Data Governance @ SneakerPark

Prepared by: RAVI KUMBAR

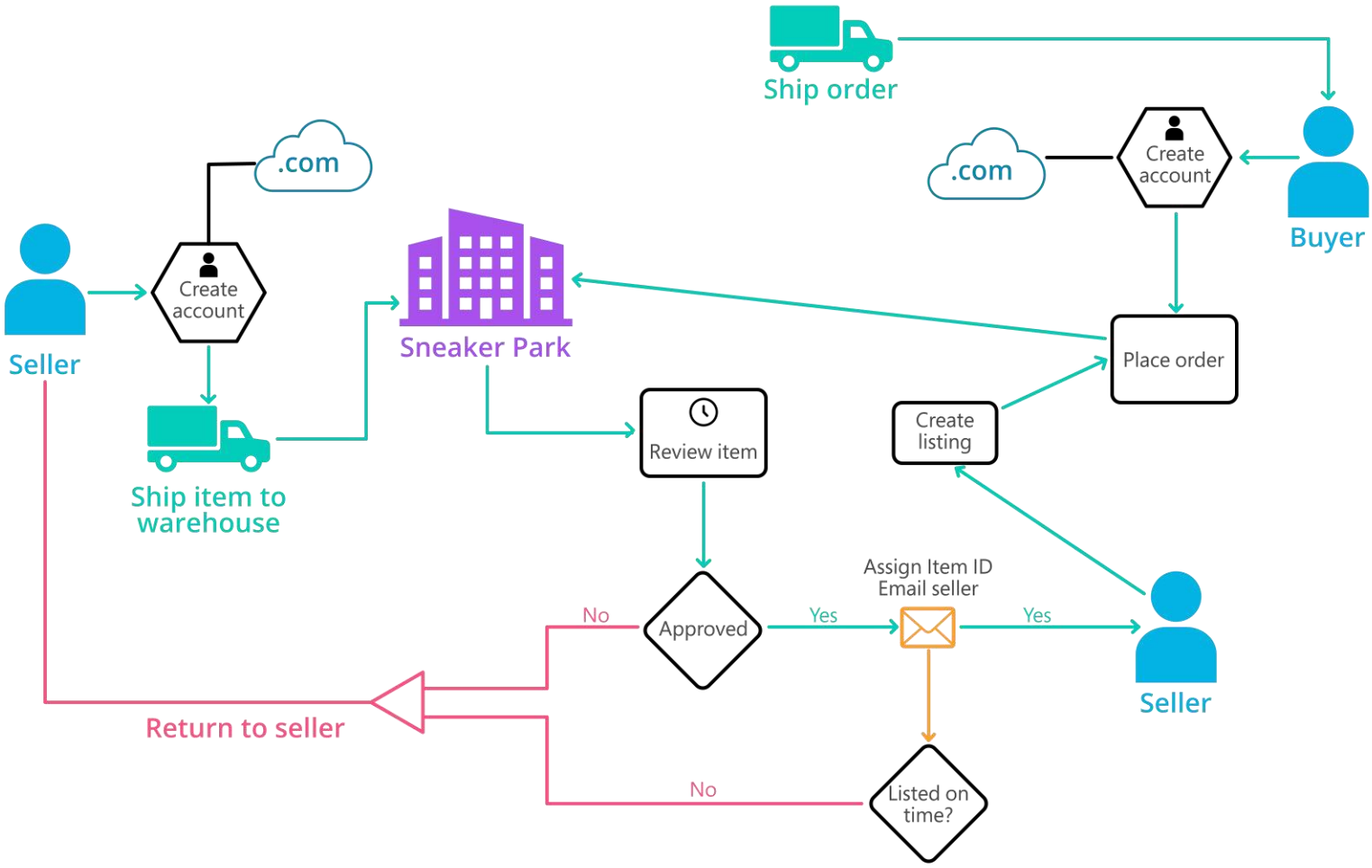
Submitted on: 30.12.2024

Background

- **SneakerPark** is an online shoe reseller that allows people to buy and sell used and new shoes. Buyers can bid for shoes or buy them outright, and sellers can set a price or sell to the highest bidder.
- Each buyer and seller must have an active account in order to sell, bid, or purchase sneakers using SneakerPark's website.
- SneakerPark authenticates the shoes before shipping them to the buyer, so before listing an item, the seller must ship it to SneakerPark's warehouse. Upon receipt, SneakerPark assigns an item number to each pair of sneakers and notifies the seller that they are now free to list their item. If the item is not listed within 45 days, SneakerPark returns it to the seller and sends an invoice to the seller for the shipping cost.
- If the item is found to be inauthentic or in an unacceptable condition, it is also returned back to the seller in a similar fashion.
- When the item sells, the buyer's account is credited with the purchase price minus the SneakerPark service fee and shipping fees to deliver the item to the buyer.
- Currently, SneakerPark only supports sales within the United States.

Background (cont'd)

- Below you can see a diagram that will hopefully help you visualize some of SneakerPark's business processes. Keep in mind that it does not capture ALL processes and every nuance, but simply serves as another artifact to use in your project.

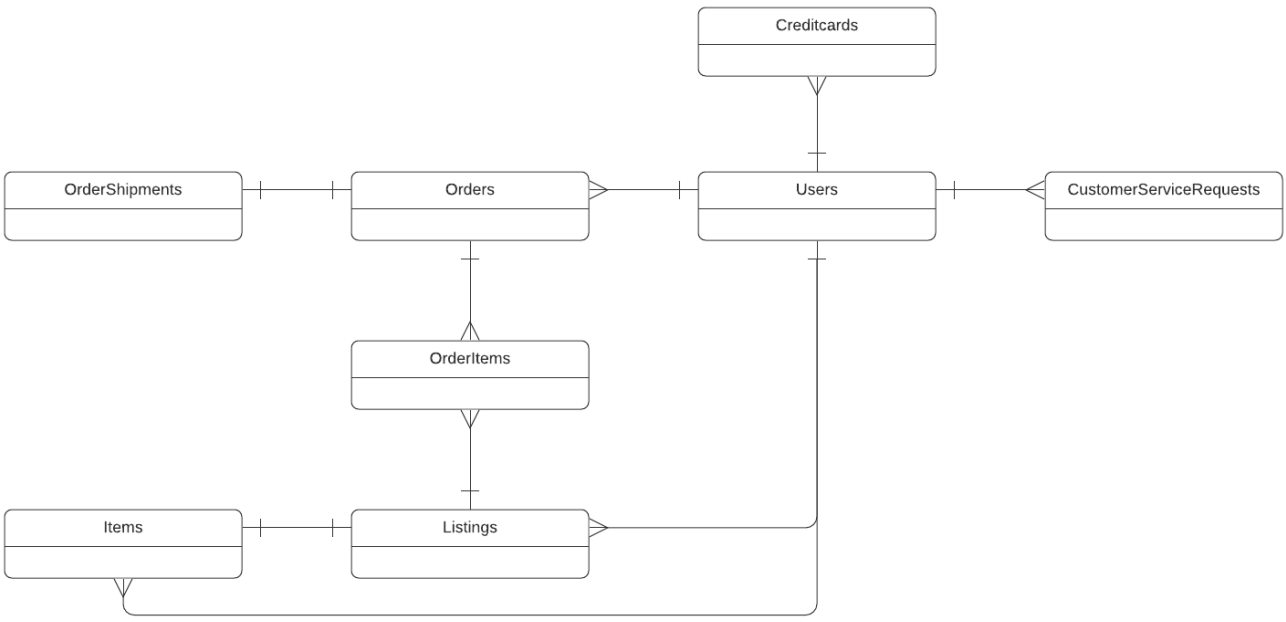


Step 1

Enterprise Data Catalog

Part 1: Enterprise Data Model

Conceptual SneakerPark Diagram



Step 2

Enterprise Data Catalog

Part 2: Metadata

First version of the Metadata Catalog created by documenting the metadata from all systems in the "Data Dictionary" and the “Enterprise Data Catalog” tabs is provided in this file.

Step 3

Data Quality

Part 1: Profiling and Cleansing

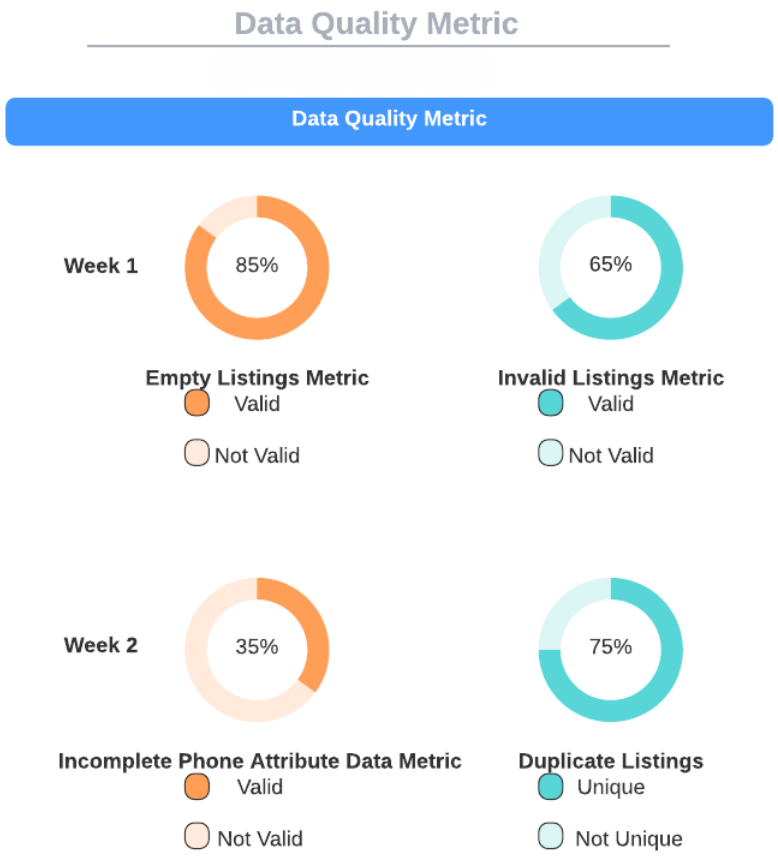
All identified "Data Quality Issues" issues can be found under this file.

Step 4

Data Quality

Part 2: Monitoring

The graph below was created based on this ‘Data Quality Metric’.

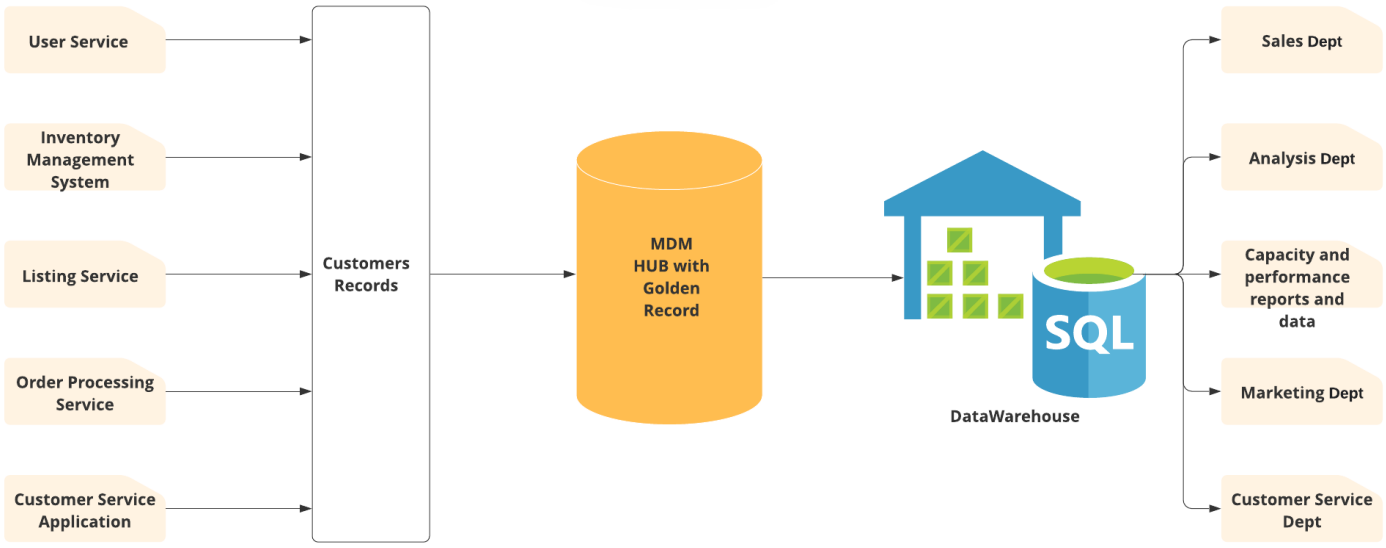


Step 5

Master Data Management

Part 1: MDM Architecture

Consolidated Master Data Management Architecture



Explanation:

I've chosen to go with Consolidated style of architecture. SneakerPark has a number of disparate systems that aggregate sellers and buyers order data which makes it difficult to trace the same customer across their order history, customers support and for example targeted marketing campaign. SneakerPark company wants to invest in an Enterprise Data Management program to gain better visibility and control over their data systems. On the other hand, it would be useful to make use of the analytical skills of staff to carry out analysis of the data collected. For that reason, I believe consolidating data from all these systems into a central MDM Hub that can feed merged, cleansed, and verified customer golden records to data warehouses and reports is the best solution here. The proposed solution provides the highest quality master data, furthermore enabling the business user to identify data analysis problems and then create rules and corrective processes. The designed MDM does not yet provide the full capability to prevent problems at the source - at the data origins. However, this will be partially possible where data quality analysis enables diagnosis of causes and easy remediation. The proposed solution, which aims to manage operational master data used by transactional applications. It relies heavily on integration technologies. As a result of operational MDM, the need for additional analytical MDM mechanisms is eliminated or very significantly reduced, which translates into simplification of analytical layers, such as Data Warehouse or Business Intelligence. It is an IT solution that performs the functions of consolidation, synchronisation and unification of master data processed in different systems of an organisation. Advanced algorithms for merging and improving data quality automate the process of data integration. The result is a centralised repository constituting a reliable source of information necessary for effective business management.

Step 6

Master Data Management

Part 2: Master Record

Customer:

- User with the same userID
- User with the same email address. This could be improved with a fuzzy rule because e.g. ‘gmail’ ignores dots in the name (abc@gmail.com = a.b.c@gmail.com) and many many mail servers ignore all information that comes after a ‘+’ sign, which some users use for provider assignment (abc+sneakerpark@gmail.com = abc@gmail.com)

Item:

- Items with the same listingId
- Items with the same combination of sellerID, brand, type, gender, color, condition and size and where the listingcreatdate is after the arrivaldate

Step 7

Data Governance:

Roles and Responsibilities

Data governance ensures data is properly classified, accessed, protected, and used. It also involves establishing strategies and policies to ensure the data lake processing environment complies with necessary regulatory requirements. Such policies also verify data quality and standardization to ensure the data is properly prepared to meet the needs of your organization. **SneakerPark** data governance policies define access and control of personal identifiable information (PII). The types of information that fall under these specific guidelines include credit card information, names, date of birth, and other such data. Implementing effective governance early in our MDM solution process will help avoid potential pitfalls, such as poor access control and metadata management, unacceptable data quality, and insufficient data security. Data governance isn't a technology. Rather, it's an organizational commitment that involves people, processes, and tools. Here are some roles that we should consider using in our system:

- Data Stewards:
    - Establish a core team of stakeholders like Jessica who is working in our company for some time, and data stewards to create a data governance framework. This begins with an audit to identify issues with current data management policies and areas needing improvement.

- Data Quality Assurance
    - Define the problems we are trying to solve: better regulatory compliance, increased data security, and improved data quality. And determine what we need to change, such as fine-tuning access rights, protecting sensitive data, or consolidating data silos. Assess tools and skills in your team will need to execute the data governance program. This may include people with skills in data modeling, data cataloging, data quality, and reporting. Due to the lack of suitable persons in the current team of the company. I recommend starting the recruitment process for the above mentioned positions immediately.
- Data Administrator:
    - The primary task of this role will be to inventory our data to see what we have, how it’s classified, where it resides, who can access it, and how it is used. Identify capabilities and gaps. Then figure out how to fill those gaps by hiring in-house specialists or by using a partner.