**Business Intelligence Tech Screening Exercise**

Hi! Thank you for expressing interest in joining our team at LinkedIn, we have the unique opportunity to be seen as data advisors to our GSO counterparts across all LOBs. The exercise below is a nice sample of what types of work our team is responsible for. Please work through the questions below and send your answers to the recruiter before your tech screening, the answers will be discussed during your next interview.

**Overview:**​

Data on LinkedIn's B2B sales is stored in two tables (see next page for table pictures):

​

**Figure 1) Opportunity** {OpportunityID, AccountID, OwnerID, CreateDate,  CloseDate, Amount, Status}​



(show top 5 sample rows)​

​

**Figure 2) Opportunity\_Line\_Item** {OpportunityID, LineItemID, Units, Amount}​

(show top 5 sample rows)​

​

**Question 1 – Data Modeling:**Using Figure 1 & Figure 2

1. What do you think these tables and columns are capturing? Describe in plain English​

**Opportunity Table**

This table is used to capture and track various sales opportunities

**Fields:**

1. OpportunityID: This field looks to be the primary key of this table. This is used to uniquely identify each opportunity.
2. AccountID: This field looks to be a ID that is associated with an account. An account would be associated with a specific customer. The AccountID looks to be a foreign key to a different table not identified.
3. OwnerID: The OwnerId is a field that relates to an internal employee who is the owner of the account. In many situations, this is the employee who manages interactions with the associated account. The OwnerId looks to be a foreign key to a different table not identified.
4. CreatedDate: The date the opportunity was created. This is field is most likely autogenerated when the opportunity was added to the system.
5. CloseDate: This field is the best estimate for when a decision will be made and the sale will close. This field will change as the opportunity owner finds out more about the opportunity.
6. Amount: This is the total summation for the product or services provided. This is also a field that changes as you understand the needs for the account.
7. Status: This field is associated with where the opportunity is with the sales cycle. This is another field that would be updated as opportunities are closed and won.

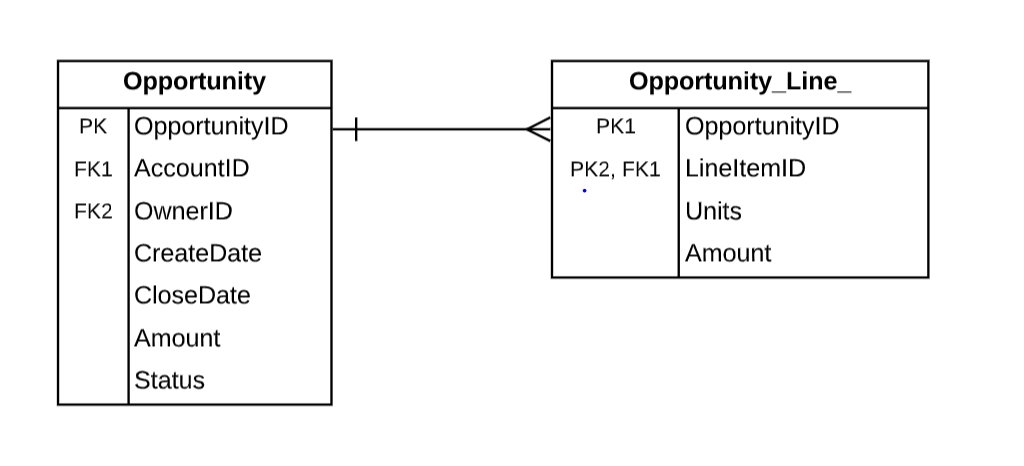
**Opportunity\_Line\_Item**

This table is at a lower granularity level than the opportunity table. For each opportunity, this holds the individual product or service that the customer is interested in.

**Fields:**

1. OpportunityID: This is used to uniquely identify each opportunity. This field is also found in the Opportunity table. This field looks to be both a composite Primary Key and Foreign Key. As a Foreign Key, it relates to the Opportunity table.
2. LineItemID: This looks to identify the individual product or service that is being sold. In a combination with OpportunityID, this field looks to make up the composite primary key. The LineItemID looks to be a foreign key to a different table not identified.
3. Units: The number of products that are being ordered.
4. TotalAmount: The total amount of each product or service being offered.

**Below is an example of the ERD Diagram for the two tables that are known. This would be expanded as more tables are identified.**



1. What relevant information is missing?

* For many of the fields (OpportunityID, OwnerID, AccountID, LineItemID) we have a combination of numbers and letters that make up the field. This combination of characters would match a description column. That description column would be more helpful to understand information from the table.
* The probability of close is missing. This field might not be needed if the opportunities are only showing those that are either closed or refused, but if there are more opportunities, understanding the opportunity to win the sale is important. Knowing the opportunity can help with forecasting as well as prioritizing what opportunities we need to focus on.
* The type of opportunity. Is this a opportunity for new business or existing business. Understanding the opportunity type can help develop a strategy.
* Next Steps. If the opportunity is not closed, what are the next steps identified to help close the opportunity?

1. If you were to design the data to have more product details, what are the tradeoffs between including them as columns in the line item table vs. in a separate dimension table? What would you choose and why?

This is asking the difference between a more normalized and demoralized database design. The goal of a normalized database is to reduce data redundancy and data integrity. This would mean breaking up large tables into multiple tables that would be joined in queries.

In a denormalized table, we would combine both the Opportunity\_Line\_Item and Opportunity table by adding LineItemID, Units, and TotalAmount to the Opportunity table. With that information, we would have one table with all related data.

**Pros to normalized design**

* There is less data duplication (which leaves fewer opportunities for errors)
* The tables are smaller as they are broken out into multiple tables. Smaller tables can be retrieved quicker as they can be stored in memory

**Cons to normalized design**

* Many joins are can be needed for queries
* Query performance maybe slower.

I would recommend a normalized form. Because of how often new opportunities are added and how often changes happen to the tables, it's important to ensure data integrity across the design. It also seems the current database structure uses a normalized design.

**Question 2 – SQL Query:** Using Figure 1 & Figure 2

1. Write a query that tracks $ sales by product by month​

In the queries below, I will assume that LineItemID is associated with a specific product.

**By Month:**

Select EXTRACT(month FROM opp.CloseDate) "Month", LineItemID, SUM(line.TotalAmount)

FROM Opportunity opp

JOIN Opportunity\_Line\_Item line

ON Opportunity. OpportunityID​ = Opportunity\_Line\_Item.OpportunityID

GROUP BY EXTRACT(month FROM opp.CloseDate), LineItemID

**By Year/Month**

Select to\_char( EXTRACT(year FROM opp.CloseDate)) || ‘’/’ || to\_char(EXTRACT(month FROM opp.CloseDate) )"Year/Month", SUM(line.TotalAmount)

FROM Opportunity opp

JOIN Opportunity\_Line\_Item line

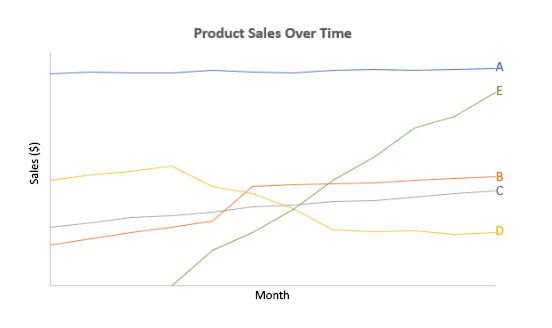
ON Opportunity. OpportunityID​ = Opportunity\_Line\_Item.OpportunityID

GROUP BY Select EXTRACT(year FROM opp.CloseDate) || ‘’/’ || EXTRACT(month FROM opp.CloseDate), LineItemID

Depending on what the business need is, I can see the need to have both of these queries. For example, if the business wants to know historically which month has the most sales, it might make sense to group everything at the month number level.

If the business wants to see historical trending, adding a year component is needed.

**Question 3 – Business Judgement:**Please use the Product Sales Over Time graph below ​

******



1. How would you describe Product A? If it were an actual LinkedIn product which would it be? ​

Product A is the product that brings in the most revenue but experiences little growth. As a LinkedIn Product, This could be revenue generated by marketing solutions related specifically to ads and sponsored content.

It could also be Talent Solutions. Based on the 2019 Microsoft annual report, this product line generates the most revenue. [https://www.microsoft.com/investor/reports/ar19/index.html#](https://www.microsoft.com/investor/reports/ar19/index.html). Also based on the report, it notes that Talent Solutions revenue has increased over time. This growth is not reflected by Product A.

1. What’s going on with Product D? ​

D had a sudden decrease in sales. This could be one of few things. First, the introduction of E had a negative impact. Second B and D are negatively correlated. Third, a external effect caused the decrease in D.

2. What might explain the jump in sales of product B? What data in these tables might you want to investigate further?​

It looks like there was some kind of disruption that leads to a quick increase in sales, then stable over time. This could be something similar to the current COVID climate, fewer companies are hiring, but more employees have time to look into new skills (example LinkedIn Learning). I would want to look at the Opportunity\_Line\_Item table to understand which LineItemID’s are causing this increase. I would also look at OwnerID and AccountID. I want to understand if it’s a specific OwnerID or AccountID that drives this change.

**Question 4 – Business Judgement:** Use the Product Sales Over Time graph

1. Based on this data, what strategic recommendations might you give?​

* Understand a new growth strategy for E. It most likely cannot continue to grow at the current rate. Being proactive on a new strategy could help if growth plateaus.
* Potentially look at A. While it is our best performing product by revenue, there has been no growth over time.
* D is falling. Does it need to be reevaluated as a product we want to continue investing resources and time on?

**Figure 1)** **Opportunity** {OpportunityID, AccountID, OwnerID, CreateDate,  CloseDate, Amount, Status}​

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **OpportunityID​** | **AccountID​** | **OwnerID​** | **CreateDate​** | **CloseDate​** | **Amount​** | **Status​** |
| 0063200001o7xB3AAI​ | 0016000000m5Io3AAE​ | 00532000005L21IAAS​ | 11/25/16​ | 2/10/17​ | $8,485.29​ | Closed Won​ |
| 0063200001o7nxtAAA​ | 0016000000lbg4FAAQ​ | 00532000005L21IAAS​ | 11/22/17​ | NULL​ | $25,673.44​ | Open​ |
| 0066000000ROPX6AAP​ | 0016000000m3a51AAA​ | 00560000001BLCsAAO​ | 06/04/17​ | 06/24/17​ | $12,448.70​ | Closed Won​ |
| 0066000000IndvoAAB​ | 0016000000dfUI9AAM​ | 00560000001BLCsAAO​ | 09/16/11​ | 03/28/12​ | $33,835.21​ | Closed Won​ |
| 0066000000Ah7rpAAB​ | 0016000000JuVPQAA3 | 005600000010BHIAA2​ | 07/01/17​ | 07/02/17​ | NULL​ | Refused​ |

**Figure 2) Opportunity\_Line\_Item** {OpportunityID, LineItemID, Units, Amount}​

|  |  |  |  |
| --- | --- | --- | --- |
| **OpportunityID​** | **LineItemID​** | **Units​** | **TotalAmount​** |
| 0066000000ROPX6AAP​ | 00k6000000JG6kCAAT​ | 1​ | $10,892.70​ |
| 0066000000ROPX6AAP​ | 00k6000000JG6kDAAT​ | 1​ | $1,556.00​ |
| 0066000000IndvoAAB​ | 00k60000009MZjmAAG​ | 1​ | $7,123.20​ |
| 0066000000IndvoAAB​ | 00k60000009MZjcAAG​ | 1​ | $10,176.00​ |
| 0066000000IndvoAAB​ | 00k60000009MZjhAAG​ | 1​ | $9,921.60​ |
| 0066000000IndvoAAB​ | 00k60000009MZjXAAW​ | 5​ | $1,322.88​ |