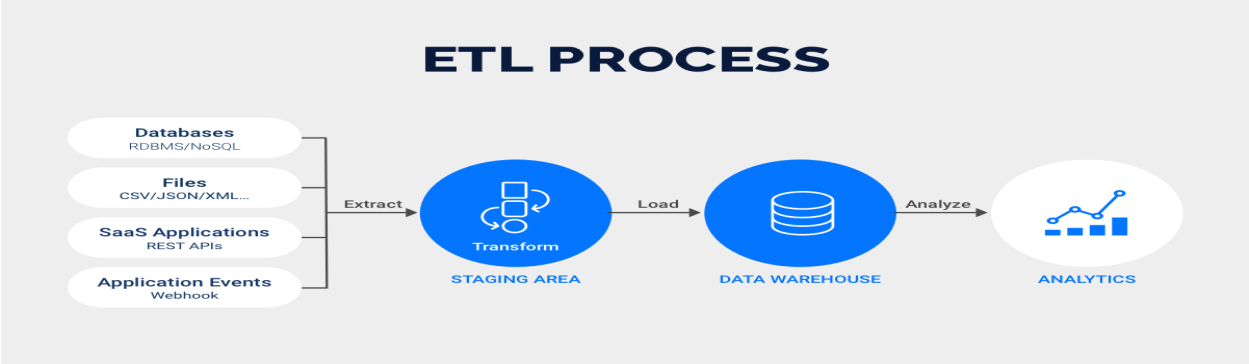
ETL(Extract, Transform, and Load) is a process that extracts data from multiple source systems, changes it and then puts it into the Data Warehouse system. It's easy to building a Data warehouse as simple as pulling data from numerous sources and feeding it into a Data warehouse database.

The ETL process, which is technically complex, involves active participation from a variety of stakeholders, including developers, analysts, testers, and senior executives.

To preserve its value as a decision-making tool, the data warehouse system must develop in sync with business developments. ETL is a regular (daily, weekly, monthly) process of a data warehouse system that must be agile, automated, and properly documented.



Steps for ETL process :

<u>Step 1) Extraction</u>

Data is extracted from the source system and placed in the staging area during extraction. If any transformations are required, they are performed in the staging area so that the performance of the source system is not harmed. Rollback will be difficult if damaged data is transferred directly from the source into the Data warehouse database.

Data warehouses can combine systems with different hardware, database management systems, operating systems, and communication protocols. Data warehouses must combine systems with disparate DBMS, hardware, operating systems, and communication protocols. Sources may include legacy programs such as mainframes, customized applications, point-of-contact devices such as ATMs and call switches, text files, spreadsheets, ERP, data from vendors and partners, and so on. Before extracting data and loading it physically, a logical data map is required.

Step 2 : Transformation

The data retrieved from the source server is raw and unusable in its original state, it must be cleaned, mapped, and transformed. It is a key ETL concept in which you apply a collection of functions to extracted data. **Direct move** or **pass through dat**a is the type of data that does not require any transformation.

Step 3 : Loading The final stage in the ETL process is to load data into the target data warehouse database. A large volume of data is loaded in a relatively short period of time in a typical data warehouse. As a result, the load process should be optimized for performance.

In the occurrence of a load failure, recovery procedures should be put in place so that operations can restart from the point of failure without compromising data integrity. Data Warehouse administrators must monitor, continue, and stop loads based on server performance.

Types of Loading:

Initial Load — filling all of the Data Warehouse tables

Incremental Load — implementing ongoing modifications as needed on a regular basis

Full Refresh — clearing the contents of one or more tables and reloading them with fresh data

Load verification

- Check that the key field data is not missing or null.
- Modelling views based on target tables should be tested.
- Examine the combined values3 and computed measures.
- Data checks in the dimension and history tables.
- Examine the BI reports on the loaded fact and dimension table.

Step 1 Create database Data Warehouse :

Step 2 Create Customer dimension table in Data Warehouse which will hold customer personal details. Fill the Customer dimension with sample Values

Step 3 Create basic level of Product Dimension table without considering any Category or Subcategory Fill the Product dimension with sample Values

Step 4 Create Store Dimension table which will hold details related stores available across various places. Fill the Store Dimension with sample Values

Step 5 Create Dimension Sales Person table which will hold details related stores available across various places. Fill the Dimension Sales Person with sample values:

Step 6 Create Date Dimension table which will create and populate date data divided on various levels.

Step 7 Create Time Dimension table which will create and populate Time data for the entire day with various time buckets.

Step 8 Create Fact table to hold all your transactional entries of previous day sales with appropriate foreign key columns which refer to primary key column of your dimensions;

```
Createdatabase Sales_DW
Create table DimCustomer (
CustomerID int primary key identity,
CustomerAltID varchar(10) not null,
CustomerName varchar(50),
Gender varchar(20)
)
Insert into DimCustomer(CustomerAltID,CustomerName,Gender) values
('IMI-001','Henry Ford','M'),
('IMI-002','Bill Gates','M'),
('IMI-003','Muskan Shaikh','F'),
('IMI-004','Richard Thrubin','M'),
('IMI-005','Emma Wattson','F');)

Create table DimProduct
```

```
(
ProductKey int primary key identity,
ProductAltKey varchar(10)not null,
ProductName varchar(100),
ProductActualCost money,
ProductSalesCost money

)
```

```
Insert into DimProduct (ProductAltKey,ProductName, ProductActualCost, ProductSalesCost)values
('ITM-001','Wheat Floor 1kg',5.50,6.50),
('ITM-002','Rice Grains 1kg',22.50,24),
('ITM-003','SunFlower Oil 1 ltr',42,43.5),
('ITM-004','Nirma Soap',18,20),
('ITM-005','Arial Washing Powder 1kg',135,139);)
```

```
Create table DimStores
(
StoreID int primary key identity,
StoreAltID varchar(10)not null,
StoreName varchar(100),
StoreLocation varchar(100),
City varchar(100),
State varchar(100),
Country varchar(100) )
```

```
Insert into DimStores(StoreAltID,StoreName,StoreLocation,City,State,Country )values
('LOC-A1','X-Mart','S.P. RingRoad','Ahmedabad','Guj','India'),
('LOC-A2','X-Mart','Maninagar','Ahmedabad','Guj','India'),
('LOC-A3','X-Mart','Sivranjani','Ahmedabad','Guj','India');)


Create table DimSalesPerson
(
SalesPersonID int primary key identity,
SalesPersonAltID varchar(10)not null,
SalesPersonName varchar(100),
StoreID int,
City varchar(100),
State varchar(100),
Country varchar(100)
)


Insert into DimSalesPerson(SalesPersonAltID,SalesPersonName,StoreID,City,State,Country
)values
('SP-DMSPR1','Ashish',1,'Ahmedabad','Guj','India'),
('SP-DMSPR2','Ketan',1,'Ahmedabad','Guj','India'),
('SP-DMNGR1','Srinivas',2,'Ahmedabad','Guj','India'),
('SP-DMNGR2','Saad',2,'Ahmedabad','Guj','India'),
('SP-DMSVR1','Jasmin',3,'Ahmedabad','Guj','India'),
('SP-DMSVR2','Jacob',3,'Ahmedabad','Guj','India');)
```