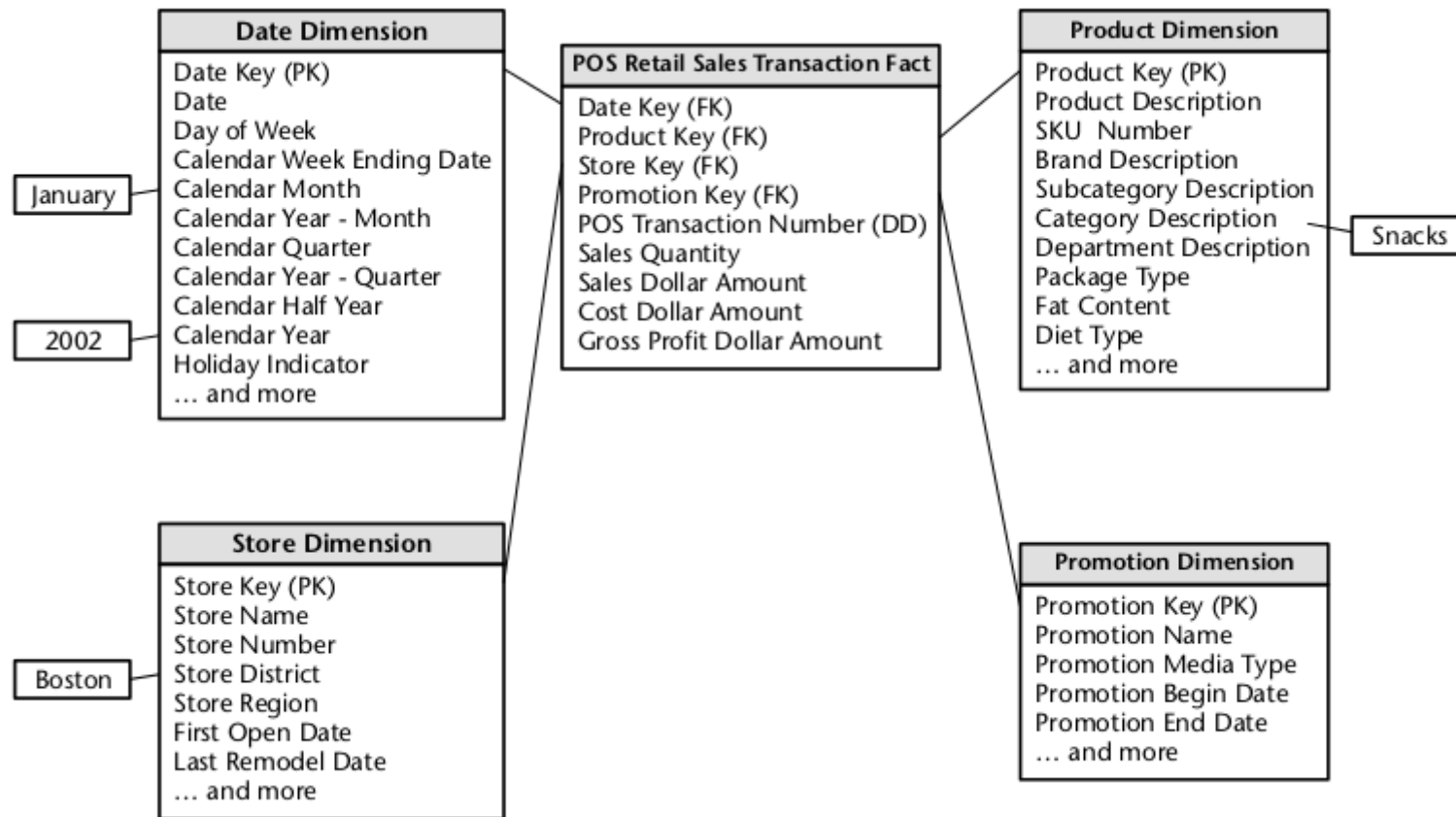


CI2355 – Almacenes de datos y OLAP

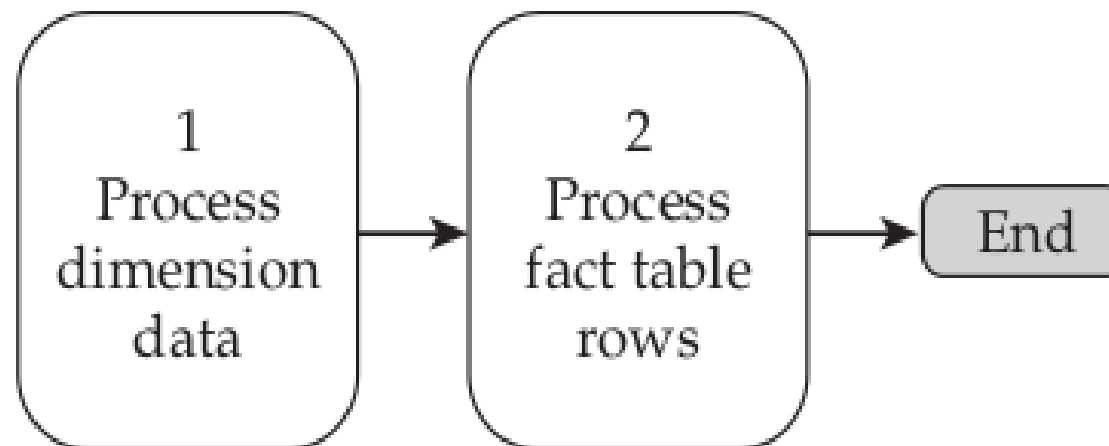


**UNIVERSIDAD DE
COSTA RICA**

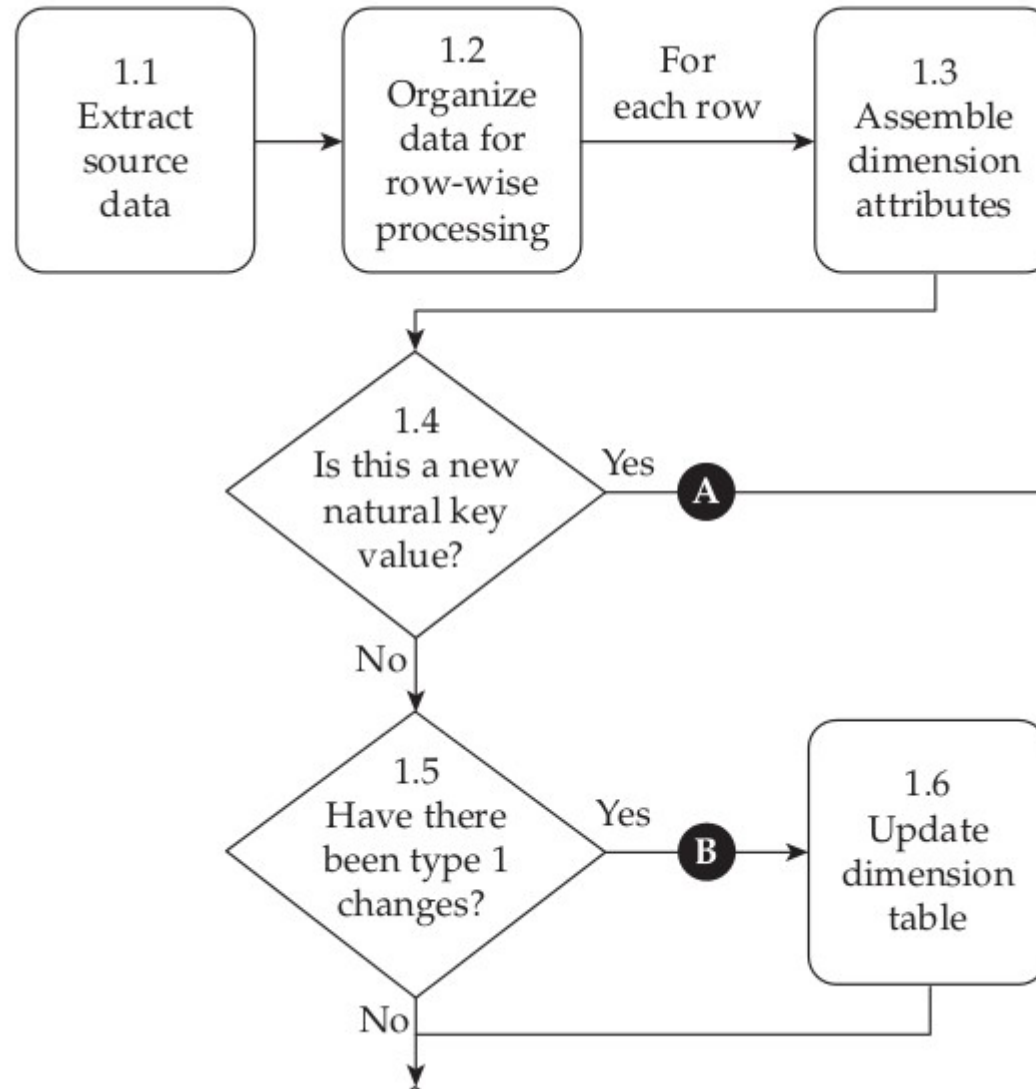
ETL



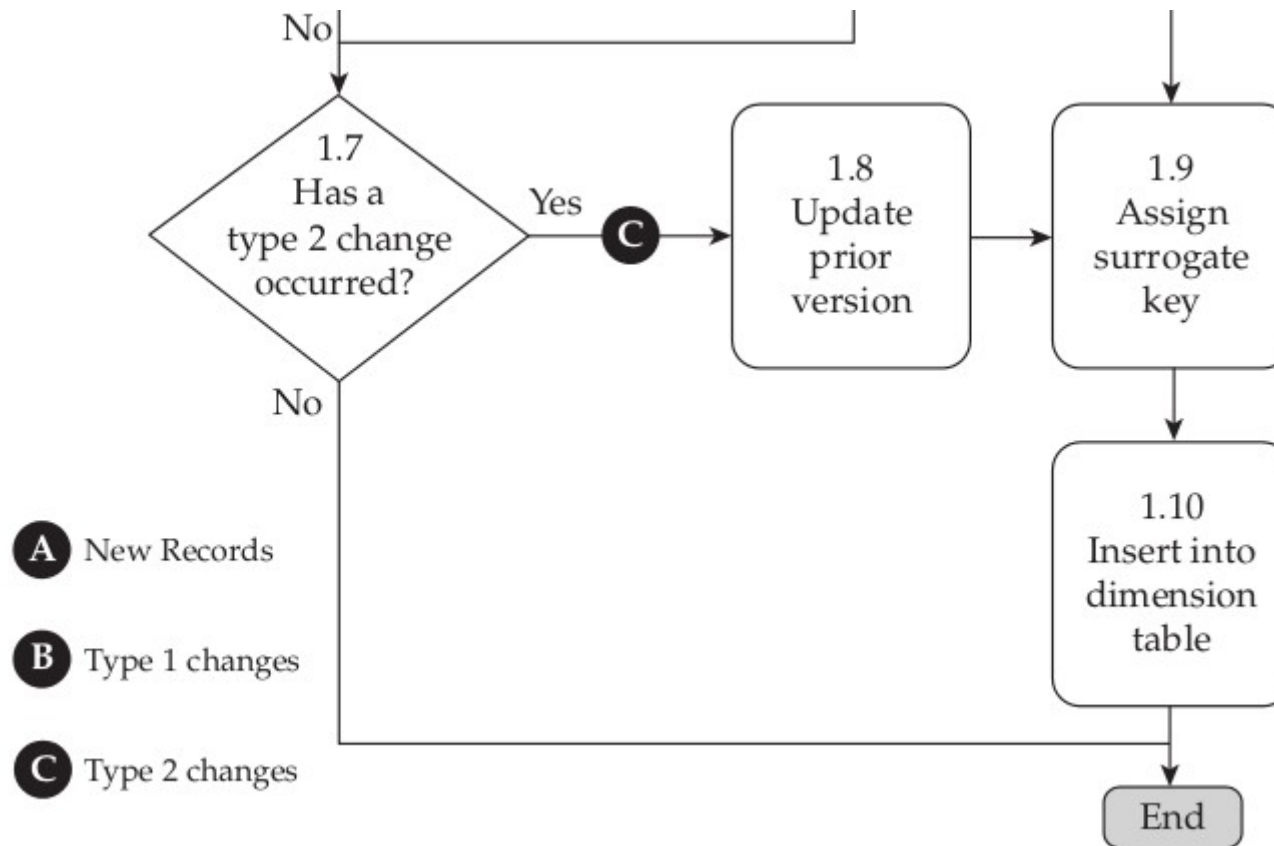
ETL



ETL – Tabla de dimensiones



ETL – Tabla de dimensiones



ETL

Design of PRODUCT dimension table:

PRODUCT	
product_key	SK
sku	NK
product_name	1
brand_code	2
brand_manager	2
.	.
.	.
current_version	HK

Legend:

- SK Surrogate Key
- NK Natural Key
- 1 Type 1 Change (Overwrite)
- 2 Type 2 Change (Insert new record)
- HK Housekeeping Column (Internal use)

ETL

Existing content of PRODUCT Dimension Table:

<i>SK</i>	<i>NK</i>	<i>1</i>	<i>2</i>	<i>2</i>	<i>HK</i>
product_ key	sku	product_ name	brand_ code	brand_ manager	current_ version
22	A-0033	Flat Box 12"	110	Lei, Michelle	Current
344	A-1011	Packing Tape	221	Jones, Paul	Current
1001	B-3691	Twine Spool	501	Smith, Dei	Not Current
1201	B-3691	Twine Spool	501	Klein, Pete	Not Current
2700	B-3691	Twine Spool	702	Jones, Fay	Current
2711	Z-3320	Envelope A5	109	Harvey, Ned	Current

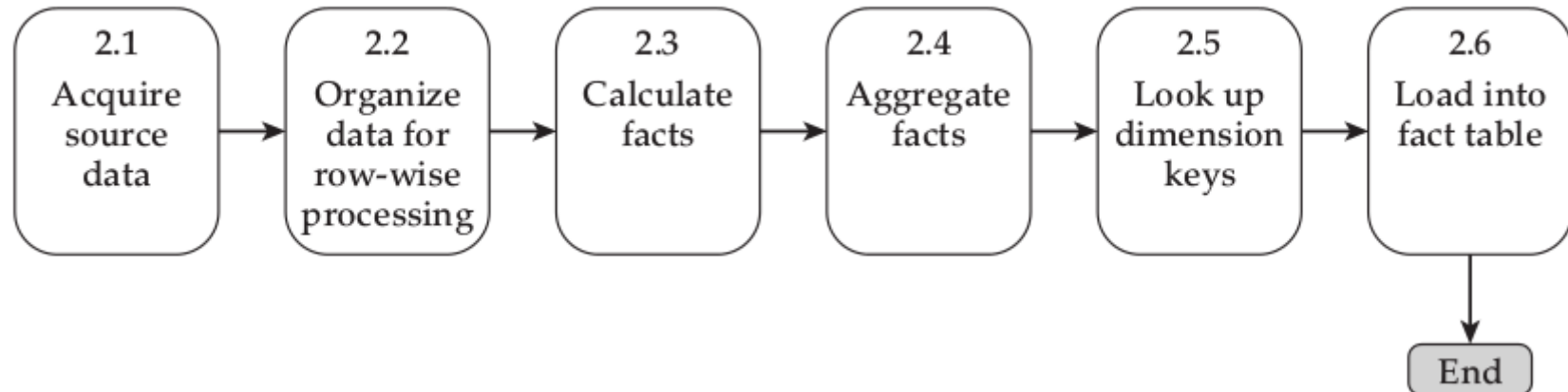
ETL

Source records to be processed:

	sku	product_name	brand_code	brand_manager	
1	S-8472	Envelope B4	109	Harvey, Ned	A new record
2	B-3691	TwineSpool 12	702	Jones, Fay	Type 1 change affecting 3 dimension rows
3	A-0033	Flat Box 12"	110	Cook, Dan	Type 2 change
4	A-1011	Clear Packing Tape	221	Davis, Jon	Type 1 and type 2 changes
5	Z-3320	Envelope A5	109	Harvey, Ned	An unchanged record

The diagram illustrates source records to be processed in an ETL process. A table with 5 rows and 5 columns is shown. The columns are sku, product_name, brand_code, and brand_manager. The rows are numbered 1 to 5. Callouts point to specific rows: 'A new record' points to row 1; 'Type 1 change affecting 3 dimension rows' points to row 2; 'Type 2 change' points to row 3; 'Type 1 and type 2 changes' points to row 4; and 'An unchanged record' points to row 5.

ETL – Tablas de hechos



ETL

date	sku	adams	benson	chiu	roberson
1/8/2009	A-1011	744.00	1,222.00	3,011.11	0
1/8/2009	Z-3320	100.00	0	600.00	200.00



date	sku	salesperson	order_dollars
1/8/2009	A-1011	Adams	744.00
1/8/2009	Z-3320	Adams	100.00
1/8/2009	A-1011	Benson	1,222.00
1/8/2009	A-1011	Chiu	3,011.11
1/8/2009	Z-3320	Chiu	600.00
1/8/2009	Z-3320	Roberson	200.00

ETL

date	order_line	sku	type	value
1/8/2009	Q1977-001	A-0033	Sales Dollars	10,000.00
1/8/2009	Q1977-001	A-0033	Margin Dollars	2,000.00
1/8/2009	Q1977-001	A-0033	Quantity Ordered	100
1/8/2009	Q1977-002	B-3691	Sales Dollars	14,000.00
1/8/2009	Q1977-002	B-3691	Margin Dollars	700.00
1/8/2009	Q1977-002	B-3691	Quantity Ordered	70



date	order_line	sku	sales_dollars	margin_dollars	quantity_ordered
1/8/2009	Q1977-001	A-0033	10,000.00	2,000.00	100
1/8/2009	Q1977-002	B-3691	14,000.00	700.00	70

Orders being processed for FACT load:

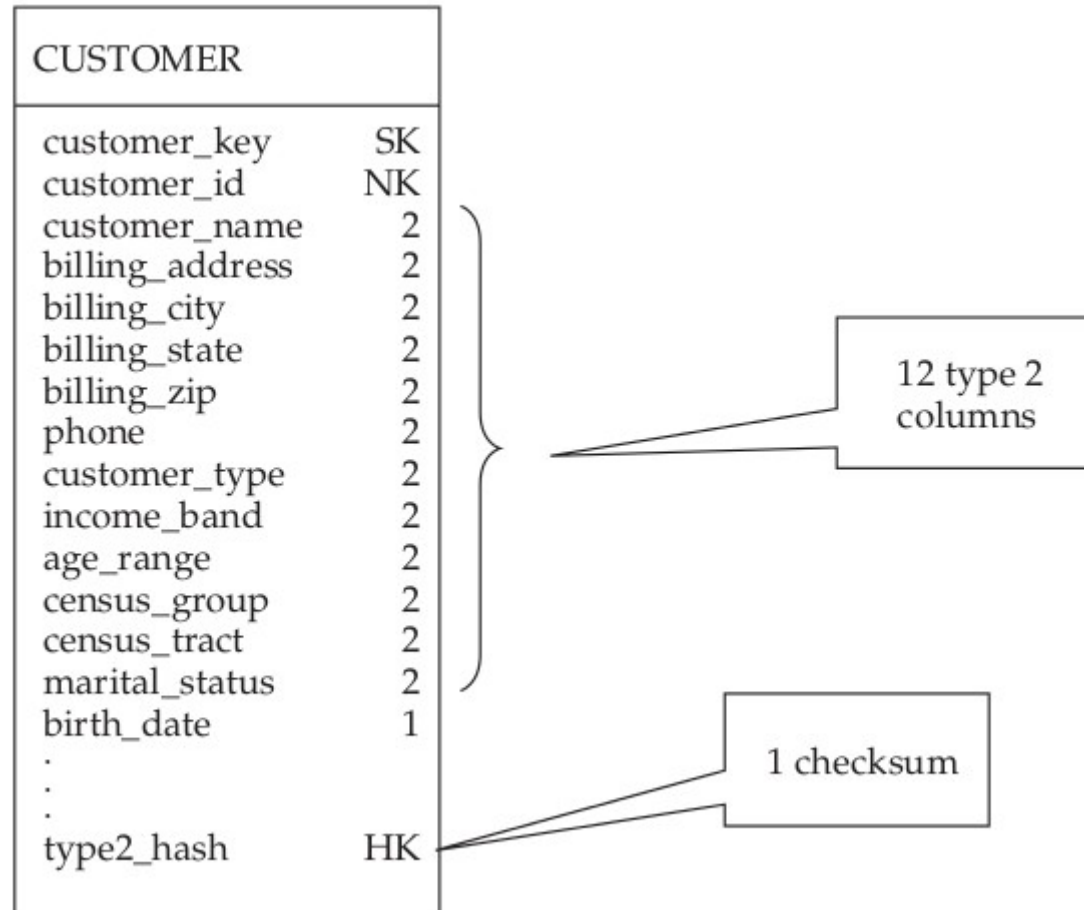
	date	order_dollars	margin_dollars	customer_id	salesrep_id	sku	product_key
1	3/12/2010	1,700.22	643.06	2201	1197	S-8472	2799
2	3/12/2010	984.31	201.64	3832	1197	B-3691	

PRODUCT Dimension Table:

SK	NK	1	2	2	HK
product_key	sku	product_name	brand_code	brand_manager	current_version
22	A-0033	Flat Box 12"	110	Lei, Michelle	Not Current
1192	A-0033	Flat Box 12"	110	Cook, Dan	Current
344	A-1011	Clear Packing Tape	221	Jones, Paul	Not Current
2828	A-1011	Clear Packing Tape	221	Davis, Jon	Current
1001	B-3691	TwineSpool 12	501	Smith, Dei	Not Current
1201	B-3691	TwineSpool 12	501	Klein, Pete	Not Current
2700	B-3691	TwineSpool 12	702	Jones, Fay	Current
2799	S-8472	Envelope B4	109	Harvey, Ned	Current
3001	Z-3320	Envelope A5	109	Harvey, Ned	Current



ETL



ETL

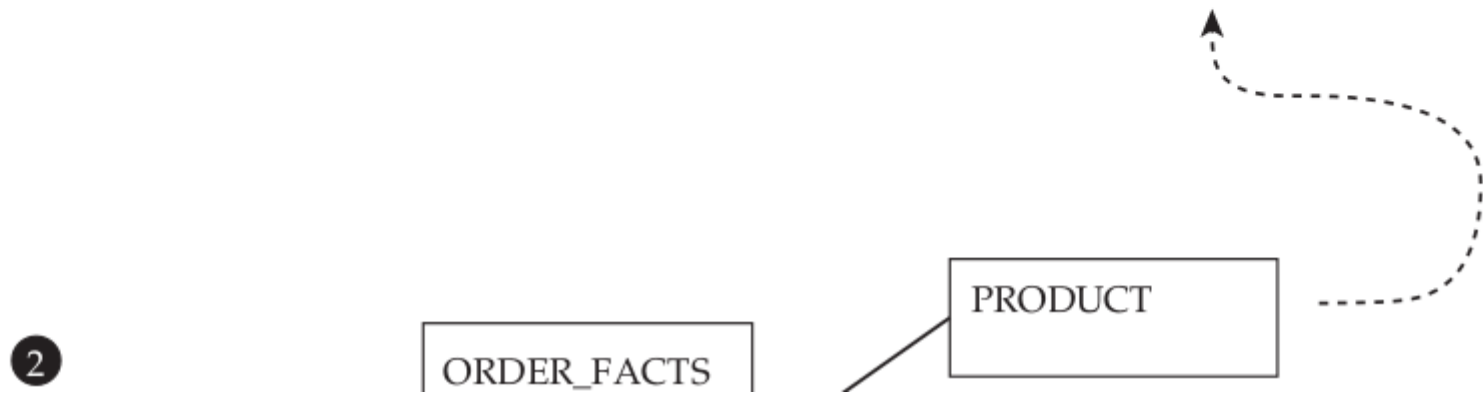
source_system	source_element_name	source_element_value	warehouse_element_name	warehouse_element_value
001	brand_id	100	brand_code	100
002	brand_num	B-100-09	brand_code	100
001	brand_id	200	brand_code	200
002	brand_num	B-200-12	brand_code	200

ETL

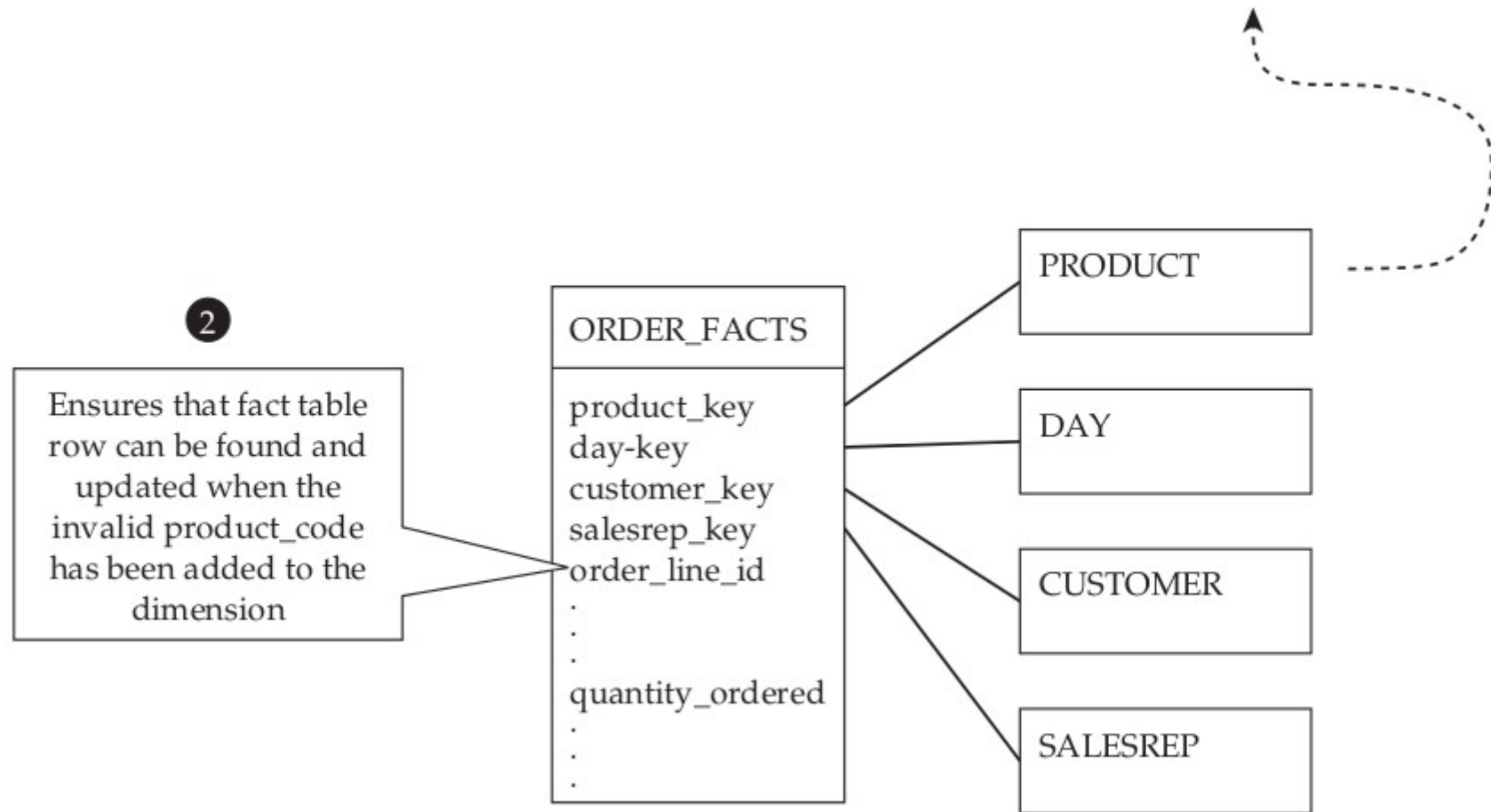
1

Used in the fact table when an order has an invalid product_code

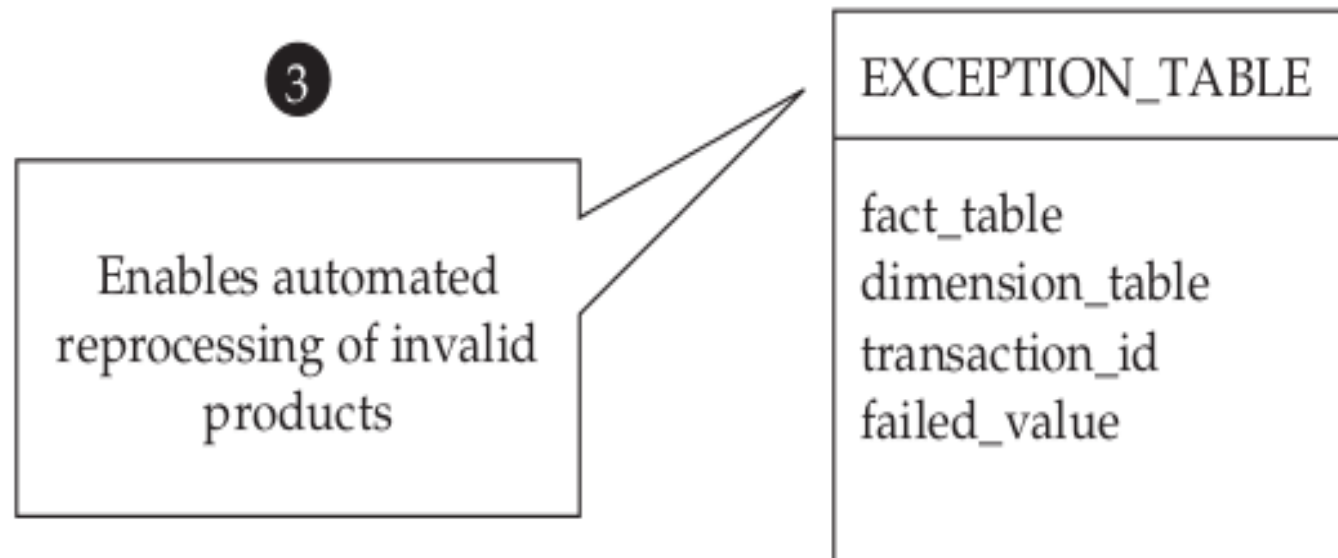
product_key	row_type	product_code	product_name
0	Invalid	[Invalid]	n/a
101	Product	B57330-1	Cardboard Box
102	Product	B47770-2	Bubble Envelope



ETL



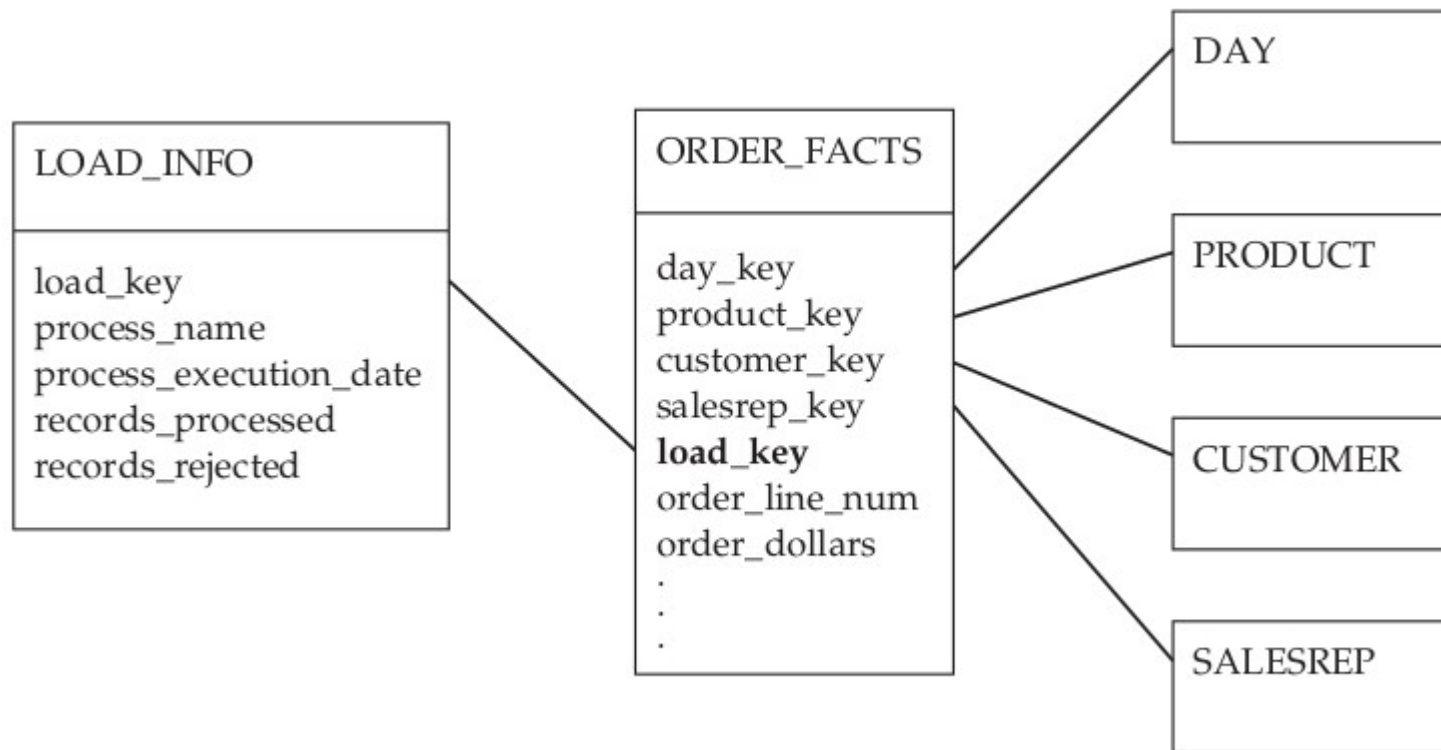
ETL



CUSTOMER	
customer_key	SK
customer_id	NK
customer_name	2
billing_address	2
billing_city	2
billing_state	2
billing_zip	2
phone	2
customer_type	2
income_band	2
age_range	2
census_group	2
census_tract	2
marital_status	2
birth_date	1
.	
.	
current_version	HK
effective_date	HK
expiration_date	HK
type2_hash	HK
type1_hash	HK
creation_date	HK
update_date	HK
update_process	HK

} Housekeeping Columns

ETL



Referencias



Adamson, C.

Star Schema - The Complete Reference.

McGraw-Hill, 2010.