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)	i i
tuoi	int	i i
diachi	varchar(64)	j i
luong	int	i i

"CREATE TABLE nhanvien (Id int, ten varchar(64), tuoi int, diachi varchar(64), luong int);"

Cập nhật các trường dữ liệu ta được:

mysql>	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	0.00 sed	:)			

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 tuong tu như id 8

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

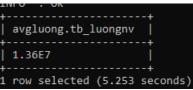
Tạo 1 view có tên là hanoi địa chỉ là ở hanoi:

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

INFO : OK	4		
hanoi.id	hanoi.ten	hanoi.tuoi	hanoi.luong
2 5	Loan Chien	25 27	15000000 12000000
7	Lam	24	18000000
8 +	Hai +	29 +	28000000 +

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;



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

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

INFO : OK				
luongcao.id	luongcao.ten	luongcao.tuoi	luongcao.diachi	luongcao.luong
2 7 8 10	Loan Lam Hai Tung	25 24 29 28	HaNoi HaNoi HaNoi Hue	15000000 18000000 28000000 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	<u>-</u>	++
	luong.tong_luong	luong.trung_binh_luong
HaNam HaiPhong Hue HCM HaNoi QuangNam	5000000 22000000 21000000 7000000 73000000 8000000	5000000.0 1.1E7 2.1E7 7000000.0 1.825E7 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 : OK	ten everneting (ommanu (quer y ru-11.
views.id_v	views.by_1	views.track
1 2 3 4 5	5 5 67 6 125	55 55 55 29
6 7 	5 5	698 698

CREATE TABLE tracks:

INFO : OK		.
tracks.id	tracks.uid	tracks.title
2 2 29	15 567	Worf is in the air Stargold
55 698	567 567	No love liers 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	1
ten	varchar(64)	İ
tuoi	int	Ì
diachi	varchar(64)	Ī
luong	int	1

"CREATE TABLE NhanVien (Id int, ten varchar(64), tuoi int, diachi varchar(64), luong int);"

Cập nhật các trường dữ liệu ta được:

mysql>	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	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 tuong tự như id 8

mysql> SELECT * FROM NhanVien;						
id	ten	tuoi	diachi	luong		
1 2	Thanh Loan	21 25	HaNoi HaNoi	10000000 15000000		
3	Nga Manh	23 25	HaNam HaiPhong	5000000 12000000		
5	Chien Cao	27 22	HaNoi HCM	12000000		
7	Lan	24	HaNoi	18000000		
8	Hai Minh	29 21	HaNoi QuangNam	28000000 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, luong FROM NhanVien WHERE diachi = 'HaNoi';

nysql> SELECT * FROM hanoi;					
id	ten	tuoi	luong		
 1 2 5	+ Thanh Loan Chien	21 25 27	10000000 15000000 12000000		
7 8	Lan Hai	24 29	18000000 28000000		
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;

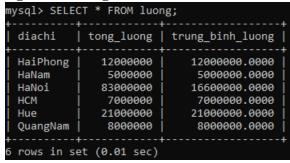
Tạo 1 view "luongcao" nhóm tất cả nhân viên có lương trên >= 15000000

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



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;



Ex2:

CREATE TABLE users:

```
mysql> SELECT * FROM users;

+----+

| idu | username |

+----+

| 567 | TrainingPuppy |

| 687 | BathroomMakeover |

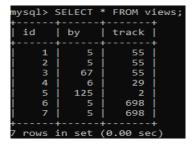
| 45 | PoshNachos |

| 15 | SewingButtons |

+----+

4 rows in set (0.00 sec)
```

CREATE TABLE views:



CREATE TABLE tracks:

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

JOIN thành bảng theo yêu cầu