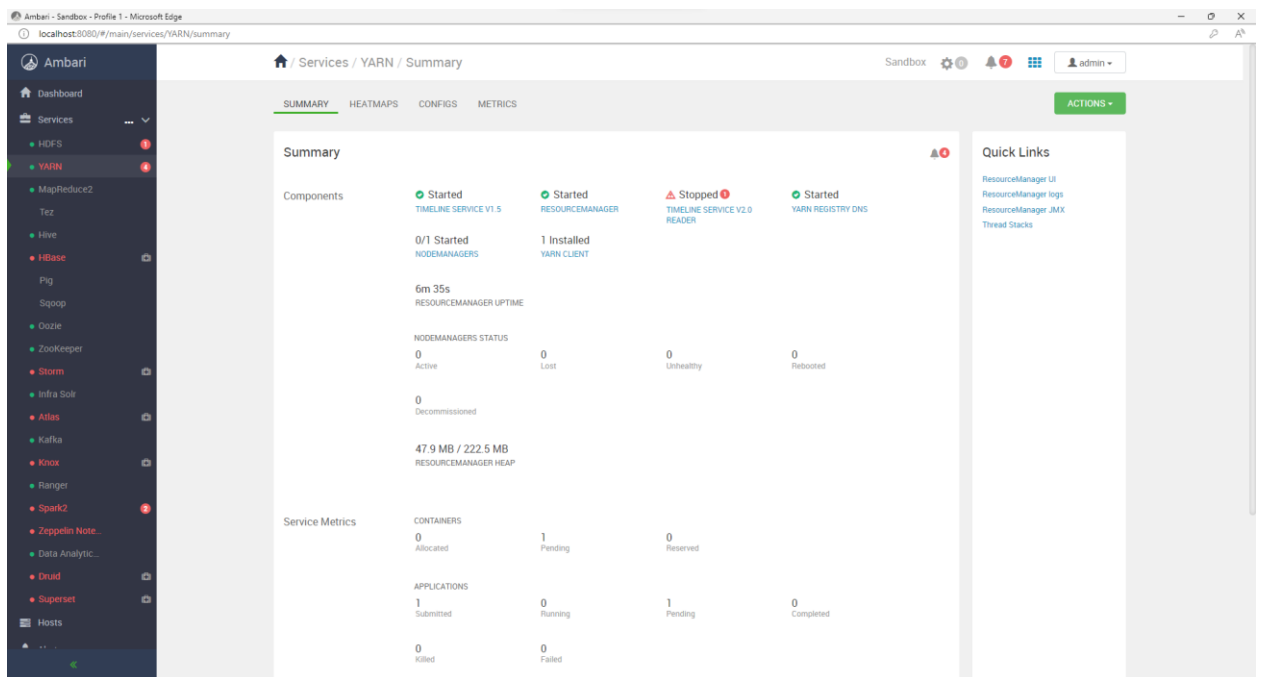


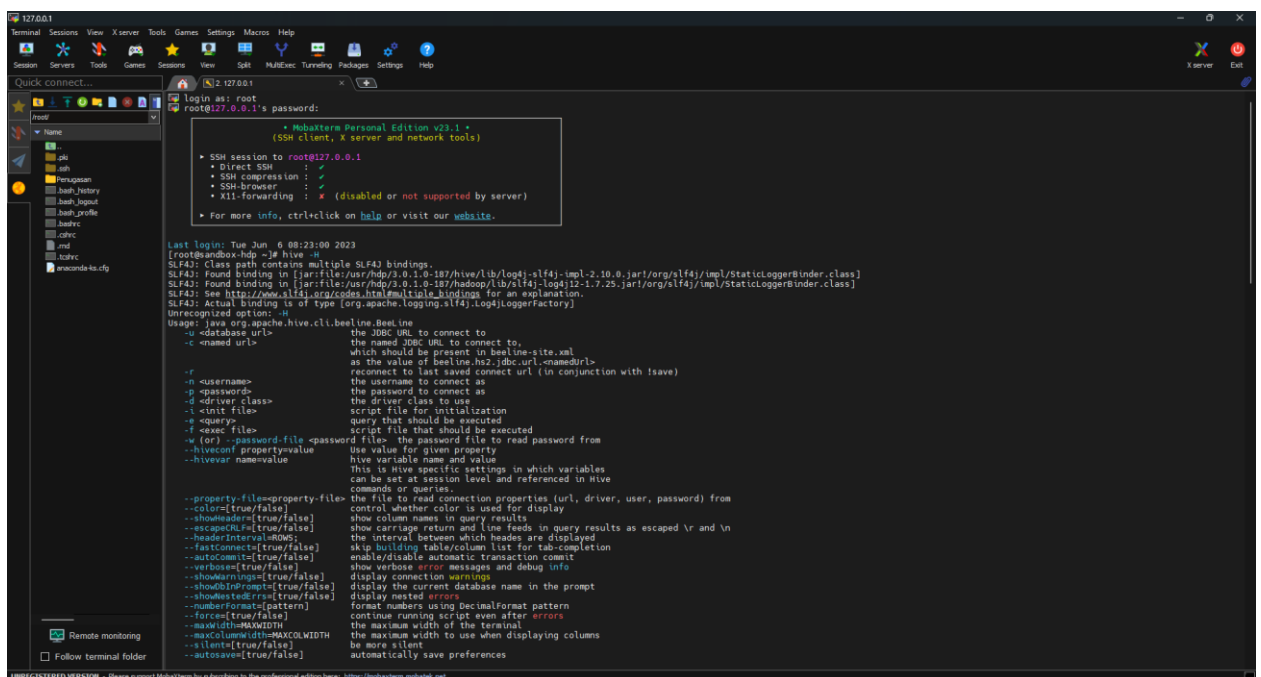
Nama : Komang Niko Romano Prodi
NIM : 222011356
Kelas : 3SD1
No Absen : 06

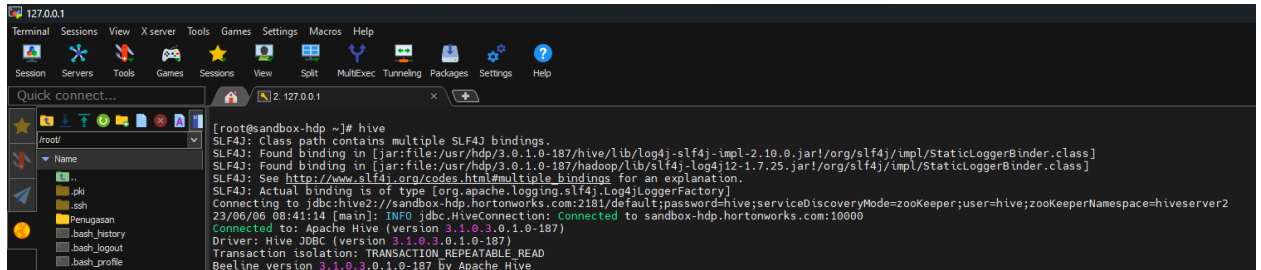
Praktikum 6 Teknologi Big Data – DISTRIBUTED DATAWAREHOUSE DENGAN HIVE

1. Memastikan service dalam keadaan hidup



2. Memastikan hive telah terinstall dan bisa digunakan

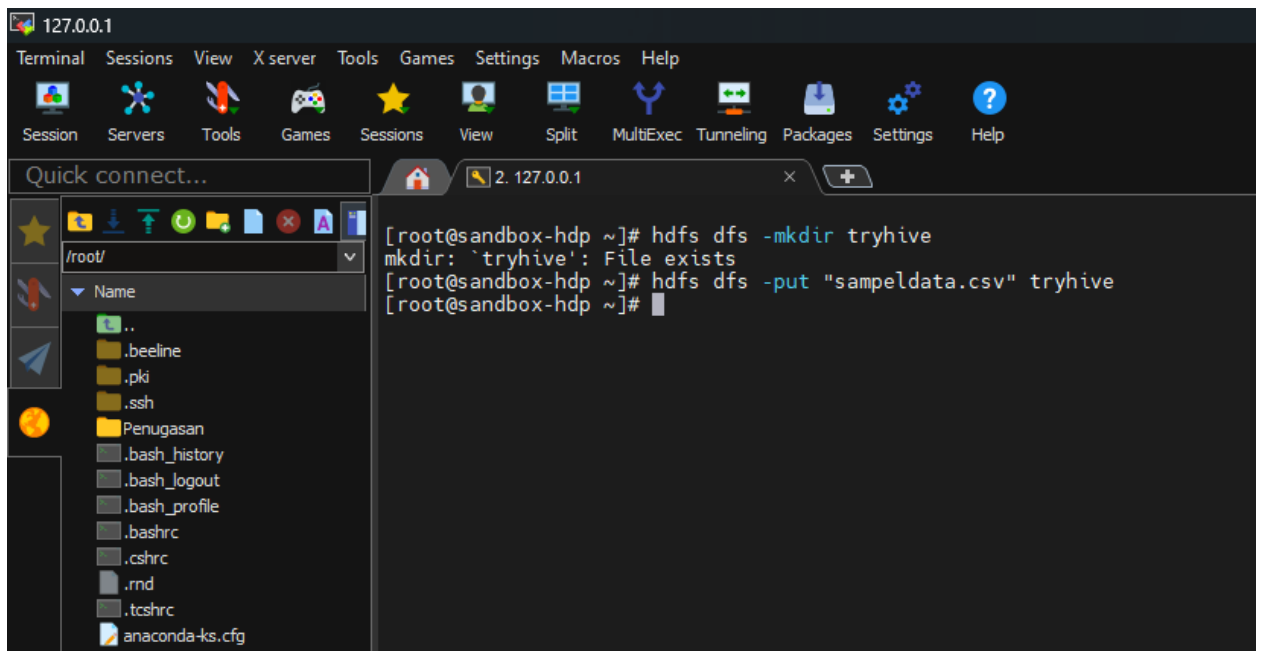




Terminal window showing Hive connection logs. The logs indicate that the class path contains multiple SLF4J bindings and that the connection to Apache Hive (version 3.1.0.3.0.1.0-187) was successful. The transaction isolation is TRANSACTION_REPEATABLE_READ and the Beeline version is 3.1.0.3.0.1.0-187.

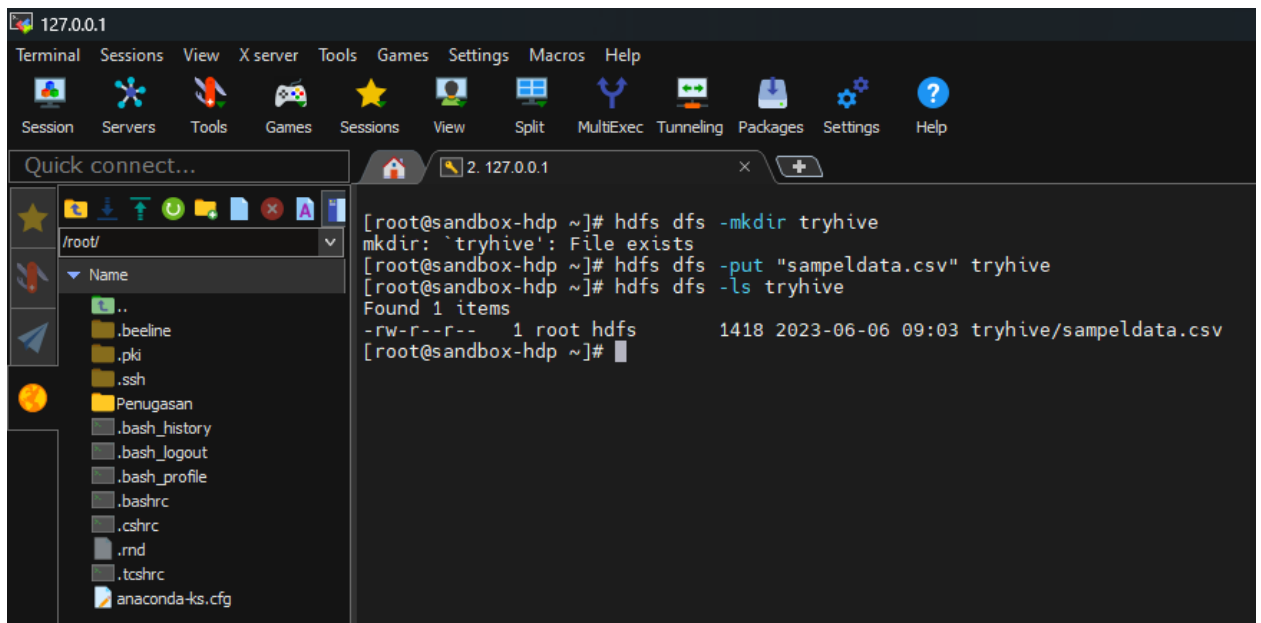
```
[root@sandbox-hdp ~]# hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://sandbox-hdp.hortonworks.com:2181/default;password=hive;serviceDiscoveryMode=zooKeeper;user=hive;zooKeeperNamespace=hiveserver2
23/06/06 08:41:14 [main]: INFO jdbc.HiveConnection: Connected to sandbox-hdp.hortonworks.com:10000
Connected to: Apache Hive (version 3.1.0.3.0.1.0-187)
Driver: Hive JDBC (version 3.1.0.3.0.1.0-187)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.0.1.0-187 By Apache Hive
```

3. Lakukan input sampel data



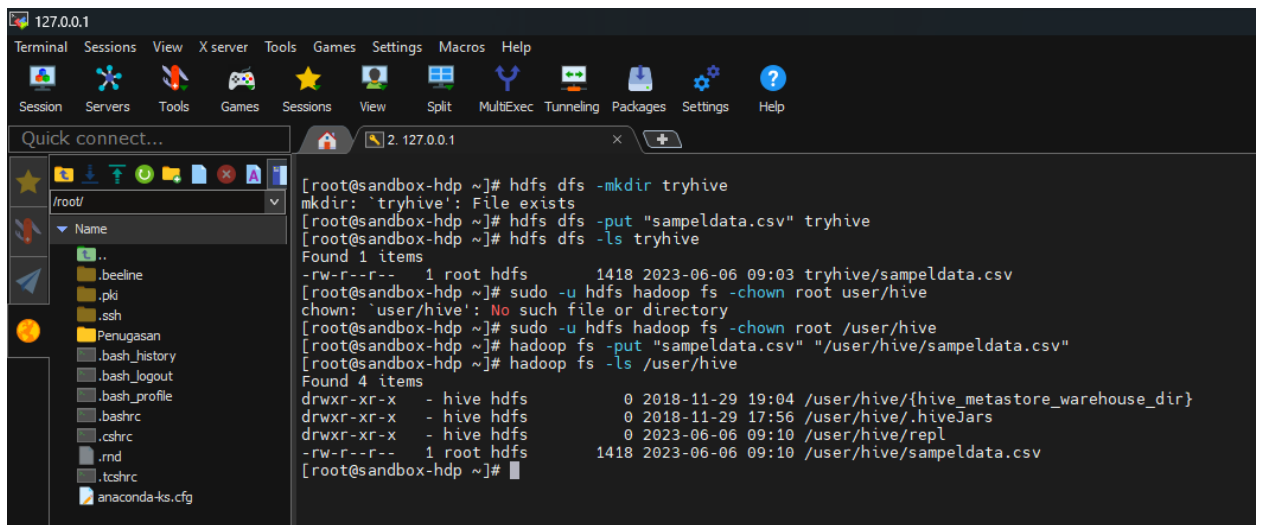
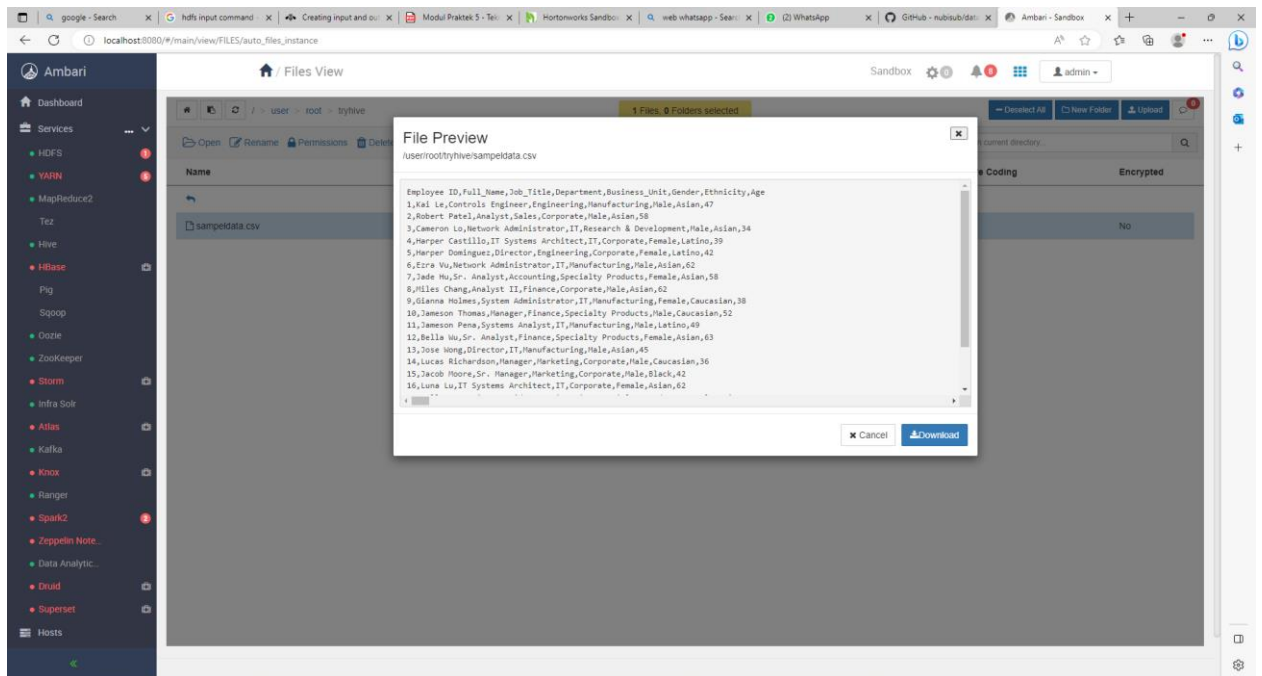
Terminal window showing HDFS commands. The user creates a directory named 'tryhive' and uploads a file named 'sampeldata.csv' to it. The file exists.

```
[root@sandbox-hdp ~]# hdfs dfs -mkdir tryhive
mkdir: `tryhive': File exists
[root@sandbox-hdp ~]# hdfs dfs -put "sampeldata.csv" tryhive
[root@sandbox-hdp ~]#
```

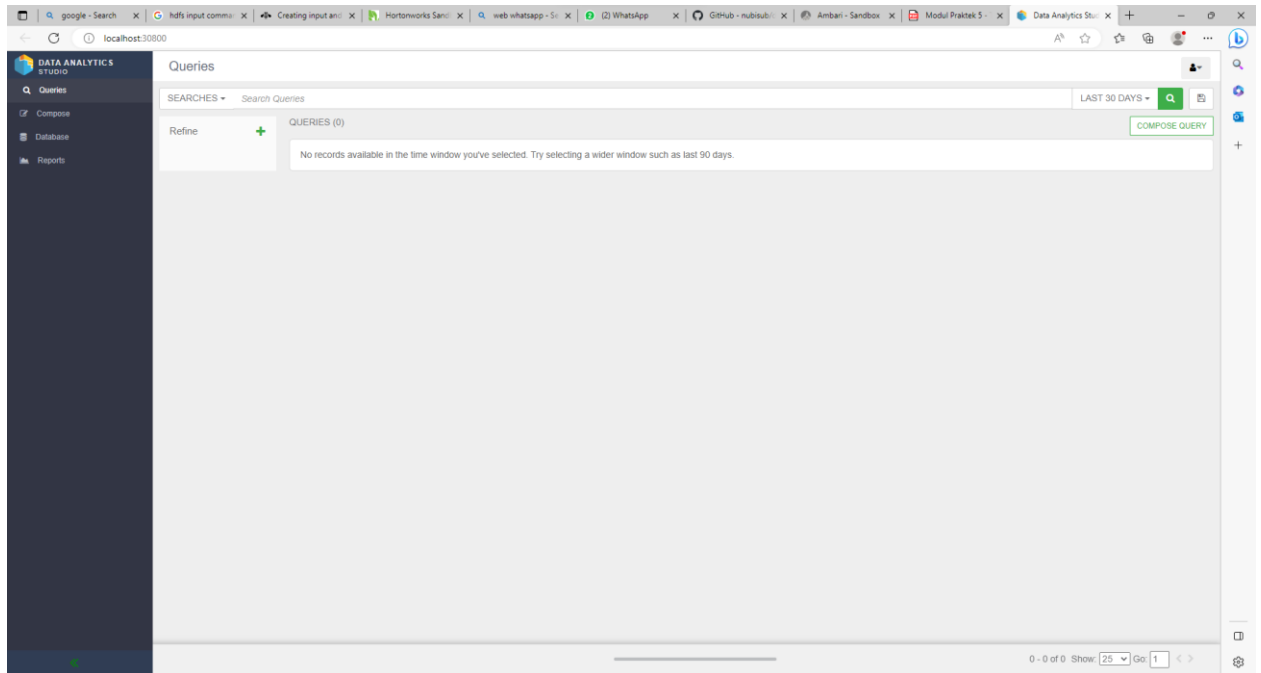


Terminal window showing HDFS commands and file listing. The user creates a directory named 'tryhive', uploads a file named 'sampeldata.csv' to it, and lists the contents of the directory. The output shows one item: 'tryhive/sampeldata.csv' with permissions '-rw-r--r--', owner 'root', group 'hdfs', size '1418', and timestamp '2023-06-06 09:03'.

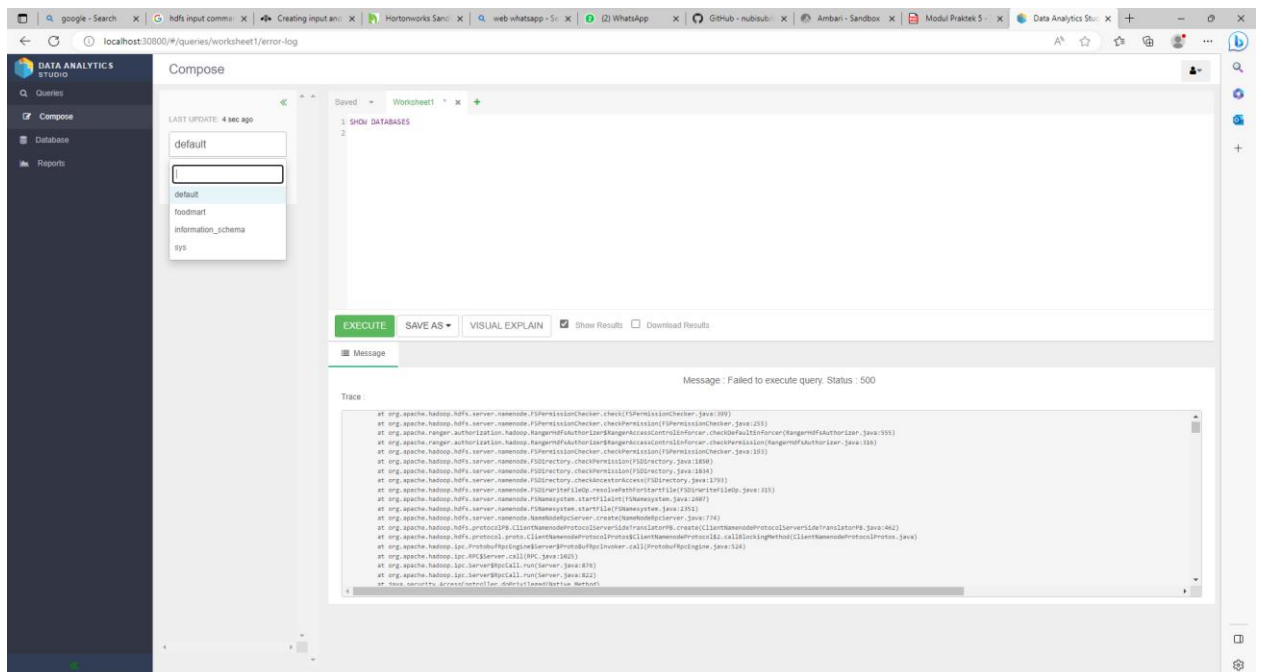
```
[root@sandbox-hdp ~]# hdfs dfs -mkdir tryhive
mkdir: `tryhive': File exists
[root@sandbox-hdp ~]# hdfs dfs -put "sampeldata.csv" tryhive
[root@sandbox-hdp ~]# hdfs dfs -ls tryhive
Found 1 items
-rw-r--r-- 1 root hdfs 1418 2023-06-06 09:03 tryhive/sampeldata.csv
[root@sandbox-hdp ~]#
```



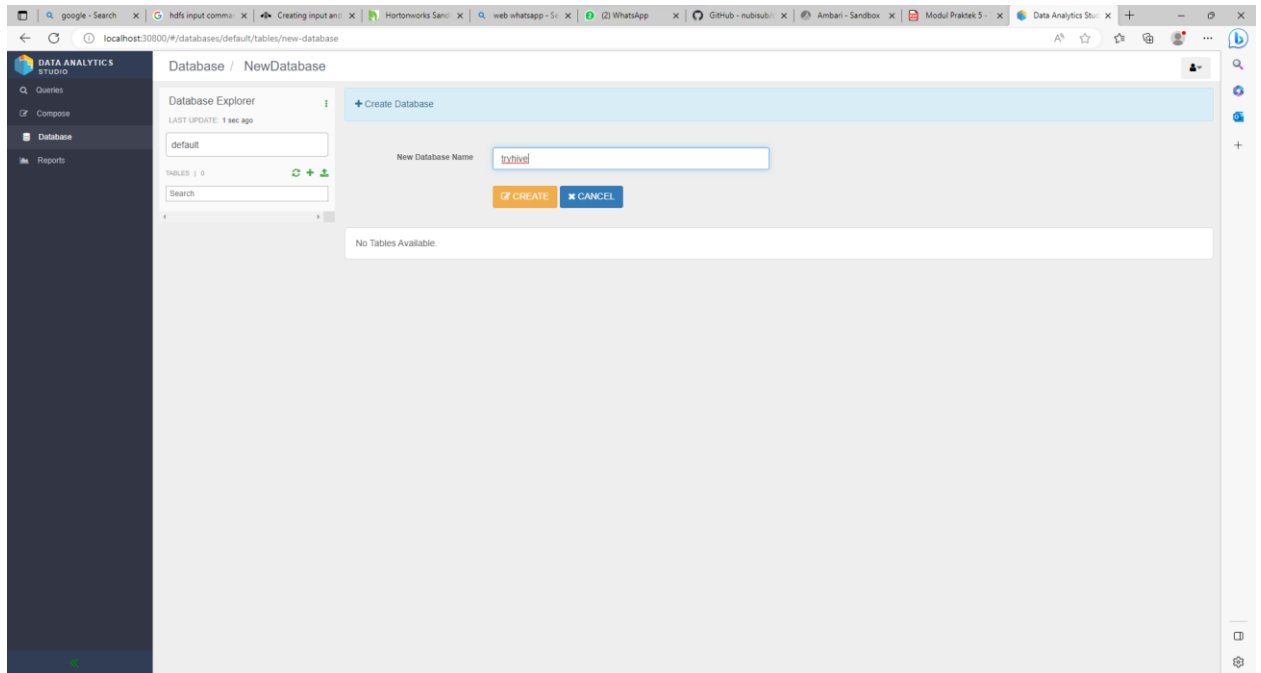
4. Analisis data menggunakan DAS UI



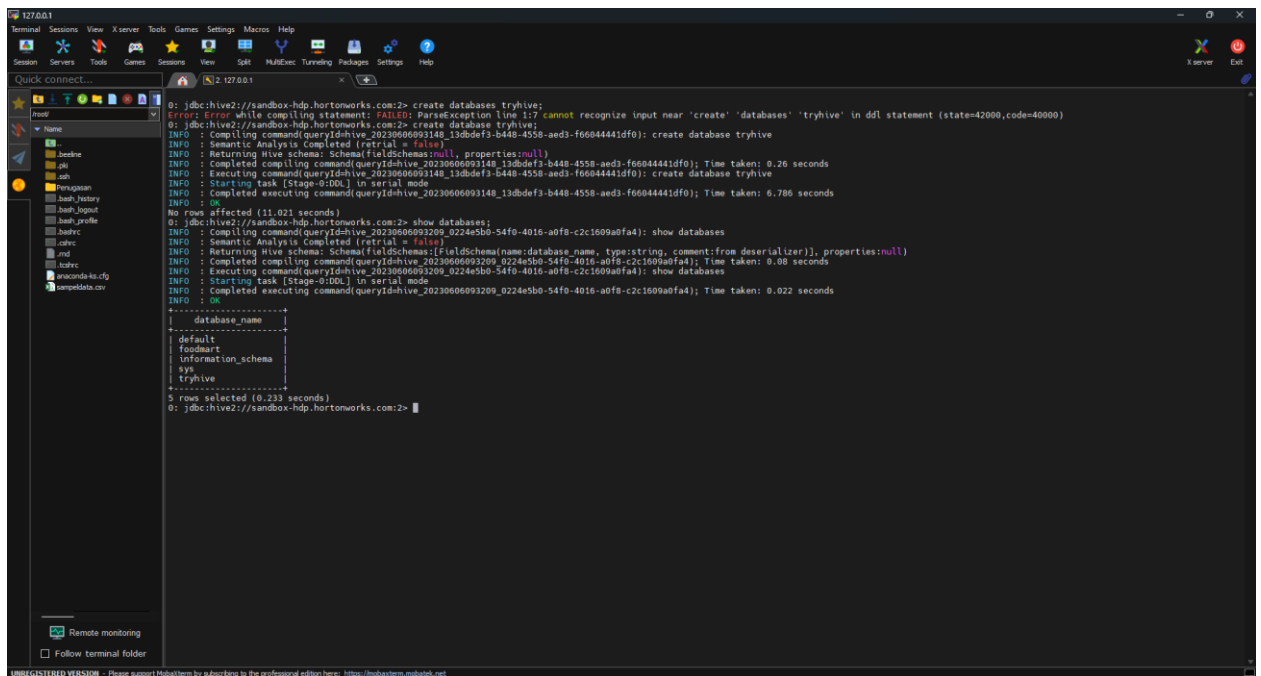
Menampilkan seluruh databases



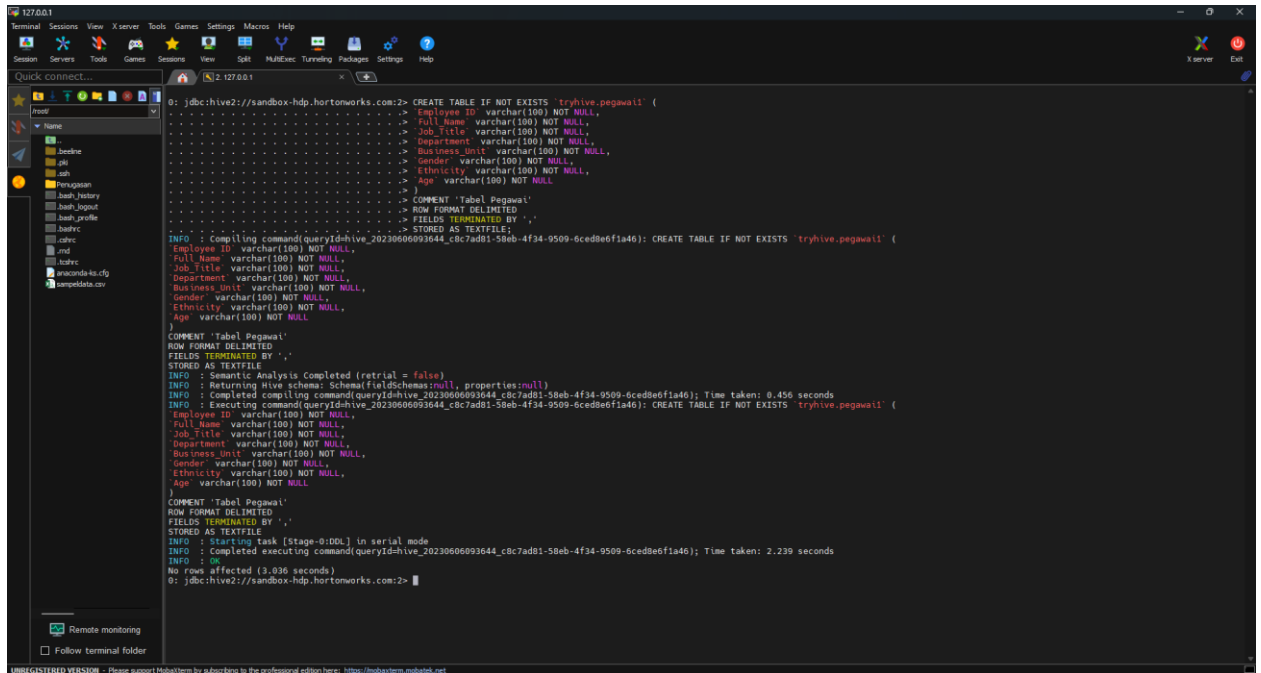
Membuat database



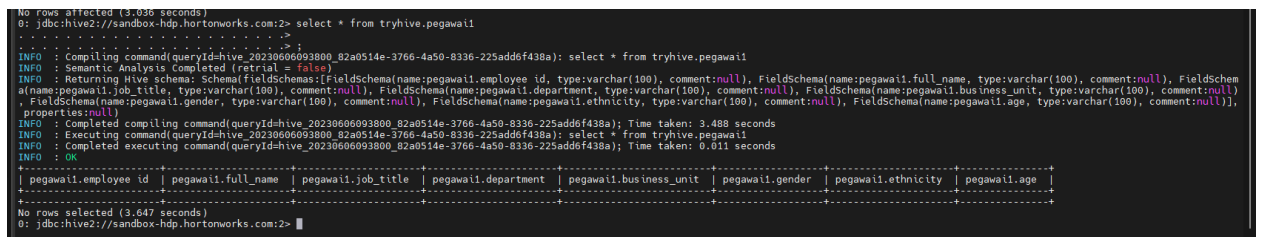
Cara lain yaitu melalui mobaxterm menggunakan syntax:



Membuat tabel

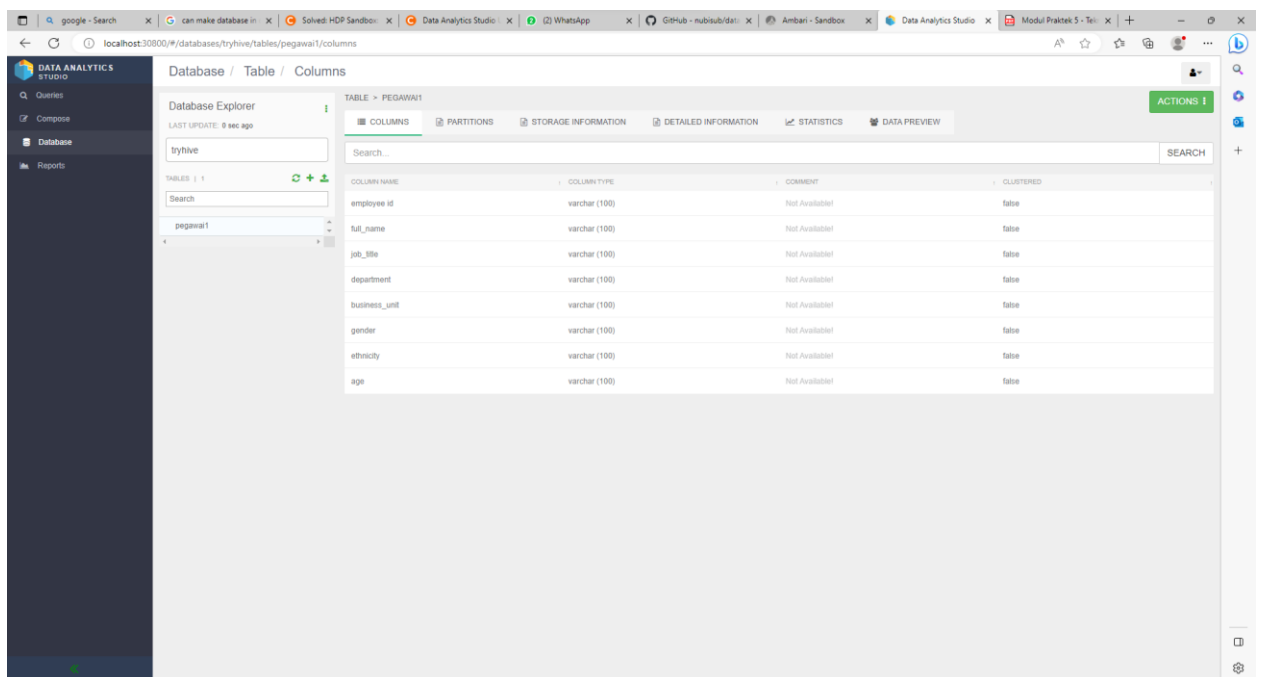


```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> CREATE TABLE IF NOT EXISTS 'tryhive.pegawai' (
  Employee ID varchar(100) NOT NULL,
  Full Name varchar(100) NOT NULL,
  Job Title varchar(100) NOT NULL,
  Department varchar(100) NOT NULL,
  Business Unit varchar(100) NOT NULL,
  Gender varchar(100) NOT NULL,
  Ethnicity varchar(100) NOT NULL,
  Age varchar(100) NOT NULL
)
COMMENT 'Tabel Pegawai'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
INFO : Compiling command(queryId=hive_20230606093644_c8c7ad81-58eb-4f34-9509-6ced8e6fia46): CREATE TABLE IF NOT EXISTS 'tryhive.pegawai' (
  Employee ID varchar(100) NOT NULL,
  Full Name varchar(100) NOT NULL,
  Job Title varchar(100) NOT NULL,
  Department varchar(100) NOT NULL,
  Business Unit varchar(100) NOT NULL,
  Gender varchar(100) NOT NULL,
  Ethnicity varchar(100) NOT NULL,
  Age varchar(100) NOT NULL
)
COMMENT 'Tabel Pegawai'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fields:schema:null, properties:null)
INFO : Completed compiling command(queryId=hive_20230606093644_c8c7ad81-58eb-4f34-9509-6ced8e6fia46); Time taken: 0.456 seconds
INFO : Executing command(queryId=hive_20230606093644_c8c7ad81-58eb-4f34-9509-6ced8e6fia46): CREATE TABLE IF NOT EXISTS 'tryhive.pegawai' (
  Employee ID varchar(100) NOT NULL,
  Full Name varchar(100) NOT NULL,
  Job Title varchar(100) NOT NULL,
  Department varchar(100) NOT NULL,
  Business Unit varchar(100) NOT NULL,
  Gender varchar(100) NOT NULL,
  Ethnicity varchar(100) NOT NULL,
  Age varchar(100) NOT NULL
)
COMMENT 'Tabel Pegawai'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
INFO : Starting Task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230606093644_c8c7ad81-58eb-4f34-9509-6ced8e6fia46); Time taken: 2.239 seconds
INFO : OK
No rows affected (3.036 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```



```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from tryhive.pegawai
+-----+-----+-----+-----+-----+-----+-----+-----+
| pegawai.employee id | pegawai.full_name | pegawai.job_title | pegawai.department | pegawai.business_unit | pegawai.gender | pegawai.ethnicity | pegawai.age |
+-----+-----+-----+-----+-----+-----+-----+-----+
No rows selected (3.647 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Cek tabel melalui DAS UI



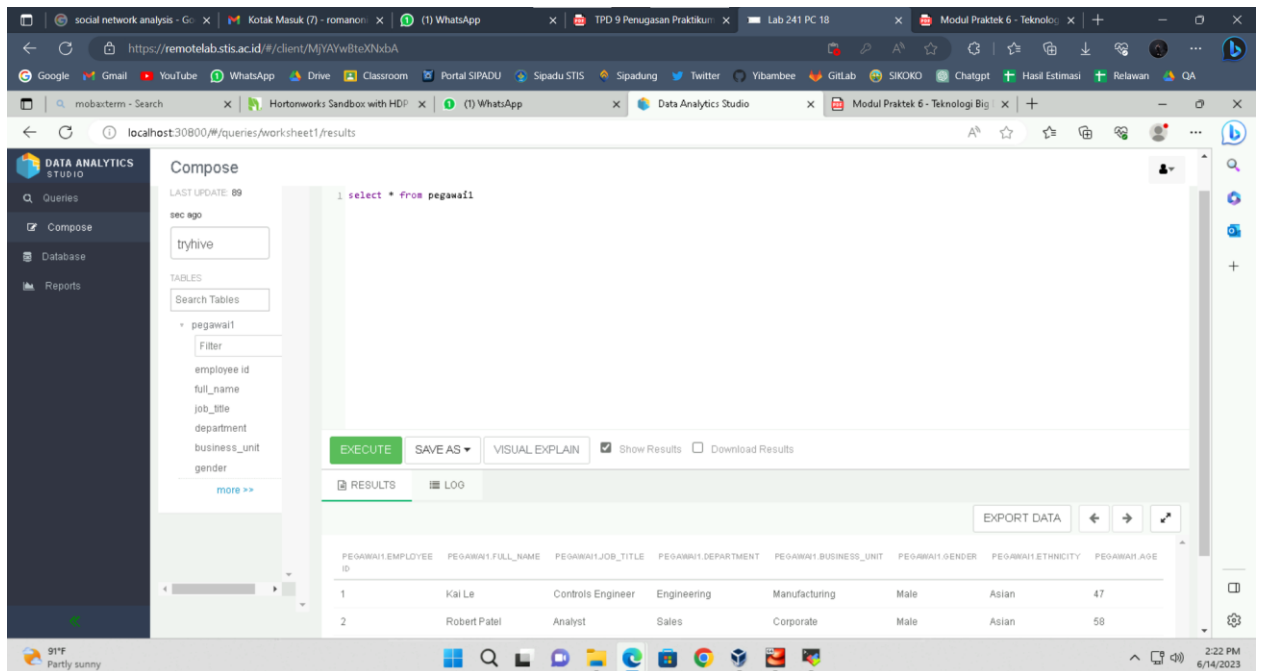
The screenshot shows the Data Analytics Studio (DAS) interface. On the left, there is a sidebar with 'Database Explorer' showing the 'tryhive' database. The main area displays the 'pegawai' table structure with columns: employee_id, full_name, job_title, department, business_unit, gender, ethnicity, and age. The table is located in the 'tryhive' database. The interface includes a search bar, a table preview, and various action buttons.

COLUMN NAME	COLUMN TYPE	COMMENT	CLUSTERED
employee_id	varchar(100)	Not Available	false
full_name	varchar(100)	Not Available	false
job_title	varchar(100)	Not Available	false
department	varchar(100)	Not Available	false
business_unit	varchar(100)	Not Available	false
gender	varchar(100)	Not Available	false
ethnicity	varchar(100)	Not Available	false
age	varchar(100)	Not Available	false

Insert data ke tabel yang telah dibuat

[illegible]

Cek data yang telah diinsert apakah sudah terisi ke dalam table



5. Penugasan

Load data kedalam HIVE untuk data dengan beberapa format yang berbeda yaitu `sampeldata.csv`, `sampel_data.parquet` dan `sampel_sata.avro`. Lakukan beberapa kueri sederhana seperti contoh yang terdapat pada slide kuliah teori. Hasil pekerjaan praktikum berupa dokumentasi atau screenshot Proses dan Hasil kueri data ke HIVE

a. Syntax create table.parquet

```
CREATE TABLE penugasan9.pegawaiparq(  
  `Employee_ID` varchar(100) NOT NULL, `Full_Name`  
  varchar(100) NOT NULL, `Job_Title` varchar(100) NOT  
  NULL, `Department` varchar(100) NOT NULL, `Business_Unit`  
  varchar(100) NOT NULL, `Gender` varchar(100) NOT  
  NULL, `Ethnicity` varchar(100) NOT NULL, `Age` varchar(100)  
  NOT NULL)  
STORED AS PARQUET;
```

b. Syntax load data.parquet

```
LOAD DATA INPATH '/user/hive/sample_data.parquet' OVERWRITE  
INTO TABLE pegawaiparq
```

c. Syntax create table.parquet

```
CREATE TABLE pegawaiavro  
ROW FORMAT SERDE  
'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
```



```

STORED AS INPUTFORMAT
'org.apache.hadoop.hive.ql.io.avro.AvroContainerInputFormat'
OUTPUTFORMAT
'org.apache.hadoop.hive.ql.io.avro.AvroContainerOutputFormat'
TBLPROPERTIES
('avro.schema.literal'='{ "type": "record", "name": "Record", "
fields": [{ "name": "Full_Name", "type": "string" }
, { "name": "Job_Title", "type": "string" }, { "name": "Department"
, "type": "string" }, { "name": "Business_Unit",
"type": "string" }, { "name": "Gender", "type": "string" }, { "name"
: "Ethnicity", "type": "string" }, { "name": "Age"
, "type": "long" } ] }');

```

d. Syntax load data .avro

```

LOAD DATA INPATH '/user/hive/sample_data.avro' OVERWRITE
INTO TABLE pegawaiavro

```

e. Import data ke hadoop

```

• MobaXterm Personal Edition v23.0 •
(SSh client, X server and network tools)

► SSH session to root@127.0.0.1
• Direct SSH      : ✓
• SSH compression : ✓
• SSH-browser     : ✓
• X11-forwarding  : ✗ (disabled or not supported by server)

► For more info, ctrl+click on help or visit our website.

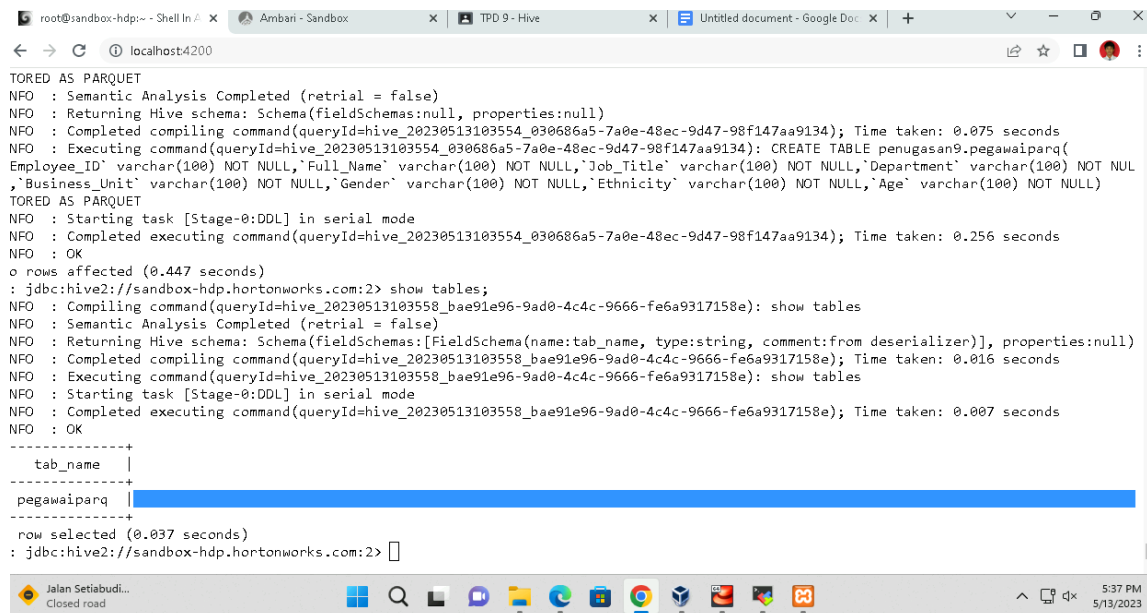
Last login: Tue May 9 07:07:43 2023 from 172.18.0.3
[root@sandbox-hdp ~]# hdfs dfs -mkdir penugasan
[root@sandbox-hdp ~]# hdfs dfs -put "sample_data.avro" "/user/root/penugasan9/"^C
[root@sandbox-hdp ~]# hdfs dfs -mkdir penugasan9
[root@sandbox-hdp ~]# hdfs dfs -put "sample_data.avro" "user/root/penugasan9"
put: 'user/root/penugasan9': No such file or directory: 'hdfs://sandbox-hdp.hortonwork
s.com:8020/user/root/user/root/penugasan9'
[root@sandbox-hdp ~]# hdfs dfs -put "sample_data.avro" "/user/root/penugasan9"
[root@sandbox-hdp ~]# hdfs dfs -ls penugasan9
Found 1 items
-rw-r--r--  1 root hdfs      1655 2023-05-09 07:52 penugasan9/sample_data.avro
[root@sandbox-hdp ~]# hdfs dfs -put "sample_data.parquet" "/user/root/penugasan9"
[root@sandbox-hdp ~]# hdfs dfs -ls penugasan9
Found 2 items
-rw-r--r--  1 root hdfs      1655 2023-05-09 07:52 penugasan9/sample_data.avro
-rw-r--r--  1 root hdfs      6878 2023-05-09 07:53 penugasan9/sample_data.parquet
[root@sandbox-hdp ~]#

```

```
[root@sandbox-hdp ~]# hadoop fs -put "sample_data.parquet" "/user/hive/sample_data.parquet"
[root@sandbox-hdp ~]# hadoop fs -put "sample_data.avro" "/user/hive/sample_data.avro"
[root@sandbox-hdp ~]# hdfs dfs -ls /user/hive
Found 5 items
drwxr-xr-x  - hive hdfs          0 2018-11-29 19:04 /user/hive/{hive_metastore_warehouse_dir}
drwxr-xr-x  - hive hdfs          0 2018-11-29 17:56 /user/hive/.hiveJars
drwxr-xr-x  - hive hdfs          0 2023-03-21 10:59 /user/hive/repl
-rw-r--r--  1 root hdfs        1655 2023-05-09 07:55 /user/hive/sample_data.avro
-rw-r--r--  1 root hdfs        6878 2023-05-09 07:55 /user/hive/sample_data.parquet
[root@sandbox-hdp ~]#
```

f. Sample_data.parque

Create table untuk format .parquet



```
root@sandbox-hdp: ~ - Shell In / x  Ambani - Sandbox x  TPD 9 - Hive x  Untitled document - Google Doc x  +
localhost4200
TOROED AS PARQUET
NFO : Semantic Analysis Completed (retrial = false)
NFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
NFO : Completed compiling command(queryId=hive_20230513103554_030686a5-7a0e-48ec-9d47-98f147aa9134); Time taken: 0.075 seconds
NFO : Executing command(queryId=hive_20230513103554_030686a5-7a0e-48ec-9d47-98f147aa9134): CREATE TABLE penugasan9.pegawaiparq(
Employee_ID` varchar(100) NOT NULL,`Full_Name` varchar(100) NOT NULL,`Job_Title` varchar(100) NOT NULL,`Department` varchar(100) NOT NUL
,`Business_Unit` varchar(100) NOT NULL,`Gender` varchar(100) NOT NULL,`Ethnicity` varchar(100) NOT NULL,`Age` varchar(100) NOT NULL)
TOROED AS PARQUET
NFO : Starting task [Stage-0:DDL] in serial mode
NFO : Completed executing command(queryId=hive_20230513103554_030686a5-7a0e-48ec-9d47-98f147aa9134); Time taken: 0.256 seconds
NFO : OK
0 rows affected (0.447 seconds)
: jdbc:hive2://sandbox-hdp.hortonworks.com:2> show tables;
NFO : Compiling command(queryId=hive_20230513103558_bae91e96-9ad0-4c4c-9666-fe6a9317158e): show tables
NFO : Semantic Analysis Completed (retrial = false)
NFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
NFO : Completed compiling command(queryId=hive_20230513103558_bae91e96-9ad0-4c4c-9666-fe6a9317158e); Time taken: 0.016 seconds
NFO : Executing command(queryId=hive_20230513103558_bae91e96-9ad0-4c4c-9666-fe6a9317158e): show tables
NFO : Starting task [Stage-0:DDL] in serial mode
NFO : Completed executing command(queryId=hive_20230513103558_bae91e96-9ad0-4c4c-9666-fe6a9317158e); Time taken: 0.007 seconds
NFO : OK
-----+
tab_name |
-----+
pegawaiparq |
-----+
row selected (0.037 seconds)
: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Load data pada table .parquet

```
root@sandbox-hdp:~ - Shell In A
Box
localhost4200

jdbcd.HiveConnection: Connected to sandbox-hdp.hortonworks.com:10000
Connected to: Apache Hive (version 3.1.0.3.0.1.0-187)
Driver: Hive JDBC (version 3.1.0.3.0.1.0-187)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Hive version 3.1.0.3.0.1.0-187 by Apache Hive
jdbcd:hive2://sandbox-hdp.hortonworks.com:2> LOAD DATA INPATH '/user/hive/sample_data.parquet' OVERWRITE INTO TABLE pegawaiparq
.....>;
Error: Error while compiling statement: FAILED: SemanticException [Error 10001]: Line 1:71 Table not found 'pegawaiparq' (state=42S02,cod
=10001)
jdbcd:hive2://sandbox-hdp.hortonworks.com:2> use penugasan9;
NFO : Compiling command(queryId=hive_20230513104729_4430fc45-0b24-4312-8288-7fd17279c80e): use penugasan9
NFO : Semantic Analysis Completed (retrial = false)
NFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
NFO : Completed compiling command(queryId=hive_20230513104729_4430fc45-0b24-4312-8288-7fd17279c80e); Time taken: 0.014 seconds
NFO : Executing command(queryId=hive_20230513104729_4430fc45-0b24-4312-8288-7fd17279c80e): use penugasan9
NFO : Starting task [Stage-0:DDL] in serial mode
NFO : Completed executing command(queryId=hive_20230513104729_4430fc45-0b24-4312-8288-7fd17279c80e); Time taken: 0.013 seconds
NFO : OK
0 rows affected (0.057 seconds)
jdbcd:hive2://sandbox-hdp.hortonworks.com:2> show tables;
NFO : Compiling command(queryId=hive_20230513104733_75c84464-8016-4f97-af9a-9aaa339262e6): show tables
NFO : Semantic Analysis Completed (retrial = false)
NFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
NFO : Completed compiling command(queryId=hive_20230513104733_75c84464-8016-4f97-af9a-9aaa339262e6); Time taken: 0.015 seconds
NFO : Executing command(queryId=hive_20230513104733_75c84464-8016-4f97-af9a-9aaa339262e6): show tables
NFO : Starting task [Stage-0:DDL] in serial mode
NFO : Completed executing command(queryId=hive_20230513104733_75c84464-8016-4f97-af9a-9aaa339262e6); Time taken: 0.006 seconds
NFO : OK
```

Select * pada table data source format parquet

```
root@sandbox-hdp:~ - Shell In A
Box
localhost4200

INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:pegawaiparq.full_name, type:varchar(100), comment:from deserializer), FieldSchema(name:pegawaiparq.department, type:varchar(100), comment:null), FieldSchema(name:pegawaiparq.gender, type:varchar(100), comment:null), FieldSchema(name:pegawaiparq.employee_id, type:varchar(100), comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20230513104750_7-9d85-9239ec019e34): select * from pegawaiparq
INFO : Executing command(queryId=hive_20230513104750_7-9d85-9239ec019e34): select * from pegawaiparq
INFO : Completed executing command(queryId=hive_20230513104750_7-9d85-9239ec019e34); Time taken: 0.009 seconds
INFO : OK

+-----+-----+-----+-----+-----+-----+-----+
| pegawaiparq.employee_id | pegawaiparq.full_name | pegawaiparq.job_title | pegawaiparq.department | pegawaiparq.business_unit | pegawaiparq.gender |
+-----+-----+-----+-----+-----+-----+-----+
| NULL | Asian | Kai Le | 47 | Controls Engineer | Engineering | Manufacturing | Mal
| NULL | Asian | Robert Patel | 58 | Analyst | Sales | Corporate | Mal
| NULL | Asian | Cameron Lo | 34 | Network Administrator | IT | Research & Development | Mal
| NULL | Latino | Harper Castillo | 39 | IT Systems Architect | IT | Corporate | Fem
| NULL | Latino | Harper Dominguez | 42 | Director | Engineering | Corporate | Fem
| NULL | Asian | Ezra Vu | 62 | Network Administrator | IT | Manufacturing | Mal
| NULL | Asian | Jade Hu | 58 | Sr. Analyst | Accounting | Specialty Products | Fem
```

g. Sample_data.avro

Membuat tabel untuk format .avro

```
root@sandbox-hdp:~ - Shell In A
Box
localhost4200

Sophia Gutierrez | Manager | Accounting | Specialty Products | Fem

+-----+
!0 rows selected (0.945 seconds)
): jdbc:hive2://sandbox-hdp.hortonworks.com:2> CREATE TABLE pegawaiavro
. . . . .
. . . . . ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
. . . . .
. . . . . STORED AS INPUTFORMAT 'org.apache.hadoop.hive ql.io.avro.AvroContainerInputFormat'
. . . . .
. . . . . OUTPUTFORMAT 'org.apache.hadoop.hive ql.io.avro.AvroContainerOutputFormat'
. . . . .
. . . . . TBLPROPERTIES ('avro.schema.literal'='{"type":"record","name":"Record","fields":[{"name":"Full_Name","type":"string"}, {"name":"Job_Title","type":"string"}, {"name":"Department","type":"string"}, {"name":"Business_Unit","type":"string"}, {"name":"Gender","type":"string"}, {"name":"Ethnicity","type":"string"}, {"name":"Age","type":"long"}]}');
INFO : Compiling command(queryId=hive_20230513105018_2c3ff219-2597-4f53-aaac-5596c6d37bc1): CREATE TABLE pegawaiavro
:OW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
:STORED AS INPUTFORMAT 'org.apache.hadoop.hive ql.io.avro.AvroContainerInputFormat'
:UTPUTFORMAT 'org.apache.hadoop.hive ql.io.avro.AvroContainerOutputFormat'
:BLPROPERTIES ('avro.schema.literal'='{"type":"record","name":"Record","fields":[{"name":"Full_Name","type":"string"}, {"name":"Job_Title","type":"string"}, {"name":"Department","type":"string"}, {"name":"Business_Unit","type":"string"}, {"name":"Gender","type":"string"}, {"name":"Ethnicity","type":"string"}, {"name":"Age","type":"long"}]}')
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20230513105018_2c3ff219-2597-4f53-aaac-5596c6d37bc1); Time taken: 0.037 seconds
INFO : Executing command(queryId=hive_20230513105018_2c3ff219-2597-4f53-aaac-5596c6d37bc1): CREATE TABLE pegawaiavro
:OW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
:STORED AS INPUTFORMAT 'org.apache.hadoop.hive ql.io.avro.AvroContainerInputFormat'
```

```
root@sandbox-hdp:~ - Shell In A
Box
localhost4200

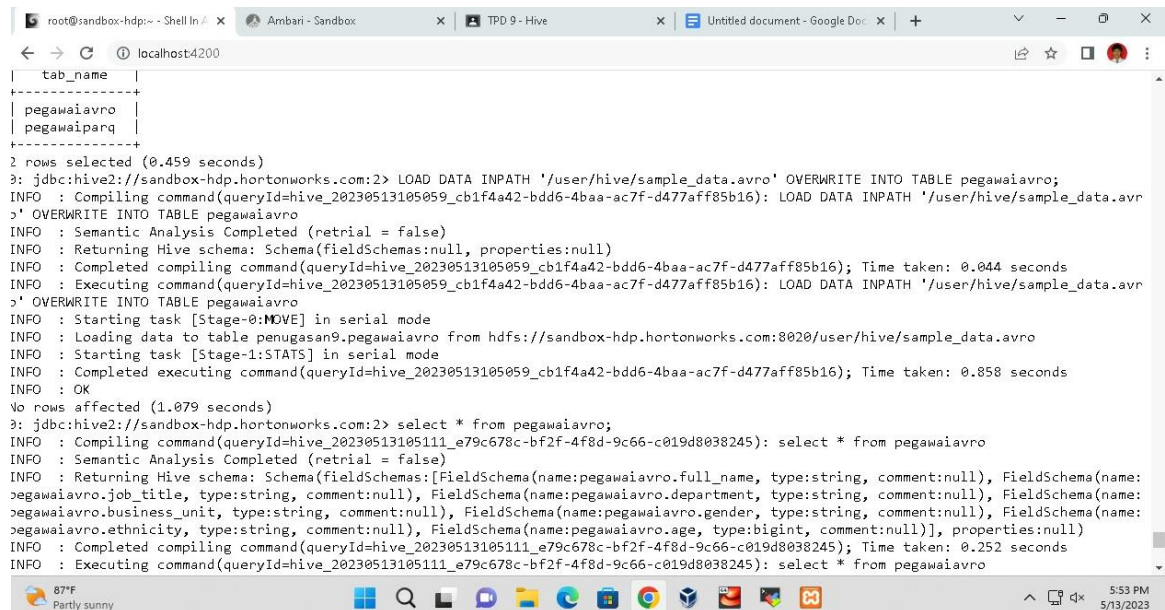
INFO : Executing command(queryId=hive_20230513105025_da8e5cef-c074-4f01-b003-9dfec59cdce0): show tables
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230513105025_da8e5cef-c074-4f01-b003-9dfec59cdce0); Time taken: 0.041 seconds
INFO : OK

+-----+
tab_name |
+-----+
pegawaiavro |
pegawaiparq |
+-----+

! rows selected (0.459 seconds)
): jdbc:hive2://sandbox-hdp.hortonworks.com:2> LOAD DATA INPATH '/user/hive/sample_data.avro' OVERWRITE INTO TABLE pegawaiavro;
INFO : Compiling command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16): LOAD DATA INPATH '/user/hive/sample_data.avr
)' OVERWRITE INTO TABLE pegawaiavro
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16); Time taken: 0.044 seconds
INFO : Executing command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16): LOAD DATA INPATH '/user/hive/sample_data.avr
)' OVERWRITE INTO TABLE pegawaiavro
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table penugasan9.pegawaiavro from hdfs://sandbox-hdp.hortonworks.com:8020/user/hive/sample_data.avro
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16); Time taken: 0.858 seconds
INFO : OK

!0 rows affected (1.079 seconds)
): jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from pegawaiavro;
INFO : Compiling command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245): select * from pegawaiavro
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name=pegawaiavro.full name, type:string, comment:null), FieldSchema(name=
```


Load data ke dalam tabel



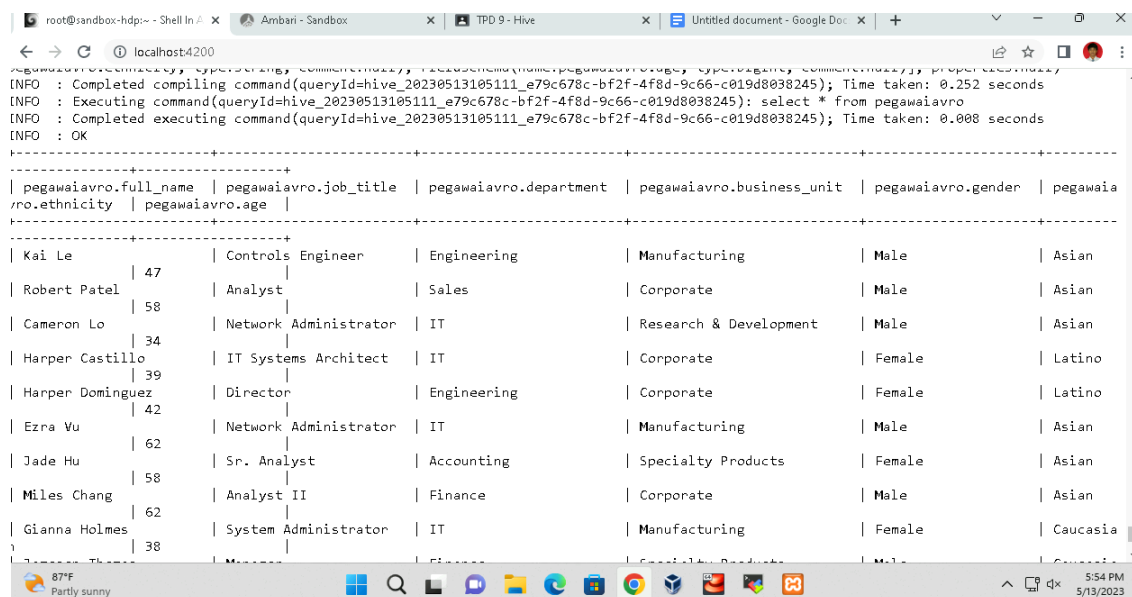
```
root@sandbox-hdp:~ - Shell In A... Ambari - Sandbox TPD 9 - Hive Untitled document - Google Doc +
localhost4200
+-----+
| tab_name |
+-----+
| pegawaiavro |
| pegawaiiparq |
+-----+

2 rows selected (0.459 seconds)

3: jdbc:hive2://sandbox-hdp.hortonworks.com:2> LOAD DATA INPATH '/user/hive/sample_data.avro' OVERWRITE INTO TABLE pegawaiavro;
INFO : Compiling command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16): LOAD DATA INPATH '/user/hive/sample_data.avro' OVERWRITE INTO TABLE pegawaiavro
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16); Time taken: 0.044 seconds
INFO : Executing command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16): LOAD DATA INPATH '/user/hive/sample_data.avro' OVERWRITE INTO TABLE pegawaiavro
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table penugasan9.pegawaiavro from hdfs://sandbox-hdp.hortonworks.com:8020/user/hive/sample_data.avro
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryId=hive_20230513105059_cb1f4a42-bdd6-4baa-ac7f-d477aff85b16); Time taken: 0.858 seconds
INFO : OK
No rows affected (1.079 seconds)

3: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from pegawaiavro;
INFO : Compiling command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245): select * from pegawaiavro
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:pegawaiavro.full_name, type:string, comment:null), FieldSchema(name:pegawaiavro.job_title, type:string, comment:null), FieldSchema(name:pegawaiavro.department, type:string, comment:null), FieldSchema(name:pegawaiavro.business_unit, type:string, comment:null), FieldSchema(name:pegawaiavro.gender, type:string, comment:null), FieldSchema(name:pegawaiavro.ethnicity, type:string, comment:null), FieldSchema(name:pegawaiavro.age, type:bigint, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245); Time taken: 0.252 seconds
INFO : Executing command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245): select * from pegawaiavro
```

Select all pada table dengan format avro



```
root@sandbox-hdp:~ - Shell In A... Ambari - Sandbox TPD 9 - Hive Untitled document - Google Doc +
localhost4200
pegawaiavro.full_name, type:string, comment:null), FieldSchema(name:pegawaiavro.job_title, type:string, comment:null), FieldSchema(name:pegawaiavro.department, type:string, comment:null), FieldSchema(name:pegawaiavro.business_unit, type:string, comment:null), FieldSchema(name:pegawaiavro.gender, type:string, comment:null), FieldSchema(name:pegawaiavro.ethnicity, type:string, comment:null), FieldSchema(name:pegawaiavro.age, type:bigint, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245); Time taken: 0.252 seconds
INFO : Executing command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245): select * from pegawaiavro
INFO : Completed executing command(queryId=hive_20230513105111_e79c678c-bf2f-4f8d-9c66-c019d8038245); Time taken: 0.008 seconds
INFO : OK
+-----+
| pegawaiavro.full_name | pegawaiavro.job_title | pegawaiavro.department | pegawaiavro.business_unit | pegawaiavro.gender | pegawaiavro.ethnicity | pegawaiavro.age |
+-----+
| Kai Le | 47 | Controls Engineer | Engineering | Manufacturing | Male | Asian |
| Robert Patel | 58 | Analyst | Sales | Corporate | Male | Asian |
| Cameron Lo | 34 | Network Administrator | IT | Research & Development | Male | Asian |
| Harper Castillo | 39 | IT Systems Architect | IT | Corporate | Female | Latino |
| Harper Dominguez | 42 | Director | Engineering | Corporate | Female | Latino |
| Ezra Vu | 62 | Network Administrator | IT | Manufacturing | Male | Asian |
| Jade Hu | 58 | Sr. Analyst | Accounting | Specialty Products | Female | Asian |
| Miles Chang | 62 | Analyst II | Finance | Corporate | Male | Asian |
| Gianna Holmes | 38 | System Administrator | IT | Manufacturing | Female | Caucasian |
```

h. Manipulasi database Cluster by department

```
select full_name, job_title, department from pegawaiavro
cluster by department;
```

```
root@sandbox-hdp:~ - Shell In A # Ambari - Sandbox TPD 9 - Hive Untitled document - Google Doc +
localhost4200
INFO : Completed executing command(queryId=hive_20230513144318_2f5833fc-f362-4324-820e-1bb2d353ddef); Time taken: 35.036 seconds
INFO : OK

-----+-----+-----+
full_name | job_title | department |
-----+-----+-----+
Jade Hu | Sr. Analyst | Accounting |
Sophia Gutierrez | Manager | Accounting |
Kai Le | Controls Engineer | Engineering |
Harper Dominguez | Director | Engineering |
Bella Tran | Vice President | Engineering |
Bella Wu | Sr. Analyst | Finance |
Miles Chang | Analyst II | Finance |
Jameson Thomas | Manager | Finance |
Jacob Moore | Sr. Manager | Marketing |
Lucas Richardson | Manager | Marketing |
Ezra Vu | Network Administrator | IT |
Cameron Lo | Network Administrator | IT |
Gianna Holmes | System Administrator | IT |
Jose Wong | Director | IT |
Luna Lu | IT Systems Architect | IT |
Jameson Pena | Systems Analyst | IT |
Harper Castillo | IT Systems Architect | IT |
Jordan Kumar | Service Desk Analyst | IT |
Robert Patel | Analyst | IT |
Ivy Chau | Analyst | IT |
-----+-----+-----+
30 rows selected (35.533 seconds)
j: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Create view berdasarkan departement

create view depart_it as

```
select full_name, job_title from pegawaiiparq where
department='IT';
```

```
root@sandbox-hdp:~ - Shell In A # Ambari - Sandbox TPD 9 - Hive Untitled document - Google Doc +
localhost4200
select full_name, job_title from pegawaiiparq where department='IT'
INFO : Starting task [Stage-1:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230513145416_d8d6f6b5-01b6-45f9-b947-a4c948c5e42b); Time taken: 0.103 seconds
INFO : OK
40 rows affected (0.568 seconds)
j: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from depart_it;
INFO : Compiling command(queryId=hive_20230513145622_e0e1fa9b-981b-409f-8c65-50a874a321fc): select * from depart_it
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:depart_it.full_name, type:varchar(100), comment:null), FieldSchema(name:depart_it.job_title, type:varchar(100), comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20230513145622_e0e1fa9b-981b-409f-8c65-50a874a321fc); Time taken: 0.225 seconds
INFO : Executing command(queryId=hive_20230513145622_e0e1fa9b-981b-409f-8c65-50a874a321fc): select * from depart_it
INFO : Completed executing command(queryId=hive_20230513145622_e0e1fa9b-981b-409f-8c65-50a874a321fc); Time taken: 0.044 seconds
INFO : OK

-----+-----+-----+
depart_it.full_name | depart_it.job_title |
-----+-----+-----+
Cameron Lo | Network Administrator |
Harper Castillo | IT Systems Architect |
Ezra Vu | Network Administrator |
Gianna Holmes | System Administrator |
Jameson Pena | Systems Analyst |
Jose Wong | Director |
Luna Lu | IT Systems Architect |
Jordan Kumar | Service Desk Analyst |
-----+-----+-----+
8 rows selected (0.503 seconds)
j: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Distributed by dan sortby

```
select full_name, job_title, department from pegawaiiavro
distribute by department sort by department asc;
```

root@sandbox-hdp:~ - Shell In / x

Ambari - Sandbox x

TPD 9 - Hive x

Untitled document - Google Doc x

+ x

localhost:4200

INFO : Completed executing command(queryId=hive_20230513145946_877caaa7-70af-4e41-b9be-12d0586a7d91); Time taken: 20.249 seconds
INFO : OK

full_name	job_title	department
Jade Hu	Sr. Analyst	Accounting
Sophia Gutierrez	Manager	Accounting
Kai Le	Controls Engineer	Engineering
Harper Dominguez	Director	Engineering
Bella Tran	Vice President	Engineering
Bella Wu	Sr. Analyst	Finance
Miles Chang	Analyst II	Finance
Jameson Thomas	Manager	Finance
Jacob Moore	Sr. Manager	Marketing
Lucas Richardson	Manager	Marketing
Ezra Vu	Network Administrator	IT
Cameron Lo	Network Administrator	IT
Gianna Holmes	System Administrator	IT
Jose Wong	Director	IT
Luna Lu	IT Systems Architect	IT
Jameson Pena	Systems Analyst	IT
Harper Castillo	IT Systems Architect	IT
Jordan Kumar	Service Desk Analyst	IT
Robert Patel	Analyst	Sales
Ivy Chau	Analyst	Sales

0 rows selected (20.499 seconds)

: jdbc:hive2:///sandbox-hdp.hortonworks.com:2>

79°F
Mostly cloudy

10:00 PM
5/13/2023

