**THEORY:**

The ETL process stands for Extraction, Transformation & Loading.

**EXTRACTION:**

For a data warehouse, we have to extract data from many discrete sources. Few separate option are available for Data Extraction Technique.

Immediate Data Extraction

Referred data Extraction.

**TRANSFORMATION:**

Transformation is the operation of moving data from one system to another system. Major Transformation types are Formal Revision, Decoding of fields, Date/Time conversion etc.

**LOADING:**

The transformed data is required to be loaded into the data warehouse.

Initial Load

Increment Load

Full Refresh.

Above are the 3 phases involved in the loading process.

**Overview of practical**

Three tables are there **salest, sales\_details, sales\_record**

The data are extracted, cleaned and transformed from **salest** and **sales\_details** and loaded in **sales\_record**

for this particular example **sales\_record** is the datawarehouse

**CREATING TABLE SAELST:**

SQL> create table **salest**

2 (s\_person varchar2(15),

3 item\_sold number(5),

4 price number(10),

5 s\_dt date,

6 item\_category varchar2(10),

7 num\_item\_sold number(10));

Table created.

**INSERTING VALUES INTO SALEST:**

SQL> insert into salest values('Mayuri','2','12000','1-feb-2012','hardware','5');

1 row created.

SQL> select \* from salest;

S\_PERSON ITEM\_SOLD PRICE S\_DT ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri 2 12000 01-FEB-12 hardware 5

Chandrika 1 1500 02-FEB-12 software 3

Pallavi 1 15000 02-FEB-12 hardware 3

Shraddha 2 10000 01-FEB-12 hardware 4

Shweta 2 8000 01-FEB-12 software 7

Siya 3 6000 01-FEB-12 software 10

Sharvari 1 2500 01-FEB-12 software 5

Nikki 3 5000 02-FEB-12 hardware 13

Rucha 2 7000 2-FEB-12 hardware 2

Neha 3 7500 02-FEB-12 hardware 4

10 rows selected.

**CREATING TABLE SAELS\_DETAILS:**

SQL> create table **sales\_details**

2 (sales\_person varchar(30),

3 item\_name varchar(30),

4 item\_price number,

5 sales\_date date,

6 branch varchar(30),

7 no\_of\_item\_sold number);

Table created.

**INSERTING VALUES INTO SALES\_DETAILS:**

SQL> insert into sales\_details values('Mayuri','HardDisk',12000,'1-feb-2012','Thane',2);

1 row created.

SQL> select \* from sales\_details;

SALES\_PERSON ITEM\_NAME ITEM\_PRICE S ALES\_DAT BRANCH NO\_OF\_ITEM\_SOLD

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

Mayuri HardDisk 12000 01-FEB-12 Thane 2

Chandrika AntiVirus 1500 02-FEB-12 CST 3

Pallavi CPU 1500 02-FEB-12 Thane 2

Shraddha RAM 10000 01-FEB-12 Andheri 5

Sharvari HardDisk 12000 11-FEB-12 Thane 4

Rucha AntiVirus 1500 11-MAR-12 CST 2

6 rows selected.

**EXTRACTION:**

**CREATING TABLE SAELS\_RECORDS:**

SQL> create table **sales\_records**

2 (supplier varchar2(15),

3 item\_name varchar2(10),

4 price number(10),

5 sales\_dt date,

6 branch varchar2(12),

7 item\_category varchar2(10),

8 num\_item\_sold number(6));

Table created.

**INSERTING VALUES INTO SAELS\_RECORDS:**

SQL> insert into sales\_records

2 (select s\_person,item\_category,price,s\_dt,**null**,

3 item\_category,num\_item\_sold from salest);

10 rows created.

**\*\* NOTE there id no branch attribute in salest therefore a NULL is inserted into branch attribute of sales\_records**. (you can insert any ohter value also)

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT **BRANCH** ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 hardware 5

Chandrika software 1500 02-FEB-12 software 3

Pallavi hardware 15000 02-FEB-12 hardware 3

Shraddha hardware 10000 01-FEB-12 hardware 4

Shweta software 8000 01-FEB-12 software 7

Siya software 6000 01-FEB-12 software 10

Sharvari software 2500 01-FEB-12 software 5

Nikki hardware 5000 02-FEB-12 hardware 13

Rucha hardware 7000 02-FEB-12 hardware 2

Neha hardware 7500 02-FEB-12 hardware 4

10 rows selected.

SQL> insert into sales\_records

2 (select sales\_person,item\_name,item\_price,sales\_date,branch,

3  **null**,no\_of\_item\_sold from sales\_details);

6 rows created.

**\*\* There is no Item\_category in sales\_details table therefore a NULL is inserted in item\_category attribute of sales\_record**

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH  **ITEM\_CATEG** NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 hardware 5

Chandrika software 1500 02-FEB-12 software 3

Pallavi hardware 15000 02-FEB-12 hardware 3

Shraddha hardware 10000 01-FEB-12 hardware 4

Shweta software 8000 01-FEB-12 software 7

Siya software 6000 01-FEB-12 software 10

Sharvari software 2500 01-FEB-12 software 5

Nikki hardware 5000 02-FEB-12 hardware 13

Rucha hardware 7000 02-FEB-12 hardware 2

Neha hardware 7500 02-FEB-12 hardware 4

Mayuri HardDisk 12000 01-FEB-12 Thane 2

Chandrika AntiVirus 1500 02-FEB-12 CST 3

Pallavi CPU 1500 02-FEB-12 Thane 2

Shraddha RAM 10000 01-FEB-12 Andheri 5

Sharvari HardDisk 12000 11-FEB-12 Thane 4

Rucha AntiVirus 1500 11-MAR-12 CST 2

16 rows selected.

**TRANSFORMATION:**

**INSERTING VALUES INTO SAELS\_DETAILS: (transforming the dates)**

SQL> insert into sales\_details values('Gargi','HardDisk',5000**,to\_date('2012-jan-03','yyyy-mm-dd')**,'mumbai',2);

1 row created.

SQL> select \* from sales\_details;

SALES\_PERSON ITEM\_NAME ITEM\_PRICE SALES\_DAT BRANCH NO\_OF\_ITEM\_SOLD

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

Mayuri HardDisk 12000 01-FEB-12 Thane 2

Chandrika AntiVirus 1500 02-FEB-12 CST 3

Pallavi CPU 1500 02-FEB-12 Thane 2

Shraddha RAM 10000 01-FEB-12 Andheri 5

Sharvari HardDisk 12000 11-FEB-12 Thane 4

Rucha AntiVirus 1500 11-MAR-12 CST 2

Gargi HardDisk 5000 03-JAN-12 mumbai 2

7 rows selected.

**UPDATING VALUES OF SALES\_RECORDS: (instead of empty field it will show the value UNKNOWN)**

SQL> update sales\_records set branch='unknown' where branch is null;

10 rows updated.

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 unknown hardware 5

Chandrika software 1500 02-FEB-12 unknown software 3

Pallavi hardware 15000 02-FEB-12 unknown hardware 3

Shraddha hardware 10000 01-FEB-12 unknown hardware 4

Shweta software 8000 01-FEB-12 unknown software 7

Siya software 6000 01-FEB-12 unknown software 10

Sharvari software 2500 01-FEB-12 unknown software 5

Nikki hardware 5000 02-FEB-12 unknown hardware 13

Rucha hardware 7000 02-FEB-12 unknown hardware 2

Neha hardware 7500 02-FEB-12 unknown hardware 4

Mayuri HardDisk 12000 01-FEB-12 Thane 2

Chandrika AntiVirus 1500 02-FEB-12 CST 3

Pallavi CPU 1500 02-FEB-12 Thane 2

Shraddha RAM 10000 01-FEB-12 Andheri 5

Sharvari HardDisk 12000 11-FEB-12 Thane 4

Rucha AntiVirus 1500 11-MAR-12 CST 2

16 rows selected.

**From the domain knowledge or metadata repository we know that antivirus is in software category therefore update the blank fields accordingly**

SQL> update sales\_records set item\_category='software' where item\_name='AntiVirus';

2 rows updated.

SQL> update sales\_records set item\_category='hardware' where item\_name in

('CPU','RAM','HardDisk');

4 rows updated.

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 unknown hardware 5

Chandrika software 1500 02-FEB-12 unknown software 3

Pallavi hardware 15000 02-FEB-12 unknown hardware 3

Shraddha hardware 10000 01-FEB-12 unknown hardware 4

Shweta software 8000 01-FEB-12 unknown software 7

Siya software 6000 01-FEB-12 unknown software 10

Sharvari software 2500 01-FEB-12 unknown software 5

Nikki hardware 5000 02-FEB-12 unknown hardware 13

Rucha hardware 7000 02-FEB-12 unknown hardware 2

Neha hardware 7500 02-FEB-12 unknown hardware 4

Mayuri HardDisk 12000 01-FEB-12 Thane hardware 2

Chandrika AntiVirus 1500 02-FEB-12 CST **software** 3

Pallavi CPU 1500 02-FEB-12 Thane **hardware**  2

Shraddha RAM 10000 01-FEB-12 Andheri  **hardware** 5

Sharvari HardDisk 12000 11-FEB-12 Thane **hardware** 4

Rucha AntiVirus 1500 11-MAR-12 CST  **software** 2

16 rows selected.

**for missing values like num-items-sold we can take average of the other rows**

SQL> update sales\_records set num\_item\_sold=(select avg(num\_item\_sold)

2 from sales\_records) where item\_name='AntiVirus';

2 rows updated.

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 unknown hardware 5

Chandrika software 1500 02-FEB-12 unknown software 3

Pallavi hardware 15000 02-FEB-12 unknown hardware 3

Shraddha hardware 10000 01-FEB-12 unknown hardware 4

Shweta software 8000 01-FEB-12 unknown software 7

Siya software 6000 01-FEB-12 unknown software 10

Sharvari software 2500 01-FEB-12 unknown software 5

Nikki hardware 5000 02-FEB-12 unknown hardware 13

Rucha hardware 7000 02-FEB-12 unknown hardware 2

Neha hardware 7500 02-FEB-12 unknown hardware 4

Mayuri HardDisk 12000 01-FEB-12 Thane hardware 2

Chandrika AntiVirus 1500 02-FEB-12 CST software 5

Pallavi CPU 1500 02-FEB-12 Thane hardware 2

Shraddha RAM 10000 01-FEB-12 Andheri hardware 5

Sharvari HardDisk 12000 11-FEB-12 Thane hardware 4

Rucha AntiVirus 1500 11-MAR-12 CST software 5

16 rows selected.

SQL> update sales\_records set num\_item\_sold=(select avg(num\_item\_sold)

2 from sales\_records) where item\_name in('CPU','RAM','HardDisk');

4 rows updated.

SQL> select \* from sales\_records;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH ITEM\_CATEG NUM\_ITEM\_SOLD

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

Mayuri hardware 12000 01-FEB-12 unknown hardware 5

Chandrika software 1500 02-FEB-12 unknown software 3

Pallavi hardware 15000 02-FEB-12 unknown hardware 3

Shraddha hardware 10000 01-FEB-12 unknown hardware 4

Shweta software 8000 01-FEB-12 unknown software 7

Siya software 6000 01-FEB-12 unknown software 10

Sharvari software 2500 01-FEB-12 unknown software 5

Nikki hardware 5000 02-FEB-12 unknown hardware 13

Rucha hardware 7000 02-FEB-12 unknown hardware 2

Neha hardware 7500 02-FEB-12 unknown hardware 4

Mayuri HardDisk 12000 01-FEB-12 Thane hardware 5

Chandrika AntiVirus 1500 02-FEB-12 CST software 5

Pallavi CPU 1500 02-FEB-12 Thane hardware 5

Shraddha RAM 10000 01-FEB-12 Andheri hardware 5

Sharvari HardDisk 12000 11-FEB-12 Thane hardware 5

Rucha AntiVirus 1500 11-MAR-12 CST software 5

**LOADING: (summarized data is loaded in views and tables)**

**CREATING VIEW CATALOG1:**

SQL> create view catalog1 as (select sum(price\*num\_item\_sold) as yearly\_sale,

2 to\_char(sales\_dt,'yyyy') as year from sales\_records group by to\_char(sales\_dt,'yyyy'));

View created.

SQL> select \* from catalog1;

YEARLY\_SALE YEAR

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

519500 2012

SQL> create view catalog2 as (select sum(price\*num\_item\_sold) as yearly\_sale,

2 to\_char(sales\_dt,'yy') as year from sales\_records group by to\_char(sales\_dt,'yy'));

View created.

SQL> select \* from catalog2;

YEARLY\_SALE YE

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

519500 12

SQL> create view category as(select \* from sales\_records where item\_category='software');

View created.

SQL> select \* from category;

SUPPLIER ITEM\_NAME PRICE SALES\_DT BRANCH ITEM\_CATEG NUM\_ITEM\_SOLD

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

Chandrika software 1500 02-FEB-12 unknown software 3

Shweta software 8000 01-FEB-12 unknown software 7

Siya software 6000 01-FEB-12 unknown software 10

Sharvari software 2500 01-FEB-12 unknown software 5