Deploying Hortonworks Sandbox on VMWare

Introduction

This tutorial walks through the general approach for installing the Hortonworks Sandbox (HDP or HDF) onto VMware on your computer.

Prerequisites

- Download the Hortonworks Sandbox
- VMWare Installed, version 14 or newer
  - VMWare Workstation For Linux/Windows
  - VMWare Fusion For macOS
- A computer with minimum **10 GB RAM** dedicated to the virtual machine

Outline

- Import Hortonworks Sandbox
  - VMWare Workstation
  - VMWare Fusion
- Update Virtual Machine Memory
- Enable Connected Data Architecture (CDA) - Advanced Topic
- Further Reading

Import Hortonworks Sandbox

VMWare Workstation

Open VMWare and select **Open a Virtual Machine**:

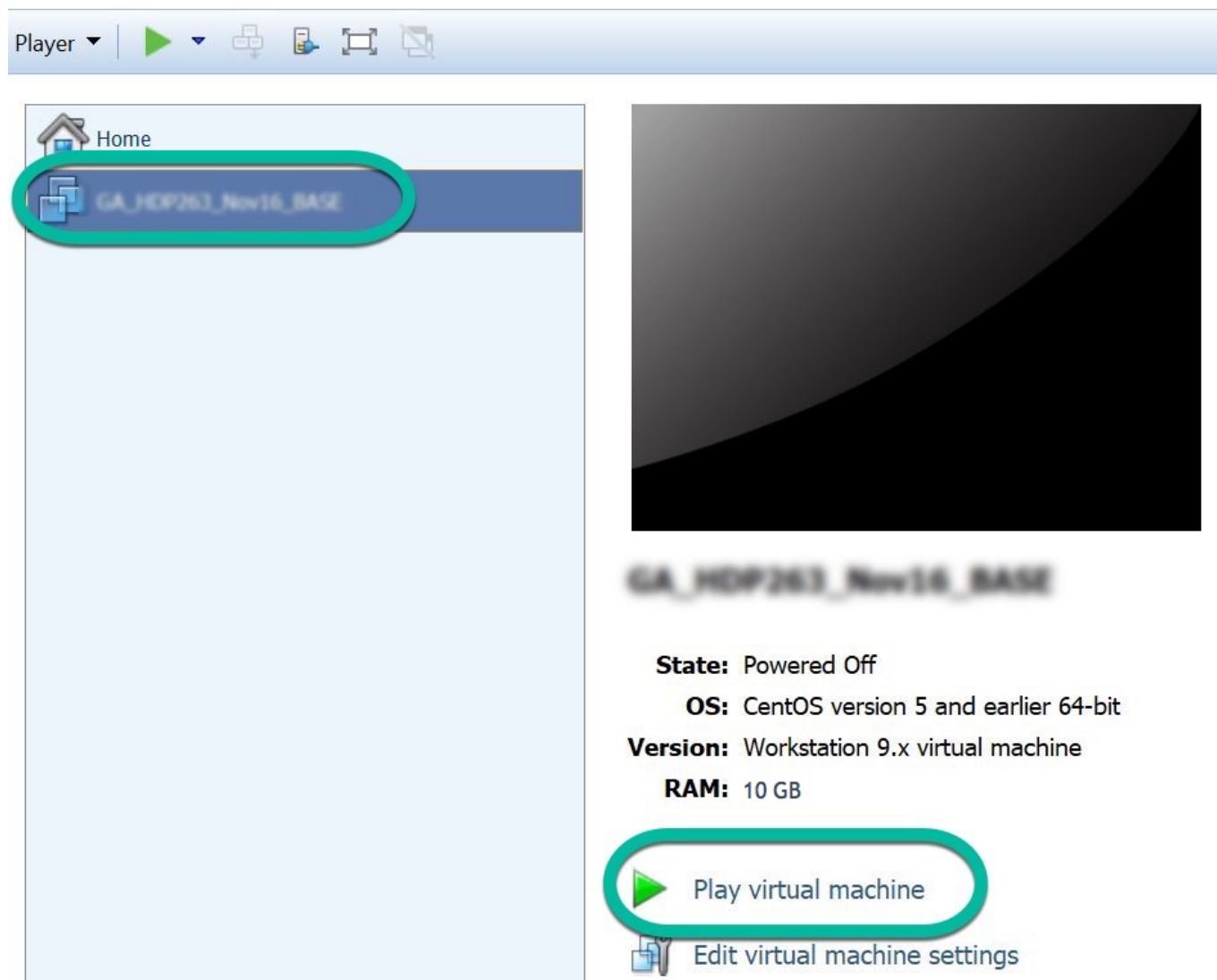Browse and select the sandbox image you downloaded and click **Open**.

Next, you're given the opportunity to save the virtual machine under a different name and/or location. After changes are made, click **Import**. You should then see the importing progress dialog:

Once import is complete, you will see the newly imported machine listed.

Select the virtual machine you want to start and click **Play virtual machine**.

When the virtual machine fully boots up, you will see a similar screen as below.

Enter the URL into your browser - you're ready to begin using the sandbox.

Welcome to the Hortonworks Sandbox!

VMWare Fusion

Open VMWare and navigate to **File -> Import... -> Choose File...**.
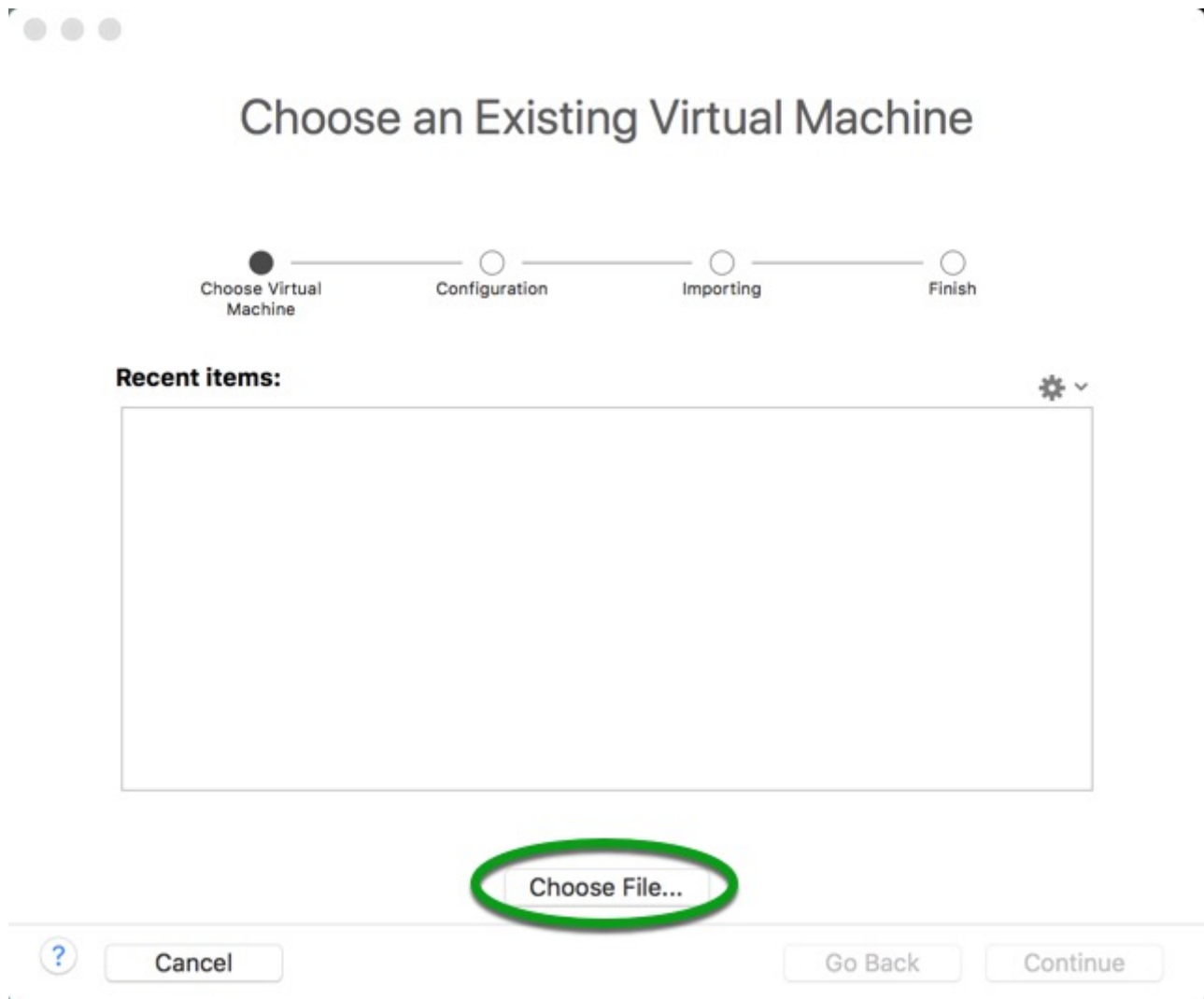
Select the sandbox image you downloaded and click **Open**, then **Continue**.

You are given the opportunity to save the virtual machine under a different name/location. If you have no preference in renaming, you can just leave the default name and click Save. You should then see the importing progress dialog:
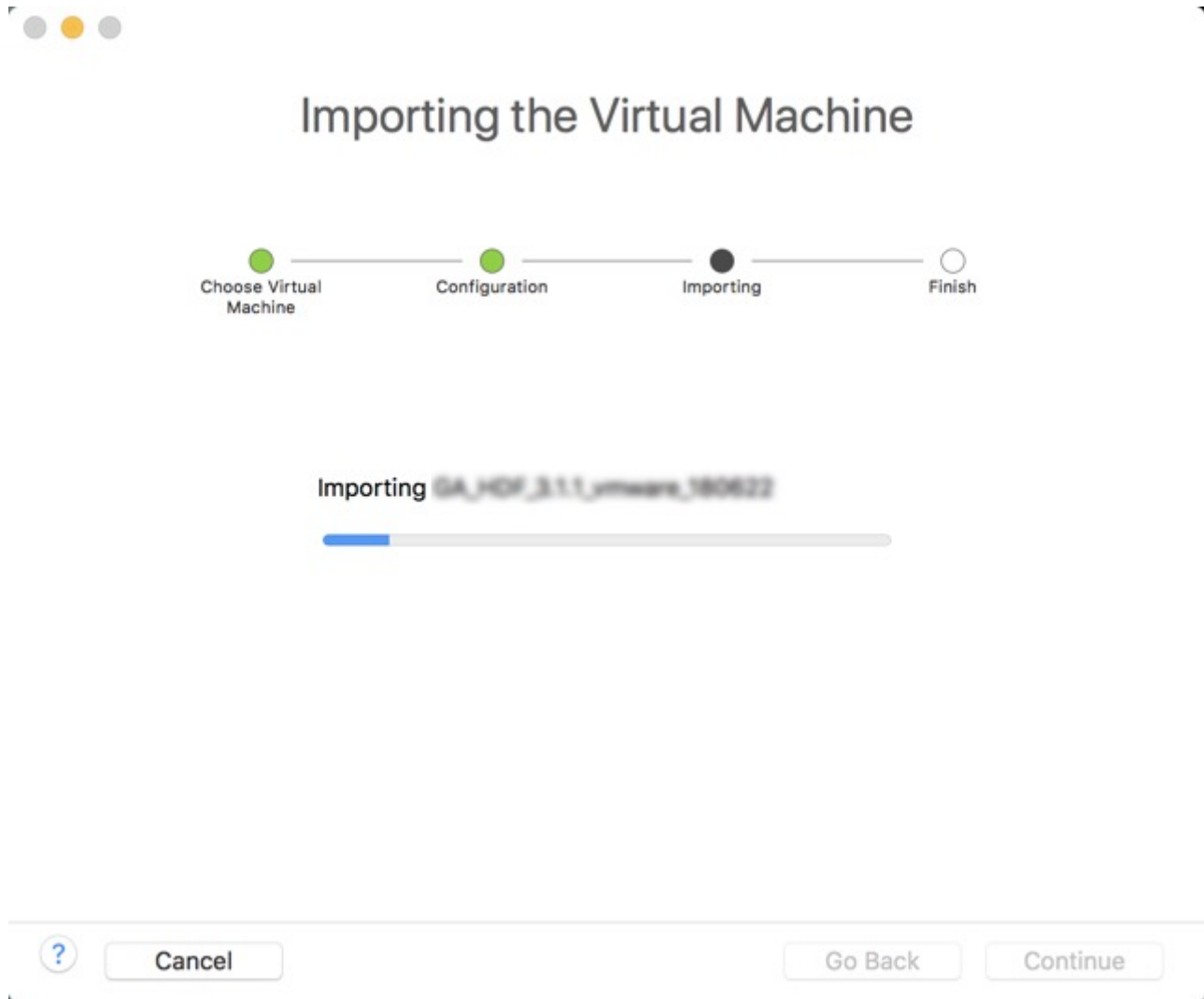
Once finished, the following screen is displayed:
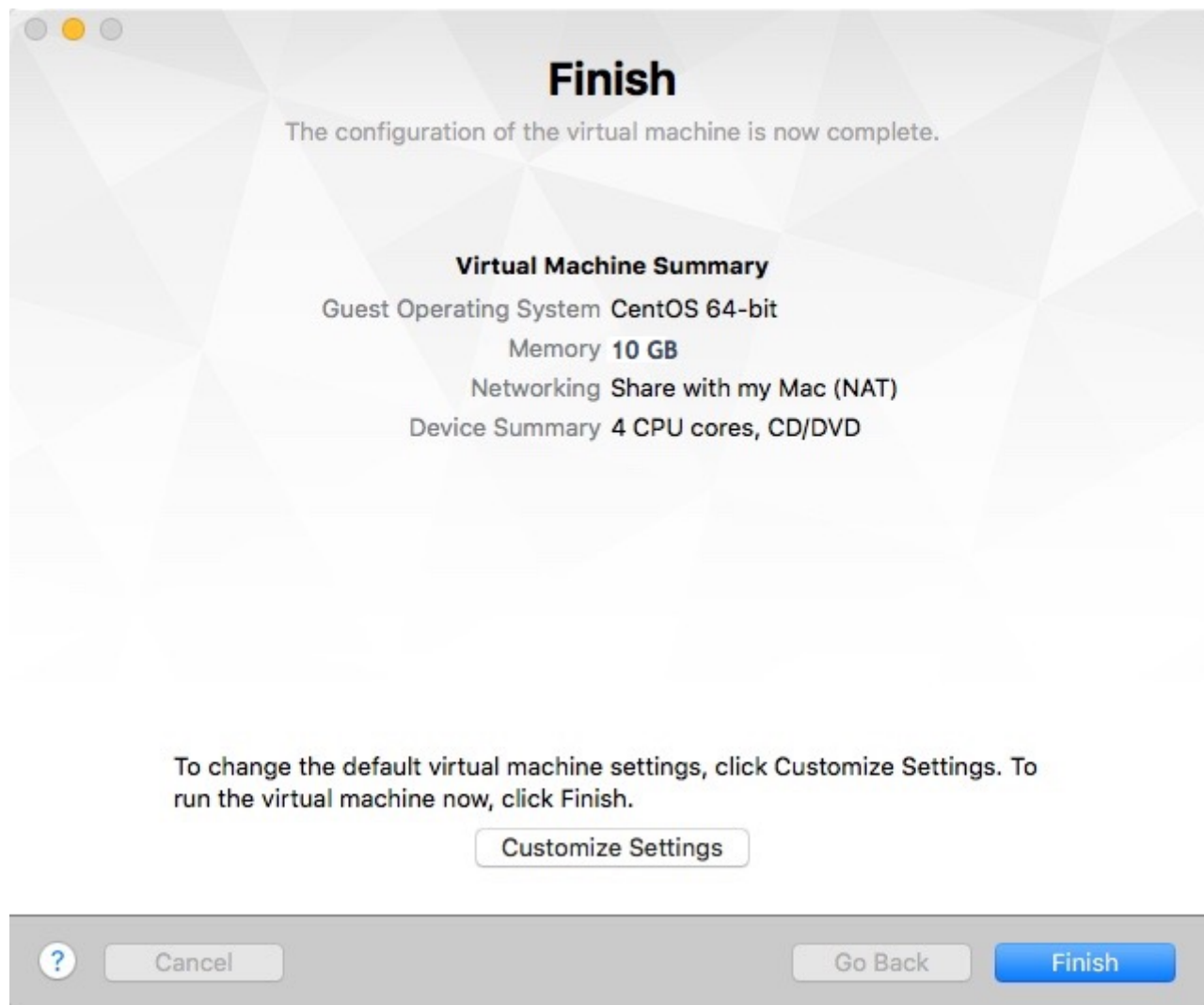
Click **Finish** to start your new virtual machine.

When the virtual machine fully boots up, you will see a similar screen as below.

Enter the URL into your browser - you're ready to begin using the sandbox.

```
Hortonworks Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/getting-started-with-hdf-sandbox




                                         8 / 12




To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
    Welcome screen:  http://localhost:1080
    SSH:  http://localhost:4200

For VMware:
    Welcome screen:  http://192.168.163.141:1080
    SSH:  http://192.168.163.141:4200
```
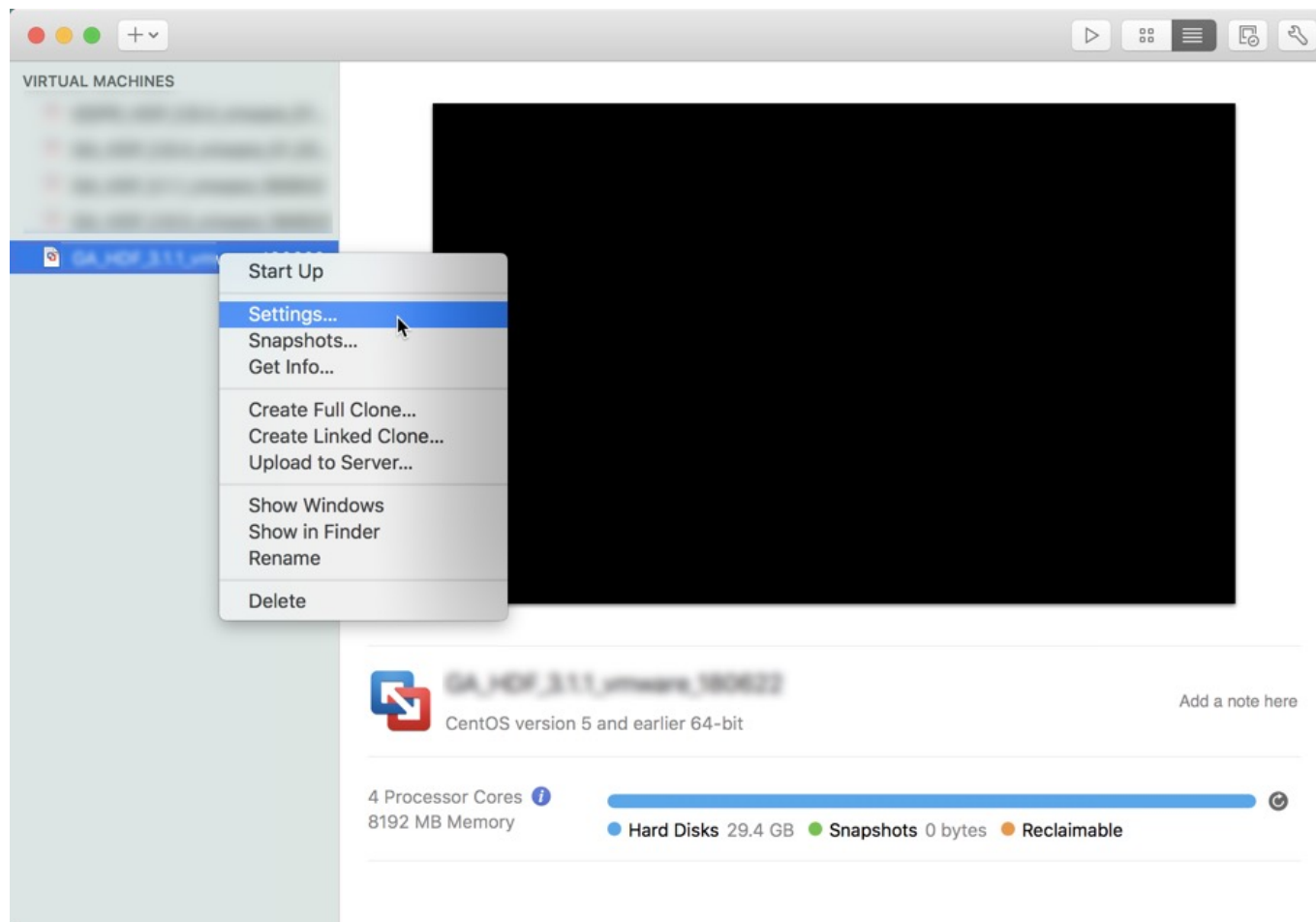
Welcome to the Hortonworks Sandbox!

Update Virtual Machine Memory

Make sure the virtual machine is shutdown before modifying configurations. **Right-click** on virtual machine and select **Settings....**
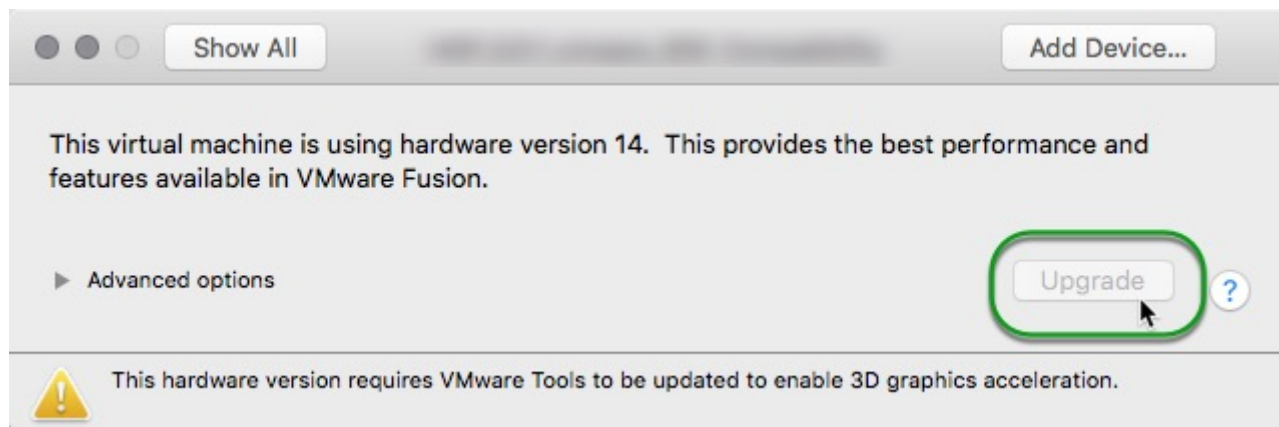
First we must upgrade the virtual image to be able to upgrade the memory click on compatibility:

Next, click upgrade:

next return to "**Show All**" and select **Processors & Memory** and modify memory accordingly.

Finally restart the virtual machine for changes to take place.

Enable Connected Data Architecture (CDA) - Advanced Topic

Prerequisite:

- A computer with minimum **22 GB of RAM** dedicated to the virtual machine
- Have already deployed the latest HDP/HDF sandbox
- Update virtual machine settings to minimum 22 GB (22528 MB)

Hortonworks Connected Data Architecture (CDA) allows you to play with both data-in-motion (HDF) and data-at-rest (HDP) sandboxes simultaneously.

HDF (Data-In-Motion)

Data-In-Motion is the idea where data is being ingested from all sorts of different devices into a flow or stream. While the data is moving throughout this flow, components or as NiFi calls them "processors" are performing actions on the data to modify, transform, aggregate and route it. Data-In-Motion covers a lot of the

preprocessing stage in building a Big Data Application. For instance, data preprocessing is where Data Engineers work with the raw data to format it into a better schema, so Data Scientists can focus on analyzing and visualizing the data.

HDP (Data-At-Rest)

Data-At-Rest is the idea where data is not moving and is stored in a database or robust datastore across a distributed data storage such as Hadoop Distributed File System (HDFS). Instead of sending the data to the queries, the queries are being sent to the data to find meaningful insights. At this stage data, data processing and analysis occurs in building a Big Data Application.

Run Script to Enable CDA

The sandbox comes prepackaged with the script needed to enable CDA. Assuming you have already deployed the **HDP sandbox**, you need to **SSH** into Sandbox VM using password `hadoop`:

- Issue command: `ssh root@sandbox-hdp.hortonworks.com -p 22`

> Note: if you originally deployed HDF sandbox, replace **sandbox-hdp** with **sandbox-hdf** in the ssh command above.

- Run bash script:

```
cd /sandbox/deploy-scripts/
sh enable-vm-cda.sh
```

The script output will be similar to:

```
[root@sandbox-host deploy-scripts]# sh enable-vm-cda.sh
hortonworks/sandbox-hdp-security    3.0              89babf832754        38 hours ago        26.4GB
3.2.0: Pulling from hortonworks/sandbox-hdf
70799bbf2226: Pull complete
40963917cdad: Pull complete
77f413c29eea: Pull complete
9332c0f89423: Pull complete
500bc38baa9a: Pull complete
6433f89da7a0: Pull complete
23b72ae56aed: Pull complete
e86bb7634bcd: Pull complete
500ba4c91e7b: Pull complete
0d20dc33e062: Pull complete
7a27fed6ccce: Pull complete
61192d7d6331: Pull complete
9178cb5038a9: Pull complete
5b5184f42eb7: Pull complete
3a63ba86263f: Pull complete
74b1de21d641: Pull complete
b5224d21e56d: Pull complete
57fe93ab5b76: Pull complete
db5299d5d378: Pull complete
0f00be75ed7d: Pull complete
a54074912ec5: Pull complete
b4943a74f8d0: Pull complete
00f7b038dd8c: Pull complete
e8e8749ccf00: Pull complete
74fec3d0de95: Pull complete
80dec10620a3: Pull complete
6d73e6bf65bd: Pull complete
656b9cd14994: Pull complete
de9078843fae: Pull complete
aef67191f76b: Pull complete
576faf28d80c: Pull complete
80b1934c252f: Pull complete
5ae0f9e4c14f: Pull complete
Digest: sha256:6a97d7479208d91db2756197cf00153f5014842710b013ccb224ee14caa1b7ce
Status: Downloaded newer image for hortonworks/sandbox-hdf:3.2.0
7eee0504de1db871aa39adb4ef435b67233f99e6a9bf81a58d37f4202c1cbd13
sandbox-proxy
bb04c4b3ff741671ebc1a930f7487d7ebf2b34e1b8b66b39775b03e89d90ac90
You need to add  'sandbox-hdp.hortonworks.com' and 'sandbox-hdf.hortonworks.com' into your hosts file
[root@sandbox-host deploy-scripts]#
```

Further Reading

- Video: How to deploy sandbox using VMWare
- Sandbox Architecture
- Follow-up with the tutorial: Learning the Ropes of the HDP Sandbox
- Browse available tutorials