**Problem 1: Data Cube Design and Dimension Mapping for CPI Card Group's Data Warehouse**

**Data Cube 1: Lead Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | Lead File |
| Customer | Customer Name | - | Lead File |
| Location | Location Name | - | Lead File |
| Sales Agent | Sales Agent Name | - | Lead File |
| Sales Class | Sales Class Desc | - | Lead File |

**Data Cube 2: Job Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | ERP Database |
| Customer | Customer Name | - | ERP Database |
| Location | Location Name | - | ERP Database |
| Sales Agent | Sales Agent Name | - | ERP Database |
| Sales Class | Sales Class Desc |  | ERP Database |

**Data Cube 3: Subjob Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | ERP Database |
| Customer | Customer Name | - | ERP Database |
| Location | Location Name | - | ERP Database |
| Sales Agent | Sales Agent Name | - | ERP Database |
| Sales Class | Sales Class Desc | - | ERP Database |
| Job | JobId | - | ERP Database |

**Data Cube 4: Invoice Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | Invoice |
| Customer | Customer Name | - | Invoice |

**Data Cube 5: Shipments Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | ERP Database |
| Job | Job Description | - | ERP Database |
| Sub Job | Subjob Description | - | ERP Database |
| Invoice | Invoice Identifier | - | ERP Database |
| Customer | Customer Location Key | - | ERP Database |

**Data Cube 6: Sales Summary Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | Financial Sales Summary |
| Location | Location Name | - | Financial Sales Summary |
| Sales Class | Sales Class Desc | - | Financial Sales Summary |

**Data Cube 7: Cost Summary Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Source** |
| Date | Date | Year > Quarter > Month > Date | Financial Cost Summary |
| Location | Location Name | - | Financial Cost Summary |
| Machine Type | Machine Type Desc | - | Financial Cost Summary |
| Sales Class | Sales Class Desc | - | Financial Cost Summary |

**Problem 2: Measures and Aggregation Design for CPI Card Group's Data Cubes**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data Source** | **Measures** | **Aggregation Properties** |  | **Data Source** | **Measures** | **Aggregation Properties** |
| Lead File | Quote\_Qty | Sum | Customer | Credit\_Limit | Max |
| Quote\_Value | Sum | Date\_First\_Order | Min |
| Success | Count | Shipment | Actual\_Quantity | Sum |
| Job | Actual\_Units | Sum | Requested\_Quantity | Sum |
| Actual\_Amount | Sum | Boxes | Sum |
| Forecast\_Units | Sum | Financial Sales Summary | Actual\_Units | Sum |
| Forecast\_Amount | Sum | Actual\_Amount | Sum |
| SubJob | Actual\_Labor\_Cost | Sum | Forecast\_Units | Sum |
| Actual\_Material\_Cost | Sum | Forecast\_Amount | Sum |
| Actual\_Machine\_Cost | Sum | Financial Cost Summary | Actual\_Units | Sum |
| Actual\_Overhead\_Cost | Sum | Actual\_Labor\_Cost | Sum |
| Invoice | Invoice\_Amount | Sum | Actual\_Material\_Cost | Sum |
| Invoice\_Quantity | Sum | Actual\_Machine\_Cost | Sum |
| Invoice\_Shipped | Sum |  | Actual\_Overhead\_Cost | Sum |

**Data Cube Design for CPI Card Group's Business Analysis**

|  |  |  |
| --- | --- | --- |
| **Cube** | **Dimensions** | **Measures** |
| Leads Analysis | Sales Agent, Customer, Location, Sales Class | Quote\_Qty, Quote\_Value, Success |
| Jobs Analysis | Customer, Location, Sales Class | Actual\_Units, Actual\_Amount, Forecast\_Units, Forecast\_Amount |
| Subjobs Analysis | Customer, Location, Machine Type, Sales Class | Actual\_Labor\_Cost, Actual\_Material\_Cost, Actual\_Machine\_Cost, Actual\_Overhead\_Cost |
| Invoice Analysis | Customer, Invoice Date | Invoice\_Amount, Invoice\_Quantity, Invoice\_Shipped |
| Shipment Analysis | Customer, Location, Job, Subjob, Invoice | Actual\_Quantity, Requested\_Quantity, Boxes |
| Sales Summary Analysis | Sales Class, Location | Actual\_Units, Actual\_Amount, Forecast\_Units, Forecast\_Amount |
| Cost Summary Analysis | Sales Class, Location, Machine Type | Actual\_Units, Actual\_Labor\_Cost, Actual\_Material\_Cost, Actual\_Machine\_Cost, Actual\_Overhead\_Cost |

**Problem 3: Data Cube Grain and Sparsity Analysis for CPI Card Group's Business Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Cube** | **Grain** | **Unadjusted Size** | **Sparsity** |
| Leads Analysis | Individual Leads | 250,000 | Calculated |
|  | Coarser Grain (Customer Postal Codes) | 3,000 | Calculated |
| Jobs Analysis | Individual Jobs | 100,000 | Calculated |
|  | Coarser Grain (Location) | 10 | Calculated |
| Subjobs Analysis | Individual Subjobs | 500,000 | Calculated |
|  | Coarser Grain (Location) | 10 | Calculated |
| Invoice Analysis | Individual Invoices | 1,000,000 | Calculated |
|  | Coarser Grain (Customer) | 3,000 | Calculated |
| Shipment Analysis | Individual Shipments | 2,500,000 | Calculated |
|  | Coarser Grain (Customer, Location, Job) | 300,000 | Calculated |
| Sales Summary Analysis | Individual Sales Transactions | 1,800 | Calculated |
|  | Coarser Grain (Location) | 10 | Calculated |
| Cost Summary Analysis | Individual Cost Transactions | 5,400 | Calculated |
|  | Coarser Grain (Location, Machine Type) | 600 | Calculated |

**Note:** The "Unadjusted Size" and "Sparsity" columns in the table above are calculated based on the provided row counts for each data source. These calculations help estimate the relative storage requirements for the grain in each data cube.

**Problem 4: Transformed Data Cube Schemas – Star/Snowflake Schema Design**

|  |  |  |  |
| --- | --- | --- | --- |
| **Schema Type** | **Dimension Tables** | **Fact Table** | **Comments** |
| Snowflake | Customer, Sales Agent, Location, Date | Lead | Snowflake schema design |
| Snowflake | Customer, Sales Agent, Location, Date, Sales Class | Job | Snowflake schema design |
| Snowflake | Job, Machine Type, Date | SubJob | Snowflake schema design |
| Snowflake | Customer, Date | Invoice | Snowflake schema design |
| Snowflake | Customer, Invoice, SubJob, Date | Shipment | Snowflake schema design |
| Star | Sales Class, Location, Date | Sales Summary | Star schema design |
| Star | Sales Class, Location, Machine Type, Date | Cost Summary | Star schema design |

**Problem 5: Potential Summarizability Problems and Resolutions in Star Schema**

1. **Problem:** In the "Sales Summary" star schema, the relationship between the "Sales Class" dimension and the fact table could lead to over-aggregation if "Sales Class" is used for more granular analyses. **Resolution:** To prevent over-aggregation, consider adding more detailed sales-related dimensions, such as "Product" or "Subjob," to the fact table to maintain the required level of granularity.
2. **Problem:** The "Sales Class" dimension in the "Sales Summary" fact table may have null values if not applicable to some sales records. **Resolution:** Allow null values in the "Sales Class" column in the "Sales Summary" fact table for records where this dimension is not applicable.
3. **Problem:** In the "Cost Summary" star schema, the relationship between "Machine Type" and the fact table could lead to over-aggregation if "Machine Type" is used for multiple analyses. **Resolution:** To avoid over-aggregation, consider introducing a separate fact table for machine-related costs if the level of detail is significantly different from other analyses.
4. **Problem:** The "Machine Type" dimension in the "Cost Summary" fact table may have null values if not applicable to some cost records. **Resolution:** Allow null values in the "Machine Type" column in the "Cost Summary" fact table for records where this dimension is not applicable.
5. **Problem:** The "Customer" dimension is present in multiple fact tables (Lead, Job, Invoice, Shipment), which could lead to redundant storage of customer-related data. **Resolution:** Instead of duplicating customer information in each fact table, create a central "Customer" dimension that can be shared across all fact tables, reducing redundancy.
6. **Problem:** In the "Shipment" fact table, the relationship between "Customer" and "Customer Location" dimensions might lead to over-aggregation if location details are also stored in the customer dimension. **Resolution:** If location details are necessary for shipment analysis, consider including the "Customer Location" dimension directly in the fact table to avoid over-aggregation.
7. **Problem:** The "Customer Location" dimension in the "Shipment" fact table may have null values if not applicable to some shipments. **Resolution:** Allow null values in the "Customer Location" column in the "Shipment" fact table for records where this dimension is not applicable.
8. **Problem:** In the "Shipment" fact table, the relationship between "Subjob" and the fact table might cause over-aggregation if "Subjob" is a lower-level detail compared to other dimensions. **Resolution:** Depending on business requirements, create a separate fact table focused solely on "Subjob" analysis, ensuring appropriate granularity.