	Assignment - 01 Pate: 01/08/2020
	Title: Analyzing and Entracking data using ETI took
	Problem: For an organization of your choice, choose Statement a set of business processes. Design Star/ snow flake schemas for analysing these
	Snow like schemes for analysing bless
	processes creat a fact constellation schema
	by combining them. Entract data from
	different dabs sources, apply suitable
-	bransformations and load data into
	destination table using ETI tools. For eg. Business Organization sales, ordering,
	Markotling process.
	V .
	Objective: Implementing of problem statement using ETL tool.
	· Star/snow flake schema for analyzing process
	S/W & . PIX, 2 GB RAM, MOD
	M/W reg . ETL openiousce tool Pentaho.
	· MySQL 5.6 & 5.7 (SQL 92)
	1119592 5.0 2 5 7 (5)
	Theory: A] Star Schemas.
	· The schemas are a way to organize data marts
	or entre data warehouse using relational databases.
	· Consider the following sales model represented
	in star sauna.
	TOD TOWN
RESGAL	FOR EDUCATIONAL USE
and the same	

dim_sales_type int PK Varchar (128) dim store din-employee Cut PK int PK Varchar (128) varchar (258) Varchar (128) varchar (178 int Varchar (128 varchar (128 Varchar C128 act sales PK FK PK FK PK FK PK FK PK FK decimal (8,2) decimal (8,2) dim-product PK PK Varchar (256 date Varchar int -week whon-month int int Varchar (16) STAR, SCHEMA FOR EDUCATIONAL USE

REEGAL NOOMING

	-> Characteristics of Stor schema:
	· levery dimension is remulsented with out
	- Characteristics of Star schema: Every dimension is represented with only one- dimensional table.
	· Fact table would contain key & measure. · It is easy to understand & provides optimal disk usage
	Fact faire would contour key a measure.
	. It is easy to understand & provides optimal
	disk usage
	• It is widely supported by BI tools. • The dimensiona babbles are not joined to each other
	· The dimensioner balalar are not raised to each
	off or
- 8	our
	-7
	B) A snouglake schema.
	- It is an extension of star schema, and it adds
	- The dimension bables are normalized which splits daba into additional table.
	della anticologia della art houndante antica some
	vala unto additional table.
	· The main benefit is that it uses smaller dick space.
	· The main benefit is that it uses smaller disk space.
	· lasier to implement, a dinession is added to echema.
	De to multiple buller and beller pressure is reduced
	· Need to nexform more maintainel efforts befause of more
	Weld to heroun more maintainte exports testaust of more
	Lookup tables.
	CJETL CEntract, Transform, Load)
	· ETL is an abbreviation for Entrat, Transform & Good.
	· In their orgers an ETL tool extracts the date
	by dillow & Roems source systems they to automat
	The management of the state of
	one water my applying radamations, concardinations,
	the and load bir baba into walla warehouse system.
)ab
	FOR EDUCATIONAL USE
REEGAL	

· 4,	ETL, data flows from source to bo	right.
	B Jrangor: A B ource C = malion C D Engine E F E	Target
• E	The is a different method of looking.	ab the tool
	etead of bransforming data before its	written.
- 7	- lets ble barget system to do bransle he data is first copied to the targe	b and blien
1	10.1102	
1	Lest of open source ETL tools	sor cloud installation
	- Clover ETC	
	- Jedon - Pentoho	
	- Talend	
Tes Sour	to Cases? Description. Temperbed 6/P	Oahral a (a
1)	Xampp/hpache Sower installation installed	Actual OfP Start Apache
2)	While installing pentaho,	shyspe server.
N	ake sure to seld PENTAHO JAVALHOME SUCCESS VTAHO_INSTALLER LICENSE. PAT Willowinend variables	Success
RESSAL	Women Variable FOR EDUCATIONAL USE	

	Bro Percription. Enjected Of Actual Of P 3) Perform transformation on successfully successfully the postal codes implements implemented. 4) Perform transformation on Null values missing-zipcodes & invalid success data is removed
	Conclusion: Thus, we have plearned to entract data from different data sources, apply suitable transformations & load into destination bable using ETI tool.
REGAL INDUSTRES	FOR EDUCATIONAL USE