

MODUL PRAKTEK 1 – TEKNOLOGI BIG DATA

1.1. Deskripsi Singkat

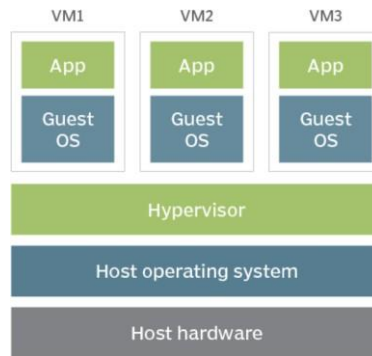
Apache Hadoop merupakan salah satu Big Data Framework atau platform *open source* berbasis Java yang berada di bawah lisensi Apache. Hadoop memungkinkan *data engineer* dan *data scientist* untuk melakukan pemrosesan data dengan cara yang dapat diskalakan (*scalable*), efisien, dan hemat biaya, serta digunakan secara luas di berbagai industri. Dalam praktikum mata kuliah Teknologi Perekayasaan Data, kita akan menggunakan Hadoop untuk menyimpan dan memproses Big Data dalam environment komputasi terdistribusi. Big Data dalam hal ini merupakan istilah yang digunakan untuk mendeskripsikan kumpulan data yang sangat besar (*volume*) dan kompleks (*variety*) dengan laju penambahan data yang cepat (*velocity*). Data akan diolah secara terdistribusi dalam waktu yang bersamaan, sehingga kecepatan prosesnya bisa lebih singkat.

Ekosistem Hadoop meliputi berbagai tools dan aplikasi yang membantu dalam pengumpulan, penyimpanan, analisis, dan pengelolaan Big Data. Tools dan aplikasi ini akan kita gunakan pada praktikum – praktikum selanjutnya diantaranya adalah Ambari, Sqoop, Flume, Hive, Hbase, dan lainnya. Meskipun Hadoop memungkinkan untuk dijalankan pada *single-node/machine cluster* untuk tujuan pengujian, tetapi untuk implementasi riil seharusnya berjalan pada *multi-node/machine clusters*. Untuk praktikum ini, kita akan menggunakan teknik virtualisasi agar dapat mengoperasikan Hadoop pada *single-node* dan menyederhanakan proses instalasi yang rumit dengan menggunakan Sandbox Image yang tersedia. Sandbox merupakan *virtual machine* yang digunakan untuk menjalankan perangkat lunak pada testing environment.

1.2. Tujuan Praktikum

Setelah praktikum pada modul 1 ini diharapkan mahasiswa mempunyai kompetensi dapat melakukan instalasi dan menjalankan Hadoop dengan virtualisasi. Secara sederhana dapat digambarkan sebagai berikut:

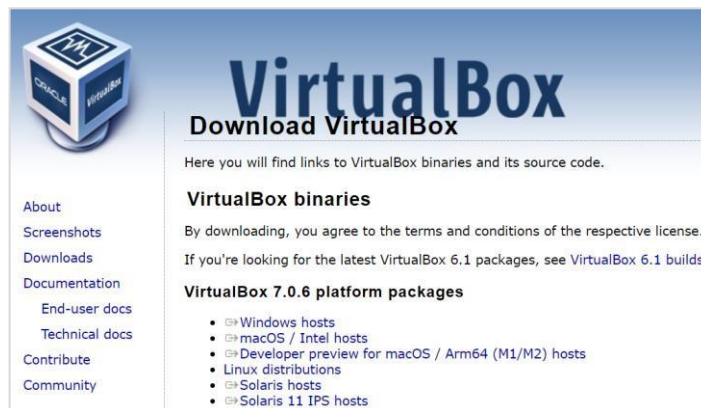
Host machine merupakan PC/laptop yang kita gunakan untuk mengerjakan praktikum dengan sistem operasi misal Windows/macOS. Hadoop akan berjalan di atas *guest virtual machine* dengan sistem operasi adalah Red Hat.



1.3. Material Praktikum

Pada kegiatan modul 1 diperlukan beberapa material berupa file, yaitu:

- 1) Internet Browser (direkomendasikan Google Chrome)
- 2) File installer VirtualBox dapat diunduh pada <https://www.virtualbox.org/wiki/Downloads> (tidak diperlukan jika VirtualBox telah terinstall pada PC/Laptop praktikum)



- 3) HDP Sandbox Image (.ova 20 GB download size) yang dapat diunduh pada <https://www.cloudera.com/downloads/hortonworks-sandbox/hdp.html> dan pilih Installation Type VirtualBox. File juga dapat diunduh langsung pada https://archive.cloudera.com/hwx-sandbox/hdp/hdp-3.0.1/HDP_3.0.1_virtualbox_181205.ova



Untuk menjalankan Hadoop pada VirtualBox diperlukan PC/Laptop dengan spesifikasi minimum sebagai berikut:

Component	Requirement
Hard Disk	Minimum of 50 GB of available hard-disk space on Drive C:
Memory	Recommended: 16 GB
Processor Type	x64 Processor (x86 is not supported)
OS	Windows 10 or Windows 8.1

1.4. Kegiatan Praktikum

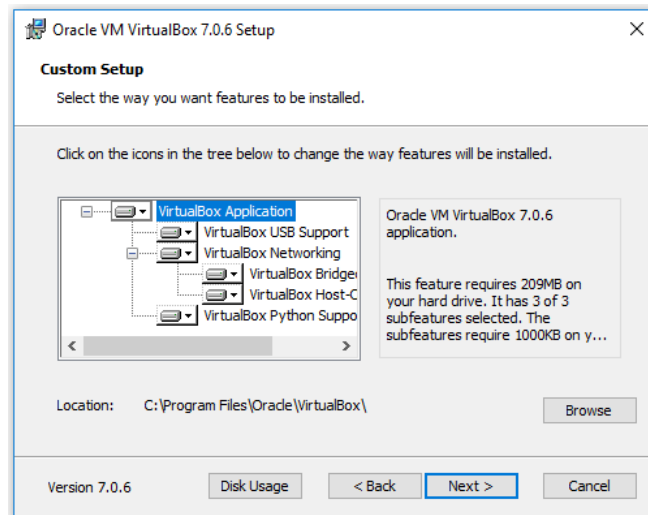
A. Instalasi VirtualBox

Tahapan berikut tidak perlu dilakukan jika VirtualBox telah terinstall pada PC/Laptop praktikum. Jika belum, maka lakukan langkah-langkah sebagai berikut:

1. Klik kanan pada file installer VirtualBox yang telah diunduh, Run as Administrator. Maka akan tampil jendela awal proses instalasi.
2. Selanjutnya, klik tombol “Next” dan lanjutkan instalasi dengan mengikuti langkah-langkah berikut ini.



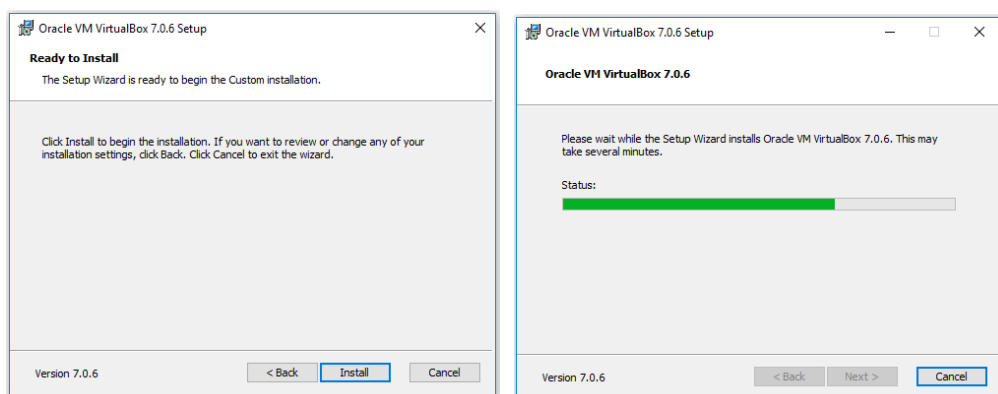
3. Selanjutnya, klik tombol “Next”



4. Pilih “Yes” ketika diminta apakah melanjutkan proses instalasi dimana proses tersebut akan mereset koneksi network sementara ketika menginstall VirtualBox Networking feature.



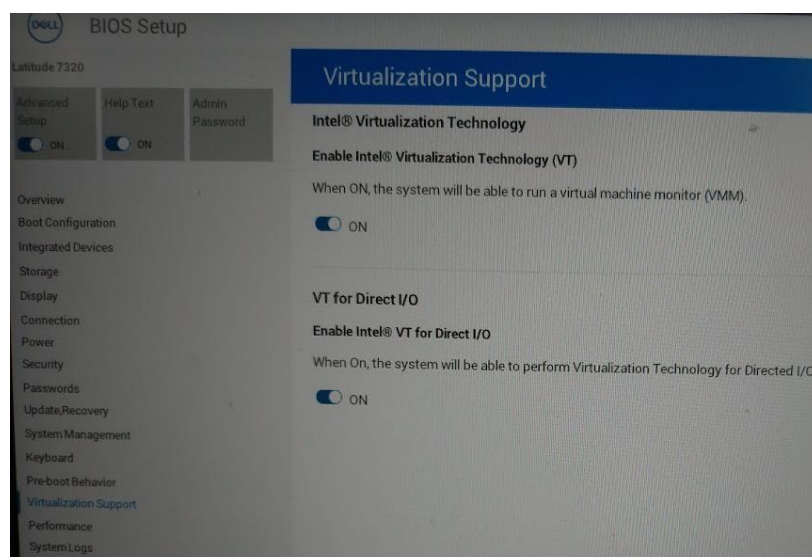
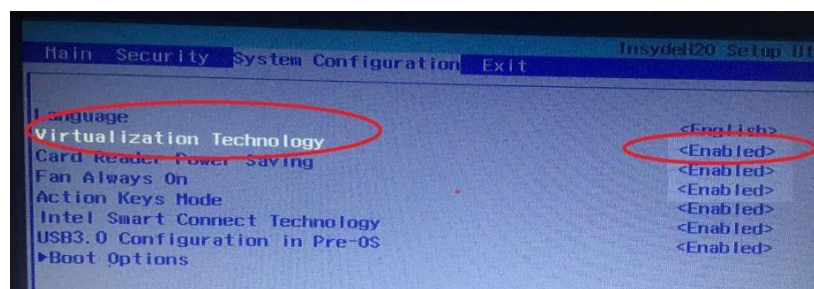
5. Kemudian klik “Install”, dan proses instalasi akan berlangsung beberapa saat.



6. Klik “Finish” untuk mengakhiri proses instalasi dan jalankan VirtualBox



7. Sebelum menjalankan VirtualBox, pastikan bahwa **Virtualization Technology** pada **BIOS host telah diatur sebagai Enabled**. Hal ini dikarenakan sistem operasi host harus dapat mendukung virtualisasi perangkat keras. Jika belum, maka silakan masuk ke Windows Machine Bios setting dan *enable* virtualization.



B. Import dan Start HDP Sandbox Image

Tahapan berikut ini dilakukan untuk meng-*import* HDP Sandbox Image yang telah diunduh.

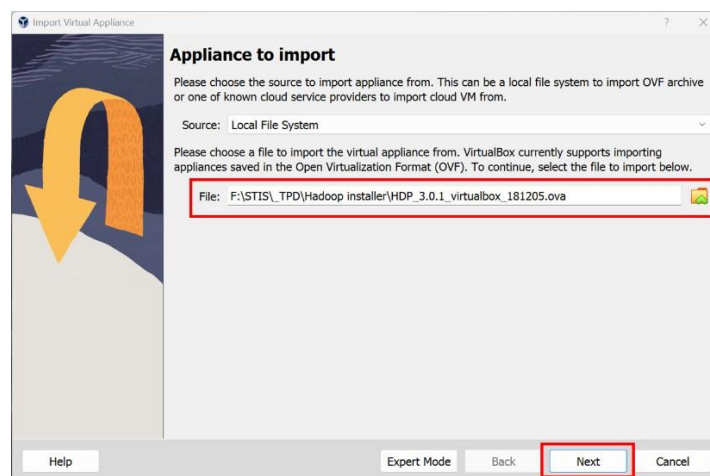
1. Buka dan jalankan VirtualBox dan kita akan mendapatkan tampilan seperti gambar di bawah ini



2. Klik pada menu “File” – “Import Appliance...”

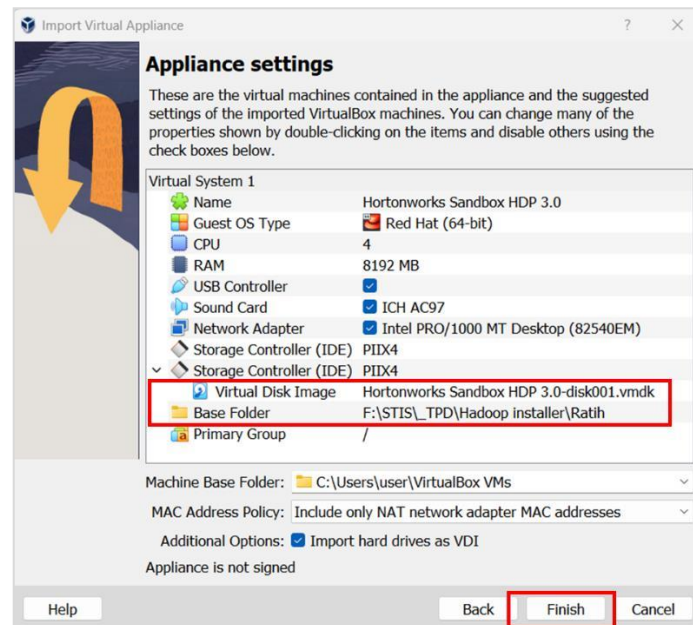


3. Pilih file HDP Sandbox Image yang telah diunduh untuk di-*import* dan klik “Next”

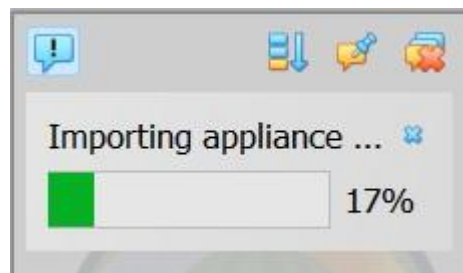


4. Pada Appliance Setting, ubah setting VirtualBox machine menjadi:
- Guest OS Type = Red Hat
 - CPU = 4
 - RAM = 8192 MB
 - Base Folder = **buat dan pilih folder baru dengan nim kalian** (agar tidak tertukar dengan praktikum mahasiswa lain yang menggunakan PC yang sama)

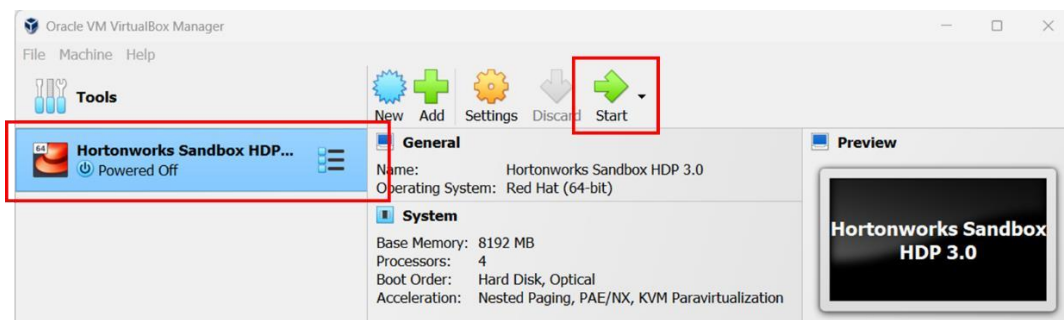
Kemudian klik “Finish”



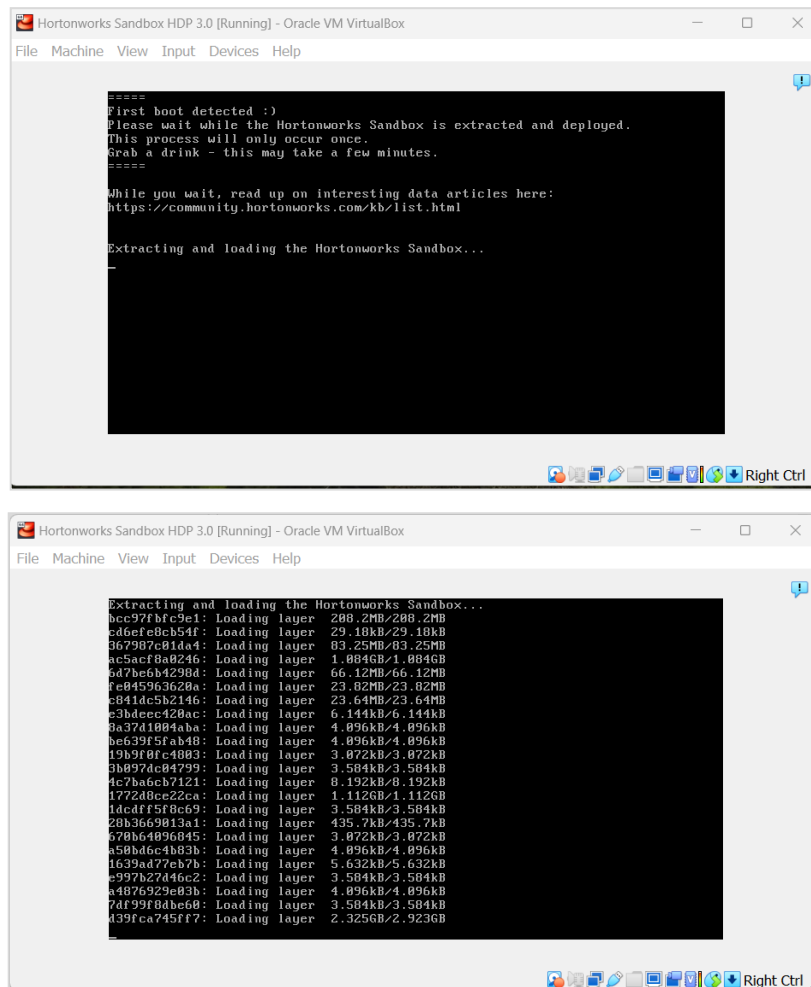
5. Tunggu proses import sampai selesai (kurang lebih membutuhkan waktu 10 menit)



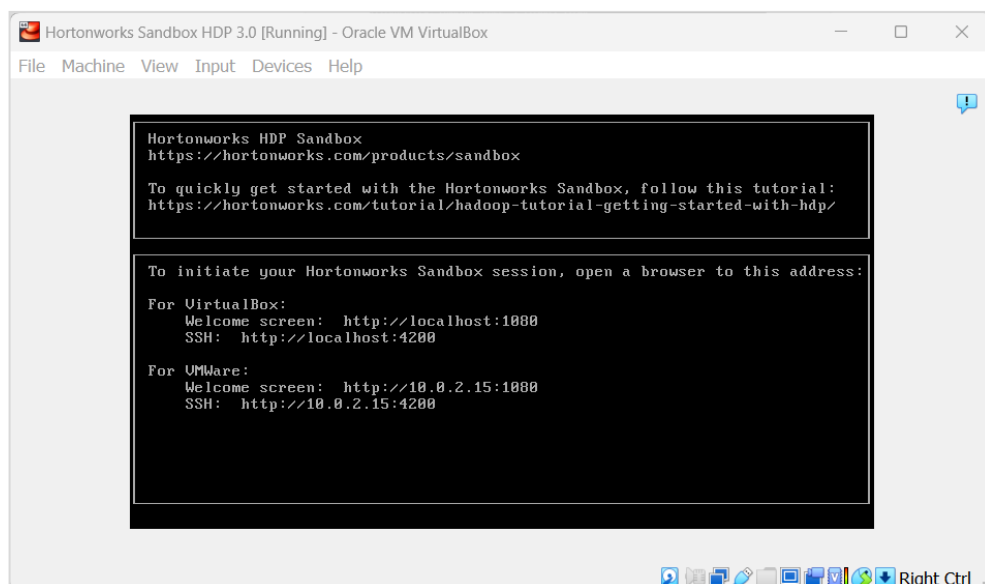
6. Setelah selesai import, pilih VirtualBox yang telah di-import, dan klik “Start”



7. Tunggu proses menjalankan VirtualBox

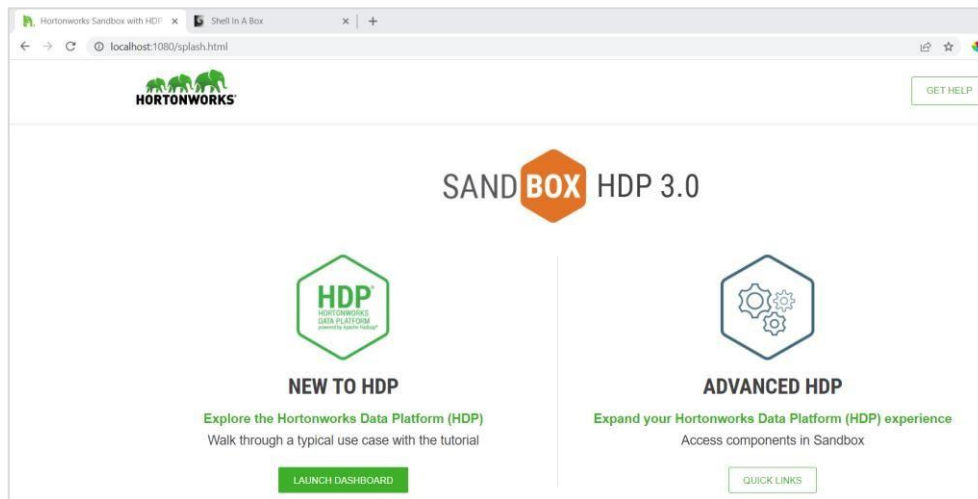


8. HDP Sandbox siap untuk digunakan ketika muncul di layar seperti gambar di bawah. Pada gambar diinformasikan bahwa HDP Sandbox dapat diakses menggunakan browser pada alamat <http://localhost:1080> dan untuk mengakses terminal/SSH dari HDP Sandbox dapat diakses menggunakan browser pada alamat <http://localhost:4200>



C. Akses HDP Sandbox – Ambari dan Terminal/SSH

Ambari merupakan salah satu komponen Hadoop yang dapat digunakan untuk instalasi, monitoring dan manajemen cluster hadoop. Setelah langkah-langkah pada poin B selesai dan sukses dilakukan, maka kita dapat mengakses Ambari pada alamat <http://localhost:1080> dan kita akan mendapatkan tampilan sebagai berikut.



Sebelum kita mengoperasikan Ambari, kita lakukan Reset Password terlebih dahulu. Caranya dengan akses pada terminal SSH pada alamat <http://localhost:4200>. Masukan username dan password awal (default) yaitu:

- Username: **root**
- Existing Password: **hadoop**

Kita akan reset password menjadi: **Tp4stis** (untuk praktikum, lebih baik disamakan agar tidak lupa di praktikum-praktikum selanjutnya)

```
root@sandbox-hdp:~ - Shell In A Box
localhost:4200
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password: (ketik hadoop dan klik Enter)
You are required to change your password immediately (root enforced)
Last login: Sat Feb  4 03:25:50 2023 from 172.18.0.2
Changing password for root.
(current) UNIX password: (ketik hadoop dan klik Enter)
New password: (ketik password baru misal Tp4stis dan klik Enter)
```

Jika sukses, akan muncul seperti di bawah ini.

```
[root@sandbox-hdp ~] #
```

Kita akan atur user untuk akses Ambari dengan cara:

```
[root@sandbox-hdp ~]# ambari-admin-password-reset
```

Atur password untuk user **admin** yaitu **admin**, re-type password yaitu **admin**. Tunggu hingga selesai dan sukses seperti gambar berikut.

```
[root@sandbox-hdp ~]# ambari-admin-password-reset
Please set the password for admin: (ketik admin dan klik Enter)
Please retype the password for admin: (ketik admin dan klik Enter)

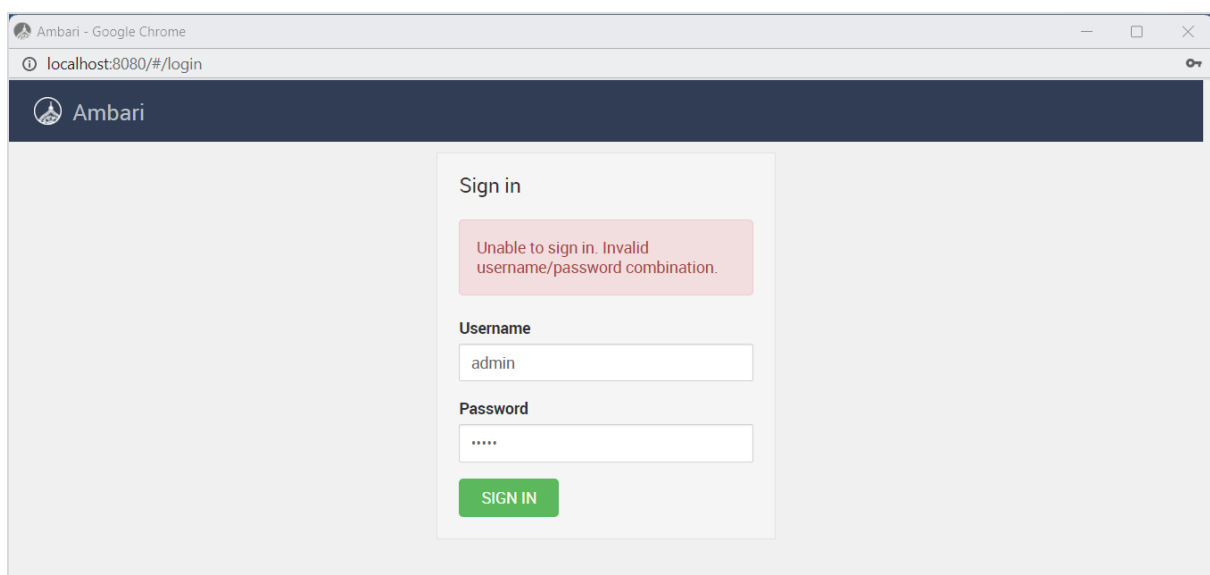
The admin password has been set.
Restarting ambari-server to make the password change effective...

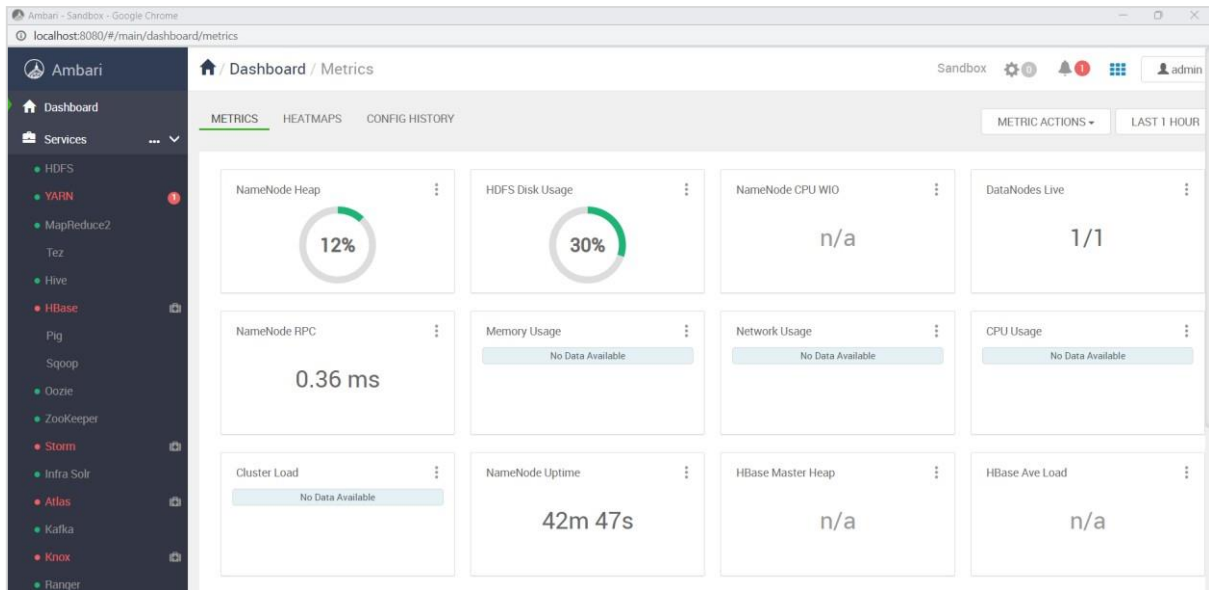
Using python /usr/bin/python
Restarting ambari-server
Waiting for server stop...
Ambari Server stopped
Ambari Server running with administrator privileges.
Organizing resource files at /var/lib/ambari-server/resources...
Ambari database consistency check started...
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start.....
Server started listening on 8080

DB configs consistency check: no errors and warnings were found.
[root@sandbox-hdp ~]#
```

Kemudian akses kembali Ambari, dengan klik tombol “LAUNCH DASHBOARD” di alamat <http://localhost:1080>. Masukkan Username dan Password yang telah kita ubah yaitu:

- Username: **admin**
- Password: **admin**





HDP Sandbox berhasil terinstall dan dapat diakses untuk digunakan pada praktikum-praktikum selanjutnya.

1.5. Penugasan

Kerjakan sesuai dengan yang dijelaskan pada bagian Kegiatan Praktikum. Hasil pekerjaan praktikum berupa screenshot dari:

1. Reset password untuk user Sandbox HDP **root** telah sukses dilakukan
2. Set password untuk user Ambari **admin** telah sukses dilakukan
3. Dashboard Ambari dapat diakses dengan username dan password yang telah dibuat

Setelah tahapan kegiatan praktikum selesai dilakukan serta telah mengerjakan penugasan. Silakan matikan/power-off VirtualBox sandbox HDP dengan klik kanan pada VirtualBox – Stop – Power Off.

