Data Governance @ SneakerPark

Prepared by: Osama Alsubaie

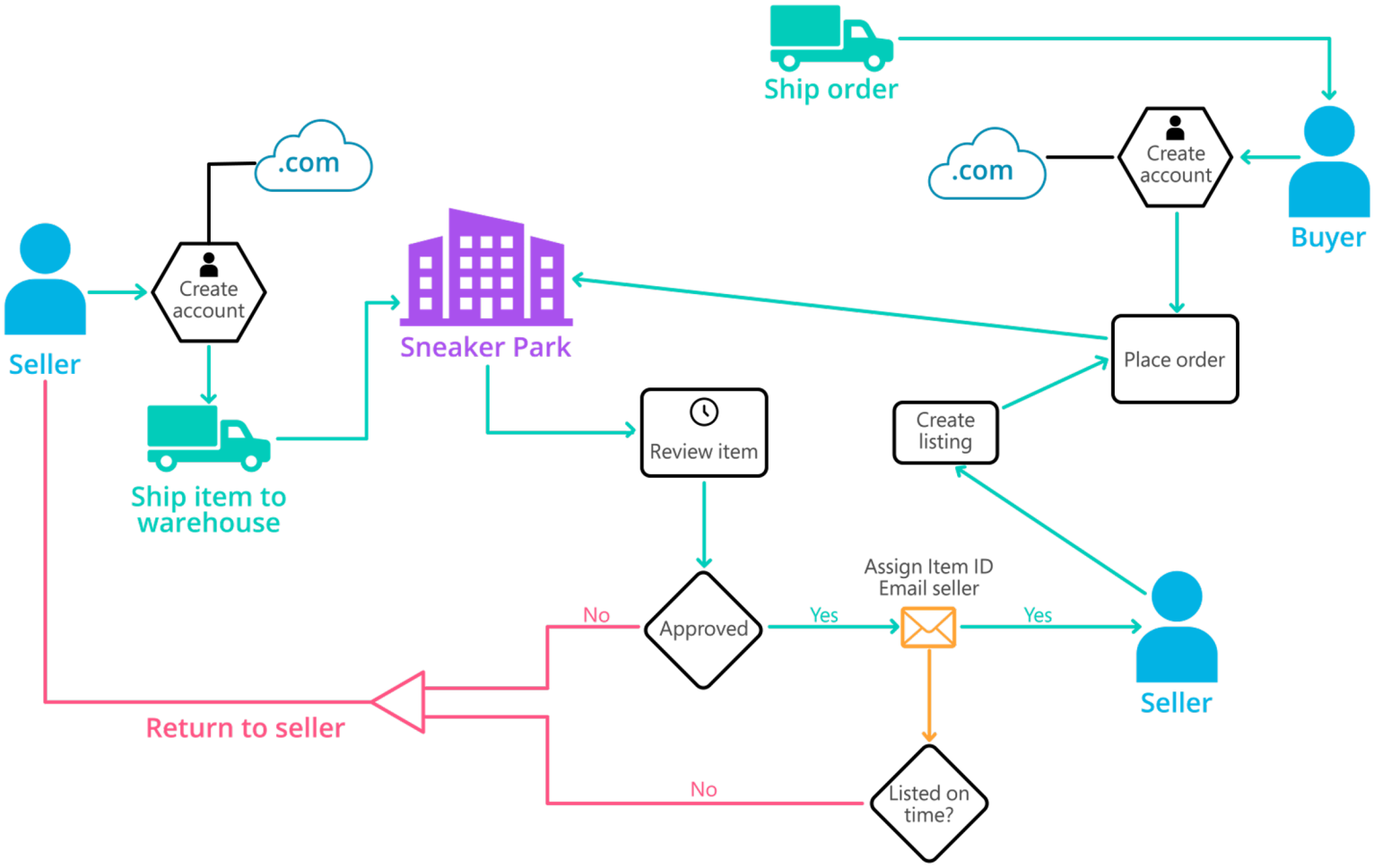
Submitted on: 28/11/2023

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.



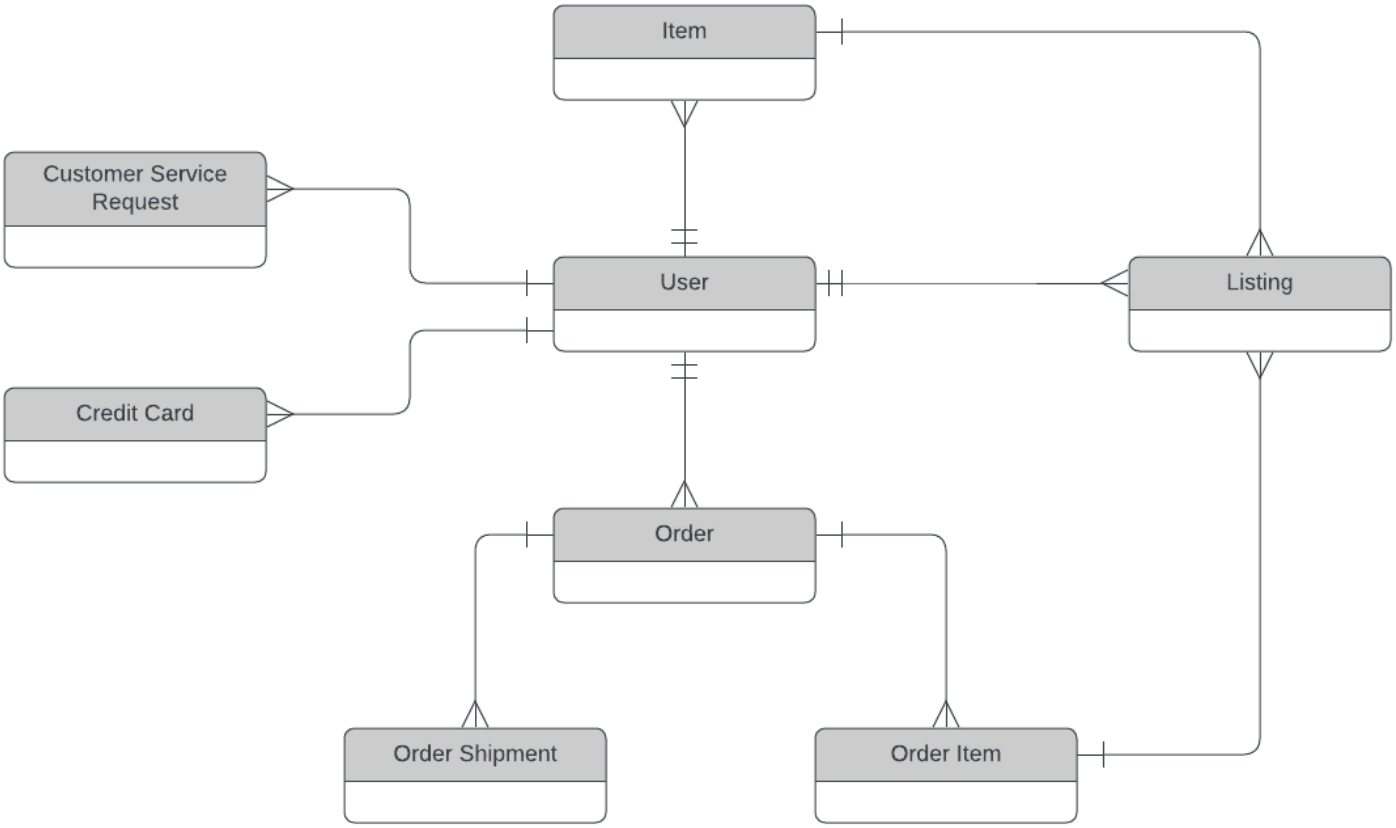
Step 1

Enterprise Data Catalog

Part 1: Enterprise Data Model

Create a **conceptual** data model that will provide SneakerPark with a holistic view of its data systems and help you grasp the organization's **important entities and relationships**, which will be instrumental as you move further in the project. You can use Lucidchart or any other diagramming tool of your choice, but please use the Crow's Foot/IE Notation and please be sure to indicate both cardinality (the type of a relationship such as 1:N or N:N) and optionality (whether the relationship is optional or mandatory).

Conceptual Data Model



Step 2

Enterprise Data Catalog

Part 2: Metadata

Metadata Catalog [“Data Dictionary”]

Entity	Source System	Table Name	Column Name	Data Type	Required	Unique	Description	Value Example	Primary Key	Foreign Key	Foreign Key Table	Foreign Key Column
Users	Customers	users	UserId	INT	Y	N	User's unique identification	80527	Y	N	N/A	N/A
Users	Customers	users	FirstName	VARCHAR(50)	Y	N	User's first name.	Emerson	N	N	N/A	N/A
Users	Customers	users	LastName	VARCHAR(50)	Y	N	User's last name.	Wire	N	N	N/A	N/A
Users	Customers	users	Email	VARCHAR(50)	Y	N	User's e-mail address.	emerson.wire@netscape.com	N	N	N/A	N/A
Users	Customers	users	Address	VARCHAR(50)	Y	N	User's address location	2 Harris Place	N	N	N/A	N/A
Users	Customers	users	ZipCode	VARCHAR(10)	Y	N	Address zip code.	13835	N	N	N/A	N/A
CreditCards	Customers	creditcards	CreditCardID	INT	Y	Y	Credit card identification	9467	Y	N	N/A	N/A
CreditCards	Customers	creditcards	CreditCardNumber	VARCHAR(50)	Y	N	Credit card number	99658816711200	N	N	N/A	N/A
CreditCards	Customers	creditcards	CreditCardExpirationDate	DATE	Y	N	Credit card expiration date	2023-11-01	N	N	N/A	N/A
CreditCards	Customers	creditcards	UserID	INT	Y	Y	Identifies the owner of the credit card	96342	N	Y	users	UserId
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	ID	INT	Y	Y	Service request identification	822950	Y	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	UserID	INT	Y	N	Identifies who is the user requesting services	3586	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	FirstName	VARCHAR(50)	Y	N	First name of the user	Bobby	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	LastName	VARCHAR(50)	Y	N	Last name of the user	Vamderheydem	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	ContactReason	VARCHAR(50)	Y	N	Reason of the request	Return	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	Email	VARCHAR(50)	N	N	E-mail address of the user	bobby.vanderheyden@fakeemail.com	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	Phone	VARCHAR(50)	N	N	Phone number of the user	(594) 811-5626	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	OrderID	INT	N	N	Order Id related to the request	12802	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	Resolution	VARCHAR(50)	Y	N	What was the resolution of the request	Provided Information	N	N	N/A	N/A
CustomerServicersRequests	CustomerServiceApplication	customerservicersrequests	ContactMethod	VARCHAR(50)	Y	N	Method of contact	Phone	N	N	N/A	N/A
Orders	OrderProcessingService	orders	OrderID	INT	Y	Y	Order unique identification	29692	Y	N	N/A	N/A
Orders	OrderProcessingService	orders	BuyerID	INT	Y	N	Identifies who is the buyer of the order	7708	N	Y	users	UserID
Orders	OrderProcessingService	orders	CreditCardID	INT	Y	N	Identifies in which credit card the order was processed	79220	N	Y	creditcards	CreditCardID
Orders	OrderProcessingService	orders	ShippingCost	DECIMAL(5,2)	Y	N	Cost of transportation	15.8	N	N	N/A	N/A
Orders	OrderProcessingService	orders	TaxRatePercent	SMALLINT	Y	N	Taxes percent	9	N	N	N/A	N/A
Orders	OrderProcessingService	orders	TotalAmount	DECIMAL(8,2)	Y	N	Total amount involved in the order process	123.78	N	N	N/A	N/A
Orders	OrderProcessingService	orders	ShippingAddress	VARCHAR(100)	N	N	Address wher the product will be shipped to	8447 Collie Hill Way	N	N	N/A	N/A
Orders	OrderProcessingService	orders	ShippingZipCode	VARCHAR(10)	Y	N	Zip code of the address where the product will be shipped to	54601	N	N	N/A	N/A
Orders	OrderProcessingService	orders	OrderDate	TIMESTAMP	Y	N	Timestamp when the order was placed	44169.46123	N	N	N/A	N/A
Orders	OrderProcessingService	orders	Status	VARCHAR(50)	N	N	Status of the order	Shipped	N	N	N/A	N/A
OrderItems	OrderProcessingService	orderitems	OrderID	INT	Y	N	Order identification that relates to teh listing	29626	Y	Y	orders	OrderID
OrderItems	OrderProcessingService	orderitems	ListingID	INT	Y	N	Listing identification that relates to the order	998839	Y	Y	listings	ListingID
OrderItems	OrderProcessingService	orderitems	ListingSoldPrice	DECIMAL(8,2)	N	N	The price in which the product was sold	45	N	N	N/A	N/A
OrderShipments	OrderProcessingService	ordershipments	ShipmentId	INT	Y	Y	Shipment identification	797095	Y	N	N/A	N/A
OrderShipments	OrderProcessingService	ordershipments	OrderID	INT	Y	N	Order that relates to the shipment	427	N	N	N/A	N/A
OrderShipments	OrderProcessingService	ordershipments	Carrier	VARCHAR(50)	Y	N	Carrier that is transporting the product	USPS	N	N	N/A	N/A
OrderShipments	OrderProcessingService	ordershipments	TrackingNumber	VARCHAR(30)	N	N	Number to track the product	78C2EF3T5KYY7283	N	N	N/A	N/A
OrderShipments	OrderProcessingService	ordershipments	OrderShipDate	DATE	Y	N	Date when the product was shipped	2020-11-09	N	Y	orders	OrderID
Listings	ListingService	listings	ListingID	INT	Y	Y	Listing identification	493279	Y	N	N/A	N/A
Listings	ListingService	listings	SellerID	INT	Y	N	Identifies who is the seller responsible for the listing	58233	N	Y	users	UserID
Listings	ListingService	listings	ProductID	INT	Y	N	Which product is being listed	509	N	N	N/A	N/A
Listings	ListingService	listings	ShoeType	VARCHAR(50)	N	N	Type of the shoe being listed	Sandals or Flip Flops	N	N	N/A	N/A
Listings	ListingService	listings	Brand	VARCHAR(50)	N	N	Brand of the shoe	UnderArmor	N	N	N/A	N/A
Listings	ListingService	listings	Color	VARCHAR(15)	N	N	Color of the shoe	black	N	N	N/A	N/A
Listings	ListingService	listings	Gender	CHAR(1)	N	N	Gender of the shoe	F	N	N	N/A	N/A
Listings	ListingService	listings	Size	VARCHAR(4)	N	N	Size of the shoe	12	N	N	N/A	N/A
Listings	ListingService	listings	Condition	VARCHAR(50)	Y	N	Tells if the product is acceptable for selling on sneakerpark	Used	N	N	N/A	N/A
Listings	ListingService	listings	ListingPrice	DECIMAL(8,2)	Y	N	Listing price	83.00	N	N	N/A	N/A
Listings	ListingService	listings	ListingType	VARCHAR(20)	Y	N	If it is a fixed price or an auction	Action	N	N	N/A	N/A
Listings	ListingService	listings	ListingCreateDate	DATE	Y	N	Start date of listing	2022-11-19	N	N	N/A	N/A
Listings	ListingService	listings	ListingEndDate	DATE	Y	N	End date of listing	2022-12-30	N	N	N/A	N/A
Items	InventoryManagementSystem	items	ItemID	INT	Y	Y	Item identification	2333	Y	N	N/A	N/A
Items	InventoryManagementSystem	items	ItemName	VARCHAR(100)	Y	N	The name of the item	Gisela	N	N	N/A	N/A
Items	InventoryManagementSystem	items	SellerID	INT	Y	N	Identifies who is selling the product	99900	N	N	N/A	N/A
Items	InventoryManagementSystem	items	Type	VARCHAR(50)	Y	N	Type of the item	Sneakers	N	N	N/A	N/A
Items	InventoryManagementSystem	items	BrandName	VARCHAR(100)	Y	N	Brand of the item	Puma	N	N	N/A	N/A
Items	InventoryManagementSystem	items	Color	VARCHAR(15)	Y	N	Color of the item	white	N	N	N/A	N/A
Items	InventoryManagementSystem	items	Size	VARCHAR(4)	Y	N	Size of the item	14	N	N	N/A	N/A
Items	InventoryManagementSystem	items	Sex	VARCHAR(10)	Y	N	Gender target of the product	Male	N	N	N/A	N/A
Items	InventoryManagementSystem	items	Condition	VARCHAR(50)	Y	N	Tells if the product is acceptable for selling on sneakerpark	Like new	N	N	N/A	N/A
Items	InventoryManagementSystem	items	ItemStatus	VARCHAR(50)	N	N	Tells at what step of the sell is the product	approved	N	N	N/A	N/A
Items	InventoryManagementSystem	items	ArrivalDate	DATE	N	N	When did it land on the warehouses	2022-10-10	N	N	N/A	N/A

*Please Review ‘SneakerPark Templates.xlsx’ File To See The Solution Clearly

Metadata Catalog [“Enterprise Data Catalog”]

Table	Data Domain	Criticality	Retention Policy	Security Classification	Data Steward
Users	Customer	High	7 years	Highly Confidential	Jessica
CreditCards	Customer	High	7 years	Highly Confidential	Jessica
CustomerServicersRequests	Customer	High	7 years	Highly Confidential	Jessica
Orders	Orders	High	7 years	Internal	Jessica
OrderItems	Orders	High	7 years	Internal	Jessica
OrderShipments	Orders	High	7 years	Internal	Jessica
Listings	Listing	Mid-High	2 years	Internal	Jessica
Items	Inventory	Medium	N/A	Internal	Jessica

*Please Review ‘SneakerPark Templates.xlsx’ File To See The Solution Clearly

Step 3

Data Quality

Part 1: Profiling and Cleansing

Profile the data to identify at least **3 data quality issues** you see in the data. Also provide **at least 1 data quality issue that you haven't yet seen** in the data, but can foresee occurring in the future. Based on the issues you've identified, come up with the data quality rule for each data quality issue, including for the one that you foresee.

*Please Review 'SneakerPark Templates.xlsx' File To See The Solution Clearly

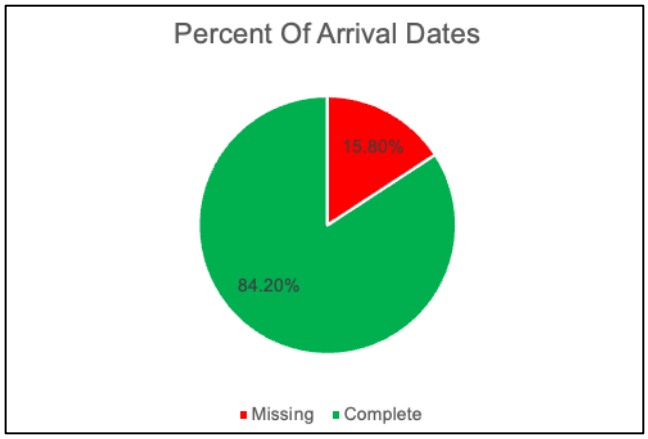
Step 4

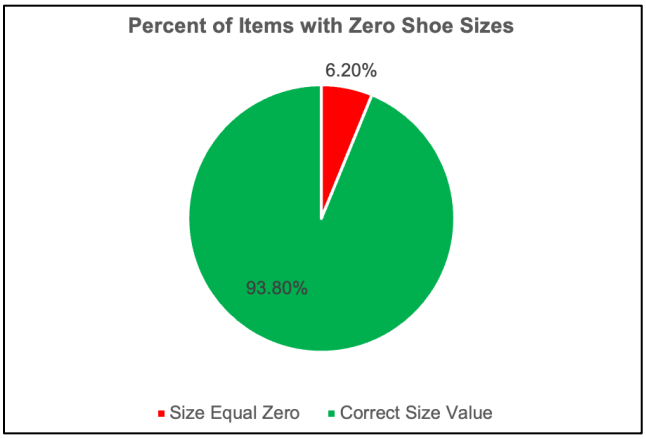
Data Quality

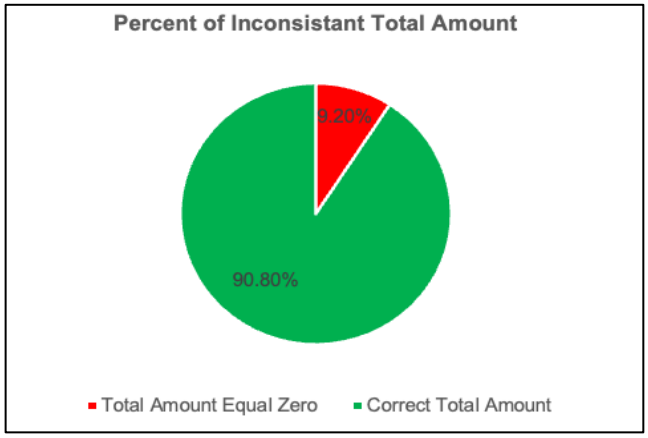
Part 2: Monitoring

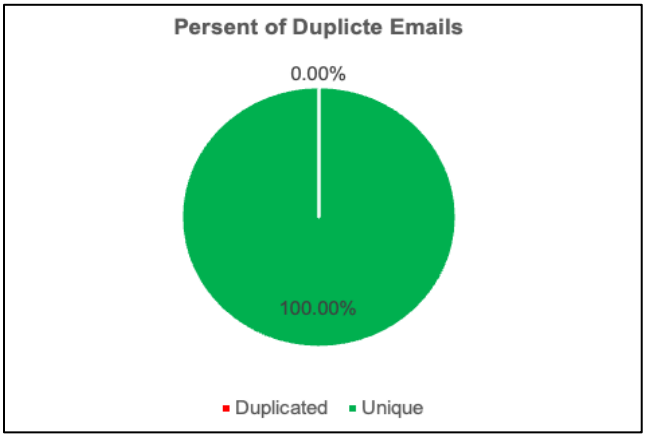
Using the metrics you've created in the last step, please create a mock-up of a data quality **monitoring dashboard** that will be used to monitor the data to ensure compliance with your data quality rules.

Please **make sure to label your metrics clearly** on your mock-up.







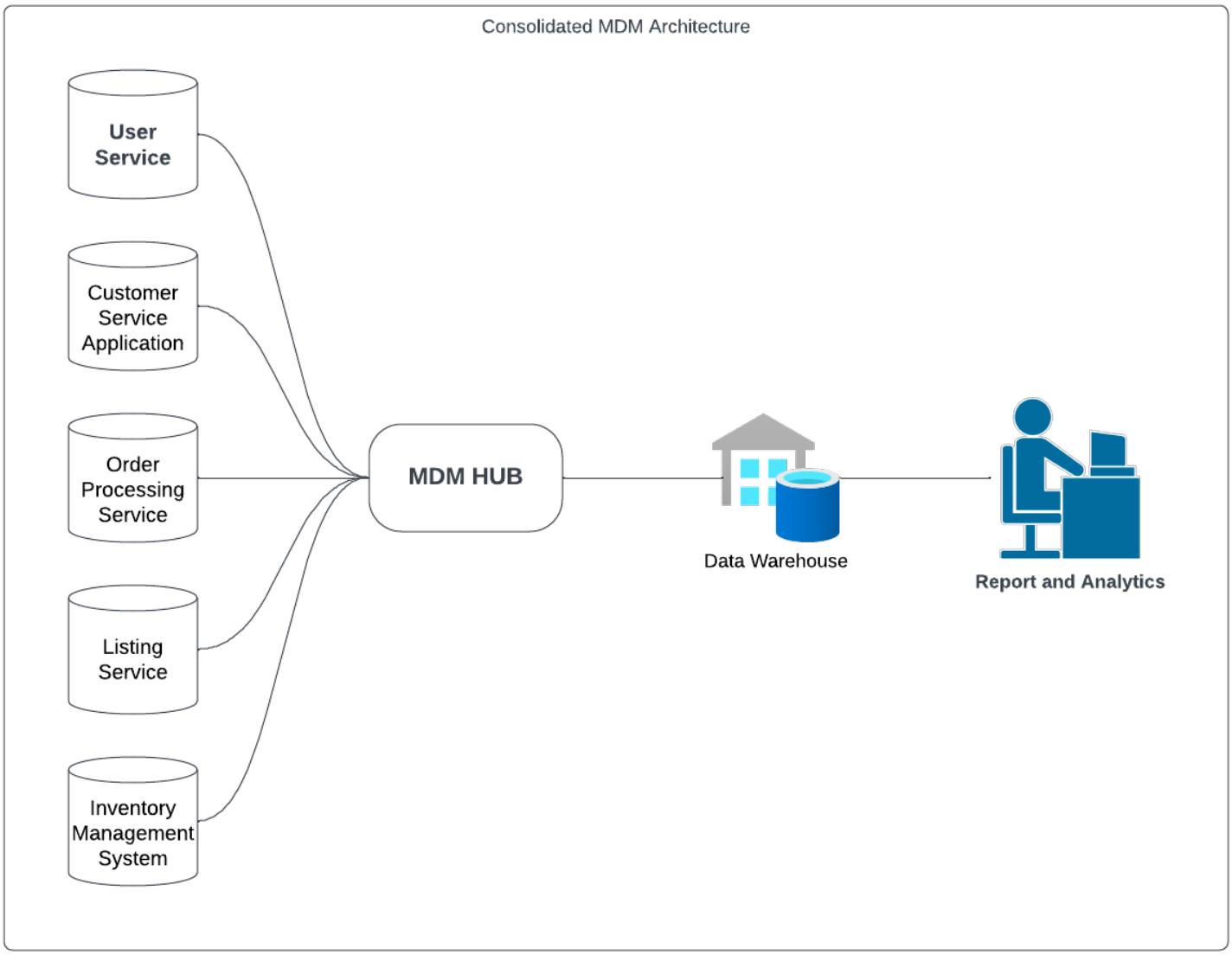


Step 5

Master Data Management

Part 1: MDM Architecture

Based on what you've read about SneakerPark's systems and business model, sketch out a proposed **MDM implementation architecture**, and write a **detailed explanation** of **why** you chose this specific approach.



SneakerPark's data is distributed across multiple sources, including User Service, Inventory Management Service, Listing Services, Order Processing Service, and Customer Service Application. However, data inconsistencies are expected due to the isolation of Inventory Management Service and Customer Service Application from other systems. As these systems hold crucial master data like customer and asset information, there's a need for a Master Data Management (MDM) system. Considering SneakerPark's business nature, implementing a Consolidated MDM architecture is proposed. This approach minimizes disruption to existing systems and is cost-effective, ensuring the generation of accurate golden records without major impact on the current setup.

Step 6

Master Data Management

Part 2: Master Record

In this step, you will define a set of **matching rules** that will be used by the SneakerPark's MDM Hub to match item and customer entities between the company's different systems.

Please come up with 4 rules - 2 for Items and 2 for Customers and list them below.

1- For the Items, Matching the brand name from the items table in the inventory management system with the listing table in the listing service system.

Query	Result

```
-- Explore the tables
SELECT * FROM im.items
SELECT * FROM li.listings


-- Match the brand name
SELECT i.brandname, l.brand
FROM im.items i
JOIN li.listings l
ON i.itemid = l.productid
```

	brandname character varying (100)	brand character varying (50)
1	Under Armour	UnderArmor
2	Under Armour	UnderArmor
3	Nike	Nike
4	Nike	Nike
5	Nike	Nike
6	Johnston & Murphy	Johnston and Murphy
7	Fila	Fila
8	Fila	Fila
9	Fila	Fila
10	Reebok	Reebok
11	Reebok	Reebok
12	Reebok	Reebok
13	Reebok	Reebok
14	New Balance	NewBalance
15	Nike	Nike
16	Skechers	Skechers
17	Skechers	Skechers
18	Reebok	Reebok
19	New Balance	New Balance
20	New Balance	New Balance
21	Berluti	Berlutti
22	New Balance	New Balance
23	New Balance	New Balance

In this step, you will define a set of **matching rules** that will be used by the SneakerPark's MDM Hub to match item and customer entities between the company's different systems.

Please come up with 4 rules - 2 for Items and 2 for Customers and list them below.

2- For the Items, matching the Condition of the Listing on the Listing Services System and the Condition of the Item from the Inventory Management Service System

Query	Result

```
-- Explore the tables
SELECT * FROM im.items
SELECT * FROM li.listings

-- Match the brand name
SELECT i.condition, l.condition
FROM im.items i
JOIN li.listings l
ON i.itemid = l.productid
```

	condition character varying (50)	condition character varying (50)
1	used	Used
2	used	Used
3	like new	Open Box
4	like new	Open Box
5	like new	Open Box
6	used	Used
7	like new	Open Box
8	like new	Open Box
9	like new	Open Box
10	new	New
11	new	New
12	new	New
13	new	New
14	new	New
15	like new	Open Box
16	like new	Open Box
17	like new	Open Box
18	new	New
19	like new	Open Box
20	like new	Open Box
21	like new	Open Box
22	new	New
23	new	New

In this step, you will define a set of **matching rules** that will be used by the SneakerPark's MDM Hub to match item and customer entities between the company's different systems.

Please come up with 4 rules - 2 for Items and 2 for Customers and list them below.

3- For the Customers, Matching the address from the users table in the user service system with the orders table in the order processing service system.

Query	Result

```
-- Explore the tables
SELECT * FROM usr.users
SELECT * FROM op.orders

-- Match the address
SELECT u.address, o.shippingaddress
FROM usr.users u
JOIN op.orders o
ON u.userid = o.buyerid
```

	address character varying (50)	shippingaddress character varying (100)
1	8447 Collie Hill Way	8447 Collie Hill Way
2	4 Shemya Drive	4 Shemya Drive
3	52 Kingfisher Drive	52 Kingfisher Drive
4	106 Garden Square	106 Garden Square
5	7038 Graham Place	7038 Graham Place
6	5 Berryhill Crescent	5 Berryhill Crescent
7	0 Courage Drive	0 Courage Drive
8	3873 Seawolf Drive	3873 Seawolf Drive
9	3873 Seawolf Drive	3873 Seawolf Drive
10	3 Dalshannon Place	3 Dalshannon Place
11	3 Dalshannon Place	3 Dalshannon Place
12	089 Hamlet Circle	089 Hamlet Circle
13	089 Hamlet Circle	089 Hamlet Circle
14	5252 Smithstone Crescent	5252 Smithstone Crescent
15	6 Connors Trail Circle	6 Connors Trail Circle
16	8677 Hawthorn Terrace	8677 Hawthorn Terrace
17	4005 Brandy Circle	4005 Brandy Circle
18	38 Geronimo Circle	38 Geronimo Circle
19	1 Scott Street	1 Scott Street
20	91 Beach Circle	91 Beach Circle
21	91 Beach Circle	91 Beach Circle
22	9 West 70th Avenue	9 West 70th Avenue
23	9 West 70th Avenue	9 West 70th Avenue

In this step, you will define a set of **matching rules** that will be used by the SneakerPark's MDM Hub to match item and customer entities between the company's different systems.

Please come up with 4 rules - 2 for Items and 2 for Customers and list them below.

4- For the Customers, Match the SellerID from the Listing on the Listing Services System and the SellerID in the Item table from the Inventory Management Service System.

Query	Result

`-- Explore the tables`
`SELECT * FROM usr.users`
`SELECT * FROM im.items`
 	
`-- Match the SallerID`
`SELECT COUNT(DISTINCT(u.userid)) "UserID Count",`
`COUNT(DISTINCT(i.sellerid)) "SellerID Count"`
`FROM usr.users u`
`JOIN im.items i`
`ON u.userid = i.sellerid`
	

Step 7

Data Governance:

Roles and Responsibilities

Write 1-2 paragraphs discussing what **data governance roles and responsibilities** will be necessary to oversee this new Data Management initiative. Please be sure to discuss the responsibilities in the context of at **least 3 different aspects** of Data Governance (such as Data Quality Management, Metadata Management, MDM, etc). Based on what you know, do SneakerPark's **current employees have the necessary skills** for these roles or should the company **make new hires**?

Data Architect: As the Data Architect, Daniel Freitas will define the workflow for Data Quality Management and Metrics. Additionally, Daniel will be responsible for designing the Master Data Management architecture to automate the identification of data breaches and quality issues. His robust experience in constructing such platforms makes him an ideal candidate for this pivotal role.

Data Steward: The individual in the Data Steward role holds responsibility for maintaining SneakerPark's Metadata Management by updating the data dictionary's business context. This involves accommodating new additions, such as tables or columns, and incorporating any changes in business descriptions. Jessica, due to her comprehensive understanding of the organization, is a natural fit for this role. To manage workload efficiently, Jessica will be joined by a new hire who will share responsibilities.

Data Engineer: The Data Engineer will oversee IT production support concerning Data Quality Management and Master Data Management. Ensuring consistency, accuracy, and timeliness of data in production is key. Jake, with a background in IT Support ownership, is well-suited for this role. However, Jake will undergo specialized training in data-related matters to ensure top-tier support.

Standout Suggestions

1. Create a Business Glossary for SneakerPark and define common terms such as Item, Buyer, etc. Think and discuss how SneakerPark can improve on the consistency of the terms that its systems currently use. (You can use the “Business Glossary” tab of the same Sheets template you’ve been using for the other parts of this project to get you started.)

1. Document SneakerPark’s current naming conventions. Can you think of any improvements? (You can use the “Standard Naming Conventions” tab of the same Sheets template you’ve been using for the other parts of this project to get you started.) Some examples of Naming Conventions include;

    - Do not use spaces or special characters.
    - Use only LOWERCASE.
    - All identifier fields should end in “_id”.
    - Avoid acronyms and abbreviations.

1. Write SQL scripts for the matching rules that you’ve created in Step 6.