

BI Data Engineer Challenge

- You have 72 hours to complete this challenge
- You can use the technologies of your choice to complete this example
- Our main technologies are Postgres, Python, LucidChart, Draw.io and Snowflake
- but we do realize not everyone uses these tools everyday like us
- All technical deliverables are expected to be in fully working order
- All code and/or other artifacts are required to be turned in
- If special configurations are required to run your solution, please include a detailed technical configuration write-up

Challenge

Background

Your Company Tumbleweed Capital is in the process of migrating from one sales system to another. The sales records will not be migrated to the new system, so your team is responsible for creating a new Data Mart to store the historical data. Once the new system is up and running your team will also need to create processes to load that data into the Data Mart for historical reporting and analysis.

Project Description

The project is running in phases, to enable continuous delivery to the stakeholders. You are assigned the initial task of creating an ETL that will upload the data from the old sales system into the Data Mart. Your Data Architect has sketched out a possible data model based on the information provided at requirements sessions. You have 3 days to create a working proof of concept using the sample file and the data model. Your Data Architect is out of the office until Monday of the next week and has not reviewed the sample. They have told you to go ahead if you need to make changes to the data model, and they will review and discuss those changes with you Monday. They do not want those changes to stand in the way of your first working deliverable.

Your assigned tasks components:

- Create a working ETL using the sample file provided from the old system (historical_orders_2B9S_LTD.csv)
- The ETL must read the data, perform any scrubbing and land it in the data model
- Update the Dimensional Model as applicable and provide the ERD


New Requirements Additional

Of course, right after the Data Architect goes on PTO and you are about to start, you get new requirements from your Project Manager, that need to be in your deliverable:


- The sample data is supposed to be reflective of the data in the old system but there will be data coming in the new system, which has a different data structure.
- The old system only stored addresses at the time of order but the new system is going to have customer address versioning
- The new sales system will also have order statuses
- Provide a sample reports for the following use cases:


- The ability to look at the total number of orders shipped to a specific state in a given year
- The ability to look up all customers current addresses
 - Summary reports that will show all orders and their order status
 - A way to see what customers did not have order in certain years
 - Marketing is really interested in evaluating customer order trends
 - Quarterly Sales reports at the product level as well a quarterly report at the customer level

Data Model

customers	
	customer_id
	first_name
	last_name
	date_of_birth
	phone
	email

order_details	
	product_name
	order_id
	price
	quantity
	comments

address_history	
	address_id
	customer_id
	street_number
	city
	postal_code
	state
	start_date
	end_date

orders	
	order_id
	customer_id
	order_date
	order_status
	payment_method_id
	comments