

Basic Steps of Designing a Data Warehouse

Ozone.bg

Teodor Chakarov
Student ID: 12141198

Intro

Since 2008, Ozone.bg online store has dedicated itself to the difficult task of helping you get the right decision and most out of moments regardless of how you prefer to spend them. The range of the products available in the online store are from Computer games, book/music store, board games, electric devices (phones, PC, laptops, PS5, etc.), sports store and much more.

OLTP

1. E-commerce Platform System:

a) Operations:

- **Product Management:** Allows easily product listings to be added, removed and modified.
- **Order Management:** Deals with customer orders until the shipment phase.
- **Inventory management:** Conducts a constant check of stock products.

b) Data Stored:

- Information about the products
- The availability of the products
- Order history
- User account data

c) Typical Questions:

- Which products are the best sellers this month?
- How many units of product X are left in stock?

- Which customers have the highest purchase frequency?
- When is the earliest date for shipment?
- What are the given product requirements/characteristics and more

d) Strategically Important Questions:

- Why are certain products more popular than others?
- What other potential products could be added to the inventory based on market trends?
- How satisfied are customers post-purchase?
- How do our sales figures compare to competitors?
- Shall the focus be on selling more expensive items?

2. Customer Relationship Management System

a) Operations:

- **Interaction tracking:** user interaction with the website
- **Marketing campaign management:** monitors marketing campaigns and results.
- **Customer segmentation management:** split users by different categories and conduct the given marketing strategy.

b) Data Stored:

- Customer data: users' profile

- Marketing campaign data
- Sales of each product

c) Typical Questions Answered:

- Which customers have not made a purchase in the last X months?
- What is the average response time to customer inquiries?
- How many times customer returned the product?
- Are customers brand biased?
- How many new customers are generated from the last marketing campaign?

d) Strategically Important Questions Not Answered:

- What are the industry trends in customer engagement?
- How can the customer journey be further optimized?
- How do competitors' customer engagement strategies compare?
- Is there a better way to segment our customers?
- How often a marketing campaign should be released?

3. Supply Chain Management System:

a) Operations:

- Supplier Management: Manages relationships with product suppliers
- Demand Forecasting: Predicts product demand to help with restocking.
- Logistics management

b) Data Stored:

- Suppliers' data: contact details, product details and prices.
- Inventory Data: Product quantities in the stock

Logistics Data: Shipment companies, services and types.

c) Typical Questions Answered:

- Which suppliers take more time for the delivery of a product?
- What is the forecasted demand for product X in the upcoming month?
- How much is being spent on logistics for a specific geographic region?
- Which products often run out of stock?
- How long does it typically take for a product to move from the supplier to the end consumer?

d) Strategically Important Questions Not Answered:

- Are there potential new suppliers that can offer better terms or products?
- Are there emerging technologies that can further streamline the supply chain?
- Are there new products which are being sold outside the country?
- Can the company become the first distributor of product X in the county?
- How can the supply chain be made more resilient to disruptions?

Data Warehouse

a) 5 Examples for Important Strategic Questions:

- Which product categories have consistently shown growth in demand over the last five years?
- Which emerging market segments (e.g., VR gaming, audiobooks) have the potential to drive significant revenue in the next five years?
- How can the supply chain be optimized to reduce holding costs while ensuring product availability?

- Which geographic regions or demographic segments currently show lower engagement with the company's offerings?
- Based on purchase histories and customer behaviors, which products are most often bought together?
- How can the organization develop targeted marketing campaigns to retain and nurture the customers?

b) 5 Key Subject Areas:

- Products information
- Customer data
- Sales transactions
- Inventory
- Supply management

More about "Sales Transactions":

Data Stored:

- ID of customer, transaction, product/s
- Date and time
- Total amount
- Payment Method
- Shipping address and Status
- Promotion applied
- Return data.

Sources:

- E-commerce Platform System
- Supply Chain Management System

c) DWH Architecture and Development:

Kimball's Bottom-Up Approach can be used. Especially when we have not so many different data storages with complex data, we can combine all sources (tabular) and rebuild them in Star schema.

Advantages:

Since ozon.bg don't provide other complex services or B2B services which need a lot of adaptations over time, the easiest and fastest way it so integrates the data into one data warehouse (Star schema). Then it can be queried easily and fast from different departments when insights are needed.

Drawbacks:

Integration Complexity: Over time, as more data marts are added, integrating them can become complex.

Consistency: Ensuring consistency across different data marts can be challenging, especially if they were developed at different times or by different teams.

d) 5 Potential Challenges:

- **Inconsistent Definitions:** Different departments or business units may define certain metrics or terms differently.
- **Data Quality:** We must be sure that the data we have is up-to-date and accurate.
- **ETL Complexity:** The processes can be complex, especially when dealing with diverse source systems and large volumes of data.
- **Complex Queries and Reporting:** As the business grows and evolves, the queries and reports can become more complex, leading to delays in delivering insights.
- **Inaccurate, Missing Data or Duplicate data:** Those aspects may occur as problematic because customer data can exist for example in both E-commerce platform and the CRM system.

Data Mart

- Data Mart for "Customer Insights":
- Transaction Details – id, date and time, total amount, payment method, status
- Product information – id, name, category, price, quantity sold, discounts applied
- Customer information – id, name, email
- Shipping details – shipping address, shipping method, delivery status, tracking number

Possible Queries:

- Which products have the highest sales in terms of quantity for a specific month?
- Who are the top 10 customers with the highest purchase amount in the last quarter?
- How many transactions were processed using each payment method in the past year?
- How many transactions were delivered using each shipping method in the last six months?
- Which categories of products have the highest returns?

Data Lake

- Cost-Effectiveness: Since we will deal mainly with structured data, more cost-effective approach will be DWH
- Use-Cases: In case we want to collect in some form unstructured data and use power of ML and Deep Learning on it for future benefits. Although as a e-commerce reseller, I don't think it will need this type of data.
- Integrating a Data Lake into the Proposed Architecture:

A data lake can keep the raw data, while the data warehouse can be used for structured data used for

regular business reporting. But since both systems are used this can lead to more complexity.

- Organizational changes:

The organization might need to hire new or train existing employees to work with the new technologies. Also we have to collect different type of data which should be adapted from the initial CRM, Supply chain management system and E-commerce Platform System.

