Data Warehousing DAY-2 :

**PROCEDURE :**

**Table creation-**

CREATE TABLE raw\_data (

patient\_id INT,

patient\_name STRING,

doctor\_id INT,

doctor\_name STRING,

appointment\_date STRING,

procedure\_name STRING,

procedure\_cost FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Loading Data-**

LOAD DATA INPATH 'C:\Users\Pavan Kalyan Reddy\Desktop\raw\_data.csv' INTO TABLE raw\_data;

**Transformed Data Table-**

CREATE TABLE transformed\_data (

patient\_id INT,

patient\_name STRING,

doctor\_id INT,

doctor\_name STRING,

appointment\_month STRING,

procedure\_count INT,

total\_cost FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Transforming the data-**

INSERT INTO transformed\_data

SELECT

patient\_id,

patient\_name,

doctor\_id,

doctor\_name,

SUBSTR(appointment\_date, 1, 7) AS appointment\_month,

COUNT(DISTINCT procedure\_name) AS procedure\_count,

SUM(procedure\_cost) AS total\_cost

FROM raw\_data

GROUP BY

patient\_id,

patient\_name,

doctor\_id,

doctor\_name,

SUBSTR(appointment\_date, 1, 7);

**Creating Hive view-**

CREATE VIEW transformed\_view AS

SELECT

patient\_id,

patient\_name,

doctor\_id,

doctor\_name,

appointment\_month,

procedure\_count,

total\_cost

FROM transformed\_data;

**Creating Hive script-**

-- Set variables

SET current\_date = FROM\_UNIXTIME(UNIX\_TIMESTAMP(), 'yyyy-MM-dd');

SET data\_path = '/user/hive/data/raw\_data.csv';

SET hive\_table = 'raw\_data';

-- Create external table for new data

CREATE EXTERNAL TABLE temp\_raw\_data (

patient\_id INT,

patient\_name STRING,

doctor\_id INT,

doctor\_name STRING,

appointment\_date STRING,

procedure\_name STRING,

procedure\_cost FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '${hiveconf:data\_path}';

-- Load new data into temporary table

INSERT OVERWRITE TABLE ${hiveconf:hive\_table}

SELECT \*

FROM temp\_raw\_data

WHERE appointment\_date > (

SELECT MAX(appointment\_date)

FROM ${hiveconf:hive\_table}

);

-- Drop temporary table

DROP TABLE temp\_raw\_data;