



IST 722 DATA WAREHOUSE

# **VBay Vault: Empowering Bidding Insights through Data Warehousing**

**A journey From Vbay Database to Vbay Data Warehouse using  
Kimball Data Warehouse Approach**

**Final project**

**December 2023**

**SUBMITTED BY**

**Somia Abdelrahman | Harshit Joshi | Vedant Devidas Patil**

## Overview

The VBay database serves as the backbone of an online marketplace known as VBay. This relational database encapsulates key components such as user profiles, item listings, bid transactions, reviews, and other pertinent data integral to the functioning of the virtual marketplace. Designed to efficiently manage e-commerce operations, the VBay database plays a pivotal role in orchestrating seamless user interactions within the platform. Leveraging Extract, Load, Transform (ELT) processes, our data warehouse project seeks to harness the rich information stored in the VBay database. Through this project, we aim to address critical business challenges by transforming raw data into meaningful insights and ultimately creating a comprehensive Power BI dashboard. This dashboard will empower decision-makers with actionable information, fostering strategic decision-making and optimizing the overall performance of the VBay online marketplace.

## Tools & Methodology

To achieve the objectives of our data warehouse project, we strategically employed a combination of advanced tools and methodologies. Our central platform was Snowflake, a cloud-based data warehouse that offered unparalleled scalability, flexibility, and high-performance analytics. Leveraging the robust architecture of Snowflake, we efficiently stored and managed the VBay data warehouse.

In conjunction with Snowflake, we implemented dbt (data build tool) to transform and model the raw data residing in Snowflake. This approach, characterized by modularity and version control, facilitated the creation of structured datasets, establishing a solid foundation for subsequent analytics.

At the heart of our methodology was the Extract, Load, Transform (ELT) process. Diverging from traditional ETL processes, ELT involves extracting raw data directly from the source, loading it into a centralized repository (in this case, Snowflake), and performing transformations within the data warehouse itself. The initial step encompassed loading raw data from S3 bucket into a raw database in Snowflake, constituting the extract process. However, this raw database represented unprocessed data that required cleaning and transformation to comply with kimball data warehouse architecture methodology , marking the second phase of the ELT process. The uncleaned data was transformed using the data build tool and subsequently loaded into Snowflake as structured, refined data into an analytics database. This analytics-ready database served as the foundation for generating meaningful insights. The analytics data warehouse, housing the transformed data, was then seamlessly integrated into the Business Intelligence tool.

For visualization and reporting, we opted for Power BI, a robust business analytics tool. Power BI played a crucial role in crafting a comprehensive and user-friendly BI dashboard, empowering stakeholders to derive valuable insights from the transformed data.

In summary, our data warehouse project's success rested on the strategic amalgamation of Snowflake, dbt, and Power BI, complemented by the ELT and kimball data warehouse methodology. This cohesive integration empowered us to effectively transform raw data from the VBay database into actionable insights, thereby facilitating informed decision-making within the context of the online marketplace.

# **Milestones & Reflections**

## **Milestone 1 - Bus Matrix**

Milestone 01, which focused on creating the Bus Matrix, marked a crucial starting point in our project. Initially, we identified four key business processes: "Bidding Behaviour," "Item Categories and Popularity," "User Satisfaction," and "Bidding Hotspots," along with relevant fact tables and dimensions. This phase laid the groundwork for a comprehensive understanding of our dataset intricacies. Following valuable feedback from our professor, highlighting the need for streamlining, we appreciated the guidance. Prioritizing efficiency and clarity, we adjusted our approach by consolidating the "Bidding Behaviour" and "Bidding Hotspots" processes into a unified business process. This refinement ensured alignment with our overarching objectives and set the stage for subsequent project milestones.

## **Milestone 2 - Detail Dimensional Model and Implementation**

In Milestone 2 of our project, we focused on the detailed implementation of the Dimensional Model. Building upon the groundwork laid in Milestone 1, which involved the conceptualization of Dimension and Fact tables, our team delved into the intricate process of implementation. The Dimensional Model was further refined to incorporate specific enhancements based on valuable feedback received during Milestone 1. Noteworthy adjustments included the addition of the "bid\_ok" column to the "fact\_bids" table, ensuring a comprehensive representation of bid-related information. Additionally, we refined the column name "item\_duration" to "item\_duration\_in\_days" to enhance clarity in the dataset. The most significant refinement was addressed in the "fact\_satisfaction" table, where we derived the "rating\_overall" and "rating\_count" to tackle the semi-additivity issue associated with the "rating" column. Consequently, we built the data pipelines and transformed the data in dbt. This milestone reflects our commitment to meticulous model implementation, incorporating iterative improvements guided by insightful feedback from the professor.

## **Milestone 3 - Business Intelligence**

For Milestone 3, our focus centered on delivering a comprehensive Business Intelligence (BI) demonstration of key business processes. In Milestone 3, we successfully established the connection between the data warehouse and the BI tool, Power BI, and crafted a meaningful BI dashboard encapsulating three vital business processes across two distinct reports. The demonstration showcased our ability to extract valuable insights from the data. Professor's feedback on this milestone emphasized refining the user experience and analytical capabilities. Recommendations included the addition of slicers for Buyer and Seller, incorporating slicers for rating, adjusting the color scheme of the pie chart to reflect bid status, introducing slicers for state and bid status, and enabling auto-zoom functionality for maps. The feedback was constructively worked upon and the team curated an even more inclusive report.

Project Github Link: [https://github.com/DewSam/dataTriUnity\\_project/tree/main/project/models](https://github.com/DewSam/dataTriUnity_project/tree/main/project/models)