

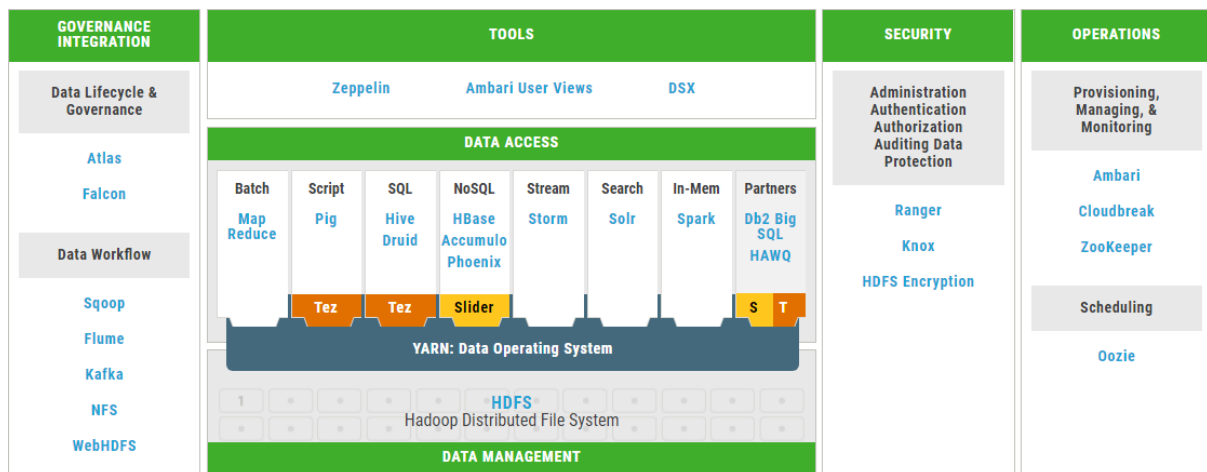
MODUL PRAKTEK 2 – TEKNOLOGI BIG DATA

1.1. Deskripsi Singkat

Seperti yang telah dijelaskan pada modul I, Apache Hadoop merupakan *framework* yang dapat digunakan oleh untuk menyimpan dan memproses data secara terdistribusi. *Deployment* Hadoop dapat dilakukan pada server *bare-metal*, *virtualization machine*, atau *container*. Setelah berhasil melakukan instalasi Hadoop Distribution Platform (HDP) pada praktikum I, konfigurasi perlu dilakukan untuk memastikan setiap *service* yang dibutuhkan saat penggunaan Hadoop berjalan dengan baik.

1.2. Tujuan Praktikum

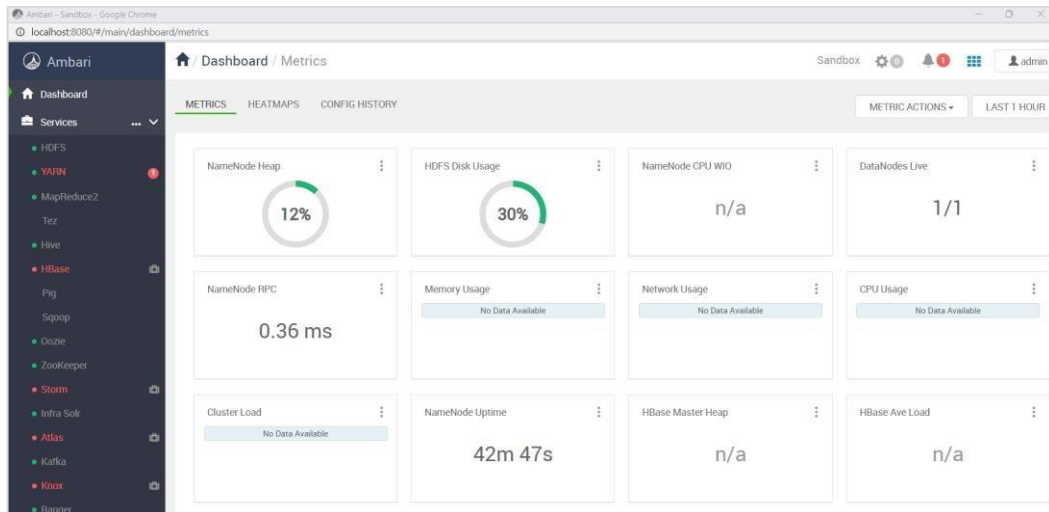
Setelah praktikum pada modul 2 ini, diharapkan mahasiswa mempunyai kompetensi dalam melakukan konfigurasi Hadoop dan mengetahui komponen yang terdapat pada HDP sebagaimana yang terdapat pada Gambar berikut.



1.3. Material Praktikum

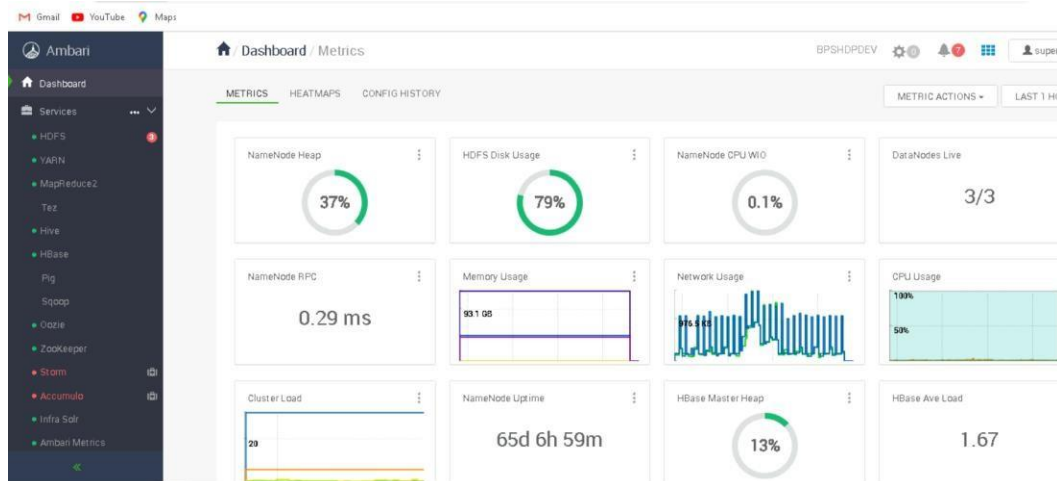
Persyaratan yang dibutuhkan untuk melakukan praktikum II yaitu :

1. HDP yang telah terinstal pada virtualbox.
2. Ambari dapat diakses



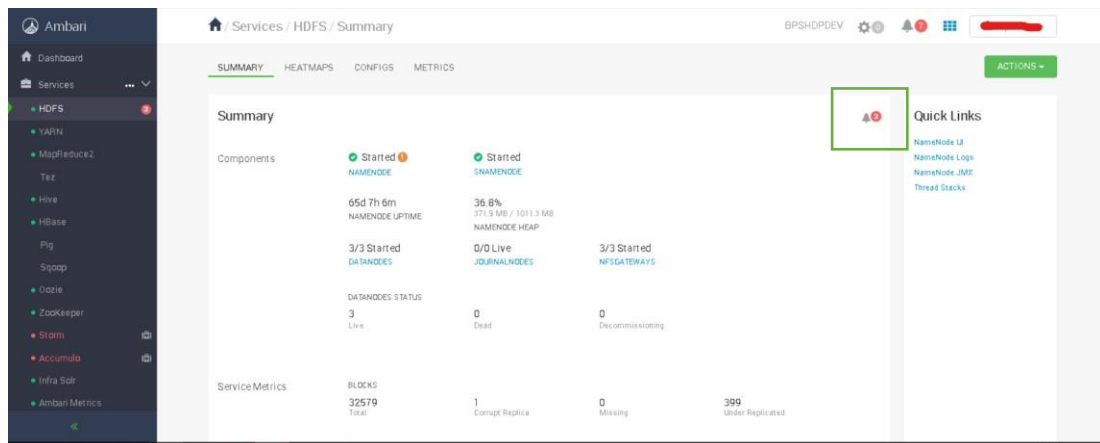
1.4. Kegiatan Praktikum

1. Akses Ambari dengan menggunakan akun yang telah dibuat sebelumnya.

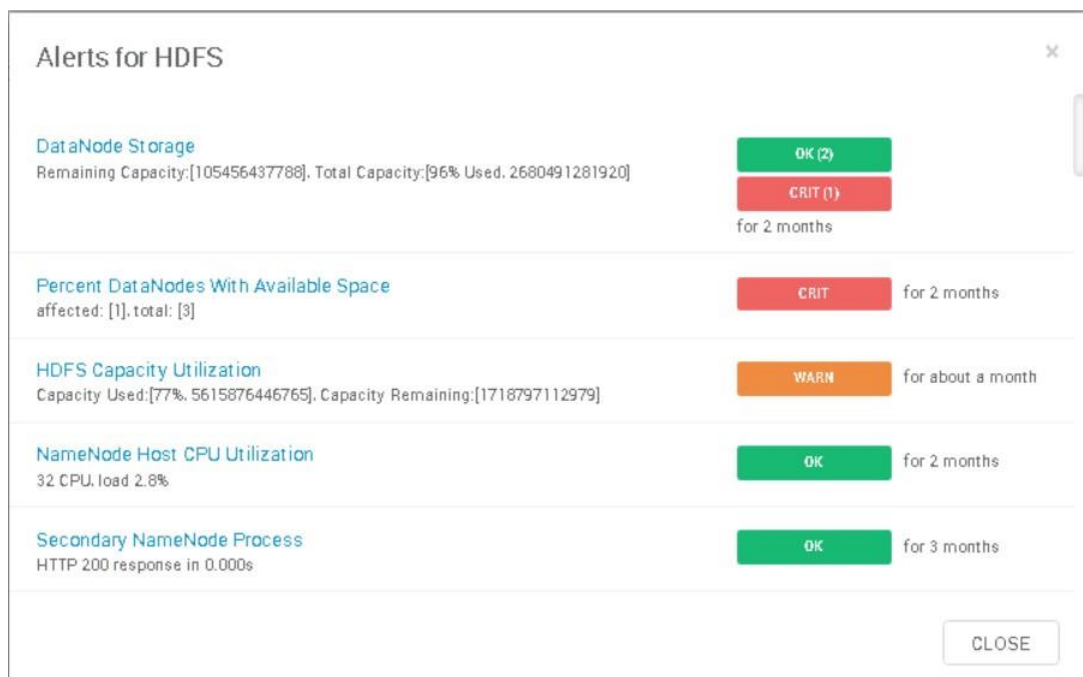


2. Identifikasi *services* pada HDP yang masih dalam status error/warning.

Beberapa *service* yang penting untuk dipastikan **on** adalah HDFS, Yarn, MapReduce2, Hive, Oozie, Zookeeper. Setiap *service* dapat dicek dengan memilih salah satu *service* di panel sebelah kanan, misal HDFS. Setelah meng-klik HDFS, maka akan muncul tampilan sebagai berikut.



Klik icon alert untuk mengetahui warning atau error pada service.



Terlihat pada gambar, bahwa permasalahan utama pada HDFS adalah kapasitas storage yang dapat disebabkan oleh *log* dari *service* dan juga data yang tersimpan di Hadoop.

3. Melakukan konfigurasi dan *troubleshoot* pada komponen Hadoop.
 - a) Stop and start ambari



```

root@sandbox-hdp.hortonworks.com's password:
Last login: Thu Jul 1 01:55:39 2021 from 172.18.0.2
[root@sandbox-hdp ~]# ambari-server status
Using python /usr/bin/python
Ambari-server status
Ambari Server running
Found Ambari Server PID: 367 at: /var/run/ambari-server/ambari-server.pid
[root@sandbox-hdp ~]# systemctl status nginx
Unit nginx.service could not be found.
[root@sandbox-hdp ~]# ambari-server stop
Using python /usr/bin/python
Stopping ambari-server
Waiting for server stop...
Ambari Server stopped
[root@sandbox-hdp ~]# ambari-server start
Using python /usr/bin/python
Starting ambari-server
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.
Ambari Server 'start' completed successfully.
[root@sandbox-hdp ~]#

```

b) Monitoring Hosts

- 1) Klik Hosts pada panel sebelah kanan, maka akan muncul tampilan sebagai berikut.

Hosts

<input type="checkbox"/>	Name		IP Address	Rack	Cores	RAM	Disk Usage	Load Avg	Versions	Components
<input checked="" type="checkbox"/>	ambaridev.bps.go.id		 10.0.45.111	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.92	HDP-3.1.5.0	44 Components
<input checked="" type="checkbox"/>	hdpdev02.bps.go.id		 10.0.45.112	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.30	HDP-3.1.5.0	32 Components
<input checked="" type="checkbox"/>	hdpdev03.bps.go.id		 10.0.45.113	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.38	HDP-3.1.5.0	32 Components
<input checked="" type="checkbox"/>	hdpdev04.bps.go.id		 10.0.45.114	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.10	HDP-3.1.5.0	26 Components
<input checked="" type="checkbox"/>	hdpdev05.bps.go.id		 10.0.45.115	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.29	HDP-3.1.5.0	26 Components
<input checked="" type="checkbox"/>	hdpdev06.bps.go.id		 10.0.45.116	/default-rack	32 (32)	125.75GB	<div><div></div></div>	0.25	HDP-3.1.5.0	26 Components

Items per page: 101 - 6 of 6

Pada layar akan tampil informasi terkait hosts yang digunakan saat deployment Hadoop. Jika terdapat 6 host, maka akan muncul list seperti pada gambar di atas.

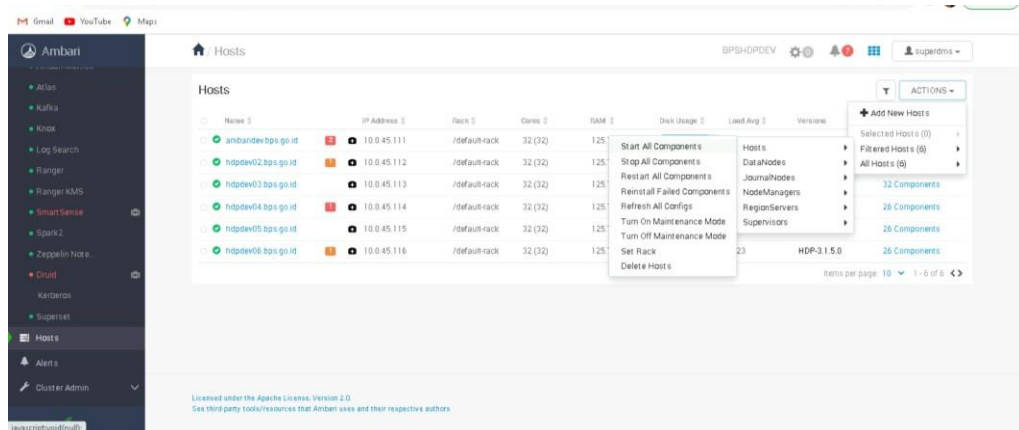
Ketika terdapat warning pada salah satu host, maka klik salah satu host untuk melihat detail dari warning.

Service	Alert Definition Name	Status	Response
Accumulo	Accumulo Tracer Process	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Accumulo	Accumulo Monitor Process	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Accumulo	Accumulo GC Process	CRT for 5 days	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Accumulo	Accumulo Master Process	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Ambari	Host Disk Usage	CRT for 13 hours	Capacity Used: [80.13%, 429.1 GB], Capacity Total: [535.5 GB], path=/u...
Druid	Druid Overlord Web UI	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Druid	Druid Broker Process	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Druid	Druid Router Process	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...
Druid	Druid Coordinator Web UI	CRT for 3 months	Connection failed: [Errno 111] Connection refused to ambaridev.bps.go...

Jika masalahnya adalah storage, maka perlu dilakukan pengarsipan terhadap data dan log yang ada.

c) Restart service

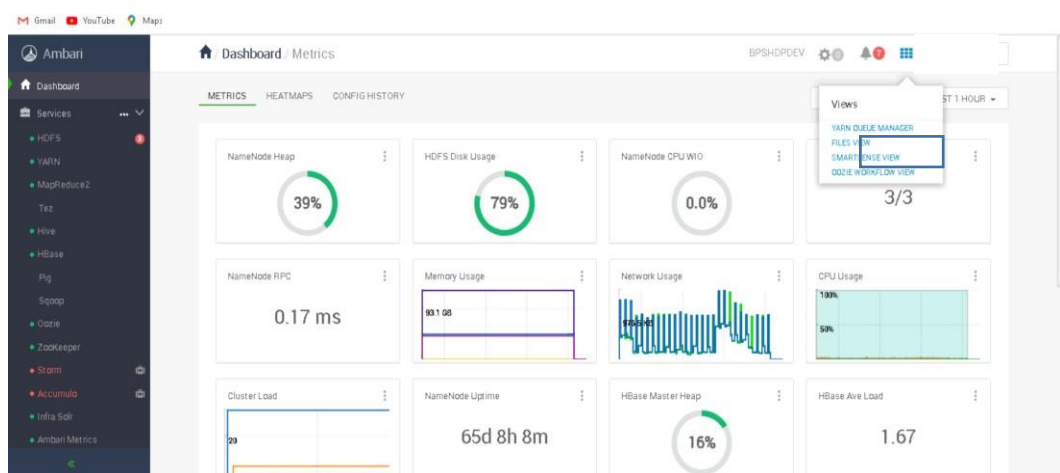
Selain itu, jika terdapat alert pada salah satu atau beberapa *services*, misal status *service stop*, maka restart dapat dilakukan pada *services* terkait atau semua *services* secara sekaligus. Caranya yaitu dengan melakukan klik pada fitur **Actions** dan pilih Host → start all components.

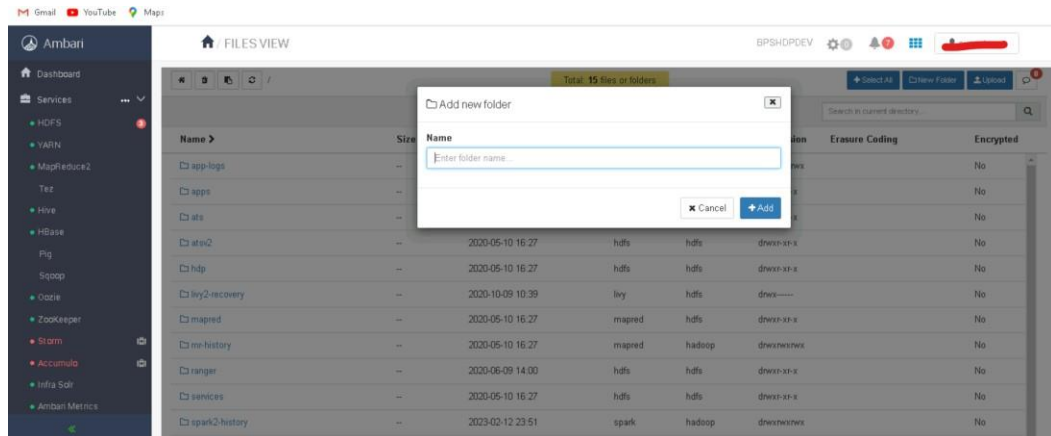


4. Menggunakan fitur Hadoop.

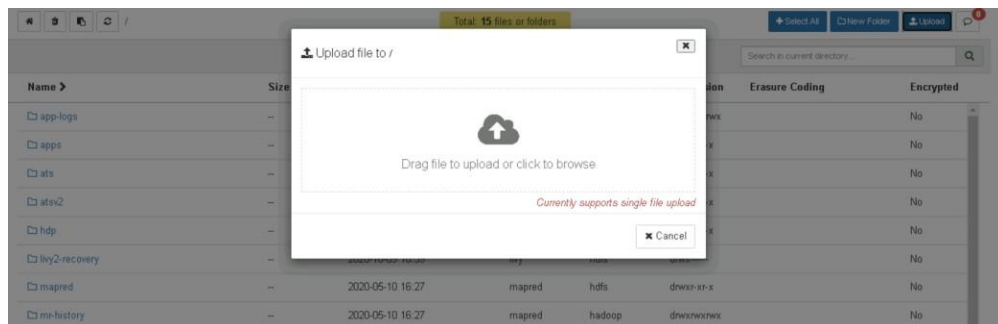
a) Load data ke hadoop (Menggunakan File View)

- 1) Download data pada link : [Geolocation.zip](#)
- 2) Buka fitur Files View pada Ambari





- Tulis nama folder → klik add
- Klik upload



- Pilih file yang hasil download yang telah di-unzip terlebih dahulu. Setelah itu, maka file akan muncul pada folder yang telah dibuat sebelumnya.
- b) Load data ke hadoop (Menggunakan Command)
- 1) Akses <http://localhost:4200> pada browser.
 - 2) Lakukan login terlebih dahulu.
Username: **root**
Existing Password: **Tp4stis**
 - 3) Jalankan command HDFS untuk melihat list file pada folder yang telah dibuat:
`hdfs dfs -ls [path]`
 - 4) Hapus directory
`hdfs dfs -rm [path]`
 - 5) Buat direktori baru
`hdfs dfs -mkdir [path]`
 - 6) Copy file ke hdfs
 - a) Instal MobaXterm portable
 - b) Akses host dari HDP sandbox dengan klik session → SSH
 - c) Masukkan IP host: 127.0.0.1 dengan port 2222
 - d) Masukkan username dan password

- e) Upload file csv ke path tertentu pada local system
- f) Jalankan command berikut:
`hdfs dfs -put [source path] [hdfs destination path]`
- 7) Cek file yang telah di-copy ke hdfs pada fitur files view

1.5. Penugasan

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

1. Hasil troubleshoot yang dilakukan pada services di HDP, terutama HDFS, Yarn, MapReduce2, Hive, Oozie, Zookeeper, yang mencakup:
 - Warning yang muncul
 - Mekanisme troubleshoot yang dilakukan
 - Hasil akhir
2. Melakukan load data ke hadoop dengan dua mekanisme yang telah dijelaskan sebelumnya.
3. Mencari command hdfs lainnya (maksimal 5) dan mempraktekkan command tersebut. Hasil dari tugas ini adalah daftar command yang digunakan, fungsi command, dan hasil eksekusi command hdfs.