**UTS PENGENALAN BIG DATA**

1. Cari dan sebutkan 3 DBMS yang bisa digunakan untuk mengelola big data.
2. Oracle Big Data SQL
3. Microsoft SQL Server
4. IBM DB2
5. SAP Saybase ASE
6. Apache Hadoop
7. Carilah contoh masalah big data yang bisa dikelola menggunakan salah satu DBMS tersebut, jelaskan mulai dari instalasi sampai CRUD untuk data menggunakan DBMS tersebut. Asumsikan anda akan memecahkan masalah big data yang sudah anda cari contoh tadi, jelaskan kira-kira bagaimana arsitektur dari solusi big data menggunakan DBMS tersebut, gambarkan diagramnya.
8. **Contoh DBMS Hadoop Dalam pengobatan Kanker dan Genomic**
9. **Instalasi Hadoop pada windows**

Software yang dibutuhkan :

1. Aplikasi Hadoop
2. Java JDK

Setalah software diatas di install, selanjutnya mengconfigurasinya.

1. Configurasi
2. Edit file **C: /Hadoop-2.8.0/etc/hadoop/core-site.xml** , rekatkan di bawah paragraf xml dan simpan file ini.

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

1. Ganti nama "mapred-site.xml.template" menjadi "mapred-site.xml" dan edit file ini C: /Hadoop-2.8.0/etc/hadoop/mapred-site.xml , tempel di bawah paragraf xml dan simpan file ini.

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

1. Buat 3 Folde

* Buat folder "data" di bawah "C: \ Hadoop-2.8.0"
* Buat folder "datanode" di bawah "C: \ Hadoop-2.8.0 \ data"
* Buat folder "namenode" di bawah "C: \ Hadoop-2.8.0 \ data"

1. Edit file **C: \ Hadoop-2.8.0 / etc / hadoop / hdfs-site.xml** , rekatkan di bawah paragraf xml dan simpan file ini.

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>/hadoop-2.8.0/data/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>/hadoop-2.8.0/data/datanode</value>

</property>

</configuration>

1. Edit file C: /Hadoop-2.8.0/etc/hadoop/yarn-site.xml , rekatkan di bawah paragraf xml dan simpan file ini.

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

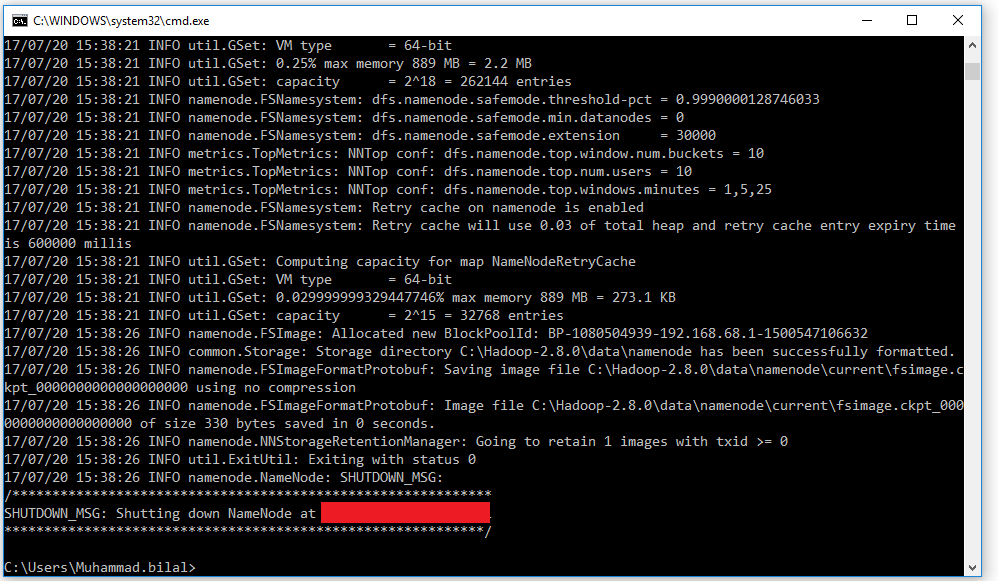
</property>

</configuration>

1. dit file C: /Hadoop-2.8.0/etc/hadoop/hadoop-env.cmd dengan menutup baris perintah “JAVA\_HOME =% JAVA\_HOME%” alih-alih mengatur “JAVA\_HOME = C: \ Java” (Pada C: \ java this path ke file jdk.18.0).

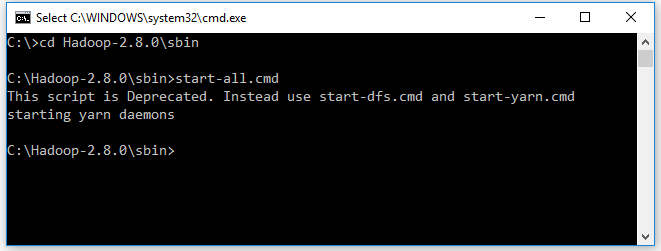
**Konfigurasi Hadoop**

1. File Dowload, Hadoop Configuration.zip
2. Hapus file bin pada C: \ Hadoop-2.8.0 \ bin, diganti dengan file bin pada file yang baru saja diunduh (dari Hadoop Configuration.zip)
3. Buka cmd dan ketikkan perintah “hdfs namenode –format” . Maka Hasilnya seperti dibawa ini.



**Menjalankan Hadoop**

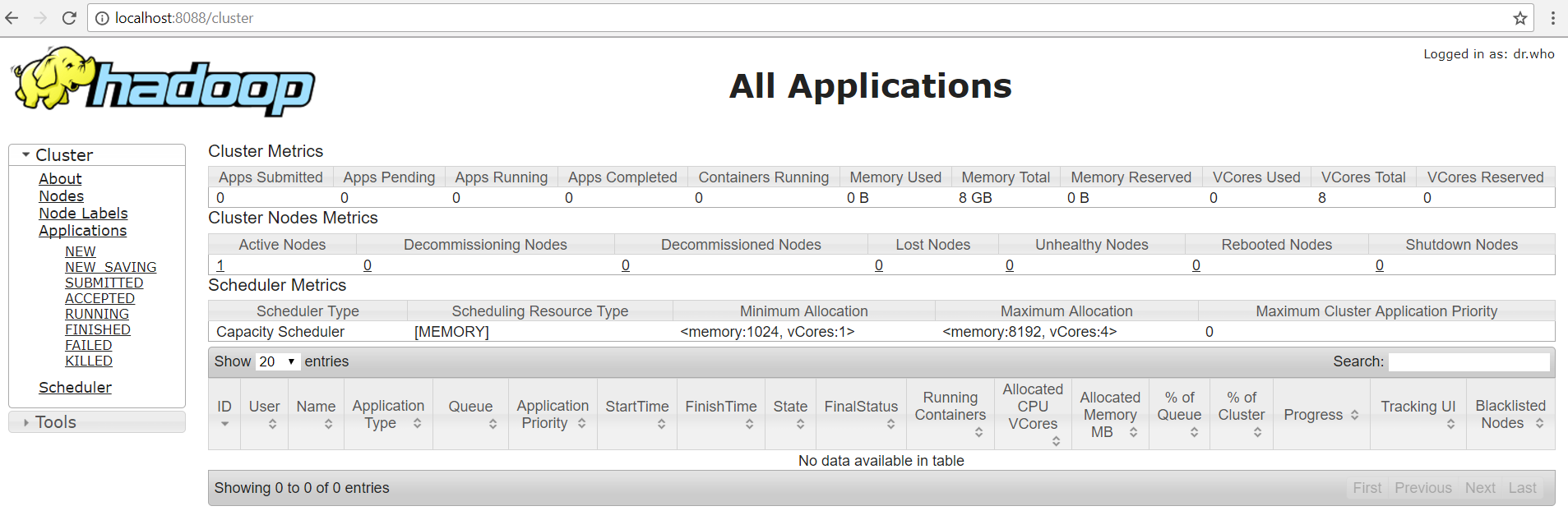
1. Buka cmd dan ubah direktori menjadi "C: \ Hadoop-2.8.0 \ sbin" dan ketik "start-all.cmd" untuk memulai apache.



1. Pastikan aplikasi ini berjalan

* Hadoop Namenode
* Dadoode Hadoop
* Manajer Sumber Daya YARN
* YARN Node Manager

1. Buka :http://localhost:8088

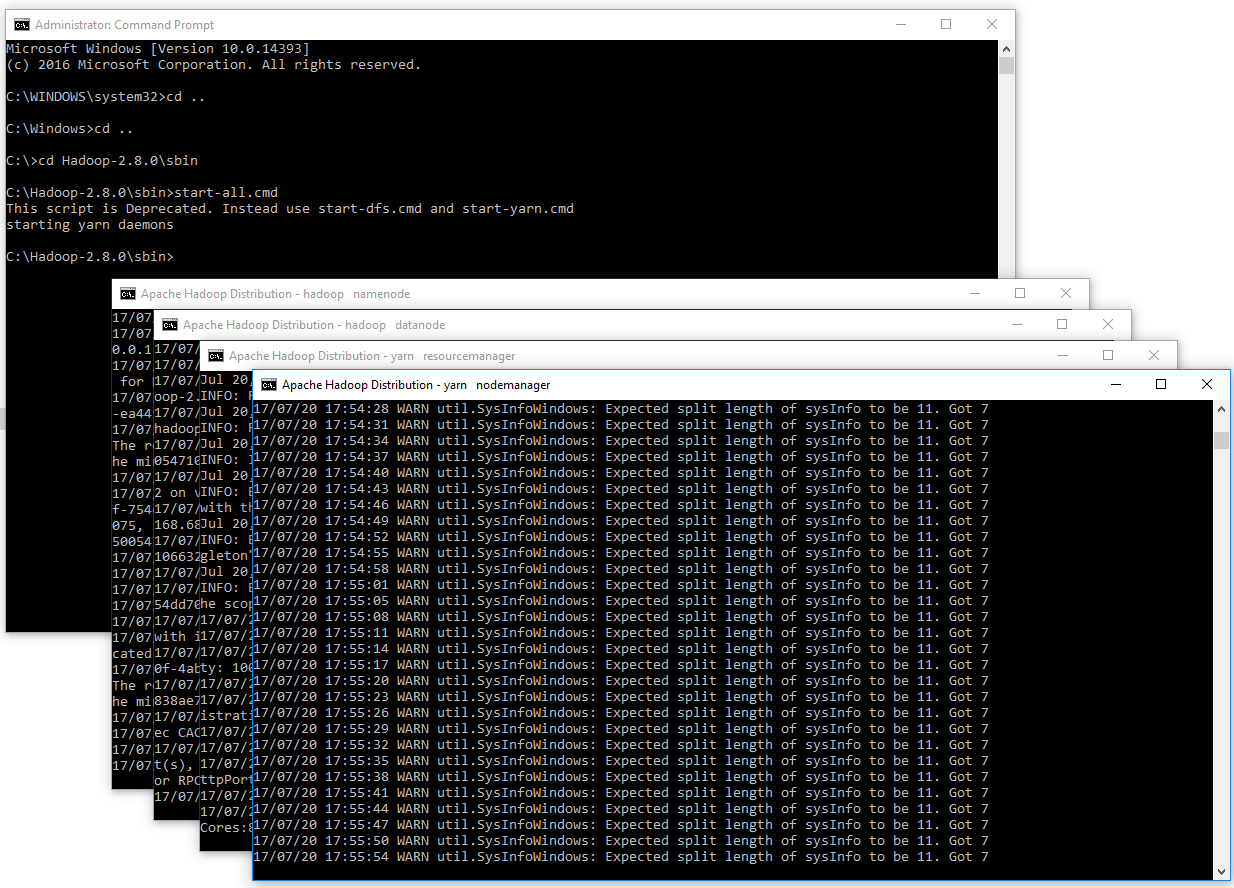


1. **Contoh CRUD Dalam Hadoop**

Agar mempermudah pengelolaan data maka kita perlu software atau framework MapReduce yang di mana MapReduce merupakan sebuah model pemograman yang didesain untuk dapat melakukan pemrosesan data dengan jumlah yang sangat besar dengan cara membagi pemrosesan tersebut ke beberapa tugas yang indipenden satu sama lain. Selanjutnya MapReduceClient.jar dan Data yang akan digunakan taruh di "C: /".

1. Buka cmd dalam mode Administratif dan pindah ke "C: /Hadoop-2.8.0/sbin" dan mulai cluster.

Start-all.cmd



1. Buat direktori input dalam HDFS.

hadoop fs -mkdir /input\_dir

1. Salin file teks input bernama input\_file.txt(data yang akan digunakan) di direktori input (input\_dir) HDFS.

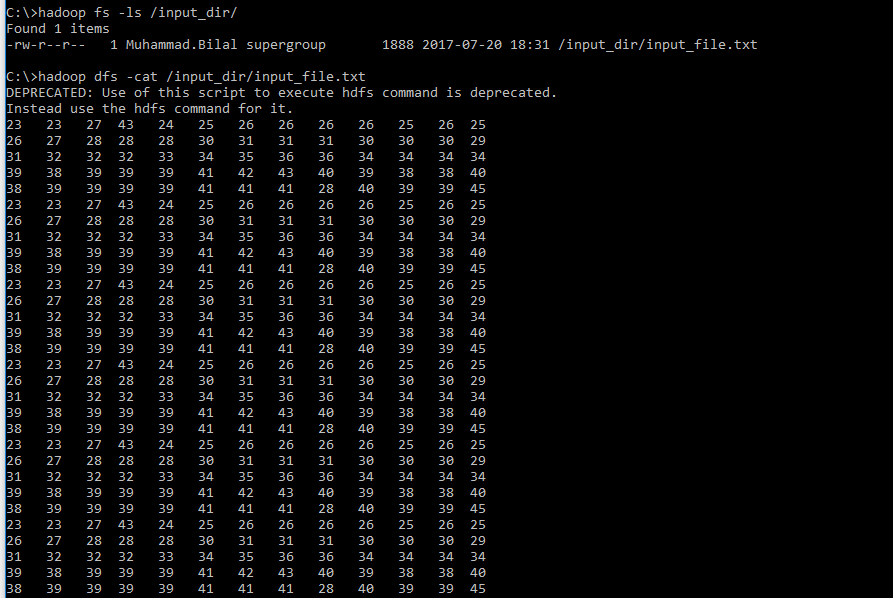
hadoop fs -put C:/input\_file.txt /input\_dir

1. Verifikasi input\_file.txt tersedia di direktori input HDFS (input\_dir).

hadoop fs -ls /input\_dir/

1. Verifikasi konten dari file yang disalin.

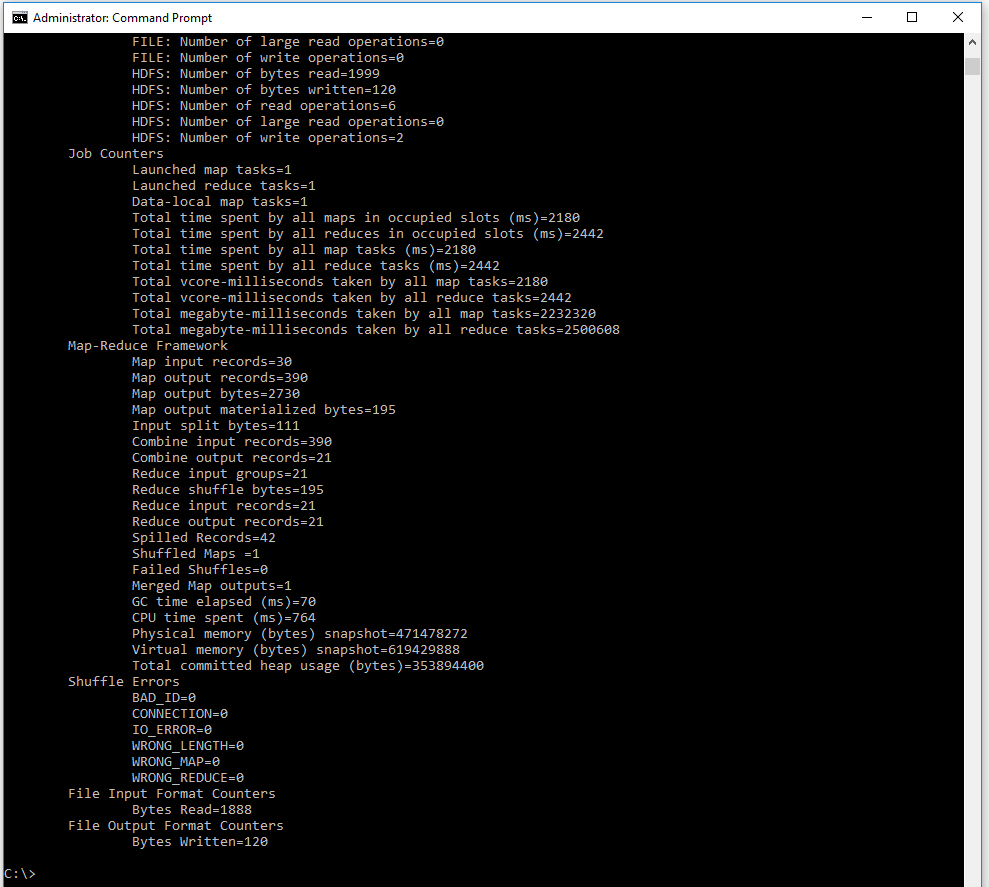
hadoop dfs -cat /input\_dir/input\_file.txt



Contoh Datanya

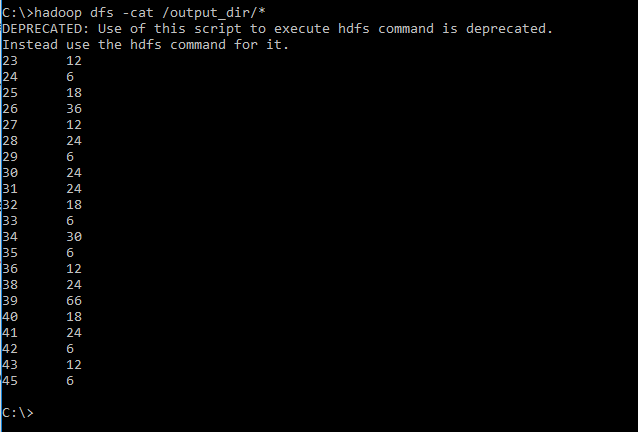
1. Jalankan MapReduceClient.jar dan berikan juga masukan dan direktori keluar.

hadoop jar C:/MapReduceClient.jar wordcount /input\_dir /output\_dir



1. Untuk melihat file output yang dihasilkan.

hadoop dfs -cat /output\_dir/\*



Hasil Outputnya

1. **Arsitektur Dari Contoh Kasus A**

Pada arsitektur ini terdapat 3 bagian utama yaitu All Data Source,Big Data Ecosystem dan Analytic Applications. Yang dimana pada data source terdapat berbagai macam data mulai dari teks,audio,video dll untuk di exstraksi ke data ecosystem sehingga menjadi data yang akan digunakan untuk menganalisis bagaimana penyebab dan pengaruh DNA dan GEN terhadapat penyakit kanker.

