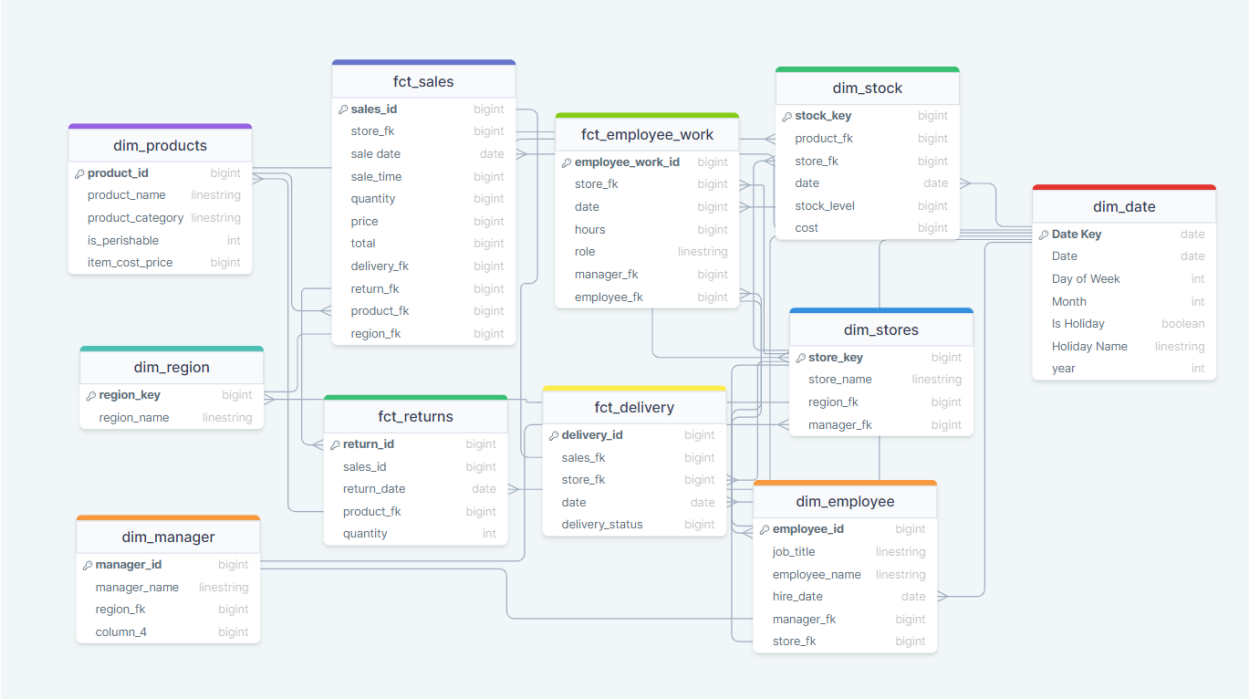
**1.C**reate a logical data model diagram that would store data for VeggieMart.




**2.** Write the SQL that selects from the model that answers the following questions:

<u>What is VeggieMart's monthly revenue?</u>

```
SELECT d.year, d.month, sum(s.sales_amount) AS monthly_revenue
FROM fact_sales s
JOIN dim_date d ON s.date_id = d.date_id
GROUP BY d.year, d.month
ORDER BY d.year, d.month
```

<u>Which manager had the highest revenue across all stores they managed last year?</u>

```
SELECT m.manager_name, SUM(s.total) AS total_revenue
FROM fact_sales s
JOIN stores st ON s.store_fk = st.store_key
JOIN managers m ON st.manager_fk = m.manager_key
```

```
WHERE date_trunc('year', s.date) = date_trunc('year', current_date - interval '1 year')
GROUP BY  m.manager_name
ORDER BY total_revenue DESC
LIMIT 1
```

What was the most-purchased item across all stores last week?

```
SELECT p.product_name, SUM(s.quantity) AS total_quantity
FROM fact_sales s
JOIN products p ON s.item_fk = p.product_id
WHERE s.date >= CURRENT_DATE - INTERVAL '7 days'
GROUP BY p.product_name
ORDER BY total_quantity DESC
LIMIT 1
```

**3.** Explain how you worked through the four steps of dimensional modeling to arrive at your logical data model.

**Selecting the business process**: The business processes i am interested in modeling is the process of selling grocery items at retail stores,processing returns and exchanges, tracking and managing the stock of products,delivering products to customers.

**Declare the grain:** I chose a granular grain that allows for more detailed analysis and better insights into the business especially in the areas highlighted in bonus 1.
Hence I declared the grain to be at the level of individual products sold, individual items returned, each product in each individual delivery made by VeggieMart would also be its own record to  allow for analysis of products that are being delivered most frequently.

**Identify dimensions**: These are some of the descriptive attributes I have highlighted to provide context for the facts.

**Stores**: This dimension describes the physical retail stores where the sales transactions take **place**.  This would provide context for trying to understand where we make the highest sales.
**Regions**: This dimension describes the geographic regions in which the stores are located. This would provide context for analyzing sales performance across different regions.
**Managers**: This dimension describes the managers who are responsible for overseeing the stores.
**Employees**: This dimension describes the employees who work at the stores. This would allow analysis of employee performance,

such as which employees are making the most sales or which employees are most efficient at restocking inventory.
**Date**: This dimension describes the date on which the sales transaction took place.
**Products**: This dimension describes the grocery products that are sold in the stores. This would allow analysis of which
products are selling the most, as well as which products are most frequently returned or have the highest profit margins.

**Identify facts**: These are the numerical values that represent some aspect of the business process selected above.
**Sales**: This fact represents the amount of revenue generated by each sales transaction.
**Inventory**: This fact represents the quantity of each product that is in stock at each store.
**Employee_work**: This fact represents the information related to the employees such as their salaries, working hours, etc.
**Delivery**: This fact represents the information related to the delivery of the products to the stores.
**Returns**: This fact represents the information related to the returns made by customers.

Based on these four steps, I could arrive at a logical data model for the VeggieMart business process that captured the relationships between them.
Using this model, I am able to analyze and answer business questions pertaining to the VeggieMart business process and gain insights into factors such as sales trends, inventory levels, and employee performance.

**Bonus 1:**

Which item accounted for the most profit last year?

```
SELECT p.product_name, SUM(f.sales_amount - (f.quantity * p.item_cost_price)) AS profit
FROM fact_sales f
JOIN dim_date d ON f.sale_date = d.date_key
JOIN dim_product p ON f.item_fk = p.product_id
WHERE d.year = 2022
GROUP BY p.product_name
ORDER BY profit DESC
LIMIT 1
```

In which state did VeggieMart make the least profit?

```
SELECT r.state, SUM(f.sales_amount - (f.quantity * p.item_cost_price)) AS profit
```

```
FROM fact_sales f
JOIN dim_region r ON s.region_key = f.region_fk
JOIN dim_product p ON f.item_fk = p.product_id
GROUP BY r.state
ORDER BY profit
LIMIT 1
```

How many frozen turkeys did VeggieMart sell last November

```
SELECT p.product_name, SUM(f.quantity) AS total_qty
FROM fact_sales f
JOIN dim_product p ON f.item_fk = p.product_id
JOIN dim_date d ON f.date_fk = d.date_key
WHERE d.year = 2022
AND d.month = 11
AND p.product_category = 'Frozen'
AND p.product_name like  '%Turkey%'
GROUP BY p.product_name
```