**Final Project**

Devkumar Chhatriwala

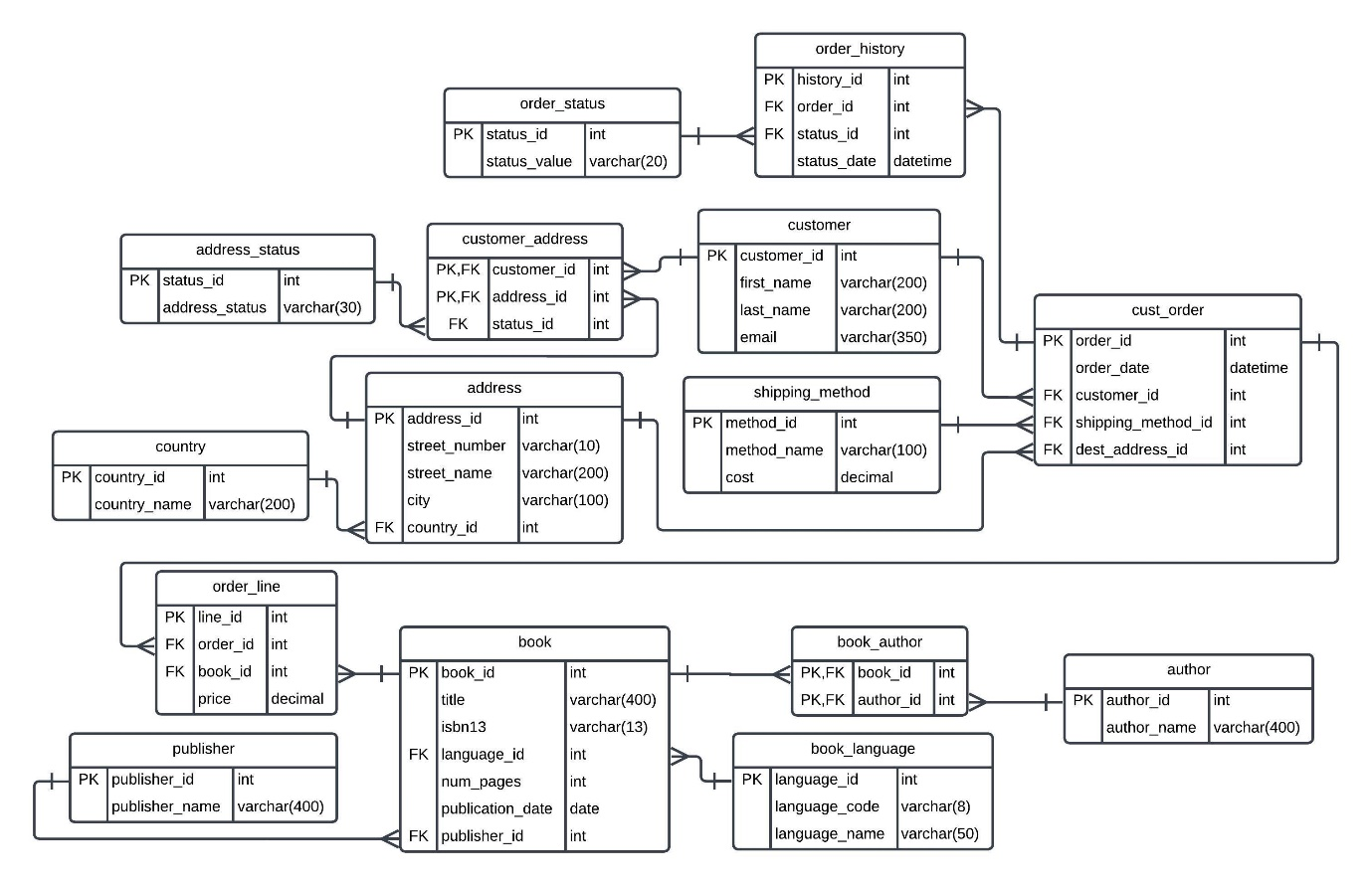
Department of Information Technology

CIS-525-05001 Applied Data Structure & DB 23FADAY

Instructor’s Name: Liz Henley

Dec 07, 2023

**ER DIAGRAM FOR GRAVITY BOOKS**



1. **Which language should they discontinue stocking?**
2. **What was the query you used, both the SQL statement and explained in plain English for someone who doesn't know SQL.**

SELECT bk\_lan.language\_name AS Language\_Available, COUNT(book.book\_id) as Total\_Ordered\_Books FROM book\_language bk\_lan   
JOIN book book ON bk\_lan.language\_id = book.language\_id  
JOIN order\_line od\_ln ON book.book\_id = od\_ln.book\_id  
JOIN cust\_order co ON od\_ln.order\_id = co.order\_id  
GROUP BY bk\_lan.language\_name having  
COUNT(book.book\_id)<5 order by Total\_Ordered\_Books asc;

1. **Identifying Book Languages (FROM book\_language bk\_lan)**: Think of a bookshelf labelled with different languages, where each shelf has books in just one language. This part is about looking at that bookshelf to see what languages are available.
2. **Matching Books to Languages (JOIN book ON bk\_lan.language\_id = book.language\_id)**: Each book has a sticker that matches it to one of the shelves. This step is about putting each book on the right shelf according to its sticker.
3. **Connecting Orders to Books (JOIN order\_line ON book.book\_id = od\_ln.book\_id)**: When a book is ordered, a receipt is made. This step is about pinning each receipt to the book it corresponds to, showing that someone bought it.
4. **Linking Orders to Customers (JOIN cust\_order ON od\_ln.order\_id = co.order\_id)**: If the receipt is part of a customer's entire shopping list, this step is like attaching the whole shopping list to the receipt, so we know who made the order.
5. **Counting Orders Per Language (COUNT(book.book\_id))**: Now, for each language shelf, we count how many receipts are pinned to the books.
6. **Focusing on Less Ordered Languages (HAVING COUNT(book.book\_id) < 5)**: We are only interested in shelves that don't have many receipts, specifically those with fewer than five orders. These are the languages that aren't ordered very often.
7. **Ordering the Languages by Popularity (ORDER BY Total\_Ordered\_Books ASC)**: Arrange the shelves starting with the one that has the least orders. It's like making a list from the least popular language to the more popular ones, but only for those with less than five orders.
8. **Why does that particular query best answer the question over pulling any other data from the database?**

The specific query is best for identifying underperforming book languages because it filters the database information down to only the essential data needed to make a decision about stocking. It avoids extraneous data which could cloud judgment or lead to less precise decisions. By focusing solely on the number of orders per language and setting a specific cut-off point (fewer than five orders), it provides a clear, straightforward metric of demand. This precision allows for more informed and efficient decision-making about which books may be candidates for discontinuation due to poor sales. Other data pulls might include irrelevant information, such as the genre or author, which are not directly related to the frequency of purchases and could complicate the decision-making process.

1. **Show a table of the results from the query and explain in plain English what those results are showing in terms of the question asked.**

A black screen with white text

Description automatically generated

* **Latin, Portuguese, Middle English, and Russian**: Each of these languages has only had one book ordered. This suggests that they are very rarely requested by customers.
* **Aleut**: This language had a slightly higher number, with two books ordered. It's still a low figure, indicating low demand.
* **Malaysian, Italian, and Canadian English**: For each of these languages, three books have been ordered. These languages are a bit more popular than the first set, but still not ordered often.

Since all the numbers are very low (all are less than five), this table is showing us that books in these languages are not selling well. If you're looking at this data to decide which books might not be worth restocking, the table suggests that books in these particular languages might be the ones to consider discontinuing, as customers hardly ever order them

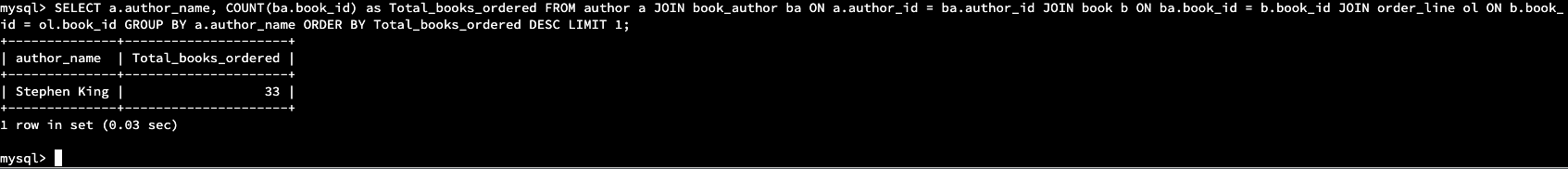
1. **What you think the business should do based on the information you gathered?**
2. **Review Inventory Levels**: For the languages with the lowest sales, evaluate if it's cost-effective to keep them in stock.
3. **Analyse Market Demand**: Conduct market research to understand why these languages have low sales. Is it due to a lack of interest, or are the books not visible or accessible to potential buyers?
4. **Marketing Strategies**: If the business decides that there is potential in these languages, they might implement targeted marketing campaigns to increase interest and sales.
5. **Stock Adjustment**: If the low demand is consistent and the research doesn't show promise for future sales, the business might reduce or discontinue stocking books in these languages.
6. **What author should they pick for a special promotion?**
7. **SQL Query Explanation**

SELECT a.author\_name, COUNT(ba.book\_id) as Total\_books\_ordered FROM author a JOIN book\_author ba ON a.author\_id = ba.author\_id JOIN book b ON ba.book\_id = b.book\_id JOIN order\_line ol ON b.book\_id = ol.book\_id GROUP BY a.author\_name ORDER BY Total\_books\_ordered DESC LIMIT 10;

1. **Identifying Authors (FROM author a)**: Think of 'author' as a list of all the writers whose books the store might sell.
2. **Matching Books to Authors (JOIN book\_author ba ON a.author\_id = ba.author\_id)**: Books don't always tell you who wrote them on the cover, so we match each book to its author from a separate list that shows which book was written by which author.
3. **Connecting Books to Orders (JOIN book ON ba.book\_id = b.book\_id JOIN order\_line ol ON b.book\_id = ol.book\_id)**: This is like saying every time someone orders a book, we make a note of it. By doing this, we can see which books are popular.
4. **Counting Orders for Each Author (COUNT(ba.book\_id) as Total\_books\_ordered)**: Now, for each author, we count how many times their books have been ordered.
5. **Organizing Authors by Popularity (GROUP BY a.author\_name ORDER BY Total\_books\_ordered DESC)**: We then make a list of authors, putting the ones whose books have been ordered the most at the top.
6. **Limiting the List to Top Ten (LIMIT 1)**: We only want to see the top most author.
7. **Why does that particular query best answer the question over pulling any other data from the database?**

This query identifies the single most popular author based on the number of books ordered. Here's why it's the ideal query:

1. **Focus**: Based on the book orders, which is a direct measure of what customers are buying and therefore who is most popular.
2. **Simplicity**: The query is straightforward, which is promoting the top seller.
3. **Top Performer**: By using **LIMIT 1**, the query ensures that only the very top author is selected, making it clear who should be featured in the promotion.
4. **Clear Metrics**: It ranks authors by a clear and simple metric: the number of books ordered, so there's no ambiguity about what "top" means.
5. **Efficiency**: The query is efficient because it avoids pulling unnecessary data about other authors who aren't as relevant for this specific promotional decision.
6. **Show a table of the results from the query and explain in plain English what those results are showing in terms of the question asked.**



The table resulting from the query shows that Stephen King is the author with the highest number of books ordered, with a total of 33 orders. In plain English, this means that among all the authors whose books this bookstore sells, Stephen King's books are the most popular or in-demand based on the number of times they've been ordered. If the bookstore is planning a special promotion and wants to feature the author whose books sell the best, Stephen King would be their top choice because his books have been ordered more than any other author's in the system. The query is designed to pick out just this top-selling author, making it perfect for deciding who to spotlight in a special promotion

1. **What you think the business should do based on the information you gathered?**
2. **Feature Stephen King**: Put a spotlight on Stephen King's books since they're proven sellers. This could include prominent displays, featured sections on the bookstore's website, or a dedicated Stephen King month with promotions.
3. **Host Events**: Organize events such as book readings, discussions, or a Stephen King-themed event to draw in fans and boost sales even further.
4. **Stock Up**: Ensure a good stock of Stephen King's books, including new releases and backlist titles, to meet customer demand.
5. **Bundle Offers**: Create bundle deals for Stephen King's books, encouraging customers to purchase more than one book at a time.
6. **Cross-Promotion**: Recommend books similar to Stephen King's to customers interested in his books, potentially increasing sales of other titles.
7. **What 10 customers should the sales team contact?**
8. **SQL Query :**

SELECT c.customer\_id, first\_name, last\_name, COUNT(DISTINCT order\_id) AS total\_orders FROM customer JOIN cust\_order c on c.customer\_id=customer.customer\_id GROUP BY c.customer\_id, first\_name, last\_name ORDER BY total\_orders DESC LIMIT 10;

1. **Identifying Customers**: SELECT Think of having a guest list for a party. This part of the command tells the database to look at the entire guest list (the customer table) for the party (the bookstore).
2. **Counting Orders**: Count() For each person on the guest list, the command asks, "How many times has this person placed an order?"
3. **Ensuring Unique Orders**: It makes sure to only count each order once, even if the customer ordered multiple things at the same time.
4. **Sorting Customers**: Order BY After counting, it lines up the guests by who placed the most orders to who placed the least.
5. **Top Ten**: Selecting the Top 10 from the list of visitors using LIMIT 10
6. **Why does that particular query best answer the question over pulling any other data from the database?**

This particular query is the best for answering the question about which customers are the most valuable in terms of orders placed because:

1. **Specific Focus**: It zeroes in on the number of unique orders by each customer, which is a direct indicator of customer engagement and loyalty.
2. **Clarity of Results**: By counting distinct orders, it ensures that each order is only counted once, providing a clear picture of each customer's ordering frequency without duplication.
3. **Top Customers Identification**: It ranks customers according to their order count, allowing the business to easily see who their top customers are.
4. **Efficiency**: The query avoids pulling unnecessary information that isn't relevant to the question, such as the details of each order or customer contact information.
5. **Actionable Data**: The resulting list of top customers can directly inform marketing and sales strategies, such as personalized promotions or rewards for frequent buyers.

In simple terms, this query acts like a filter that sifts through all the customer data and finds the most important pieces for the business question at hand – who is ordering the most.

1. **Show a table of the results from the query and explain in plain English what those results are showing in terms of the question asked.**

**A black screen with white text

Description automatically generated**

The results show that these customers are the most engaged with the business, having placed the most unique orders. They are valuable due to their repeat patronage and the sales team should consider them a priority for any outreach or promotional efforts.

* **Erena Glede** is at the top of the list with **35 orders**, making her the most frequent customer.
* **Herc Matthewson** follows with **30 orders**.
* **Francine Sier**, **Johnette Entwistle**, and **Mick Sever** have all placed **28 orders**.
* **Jermain Giraudeau** and **Rob Handes** each have **27 orders**.
* **Noby Burtenshaw** comes next with **26 orders**.
* **Gaylor Polack** has made **25 orders**.
* **Tine Bastow** rounds out the list with **24 orders**.

These results are important because they show who the business's most loyal and frequent buyers are, which can help the company decide who to target for special offers, loyalty programs, or other customer appreciation efforts, that sales team can contact.

1. **What you think the business should do based on the information you gathered?**

Based on the information gathered from the query identifying the top 10 customers with the most orders, here’s what the business should consider doing:

1. **Personalized Communication**: Reach out to these top customers with personalized messages to thank them for their loyalty.
2. **Exclusive Offers**: Provide them with special offers, discounts, or early access to new products as a reward for their frequent purchases.
3. **Feedback and Insights**: Ask for their feedback on what they enjoy about the business and any improvements they would like to see. This information can be invaluable for shaping the business’s future strategies.
4. **Loyalty Program**: If not already in place, consider enrolling these customers in a loyalty program that offers incremental rewards based on order frequency or volume.
5. **Customer Appreciation Events**: Invite these top customers to exclusive events or VIP sales, which can enhance their connection to the business.
6. **Referral Program**: Encourage them to refer friends or family to the business by offering a referral discount or benefit.
7. **Custom Experiences**: Create custom shopping experiences for them, such as personal shopping services or previews of upcoming products.

By taking these steps, the business can not only maintain the loyalty of these top customers but also leverage their satisfaction to attract new customers and improve the overall customer experience.

1. **Gravity Books is looking to add a warehouse to help optimize shipping. Where should they locate that warehouse? Name the specific city and country for the location.**
   1. **SQL query**

SELECT adr.city, ctry.country\_name, COUNT(\*) AS total\_orders

FROM cust\_order AS co

JOIN address AS adr ON co.dest\_address\_id = adr.address\_id

JOIN country AS ctry ON adr.country\_id = ctry.country\_id

GROUP BY adr.city, ctry.country\_name

ORDER BY total\_orders DESC

LIMIT 1;

1. **Joining Order and Address Information**: It connects each order with the destination address to know where the books are being sent.
2. **Joining Address and Country Information**: Then, it links the address with the country to ensure we know the country of each order's destination.
3. **Counting Orders per Destination**: It counts all the orders going to each city and country combination.
4. **Grouping by City and Country**: It groups the results so that all orders for each unique city and country pair are tallied together.
5. **Ordering the Results**: It sorts these groups to find out which city and country pair has the highest total number of orders.
6. **Selecting the Top Result**: Finally, it limits the results to the top row, which represents the city and country with the most orders.
   1. **Why does that particular query best answer the question over pulling any other data from the database?**

The SQL query provided is the best for deciding where to place a new warehouse because it focuses specifically on the number of orders going to each location. It doesn't get side tracked with other details like what books are being ordered or who is ordering them. By counting how many orders are sent to each city and country, the business can easily see where the most orders are being shipped. This helps pinpoint the busiest shipping destination, which is likely the best place to build a new warehouse to speed up deliveries and cut down on shipping costs. It's a straightforward way to base the decision on actual customer order data, which is a direct indicator of where demand is highest.

* 1. **Show a table of the results from the query and explain in plain English what those results are showing in terms of the question asked.**

A screen shot of a computer

Description automatically generated

The table shows the result of a query that aimed to determine the best location for a new warehouse by analysing shipping data. It tells us that Palmerston North in New Zealand is the city with the highest number of orders, having a total of 46 orders.

In plain English, this means that out of all the places Gravity Books ships to, Palmerston North is where the most customers have placed orders. So, according to this data, if the company wants to optimize its shipping by building a new warehouse closer to where most of their orders are going, Palmerston North would be the top choice. This is based on the premise that having a warehouse closer to this location could make shipping faster and more cost-effective.

* 1. **What you think the business should do based on the information you gathered?**

1. **Feasibility Study**: Conduct a detailed feasibility study of Palmerston North to ensure that other factors, such as shipping logistics, labor costs, and warehouse expenses, support the decision.
2. **Market Analysis**: Perform a market analysis to understand the potential for growth in Palmerston North and surrounding areas.
3. **Infrastructure Assessment**: Evaluate the local infrastructure to ensure it can support the logistical needs of a warehouse.
4. **Cost-Benefit Analysis**: Compare the costs of setting up and operating a warehouse in Palmerston North against the expected savings in shipping time and costs.
5. **Customer Service Improvement**: Consider how the warehouse could improve service levels for New Zealand customers, potentially offering faster delivery times.
6. **Local Partnerships**: Look into partnerships with local businesses for shared logistics or warehousing to reduce initial costs.
7. **Sustainability Considerations**: Ensure that the new warehouse operations align with any sustainability goals of the business, such as reducing the carbon footprint of shipping.

**Additional Insights for Gravity Books:**

1. **Seasonal Sales Trends**: By examining historical order data, it becomes apparent that there are distinct seasonal sales trends. This information can be leveraged to plan targeted marketing campaigns and promotions during peak buying seasons, potentially boosting revenue.
2. **Customer Demographics**: The database contains customer information, including names and addresses. Analyzing this data could help Gravity Books understand the geographic distribution of their customer base, allowing for more precise marketing efforts and potentially influencing inventory decisions.
3. **Popular Book Categories**: By analyzing which book categories (e.g., fiction, non-fiction, sci-fi) have the highest sales, Gravity Books can make informed decisions about which genres to focus on when expanding their inventory or promotions.
4. **Preferred Shipping Methods**: Examining the shipping method data can reveal which delivery options customers prefer. This information can help improve the shipping experience and optimize costs.
5. **Author Collaborations**: The database contains information about authors and books. By identifying which authors' works are frequently purchased together, the company can explore opportunities for collaborations or bundled promotions.
6. **Customer Loyalty Analysis**: Beyond identifying the top 10 customers for sales team contact, a more extensive loyalty analysis can be conducted. This includes categorizing customers based on their order frequency, total spend, and recency. It can lead to the development of tailored loyalty programs.
7. **Inventory Management**: Further analysis of inventory data can reveal slow-moving items, allowing the company to make informed decisions about discontinuing specific books or categories to free up resources for more popular items.
8. **Supply Chain Optimization**: By examining the time between order placement and delivery, the company can optimize its supply chain and reduce delivery times, potentially improving customer satisfaction.

**References**

Coronel, C., & Morris, S. (2014). *Database Systems: Design, Implementation, and Management* (12th ed.). Cengage Learning.

Kleppmann, M. (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media.

Sadalage, P. J., & Fowler, M. (2012). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley.

van de Laar, E. (2015). *Pro SQL Server Wait Statistics*. Apress.