

#2 Laporan Practice Big Data – *Perintah-Perintah Hadoop*

Raffi Alwinsyah (2021903430017)

Kelompok 5 (TRKJ 3.B)

Prodi Teknologi Rekayasa Komputer Jaringan

Jurusan Teknologi Informasi dan Komputer

Politeknik Negeri Lhokseumawe

Ringkasan

Apache Hadoop adalah kerangka kerja sumber terbuka yang digunakan untuk pemrosesan terdistribusi dan analisis himpunan data besar pada kluster. Hadoop menyediakan kemampuan penyimpanan data besar dan menjalankan aplikasi pada kluster perangkat keras komoditas. Kerangka kerja ini juga mencakup modul Apache Hadoop MapReduce, yang memungkinkan penulisan pekerjaan untuk memproses sejumlah besar data dengan membagi data input menjadi gugus-gugus independen yang diproses secara paralel di seluruh simpul dalam kluster. Selain itu, ekosistem Hadoop mencakup perangkat lunak dan utilitas terkait seperti Apache Hive, Apache HBase, Spark, Kafka, dan banyak lagi. Dengan kemampuannya untuk menyimpan dan memproses dataset besar secara efisien, Hadoop menjadi penting karena dapat mengatasi peningkatan volume dan variasi data, terutama dari media sosial dan internet.

Daftar Isi

1	Pendahuluan	1
1.1	Latar Belakang	1
1.2	Tujuan	2
1.3	Tinjauan Pustaka	2
2	Alat dan Bahan	2
2.1	Alat	2
2.2	Bahan	3
3	Prosedur Kerja	3
4	Hasil dan Pembahasan	4
5	Kesimpulan	6

1 Pendahuluan

Hadoop adalah sebuah framework perangkat lunak open source yang digunakan untuk menyimpan dan memproses data secara terdistribusi. Framework ini dirancang untuk mengatasi masalah pemrosesan data yang sangat besar di lingkungan yang terdiri dari sejumlah besar komputer. Hadoop didasarkan

pada model pemrograman MapReduce, yang dikembangkan oleh Google. Hadoop adalah sebuah framework perangkat lunak open source yang digunakan untuk menyimpan dan memproses data secara terdistribusi. Framework ini dirancang untuk mengatasi masalah pemrosesan data yang sangat besar di lingkungan yang terdiri dari sejumlah besar komputer. Hadoop didasarkan pada model pemrograman MapReduce, yang dikembangkan oleh Google.

1.1 Latar Belakang

Dilihat dari perkembangan teknologi informasi yang cepat, yang memberikan konsekuensi pertumbuhan data yang besar, dikenal sebagai Big Data. Big Data merupakan media penyimpanan data yang menawarkan ruang tak terbatas dan kemampuan untuk mengakomodasi serta memproses berbagai jenis data dengan sangat cepat. Dalam konteks ini, Hadoop Distributed File System (HDFS) menjadi sebuah media penyimpanan data utama dengan kapasitas data besar yang

digunakan oleh Hadoop. Seiring dengan kebutuhan akan penyimpanan dan pemrosesan data yang semakin besar, Hadoop menjadi penting dalam menangani volume dan variasi data yang terus meningkat, terutama dari media sosial dan internet.

1.2 Tujuan

Praktikum ini bertujuan memperkenalkan langkah-langkah instalasi dan konfigurasi Apache Hadoop dalam lingkungan virtual menggunakan VirtualBox dan sistem operasi Ubuntu. Dengan tujuan utama memberikan pemahaman praktis tentang penggunaan Apache Hadoop dalam pengolahan Big Data, praktikum ini dimaksudkan untuk membekali mahasiswa dengan keterampilan instalasi dan konfigurasi yang mendalam.

1.3 Tinjauan Pustaka

Dalam praktikum ini, tinjauan pustaka diarahkan pada beberapa sumber kunci yang memberikan pemahaman mendalam kepada

mahasiswa mengenai instalasi dan konfigurasi Apache Hadoop, serta konsep dasar yang melingkupinya. Buku "Hadoop: The Definitive Guide" oleh Tom White dari O'Reilly Media menjadi sumber utama yang memberikan wawasan terperinci tentang Apache Hadoop. Buku ini tidak hanya menjelaskan konsep fundamental Hadoop, tetapi juga memberikan panduan langkah demi langkah tentang instalasi dan konfigurasi. Selain itu, dokumentasi resmi dari situs web Apache Hadoop menjadi rujukan utama yang menyediakan informasi yang sah dan terkini mengenai instalasi dan konfigurasi Hadoop. Tutorial dan dokumentasi resmi Ubuntu juga dijadikan acuan untuk mahasiswa dalam memahami langkah-langkah instalasi dan konfigurasi Ubuntu sebagai sistem operasi pada praktikum. Dengan menggabungkan sumber-sumber ini, diharapkan mahasiswa dapat memperoleh pemahaman yang holistik tentang Apache Hadoop, termasuk konfigurasi dan penerapannya dalam lingkungan virtual menggunakan VirtualBox dan Ubuntu.

2 Alat dan Bahan

Sebutkan software dan hardware yang digunakan

2.1 Alat

Dalam praktikum instalasi Apache Hadoop, alat yang digunakan utamanya adalah personal komputer atau laptop. Berikut adalah jабaran rinci tentang penggunaan alat tersebut:

Personal Komputer atau Laptop:

Fungsi Utama: Alat ini digunakan sebagai pusat pengendalian dan pelaksanaan seluruh proses praktikum. Semua langkah-langkah instalasi dan konfigurasi dilakukan pada komputer atau laptop ini. Sistem Operasi: Personal komputer atau laptop umumnya menjalankan sistem operasi Linux atau distribusi Linux tertentu seperti Ubuntu. Spesifikasi Minimum: Disarankan untuk menggunakan

komputer atau laptop dengan spesifikasi yang memadai untuk menjalankan Apache Hadoop. Hal ini termasuk kapasitas RAM yang cukup, ruang penyimpanan yang mencukupi, dan prosesor yang dapat menangani tugas-tugas yang diperlukan.

Akses Internet:

Fungsi Utama: Diperlukan koneksi internet untuk mengunduh perangkat lunak yang diperlukan selama praktikum, seperti OpenJDK, Apache Hadoop, dan paket-paket lainnya. Proses Pengunduhan: Melibatkan penggunaan perintah seperti `wget` untuk mengunduh file Apache Hadoop dan `apt` untuk mengunduh dan menginstal paket-paket di Ubuntu.

Peralatan Input dan Output (Mouse, Keyboard, Monitor):

Fungsi Utama: Digunakan untuk interaksi langsung dengan komputer atau laptop selama proses instalasi dan konfigurasi. Keyboard dan mouse digunakan untuk memasukkan perintah, sedangkan monitor menampilkan hasil dan informasi yang diperlukan.

Terminal atau Command Line Interface (CLI):

Fungsi Utama: Sebagian besar langkah-langkah praktikum dilakukan melalui terminal atau command line interface. Pengguna akan memasukkan perintah-perintah untuk membuat grup, menginstal perangkat lunak, mengedit file konfigurasi, dan menjalankan perintah-perintah Hadoop. Dengan menggunakan alat-alat tersebut, praktikan dapat secara langsung mengelola dan mengontrol proses instalasi dan konfigurasi Apache Hadoop pada lingkungan personal komputer atau laptop mereka.

2.2 Bahan

Untuk membuat praktikum perintah Hadoop menggunakan Virtual Box dan ISO Ubuntu, berikut adalah bahan-bahan yang umumnya digunakan:

Virtualization Software (Virtual Box):

Fungsi Utama: Digunakan untuk membuat dan mengelola mesin virtual. Sumber: Dapat diunduh dan diinstal secara gratis dari situs resmi VirtualBox: <https://www.virtualbox.org/>

Image ISO Ubuntu:

Fungsi Utama: Diperlukan untuk menginstal sistem operasi Linux pada mesin virtual. Sumber: Image ISO Ubuntu

dapat diunduh dari situs resmi Ubuntu: <https://ubuntu.com/download/desktop>

Mesin Virtual:

Fungsi Utama: Dibuat menggunakan Virtual Box untuk menjalankan sistem operasi Ubuntu secara terisolasi. Langkah-langkah: Buat mesin virtual baru pada Virtual Box. Atur parameter seperti jumlah RAM, kapasitas penyimpanan, dan jumlah prosesor. Pilih image ISO Ubuntu yang telah diunduh sebagai media instalasi.

Koneksi Internet:

Fungsi Utama: Diperlukan untuk mengunduh paket dan perangkat lunak yang dibutuhkan selama instalasi dan konfigurasi. Konfigurasi: Pastikan mesin virtual terhubung ke internet, baik melalui koneksi Wi-Fi atau kabel.

Terminal atau Command Line Interface (CLI):

Fungsi Utama: Digunakan untuk memasukkan perintah-perintah Hadoop dan menjalankan langkah-langkah praktikum. Pada Ubuntu: Terminal dapat diakses langsung pada antarmuka grafis atau melalui pintasan keyboard (Ctrl + Alt + T).

Perangkat Lunak Apache Hadoop:

Fungsi Utama: Apache Hadoop perlu diunduh dan diinstal pada mesin virtual Ubuntu. Sumber: Paket Apache Hadoop dapat diunduh dari situs resmi Apache Hadoop: <https://hadoop.apache.org/> Dengan menggunakan kombinasi Virtual Box, ISO Ubuntu, dan perangkat lunak Apache Hadoop, Anda dapat membuat lingkungan virtual yang memungkinkan praktikan untuk belajar dan mengimplementasikan perintah-perintah Hadoop tanpa mempengaruhi sistem operasi utama pada komputer mereka.

3 Prosedur Kerja

Jelaskan sesuai kondisi pada saat praktikum

1. Menginstall Ubuntu
 2. Menginstall dan mengkonfigurasi Apache Hadoop
 3. Membuat User baru untuk HDFS
 4. Menginstall Java
 5. Menginstall OpenSSH
 6. Menjalankan Hadoop Service dengan perintah :
 - start-dfs.sh
 - start-yarn.sh
 7. Menjalankan program WordCount bawaan Hadoop dan Java
 8. Menginstall Apache Spark
-

4 Hasil dan Pembahasan

Hasil dapat berupa gambar hasil praktikum, maupun data hasil praktikum

1. Menjalankan Ubuntu dan Instalasi Hadoop:

```
wget https://dldn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
```

2. Beralih ke user hdoop:

```
su - hdoop
```

```
rafli@rafli-VirtualBox:~$ su - hdfs
Password:
hdfs@rafli-VirtualBox:~$
```

3. Verifikasi Instalasi Java:

```
java -version
```

```
hdfs@rafli-VirtualBox:~$ java -version
openjdk version "1.8.0_392"
OpenJDK Runtime Environment (build 1.8.0_392-8u392-ga-1~20.04-b08)
OpenJDK 64-Bit Server VM (build 25.392-b08, mixed mode)
hdfs@rafli-VirtualBox:~$
```

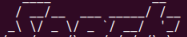
4. Menjalankan Peintah-Perintah Hadoop:

```
start-dfs.sh
```

```
hdfs@rafli-VirtualBox:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [rafli-VirtualBox]
```

5. Verifikasi Hasil Instalasi Apache Spark:

```
pyspark --version
```

```
hdfe@raffli-VirtualBox:~$ pyspark
Python 3.8.10 (default, Nov 22 2023, 10:22:35)
[GCC 9.4.0] on linux
Type "help", "copyright", "credits" or "license()" for more information.
3/12/25 01:45:16 WARN Utils: Your hostname, raffli-VirtualBox resolves to a loo
pbback address: 127.0.0.1; using 10.0.2.15 instead (on interface enp0s3)
23/12/25 01:45:16 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel
(newLevel).
Welcome to
 version 3.4.2

Using Python version 3.8.10 (default, Nov 22 2023 10:22:35)
Spark context Web UI available at http://10.0.2.15:4040
Spark context available as 'sc' (master = local[*], app id = local-170344353090
9).
SparkSession available as 'spark'.
>>>
```

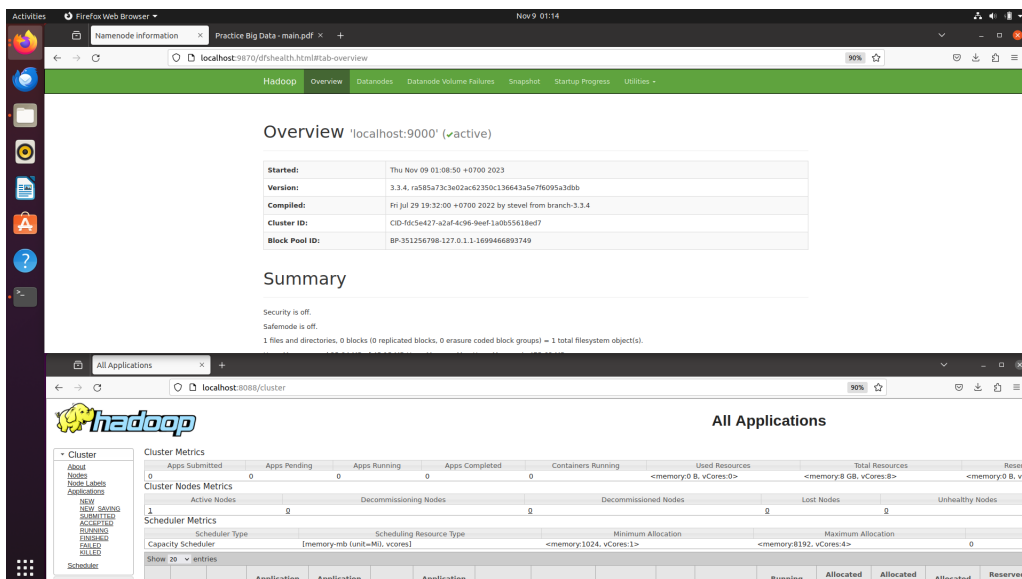
6. Menjalankan perintah jps:

jps

```
hdfs@rafl-1-VirtualBox:~$ jps
2146 Jps
1716 SparkSubmit
hdfs@rafl-1-VirtualBox:~$
```

7. Menjalankan hadoop service di localhost:

Akses melalui web browser dengan alamat [http://localhost: 98705](http://localhost:98705) atau <http://localhost:80886> .



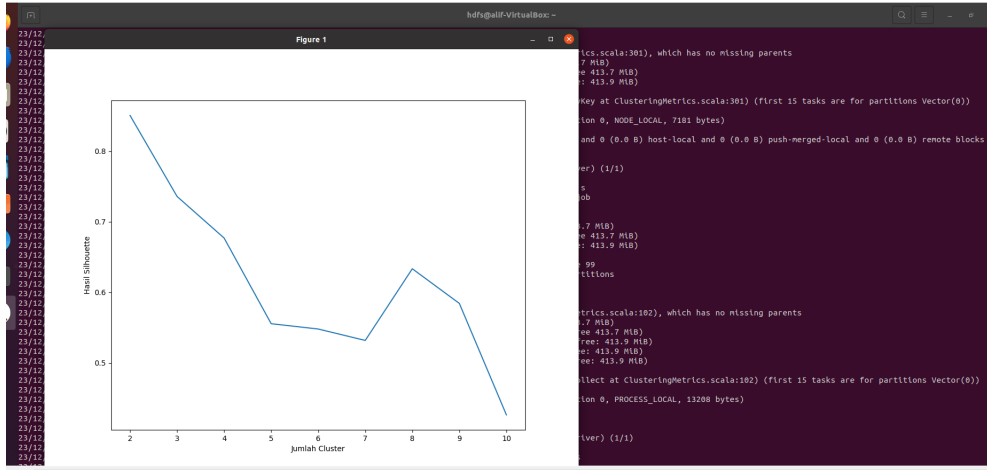
8. Melihat hasil part 3:

```
hadoop fs -ls /output
```

```
hadoop@hadoop:~$ hadoop fs -ls /output
Found 2 items
-rw-r--r-- 1 hdfs supergroup 0 2023-12-01 09:46 /output/_SUCCESS
-rw-r--r-- 1 hdfs supergroup 527522 2023-12-01 09:46 /output/part-r-00000
```

9. Melihat hasil grafik:

```
spark - submit MLPySpark.py
```



5 Kesimpulan

kesimpulan hasil praktikum

Pada pertemuan kedua, mahasiswa melakukan proses instalasi Apache Hadoop di lingkungan yang telah di persiapkan sebelumnya. Langkah-langkahnya termasuk membuat grup dan pengguna baru, menginstal Java, melakukan konfigurasi SSH, mengunduh dan mengekstrak Apache Hadoop, serta menyesuaikan hak akses. Setelah itu, kita melakukan konfigurasi berbagai file Hadoop dan membuat folder khusus untuk menyimpan data sementara dan direktori untuk Namenode/Datanode. Proses instalasi diakhiri dengan memformat HDFS. Hasil instalasi kemudian diverifikasi untuk memastikan bahwa semuanya berjalan dengan baik. Selain itu, kita juga melakukan beberapa konfigurasi pada file-file Hadoop agar memudahkan pemantauan ekosistem Hadoop yang telah diinstal. Secara keseluruhan, langkah-langkah ini membentuk dasar instalasi dan konfigurasi Apache Hadoop yang siap digunakan dalam pengolahan Big Data.

Daftar Pustaka

- [1] Mike Bowker and Phil Williams. Helsinki and west european security. *International Affairs*, 61(4):607–618, 1985.