**SECTION 1 : BASICS**

Created Table

**CREATE TABLE Migrants (**

**Web\_ID INT,**

**Incident\_Region STRING,**

**Reported\_Date STRING,**

**Reported\_Year INT,**

**Reported\_Month VARCHAR(5),**

**No\_Dead INT,**

**Estd\_No\_Missing INT,**

**Tot\_Dead\_Missing INT,**

**No\_Survivors INT,**

**No\_Females INT,**

**No\_Males INT,**

**No\_Children INT,**

**Death\_Cause String,**

**Loc\_Desc String,**

**Info\_Src String**

**);**

------------------------------------------------------------------

Table Description

**describe migrants;**

------------------------------------------------------------------

Before reading from file, set required delimiter. (Default Delimiter - Ctrl A)

**alter table migrants set serdeproperties ('field.delim' = '\t');**

------------------------------------------------------------------

Insert Data

**load data inpath '/user/hive/warehouse/Migrants.txt'**

**overwrite into table migrants;**

------------------------------------------------------------------

Display Added data

**select \* FROM migrants;**

------------------------------------------------------------------

Display maximum dead and missing each month.

**SELECT max(tot\_dead\_missing),reported\_month**

**FROM migrants**

**GROUP BY reported\_month;**

------------------------------------------------------------------

Display no.of migrants died in :

USA

**Select sum(no\_dead) as Total\_Migrants\_Death\_USA**

**FROM migrants**

**WHERE incident\_region like "US-Mexico Border" OR incident\_region = "Central America";**

Africa

**Select sum(no\_dead) as Total\_Migrants\_Death\_Africa**

**FROM migrants**

**WHERE incident\_region like "North Africa" OR incident\_region = "Horn of Africa" OR incident\_region = "Sub-Saharan Africa";**

Europe

**Select sum(no\_dead) as Total\_Migrants\_Death\_Europe**

**FROM migrants**

**WHERE incident\_region like "Mediterranean" OR incident\_region = "Europe"**

------------------------------------------------------------------

Display total no.of migrants

**SELECT sum(tot\_dead\_missing) Total\_Dead\_Missing,sum(no\_survivors) Total\_Survivors,**

**(sum(tot\_dead\_missing) + sum(no\_survivors)) Total\_Migrants**

**FROM migrants;**

------------------------------------------------------------------

Display total no. of incidents occurred in a region :

**SELECT incident\_region,count(incident\_region) Region\_Count**

**FROM migrants**

**GROUP BY incident\_region**

------------------------------------------------------------------

Display :

maximum no. of migrants dead & missing in each region

along with the month in each it occurred,

the death cause,

no. of males,females & children reported.

**WITH MAX\_DEATH AS ( SELECT incident\_region, max(tot\_dead\_missing) MAXD**

**FROM migrants**

**GROUP BY incident\_region**

**)**

**SELECT migrants.incident\_region, migrants.reported\_month, MAX\_DEATH.MAXD,**

**no\_males C\_males,no\_females C\_females,no\_children C\_children,death\_cause**

**FROM migrants,MAX\_DEATH**

**WHERE migrants.incident\_region = MAX\_DEATH.incident\_region AND**

**migrants.tot\_dead\_missing = MAX\_DEATH.MAXD;**

------------------------------------------------------------------

Display :

the no.of of times an incident occurred around all the regions,

maximum dead migrants,

maximum survivors for that particular incident and

display the result in descending order.

**SELECT count(death\_cause) as c, death\_cause, max(no\_dead) as d ,**

**max(no\_survivors) as s**

**FROM migrants**

**GROUP BY death\_cause**

**ORDER BY d DESC**

------------------------------------------------------------------

Display total survivors in an incident region.

**SELECT sum(no\_survivors) as max\_surv,Incident\_Region**

**FROM migrants**

**GROUP BY incident\_region**

**ORDER BY max\_surv desc;**

------------------------------------------------------------------

Note : SIMILARLY MONTH WISE AS WELL i.e. GROUP BY MONTH

Display whether total no. of migrants died & missing in each region is less or more than the average dead & missing migrants over all regions..

**CREATE VIEW Total\_Migrants\_Region AS**

**SELECT cast(sum(tot\_dead\_missing) as DOUBLE) tot\_D\_M, m2.incident\_region**

**FROM migrants m2**

**GROUP BY incident\_region;**

**DESCRIBE EXTENDED total\_migrants\_region;**

**show tables;**

**WITH AVG\_DEATH AS ( SELECT avg(tot\_d\_m) AVGD**

**FROM total\_migrants\_region**

**)**

**SELECT tmr2.incident\_region,tmr2.tot\_d\_m, AVG\_DEATH.AVGD,**

**CASE**

**WHEN tmr2.tot\_d\_m < AVG\_DEATH.AVGD THEN 'Less than average'**

**WHEN AVG\_DEATH.AVGD < tmr2.tot\_d\_m THEN 'More than average'**

**ELSE 'Nada'**

**END AS Migrants\_Affected\_Regions**

**FROM Total\_Migrants\_Region tmr2, AVG\_DEATH**

**GROUP BY incident\_region,tot\_d\_m,AVG\_DEATH.AVGD;**

------------------------------------------------------------------

EXTRAS

**truncate table migrants;**

**drop table migrants;**

**drop view Total\_Migrants\_Region;**

------------------------------------------------------**THE END** -----------------------------------------------