

# Data Mart case study

You are assigned the task of designing the backend ETL pipeline and data warehouse for large e-commerce players like Amazon Shopping App.

Here you are required to store all the information about products, brands, categories, sub-categories, sellers, customers, orders, sales, shipping and delivery.

The solution should include the tech stack for batch processing where the source of the data is product OLTP system.

The backend data warehouse should be able to solve for the following business scenarios:

1. Provide the Product Organisations about the growth/depreciation trends and categorise the products by sales quantities Month over Month
2. Study Seller Behaviour and Statistics (both at individual and demographic level) to categorise them into high, medium and low buckets based on quality of service (consider sales, returns & refunds)
3. Study Consumer Behaviour and Statistics (both at individual and demographic level) to categorise them into high, medium and low buckets based on quality (consider buy, return & refund), spends (consider sales transaction amount) and frequency (consider transaction counts)

Please come up with a relevant design for

1. Component of the end to end data pipeline with the selection for correct data processing platform and tools.
2. preferably with a set of dimensional model diagrams.

You can assume the table and column names as per your interpretation of the amazon.com website. Feel free to make relevant assumptions and please call out your assumptions considered, if any.

Please note that you will be expected to explain in detail the reasons to support your design and the solutions for the scenarios at the time of the interview.