

Source to Stage Mapping

Field Name	Data Type	Example Data	Max Length in dataset	Mapping to Stage	Stage Table Name	Stage Column	Data Type
Category	object	office_supplies, technology, furniture	15	straight copy	stg_orders	category	varchar(50)
City	object	new_york_city, los_angeles, philadelphia	35	straight copy	stg_orders	city	varchar(75)
Country	object	united_states, australia, france	32	straight copy	stg_orders	country	varchar(75)
Customer ID	object	jg-158051 wb-218504 bc-111252	9	drop record if value is NULL, else straight copy	stg_orders	customer_id	varchar(25)
Customer Name	object	muhammed_yedwa b, steven_ward, gary_hwang	22	straight copy	stg_orders	customer_name	varchar(75)
Discount	float64			straight copy	stg_orders	discount	float
Market	object	APAC, LATAM, EU	6	straight copy	stg_orders	market	varchar(50)
记录数	int64	1	1	straight copy	stg_orders	记录数	int
Order Date	object	2014-06-18 00:00:00.000	23	straight copy	stg_orders	order_date	timestamp
Order ID	object	CA-2014-100111	15	drop record if value is NULL, else straight copy	stg_orders	order_id	varchar(50)
Order Priority	object	Medium, High, Critical	8	straight copy	stg_orders	order_priority	varchar(50)
Product ID	object	OFF-AR-10003651	16	drop record if value is NULL, else straight copy	stg_orders	product_id	varchar(50)
Product Name	object	Staples, Cardinal Index Tab, Clear, Zebra GK420t Direct Thermal/ Thermal Transfer Printer	127	straight copy	stg_orders	product_name	varchar(200)
Profit	float64			straight copy	stg_orders	profit	float
Quantity	int64		14	straight copy	stg_orders	quantity	int
Region	object	Central, South, EMEA	14	straight copy	stg_orders	region	varchar(50)
Row ID	int64		51290	straight copy	stg_orders	row_id	int
Sales	int64		22638	straight copy	stg_orders	sales	float
Segment	object	Consumer, Corporate, Home Office	11	straight copy	stg_orders	segment	varchar(50)
Ship Date	object	2014-11-22 00:00:00.000	23	straight copy	stg_orders	ship_date	timestamp
Ship Mode	object	Standard Class, Second Class, First Class	14	straight copy	stg_orders	ship_mode	varchar(50)
Shipping Cost	float64			straight copy	stg_orders	ship_cost	float
State	object	California, England, New York	36	straight copy	stg_orders	state	varchar(100)
Sub-Category	object	Binders, Storage, Art	11	straight copy	stg_orders	sub_category	varchar(50)
Year	int64		2014	straight copy	stg_orders	year	int

Field Name	Data Type	Example Data	Max Length in dataset	Mapping to Stage	Stage Table Name	Stage Column	Data Type
Market2	object	APAC, North America, LATAM	13	straight copy	stg_orders	market2	varchar(50)
weeknum	int64		53	straight copy	stg_orders	weeknum	int
				File ID associated with the staging file from the files table	stg_orders	file_id	int
						insert_date	timestamp

Stage to Target Mapping - customers

Field Name	Data Type	PK/FK details	Mapping from Stage	Stage Table Name	Stage Column	Data Type
customer_id	varchar(50)	primary key	Check to see if customer_id already exists on the customers table, if it doesn't then a new customer record will be added with the value from stg_orders as the customer_id, if it does then the record will be updated with the latest data, for records with the same customer_id value on a given file, take the record that appears last in the file	stg_orders	customer_id	varchar(50)
customer_name	varchar(75)		direct copy from stg_orders	stg_orders	customer_name	varchar(75)
segment	varchar(50)		direct copy from stg_orders	stg_orders	segment	varchar(50)
market	varchar(50)		direct copy from stg_orders	stg_orders	market	varchar(50)
file_id	int		direct copy from stg_orders	stg_orders	file_id	int
record_id	int		direct copy from stg_orders	stg_orders	id	int
insert_date	timestamp		date/time when row is inserted			

Stage to Target Mapping - orders

Field Name	Data Type	PK/FK details	Mapping from Stage	Stage Table Name	Stage Column	Data Type
der_order_id	int	PK (auto incrementing field)				
order_id	varchar(50)	natural key	A distinct order consists of a distinct combination of order_id / customer_id / order_date. Check to see if this combination already exists on the orders table, if it doesn't then a new order record will be added, if it does then the record will be updated with the latest data, for dupe records, take the record with the largest row_id	stg_orders	order_id	varchar(50)
customer_id	varchar(50)	natural key and foreign key - customers table	See order_id for logic	stg_orders	customer_id	int
order_date	timestamp	natural key	direct copy from stg_orders	stg_orders	order_date	timestamp
priority	varchar(50)		direct copy from stg_orders	stg_orders	priority	varchar(50)
city	varchar(75)		direct copy from stg_orders	stg_orders	city	varchar(75)
state	varchar(100)		direct copy from stg_orders	stg_orders	state	varchar(100)
country	varchar(100)		direct copy from stg_orders	stg_orders	country	varchar(100)
region	varchar(50)		direct copy from stg_orders	stg_orders	region	varchar(50)
file_id	int		direct copy from stg_orders	stg_orders	file_id	int
record_id	int		direct copy from stg_orders	stg_orders	id	int
insert_date	timestamp		date/time when row is inserted			

Stage to Target Mapping - orders_products

Field Name	Data Type	PK/FK details	Mapping from Stage	Stage Table Name	Stage Column	Data Type
id	int	PK (auto incrementing field)				
der_order_id	int	foreign key - orders table	For a given order / product_id combination on the stage table, a new row will be written to this table			
der_product_id	int	foreign key - orders table	See der_order_id for logic			
quantity	int		direct copy from stg_orders	stg_orders	quantity	int
sale_amt	float		direct copy from stg_orders	stg_orders	total_sale_amt	float
discount	float		direct copy from stg_orders	stg_orders	discount	float
profit	float		direct copy from stg_orders	stg_orders	profit	float
ship_date	timestamp		direct copy from stg_orders	stg_orders	shipment_date	timestamp
ship_mode	varchar(50)		direct copy from stg_orders	stg_orders	mode	varchar(50)
ship_cost	float		direct copy from stg_orders	stg_orders	cost	float
file_id	int		direct copy from stg_orders	stg_orders	file_id	int
record_id	int		direct copy from stg_orders	stg_orders	id	int
insert_date	timestamp		date/time when row is inserted			

Stage to Target Mapping - products

Field Name	Data Type	PK/FK details	Mapping from Stage	Stage Table Name	Stage Column	Data Type
der_product_id	int	primary key				
product_id	varchar(50)	natural key	Check to see if product_id / product_name combination already exists on the products table, if it doesn't then a new product record will be added from stg_orders, if it does then the record will be updated with the latest data, for records with the same product_id / product_name value on a given file, take the record that appears last in the file	stg_orders	product_id	varchar(50)
product_name	varchar(200)	natural key	direct copy from stg_orders	stg_orders	product_name	varchar(200)
category	varchar(50)		direct copy from stg_orders	stg_orders	category	varchar(50)
subcategory	varchar(50)		direct copy from stg_orders	stg_orders	sub_category	varchar(50)
file_id	int		direct copy from stg_orders	stg_orders	file_id	int
record_id	int		direct copy from stg_orders	stg_orders	id	int
insert_date	timestamp		date/time when row is inserted			