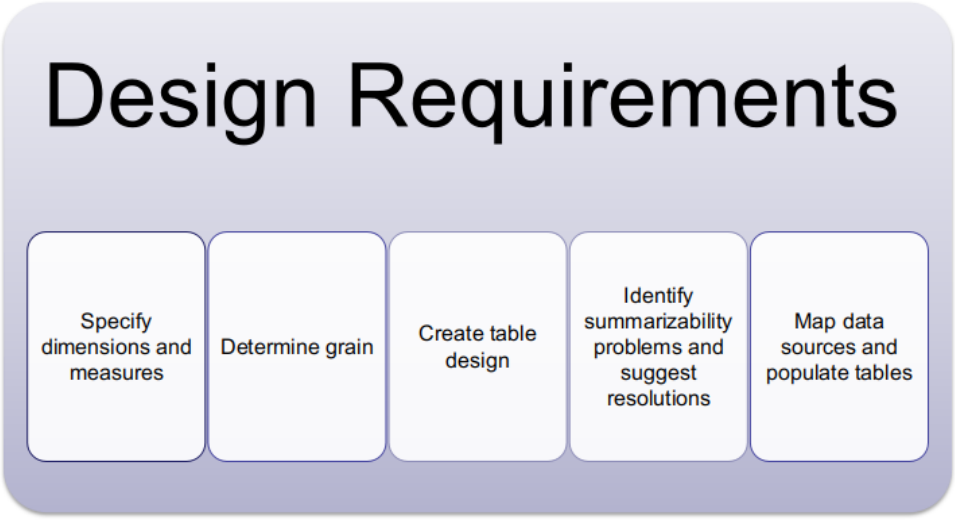
Data Warehouse Design Assignment

Design Process



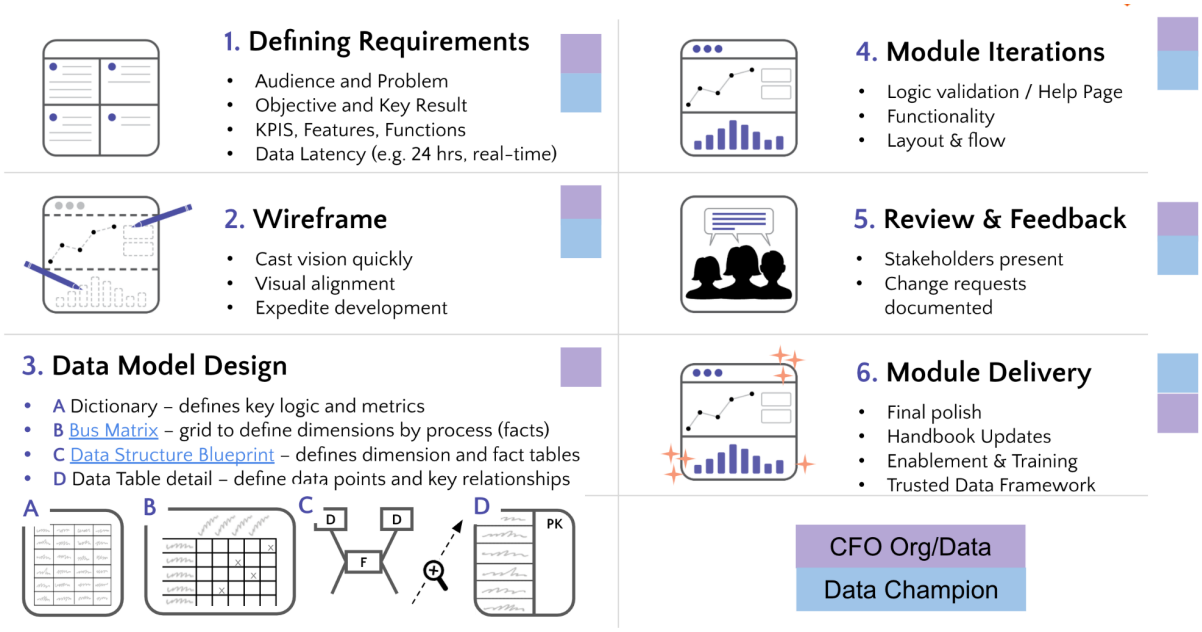
Source: Course Material

Four-Step Dimensional Design Process

The four key decisions made during the design of a dimension

1. Select the business process.
2. Declare the grain.
3. Identify the dimensions.
4. Identify the facts.

Source: Kimball's Book



Source: [Gitlab's Trusted Data Development]

BUSINESS PROCESSES	COMMON DIMENSIONS: Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Source: [Enterprise Data Warehouse Bus Architecture]

Questions

Details in requirement document (6)

1. Job and Shipment Performance and Trends
    a. Job Revenue (per Location, over Time)
    b. Sales Agent Productivity (from Leads to Jobs, over Time)
    c. Production Trends for Jobs for Entities (time to subjob production)
    d. Shipment Trends for Jobs (contract time to shipment) (compared to shipment promised dates and first shipping dates)
2. Invoice Trends
    a. Highest Invoice Amounts (over Time)
    b. Trends for Invoicing of Job Amounts for Entities (time to invoice) (over Time)
    c. Returns (difference between invoice quantity and shipping quantity) for Products, Machines, and Locations
3. Financial Performance
    a. Gross Margins (per Location)
    b. Locations Gross Margins Variation by Forecast/Budget (per Month)
    c. Products Most Difficult to Forecast
    d. Most Profitable Products and Locations (over Time)

Business Process, Dimensions, Measures

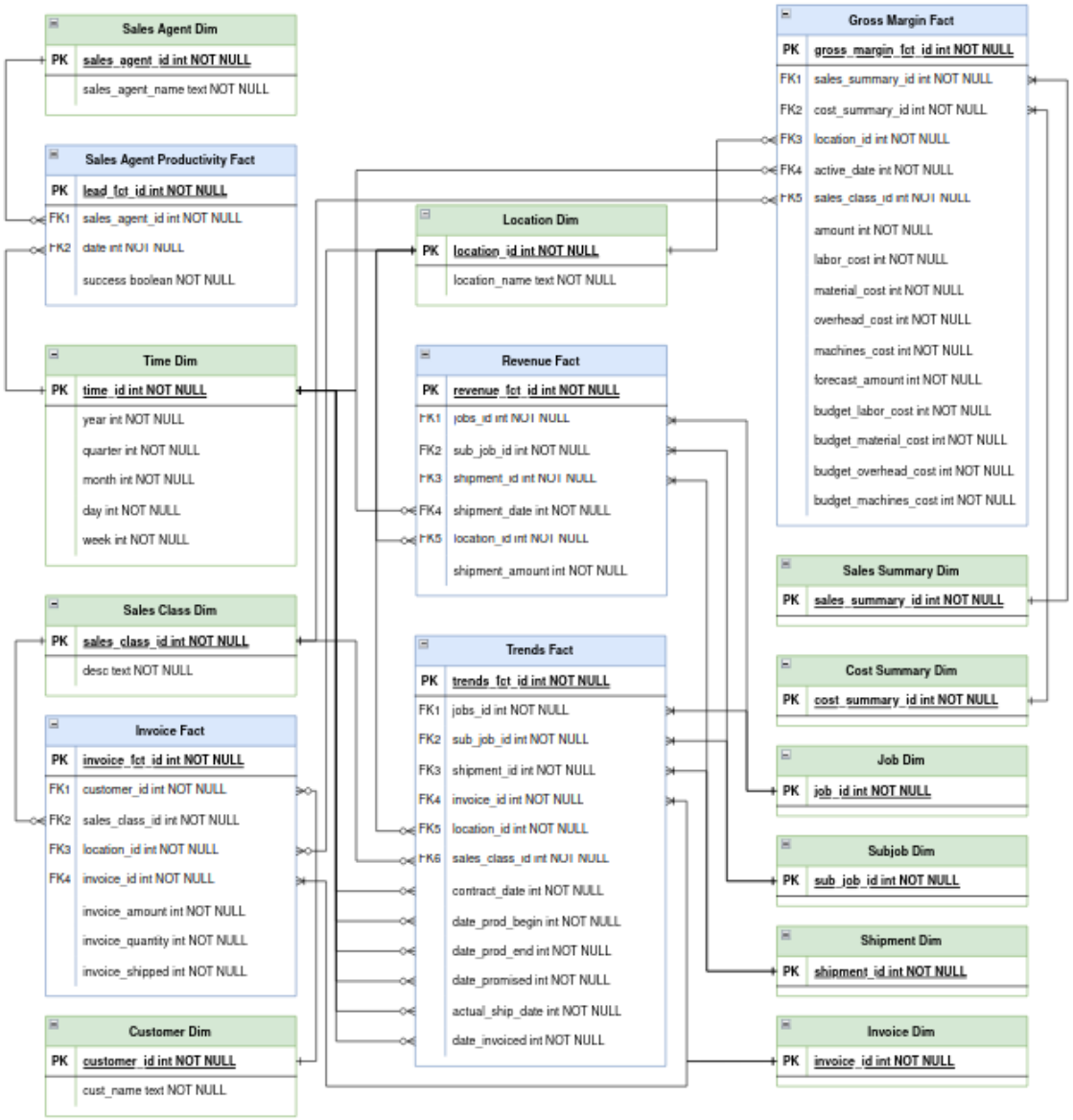
Q	Dims	Attr.	Desc.	Measures	Sources	Agg. Prop.
Revenue	Jobs, Location, Time	Jobs.Unit Price Jobs.Quantity Ordered OR Shipment.Shipment Amount	One table. Price * Quantity	Additive	ERP	SUM Q
Sales Agent Productivity	Sales Agent, Lead, Time	Lead.Success	One table. Group By Success on Distinct Customer ID	Additive	Lead File	SUM Q
Time to Sub Job Production for Entities	Sub Job, Jobs Location, Sales Class	Jobs.Contract Date Sub Job.Date Prod Begin Sub Job.Date Prod End	Jobs Dim. Date Prod Begin - Contract Date	Semi-additive	ERP	AVG Q
Shipment Trends for Jobs	Shipment, Jobs	Jobs.Contract Date Jobs.Date Promised Shipment.Actual Ship Date	Jobs. Dim. Actual Ship Date - Contract Date	Semi-additive	ERP	AVG Q
Entities with Highest Invoice Amounts	Invoice, Customers, Locations, Products	Invoice.Invoice Amount	Entities Dim. Group By Entities	Additive	ERP	SUM Invoice Amounts
Time to Invoice for Locations & Products	Invoice, Locations, Products	Jobs.Contract Date Invoice.Date Invoiced	Entities Dim.	Semi-additive	ERP	AVG Q

Returns	Invoice	Quantity, Shipped	One Table. Quantity - Shipped	Additive	ERP	SUM, AVG Q
Gross Margins per Locations	Sales Summary, Cost Summary	Sales.Amount Cost.Cost * Locations	Amount - Cost	Semi-additive	Summary	AVG Q
Locations Gross Margins Variation by Forecast/Budget (Monthly)	Sales Summary, Cost Summary	Sales.Amount Cost.* Cost Locations Month	Amount - Cost	Non-additive	Summary	AVG, STDEV Q
Products Most Difficult to Forecast	Same as above, Sales Class	Sales.Amount Sales.Forecast Amount Sales.Sales Class	ABS(Amount - Forecast Amount)	Non-additive	Summary	-
Most Profitable Products and Locations	Same as above, Locations	Sales.Amount Cost.* Cost Locations Sales.Sales Class	Amount - Cost	Non-additive	Summary	MAX, AVG Q

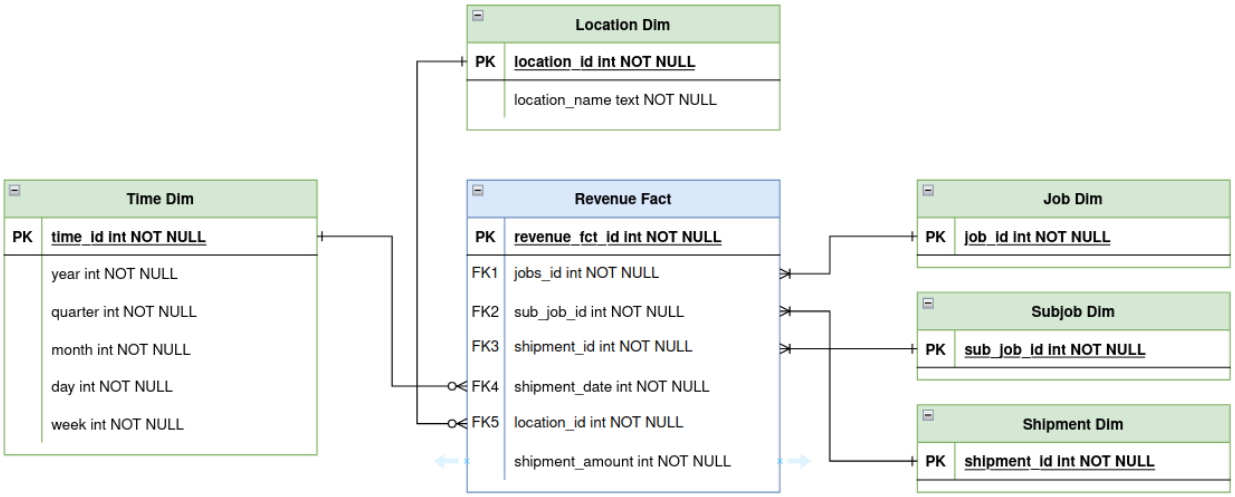
Cube and Grain

Cube	Grain	Dims	Measures	Sparsity
Revenue	Shipment (2,500,000)	Job + Subjob Time	Shipment Amount	0.986
Sales Agent Productivity	Lead (250,000)	Sales Agent Time	Success	-
Trends	Shipment (2,500,000)	Job + Subjob Invoice Location Sales Class Time	Contract Date Date Prod Begin Date Prod End Date Promised Actual Ship Date Date Invoiced	-
Invoice	Invoice (1,000,000)	Location Sales Class Time	Invoice Amount Invoice Quantity Invoice Shipped	-
Gross Margins per Locations	Sales + Cost Summary (5,400)	Location Sales Class Time	Amount Labor Cost Material Cost Overhead Cost Machines Cost Forecast Amount Budget Labor Cost Budget Material Cost Budget Overhead Cost Budget Machines Cost	-

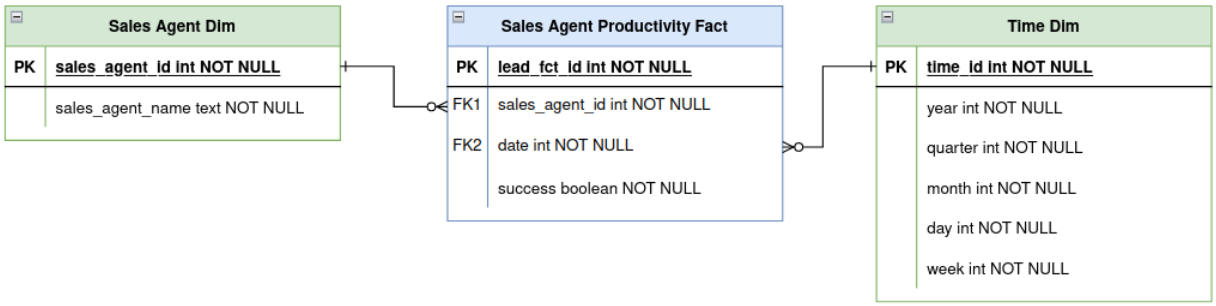
Schemas



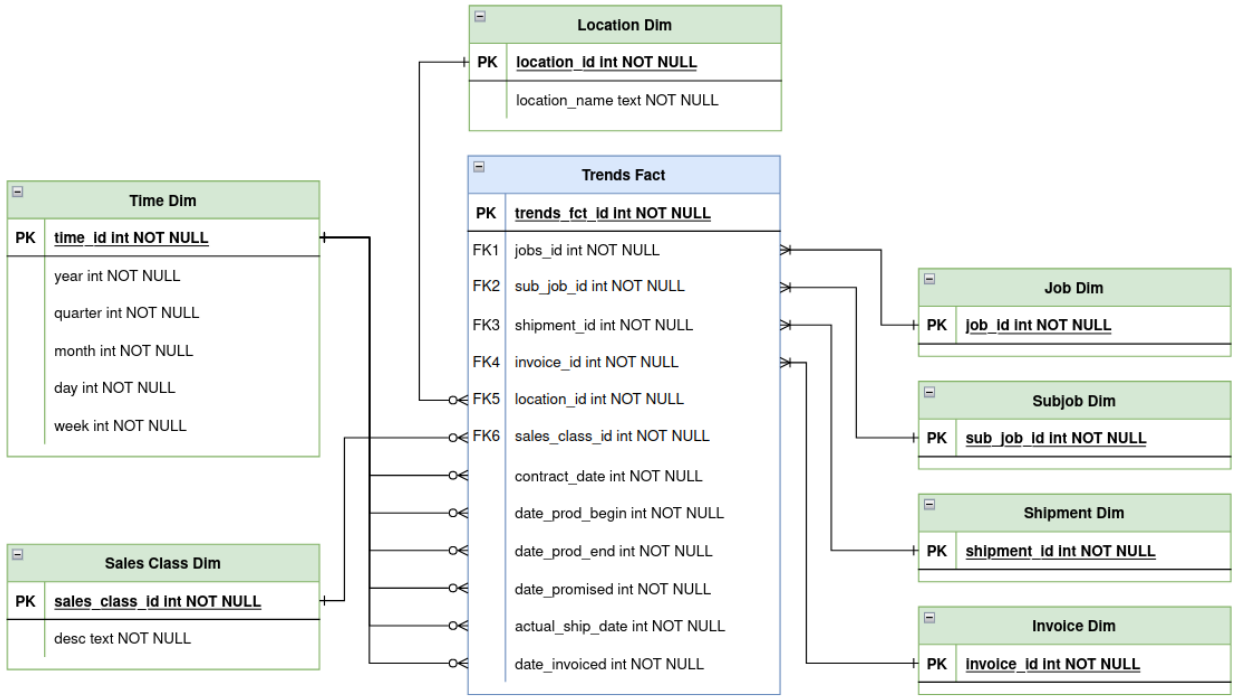
All in One
(blue is fact tables, green is dimension tables. Long dimension details are omitted)



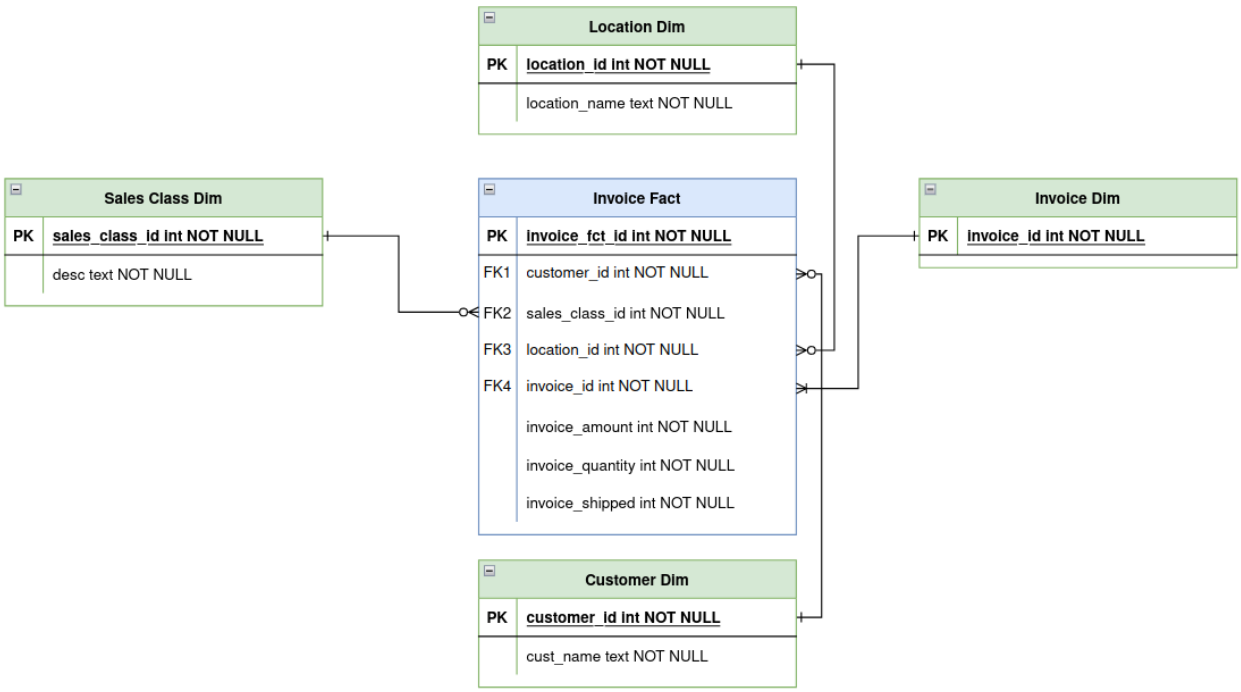
Revenue Fact Table



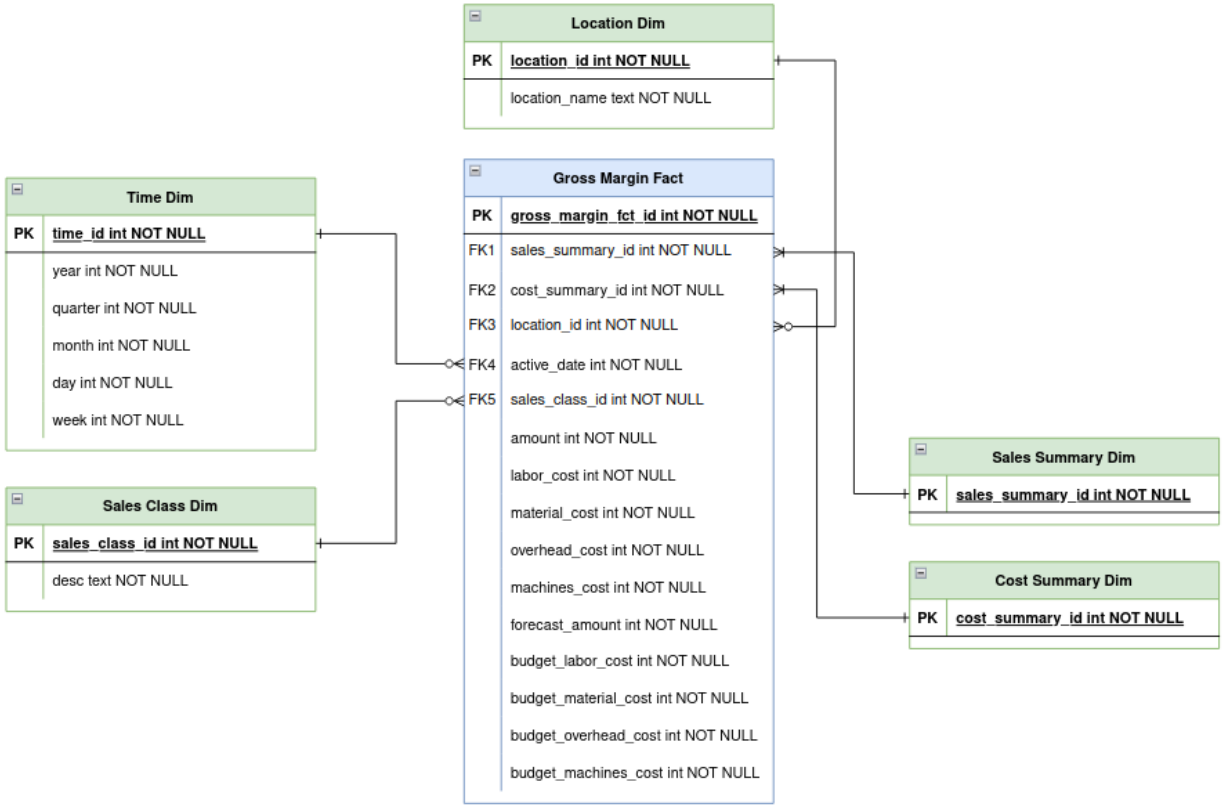
Sales Agent Productivity Table



Trends Fact Table



Invoice Table



Gross Margin Table

Summarizability

No known problems might arise because I designed them with a more complex integration schemes in mind, so that:

- It's *complete* at the finest grain, that is all information needed is there. `NOT NULL` is safe to be given in all columns. Therefore, **incomplete relationship might not happen**
- There's no *non-strict dimension-fact relationship*, because
    - I'm making sure each row is *single-valued* (1NF)
    - There's no *many-to-many* relationship in the source data, so no further M-N simplifications are required

Sample Data

Revenue Fact

revenue_fct_id	jobs_id	sub_job_id	shipment_id	shipment_date	location_id	shipment_amount
1	111111	1	111111	20140115	1	15.000
2	111111	1	111112	20140117	1	22.500
3	111111	2	111113	20140202	1	9.000
4	222222	1	111114	20150126	1	13.500
5	222222	1	111115	20150202	1	13.500
6	222222	2	111116	20150215	1	15.500
7	333333	1	111117	20150410	3	7.000

Time Dim

time_id	year	quarter	month	day	week	desc
20130608	2013	2	6	8	23	lead_fct
20131205	2013	4	12	5	49	lead_fct
20131220	2013	4	12	20	51	
20140101	2014	1	1	1	1	gross_margin_fct
20140103	2014	1	1	3	1	trends_fct
20140105	2014	1	1	5	2	trends_fct

20140110	2014	1	1	10	2	
20140115	2014	1	1	15	3	revenue_fct
20140117	2014	1	1	17	3	revenue_fct
20140124	2014	1	1	24	4	trends_fct
20140125	2014	1	1	25	4	trends_fct
20140131	2014	1	1	31	5	
20140201	2014	1	2	1	5	gross_margin_fct
20140202	2014	1	2	2	6	revenue_fct
20140203	2014	1	2	3	6	
20140215	2014	1	2	15	7	trends_fct
20140224	2014	1	2	24	9	trends_fct
20140228	2014	1	2	28	9	
20140301	2014	1	3	1	9	trends_fct, gross_margin_fct
20140314	2014	1	3	14	11	
20140315	2014	1	3	15	11	lead_fct
20140331	2014	1	3	31	14	
20140401	2014	2	4	1	14	
20140430	2014	2	4	30	18	
20140501	2014	2	5	1	18	
20140517	2014	2	5	17	20	lead_fct

20140531	2014	2	5	31	22	
20140601	2014	2	6	1	23	
20140603	2014	2	6	3	23	
20140615	2014	2	6	15	25	
20140630	2014	2	6	30	27	
20140701	2014	3	7	1	27	
20140707	2014	3	7	7	28	
20140731	2014	3	7	31	31	
20140801	2014	3	8	1	31	
20140810	2014	3	8	10	33	lead_fct
20140831	2014	3	8	31	36	
20140901	2014	3	9	1	36	
20140905	2014	3	9	5	36	
20140915	2014	3	9	15	38	
20140930	2014	3	9	30	40	
20141001	2014	4	10	1	40	
20141005	2014	4	10	5	41	
20141031	2014	4	10	31	44	
20141101	2014	4	11	1	44	
20141120	2014	4	11	20	47	lead_fct

20141130	2014	4	11	30	49	
20141201	2014	4	12	1	49	
20141208	2014	4	12	8	50	
20141215	2014	4	12	15	51	
20141231	2014	4	12	31	53	
20150103	2015	1	1	3	1	trends_fct
20150110	2015	1	1	10	2	trends_fct
20150125	2015	1	1	25	5	trends_fct
20150126	2015	1	1	26	5	revenue_fct
20150202	2015	1	2	2	6	revenue_fct
20150203	2015	1	2	3	6	
20150213	2015	1	2	13	7	
20150214	2015	1	2	14	7	trends_fct
20150215	2015	1	2	15	8	revenue_fct
20150224	2015	1	2	24	9	trends_fct
20150228	2015	1	2	28	9	trends_fct
20150303	2015	1	3	3	10	trends_fct
20150304	2015	1	3	4	10	
20150315	2015	1	3	15	12	
20150403	2015	2	4	3	15	trends_fct

20150407	2015	2	4	7	15	trends_fct
20150410	2015	2	4	10	15	revenue_fct
20150411	2015	2	4	11	15	
20150417	2015	2	4	17	16	lead_fct
20150501	2015	2	5	1	18	trends_fct
20150505	2015	2	5	5	19	
20150510	2015	2	5	10	20	
20150515	2015	2	5	15	20	trends_fct
20150529	2015	2	5	29	22	
20150714	2015	3	7	14	29	lead_fct
20150820	2015	3	8	20	34	
20150901	2015	3	9	1	36	
20151017	2015	4	10	17	42	
20151030	2015	4	10	30	44	

Location Dim

location_id	name
1	Denver HQ
2	Atlanta Division
3	Montreal Plant

4	London Plant
5	Charlotte Division

Jobs Dim

job_id
111111
222222
333333
444444
555555
666666

Subjob Dim

sub_job_id
111111
111111
222222
222222
333333
444444
555555
666666

Shipment Dim

shipment_id
111111
111112
111113
111114
111115
111116
111117

Sales Agent Productivity Fact

lead_fct_id	sales_agent_id	date	success
555555	1	20131205	TRUE
665556	2	20141120	TRUE
690006	3	20150417	TRUE
734033	1	20140810	TRUE

Sales Agent Dim

sales_agent_id	sales_agent_name
1	Ron Jones
2	Tammy Ralston
3	Theo Jackson

633557	1	20140517	TRUE
756778	1	20150714	TRUE
333125	3	20130608	FALSE
673125	2	20140315	FALSE

Trends Fact

trends_fct_id	jobs_id	sub_job_id	shipment_id	invoice_id	location_id	sales_class_id	contract_date	date_prod_begin	date_prod_end	date_promised	actual_ship_date	date_invoiced
1	111111	1	111111	111111	1	1	20140103	20140105	20140117	20140301	20140115	20140124
2	111111	1	111112	111111	1	1	20140103	20140105	20140117	20140301	20140117	20140124
3	111111	2	111113	111112	1	1	20140103	20140125	20140215	20140301	20140202	20140224
4	222222	1	111114	111113	1	2	20150103	20150110	20150214	20150303	20150126	20150224
5	222222	1	111115	111113	1	2	20150103	20150110	20150214	20150303	20150202	20150224
6	222222	2	111116	111113	1	2	20150103	20150125	20150228	20150303	20150215	20150224
7	333333	1	111117	111114	3	3	20150403	20150407	20150410	20150515	20150410	20150501

Invoice Dim

invoice_id
111111
111112

Sales Class Dim

sales_class_id	sales_class_desc
1	Credit Card Smart
2	Credit Card Mag

Customer Dim

customer_id	cust_name
111111	First Credit
111112	Second Credit

111113
111114

3	Debit Card Smart
4	Debit Card Mag
5	Prepaid Mag
6	Loyalty Mag

111113	Third Credit
111114	First Bank

Invoice Fact

invoice_fct_id	customer_id	sales_class_id	location_id	invoice_id	amount	quantity	shipped
1	111111	1	1	111111	37.500	25.400	25.500
2	111111	1	1	111112	9.000	9.000	9.000
3	111112	2	1	111113	42.500	30.900	32.000

Gross Margin Fact - 1st part

gross_margin_fct_id	sales_summary_id	cost_summary_id	location_id	active_date	sales_class_id	amount	labor_cost
1	1	1	1	20140101	1	993.800	326.250
2	2	2	1	20140201	1	1.883.800	603.563
3	3	3	1	20140301	1	3.103.800	880.875
4	4	4	1	20140401	1	3.896.800	1.158.188
5	5	5	1	20140501	1	5.116.800	1.435.500
6	6	6	1	20140601	1	6.336.800	1.712.813

7	7	7	1	20140701	1	7.556.800	1.990.125
8	8	8	1	20140801	1	8.776.800	2.332.688
9	9	9	1	20140901	1	9.996.800	2.610.000
10	10	10	1	20141001	1	11.216.800	2.926.463
11	11	11	1	20141101	1	12.436.800	3.203.775
12	12	12	1	20141201	1	13.656.800	3.523.500

Gross Margin Fact - 2nd part

material_cost	overhead_cost	machines_cost	forecast_amount	budget_labor_cost	budget_material_cost	budget_overhead_cost	budget_machines_cost
191.250	135.000	202.500	13.500.000	3.915.000	`2.295.000	1.620.000	2.430.000
353.813	249.750	374.625	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
516.375	364.500	546.750	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
678.938	479.250	718.875	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
841.500	594.000	891.000	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
1.004.063	708.750	1.063.125	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
1.166.625	823.500	1.235.250	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
1.367.438	965.250	1.447.875	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
1.530.000	1.080.000	1.620.000	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
1.715.513	1.210.950	1.816.425	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000

1.878.075	1.325.700	1.988.550	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000
2.065.500	1.458.000	2.187.000	13.500.000	3.915.000	2.295.000	1.620.000	2.430.000