

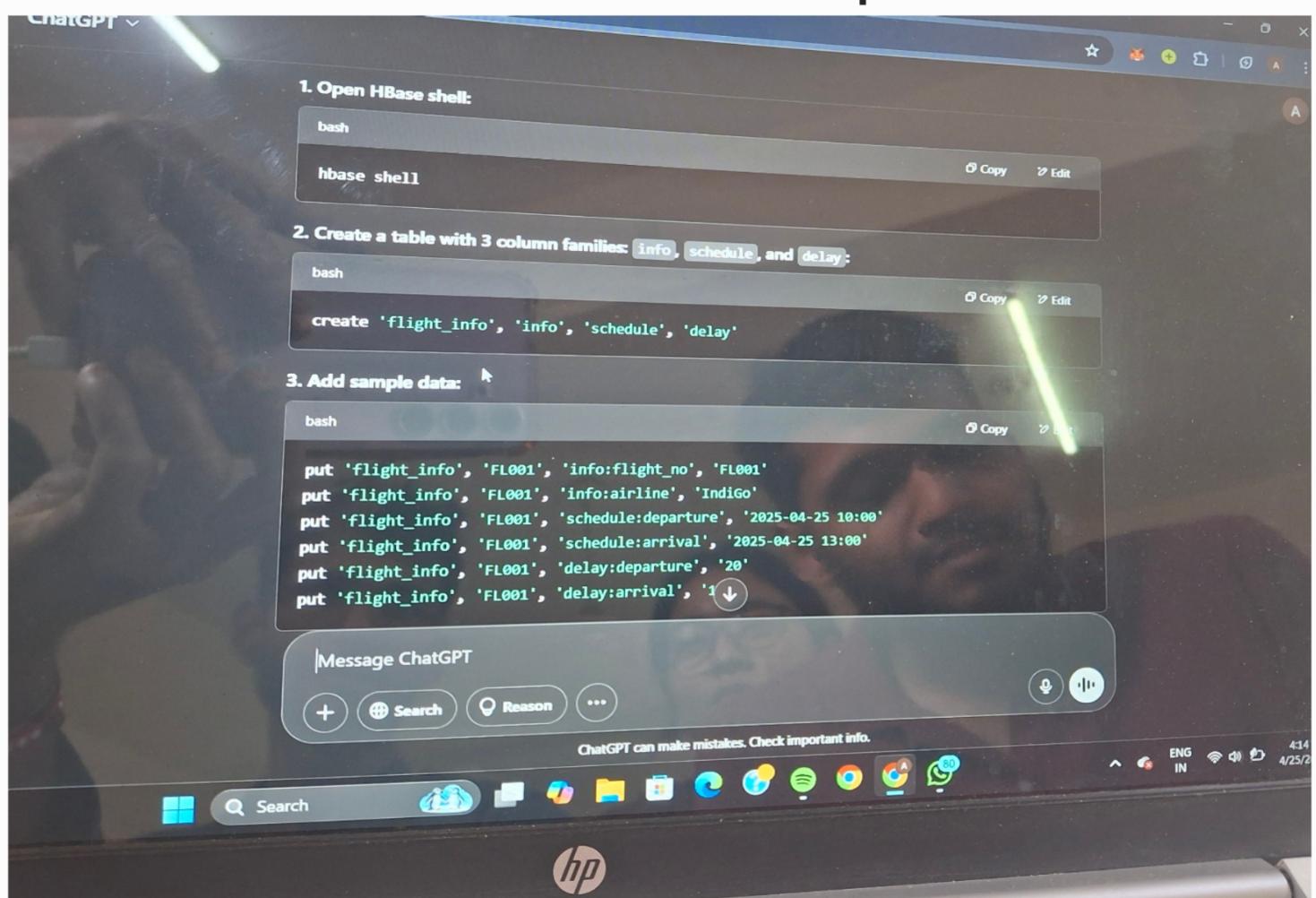
# 1. Open Terminal in Cloudera on VM Chk:

```
> hadoop version  
> hdfs version
```

Start

```
> hbase shell
```

# 2. Create Table & Add Sample Data



Here flight\_info is table

### 3. To chk content of table

> scan 'flight\_info'

```
=> Hbase::Table - flight_info
hbase(main):002:0> put 'flight_info', 'FL001', 'info:flight_no', 'FL001'
0 row(s) in 0.6600 seconds
hbase(main):003:0> put 'flight_info', 'FL001', 'info:airline', 'IndiGo'
0 row(s) in 0.0210 seconds
hbase(main):004:0> put 'flight_info', 'FL001', 'schedule:departure', '2025-04-25 10:00'
0 row(s) in 0.0300 seconds
hbase(main):005:0> put 'flight_info', 'FL001', 'schedule:arrival', '2025-04-25 13:00'
0 row(s) in 0.0070 seconds
hbase(main):006:0> put 'flight_info', 'FL001', 'delay:departure', '20'
0 row(s) in 0.0130 seconds
hbase(main):007:0> put 'flight_info', 'FL001', 'delay:arrival', '15'
0 row(s) in 0.0130 seconds
hbase(main):008:0> scan 'flight_info'
ROW
  FL001
  FL001
  FL001
  FL001
  FL001
  FL001
1 row(s) in 0.1020 seconds
hbase(main):009:0> put ■
cloudera@quickstart:~
```

### 4. Alter (adding status column in flight\_info)

> alter 'flight\_info', 'status'

### 5. To delete particular column(status)

>

```
FL001
FL001
FL001
FL001
FL001
2 row(s) in 0.0280 seconds
hbase(main):016:0> alter 'flight_info', 'status'
Updating all regions with the new schema...
0/1 regions updated.
1/1 regions updated.
Done.
0 row(s) in 3.1730 seconds
hbase(main):017:0> disable 'flight_info'
0 row(s) in 2.3600 seconds
hbase(main):018:0> alter 'flight_info', {NAME => 'status', METHOD=> 'delete'}
Updating all regions with the new schema...
1/1 regions updated.
Done.
0 row(s) in 2.2260 seconds
hbase(main):019:0> enable 'flight_info'
0 row(s) in 1.2940 seconds
hbase(main):020:0> exit
[cloudera@quickstart ~]$ hive■
cloudera@quickstart:~
```



Search



# Start HIVE

1.

```
> hive
```

2.

```
> CREATE EXTERNAL TABLE
```

```
flight_info_hive (  
rowkey STRING,  
flight_no STRING,  
airline STRING,  
departure STRING,  
arrival STRING,  
dep_delay INT,  
arr_delay INT
```

```
)
```

```
STORED BY
```

```
'org.apache.hadoop.hive.hbase.HBaseSt  
orageHandler'
```

```
WITH SERDEPROPERTIES (
```

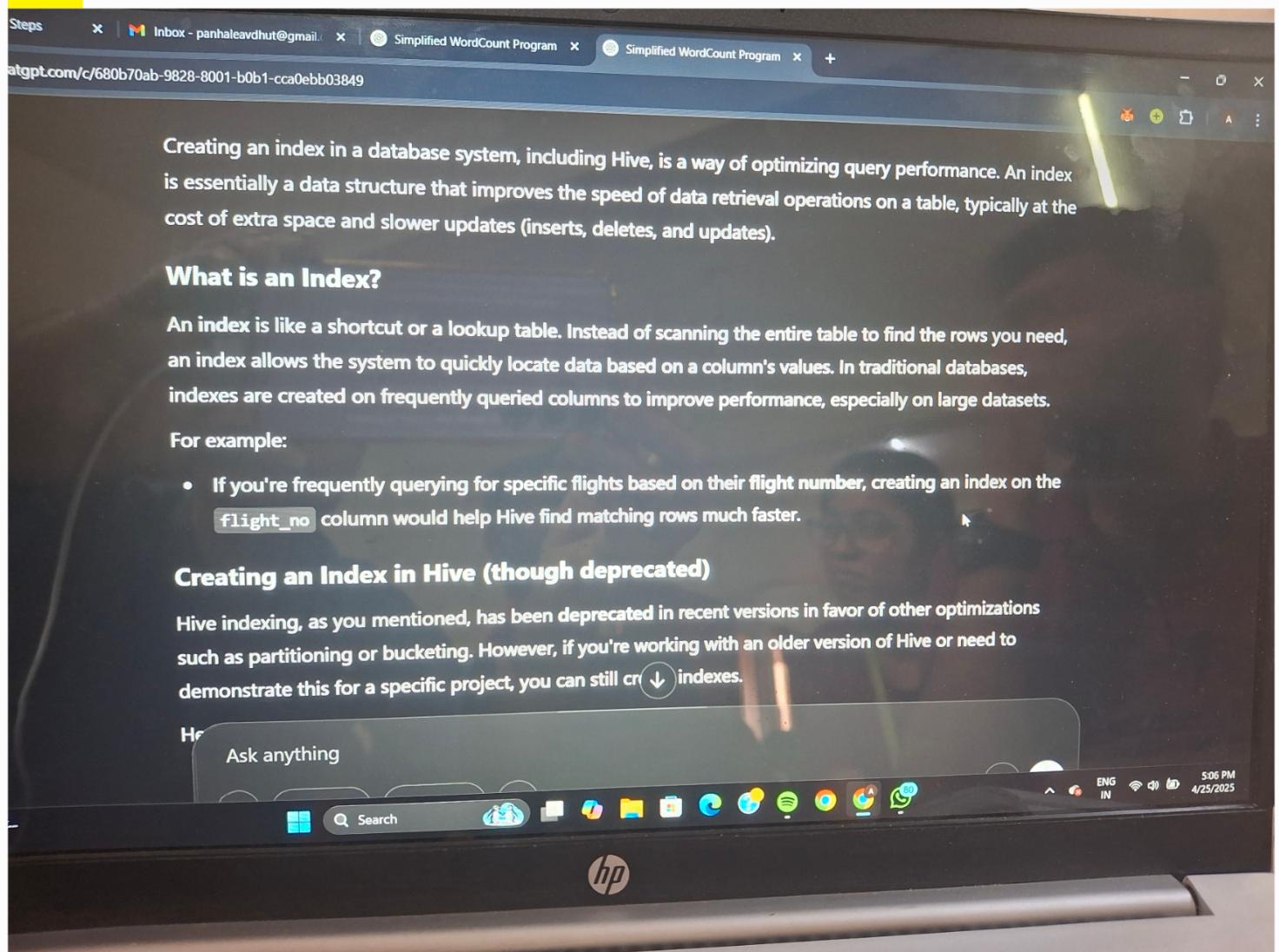
```
"hbase.columns.mapping" =  
":key,info:flight_no,info:airline,schedule:d  
eparture,schedule:arrival,delay:departure,  
delay:arrival"  
)  
TBLPROPERTIES ("hbase.table.name" =  
flight_info");
```

### 3. Find Total Departure Delay in HIVE Find Average Departure Delay in HIVE

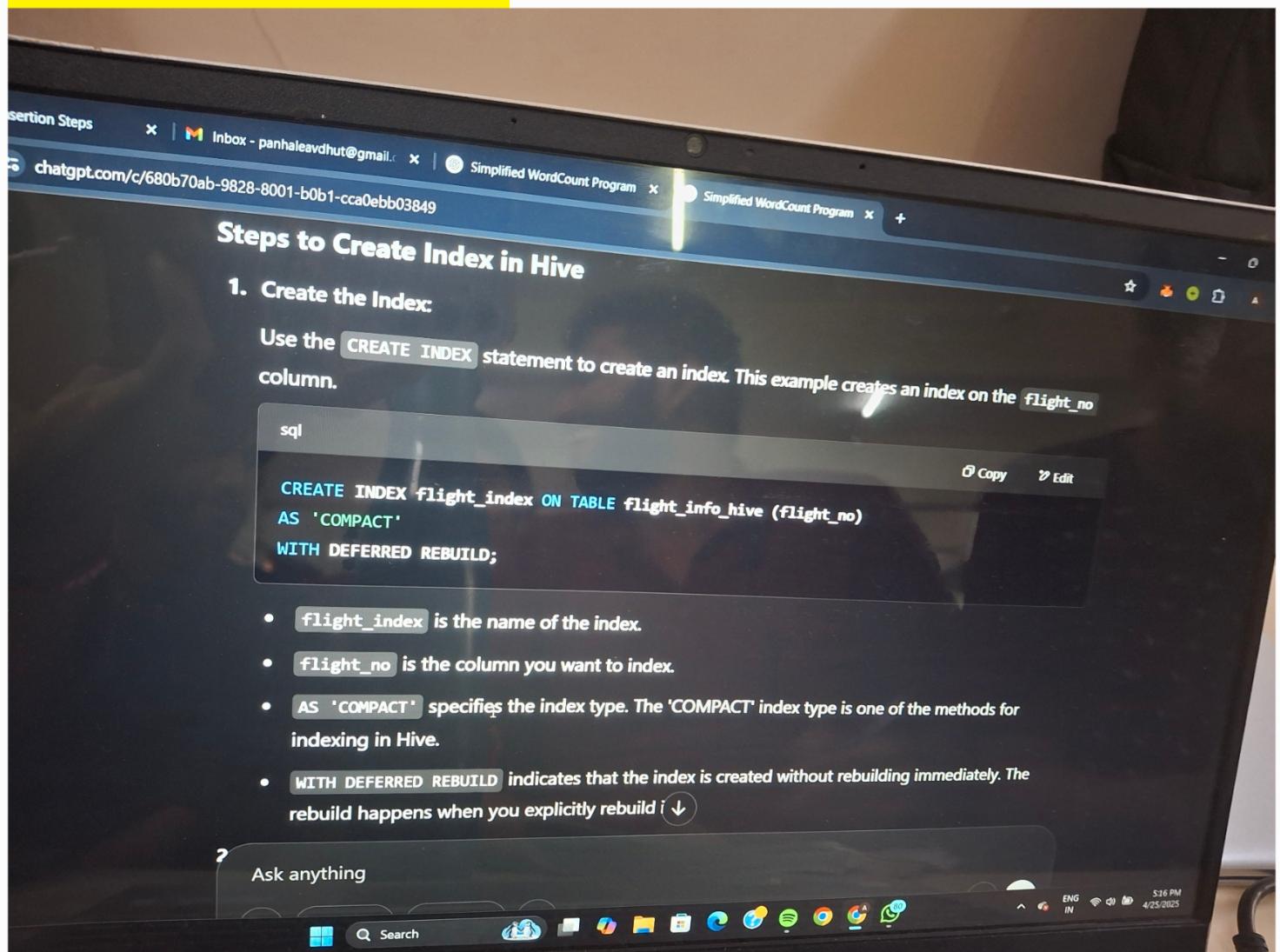
The screenshot shows a terminal window with three distinct sections, each containing an SQL query:

- d. Find Total Departure Delay in Hive**  
sql  
SELECT SUM(dep\_delay) AS total\_departure\_delay FROM flight\_info\_hive;
- e. Find Average Departure Delay in Hive**  
sql  
SELECT AVG(dep\_delay) AS average\_departure\_delay FROM flight\_info\_hive;
- f. Create Index on Flight Information Table in Hive**  
Ask anything  
+ Search Reason ...  
ChatGPT can make mistakes. Check important info.  
Search

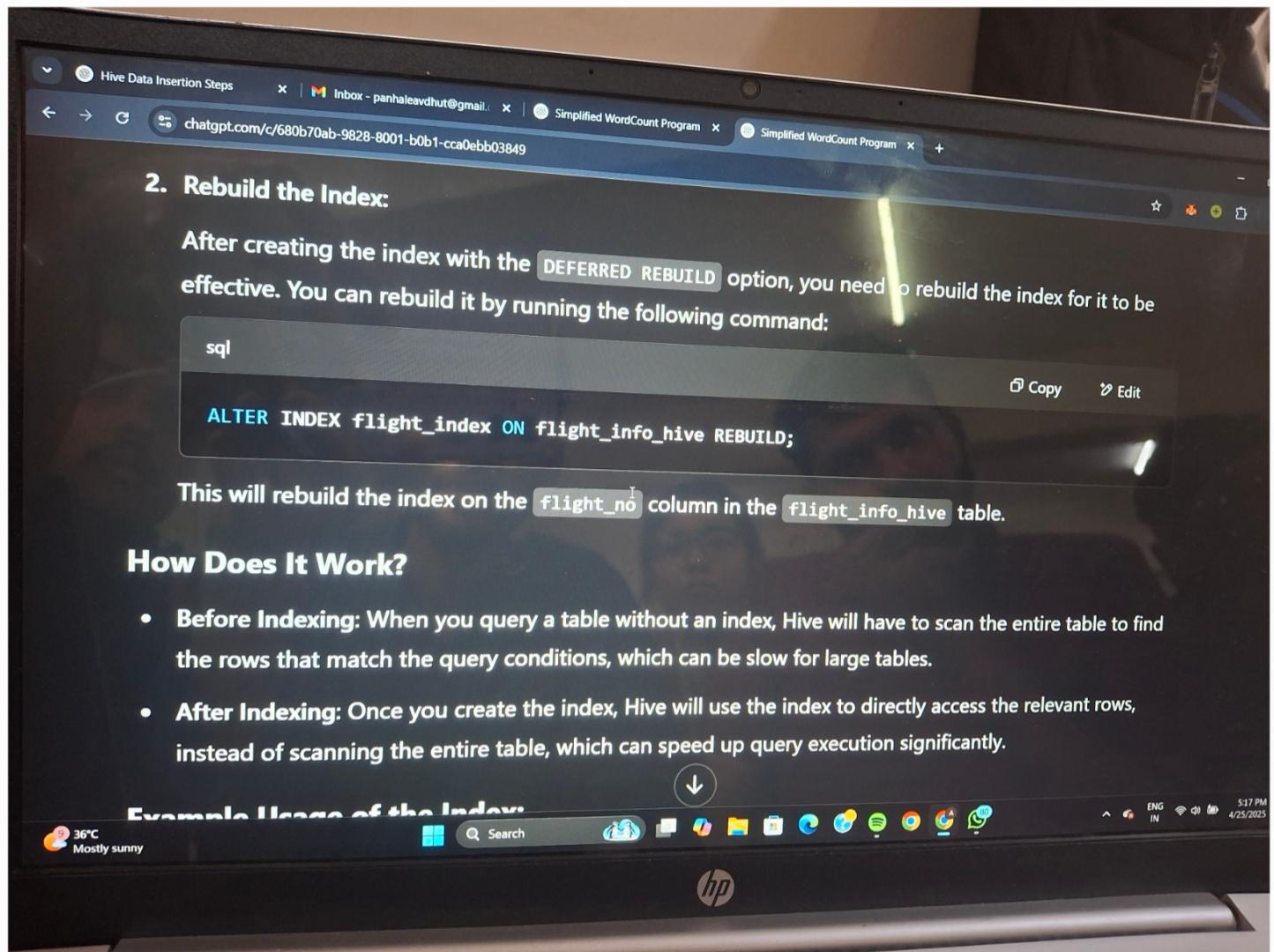
# 4. INDEX ON Flight\_info in HIVE



## 4.1 Create Index



## 4.2 Rebuild the index



## 5. Select \* from flight\_info;

### Restart HBase Shell Commands:

