```

1.4 Gather IP addresses or FQDNs

AEN is very sensitive to the IP address or domain name used to connect to the Server and Gateway components. If users will be using the domain name, you should install thecomponents using the domain name instead of the IP addresses. The authentication systemrequires the proper hostnames when authenticating users between the services.

Fill in the domain names or IP addresses of the components below and record the autogenerated wakari password in the box below after installing the AEN Server component.

Component	**Name or IP address**
AEN Server	
AEN Gateway	

1.5 Install AEN Server

The AEN server is the administrative frontend to the system. This is where users login to the system, where user accounts are stored, and where admins can manage the system.

1.5.1 AEN Server Preparation Prerequisites

1.5.1.1 Download Prerequisite RPMs

- **Air Gap Installation:** Copy RPMs from the Air Gap media
- **Regular Installation:**

```
RPM_CDN="https://820451f3d8380952ce65-4cc6343b423784e82fd202bb87cf87cf.ssl.cf1.rackcdn.com"
curl -O $RPM_CDN/nginx-1.6.2-1.el6.ngx.x86_64.rpm
curl -O $RPM_CDN/mongodb-org-tools-2.6.8-1.x86_64.rpm
curl -O $RPM_CDN/mongodb-org-shell-2.6.8-1.x86_64.rpm
curl -O $RPM_CDN/mongodb-org-server-2.6.8-1.x86_64.rpm
curl -O $RPM_CDN/mongodb-org-mongos-2.6.8-1.x86_64.rpm
curl -O $RPM_CDN/mongodb-org-2.6.8-1.x86_64.rpm
curl -O $RPM_CDN/elasticsearch-1.7.2.noarch.rpm
curl -O $RPM_CDN/jre-8u65-linux-x64.rpm
```

1.5.1.2 Install Prerequisite RPMs

```
sudo yum install -y *.rpm
sudo /etc/init.d/mongod start
sudo /etc/init.d/elasticsearch stop
sudo chkconfig --add elasticsearch
```

1.5.2 Run the AEN Server Installer

1.5.2.1 Set Variables and Change Permissions

```
export AEN_SERVER=<FQDN HOSTNAME> # Use the real FQDN
chmod a+x aen-*.sh                # Set installer to be executable
```

1.5.2.2 Run AEN Server Installer

```
sudo ./aen-server-4.0.0-Linux-x86_64.sh -w $AEN_SERVER
<license text>
...
...

PREFIX=/opt/wakari/wakari-server
Logging to /tmp/wakari_server.log
Checking server name
Ready for pre-install steps
Installing miniconda
...
...
Checking server name
Loading config from /opt/wakari/wakari-server/etc/wakari/config.json
Loading config from /opt/wakari/wakari-server/etc/wakari/wk-server-config.json


==================================

Created password '<RANDOM_PASSWORD>' for user 'wakari'

==================================


Starting Wakari daemons...
installation finished.
```

After successfully completing the installation script, the installer will create the administrator account (wakari user) and assign it a password:

```
Created password '<RANDOM_PASSWORD>' for user 'wakari'
```

Record this password. It will be needed in the following steps. It is also available in the installation log file found at `/tmp/wakari_server.log`

1.5.2.3 Start ElasticSearch

Start elasticsearch to read the new config file

```
sudo service elasticsearch start
```

1.5.2.4 Test the AEN Server install

Visit http://$AEN_SERVER. You should be shown the **“license expired”** page.

1.5.2.5 Update the License

From the **“license expired”** page, follow the onscreen instructions to upload your license file. After submitting, you should see the login page.

1.6 Install AEN Gateway

The gateway is a reverse proxy that authenticates users and automatically directs them to the proper AEN Compute machine for their project. Users will not notice this component as it automatically routes them.

1.6.1 Set Variables and Change Permissions

```
export AEN_SERVER=<FQDN HOSTNAME> # Use the real FQDN
export AEN_GATEWAY_PORT=8089
export AEN_GATEWAY=<FQDN HOSTNAME>  # will be needed shortly
chmod a+x aen-*.sh              # Set installer to be executable
```

1.6.2 Run Wakari Gateway Installer

```
sudo ./aen-gateway-4.0.0-Linux-x86_64.sh -w $AEN_SERVER
<license text>
...
...

PREFIX=/opt/wakari/wakari-gateway
Logging to /tmp/wakari_gateway.log
...
...
Checking server name
Please restart the Gateway after running the following command
to connect this Gateway to the AEN Server
...
```

Note

replace **password** with the password of the wakari user that was generated during server installation.

1.6.3 Register the AEN Gateway

The AEN Gateway needs to register with the AEN Server. This needs to be authenticated, so the wakari user’s credentials created during the AEN Server install need to be used. **This needs to be run as sudo or root** to write the configuration file: `/opt/wakari/wakari-gateway/etc/wakari/wk-gateway-config.json`

```
sudo /opt/wakari/wakari-gateway/bin/wk-gateway-configure \
--server http://$AEN_SERVER --host $AEN_GATEWAY \
--port $AEN_GATEWAY_PORT --name Gateway --protocol http \
--summary Gateway --username wakari \
--password '<USE PASSWORD SET ABOVE>'
```

1.6.3.1 Ensure Proper Permissions

```
sudo chown wakari /opt/wakari/wakari-gateway/etc/wakari/wk-gateway-config.json
```

1.6.3.2 start the gateway

```
sudo service wakari-gateway start
```

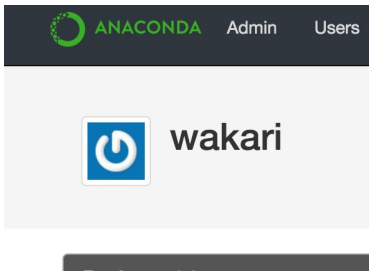
Note

Ignore any errors about missing /lib/lsb/init-functions

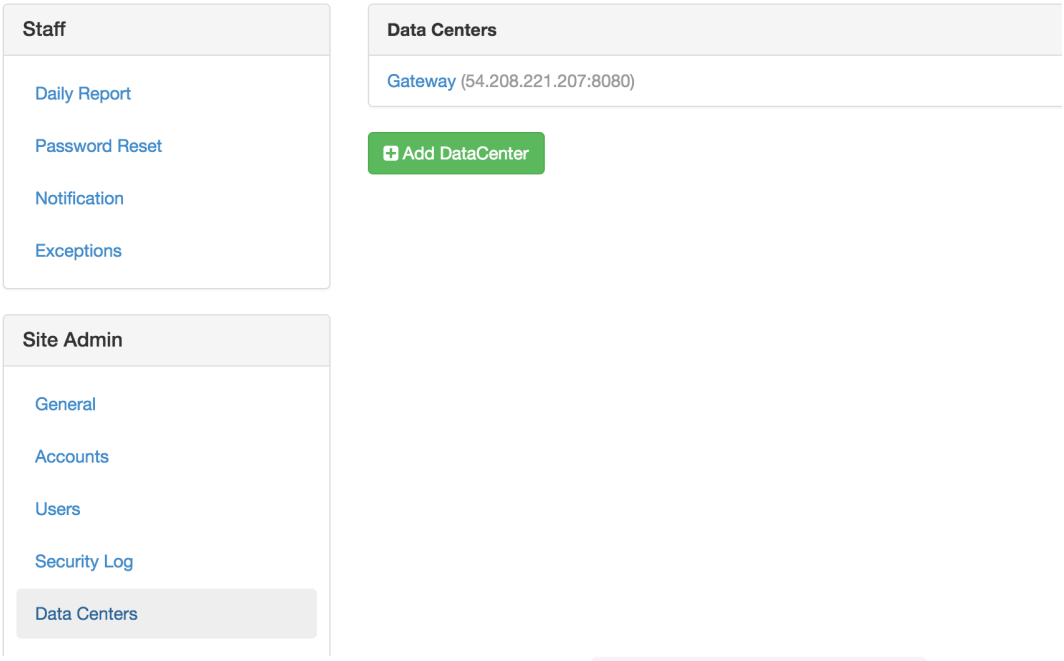
1.6.3.3 Verify the AEN Gateway has Registered

1. Login to the AEN Server using Chrome or Firefox browser using the wakari user.

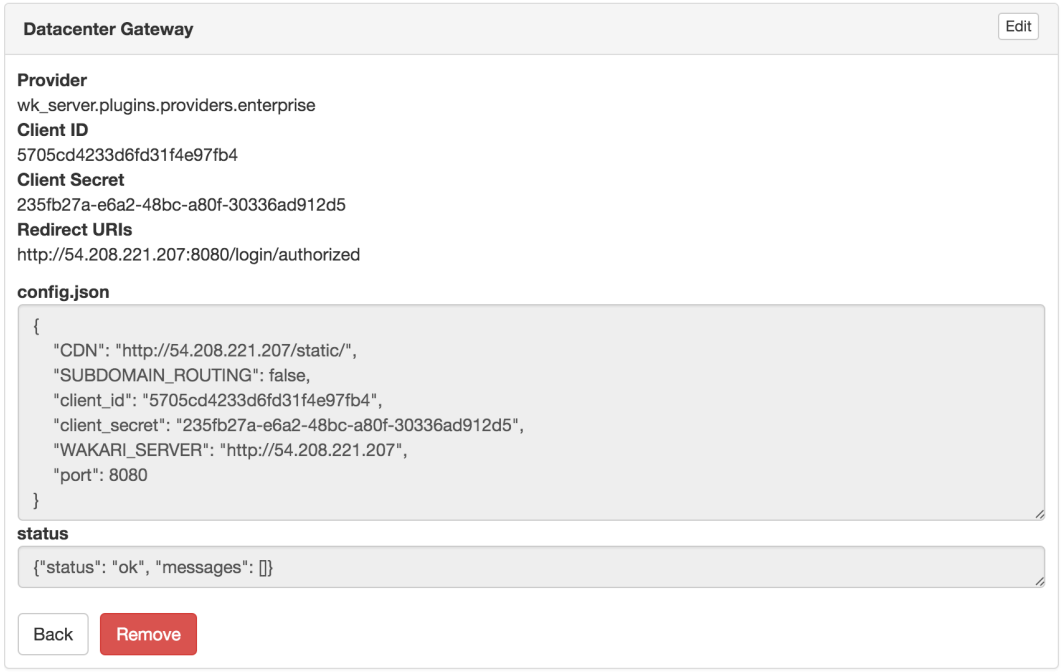
2. Click the Admin link in the toolbar



3. Click the Datacenters subsection and then click your datacenter:



4. Verify that your datacenter is registered and status is `{"status": "ok", "messages": []}`



1.7 Install AEN Compute

This is where projects are stored and run. Adding multiple AEN Compute machines allows one to scale-out horizontally to increase capacity. Projects can be created on individual compute nodes to spread the load.

1.7.1 Set Variables and Change Permissions

```
export AEN_SERVER=<FQDN HOSTNAME> # Use the real FQDN
chmod a+x aen-*.sh                # Set installer to be executable
```

1.7.2 Run AEN Compute Installer

```
sudo ./aen-compute-4.0.0-Linux-x86_64.sh -w $AEN_SERVER
...
...
PREFIX=/opt/wakari/wakari-compute
Logging to /tmp/wakari_compute.log
Checking server name
...
...
Initial clone of root environment...
Starting Wakari daemons...
installation finished.
Do you wish the installer to prepend the wakari-compute install location
to PATH in your /root/.bashrc ? [yes|no]
[no] >>> yes
```

1.7.3 Configure AEN Compute Node

Once installed, you need to configure the Compute Launcher on AEN Server.

1. Point your browser at the AEN Server
2. Login as the wakari user
3. Click on the Admin link in the top navbar
4. Click on Enterprise Resources in the left navbar
5. Click on Add Resource
6. Select the correct (probably the only) Data Center to associate this Compute Node with
7. For URL, enter **http://$AEN_COMPUTE:5002**.

   Note

   If the Compute Launcher is located on the same box as the Gateway, we recommend using **http://localhost:5002** for the URL value.

8. Add a Name and Description for the compute node
9. Click the Add Resource button to save the changes.

1.7.4 Configure conda to use local on-site Anaconda Enterprise Repo

This integrates Anaconda Enterprise Notebooks to use a local onsite Anaconda Enterprise Repository server instead of Anaconda.org.

1.7.4.1 Edit the condarc on the Compute Node

Note

If there are some channels below that you haven't mirrored, you should remove them from the configuration.

```
#/opt/wakari/anaconda/.condarc
channels:
   - defaults

create_default_packages:
   - anaconda-client
   - python
   - ipython-we
   - pip

# Default channels is needed for when users override the system .condarc
# with ~/.condarc.  This ensures that "defaults" maps to your Anaconda Server and not
# repo.continuum.io
default_channels:
   - http://<your Anaconda Server name:8080/conda/anaconda
   - http://<your Anaconda Server name:8080/conda/wakari
   - http://<your Anaconda Server name:8080/conda/anaconda-cluster
   - http://<your Anaconda Server name:8080/conda/r-channel

# Note:  You must add the "conda" subdirectory to the end
channel_alias: http://<your Anaconda Server name:8080/conda
```

1.7.4.2 Configure Anaconda Client

Anaconda client lets users work with the Anaconda Repository from the command-line. Things like the following: search for packages, login, upload packages, etc. The command below will set this value globally for all users.

Run the following command filling in the proper value. Requires *sudo* since config file is written to root file system: */etc/xdg/binstar/config.yaml*. This sets the default config for *anaconda-client* for all users on compute node.

```
sudo /opt/wakari/anaconda/bin/anaconda config --set url http://<your Anaconda Server>:8080/api -s
```

Congratulations! You've now successfully installed and configured Anaconda Enterprise Notebook.

1.7.4.3 Test AEN Compute conda config

Test the conda config is correct by ensuring the default environment gets constructed correctly.

```
sudo -u wakari /opt/wakari/anaconda/bin/conda create -p /opt/wakari/test_default --clone root
sudo rm -rf /opt/wakari/test_default
```

1.8 PAM Authentication (optional)

This documents the process to configure the preliminary PAM authentication mechanism for AEN 4.0.0.

1. remove *wk-server* from the current *supervisord* set:

   ```
   /opt/wakari/wakari-server/bin/supervisorctl stop wk-server
   mv /opt/wakari/wakari-server/etc/supervisord/conf.d/wk-server.conf      \
   /opt/wakari/wakari-server/etc/supervisord/conf.d/wk-server.conf.SUSPEND \
   /opt/wakari/wakari-server/bin/supervisorctl reload
   ```

2. install the new package:

   ```
   /opt/wakari/miniconda/bin/conda install \
       -p /opt/wakari/wakari-server \
       /path/to/wakari-server-1.10.9-0.tar.bz2
   ```

3. modify the configuration to utilize the new auth method:

   ```
   vim /opt/wakari/wakari-server/etc/wakari/wk-server-config.json
   ```

   change the entry for the line *"accounts":* to read instead:

   ```
   "accounts": "wk_server.plugins.accounts.pam",
   ```

4. connect the new *initd* script:

   ```
   cd /etc/init.d
   ln -s /opt/wakari/wakari-server/etc/init.d/wakari-server-root
   chkconfig --add wakari-server-root
   service wakari-server-root start
   service wakari-server-root status
   ps aux | grep wk-server
   ```

5. restart/check the *worker* service is running (and only the *worker*) under the *supervisord* watchdog:

   ```
   /opt/wakari/wakari-server/bin/supervisorctl start all
   /opt/wakari/wakari-server/bin/supervisorctl status
   ```

Back to top