Intro to Data Management Project

The Home Depot

By Bhavna Kaparaju, Brett Nesfeder,
Callie Gilmore & Dawson Cook

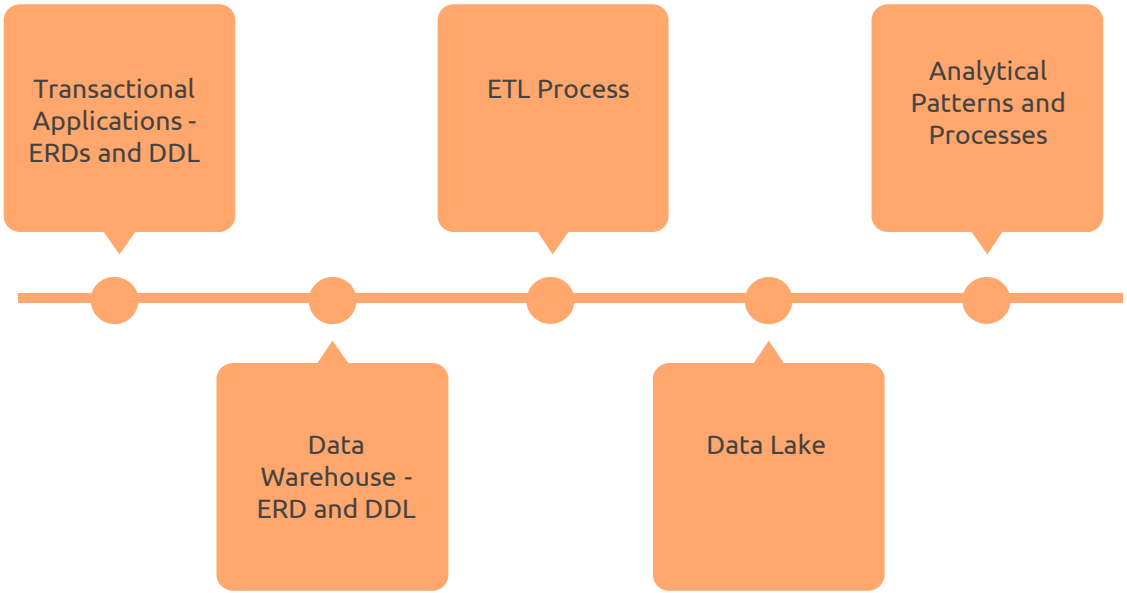
The Home Depot

- The Home Depot is the largest home improvement retailer in America
- They supply tools, construction products, and home improvement services.
- Founded in 1978 in Atlanta, Georgia
- Currently has 2,278 stores across North America
- WIth a large amount of stores and locations, The Home Depot deals with a huge amount of data on a daily basis.

Data Strategy

- Offensive Strategy
  - Highly Competitive Market (Lowe’s, Walmart, Amazon)
- Focus on supporting revenue growth
- Generate customer insights
- Integrate customer and market data to support managerial decision making

Steps


Transactional Applications - ERDs and DDL
ETL Process
Analytical Patterns and Processes
Data Warehouse - ERD and DDL
Data Lake

Transactional Applications

3 Transactional Applications

Customer POS

Point of sale system to track customer purchases

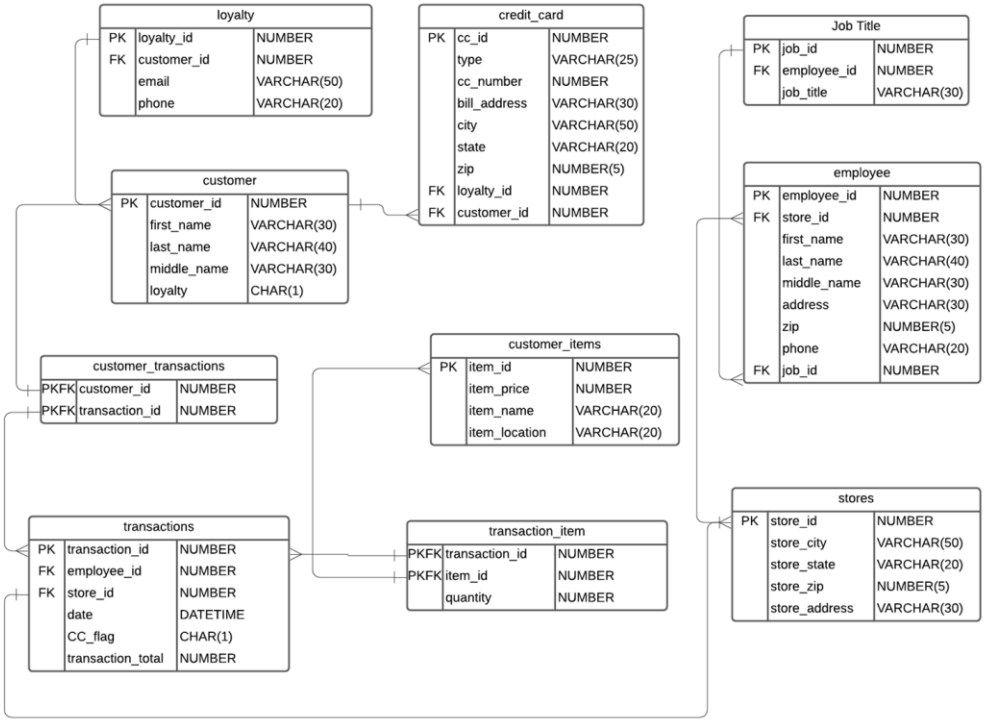
Inventory

Inventory system to keep track of stock and determine when to re-purchase

Wholesale Purchasing

Wholesale purchasing to re-new stock when necessary

Customer POS ERD



Customer DDL & Inserts

	CUSTOMER_ID	FIRST_NAME	MIDDLE_NAME	LAST_NAME	LOYALTY
1	1	Jeffrey	John	Jacobs	Y

	CUSTOMER_ID	TRANSACTIONS_ID
1	1	1000

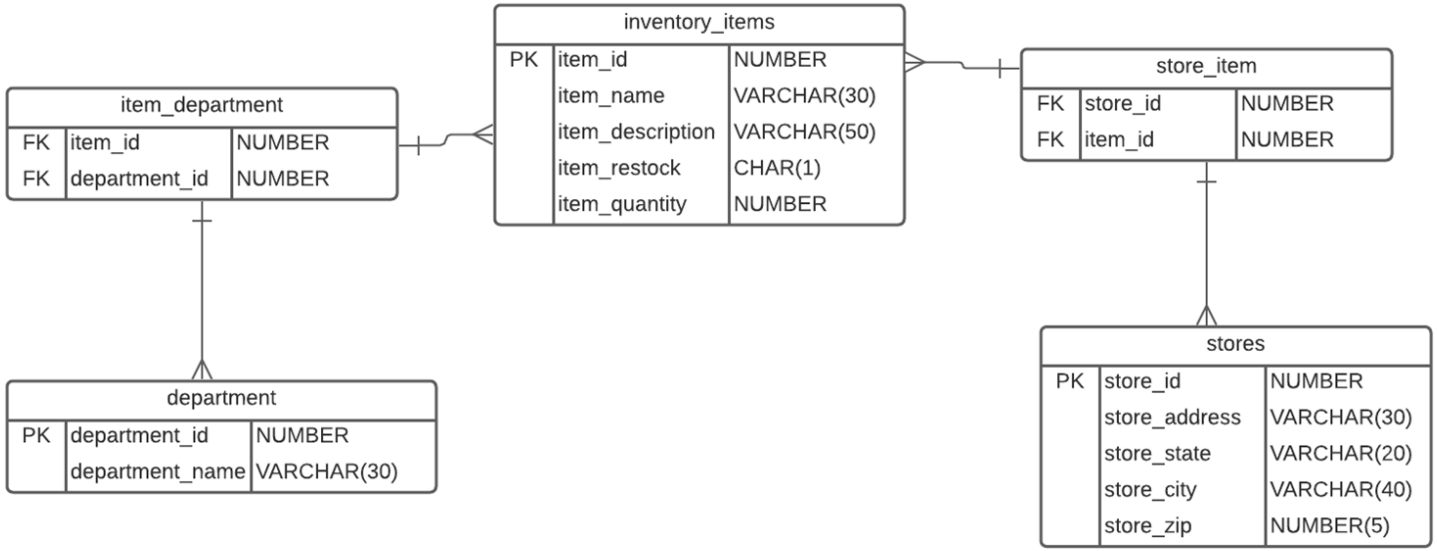
	STORE_ID	STORE_CITY	STORE_STATE	STORE_ZIP	STORE_ADDRESS
1	2000	Los Angeles	California	90210	1200 Home Depot Blvd

	TRANSACTIONS_ID	TRANSACTION_DATE	CREDIT_FLAG	TRANSACTION_TOTAL
1	1000	25-JUN-20	Y	6000

	TRANSACTIONS_ID	ITEM_ID	QUANTITY
1	1000	1000	2

	EMPLOYEE_ID	STORE_ID	FIRST_NAME	LAST_NAME	MIDDLE_NAME	ADDRESS	ZIP	PHONE	JOB_ID
1	1000	2000	Pete	Plunker	Patrick	23 Champion Dr	90210	5126157824	1000

Inventory ERD



Inventory DDL & Inserts

	STORE_ID	STORE_ADDRESS	STORE_CITY	STORE_STATE	STORE_ZIP
1	1000	1200 Home Depot Blvd	Los Angeles	California	90210

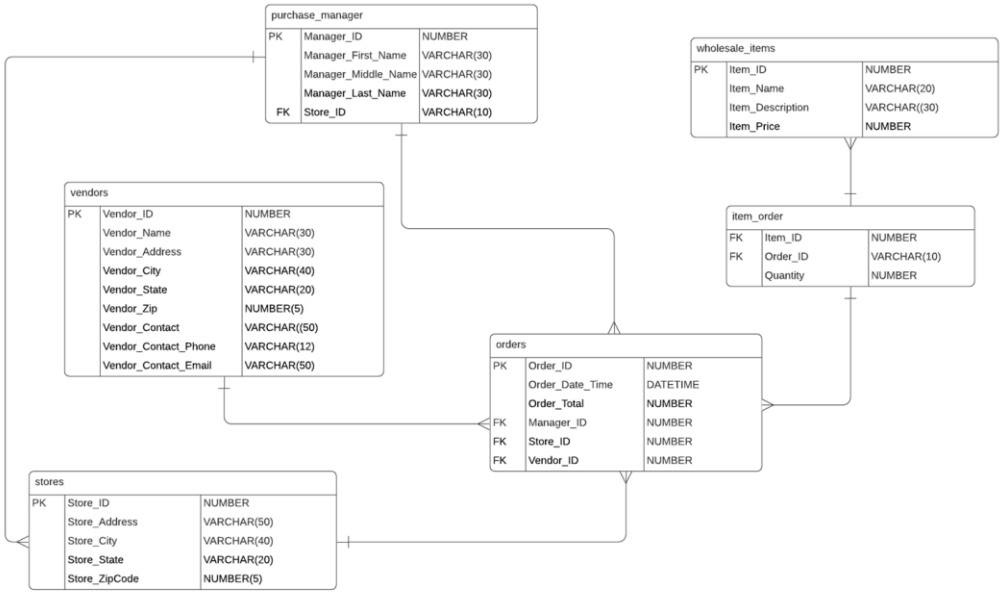
	DEPARTMENT_ID	DEPARTMENT_NAME
1	1000	Doorbells

	STORE_ID	ITEM_ID
1	1000	1000

	DEPARTMENT_ID	ITEM_ID
1	1000	1000

	ITEM_ID	ITEM_NAME	ITEM_DESCRIPTION	ITEM_QUANTITY	ITEM_RESTOCK
1	1000	Nest Hello Video Doorbell	A wired video doorbell	50	N

Wholesale Purchasing ERD



Wholesale Purchasing DDL & Inserts

	ITEM_ID	ITEM_NAME	ITEM_DESCRIPTION	ITEM_PRICE
1	1000	Lamp	Rainbow Colorful Lamp	100

	ORDER_ID	ITEM_ID	QUANTITY
1	1000	1000	200

	STORE_ID	STORE_CITY	STORE_STATE	STORE_ZIP	STORE_ADDRESS
1	3000	Los Angeles	California	90210	1300 Home Depot Ave

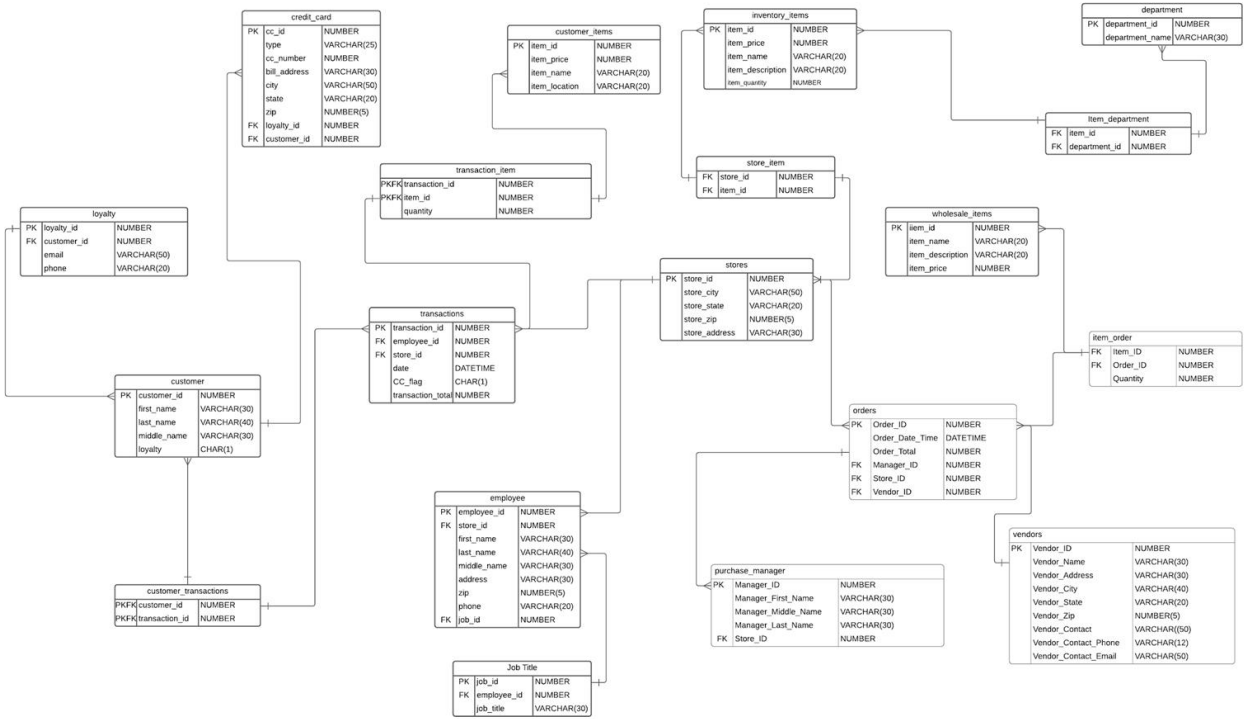
	MANAGER_ID	MANAGER_FIRSTNAME	MANAGER_MIDDLENAME	MANAGER_LASTNAME	STORE_ID
1	1000	Jeff	Keith	Adams	3000

	VENDOR_ID	VENDOR_NAME	VENDOR_ADDRESS	VENDOR_CITY	VENDOR_STATE	VENDOR_ZIP	VENDOR_CONTACT_NAME	VENDOR_CONTACT_PHONE	VENDOR_CONTACT_EMAIL
1	1000	Lamps Inc	101 Bunny Run	Los Angeles	California	90210	Bob Roberts	7816549087	bob.roberts@gmail.com

	ORDER_ID	ORDER_DATE_TIME	ORDER_TOTAL	MANAGER_ID	STORE_ID	VENDOR_ID
1	1000	25-SEP-14	4000	1000	3000	1000

Data Warehouse

Data Warehouse ERD



ETL Process

ETL - Store Example

```
CREATE TABLE stores_dw
(
    store_id                NUMBER          DEFAULT dw_storeid_sequence.NEXTVAL NOT NULL   PRIMARY KEY,
    store_city              VARCHAR(40)     NOT NULL,
    store_state             VARCHAR(20)     NOT NULL,
    store_zip               NUMBER(5)       NOT NULL,
    store_address           VARCHAR(30)     NOT NULL,
    data_source             VARCHAR(12)

);
```

Data Lake

Data Lake

```sql
%sql
SELECT store_zip, store_city, store_state, count(employee_id), v.vendor_name, v.vendor_city, v.vendor_state, v.vendor_zip, c.TotalPopulation, c.MedianAge, c.AverageHouseholdSize
FROM stores s
LEFT JOIN vendor v ON v.vendor_zip = s.store_zip
LEFT JOIN census c ON c.ZipCode = s.store_zip
LEFT JOIN employee e ON e.zip = s.store_zip
GROUP BY store_zip, store_city, store_state, v.vendor_name, v.vendor_city, v.vendor_state, v.vendor_zip, c.TotalPopulation, c.MedianAge, c.AverageHouseholdSize
ORDER BY count(employee_id);
```

▸ (2) Spark Jobs

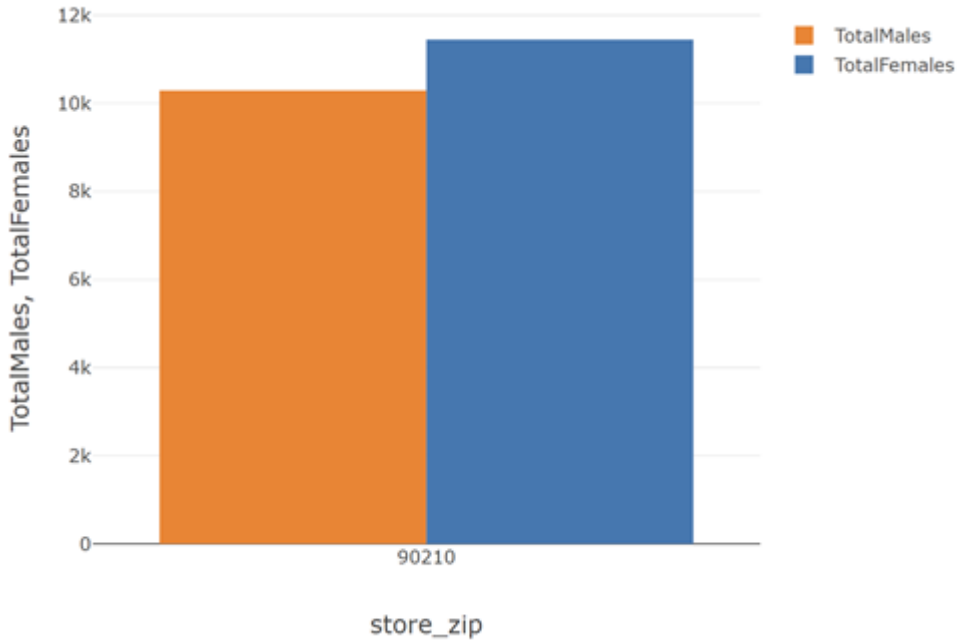
	store_zip	store_city	store_state	count(employee_id)	vendor_name	vendor_city	vendor_state	vendor_zip	TotalPopulation	MedianAge	AverageHouseholdSize
1	90210	Los Angeles	California	1	Lamps Inc	Los Angeles	California	90210	21741	47.5	2.49

Showing all 1 rows.

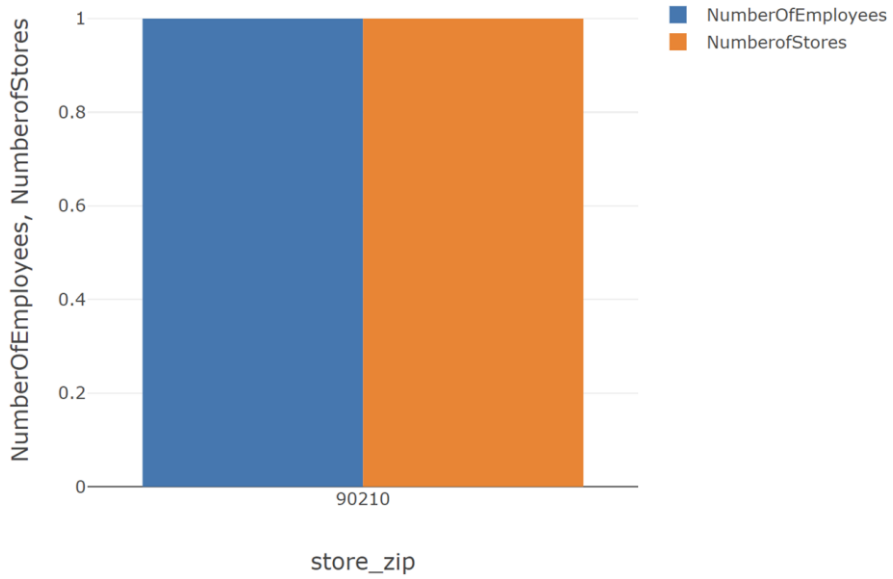
ZipCode	TotalPopulation	MedianAge	TotalMales	TotalFemales	TotalHouseholds	AverageHouseholdSize
91371	1	73.5	0	1	1	1.0
90001	57110	26.6	28468	28642	12971	4.4
90002	51223	25.5	24876	26347	11731	4.36
90003	66266	26.3	32631	33635	15642	4.22
90004	62180	34.8	31302	30878	22547	2.73
90005	37681	33.9	19299	18382	15044	2.5
90006	59185	32.4	30254	28931	18617	3.13
90007	40920	24.0	20915	20005	11944	3.0
90008	32327	39.7	14477	17850	13841	2.33
90010	3800	37.8	1874	1926	2014	1.87
90011	103892	26.2	52794	51098	22168	4.67
90012	31103	36.3	19493	11610	10327	2.12
90013	11772	44.6	7629	4143	6416	1.26
90014	7005	44.8	4471	2534	4109	1.34
90015	18986	31.3	9833	9153	7420	2.45
90016	47596	33.9	22778	24818	16145	2.93
90017	23768	29.4	12818	10950	9338	2.53

Analytical Patterns

Total Males vs Total Females by Zip Code



Total Employee Count vs Total Number of Stores



Conclusion

Final Thoughts

Corporate Data Environments are very complex

Clear & organized naming conventions are required to keep everything straight

Clear & concise documentation on what you did and how you did it will not only help yourself but will also help team members and overall organization

Value: Data only matters if you can access it. This project taught us the importance of maintaining complex data environments so that we can access it and perform analytics.

Going forward: practicing documentation, organization and clear naming conventions

Thanks!

Any Questions?