Anaconda Enterprise Notebook Runbook

Citi Air-Gap Install (minimal) 2016-05-13

Anaconda Enterprise Notebook (AEN) is a Python data analysis environment from Continuum Analytics. Accessed through a browser, Anaconda Enterprise Notebook 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 Notebook 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. This runbook is designed for offline environments where public Internet 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.

NOTE: This component of Anaconda Enterprise was formerly known as *Wakari*, and that name still appears in the software and the installation and configuration information below.

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.

---

1. Requirements

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

1.2 OS 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.

  NOTE: This directory needs the filesystem mounted with Posix ACL support (Posix.1e). Check with `mount` and `tune2f -l /path/to/filesystem | grep options`

1.3 Software Prerequisites

AEN Server

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

NOTE: For Air Gap installations, Oracle JRE must already be installed, and if using Python 3.5 then Anaconda Repository must also be installed.

AEN Compute

- git

1.4 Security Requirements

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

1.5 Network Requirements

TCP Ports

- Server: 80
- Gateway: 8080
- Compute: 5002

1.6 Other Requirements

Assuming the above requirements are met, there are no additional dependencies necessary for Anaconda Enterprise Notebooks.

---

2. Download the Installers

Download the installers and copy them to the corresponding servers. The Publisher should be installed on the AEN Server machine.

```
1.  RPM_CDN="https://820451f3d8380952ce65-4cc6343b423784e82fd202bb87cf87cf.ssl.cf1.rackcdn.com"
```

```
curl -O $RPM_CDN/wakari-server-0.10.0-Linux-x86_64.sh
curl -O $RPM_CDN/wakari-gateway-0.10.0-Linux-x86_64.sh
curl -O $RPM_CDN/wakari-compute-0.10.0-Linux-x86_64.sh
curl -O $RPM_CDN/wakari-publisher-0.10.0-Linux-x86_64.sh
```

3. 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 wi 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	
AEN Compute	

4. Install AEN Server

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

4.1 AEN Server Preparation Prerequisites

Download Prerequisite RPMs

```
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
```

Install Prerequisite RPMs

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

```
4.  sudo chkconfig --add elasticsearch
```

4.2 Run the AEN Server Installer

Set Variables and Change Permissions

```
1.  export AEN_SERVER=<FQDN HOSTNAME> # Use the real FQDN
2.  chmod a+x wakari-*.sh               # Set installer to be executable
3.
4.  sudo ./wakari-server-0.10.0-Linux-x86_64.sh -w $AEN_SERVER
```

Run AEN Server Installer

```
1.  sudo ./wakari-server-0.10.0-Linux-x86_64.sh -w $AEN_SERVER
2.  <license text>
3.  ...
4.  ...
5.
6.  PREFIX=/opt/wakari/wakari-server
7.  Logging to /tmp/wakari_server.log
8.  Checking server name
9.  Ready for pre-install steps
10. Installing miniconda
11. ...
12. ...
13. Checking server name
14. Loading config from /opt/wakari/wakari-server/etc/wakari/config.json
15. Loading config from /opt/wakari/wakari-server/etc/wakari/wk-server-config.json
16.
17. ====================================
18. Created password '<RANDOM_PASSWORD>' for user 'wakari'
19. ====================================
20.
21. Starting Wakari daemons...
22. installation finished.
```

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

```
1.  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` .

Restart ElasticSearch

Restart ElasticSearch to read the new config file

```
sudo service elasticsearch restart
```

Test the AEN Server install

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

Update the License

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

---

5. 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.

5.1 Set Variables and Change Permissions

```
export AEN_SERVER=<FQDN HOSTNAME> # Use the real FQDN
export AEN_GATEWAY_PORT=8080
export AEN_GATEWAY=<FQDN HOSTNAME>  # will be needed shortly
chmod a+x wakari-*.sh               # Set installer to be executable
```

5.2 Run AEN Gateway Installer

```
sudo ./wakari-gateway-0.10.0-Linux-x86_64.sh -w $WAKARI_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 Wakari Server

PATH=/opt/wakari/wakari-gateway/bin:$PATH \
/opt/wakari/wakari-gateway/bin/wk-gateway-configure \
--server http://1.1.1.1 --host 1.1.1.2 --port 8080 --name Gateway \
--protocol http --summary Gateway --username wakari --password password
```

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

5.3 Start the AEN Gateway

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

5.4 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 root** to write the configuration file: `/opt/wakari/wakari-gateway/etc/wakari/wk-gateway-config.json`

```
PATH=/opt/wakari/wakari-gateway/bin:$PATH \
/opt/wakari/wakari-gateway/bin/wk-gateway-configure \
--server http://$WAKARI_SERVER --host $WAKARI_GATEWAY \
--port $WAKARI_GATEWAY_PORT --name Gateway --protocol http \
--summary Gateway --username wakari \
--password '<USE PASSWORD SET ABOVE>'
```

5.5 Ensure Proper Permissions

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

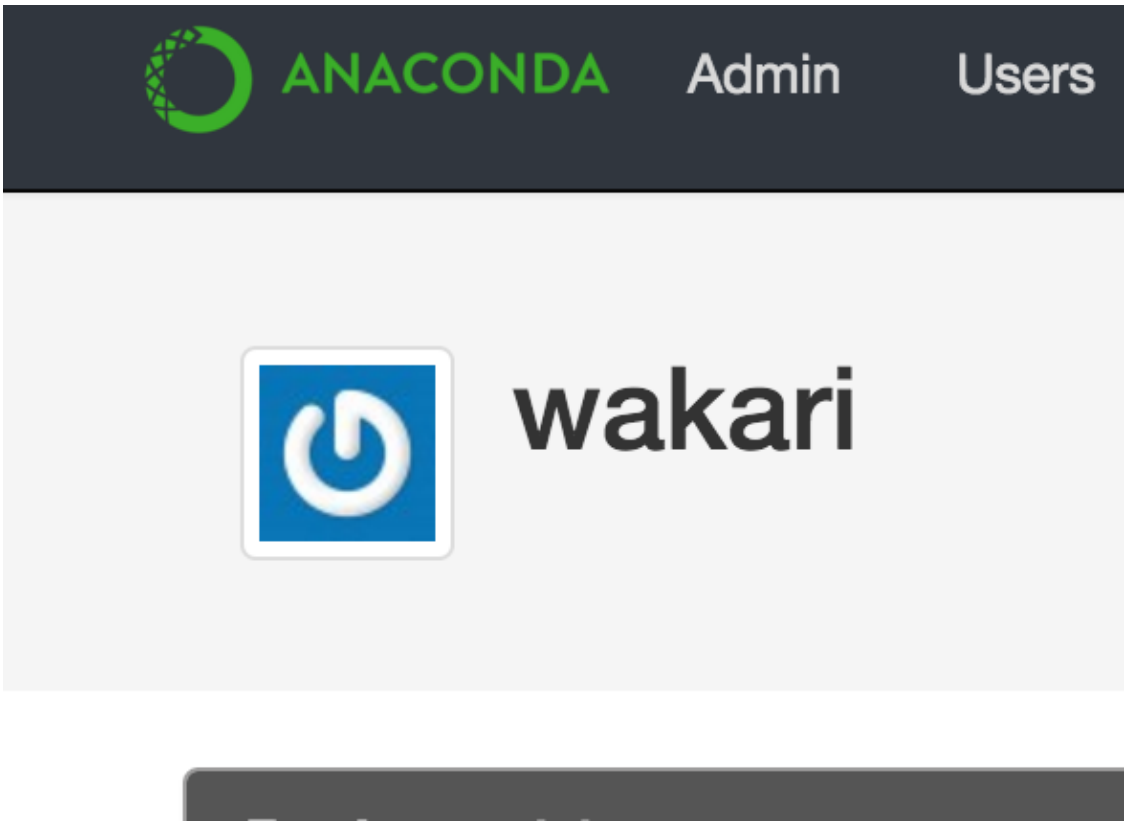
5.6 Restart the gateway to load the new configuration file

```
sudo service wakari-gateway restart
```

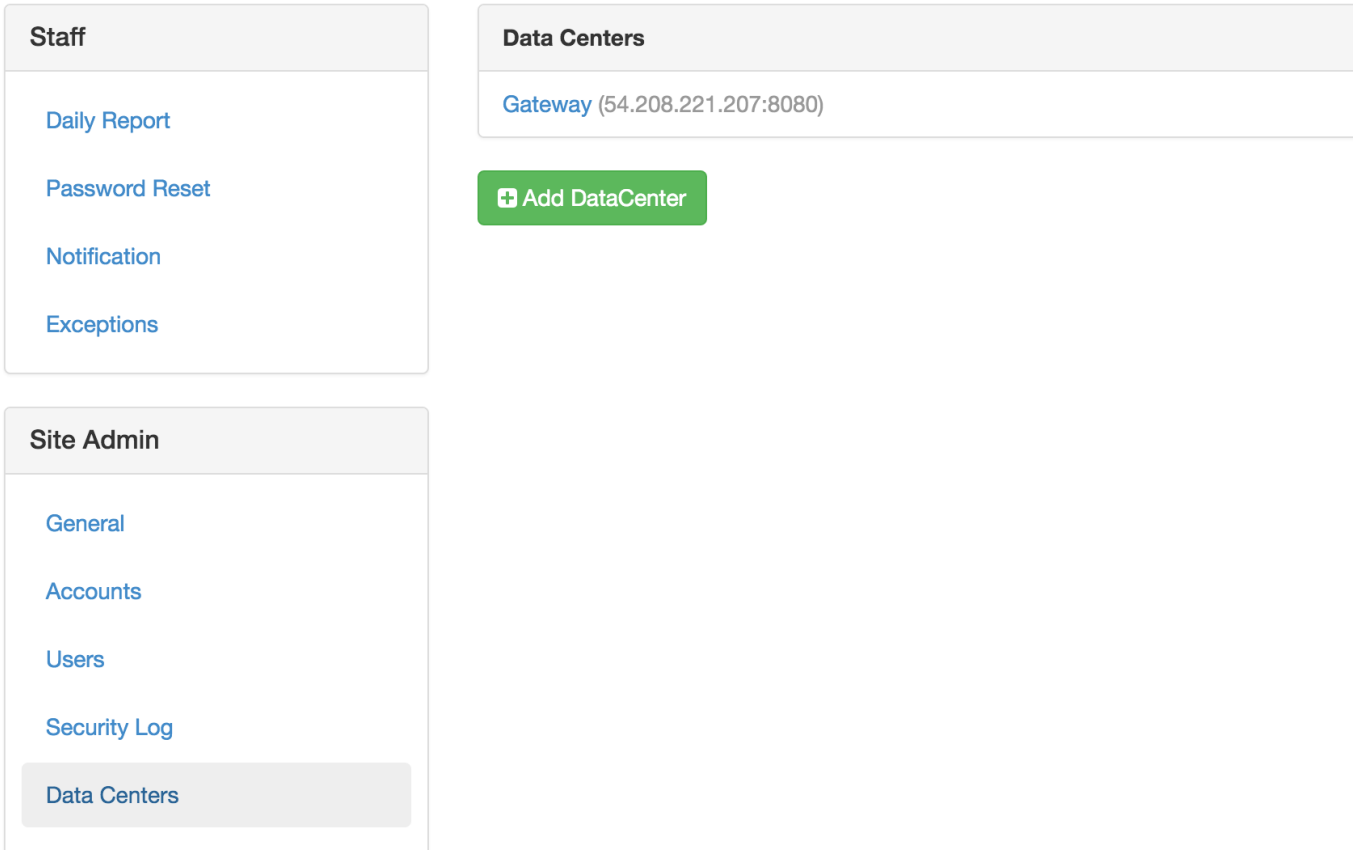
NOTE: Ignore any errors about missing /lib/lsb/init-functions

5.7 Verify the AEN Gateway has Registered

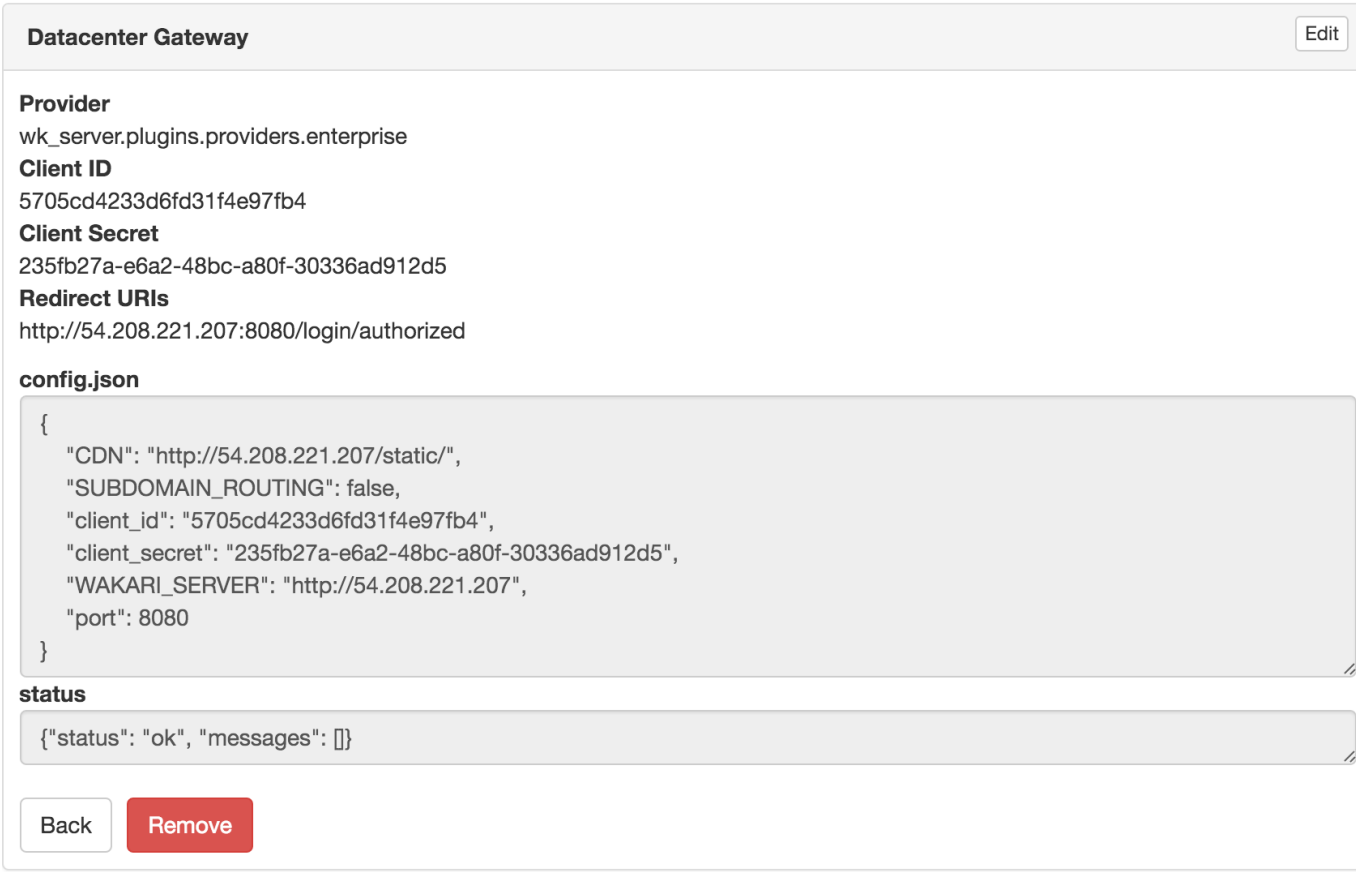
1. Login to the AEN Server using Chrome or Firefox browser using the wakari user.
2. Click the Admin link in the toolbar



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



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



6. 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.

6.1 Set Variables and Change Permissions

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

6.2 Run AEN Compute Installer

```
sudo ./wakari-compute-0.10.0-Linux-x86_64.sh -w $WAKARI_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
```

6.3 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://$WAKARI_COMPUTE:5002**.

   NOTE: If the Compute Launcher is located on the same box as the Gateway, we recommend using **http://localhost:500** for the URL value.
8. Add a Name and Description for the compute node
9. Click the Add Resource button to save the changes.

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