



# Data Warehousing on the MPE Platform

**Presentation #272**

*Miklos Boldog*

*Speedware Corporation*

*9999 Boulevard Cavendish, #100*

*St. Laurent, Quebec Canada H4M 2X5*

*1.800.361.6782 Fax: 1.514.747.3320*

*Mboldog@speedware.com*



# Abstract

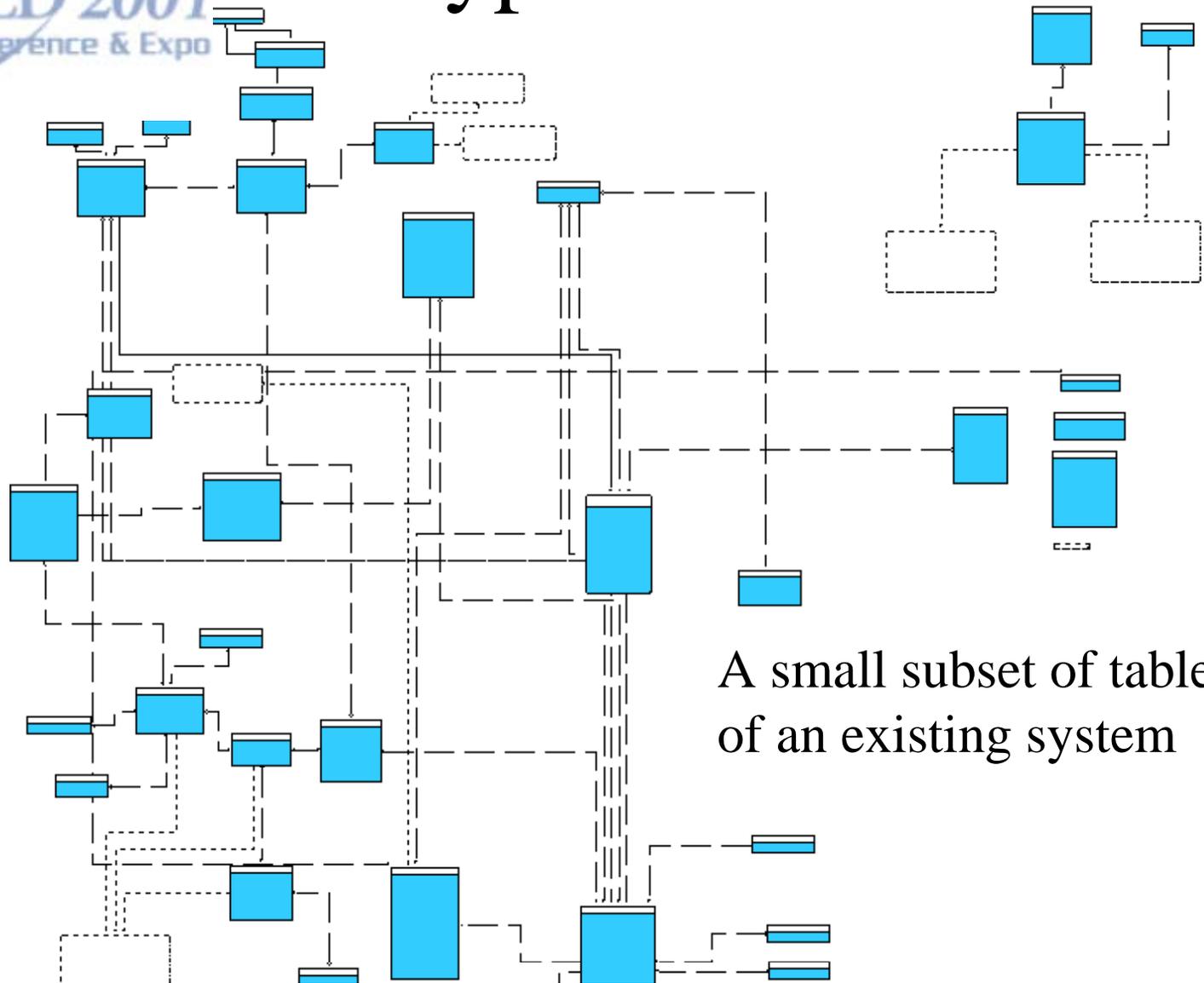
- Company Data spread over multiple systems
  - Operations, Human Resources, Customers, Sales etc..
- ‘Fixed’ reports, change is difficult
- Ad-Hoc questions may not be possible or difficult
- Merging Disparate reports data time consuming
- Most system tuned to getting data in, not getting data out



# Client Concerns

- Complex entity relation data models are difficult for Client's to understand.
- Ad-hoc queries comprising of millions of rows may not return any result sets in a timely manner if at all.
- Running queries against huge tables may 'disrupt' DBA relationships

# Typical ERD's



A small subset of tables  
of an existing system



CUSTOMER	
<b>PK</b>	<b><u>CUSTKEY</u></b>
	NAME STREET CITY STATE ZIP

SHIPMENTS	
<b>PK,FK4</b>	<b><u>PRODKEY</u></b>
<b>PK</b>	<b><u>INVOICE</u></b>
FK1,11 FK2,13 FK3,14	PERKEY CUSTKEY SHIPKEY DOLLARS WEIGHT

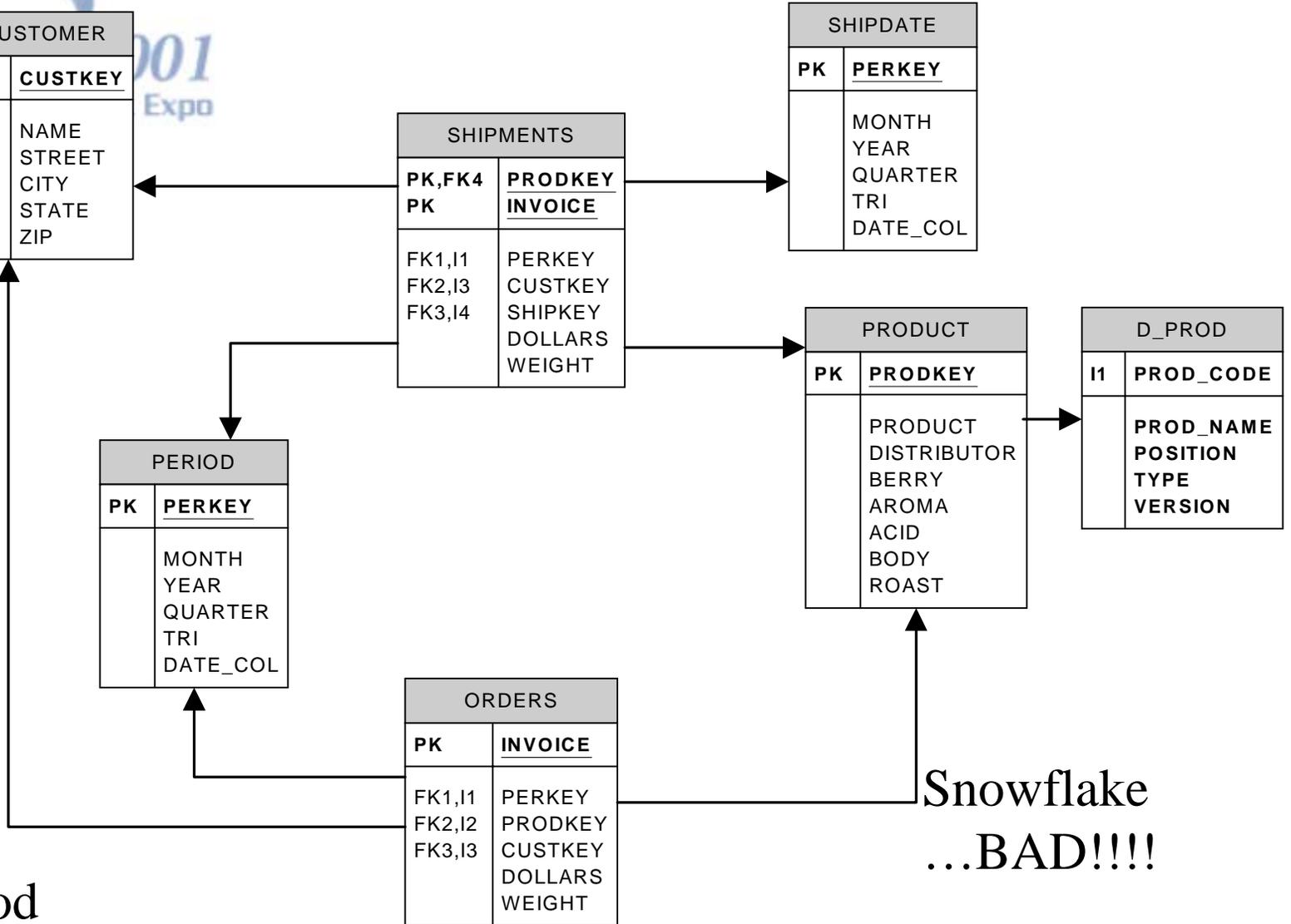
SHIPDATE	
<b>PK</b>	<b><u>PERKEY</u></b>
	MONTH YEAR QUARTER TRI DATE_COL

PERIOD	
<b>PK</b>	<b><u>PERKEY</u></b>
	MONTH YEAR QUARTER TRI DATE_COL

PRODUCT	
<b>PK</b>	<b><u>PRODKEY</u></b>
	PRODUCT DISTRIBUTOR BERRY AROMA ACID BODY ROAST

D_PROD	
<b>I1</b>	<b><u>PROD_CODE</u></b>
	PROD_NAME POSITION TYPE VERSION

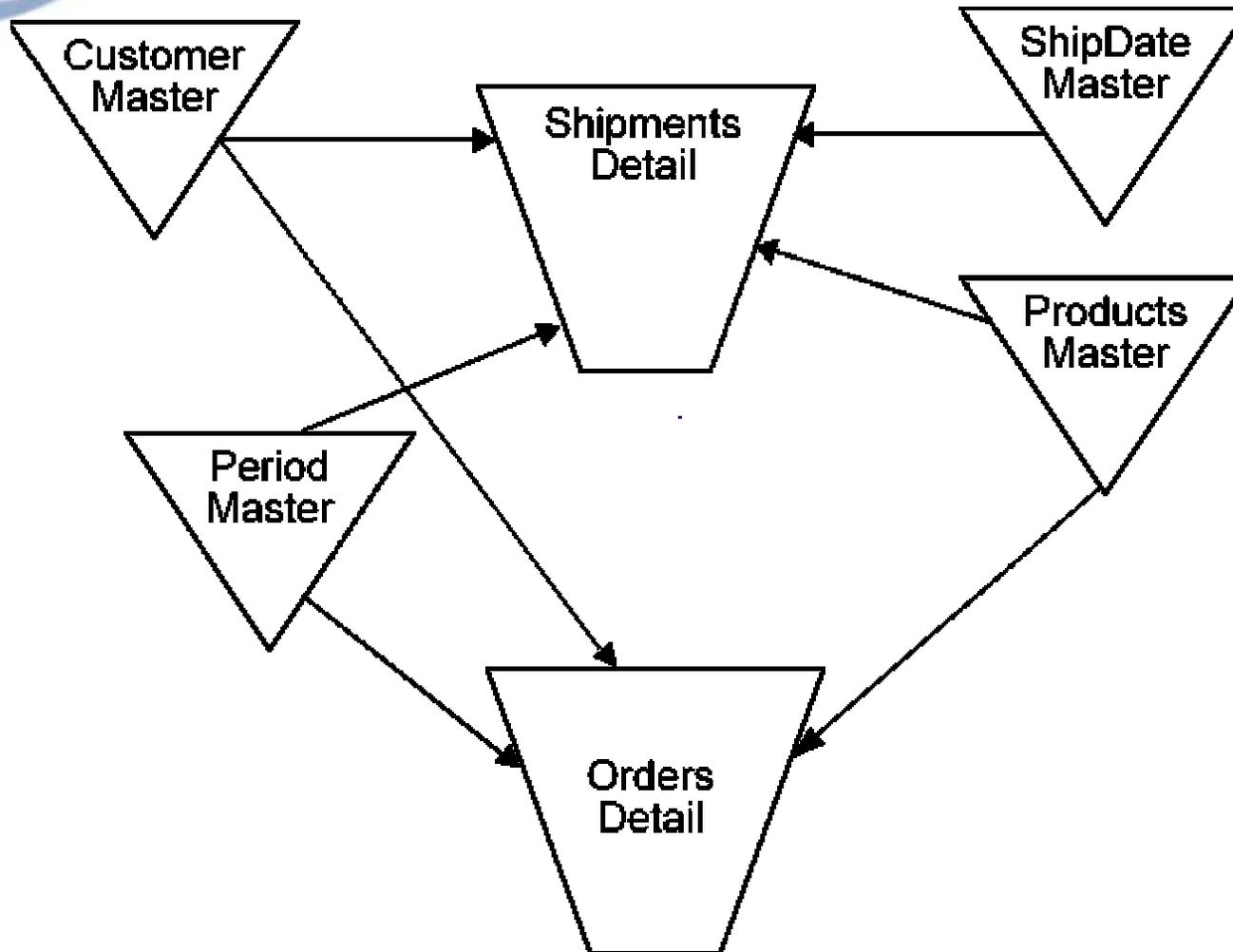
ORDERS	
<b>PK</b>	<b><u>INVOICE</u></b>
FK1,11 FK2,12 FK3,13	PERKEY PRODKEY CUSTKEY DOLLARS WEIGHT



Star...good

Snowflake  
...BAD!!!!

# “When I wish upon a Star”





# Data Warehouse System

- Data
- Connection(s) Layer
- ETL
- Query Tools
- Analysis Tools
- Presentation Interface
- **Quality Assurance procedures**
- \*Politics\*



# Design Process

- Choose a business process to model
- Choose the grain of the business process
- Choose the dimensions
- Choose the measured facts



## Consolidation of Disparate Data Sources

- Excel Spreadsheets
- Access database
- Other Image data sources
- A plethora of other RDBMSs

Most of your work will be in the ETL,  
data staging area. This will make or  
break your project!



## Get the data

- Examine existing reports, where does the data come from?
- How is the data extracted?
- Which data types are not supported?
- Period Table



F26 = 4

A	B	C	D	E	F	G	H	I	J	K	L
DATE_OPEN	DAY_NUMBER	DAY_OF_WEEK	DAY_NMBR_OVRALL	WK_NMBR_IN_YEAR	WK_NMBR_OVRALL	MNTH_NMBR	MNTH_NAME	MNTH_NMBR_OVRALL	QTR	YEARS	WEEKD'
19800101	1	Tue	1	1	1	1	Jan	1	Q1	1980	Y
19800102	2	Wed	2	1	1	1	Jan	1	Q1	1980	Y
19800103	3	Thu	3	1	1	1	Jan	1	Q1	1980	Y
19800104	4	Fri	4	1	1	1	Jan	1	Q1	1980	Y
19800105	5	Sat	5	1	1	1	Jan	1	Q1	1980	N
19800106	6	Sun	6	2	2	1	Jan	1	Q1	1980	N
19800107	7	Mon	7	2	2	1	Jan	1	Q1	1980	Y
19800108	8	Tue	8	2	2	1	Jan	1	Q1	1980	Y
19800109	9	Wed	9	2	2	1	Jan	1	Q1	1980	Y
19800110	10	Thu	10	2	2	1	Jan	1	Q1	1980	Y
19800111	11	Fri	11	2	2	1	Jan	1	Q1	1980	Y
19800112	12	Sat	12	2	2	1	Jan	1	Q1	1980	N
19800113	13	Sun	13	3	3	1	Jan	1	Q1	1980	N
19800114	14	Mon	14	3	3	1	Jan	1	Q1	1980	Y
19800115	15	Tue	15	3	3	1	Jan	1	Q1	1980	Y
19800116	16	Wed	16	3	3	1	Jan	1	Q1	1980	Y
19800117	17	Thu	17	3	3	1	Jan	1	Q1	1980	Y
19800118	18	Fri	18	3	3	1	Jan	1	Q1	1980	Y
19800119	19	Sat	19	3	3	1	Jan	1	Q1	1980	N
19800120	20	Sun	20	4	4	1	Jan	1	Q1	1980	N
19800121	21	Mon	21	4	4	1	Jan	1	Q1	1980	Y
19800122	22	Tue	22	4	4	1	Jan	1	Q1	1980	Y
19800123	23	Wed	23	4	4	1	Jan	1	Q1	1980	Y
19800124	24	Thu	24	4	4	1	Jan	1	Q1	1980	Y
19800125	25	Fri	25	4	4	1	Jan	1	Q1	1980	Y
19800126	26	Sat	26	4	4	1	Jan	1	Q1	1980	N
19800127	27	Sun	27	5	5	1	Jan	1	Q1	1980	N
19800128	28	Mon	28	5	5	1	Jan	1	Q1	1980	Y
19800129	29	Tue	29	5	5	1	Jan	1	Q1	1980	Y
19800130	30	Wed	30	5	5	1	Jan	1	Q1	1980	Y
19800131	31	Thu	31	5	5	1	Jan	1	Q1	1980	Y
19800201	1	Fri	32	5	5	2	Feb	2	Q1	1980	Y
19800202	2	Sat	33	5	5	2	Feb	2	Q1	1980	N
19800203	3	Sun	34	6	6	2	Feb	2	Q1	1980	N
19800204	4	Mon	35	6	6	2	Feb	2	Q1	1980	Y
19800205	5	Tue	36	6	6	2	Feb	2	Q1	1980	Y
19800206	6	Wed	37	6	6	2	Feb	2	Q1	1980	Y
19800207	7	Thu	38	6	6	2	Feb	2	Q1	1980	Y
19800208	8	Fri	39	6	6	2	Feb	2	Q1	1980	Y



# Start Small

- Project Management
- Select a few metrics/indicators
- Prototype on MS Access
  - Test, experiment, hack
  - SIMPLICITY, SIMPLICITY, SIMPLICITY
  - Sign-off on Prototype



# Awareness

## Project Management

- Keep communication open at all times
- Make sure to ‘pad’ task timelines
- Benchmarks, benchmarks and finally, sign-off of benchmarks



# Core Pieces

- Select Reporting Tool
  - Must be simple yet robust for Clients
  - Performance, server/client work load
  - Security, server/client layers
- Select ETL method
  - Use what you know best
  - Ease of maintenance



## Clean the data

- How many spelling variations can 'San Francisco' have?
- Transformations will take care of Dimension (master) tables



## Ship the data

- Load the presentation server
- Verify the data
  - The data warehouse is only as good as it's data
  - Any doubt will kill the project



# View the data

Esperant Query System - [ESPERANT Retail Demo] - [Customer.eqp]

Visiolyze.com - Customer Service Analyzer - Performance Metrics for Apr 99

Microsoft Excel - Book1

Customer Service Analyzer

Y	A	B	C	D	E	F	G	H	I	J	K	L
	DATE_OPEN	DAY_NUMBER	DAY_OF_WEEK	DAY_NUMB_OVALL	WK_NUMB_IN_YEAR	WK_NUMB_OVALL	MONTH_NUMB	MONTH_NAME	MONTH_NUMB_OVALL	QTR	YEARS	WEEKID
2	19000101	1	Tue	1	1	1	1	Jan	1	1 Q1	1900 Y	
3	19000102	2	Wed	2	1	1	1	Jan	1	1 Q1	1900 Y	
4	19000103	3	Thu	3	1	1	1	Jan	1	1 Q1	1900 Y	
5	19000104	4	Fri	4	1	1	1	Jan	1	1 Q1	1900 Y	
6	19000105	5	Sat	5	1	1	1	Jan	1	1 Q1	1900 Y	
7	19000106	6	Sun	6	2	2	1	Jan	1	1 Q1	1900 Y	
8	19000107	7	Mon	7	2	2	1	Jan	1	1 Q1	1900 Y	
9	19000108	8	Tue	8	2	2	1	Jan	1	1 Q1	1900 Y	
10	19000109	9	Wed	9	2	2	1	Jan	1	1 Q1	1900 Y	
11	19000110	10	Thu	10	2	2	1	Jan	1	1 Q1	1900 Y	
12	19000111	11	Fri	11	2	2	1	Jan	1	1 Q1	1900 Y	
13	19000112	12	Sat	12	2	2	1	Jan	1	1 Q1	1900 Y	
14	19000113	13	Sun	13	3	3	1	Jan	1	1 Q1	1900 Y	
15	19000114	14	Mon	14	3	3	1	Jan	1	1 Q1	1900 Y	
16	19000115	15	Tue	15	3	3	1	Jan	1	1 Q1	1900 Y	
17	19000116	16	Wed	16	3	3	1	Jan	1	1 Q1	1900 Y	
18	19000117	17	Thu	17	3	3	1	Jan	1	1 Q1	1900 Y	
19	19000118	18	Fri	18	3	3	1	Jan	1	1 Q1	1900 Y	
20	19000119	19	Sat	19	3	3	1	Jan	1	1 Q1	1900 Y	
21	19000120	20	Sun	20	4	4	1	Jan	1	1 Q1	1900 Y	
22	19000121	21	Mon	21	4	4	1	Jan	1	1 Q1	1900 Y	
23	19000122	22	Tue	22	4	4	1	Jan	1	1 Q1	1900 Y	
24	19000123	23	Wed	23	4	4	1	Jan	1	1 Q1	1900 Y	
25	19000124	24	Thu	24	4	4	1	Jan	1	1 Q1	1900 Y	
26	19000125	25	Fri	25	4	4	1	Jan	1	1 Q1	1900 Y	
27	19000126	26	Sat	26	4	4	1	Jan	1	1 Q1	1900 Y	
28	19000127	27	Sun	27	5	5	1	Jan	1	1 Q1	1900 Y	
29	19000128	28	Mon	28	5	5	1	Jan	1	1 Q1	1900 Y	
30	19000129	29	Tue	29	5	5	1	Jan	1	1 Q1	1900 Y	
31	19000130	30	Wed	30	5	5	1	Jan	1	1 Q1	1900 Y	
32	19000131	31	Thu	31	5	5	1	Jan	1	1 Q1	1900 Y	
33	19000201	1	Fri	32	5	5	2	Feb	2	2 Q1	1900 Y	
34	19000202	2	Sat	33	5	5	2	Feb	2	2 Q1	1900 Y	
35	19000203	3	Sun	34	6	6	2	Feb	2	2 Q1	1900 Y	
36	19000204	4	Mon	35	6	6	2	Feb	2	2 Q1	1900 Y	
37	19000205	5	Tue	36	6	6	2	Feb	2	2 Q1	1900 Y	
38	19000206	6	Wed	37	6	6	2	Feb	2	2 Q1	1900 Y	
39	19000207	7	Thu	38	6	6	2	Feb	2	2 Q1	1900 Y	
40	19000208	8	Fri	39	6	6	2	Feb	2	2 Q1	1900 Y	

Customer Service Analyzer

Bar chart showing performance metrics for Mar 99 and Apr 99. Values: 172 (Mar 99), 167 (Apr 99).

Table showing performance metrics for Apr 99:

Yr	Value
1999	166.99
1998	1.82
1997	1.22
1996	6.57
1995	\$47.42
1994	6.24
1993	6.73
1992	1.84
1991	6.95

Visiolyze.com

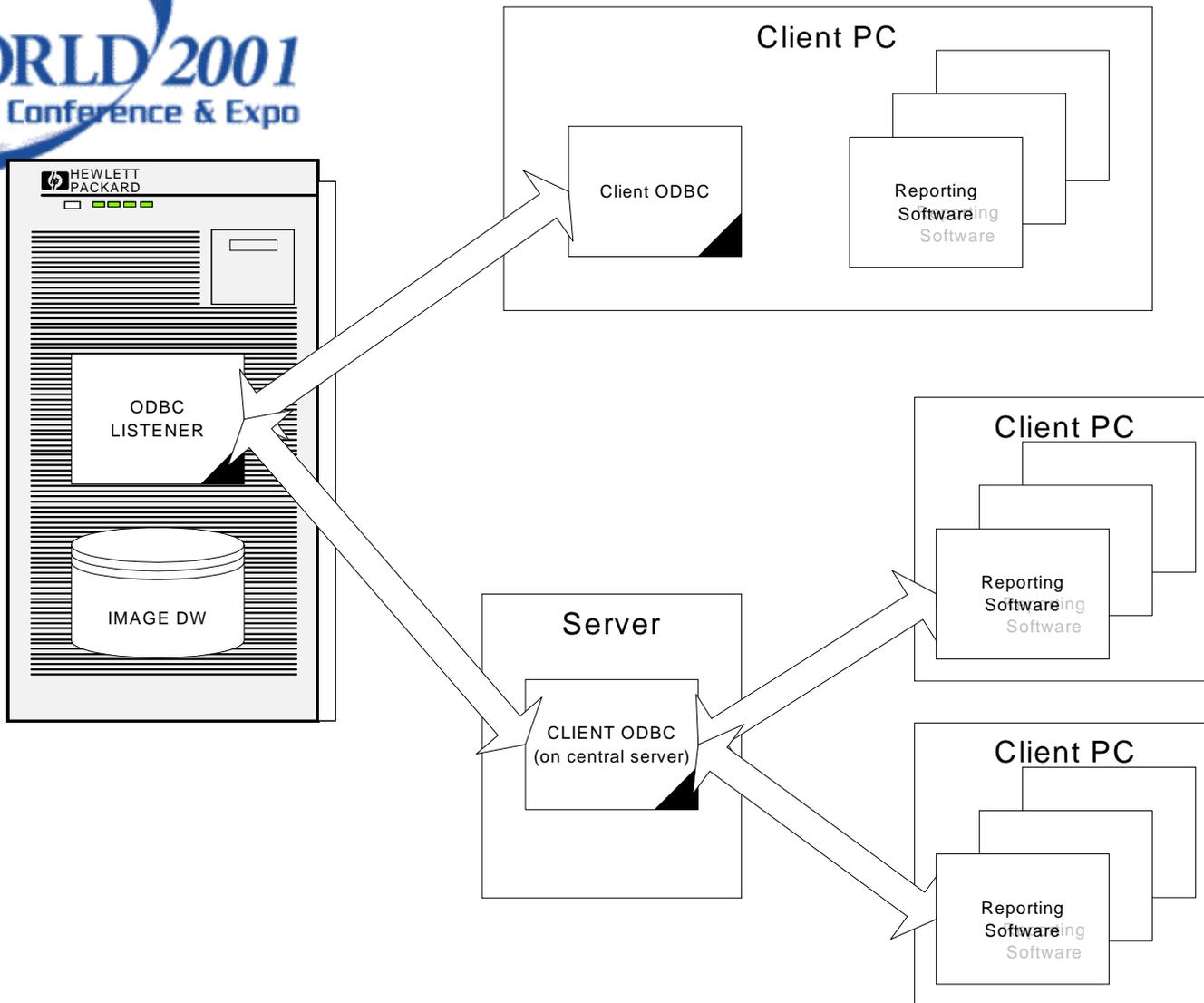


# Open Database Connectivity ODBC

- Plenty of documentation exists (sarcasm)
- May already exist in your system
- User groups and forums

# HP WORLD 2001

Conference & Expo





# Suggested Readings

## The Data Warehouse Toolkit

Ralph Kimball Foreword by W. H. Inmon



Practical  
Techniques for  
Building  
Dimensional  
Data  
Warehouses

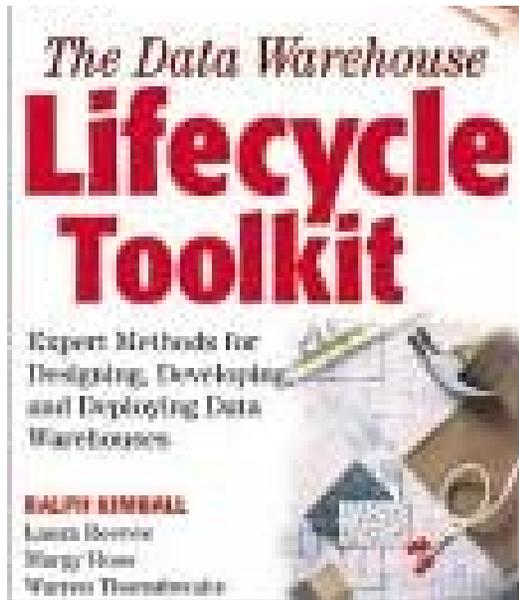


Includes  
CD-ROM

The Data Warehouse Toolkit  
Practical Techniques for Building  
Dimensional Data Warehouses  
by Ralph Kimball



# Suggested Readings



The Data Warehouse Lifecycle Toolkit : Expert Methods for Designing, Developing, and Deploying Data Warehouses by Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite