Case Study Part 1

**Four Modules:**

* Lead Module
* Job Module
* Bill\_Shipment Module
* Performance Module

**1. Dimensions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Attributes** | **Hierarchies** | **Data Sources** |
| Agent\_Dim | Sales\_Agent\_Id |  | SalesAgent |
| Agent\_Dim | Sales\_Agent\_Name |  | SalesAgent |
| Job\_SubJob\_Dim | Job\_SubJob\_Id |  | Job, Subjob |
| Job\_SubJob\_Dim | Cust\_Name |  | Customer |
| Job\_SubJob\_Dim | Manufacturer |  | MachineType |
| Location\_Dim | Location\_Id |  | Location |
| Location\_Dim | Location\_Name |  | Location |
| Product\_Dim | Sales\_Class\_Id |  | SalesClass |
| Product\_Dim | Sales\_Class\_Desc |  | SalesClass |
| Shipment\_Dim | Shipment\_Id |  | Shipment |
| Shipment\_Dim | Zip | Country →State →County→ City→Area | CustLocation |
| Calendar\_Dim | Time\_Id |  | Time Dimension Table |
| Calendar\_Dim | Year |  | Time Dimension Table |
| Calendar\_Dim | Quarter |  | Time Dimension Table |
| Calendar\_Dim | Month |  | Time Dimension Table |
| Calendar\_Dim | Day |  | Time Dimension Table |
| Calendar\_Dim | Week |  | Time Dimension Table |

**2. Measures**

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Measures** | **Aggregation Properties** |
| Lead file | Quote\_Qty | Additive |
| Lead file | Quote\_Price | Snapshot |
| Lead file | Quote\_Value | Additive |
| Job | Unit\_Price | Snapshot |
| Job | Quantity\_Ordered | Additive |
| Invoice | Invoice\_Amount | Additive |
| Sales Summary | Actual\_Amount | Additive |
| Sales Summary | Forecast\_Amount | Additive |
| Cost Summary | Actual\_Labor\_Cost | Additive |
| Cost Summary | Actual\_Material\_Cost | Additive |
| Cost Summary | Actual\_Overhead\_Cost | Additive |
| Cost Summary | Actual\_Machine\_Cost | Additive |
| Cost Summary | Budget\_Labor\_Cost | Additive |
| Cost Summary | Budget\_Material\_Cost | Additive |
| Cost Summary | Budget\_Overhead\_Cost | Additive |
| Cost Summary | Budget\_Machine\_Cost | Additive |

|  |  |  |
| --- | --- | --- |
| **Cube** | **Dimensions** | **Measures** |
| Lead Cube | Agent\_Dim | Quote\_Qty |
|  | Job\_SubJob\_Dim | Quote\_Price |
|  | Calendar | Quote\_Value |
| Job Cube | Location\_Dim | Unit\_Price |
|  | Product\_Dim | Quantity\_Ordered |
|  | Job\_SubJob\_Dim | Job\_Revenue |
|  | Calendar | Invoice\_Amount |
| Bill\_Shipment Cube | Shipment\_Dim | Contract\_Time\_Ship |
|  | Job\_SubJob\_Dim | Requested\_Time\_Ship |
|  | Product\_Dim | Invoice\_Time |
|  | Location\_Dim | Invoice\_Trend |
|  | Calendar |  |
| Performance Cube | Location\_Dim | Actual\_Amount |
|  | Product\_Dim | Forecast\_Amount |
|  | Calendar | Actual\_Labor\_Cost |
|  |  | Actual\_Material\_Cost |
|  |  | Actual\_Overhead\_Cost |
|  |  | Actual\_Machine\_Cost |
|  |  | Budget\_Labor\_Cost |
|  |  | Budget\_Material\_Cost |
|  |  | Budget\_Overhead\_Cost |
|  |  | Budget\_Machine\_Cost |
|  |  | Actual\_Cost |
|  |  | Budget\_Cost |
|  |  | Gross\_Margin |
|  |  | Gross\_Margin\_Forecast |
|  |  | Gross\_Margin\_Budget |

**3. Grain Size Calculation:**

Lead Cube:

The most detail grain is the combination of individual Agent\_Dim, Lead\_FT, Job\_SubJob\_Dim and date.

* 50 Sales Agent
* 50,000 X 5 Job\_SubJob rows
* 365 days
* 125000 leads per year
* Sparcity:
  + 1 - ( fact table size / product of dimensions )
  + 1 -(125000/(50 X 50000 X 5 X 365))
  + 0.9999726
* Grain= 0.000027
* Unadjusted Size=50 X 50000 X 5 X 365

Order Cube:

The most detail grain is the combination of individual Location, Product, Customer, Manufacturer, and date and Order\_FT.

* 10 Location
* 6 Product
* 365 days
* 3000 customers
* 10 manufacturer (MachineTypes)
* 50000 order per year
* Sparcity:
  + 1 - ( fact table size / product of dimensions )
  + 1 -(50000/(10 X 6 X 3000 X 10 X 365))
  + 0.99992
* Grain= 0.0000761
* Unadjusted Size=10 X 6 X 3000 X 10 X 365=657000000

Bill\_Shipment Cube:

The most detail grain is the combination of individual Location, Product, JobSubJob, shipment and date and Invoice\_FT.

* 10 Location
* 6 Product
* 365 days
* 50000 X 5 JobsSubJobs rows
* 50000 X 5 X 5 shipments
* (50000 X 5 X 5 )/2.5 Invoice rows
* Sparcity:
  + 1 - ( fact table size / product of dimensions )
  + 1 -(50000 X 5 X 5/(2.5 X 50000 X 5 X 5 X 50000 X 5 X 365 X 6 X 10))
  + 0.999999
* Grain= 0.000001
* Unadjusted Size=50000 X 5 X 5 X 50000 X 5 X 365 X 6 X 10=6.84e15

Performance Cube:

The most detail grain is the combination of individual Location, Product, dates and Summary Fact table.

* 10 Location
* 6 Product
* 365 days
* Sales = .70 X 1800/5 =252 rows
* Cost=.85 X 5400 /5 = 918 rows
* Sparcity:
  + 1 - ( fact table size / product of dimensions )
  + 1 -((252+918)/(10 X 6 X 365))
  + 0.05342
* Grain= 0.9465
* Unadjusted Size=21900

|  |  |  |  |
| --- | --- | --- | --- |
| **Cube** | **Grain** | **Unadjusted Size** | **Sparsity** |
| Lead Cube | 0.00002739 | 4562500000 | 0.9999726 |
| Order Cube | 0.0000761 | 657000000 | 0.99992 |
| Bill\_Ship\_Cube | 0.000001 | 6.84e15 | 0.999999 |
| Performance Cube | 0.9465 | 21900 | 0.05342 |

**4. Schema Details**

**Lead Module:**

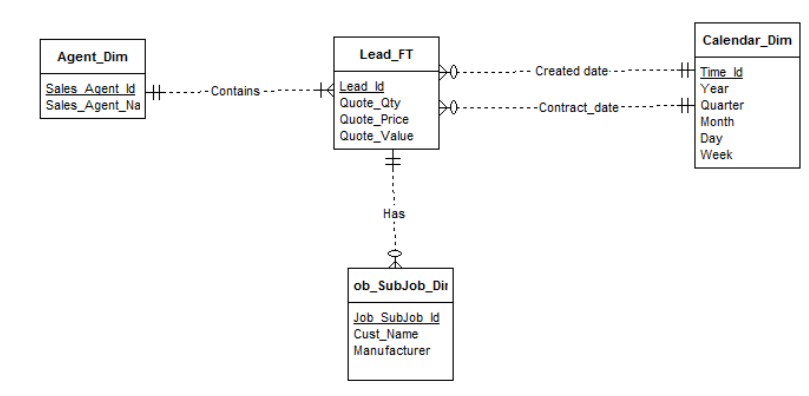
Agent\_Dim = {Sales\_Agent\_Id, Sales\_Agent\_Name}

Lead\_FT = {Lead\_Id, Sales\_Agent\_Id (F.K.), Time\_Id (F.K.), Job\_SubJob\_Id (F.K.), Quote\_Qty, Quote\_Price, Quote\_Value}

Calandar\_Dim = {Time\_Id, Year, Quarter, Month, Week}

Job\_SubJob\_Dim = {Job\_SubJob\_Id, Cust\_Name, Manufacturer}

Forign Keys (Sales\_Agent\_Id, Time\_Id, Job\_SubJob\_Id)



**Job Module:**

Location\_Dim = {Location\_Id, Location\_Name}

Order\_FT = {Order\_Id, Location\_Id (F.K.), Time\_Id (F.K.), Sales\_Class\_Id (F.K.), Job\_Sub\_Id (F.K.), Unit\_Price, Quantity\_Ordered, Job\_Revenue}

Calandar\_Dim = {Time\_Id, Year, Quarter, Month, Week}

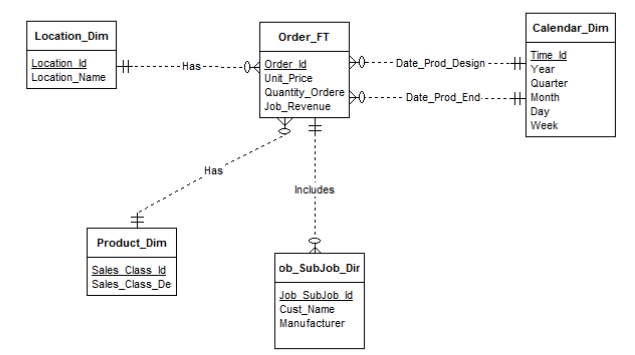
Product\_Dim = {Sales\_Class\_Id, Sales\_Class\_Desc}

Job\_SubJob\_Dim = {Job\_Sub\_Id, Cust\_Name, Manufacturer}

Forign Keys (Location\_Id, Time\_Id, Sales\_Class\_Id, Job\_Sub\_Id)

**Calculation:**

Job\_Revenue (Unit\_Price \* Quantity)



**Bill\_Shipment Module:**

Shipment\_Dim = {Shipment\_Id, Zip}

Invoice\_FT = {Invoice\_Id, Shipment\_Id(F.K.), Time\_Id (F.K.), Job\_Sub\_Id (F.K.), Sales\_Class\_Id (F.K.), Location\_Id(F.K.), Invoice\_Amt, Contract\_Time\_Ship, Requested\_Time, Invoice\_Time, Invoice\_Trend}

Calandar\_Dim = {Time\_Id, Year, Quarter, Month, Week}

Job\_SubJob\_Dim = {Job\_Sub\_Id, Cust\_Name, Manufacturer}

Product\_Dim = {Sales\_Class\_Id, Sales\_Class\_Desc}

Location\_Dim = {Location\_Id, Location\_Name}

Forign Key (Shipment\_Id, Time\_Id, Job\_Sub\_Id, Sales\_Class\_Id, Location\_Id)

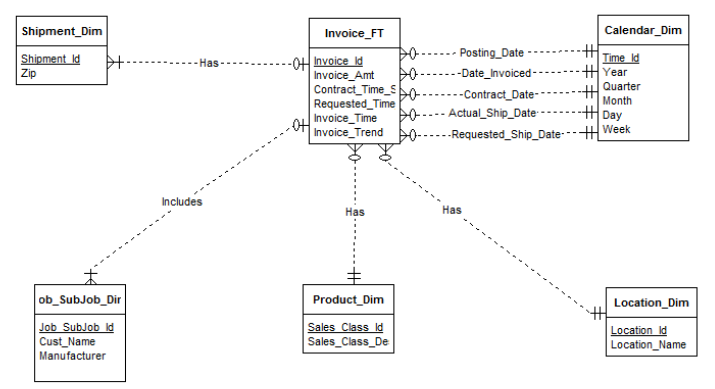
**Calculation:**

Contract\_Time\_Ship (Actual\_Shi\_Date – Contract\_Date)

Requested\_Time\_Ship (Actual\_Ship\_Date – Requested\_Ship\_Date)

Invoice\_Time (Posting\_Date – Date\_Invoiced)

Invoice\_Trend (Invoice\_Shipped – Invoice\_Quantity)



**Performance Module:**

Location\_Dim = {Location\_Id, Location\_Name}

Summary\_FT = {Summary\_Id, Location\_Id (F.K.), Time\_Id(F.K.), Sales\_Class\_Id(F.K.) Actual\_Amount, Forecast\_Amount, Actual\_Labor\_Cost, Actual\_Material\_Cost, Actual\_Overhead\_Cost, Actual\_Machine\_Cost, Budget\_Labor\_Cost, Budget\_Material\_Cost, Budget\_Overhead\_Cost, Budget\_Machine\_Cost, Actual\_Cost, Budget\_Cost, Gross\_Margin, Gross\_Margin\_Forecast, Gross\_Margin\_Budget}

Calandar\_Dim = {Time\_Id, Year, Quarter, Month, Week}

Product\_Dim = {Sales\_Class\_Id, Sales\_Class\_Desc}

Forign Key (Location\_Id, Time\_Id, Sales\_Class\_Id)

**Calculation:**

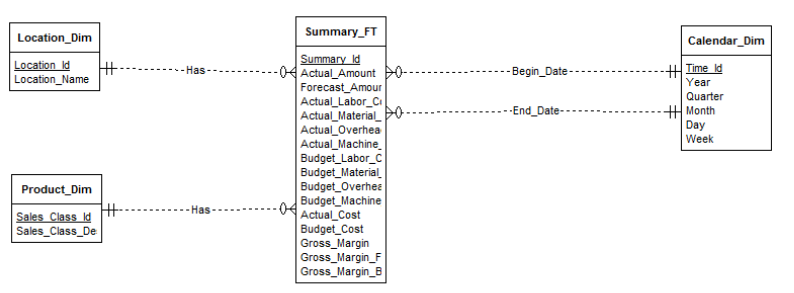
Actual\_Cost (Actual\_Labor\_Cost + Actual\_Material\_Cost + Actual\_Overhead\_Cost + Actual\_Machine\_Cost)

Budget\_Cost (Budget\_Labour\_Cost + Budget\_Material\_Cost + Budget\_Overhead\_Cost + Budget\_Machine\_Cost)

Gross\_Margin (Actual\_Amount – Actual\_Cost)

Gross\_Margin\_Forecast (Gross\_Margin – Forecast\_Amount)

Gross\_Margin\_Budget (Gross\_Margin – Budget\_Cost)



|  |  |  |  |
| --- | --- | --- | --- |
| **Schema type** | **Dimension tables** | **Fact table** | **Comments** |
| Star | Agent\_Dim, Job\_SubJob\_Dim, Calendar\_Dim | Lead\_FT | Sales Agent Productivity |
| Star | Location\_Dim, Product\_Dim, Job\_SubJob\_Dim, Calendar\_Dim | Order\_FT | Job Revenue and production trends |
| Star | Shipment\_Dim,Location\_Dim,Product\_Dim,Job\_SubJob\_Dim, Calendar\_Dim | Invoice\_FT | Shipping and Invoice trends |
| Star | Location\_Dim,Product\_Dim, Calemdar\_Dim | Summary\_FT | Financial Summary trends |

**5. Summarizability Problems:**

1. Week – Month has Non strict non- summarizability issue because, Week has overlapping months in Calendar\_Dim.
2. Manufacturer has null values in the Job\_SubJob\_Dim. This may lead to rollup incompleteness, which can be avoided by adding a default type.
3. Contract\_Date has null values in Lead\_FT. This may lead to rollup incompleteness.
4. Quote\_Price will support null values.