

ETL Step by step Example

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

DB1

Customer Table 1

CustID	CustName	Gender
1	Mark	Male
2	John	Male
3	Mila	Female
4	Ann	Female
5	Philip	Male

Order Table 1

ItemID	CustID	Date	Qty
1	1	1/1/10	8
1	2	1/10/10	6
2	2	4/5/11	1
2	4	2/2/09	5
2	5	1/1/12	9
3	5	2/2/13	1
3	3	1/1/11	3
4	4	15/7/12	5

Items Table 1

ItemID	Description	Price
1	iPad	899
2	Laptop	688
3	Lemon	1.5
4	Pen	5
5	Table	59

DB2

Customer Table 2

CustID	CustName	Gender
1	Mark	1
2	Joe	1
3	Martha	0
4	Paula	0
5	Philip	1

Order Table 2

ItemID	CustID	Date	Qty
1	1	5/1/10	8
1	5	4/10/10	6
2	2	6/5/11	1
2	4	7/2/12	5
2	4	3/1/12	9
3	5	21/2/13	1
3	3	13/1/14	3
4	4	15/7/12	5

Items Table 2

ItemID	Description	Price
1	chair	56
2	Laptop	688
3	Apple	1.5
4	Spoon	0.7
5	Table	59

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

TARGET STAR SCHEMA

Customer DIM

<u>CustSK</u>	CustName	Gender
		Male

Item DIM

<u>ItemSK</u>	Description	Price

FACT TABLE

<u>ItemSK</u>	<u>CustSK</u>	<u>DateSK</u>	Total	Qty

DATE Dim

<u>DateSK</u>	Day	Month	Year

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

CUSTOMER DIMENSION

<https://tutorcs.com>

WeChat: cstutorcs

DB1

Customer Table 1

CustID	CustName	Gender
1	Mark	Male
2	John	Male
3	Mila	Female
4	Ann	Female
5	Philip	Male

DB2

Customer Table 2

CustID	CustName	Gender
1	Mark	1
2	Joe	1
3	Martha	0
4	Paula	0
5	Philip	1

Customer Staging Table

CustID	CustName	Gender	SourceDB	SK
1	Mark	Male	1	
2	John	Male	1	
3	Mila	Female	1	
4	Ann	Female	1	
5	Philip	Male	1	
1	Mark	1	2	
2	Joe	1	2	
3	Martha	0	2	
4	Paula	0	2	
5	Philip	1	2	

Assignment Project Exam Help

LOAD



<https://tutorcs.com>

WeChat: cstutorcs

Customer Staging Table

CustID	CustName	Gender	SourceDB	CustSK
1	Mark	Male	1	
2	John	Male	1	
3	Mila	Female	1	
4	Ann	Female	1	
5	Philip	Male	1	
1	Mark	1	2	
2	Joe	1	2	
3	Maria	0	2	
4	Paula	0	2	
5	Philip	1	2	

1. Consolidation /
Entity Matching.
Assume same name =
same customer. Assign
same surrogate key

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Customer Staging Table

CustID	CustName	Gender	SourceDB	CustSK
1	Mark	Male	1	1
2	John	Male	1	2
3	Mila	Female	1	3
4	Ann	Female	1	4
5	Philip	Male	1	5
1	Mark	Male	2	1
2	Joe	Male	2	6
3	Martina	Female	2	7
4	Paula	Female	2	8
5	Philip	Male	2	5

1. Consolidation / Entity Matching.
Assume same name = same customer. Assign same surrogate key

2. Assign Surrogate Keys. This information is used for future load of new data, it is not temporary!

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Standardize data

Customer Staging Table

CustID	CustName	Gender	SourceD B	CustSK
1	Mark	Male	1	1
2	John	Male	1	2
3	Mila	Female	1	3
4	Ann	Female	1	4
5	Philip	Male	1	5
1	Mark	1	2	1
2	Joe	1	2	6
3	Martha	0	2	7
4	Paula	0	2	8
5	Philip	1	2	5

Customer Dimension

CustSK	CustName	Gender
1	Mark	Male
2	John	Male
3	Mila	Female
4	Ann	Female
5	Philip	Male
6	Joe	Male
7	Martha	Female
8	Paula	Female

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



LOAD DATA INTO THE DIMENSION CUSTOMER

Assignment Project Exam Help

<https://tutorcs.com>
DIMENSION ITEM

WeChat: cstutorcs

DB1

Items Table 1

ItemID	Description	Price
1	iPad	899
2	Laptop	688
3	Lemon	1.5
4	Pen	5
5	Table	59

DB2

Items Table 2

ItemID	Description	Price
1	chair	56
2	Laptop	688
3	Apple	1.5
4	Spoon	0.7
5	Table	59

Item Staging Table

ItemID	CustName	Price	SourceDB	SK
1	iPad	899	1	
2	Laptop	688	1	
3	Lemon	1.5	1	
4	Pen	5	1	
5	Table	59	1	
1	chair	56	2	
2	Laptop	688	2	
3	Apple	1.5	2	
4	Spoon	0.7	2	
5	Table	59	2	

Assignment Project Exam Help

LOAD



<https://tutorcs.com>

WeChat: cstutorcs

Item Staging Table

ItemID	CustName	Price	SourceDB	SK
1	iPad	899	1	1
2	Laptop	688	1	2
3	Lemon	1.5	1	3
4	Pen	5	1	4
5	Table	59	1	5
1	chair	56	2	6
2	Laptop	688	2	2
3	Apple	1.5	2	7
4	Spoon	0.7	2	8
5	Table	99	2	9

1. Consolidation /
Entity Matching

2. Assign surrogate keys

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Item Staging Table

ItemID	CustName	Price	SourceDB	SK
1	iPad	899	1	1
2	Laptop	688	1	2
3	Lemon	1.5	1	3
4	Pen	5	1	4
5	Table	59	1	5
1	chair	56	2	6
2	Laptop	688	2	2
3	Apple	1.5	2	7
4	Spoon	0.7	2	8
5	Table	99	2	9

Item Dimension

ItemSK	CustName	Price
1	iPad	899
2	Laptop	688
3	Lemon	1.5
4	Pen	5
5	Table	59
6	chair	56
7	Apple	1.5
8	Spoon	0.7
9	Table	99

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutores



LOAD DATA INTO THE DIMENSION ITEM

Assignment Project Exam Help

<https://tutorcs.com>
DATE DIMENSION

WeChat: cstutorcs

Date dimension

- It can be pre-populated (the same for time dimension).

```
CREATE TABLE d_date AS  
SELECT
```

```
  n AS Date_SK,  
  TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day') AS Full_Date,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'DD') AS Days,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'Mon') AS Month_Short,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'MM') AS Month_Num,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'Month') AS Month_Long,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'Q') AS Quarter,  
  TO_CHAR(TO_DATE('31/12/2007','DD/MM/YYYY') + NUMTODSINTERVAL(n,'day'),'YYYY') AS Year
```

```
FROM (  
  select level n  
  from dual  
 connect by level <= 2000);
```

Assignment Project Exam Help

<https://tutores.com>

WeChat: cstutores

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Order Table 1

ItemID	CustID	Date	Qty
1	1	1/1/10	8
1	2	1/10/10	6
2	2	4/5/11	1
2	4	2/2/09	5
2	5	1/1/12	9
3	5	2/2/13	1
3	3	1/1/11	3
4	4	15/7/12	5

Order Table 2

ItemID	CustID	Date	Qty
1	1	5/1/10	8
1	5	4/10/10	6
2	2	6/5/11	1
2	4	7/2/12	5
2	4	3/1/12	9
3	5	21/2/13	1
3	3	13/1/14	3
4	4	15/7/12	5

Stage Fact Table

ItemID	CustID	Date	Qty	Price	Item SK	Cust SK	Date SK	Source DB
1	1	1/1/10	8					1
1	2	1/10/10	6					1
2	2	4/5/11	1					1
2	4	2/2/09	5					1
2	5	1/1/12	9					1
3	5	2/2/13	1					1
3	3	1/1/11	3					1
4	4	15/7/12	5					1
1	5	5/1/10	8					2
1	5	4/10/10	6					2
2	2	6/5/11	1					2
2	4	7/2/12	5					2
2	4	3/1/12	9					2
3	5	21/2/13	1					2
3	3	13/1/14	3					2
4	4	15/7/12	5					2

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Stage Fact Table

Assign Surrogate Key

1. Join the stage Fact Table with the Customer Stage Table to get the right Customer SK. Join on the CustID and the SourceDB
2. Join the stage Fact Table with the Item Stage Table to get the right Item SK. Join on the ItemID and the SourceDB
3. Join the stage Fact Table with the Date dimension to get the right DateSK. Join on the day, month and year parsed from the field Date of the stage Fact Table.
4. Price is a derived field (Qty X Price). In the fact table it is called total

ItemID	CustID	Date	Qty	Price	Item SK	Cust SK	Date SK	Source DB
1	1	1/1/10	8	7192	1	1	34	1
1	2	1/10/10	6	5394	1	2	45	1
2	2	4/5/11	1	688	2	2	163	1
2	4	2/2/09	5	3440	2	4	54	1
2	5	1/1/12	9	6192	2	5	34	1
3	5	2/2/13	1	1.5	3	5	65	1
3	3	1/1/11	3	4.5	3	3	34	1
4	4	15/7/12	5	25	4	4	54	1
1	1	5/1/10	8	448	6	1	39	2
1	5	4/10/10	6	336	6	5	59	2
2	2	6/5/11	1	688	2	6	133	2
2	4	7/2/12	5	3440	2	8	178	2
2	4	3/1/12	9	6192	2	8	33	2
3	5	21/2/13	1	1.5	7	5	61	2
3	3	13/1/14	3	4.5	7	7	342	2
4	4	15/7/12	5	3.5	8	8	128	2

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Stage Fact Table

ItemID	Cust ID	Date	Qty	Price	Item SK	Cust SK	Date SK	Source DB
1	1	1/1/10	8	7192	1	1	34	1
1	2	1/10/10	6	5394	1	2	45	1
2	2	4/5/11	1	688	2	2	163	1
2	4	2/2/09	5	3440	2	4	54	1
2	5	1/1/12	9	6192	2	5	34	1
3	5	2/2/13	1	1.5	3	5	65	1
3	3	1/1/11	3	4.5	3	3	34	1
4	4	15/7/12	5	25	4	4	54	1
1	1	5/1/10	8	448	6	1	39	2
1	5	4/10/10	6	336	6	5	59	2
2	2	6/5/11	1	688	2	6	133	2
2	4	7/2/12	5	3440	2	8	178	2
2	4	3/1/12	9	6192	2	8	33	2
3	5	21/2/13	1	1.5	7	5	61	2
3	3	13/1/14	3	4.5	7	7	342	2
4	4	15/7/12	5	3.5	8	8	128	2

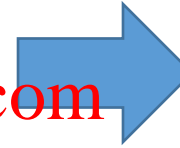
Fact Table

Qty	TOTAL	Item SK	Cust SK	Date SK
8	7192	1	1	34
6	5394	1	2	45
1	688	2	2	163
5	3440	2	4	54
9	6192	2	5	34
1	1.5	3	5	65
3	4.5	3	3	34
5	25	4	4	54
8	448	6	1	39
6	336	6	5	59
1	688	2	6	133
5	3440	2	8	178
9	6192	2	8	33
1	1.5	7	5	61
3	4.5	7	7	342
5	3.5	8	8	128

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



LOAD INTO THE FACT TABLE

Assignment Project Exam Help

WHAT HAPPENS IF THERE ARE NEW DATA AND I
NEED TO LOAD THEM IN THE STAR SCHEMA?

<https://tutorcs.com>

WeChat: cstutorcs

NEW DATA FROM DB2

Customer Staging Table

DB2

CustID	CustName	Age	Gender
1	Mark	34	1
2	Joe	30	1
3	Martha	34	0
4	Pauline	19	0
5	Philip	25	1
6	Ryan	28	1
7	Patricia	19	0

LOAD



CustID	CustName	Age	Gender	SourceDB	CustSK
1	Mark	34	Male	1	1
2	John	33	Male	1	2
3	Mila	51	Female	1	3
4	Ann	19	Female	1	4
5	Philip	25	Male	1	5
1	Mark	34	Male	2	1
2	Joe	30	Male	2	6
3	Martha	34	Female	2	7
4	Paula	19	Female	2	8
5	Philip	25	Male	2	5

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

NEW DATA FROM DB2

Customer Staging Table

CustID	CustName	Gender	SourceDB	CustSK
1	Mark	Male	1	1
2	John	Male	1	2
3	Mila	Female	1	3
4	Ann	Female	1	4
5	Philip	Male	1	5
1	Mark	Male	2	1
2	Joe	Male	2	6
3	Martina	Female	2	7
4	Pauline	Female	2	8
5	Philip	Male	2	5
6	Ryan	Male	2	9
7	Patricia	Female	2	10

Assignment Project Exam Help

LOAD

<https://tutorcs.com>

WeChat: cstutorcs

ASSIGN SURROGATE KEYS AND STANDARDIZE DATA

Customer Staging Table

CustID	CustName	Gender	SourceDB	CustSK
1	Mark	Male	1	1
2	John	Male	1	2
3	Mila	Female	1	3
4	Ann	Female	1	4
5	Philip	Male	1	5
1	Mark	Male	2	1
2	Joe	Male	2	6
3	Martha	Female	2	7
4	Pauline	Female	2	8
5	Philip	Male	2	5
6	Ryan	Male	2	9
7	Patricia	Female	2	10

Updated Customer Dimension

CustSK	CustName	Gender
1	Mark	Male
2	John	Male
3	Mila	Female
4	Ann	Female
5	Philip	Male
6	Joe	Male
7	Martha	Female
8	Pauline	Female
9	Ryan	Male
10	Patricia	Female

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



INCREMENTAL LOAD DATA INTO THE DIMENSION CUSTOMER

Changes in the Fact Table and Dimensions

- New facts are 99% appended in the fact table (very few exceptions)
- Dimensions change slower than the fact table. However, when a dimension is modified by an update, multiple strategies are possible (see next presentation)

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs