**Group A**

**Lab 2c**

**Hive part 2 and file format.**

**Participants:**

**1. Aadarsha chapagain**

**2.Jyoti shukla**

**3.Rishi Phaneendra Varma**

**4.Priti Bhale**

**5.Sreya Treesa Johny**

**6.Piyush Bhatia**

**Slide 2:**

**create table employee\_data(id int, name string,**

**compensation Map<string, int>)**

**row format delimited**

**fields terminated by ','**

**collection items terminated by '$'**

**map keys terminated by ':'**

**Stored as textfile;**

**Graphical user interface, text, application

Description automatically generated**

**Slide 7:**

**Select name, compensation['salary'] + compensation['comm'] + compensation['bonus']**

**as totalcompensation from employee\_data where name='Benjamin';**

**Graphical user interface, text, application

Description automatically generated**

**Slide 8:**

**Create view sales\_by\_US\_view as select \***

**From salesjan2009\_ext where country='United States';**

**Graphical user interface, text, application

Description automatically generated**

**Slide 9:**

**Select productid, salesamount, country from sales\_by\_US\_view.**

**Graphical user interface, text, application

Description automatically generated**

**Slide 10**:

**Select paymenttype, sum(salesamount) from sales\_by\_US\_view group by paymenttype**

Graphical user interface, text

Description automatically generated

**Slide 11:**

**Create a script to create a table and change permission and run via hive -f**

Graphical user interface, text, application

Description automatically generated

**Slide 26:**

**create table salesjan2009\_clustered(**

**productid string,salesamount int,paymenttype string,customername string,city string,region string,country string)**

**clustered by (country) into 10 buckets**

**stored as ORCFILE**

Text

Description automatically generated

**Slide 27:**

**insert into salesjan2009\_clustered**

**select \* from salesjan2009\_ext;**

**Graphical user interface, text, application

Description automatically generated**

**Slide 28:**

**Graphical user interface

Description automatically generated**

**Slide 29:**

**10 buckets are created and orc file is not in human readable form.**

**Text

Description automatically generated**

**Slide 32:**

**select country, count(\*) as count from SalesJan2009\_ext**

**group by (country)**

**order by count;**

**Here we can see that the data is skewed and do not have normal distribution.**

**Graphical user interface, text, application

Description automatically generated**

**Slide 33:** **create table salesjan2009\_skewed(**

**productid string,salesamount int,paymenttype string,customername string,city string,region string,country string)**

**skewed by (country) on ('United States', 'United Kingdom') stored as directories**

**row format delimited**

**fields terminated by ',';**

**We need to specify by which column and value for skewed table;**

**A screenshot of a computer

Description automatically generated**

**Slide 34: Skewed values are stored as sub directories**

**Graphical user interface, text

Description automatically generated**

**Slide 35:**

**insert into salesjan2009\_skewed**

**select \* from salesjan2009\_ext;**

**Graphical user interface, text

Description automatically generated**

**Slide 89: create table salesjan2009\_parquet1(**

**productid string,salesamount int,paymenttype string,customername string,city string,region string,country string)**

**stored as PARQUET**

**TBLPROPERTIES ("parquet.compress"="snappy");**

**Storing file as Parquet format. It is columnar file format and data are partitioned by columns.**

**Graphical user interface, text

Description automatically generated**