	DATE:
	Assignment No: DMW-1
*	Title: ETL: Tool of the private of silving with an estate
	of the house of the control of the c
<del>*</del>	Problem Statement:
	for an organisation of your choice, choose a set of business process. Design star snowflake schemas for analyzing
104	these processes create a fact constitution 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 organization: sales, order, marketing
	process.
	Vise integ orbiting hadress Kin
*	Objectives a To, begut Es ATE amosphe beninger United to
	is rearn data modeling concepts and schemas like star snowflake
	fast constellation.
	2) apply suitable transforms on data using ETL tool.
12.5	nandora storegiz x ocusa enega A K
*	outromes: Students will be able to
	star and implement data modeling concepts and schemas like
	grave, strong transcription, free discourse
	27 Lie ETL tool.
*	sion and Him requirements: is in what contain rate perhants
	- Java JDK 1.8, MySQL, 5.5 Pentaho pdi-ce-7, MySQL connector
	- 64-bit 09-Windows 10, Intel core 15 processor, 500 98 HDD.
*	concepts related to theory: : (det an environment)
	Datawarehousing - soly called Malantin to
	Generalize and consolidate data on multidimensional
-	FOR EDUCATIONAL USE

	DATE:					
	space. I - Gapta : cold interactions.					
	OLAP- Online Analytic Processing took, information processing					
	from historical data.					
	: the state of the					
	ETKNTool - const while man is will est more in rel					
the file	Pentaho Data Integration (PDI) provides access to an					
CHILDY	extraction, transformation and loading engine that captures					
drain-						
J12 11	format. estiled animolitate and topol time evolutional and					
Marx X 800	v ration ration improver saint start int int					
	wind ows Installation					
	1> Download pentaho pdi-ce-7					
	27 Install required softwares - JDK 1.8, MySQL server					
Irli or	3> Run spoon bat file young Data Integration folder					
U	and other than					
	Assignment steps FTL wing Pentaho on sales data cur					
**	y Preprocessing → zip code resolving.					
	of older ad Illian standard season die o					
Sip .	data missing zip code data					
	data missing zip code data					
	no les my les se					
	TIG .					
	2> Lookup for missing data in zipcodes while per will have some					
Hagga						
rts	3> road data to Mysal database - then point to a csv file.					
	Transformations on data: : possedi at hodeler timener of					
	a) calculate selling price					
11111	profit loss calculation					
	FOR EDUCATIONAL USE					

	Y .(C)			DATE:				
tro	concatenation of fields related to product.							
1,1	when is tweeter our undilies without of historien ann							
	Data preprocessing before minimed (ETL) 1100000 110000 del							
	· Data extraction: gathering data.							
	· Data deaning: find errors - mode it tens but -							
	· Data transformation : convert to different format							
	· load : summarize views							
	· Refresh or update.	and .	11 /	Labor William				
	Outro dang	Gridos artis)	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	and in the				
	Modeling i data - schemas : part mode							
	> star schema - A large				,			
	and control			noite of				
	ast. at	sales (fact table)		rollet				
	Kento/ah	time-key		Elim				
	time vin	item_key	item					
	(dim table)	branch_key		(dim.table)				
•	time_key	location_key		item-key				
	day saling	unity sold		item=name				
	day-ofweek will	(3/du.) arith)		brand				
ý	month & region -	Lead notherol	The second	typent				
	quarter	Janesta		supplier type				
	years with	<u>asy</u>		ild. Tarry				
	Topy report	ustano		7010				
	* , * .							
	branch			tion (dim. table)				
	(dlm. table)			location ckey brown				
40	1 branch Lkey or miles	a religion in		treet & will				
P	branch 2.112	or igaldnamal	and how	itylera 1, 40				
	branch-type			ce or state				
	J'		, 0	unby				
_								

2> SnowPlake scheme	2 - variant of	star schema. some dim tables					
are normalized by further splitting are reduced to redundancy							
test may increase it fetch time. moled pales any mod							
J	subb crime	dun : ralbindes etia					
3> Fact constellation		ball : palant etal .					
11		more than ma fact table					
, 5		aiv arisomorus + brul *					
sales	item	Chact table)					
(fact_table)	(dim table)	(fact table)					
item_key	itom kou	· xamato - item keyilahali					
11 1 3 1							
	aupplier	Sam Los					
location_key							
dollar sold	( 2 lot if I to 1) 16 le	to_loc  dollar cost					
bloz_tinu	us/.200il						
(+) (tal-mits	110 J (risel)	unite_shipped					
	red domail						
Hann "	ried Hellionel	1191.1011.10					
(dim table)	(dim_table)	shipper 1116					
		cdim table)					
- time skey	location_key	shipper key					
1 y day por	street	shipper-name					
month	key	location_key					
year	country	shipper_type					
old and address		Sauce					
* Conclusion dimet		Colled mite					
Thus, we success	Thus see successfully installed pentaho pair-ce-7 ET tool						
and performed to	and performed transformations on sales data						
· In (= ro Gor) history	•	mus knowd					
pilgeo		<u> </u>					