

---

# Building a Data Warehouse

---

## A Practical Example

Victoria Shoemaker

Taurus Software

[vas@taurus.com](mailto:vas@taurus.com)

# **Why A Data Warehouse Or Data Mart?**

Keys to success in a data warehouse

# **What Is A Data Warehouse Or Data Mart?**

Technical goals of a data warehouse

Structure of a data warehouse

# **Building A Data Warehouse Step By Step.**

Case Study

# Why A Data Warehouse Or Data Mart?

Competitive advantage

 *Shopping Cart*

Picture of hospital bed

Picture of mfg

\$

Value - ROI is 5 times cost of mart in first year.

# Keys To Success In a Data Warehouse Or Data Mart

Senior Business V.P. + ROI = Project

High-level  
management  
commitment

= Funding

= Building something that will be used; “if you build it they will come” is NOT true!

= Building something that will make a \$ difference to the business, and know what that \$ difference will be.

# Finding Value In A Data Warehouse Or Data Mart



**Taurus**

software

*We Make Data Liquid*

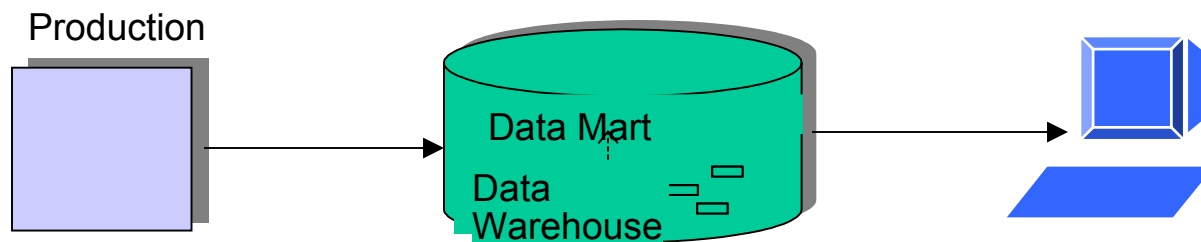
## SalesAnalysis: Drivers of Profit

© Taurus Software, Inc. 1999.

Issues/Opportunity	Remedy	Capability Needed
<p>Our revenue is growing at less than the plan rate.</p> <p>We are not capitalizing on best customers.</p>	<p>Sell more to same customers by cross selling products customer does not currently buy.</p> <p>Get customers to commit to volume purchases, individually or by Parent/Group.</p> <p>Get all Sales people or channels to emulate their most successful colleagues.</p> <p>Apply more sales and support to top customers.</p> <p>Deal direct with largest customers sold by our Channels.</p>	<p>Analyze product sales across similar customers based on size, geography, industry.</p> <p>Understand revenue by customer by product over three years.</p> <p>Informally associate customers with their Parent to see total revenue value.</p> <p>Measure Sales by customer by rep by rep experience level</p> <p>Know who the top 20 customers are, or the 80/20 set.</p> <p>Measure sales made by Channel to and match to our customer or Parent.</p>

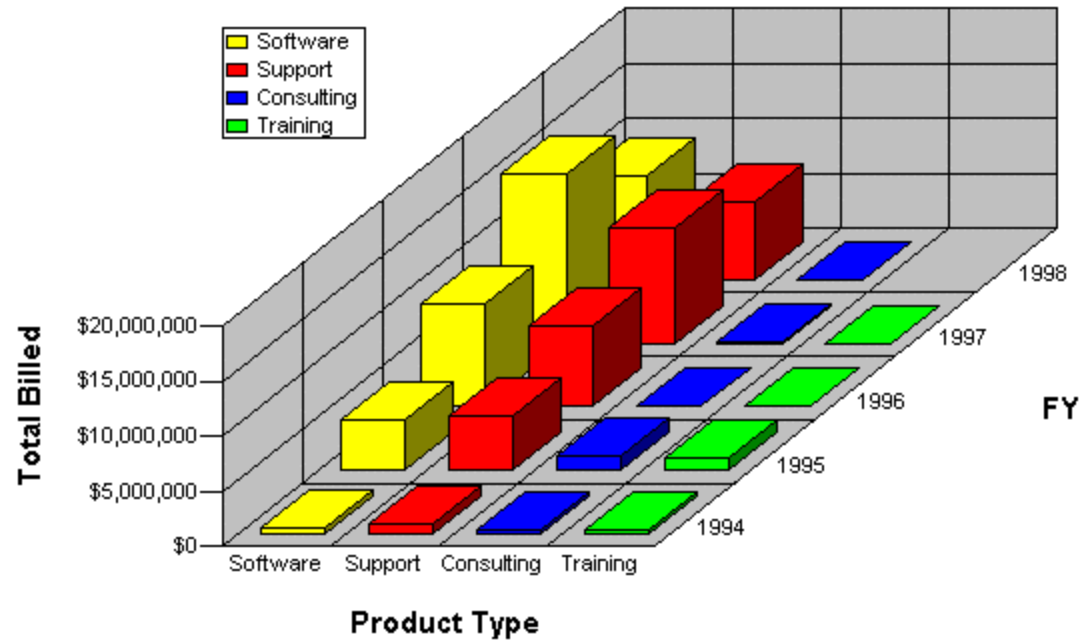
# What Is A Data Warehouse Or Data Mart?

- A “place” for top executives, managers, analysts, and other end-users to mine a rich source of company information from their desktops using GUI OLAP (On-Line Analytical Processing) tools.
- Allows managers to make decisions based on facts, not intuition.



Sort FY by Label using Sum

### Product Revenue by Type



# Technical Goals Of A Data Warehouse Or Data Mart

Meet needs of decision maker:

- Stable environment to allow consistency during analysis.
- Source conflicts are resolved once.
- Historical analysis is possible.
- Minimal interference with operational systems.
- Read-only environment allows better performance.

Provides easy access to corporate data:

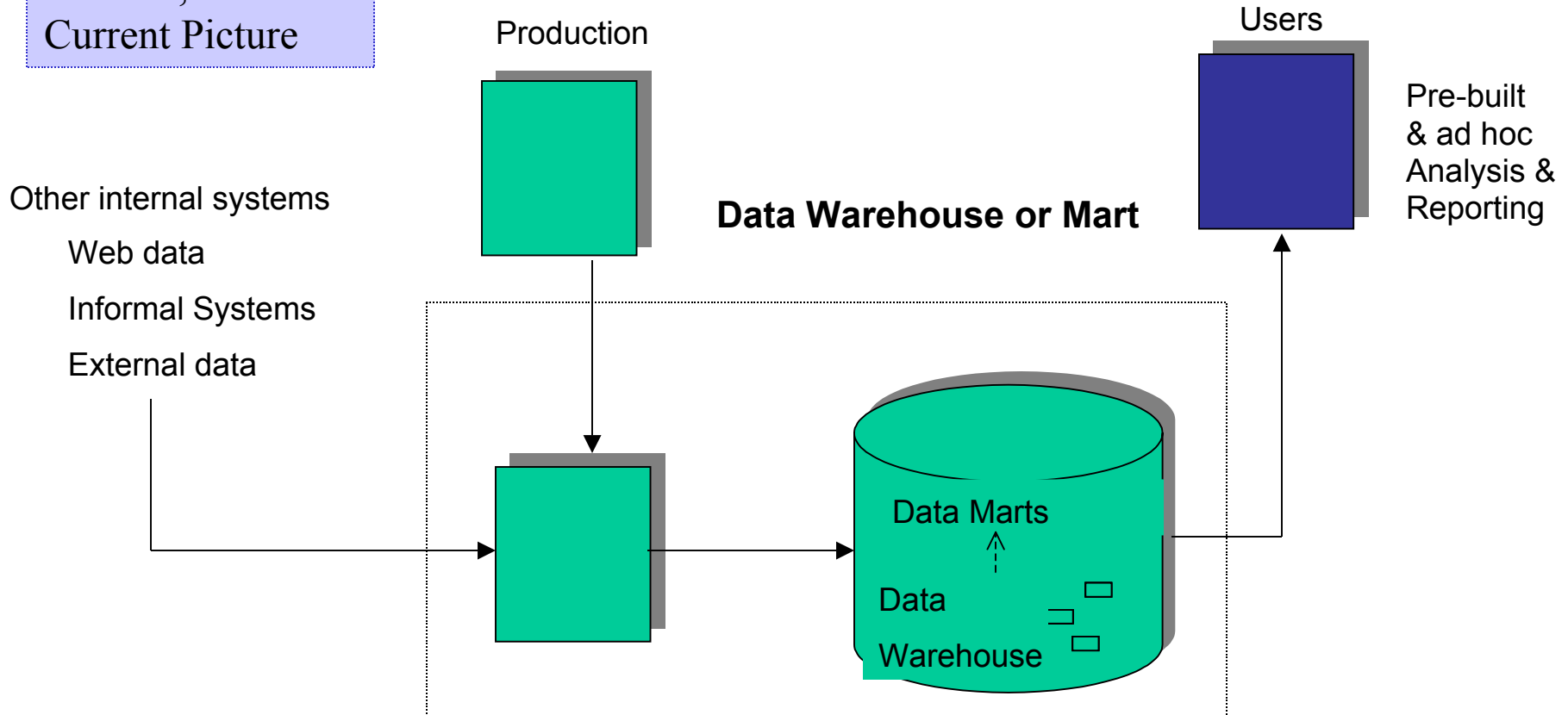
- End-user access tools must be easy to use.
- Access should be graphic, so it is easy to spot trends.
- Access should be manageable by end users.
- Queries must be fast.
- No IT intervention between the user and the mart.



# Structure Of A Data Warehouse

Data is:  
Integrated,  
Subject-oriented,  
Non-volatile,  
Time Referenced

Data is:  
Unintegrated,  
Detailed  
Process-oriented,  
Volatile,  
Current Picture



# Structure Of A Data Warehouse

## The Data

Subject oriented	Not transaction oriented. Re-group data by subject: customers, sales, products
Time referenced	Prior states of being. GMROI, backlog this month last year vs this year.
Integrated	Merged from many sources.
Read only	Never modify the Warehouse. Don't want to change history. Not designed to handle updates (de-normalized data)

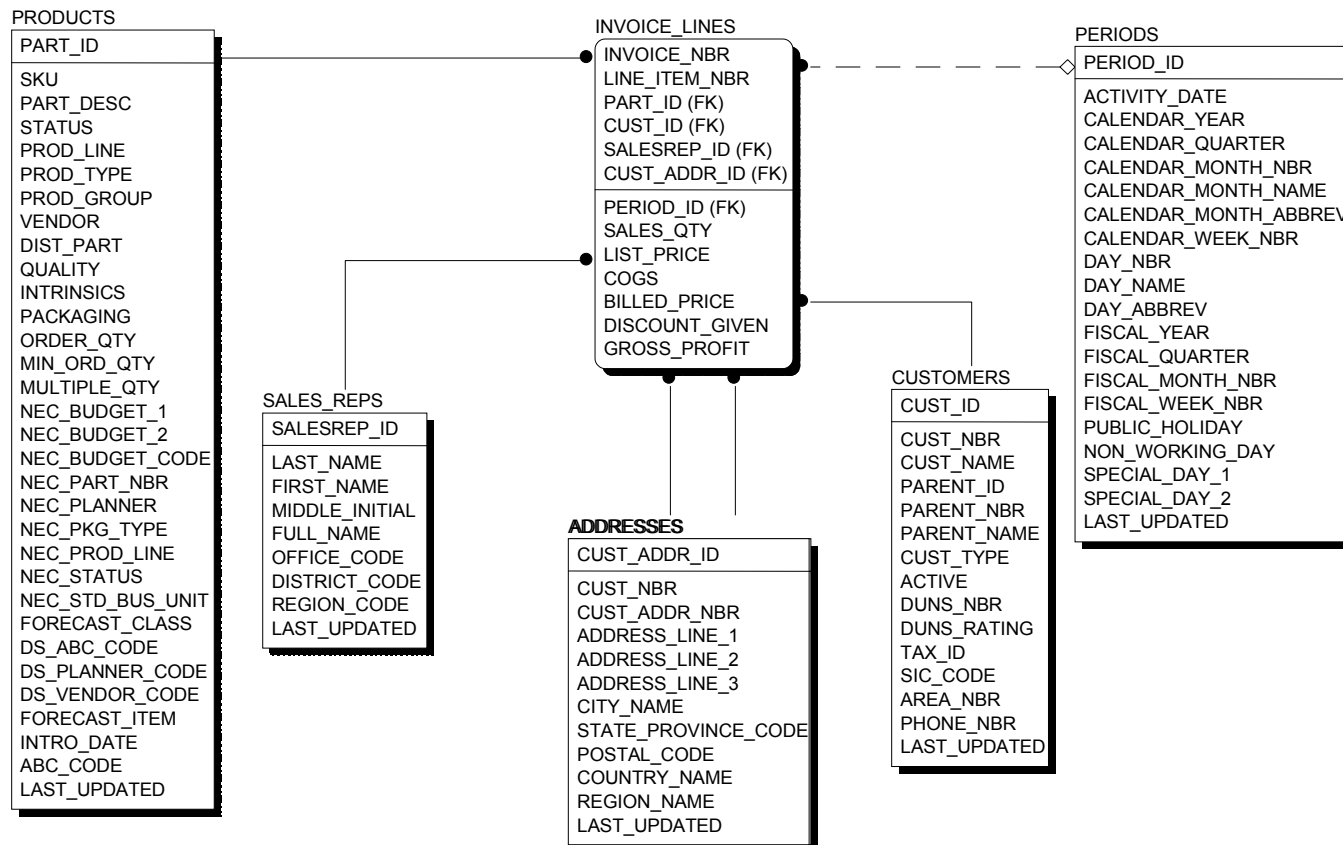
# Structure Of A Data Warehouse

The Data Model

Star Schema

One fact table: Invoices, Shipments, and Returns

Dimensions: Contain data descriptions and keys into the fact tables. Used to query about facts.

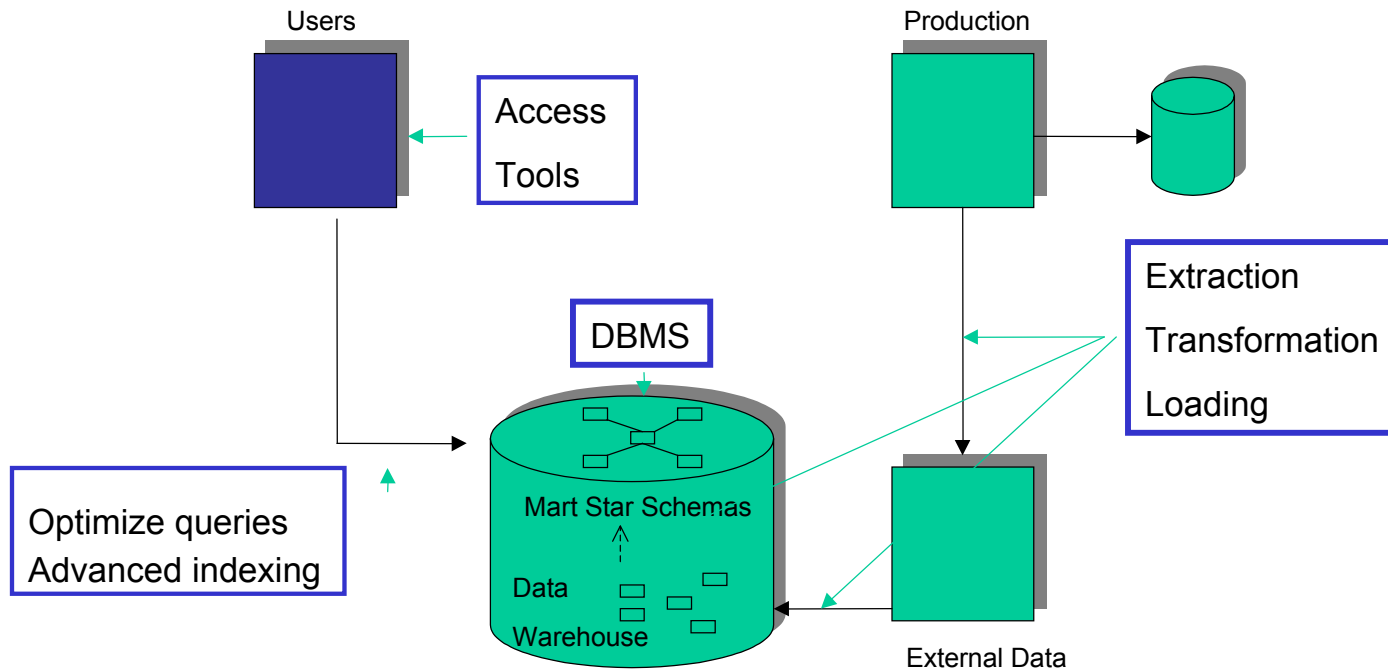


# Structure Of A Data Warehouse

The Environment

- ETL
- DBMS
- Query Optimization
- Access Tools

## Data Warehouse & Marts



# Building A Data Warehouse Step By Step

Case Study -

- 1) Analyze the business need and determine value.
- 2) Pick the scope of initial deployment.
- 3) Build the target model.
- 4) Map data from production to the mart.
- 5) Extract operational and external data and load the mart.
- 6) Optimize queries into the mart.
- 7) Deploy end user interface.
- 8) Evaluate value of the mart to the end user.

## **Our Case Study :**

Distributes electric components

Sales channel has direct sales and distributors

Wanted to create reporting and analysis environment

## Step 1: Analyze the business need and determine value

1) Get executive sponsorship

Critically important for MIS to work with users/management

MIS then knows what users want

MIS has management sponsorship

MIS has non-MIS funding

A) Interview key executives to uncover issues

“We will decide to....”

B) Assign value to “decide to’s”

# Step 1: Analyze the business need and determine value



**Taurus**

software

*We Make Data Liquid*

## Sales Analysis Drivers of Profit

©Taurus Software, Inc. 1999.

Who Cares? About what?	Issues/Opportunity	Remedy	Value
<p><b>VP</b></p>	<p>Paid on Gross Margin and can't calculate it from production system</p> <p>Want to increase sales of high profit margin products</p> <p>Want to target successful sales channels and create more like them</p>	<p>Provide details on sales including Gross Margin amounts.</p> <p>Determine if can swap in products with high profit margins for sales of similar products with low margins</p> <p>Analyze successful channels to determine what demographics make them up then target similar channels</p>	<p><b>VP Gets his bonus</b></p> <p><b>5% Increase in profit</b></p> <p><b>5% Increase in profit</b></p>



## Step 2: Pick scope of initial deployment

Smaller is better

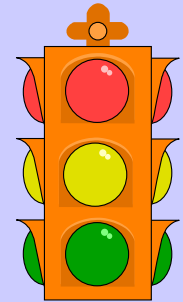
Quick success

User excitement and buy-in

Easier

A) Take each issue uncovered by Step 1 and make certain data is currently available in source systems.

Green light, yellow light, red light



B) Pick issue with most value and green lights.

# Step 2: Our Case Study's scope of initial deployment



## Sales Analysis Drivers of Profit

© Taurus Software, Inc. 1999.

Who Cares? About what?	Issues/Opportunity	Remedy	Technical Data Available?
VP	<p>Paid on Gross Margin and can't calculate it from production system</p> <p>Want to increase sales of high profit margin products</p> <p>Want to target successful sales channels and create more like them</p>	<p>Provide details on sales including Gross Margin amounts.</p> <p>Analyze product sales by margin. Determine if can swap high margin products for high selling low margin products.</p> <p>Analyze successful channels to determine what demographics make them up then target similar channels</p>	<p>Three circles are shown vertically. The top two are green, and the bottom one is red with an asterisk next to it.</p>

\* No demographic data in source systems

# Step 3: Build the target model

## 1) Select grain for fact table

Lowest level data (e.g. Invoice line item) available for analysis

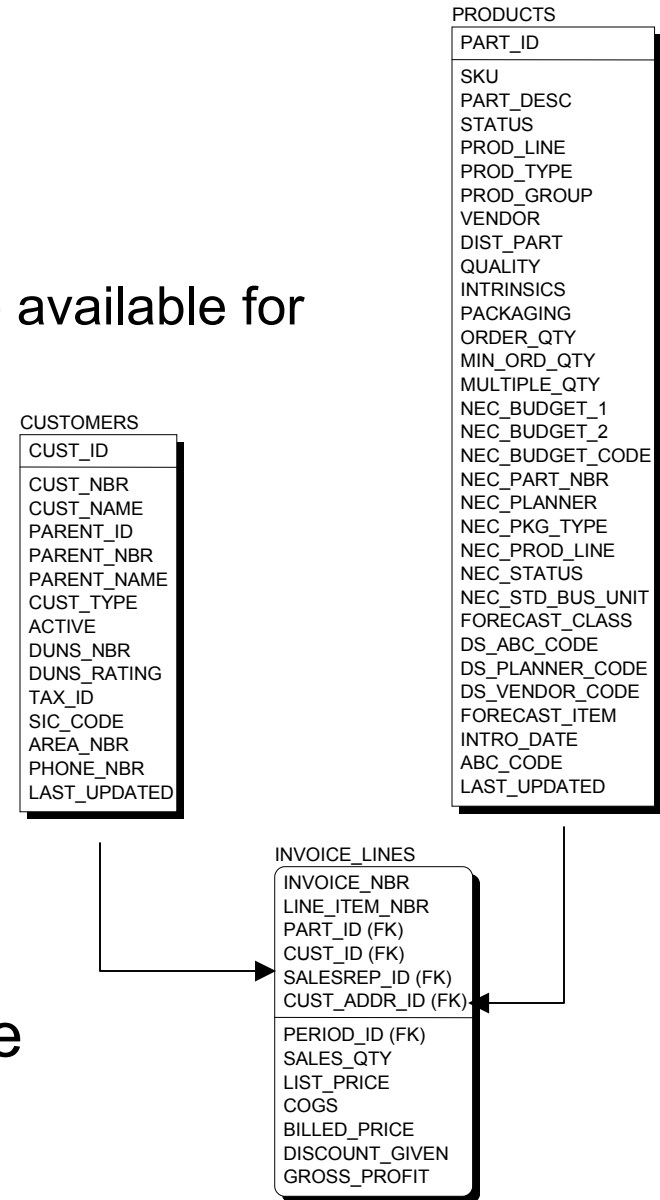
## 2) Select the dimensions

Attributes by which to analyze

## 3) Select individual facts for fact table

Numeric indicators for fact table

## 4) Complete design & walk through use



# Step 4: Map Data From Production To The Mart

“The process of sourcing data to feed the data warehouse typically takes over 75% of the project in terms of effort”

Map Facts  
Then Dimensions

Target Table: CUSTOMERS

Source Field	Transformation Rules	Target Field
cust.cust name		cust name
parent.parent id	Check to see if parent already exists, if not then add	parent id
parent.user cust no		parent nbr
parent.cust name		parent name
no source	constant: SRCAPPL	source system
no source	constant: current time/date	last updated
cust seg.active flag	Retrieve from cust_seg using cust_no and seg no = 317	active flag
cust commcn.area cd	Retrieve from cust_commcn using cust no	area nbr
cust commcn.phone no	Retrieve from cust_commcn using cust no	phone no

## Step 5: Extract Operational Data And Load Mart

### Pick Population Strategy

Determine data volumes

Determine how long full load takes

Determine refresh frequency

Defined by users requirements

Frequency of update -  
daily, weekly?

Is the change valuable?

(Multiple adjustments to inventory)

### Our Case Study:

Small volumes but growing rapidly

Full load takes 6 hours

Wanted daily updates

- Log files incomplete
- No date and time stamps on data
- Too much data to move full db daily
- Wanted to know value of changes

Trapped changes at system level

- Posted changes at end of day
- Date and time stamp data
- Date and time stamp effective dates

# Step 5: Extract Operational Data And Load Mart

## Data Cleansing

### Standardization

- Value Substitution

- Lookup Xref Tables

### Filter/Cleanse

- Updating Existing Data

- Standard Functions For  
(e.g.) State/Code  
Substitutions

- Lookups

### Aggregation

- Consolidation

- Summarization

### Error detection and correction

## Our Case Study:

Decode product codes to product names

Remove duplicate customers

Put consistent geographical identifiers

Consolidate all drop-ship to single customer

Minimum requirements: customer has valid address, products must be valid

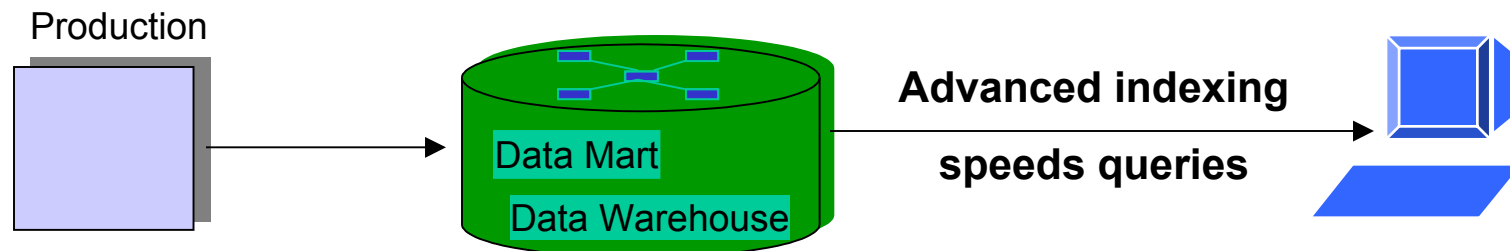
## Step 6: Optimize queries into the mart

***“I should have paid more attention to performance”***

is the #1 complaint of Data Warehousing managers, according to Alan Paller at The Data Warehousing Institute (TDWI)

### Advanced Multidimensional and Aggregation Indexes:

- Eliminate slow serial reads on large “fact” data sets / tables
- Eliminate the need to pre-build summary tables to resolve queries
- Enhance existing databases; work with PC tools



## **Step 6: Optimize queries into the mart**

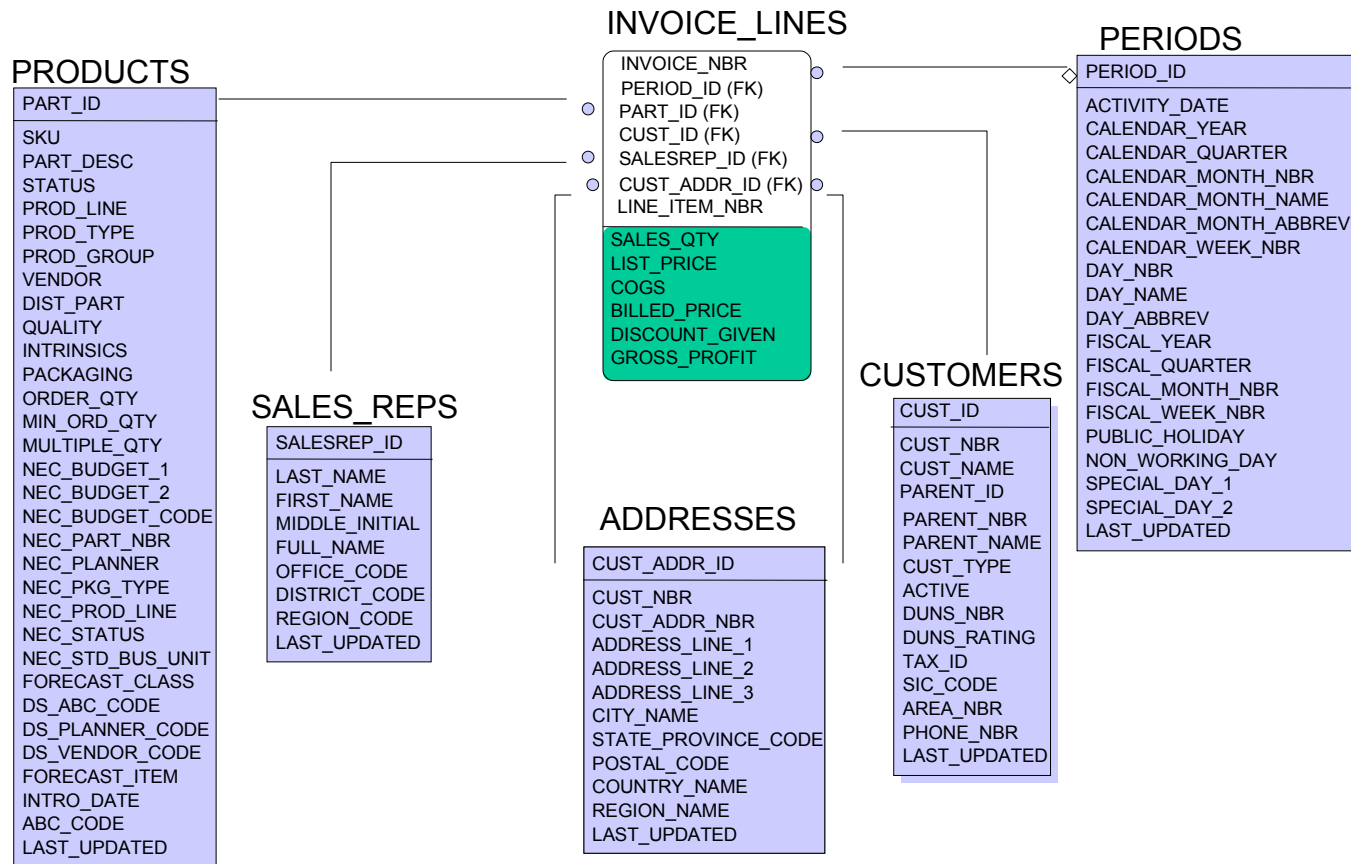
### **Our Case Study - Needs Gross Margin Analysis**

Options:

- **Cubes**
- **Summary tables**
- **Aggregated index technology**



# Step 6: Build Aggregation Scheme



## Step 7: End user interface and deployment

- a) VP wants to know what was his Gross Margin
- b) Decide to sell more high margin products

BrioQuery - EISprod.BQY

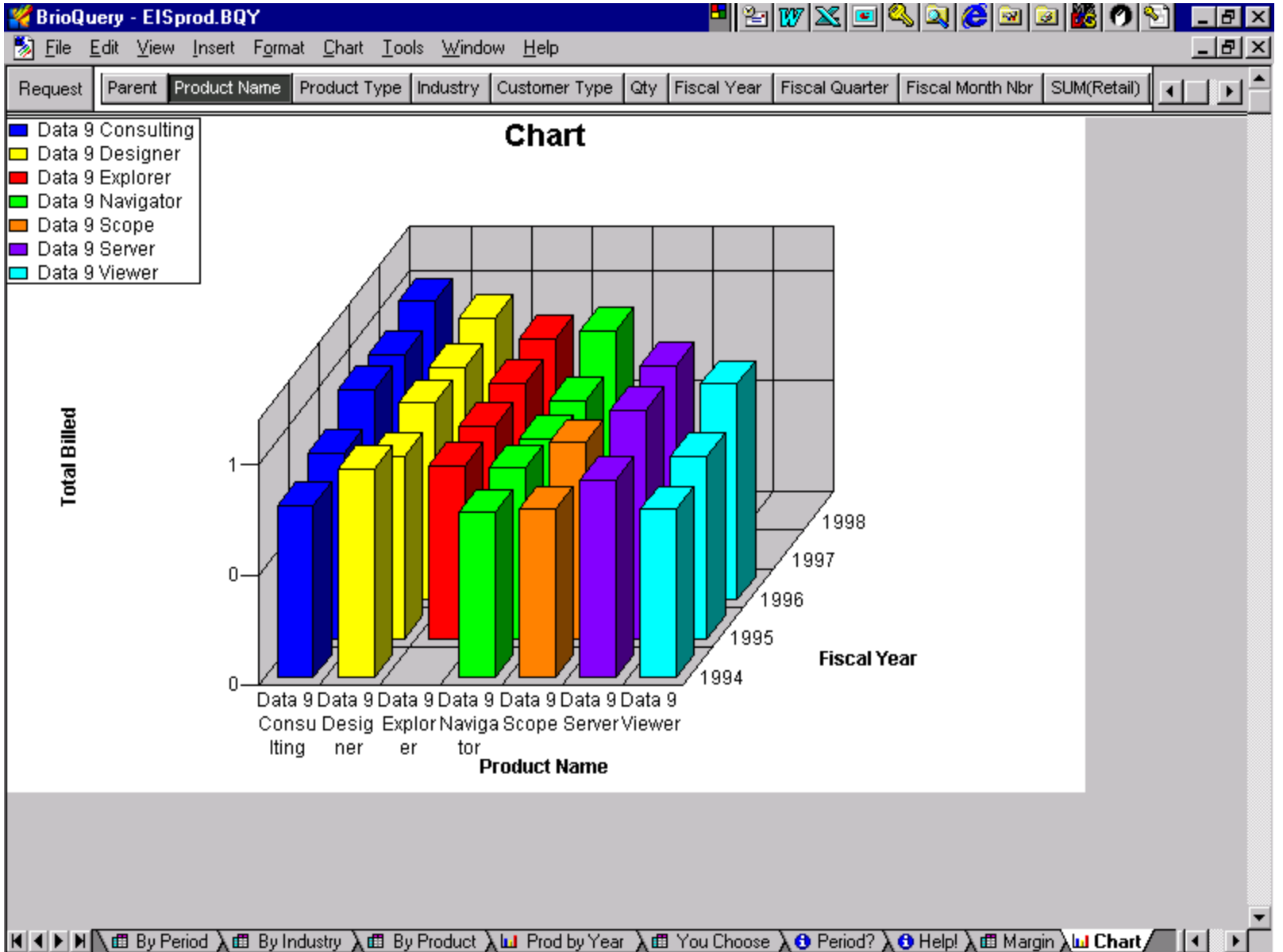
File Edit View Insert Format Pivot Tools Window Help

Sort Product Name by Margin Pct using Sum

Fiscal Year	1994	1995	1996	1997	1998	Total
Product Name	Margin Pct	Margin Pct	Margin Pct	Margin Pct	Margin Pct	Margin Pct
Data 9 Navigator	36%	36%	36%	36%	42%	37%
Data 9 Explorer		38%	39%	39%	41%	40%
Data 9 Consulting	37%	40%	48%	47%	50%	40%
Data 9 Scope	38%	44%				41%
Data 9 Viewer	38%	41%	49%			42%
Data 9 Designer	47%	41%	44%	44%	45%	44%
Data 9 Server	45%	52%	53%			50%
Total	40%	41%	40%	40%	42%	41%

Results By Period By Industry By Product Prod by Year You Choose Period? Help! Margin

No Results Yet 6x8



		Fiscal Year	1994	1995	1996	1997	1998	Total	Total
Industry	Customer Type	Product Name	Margin Pct	Margin Pct	Margin Pct	Margin Pct	Margin Pct	Margin Pct	Margin Pct
		Data 9 Consulting	50%	50%	50%			50%	50%
		Data 9 Server			55%			55%	55%
		<b>Total</b>	35%	43%	44%	43%	46%	43%	43%
Services	Direct Customer	Data 9 Explorer		44%	44%	43%	47%	44%	44%
		Data 9 Navigator		46%	45%	45%	46%	46%	46%
		Data 9 Designer		51%	48%	47%	46%	47%	47%
		Data 9 Consulting	40%	50%	46%			48%	48%
		Data 9 Server		55%				55%	55%
		<b>Total</b>	40%	49%	45%	45%	46%	46%	46%
Telecom	Direct Customer	Data 9 Navigator	38%	19%	36%	33%	40%	32%	32%
		Data 9 Viewer		38%				38%	38%
		Data 9 Explorer		47%	37%	36%	46%	40%	40%
		Data 9 Consulting	33%	43%	50%			42%	42%
		Data 9 Designer		44%	40%	43%	44%	43%	43%
		Data 9 Server		50%				50%	50%
<b>Total</b>	35%	38%	38%	39%	45%	40%	40%		
Transportation	Direct Customer	Data 9 Navigator		21%	44%	37%	35%	30%	30%
		Data 9 Consulting	40%	42%				41%	41%
		Data 9 Explorer		46%	39%	45%	50%	42%	42%
		Data 9 Designer		44%	44%	47%	47%	46%	46%
		Data 9 Server	52%	53%				53%	53%
<b>Total</b>	47%	36%	40%	44%	46%	42%	42%		
Utilities	Direct Customer	Data 9 Navigator		48%		27%	13%	20%	20%
		Data 9 Consulting	40%	33%				35%	35%
					46%	27%	46%	44%	44%

**BrioQuery - EISprod.BQY**

File Edit View Insert Format Pivot Tools Window Help

Add Remove Modify Limit Process  $\sum$  Outliner

Request Product Name Product Type Industry Customer Type Qty Fiscal Year Fiscal Quarter Fiscal Month Nbr SUM(Retail) SUM(Bill)

Sort Industry by Yield using Sum

Industry	Product Name	Yield
Reseller	Data 9 Consult	38%
	Data 9 Design	43%
	Data 9 Explore	36%
	Data 9 Navigat	36%
	Data 9 Scope	48%
	Data 9 Server	49%
	Data 9 Viewer	38%
	<b>Total</b>	<b>38%</b>
Hi Tech	Data 9 Consult	40%
	Data 9 Design	41%
	Data 9 Explore	38%
	Data 9 Navigat	35%
	Data 9 Scope	38%
	Data 9 Server	51%
	Data 9 Viewer	38%
	<b>Total</b>	<b>39%</b>
Intra Company	Data 9 Consult	30%
	Data 9 Design	41%
	Data 9 Explore	38%
	Data 9 Navigat	41%
	<b>Total</b>	<b>39%</b>
		<b>44%</b>

**Outliner**

Drag request items here to create top labels

Industry  
Product Name  
Billed  
Margin  
Yield

Results By Period By Industry **By Product** Prod by Year You Choose Period? Help! Margin

7856 of 7856 Rows 1x132

**BrioQuery - EISprod.BQY**

File Edit View Insert Format Pivot Tools Window Help

Add Remove Modify Limit Process  $\uparrow\downarrow$   $\uparrow\downarrow$   $\Sigma$  Outliner

Request Product Name Product Type Industry Customer Type Qty Fiscal Year Fiscal Quarter Fiscal Month Nbr SUM(Retail) SUM(Bill)

Sort Industry by Yield using Sum  $\uparrow\downarrow$   $\uparrow\downarrow$

Industry	Product Name	Yield
Reseller	Data 9 Consult	38%
	Data 9 Design	43%
	Data 9 Explore	36%
	Data 9 Navigat	36%
	Data 9 Scope	48%
	Data 9 Server	49%
	Data 9 Viewer	38%
	<b>Total</b>	<b>38%</b>
Hi Tech	Data 9 Consult	40%
Intra Company		
		44%

**Outliner**

Drag request items here to create top labels

Industry Product Name Billed Margin Yield

**Drilldown Into**

- Add Computed Item...
- Focus on Items
- Hide Items
- Show All Items
- Font...
- Style
- Justify
- Data Labels
- Corner Labels
- Use Surface Values

- Product Type
- Customer Type
- Qty
- Fiscal Year
- Fiscal Quarter
- Fiscal Month Nbr
- Retail
- Cost
- State/Province
- Region

Results By Period By Industry **By Product** Prod by Year You Choose Period? Help Margin

Move, size or close window 7856 of 7856 Rows 1x132

**BrioQuery - EISprod.BQY**

File Edit View Insert Format Pivot Tools Window Help

Add Remove Modify Limit Process Outline

Request Parent Product Name Product Type Industry Customer Type Qty Fiscal Year Fiscal Quarter Fiscal Month Nbr SUM(Retail)

Sort Industry by Label using Sum

Industry	Product Name	Parent	Yield
Hi Tech	Data 9 Scope	Honeywell	38%
	Data 9 Viewer	Aula de Noves Technologies Informa	38%
	Data 9 Navigator	3DO Company	48%
		Abacus Concepts	48%
		About Software	48%
		Ace Computer	34%
		Advanced Micro Devices	53%
		Aha Computer	48%
		Allied Signal	44%
		Amdahl	46%
		AMSC	44%
		Andersen Consulting	35%
		Apple Computer	44%
		Arthur Andersen	48%
		Aula de Noves Technologies Informa	21%
		Automation Group	22%
		Booze, Allen & Hamilton	29%
		By Design Computer Systems	48%
		Client Server Professional	44%
	Complete Data Solutions	48%	
	Computer Dynamics	48%	
			43%

8095 of 8095 Rows 1x2287