Assignment 1

Date of Completion:

Role of Submission: 29.8.2020

Title: - Design star/800w flake schemas for analyzing the process.

Problem Statement: - De for an organization of your choice, choose a set of business processes. Design star/snow flake schemas for analyzing these processes. Greate a fact constellation schema by combining them. Extract data from different data sources, apply suitable transformations and load into destination tables using an ET tool for example: Business Organition: Sales, Order, Marketing Process.

Learning Objective: - To understand star/snow flake schema.
To understand ETI Look.

Learning outcomes: - Students will be able to use FIL

tools to design and also be able to

analyze different schemas (snow, star et)

Software / Hardware Requirement: - An ETI eg Pentaho Data Integration tool.

Invory:-Snowflake schema:-It is variant of star schema:

The centralized fact table is connected to multiple dimensions. In the snowflake schema, dimension are present in a normalized form in multiple related talles. This effect affects only the dimension tables and not the fact table Time Dimension Ordulo D Dimension BrodudD Catgory Rmonsion Sales Name Category ID ProductD CategoniTo OrderID Description Customer CustomerID EmployeeID Employer total Vimension Ruantity Department EmployeeID Discount Name Pimension Veganmento City Dimension (Fact table) DepartmentID Name Region CityID . Location 1 Name Zipende Pinensions are maintained in normalized form to heduce dependancy. Pables are easy to maintain & some storage space.

Characteristics	
1. Use small disk space 2. Eary to implement 3. Performance reduces du to multiple tal	The second secon
2. Easy to implement	Marie
3. Performance reduces du to multiple to	14.
4.	
Advantages 1. Provides structed doto which reduces doto integrity. 2. Mighly structured data, reduces disk	
1. Travioles structed data which reduces	the problem of
acta integrity.	
2. Highly structured data, reduces disk	space
Nicado a huga	
Higher Alan al 11 11	1. 110 1
Hierar chies should belong to the dimension table only and should never be snowflaked.	
status po snowpaked.	0
Star schema: -	
Can have one but falls and n	1. 1. 1.
Can have one fout table and no Simplest type of data warshouse	Selvenson Table.
	June 7 Dia
City	Country
CityTo	(austrut)
Description Country ID City D City D	Cambry to Description
(ituTO	
EMOTO	
/ Produtto	
Empto Production Total	
Emphyces Sales	
	Produck
Rane. (Fact table)	ProductID
	I Now I

Only one dimension table
Dimension table ar not joined to each other
Easy to understand Dimension table are not normalized. Selected Organization
Collège Snow flake Schema Employee Pimension Employee DD Sales Fact Table Department ID EmployeeD Producted Department Dr. Timeto Dimension City Dia StoreID Sales ID Price CityTD Name Quantity Zipcoole Product Dim Time Dimension Name Time ID Description Month Year

ETC Tools:

| ETC stands for extract transform Load tool
| Entract means exet extracting the data from huterageneous
| or homogeneous sources into our environment for integeration
| and generak insights from it.
| ETC tools extract data from different sources (tables flat
| files, etc.) and process this data.
| Transformation phase data is chansed according to need. Data can be trimmed appended, filtered, joins and be generated to
| In the bad phase, final data is loaded into the target DB
| or into flat files of or in the form of web service.
| Eg of ETC tools are Informatica doud, Abbritio etc. Conclusion: Thus I understood the different ochema used (8 tau , snow flake etc), used Pentaho tool for form performing translation.