

## Activity 1:

Tạo folder “day5” và “data” trong day 4 trên hệ thống HDFS và Sandbox

- Trong folder big\_data: mkdir day5, mkdir /day4/data
- Hệ thống HDFS: `hadoop fs -mkdir /user/20010801/day5`, `hadoop fs -mkdir /user/20010801/day5/data`

Copyfile u.data, u.item, u.user vào folder “data” Sandbox, HDFS

- **Sandbox:** Trong folder ml-100k:
  - o `cp big_data/ml-100k/u.data day5/data/`
  - o `cp big_data/ml-100k/u.item day5/data/`
  - o `cp big_data/ml-100k/u.user day5/data/`
- **HDFS:** Trong folder ml-100k
  - o `hadoop fs -put u.data /user/20010801/day5/data`
  - o `hadoop fs -put u.item /user/20010801/day5/data`
  - o `hadoop fs -put u.user /user/20010801/day5/data`

## Activity 2:

### Ex1:

Vào sandbox nhập lệnh vào hive: **hive**

Kiểm tra database: “**SHOW DATABASES;**”

Tạo cơ sở dữ liệu: “**CREATE DATABASE Le\_Hoang\_Ngoc\_Tu;**”

Vào Cơ sở dữ liệu: **USE Le\_Hoàng\_Ngoc\_Tu;**

Tạo 1 table “**nhanvien**” có các trường dữ liệu như sau

col_name	data_type	comment
id	int	
ten	varchar(64)	
tuoi	int	
diachi	varchar(64)	
luong	int	

“**CREATE TABLE nhanvien (Id int, ten varchar(64), tuoi int, diachi varchar(64), luong int);**”

Cập nhật các trường dữ liệu ta được:

```
mysql> SELECT * FROM NhanVien;
```

id	ten	tuoi	diachi	luong
1	Thanh	21	HaNoi	12000000
2	Loan	25	HaNoi	15000000
3	Nga	23	HaNam	5000000
4	Manh	25	HaiPhong	12000000
5	Chien	27	HaNoi	12000000
5	Cao	22	HCM	7000000
5	Lan	24	HaNoi	18000000

```
7 rows in set (0.00 sec)
```

Chỉnh sửa thông tin cột id không bị trùng lặp và sắp xếp thứ tự tăng dần

- **UPDATE nhanvien SET id = 6 WHERE ten = 'Cao';**
- **UPDATE nhanvien SET id = 7 WHERE ten = 'Lan';**
- **UPDATE nhanvien SET luong = 10000000 WHERE id = 1;**

Sắp xếp thứ tự: “**SELECT \* FROM nhanvien ORDER BY id ASC;**”

```
INFO : OK
```

nhanvien.id	nhanvien.ten	nhanvien.tuoi	nhanvien.diachi	nhanvien.luong
1	Thanh	35	HaiPhong	10000000
2	Loan	25	HaNoi	15000000
3	Nga	23	HaNam	5000000
4	Manh	25	HaiPhong	12000000
5	Chien	27	HaNoi	12000000
6	Cao	22	HCM	7000000
7	Lam	24	HaNoi	18000000

Thêm vào table các trường thông tin:

- **INSERT INTO nhanvien (id, ten, tuoi, diachi, luong) VALUES (8, 'Hai', 29, 'HaNoi', 28000000);**

+ id 9, 10 tương tự như id 8

```
INFO : OK
```

nhanvien.id	nhanvien.ten	nhanvien.tuoi	nhanvien.diachi	nhanvien.luong
1	Thanh	35	HaiPhong	10000000
2	Loan	25	HaNoi	15000000
3	Nga	23	HaNam	5000000
4	Manh	25	HaiPhong	12000000
5	Chien	27	HaNoi	12000000
6	Cao	22	HCM	7000000
7	Lam	24	HaNoi	18000000
8	Hai	29	HaNoi	28000000
9	Minh	21	QuangNam	8000000
10	Tung	28	Hue	21000000

```
10 rows selected (0.717 seconds)
```

Tạo 1 view có tên là hanoi địa chỉ là ở hanoi:

- **CREATE VIEW hanoi AS SELECT id, ten, tuoi, lương FROM nhanvien WHERE diachi = 'HaNoi';**

```
INFO : OK
```

hanoi.id	hanoi.ten	hanoi.tuoi	hanoi.luong
2	Loan	25	15000000
5	Chien	27	12000000
7	Lam	24	18000000
8	Hai	29	28000000

1 row selected (0.618 seconds)

Tạo 1 view “avgluong” tính trung bình lương của tất cả nhân viên

- **CREATE VIEW avgluong AS SELECT AVG(luong) AS tb\_luongnv FROM nhanvien;**

```
INFO : OK
```

avgluong.tb_luongnv
1.36E7

1 row selected (5.253 seconds)

Tạo 1 view “luongcao” nhóm tất cả nhân viên có lương trên >= 15000000

- **CREATE VIEW luongcao AS SELECT \* FROM nhanvien WHERE lương >= 15000000;**

```
INFO : OK
```

luongcao.id	luongcao.ten	luongcao.tuoi	luongcao.diachi	luongcao.luong
2	Loan	25	HaNoi	15000000
7	Lam	24	HaNoi	18000000
8	Hai	29	HaNoi	28000000
10	Tung	28	Hue	21000000

Tạo 1 view “luong” => nhóm tất cả những nhân viên có chung địa chỉ và tính tổng thu nhập và trung bình thu nhập của mọi người:

- **CREATE VIEW luong AS SELECT diachi, SUM(luong) AS tong\_luong, AVG(luong) AS trung\_binh\_luong FROM nhanvien GROUP BY diachi;**

```
INFO : OK
```

luong.diachi	luong.tong_luong	luong.trung_binh_luong
HaNam	5000000	5000000.0
HaiPhong	22000000	1.1E7
Hue	21000000	2.1E7
HCM	7000000	7000000.0
HaNoi	73000000	1.825E7
QuangNam	8000000	8000000.0

6 rows selected (8.781 seconds)

Ex2:

**CREATE TABLE users:**

```
INFO : OK
```

users.idu	users.username
567	TrainingPuppy
687	BathroomMakeover
45	PoshNachos
15	SewingButtons

**CREATE TABLE views:** do em không gắn dc “id” với “by” nên thay thành “id\_v” và “by\_1”

```
INFO : Completed executing Command(queryId=1)
INFO : OK
```

views.id_v	views.by_1	views.track
1	5	55
2	5	55
3	67	55
4	6	29
5	125	2
6	5	698
7	5	698

**CREATE TABLE tracks:**

```
INFO : OK
```

tracks.id	tracks.uid	tracks.title
2	15	Worf is in the air
29	567	Stargold
55	567	No love liars
698	567	Lofe

JOIN thành Table theo yêu cầu: ..

### Activity 3:

**Ex1:**

Kiểm tra database: “ **SHOW DATABASES;** ”

Tạo cơ sở dữ liệu: “ **CREATE DATABASE Le\_Hoang\_Ngoc\_Tu;** ”

Vào Cơ sở dữ liệu: **USE Le\_Hoàng\_Ngoc\_Tu;**

Tạo 1 table “**NhanVien**” có các trường dữ liệu như sau

col_name	data_type	comment
id	int	
ten	varchar(64)	
tuoi	int	
diachi	varchar(64)	
luong	int	

**“CREATE TABLE NhanVien (Id int, ten varchar(64), tuoi int, diachi varchar(64), luong int);”**

Cập nhật các trường dữ liệu ta được:

```
mysql> SELECT * FROM NhanVien;
```

id	ten	tuoi	diachi	luong
1	Thanh	21	HaNoi	12000000
2	Loan	25	HaNoi	15000000
3	Nga	23	HaNam	5000000
4	Manh	25	HaiPhong	12000000
5	Chien	27	HaNoi	12000000
5	Cao	22	HCM	7000000
5	Lan	24	HaNoi	18000000

7 rows in set (0.00 sec)

Chỉnh sửa thông tin cột id không bị trùng lặp và sắp xếp thứ tự tăng dần

- **UPDATE NhanVien SET id = 6 WHERE ten = 'Cao';**
- **UPDATE NhanVien SET id = 7 WHERE ten = 'Lan';**
- **UPDATE NhanVien SET luong = 10000000 WHERE id = 1;**

Sắp xếp thứ tự: **“SELECT \* FROM NhanVien ORDER BY id ASC;”**

Thêm vào table các trường thông tin:

- **INSERT INTO NhanVien (id, ten, tuoi, diachi, luong) VALUES (8, 'Hai', 29, 'HaNoi', 28000000);**

+ id 9, 10 tương tự như id 8

```
mysql> SELECT * FROM NhanVien;
```

id	ten	tuoi	diachi	luong
1	Thanh	21	HaNoi	10000000
2	Loan	25	HaNoi	15000000
3	Nga	23	HaNam	5000000
4	Manh	25	HaiPhong	12000000
5	Chien	27	HaNoi	12000000
6	Cao	22	HCM	7000000
7	Lan	24	HaNoi	18000000
8	Hai	29	HaNoi	28000000
9	Minh	21	QuangNam	8000000
10	Tung	28	Hue	21000000

```
10 rows in set (0.00 sec)
```

Tạo 1 view có tên là hanoi địa chỉ là ở hanoi:

- **CREATE VIEW hanoi AS SELECT id, ten, tuoi, lương FROM NhanVien WHERE diachi = 'HaNoi';**

```
mysql> SELECT * FROM hanoi;
```

id	ten	tuoi	luong
1	Thanh	21	10000000
2	Loan	25	15000000
5	Chien	27	12000000
7	Lan	24	18000000
8	Hai	29	28000000

```
5 rows in set (0.00 sec)
```

Tạo 1 view “avgluong” tính trung bình lương của tất cả nhân viên

- **CREATE VIEW avgluong AS SELECT AVG(luong) AS tb\_luongnv FROM NhanVien;**

```
mysql> SELECT * FROM avgluong;
```

tb_luongnv
13600000.0000

```
1 row in set (0.01 sec)
```

Tạo 1 view “luongcao” nhóm tất cả nhân viên có lương trên >= 15000000

- **CREATE VIEW luongcao AS SELECT \* FROM NhanVien WHERE lương >= 15000000;**

```
mysql> SELECT * FROM luongcao;
```

id	ten	tuoi	diachi	luong
2	Loan	25	HaNoi	15000000
7	Lan	24	HaNoi	18000000
8	Hai	29	HaNoi	28000000
10	Tung	28	Hue	21000000

Tạo 1 view “luong” => nhóm tất cả những nhân viên có chung địa chỉ và tính tổng thu nhập và trung bình thu nhập của mọi người:

- **CREATE VIEW luong AS SELECT diachi, SUM(luong) AS tong\_luong, AVG(luong) AS trung\_binh\_luong FROM NhanVien GROUP BY diachi;**

```
mysql> SELECT * FROM luong;
```

diachi	tong_luong	trung_binh_luong
HaiPhong	12000000	12000000.0000
HaNam	5000000	5000000.0000
HaNoi	83000000	16600000.0000
HCM	7000000	7000000.0000
Hue	21000000	21000000.0000
QuangNam	8000000	8000000.0000

6 rows in set (0.01 sec)

Ex2:

**CREATE TABLE users:**

```
mysql> SELECT * FROM users;
```

idu	username
567	TrainingPuppy
687	BathroomMakeover
45	PoshNachos
15	SewingButtons

4 rows in set (0.00 sec)

**CREATE TABLE views:**

```
mysql> SELECT * FROM views;
```

id	by	track
1	5	55
2	5	55
3	67	55
4	6	29
5	125	2
6	5	698
7	5	698

7 rows in set (0.00 sec)

**CREATE TABLE tracks:**

```
mysql> SELECT * FROM tracks;
+-----+-----+-----+
| id  | uid  | title                |
+-----+-----+-----+
| 2   | 15   | Worf is in the air  |
| 29  | 567  | Stargold             |
| 55  | 567  | No love liars        |
| 698 | 567  | Lofe                 |
+-----+-----+-----+
4 rows in set (0.00 sec)
```

JOIN thành bảng theo yêu cầu