Amazon E-Commerce Services

Sidd Chauhan, Samir Epili, Ryan Rue,
Paridhi Sheth & Grayson Taylor

Why Amazon?

- Largest e-commerce retailer
- Size and Breadth of the enterprise
  - AmazonBasics
  - Amazon Fresh
  - Amazon Studios
  - Amazon Warehouse
- Industry built around constant innovation and change
- Understand organizational processes and various aspects of the company
- Write interesting queries to answer realistic questions (i.e. COVID -19 and amazon operations)

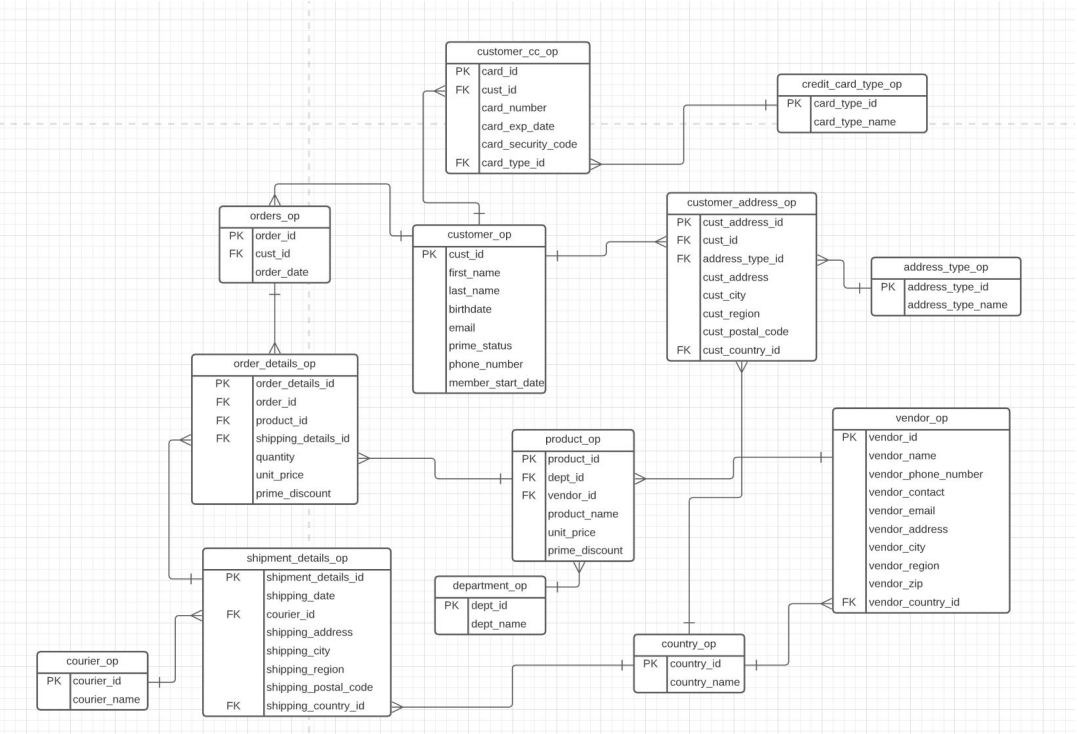
Data Strategy

- Primarily Offensive Strategy
- Aggressively expanding the business
- Architecture that can manage data volume and velocity

Transaction Management Applications

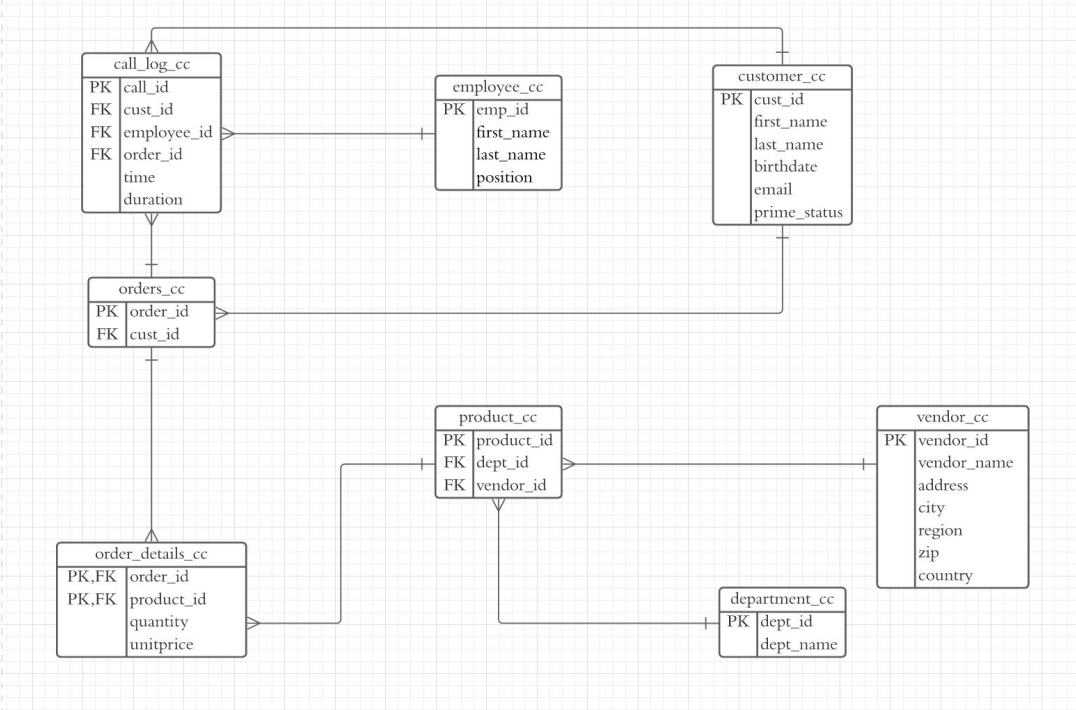
- Order Processing
- Call Center Operations
- Inventory Management

Order Processing ERD

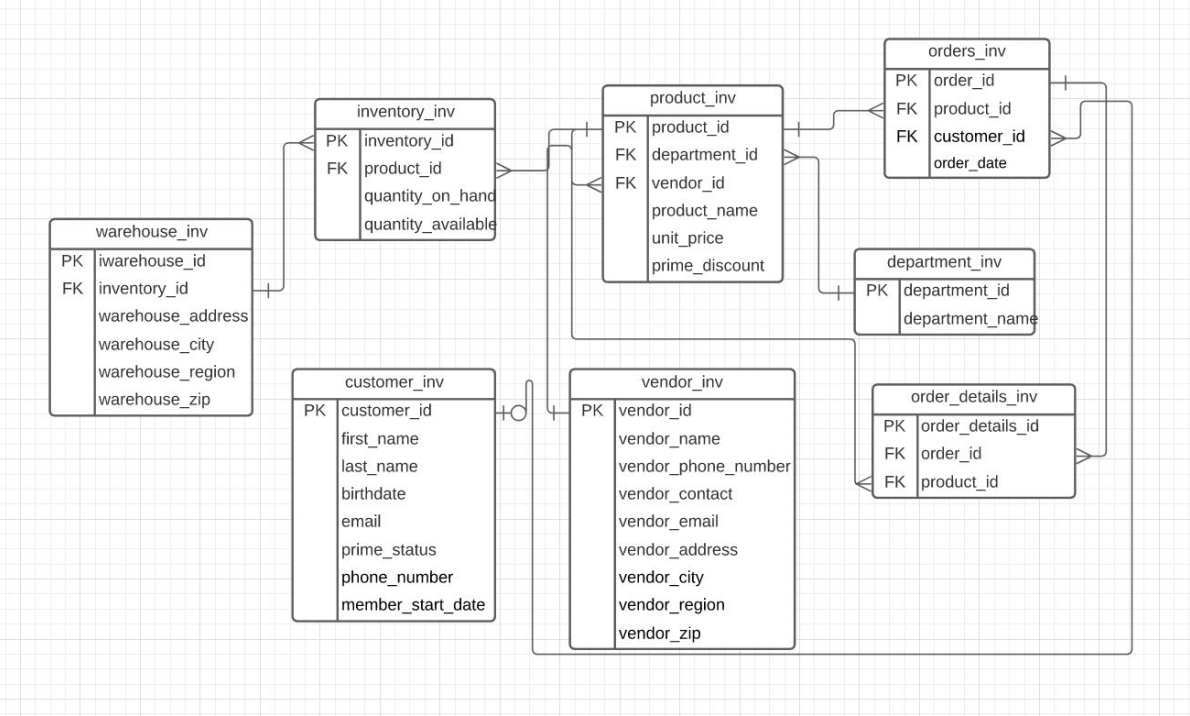


amazon

Call Center Operations ERD



Inventory Management ERD

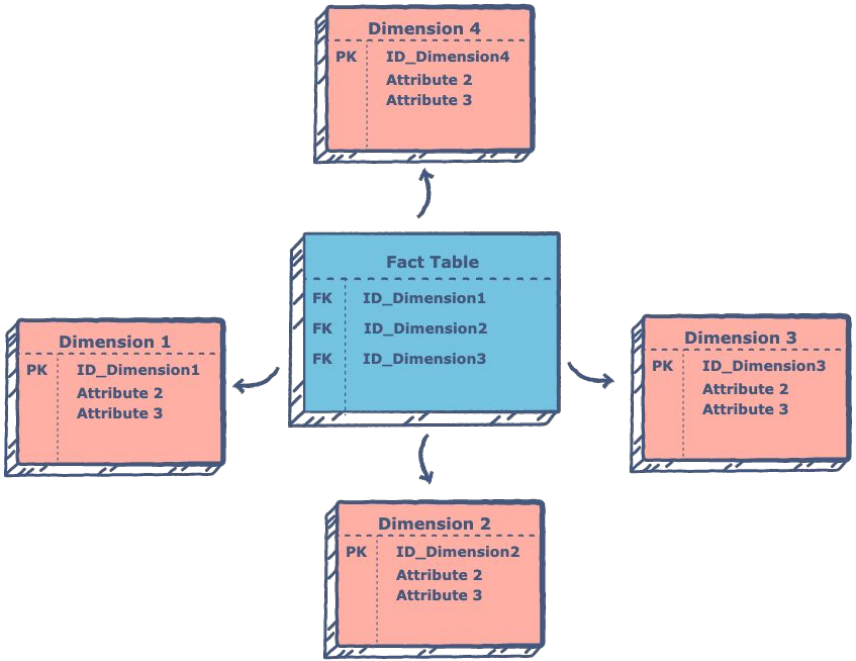


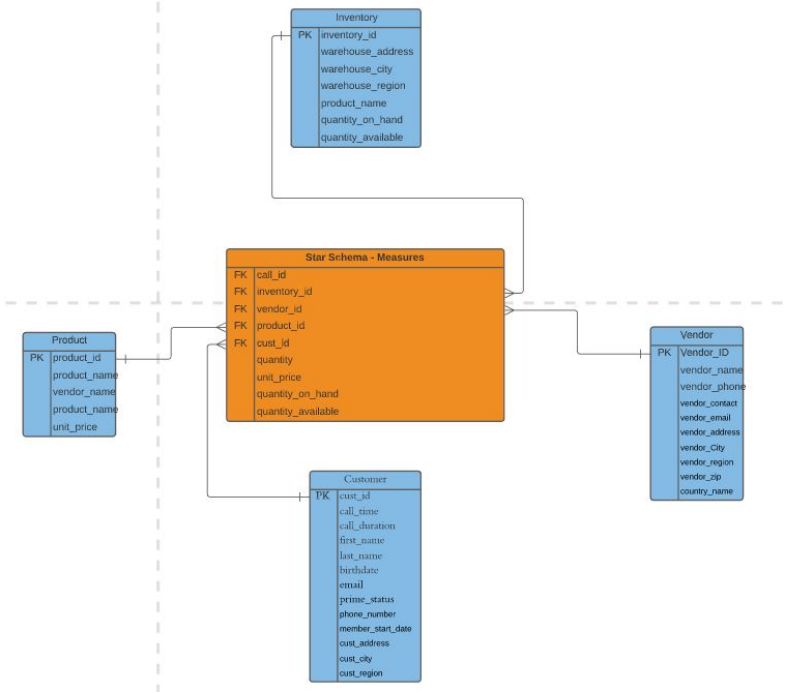
amazon

Data Warehouse Design Decisions

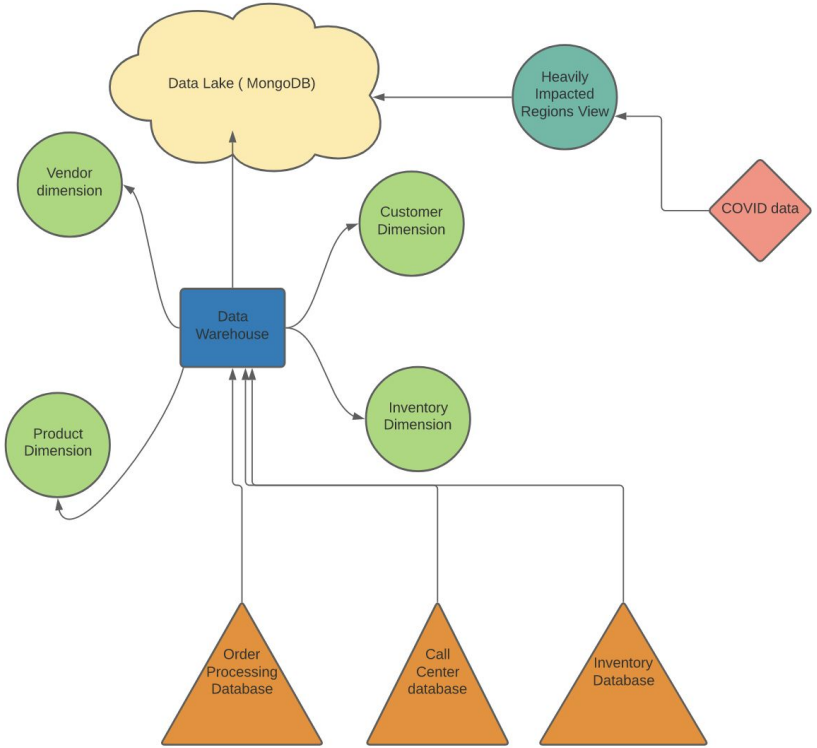
- Star Schema – Product, Vendor, Inventory, Customer Dimensions
    - Analyze call center call volume, warehouse inventory, location, etc.
- Isolate factors that can be used to reduce call frequency
- Essential in mapping trends to better understand transactional operations

Design Decisions - Star Schema

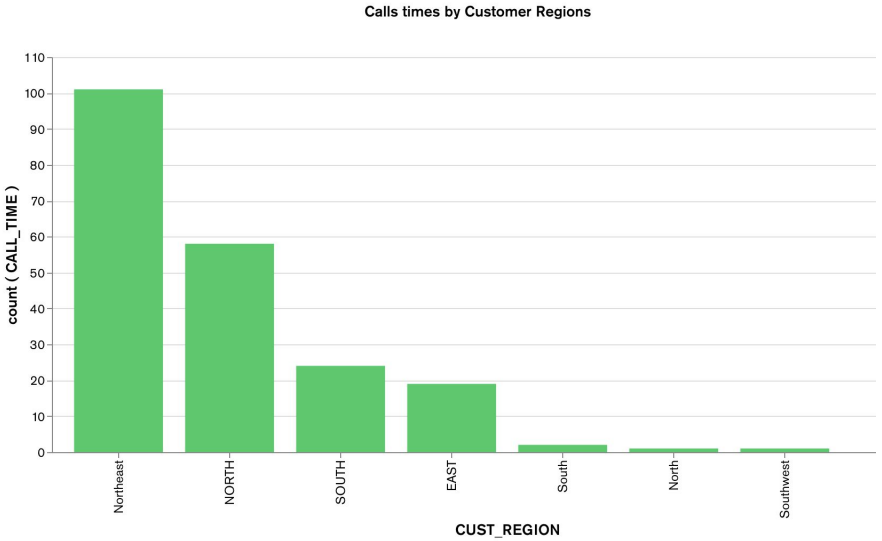


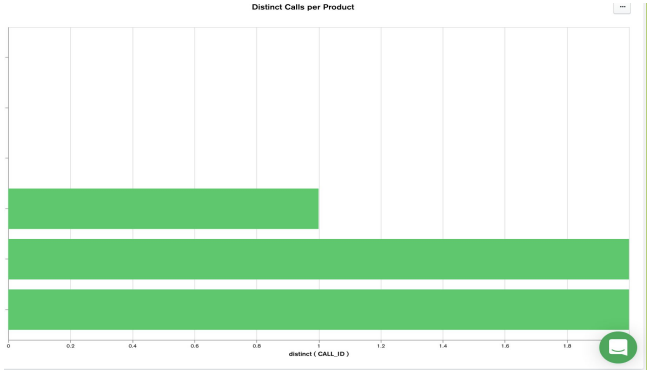


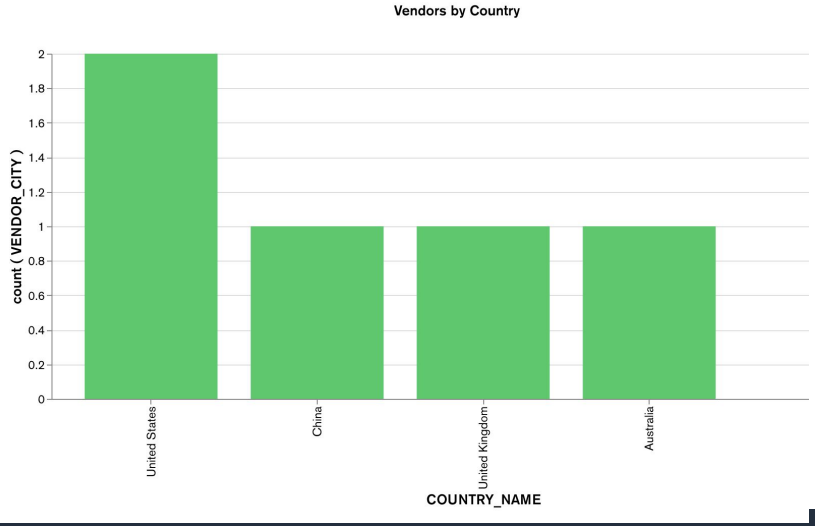
Data Lake Modeling - MongoDB



Analysis Patterns Modeled







amazon

Lessons Learned & Key Insights

- Access to data was easy in our case but this may not apply for real Amazon e-commerce
- Refined and practiced our DDL and ETL
- Thoroughly understood amazon operations and processes
- Focus group with team members and their experience with call centers, inventory management issues and on-call/online order processing
- Important to align the linking tables and inserts across ERDs
- Warehouse compensates for any inefficiency and normalization

Additional Opportunities and Potential Changes

Opportunities

- Times series data
- Additional work with data lakes
- Emphasize dealing with big data
- Streaming data
- Access to more data for various departments

Changes

- Updating and inserting the data was difficult - identify a more efficient method
- Could have utilized an alternate platform to create a data lake
- Identifying convenient ways to seed realistic data

Thank you!

Questions?

Appendix

Sidd: ERD, Data Lake model, Analysis, Powerpoint slides, Data Warehouse

Samir: ERD and DDL for call center and data warehouse, ETL, and Powerpoint slides

Ryan: Data Strategy; ERD, DDL, and inserts for inventory; Powerpoint slides

Paridhi: ERD, DDL, and inserts for call center operations; Powerpoint slides

Grayson: ERD, DDL, and inserts for order_proccessing; Powerpoint slides