EXP: 5 Create tables in Hive and write queries to access the data in the table

AIM:

To create tables in Hive and write queries to access the data in the table.

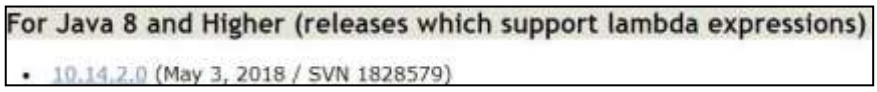
PROCEDURE:

Hive Download and installation:

1. Hive Installation setup:

- Download and install Apache Derby version 10.14.2.0:

https://db.apache.org/derby/derby_downloads.html#For+Java+8+and+Higher

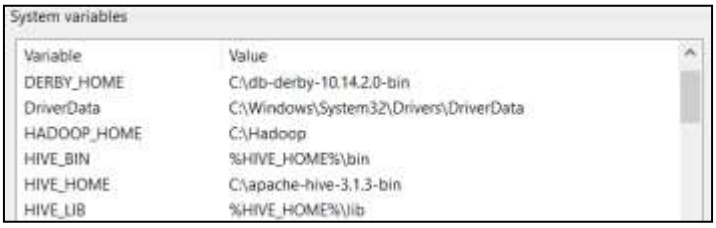


-Download and install Apache Hive version 3.1.3:

https://downloads.apache.org/hive/hive-3.1.3/

2. Add environment variables:

Environment variables > System variables > Add the below paths



-> (Inside Path)

3. Copy Derby libraries

Go to the Derby libraries directory (db-derby-10.14.2.0\lib) and copy all *.jar files. Then, paste them within the Hive libraries directory.

4. Configuring hive-site.xml and Hive's Bin folder:

Refer following link to download the file. Also download the guava file. Put hive-site.xml file to hive's conf location and replace hive's current guava file with this one in lib location. Also download the bin folder from link and replace the existing hive's bin folder.

https://1drv.ms/f/s!ArSg3Xpur4Grmw0SDqW0g44T7HYU?e=wDsoBn

5. Starting Hadoop Services

Open PowerShell as administrator and go to Hadoop sbin directory and start hadoop services using the following commands:

```
start-dfs.cmd
```

```
start-yarn.cmd
```

6. Derby Network Server:

Open another PowerShell window and run the following command to open Derby:

```
StartNetworkServer -h 0.0.0.0
```

Go to first PowerShell window and check whether NetworkServerControl is running.

7. Starting Apache Hive:

Go to Apache Hive's bin location with cd command and run the following command:

```
hive --service schematool -dbType derby –initSchema
```

8. Open Hive shell by typing:

```
hive
```

Create a Database:

Start by creating a database. Open the Hive CLI and follow the steps below:

1. Use the **CREATE DATABASE** statement to create a new database:

```
CREATE DATABASE financials;
```

2. Verify the database is present:

```
SHOW DATABASES;
```

3. Switch to the new database:

Switch to the newly created database:

```
USE financials;
```

Create a Table in Hive:

Create a simple table in your database:

```
CREATE TABLE students_table (id INT, name STRING);
```

Insert values into the table:

You can insert sample data into the table:

```
INSERT INTO finance_table VALUES (1, 'Alice'), (2,'Bob'), (3, 'Charlie');
```

Query your data:

Use SQL-like queries to retrieve data from your table:

```
CREATE VIEW myview AS SELECT name, id FROM finance_table;
```

List Hive Tables and Data:

To show all tables in a selected database, use the following statement:

```
SHOW TABLES;
```

To show table column names and data types, run:

```
DESCRIBE finance_table;
```

To display table data, use a **SELECT** statement. For example, to select everything in a table, run:

```
SELECT*FROM  myview;
```

Alter a Table:

You can alter the table structure by adding a new column:

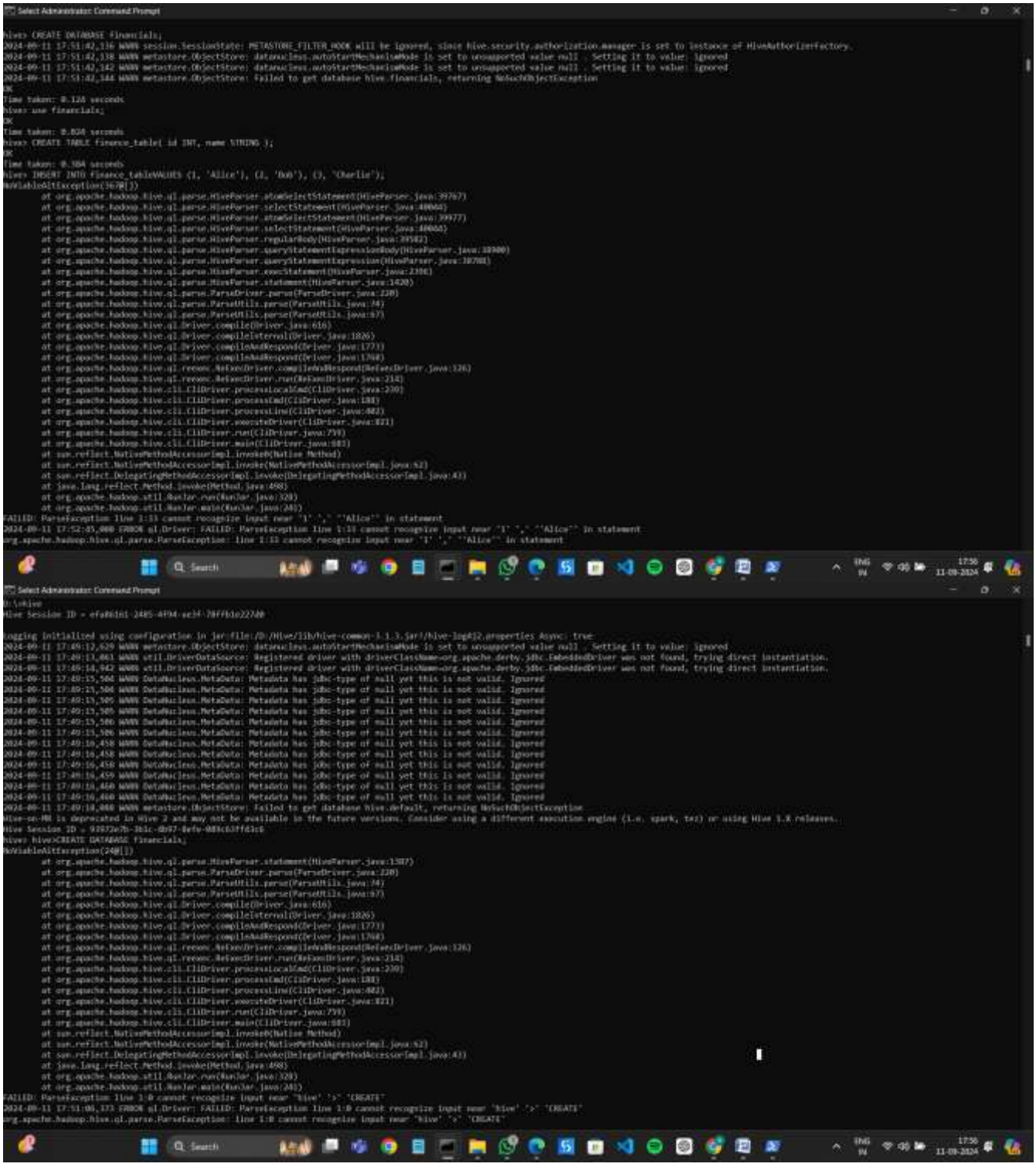
```
ALTER TABLE finance_table ADD COLUMNS (age INT);
```

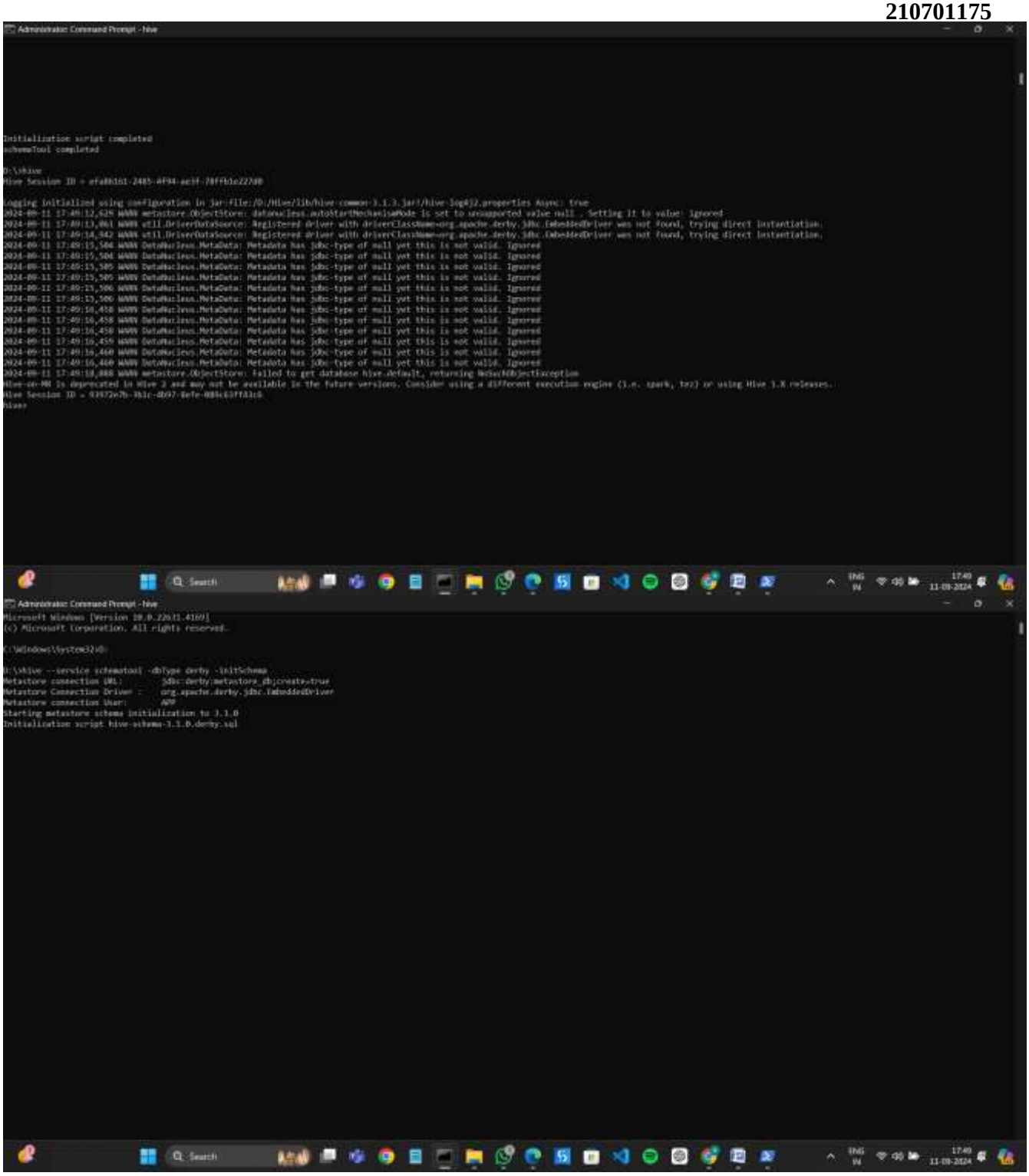
Quit Hive:

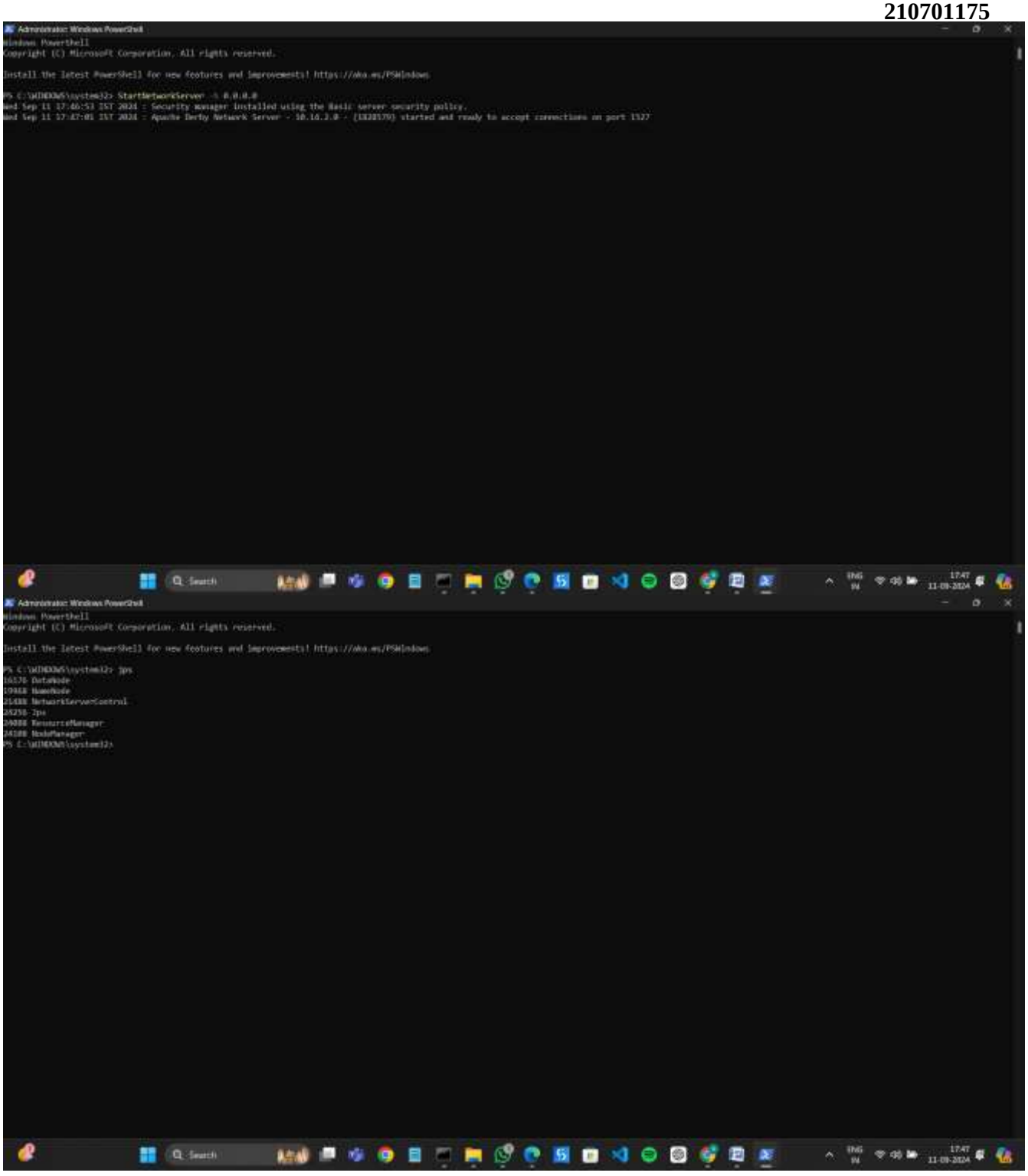
To exit the Hive CLI, simply type:

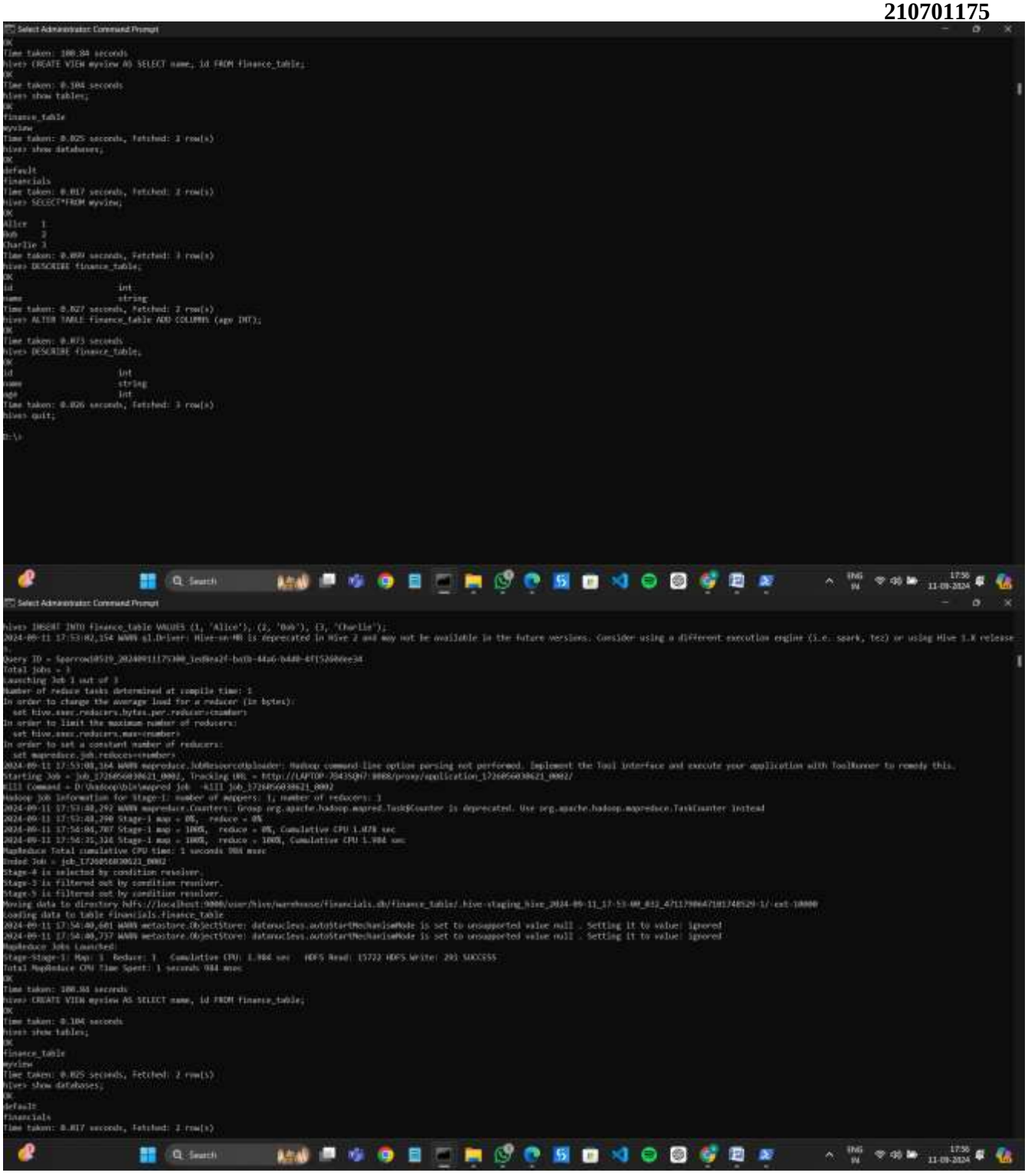
```
quit;
```

OUTPUT SCREENSHOTS:




Administrator: Command Prompt - hive
Initialization script completed
schemaTool completed
D:\hive
Hive Session ID = efa88161-2485-4f94-ac1f-78ffb1e22f40
Logging initialized using configuration in jar:file:/D:/Hive/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
2024-09-11 17:49:12,629 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-09-11 17:49:13,861 WARN util.DriverDataSource: Registered driver with driverClassName=org.apache.derby.jdbc.EmbeddedDriver was not found, trying direct instantiation.
2024-09-11 17:49:14,942 WARN util.DriverDataSource: Registered driver with driverClassName=org.apache.derby.jdbc.EmbeddedDriver was not found, trying direct instantiation.
2024-09-11 17:49:15,504 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:15,504 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:15,505 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:15,505 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:15,506 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:15,506 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,458 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,458 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,458 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,459 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,460 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:16,460 WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2024-09-11 17:49:18,888 WARN metastore.ObjectStore: Failed to get database hive.default, returning NoSuchObjectException
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hive Session ID = 93972e7b-3b1c-4b97-8efe-089c63ff83c8
hive>
Search
ENG
IN
17:49
11-09-2024
Administrator: Command Prompt - hive
Microsoft Windows [Version 10.0.22631.4169]
(c) Microsoft Corporation. All rights reserved.
C:\Windows\System32>D:
D:\hive --service schematool -dbType derby -initSchema
Metastore connection URL: jdbc:derby:metastore_db;create=true
Metastore Connection Driver :   org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User: APP
Starting metastore schema initialization to 3.1.0
Initialization script hive-schema-3.1.0.derby.sql
Search
ENG
IN
17:49
11-09-2024


Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\WINNT\system32> StartNetworkServer -h 0.0.0.0
Wed Sep 11 17:46:53 IST 2024 : Security manager installed using the basic server security policy.
Wed Sep 11 17:47:01 IST 2024 : Apache Derby Network Server - 10.16.1.0 - (1328579) started and ready to accept connections on port 1527
Search
ENG
IN
17:47
11-09-2024
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\WINNT\system32> jps
16176 DataNode
19468 NameNode
21488 NetworkServerControl
26256 Jps
24008 ResourceManager
24108 NodeManager
PS C:\WINNT\system32>
Search
ENG
IN
17:47
11-09-2024



RESULT:

Thus, to create tables in Hive and write queries to access the data in the table was completed successfully.