Assignment No : 1

Aim : For an organization of your choice, choose a set of business Processes. Design star/snow flake Schemas for analyzing these Processes create a fact constellation Schema by combining them. extract data from different data Sources, apply suitable transformations and load into destination tables using an ETL tool.

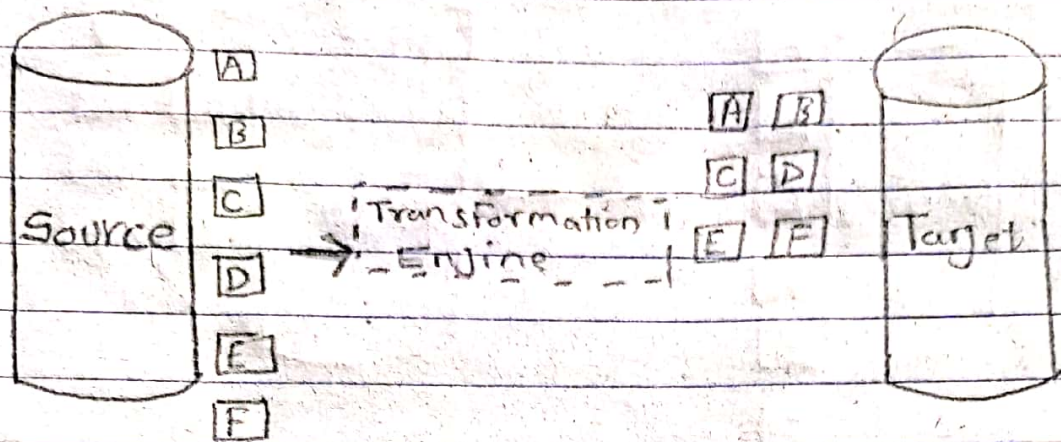For Example : Business origination : Sales, Order, Marketting Process

Objective :

1. Implementation of the Problem Statement using ETL Tool.

2. Star/Snow flake Schemas for analy
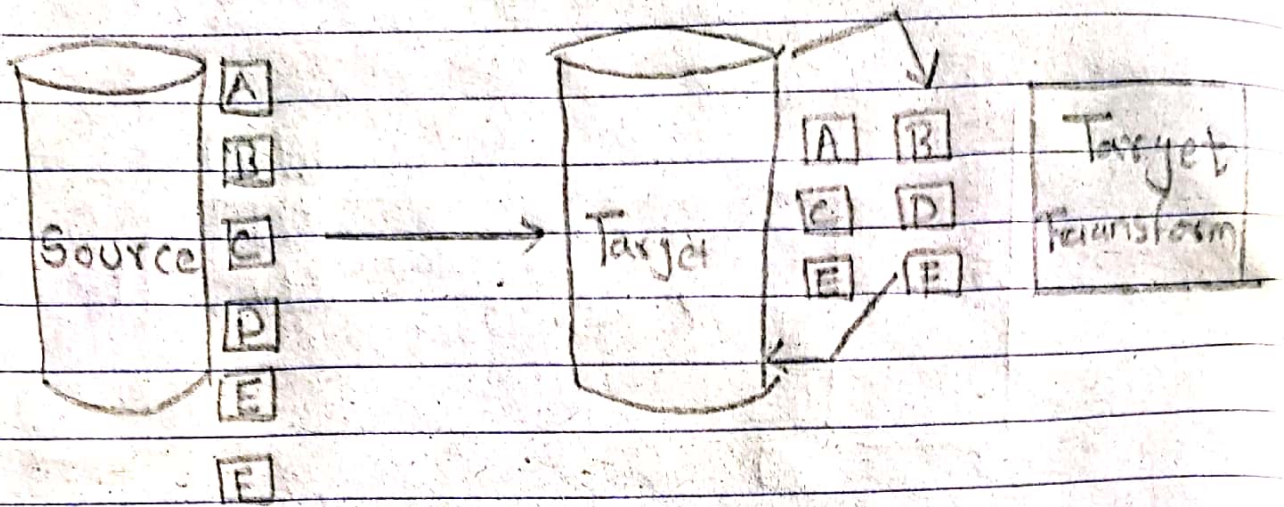
Theory :

What is ETL

1. ETL is an abbreviation of Extract, transform and load. in this Process, an ETL tool extracts the data from different RDBMS Source Systems then transforms the data like applying Calculations, Concatenations, etc and then load data into Data warehouse System.

2. In ETL data is flows from the Source to the target.

3. In ETL Process transformation engine takes care of any data changes.

Source [A] [B] [C] [D] [E] [F] → Transformation Engine [A] [B] [C] [D] [E] [F] Target

ELT :

1. ELT is a different method of looking at the tool 1 approach to data movement.
2. Instead of transforming the data before it's written, ELT lets the target System to do the transformation.
3. The data first copied to the target and then transformed in Place.
4. ELT usually used with no-sql databases like hadoop cluster, data appliance or cloud installation.

List of open Source ETL Tools

1. Clover ETL

1. Clover ETL is a pure data integration S/W suite making rapid development and enterprise capabilities available in a light footprint package.

2. It allows you to efficiently develop, deploy and automate transparent data transformations, from file to database loads to automating complex data movement between databases, files and web service APIs.

3. It provides an effective blend of rapid visual design of transformations and workflows with full coding customization capabilities and automation.

2] Jedox

1. With the user-friendly and powerful tool Jedox Integrator (ETL), you have the opportunity to combine all database systems with the multidimensional Jedox OLAP server and thus integrate BI/PM applications with Jedox fast and easy into existing IT landscapes.

2. Optimize with the Jedox Integrator data integration in order to have important data wherever they are needed.

3] Pentoho

1. With an intuitive, graphical, drag and drop design environment and a proven, scalable, standards-based architecture, data integration is increasingly the choice for organizations over traditional, proprietary ETL or Data Integration tools.

2. Data Integration delivers powerful Extraction transformation, and loading ETL, capabilities using a groundbreaking, metadata-driven approach.

5] Talend

1. Talend is the leading developer of open source data integration Systems

2. Recognized as "Visionary" in recent gartner reports on the data integration Sector, Talend makes Powerful ETL and data management technology easily available to organizations of all sizes with Talend open Studio for data integration.

3. Rather than bogging down in the business case writing and budgetary haggling associated with buying soln from one of the few behemoths of the Proprietary data management s[w market, one can download talend open studio for data integration today, for free, and dig right into Productive work on ETL and other Projects.

Conclusion : Thus we have learned to extract data from different data sources, apply suitable transformations and load into destination tables using an ETL tools.