

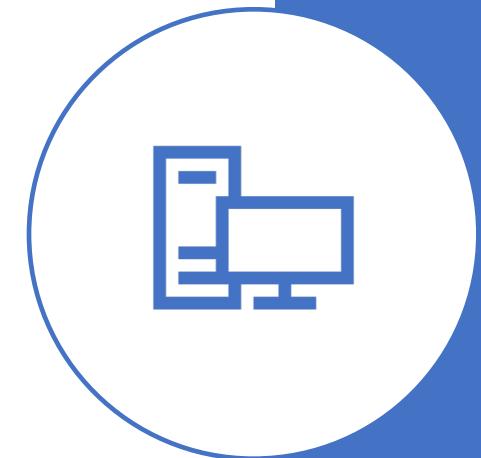
# CSCI 5408



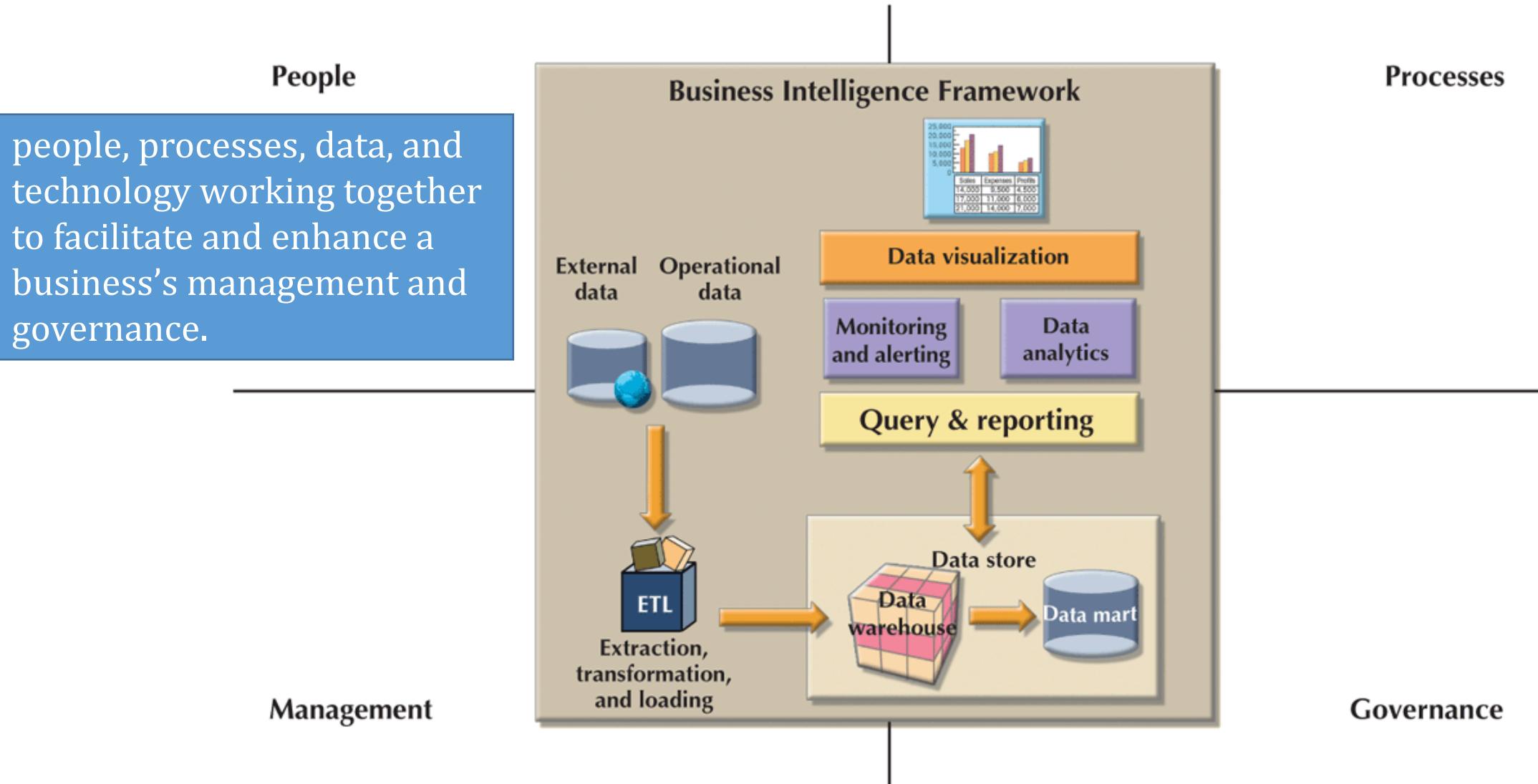
**Dr. Saurabh Dey**  
[saurabh.dey@dal.ca](mailto:saurabh.dey@dal.ca)

# Outline

- What is Data Warehouse?
- How does dimensional modelling help a business?



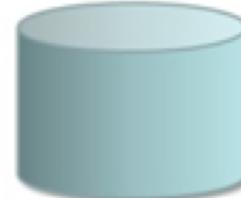
# Business Intelligence Architecture



# DW vs Data Marts



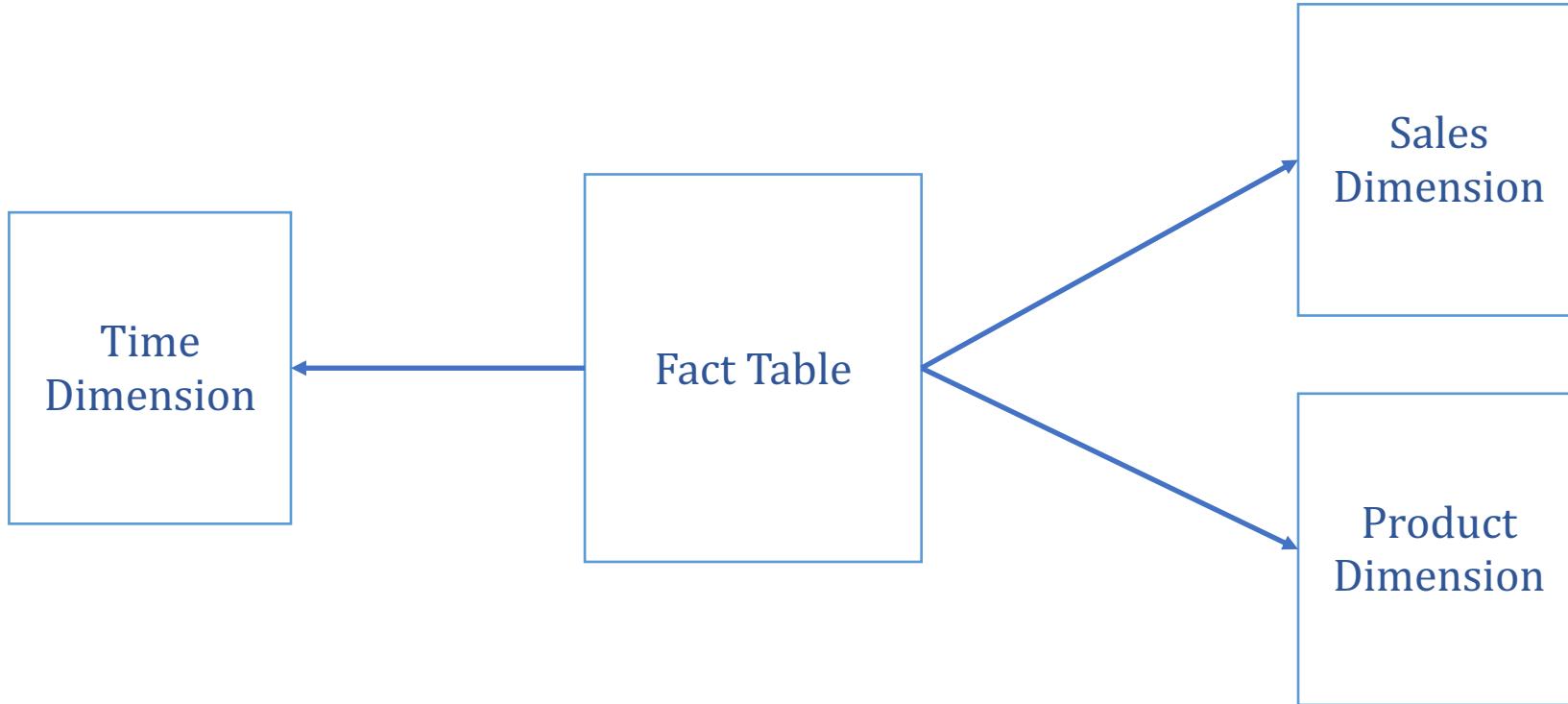
"And I'm a  
data mart."



"Hello, I'm a  
data warehouse."

Data Mart	Data Warehouse
<ul style="list-style-type: none"><li>Often holds only one subject area- for example, Finance, or Sales</li></ul>	<ul style="list-style-type: none"><li>Holds multiple subject areas</li></ul>
<ul style="list-style-type: none"><li>May hold more summarized data (although many hold full detail)</li></ul>	<ul style="list-style-type: none"><li>Holds very detailed information</li></ul>
<ul style="list-style-type: none"><li>Concentrates on integrating information from a given subject area or set of source systems</li></ul>	<ul style="list-style-type: none"><li>Works to integrate all data sources</li></ul>
<ul style="list-style-type: none"><li>Is built focused on a dimensional model using a star schema.</li></ul>	<ul style="list-style-type: none"><li>Does not necessarily use a dimensional model but feeds dimensional models</li></ul>

Source: <https://www.quora.com/What-does-a-Data-Mart-contain>



HEY BOB, DO YOU HAVE MY REPORT?

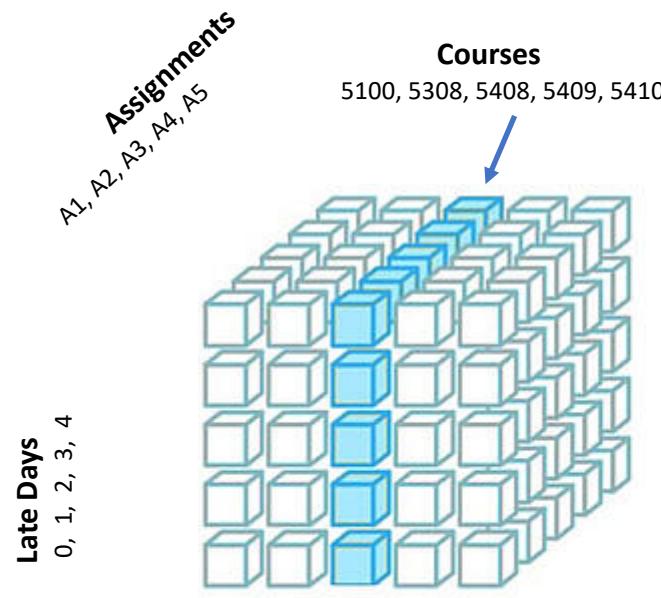


YEAH, THREE VERSIONS OF IT. YOU CAN PICK  
WHICHEVER YOU PREFER.



Dataedo /cartoon

Piotr@Dataedo

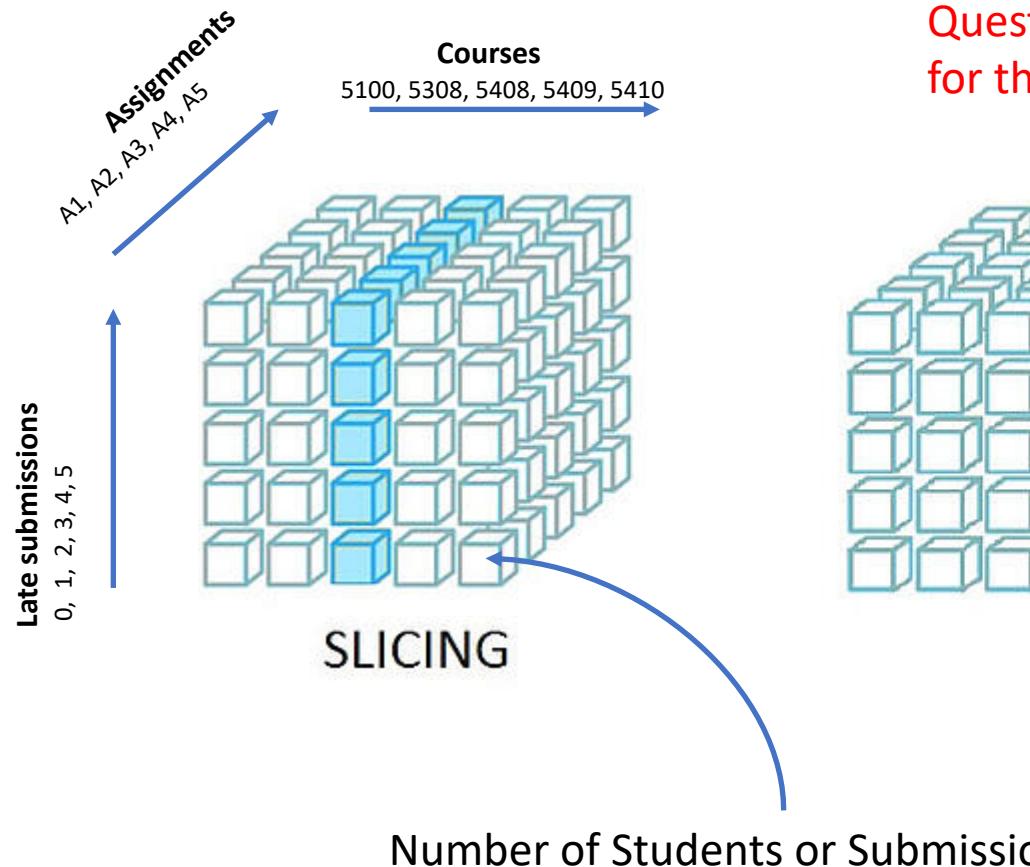


Course	Assignment	Late Days	Submission
5408	A1	0	2
5408	A1	1	4
5408	A1	2	5
5408	A1	3	7
5408	A1	4	2
5408	A2	0	0
5408	A2	1	2

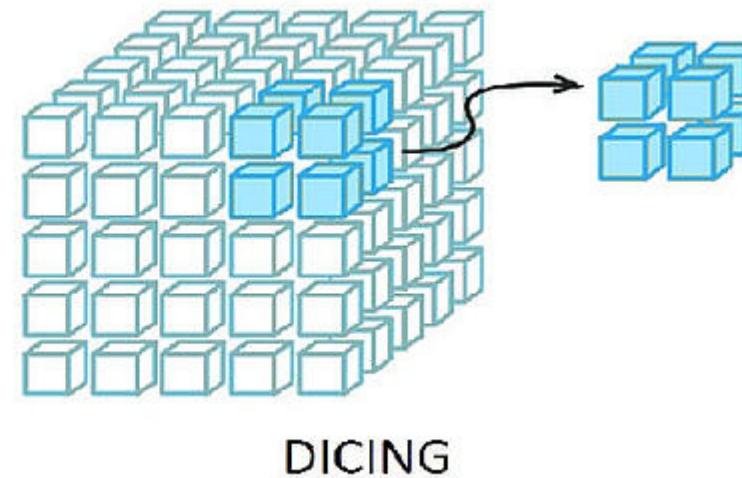
5408	A5	4	0
5408	A5	4	0

Question: How many students have submitted assignments?

Question: How many students have submitted CSCI 5408 assignments?



Question: How many students have used SDAs or submitted late for the initial assignments of cloud computing courses?



# Multidimensional Data Analysis Techniques

Data are processed and viewed as part of a multidimensional structure

Particularly attractive to business decision makers who tend to view business data as being related to other business data

Table name: DW\_INVOICE

INV_NUM	INV_DATE	CUS_NAME	INV_TOTAL
2034	15-May-18	Dartonik	1400.00
2035	15-May-18	Summer Lake	1200.00
2036	16-May-18	Dartonik	1350.00
2037	16-May-18	Summer lake	3100.00
2038	16-May-18	Trydon	400.00

Database name: Ch13\_Text

Operational Data

Table name: DW\_LINE

INV_NUM	LINE_NUM	PROD_DESCRIPTION	LINE_PRICE	LINE_QUANTITY	LINE_AMOUNT
2034	1	Optical Mouse	45.00	20	900.00
2034	2	Wireless RF remote and laser pointer	50.00	10	500.00
2035	1	Everlast Hard Drive, 60 GB	200.00	6	1200.00
2036	1	Optical Mouse	45.00	30	1350.00
2037	1	Optical Mouse	45.00	10	450.00
2037	2	Roadster 56KB Ext. Modem	120.00	5	600.00
2037	3	Everlast Hard Drive, 60 GB	205.00	10	2050.00
2038	1	NoTech Speaker Set	50.00	8	400.00

Multidimensional View of Sales  
(using MS Excel PivotTable)

CUS_NAME	INV_DATE			Grand Total
	15-May-18	16-May-18	Grand Total	
Dartonik	\$ 1,400.00	\$ 1,350.00	\$ 2,750.00	
Summer Lake	\$ 1,200.00	\$ 3,100.00	\$ 4,300.00	
Trydon		\$ 400.00	\$ 400.00	
Grand Total	\$ 2,600.00	\$ 4,850.00	\$ 7,450.00	

Customer Dimension

Time Dimension

Sales are located in the intersection of a customer row and date (time) column

Aggregations (grand total sales) are provided for both dimensions (time and customer)

# Example: SaleCoDW Data (Using PowerBI)

The screenshot shows the Power BI Desktop interface with the 'Home' tab selected. A data import dialog box is open, displaying the contents of 'SaleCoCustomer.csv'. The dialog includes settings for 'File Origin' (1252: Western European (Windows)), 'Delimiter' (Comma), and 'Data Type Detection' (Based on first 200 rows). The data preview shows 19 rows with columns: CUS\_CODE, CUS\_LNAME, CUS\_FNAME, CUS\_INITIAL, CUS\_STATE, and REG\_ID. The Power BI visualizations pane on the right shows various chart and report options.

SaleCoDSaleCo\_Customer.csv

CUS_CODE	CUS_LNAME	CUS_FNAME	CUS_INITIAL	CUS_STATE	REG_ID
10010	Ramas	Alfred	A	TN	3
10011	Dunne	Leona	K	GA	3
10012	Smith	Kathy	W	NY	1
10013	Olowksi	Paul	F	NJ	1
10014	Orlando	Myron	NULL	CO	2
10015	O'Brian	Amy	B	TN	3
10016	Brown	James	G	GA	3
10017	Williams	George	NULL	CA	4
10018	Farris	Anne	G	CA	4
10019	Smith	Olette	K	CO	2

Load Edit Cancel

Visualizations > Fields >

Search

Values

Add data fields here

Filters

Page level filters

Add data fields here

Report level filters

Add data fields here

Drillthrough

Cross-report

Off

Keep all filters

On

Add drillthrough fields here

Page 1 +

PAGE 1 OF 1

# Example: SaleCoDW Data (Using PowerBI)

The screenshot shows the Power BI Desktop interface with the 'File' tab selected. A modal dialog box titled 'SaleCoDWSaleCo\_Product.csv' is open, showing the 'Get Data' step for loading data from a CSV file. The dialog includes fields for 'File Origin' (1252: Western European (Windows)), 'Delimiter' (Comma), and 'Data Type Detection' (Based on first 200 rows). The main area displays a preview of the data, which consists of four columns: P\_CODE, P\_DESCRPT, P\_CATEGORY, and V\_CODE. The data includes various product entries like 'Power painter, 15 psi., 3-nozzle' and 'Hrd. cloth, 1/4-in., 2x50'. On the right side of the screen, the 'Visualizations' and 'Fields' panes are visible, showing available data sources and fields for further analysis.

P_CODE	P_DESCRPT	P_CATEGORY	V_CODE
11QER/31	Power painter, 15 psi., 3-nozzle	CAT1	25595
13-Q2/P2	7.25-in. pwr. saw blade	CAT1	21344
14-Q1/L3	9.00-in. pwr. saw blade	CAT1	21344
1546-QQ2	Hrd. cloth, 1/4-in., 2x50	CAT2	23119
1558-QW1	Hrd. cloth, 1/2-in., 3x50	CAT2	23119
2232/QTY	B&D jigsaw, 12-in. blade	CAT2	24288
2232/QWE	B&D jigsaw, 8-in. blade	CAT3	24288
2238/QPD	B&D cordless drill, 1/2-in.	CAT3	25595
23109-HB	Claw hammer	CAT4	21225
23114-AA	Sledge hammer, 12 lb.	CAT4	21225
54778-2T	Rat-tail file, 1/8-in. fine	CAT1	21344
89-WRE-Q	Hicut chain saw, 16 in.	CAT2	24288
PVC23DRT	PVC pipe, 3.5-in., 8-ft	CAT3	21225
SM-18277	1.25-in. metal screw, 25	CAT4	21225
SW-23116	2.5-in. wd. screw, 50	CAT2	21231
WR3/TT3	Steel matting, 4x8'x1/6", .5" mesh	CAT3	25595

# Example: SaleCoDW Data (Using PowerBI)

The screenshot shows the Microsoft PowerBI desktop application interface. The main area displays a data preview window titled "SaleCoDWSaleCo\_Region.csv". The preview shows a table with four rows and two columns:

REG_ID	REG_NAME
1	NE
2	NW
3	SE
4	SW

The "File Origin" dropdown shows "1252: Western European (Windows)". The "Delimiter" is set to "Comma" and "Data Type Detection" is set to "Based on first 200 rows". At the bottom of the preview window are "Load", "Edit", and "Cancel" buttons.

To the right of the preview window is the "Visualizations" pane, which contains various chart and report icons. Below it is the "Fields" pane, which lists fields from three tables:

- SaleCoDWSaleCo\_Customer:
  - CUS\_CODE
  - CUS\_FNAME
  - CUS\_INITIAL
  - CUS\_LNAME
  - CUS\_STATE
  - REG\_ID
- SaleCoDWSaleCo\_DaySalesFact
- SaleCoDWSaleCo\_Product

The "Filters", "Drillthrough", and "Cross-report" sections are currently empty.

At the bottom left, there are navigation buttons for "Page 1" and a plus sign. The bottom status bar shows "PAGE 1 OF 1" and a taskbar with various icons.

# Example: SaleCoDW Data (Using PowerBI)

The screenshot shows the Power BI Desktop interface with the following details:

- File Bar:** File, Home, View, Modeling, Help.
- Clipboard:** Cut, Copy, Format Painter, Paste.
- External Data:** Get Data (Recent Sources, Enter Data, Refresh), New Page, New Visual, Ask A Question, Buttons, Text box, Image, Shapes, From Marketplace, From File, Switch Theme, Manage Relationships, New Measure, New Column, New Quick Measure, Calculations, Share.
- Visualizations:** A grid of visualization icons.
- Fields:** A search bar and a list of fields categorized by table:
  - SaleCoDWSaleCo\_Customer:** CUS\_CODE, CUS\_FNAME, CUS\_INITIAL, CUS\_LNAME, CUS\_STATE, REG\_ID.
  - SaleCoDWSaleCo\_DaySalesFact**
  - SaleCoDWSaleCo\_Product**
  - SaleCoDWSaleCo\_Region**
- Current View:** A preview window titled "SaleCoDWSaleCo\_Time.csv" showing a table with columns TM\_ID, TM\_YEAR, TM\_MONTH, TM\_DAY, and TM\_QTR. The data in the preview is:

TM_ID	TM_YEAR	TM_MONTH	TM_DAY	TM_QTR
201	2015	9	29	3
202	2015	9	30	3
203	2015	9	31	3
206	2015	10	3	4
207	2015	10	4	4
- Bottom Navigation:** Page 1, +, PAGE 1 OF 1.
- System Taskbar:** Windows, Search, File Explorer, Start, Task View, Taskbar settings, 1:24 PM, 7/16/2019.

# Example: SaleCoDW Data (Using PowerBI)

SaleCoDWSaleCo\_Vendor.csv

File Origin: 1252: Western European (Windows) Delimiter: Comma Data Type Detection: Based on first 200 rows

V_CODE	V_NAME	V_AREACODE	V_STATE
21225	Bryson, Inc.	615	TN
21226	SuperLoo, Inc.	904	FL
21231	D&E Supply	615	TN
21344	Gomez Bros.	615	KY
22567	Dome Supply	901	GA
23119	Randssets Ltd.	901	GA
24004	Brackman Bros.	615	TN
24288	ORDVA, Inc.	615	TN
25443	B&K, Inc.	904	FL
25501	Damal Supplies	615	TN
25595	Rubicon Systems	904	FL

Visualizations >

Fields >

Search: SaleCoDWSaleCo\_Customer

- CUS\_CODE
- CUS\_FNAME
- CUS\_INITIAL
- CUS\_LNAME
- CUS\_STATE
- REG\_ID

Visualizations >

Fields >

Search: SaleCoDWSaleCo\_DaySalesFact

Values

Add data fields here

Filters

Page level filters

Add data fields here

Report level filters

Add data fields here

Drillthrough

Cross-report

Off

Keep all filters

On

Add drillthrough fields here

Load Edit Cancel

Page 1 +

PAGE 1 OF 1

1:25 PM 7/16/2019

# Example: SaleCoDW Data (Using PowerBI)

The screenshot shows the Power BI Desktop interface with the following details:

- File**: Untitled - Power BI Desktop
- Home**: Selected tab
- View**, **Modeling**, **Help**: Other tabs
- Clipboard**: Standard clipboard icons (Paste, Cut, Copy, Format Painter)
- External data**: Options for getting data from various sources (Get Data, Recent Sources, Enter Data, Edit Queries, Refresh, New Page, New Visual, Ask A Question, Buttons, Text box, From Marketplace, From File, Switch Theme, Manage Relationships, New Measure, New Column, New Quick Measure, Publish, Share)
- SaleCoDSaleCo\_DaySalesFact.csv**: Preview window showing the contents of the CSV file.

  - File Origin**: 1252: Western European (Windows)
  - Delimiter**: Comma
  - Data Type Detection**: Based on first 200 rows

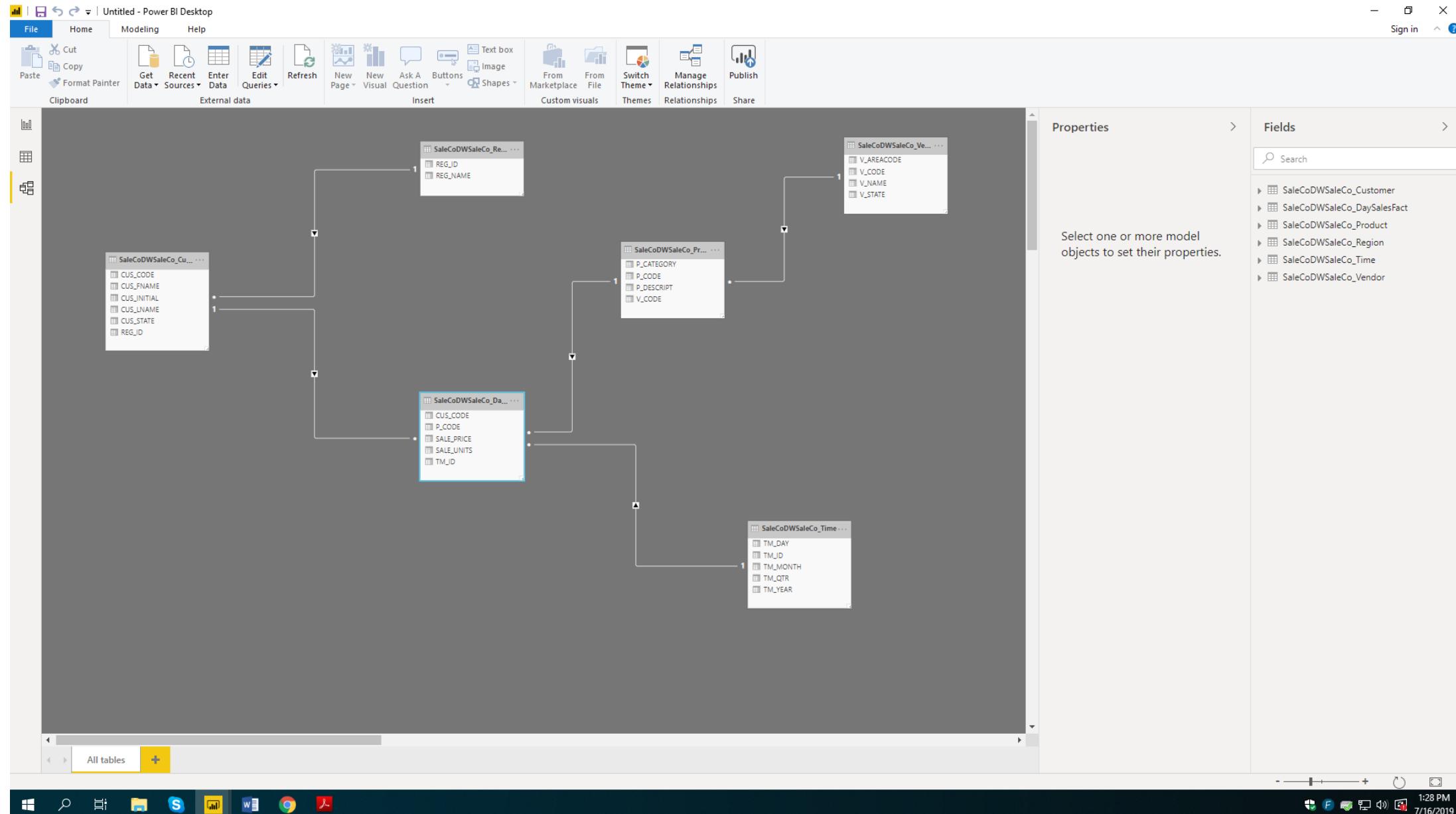
TM_ID	CUS_CODE	P_CODE	SALE_UNITS	SALE_PRICE
201	10014	13-Q2/P2	1	14.99
201	10014	23109-HB	1	9.95
201	10015	2238/QPD	1	38.95
201	10015	54778-2T	2	4.99
202	10016	13-Q2/P2	5	14.99
202	10016	1546-QQ2	1	39.95
202	10017	23109-HB	2	9.95
202	10017	54778-2T	3	4.99
202	10018	PVC23DRT	12	5.87
203	10012	SM-18277	3	6.99
203	10014	2232/QTY	1	109.92
203	10015	23109-HB	1	9.95
203	10015	89-WRE-Q	1	256.99
203	10016	13-Q2/P2	2	14.99
203	10016	54778-2T	1	4.99
203	10016	PVC23DRT	5	5.87
203	10017	13-Q2/P2	1	14.99
203	10017	23109-HB	1	9.95
203	10017	WR3/TT3	3	119.95
203	10018	2238/QPD	1	38.95

- Visualizations**: Preview of available visual types (Bar, Line, Clustered Bar, etc.)
- Fields**: List of fields from the loaded data source (SaleCoDW\_SaleCo\_Customer).

  - Search**: Search bar for fields.
  - SaleCoDW\_SaleCo\_Customer**: Field list for the customer dimension.
  - Values**: Placeholder for selected data fields.
  - Add data fields here**: Placeholder for additional data fields.
  - Filters**: Options for page and report level filters.
  - Page level filters**: Placeholder for page-level filters.
  - Report level filters**: Placeholder for report-level filters.
  - Drillthrough**: Options for cross-report and keep all filters.
  - Cross-report**: Off (radio button).
  - Keep all filters**: On (radio button).
  - Add drillthrough fields here**: Placeholder for drillthrough fields.

- Page**: Navigation buttons for pages (Page 1, +).
- PAGE 1 OF 1**: Page status.
- System tray**: Standard Windows system tray icons (File Explorer, Task View, Start, Search, Taskbar, etc.).
- Power BI status**: 1:19 PM, 7/16/2019.

# Example: SaleCoDW Data (Using PowerBI)



# Example: SaleCoDW Data (Using PowerBI)

The screenshot displays three distinct Power BI visualizations:

- Bar Chart:** A horizontal bar chart titled "SALE\_PRICE by V\_STATE and P\_CATEGORY". The Y-axis represents "SALE\_PRICE" from 0 to 800. The X-axis lists states: TN, FL, GA, and KY. The legend indicates four categories: CAT1 (teal), CAT2 (black), CAT3 (red), and CAT4 (yellow). The chart shows significant sales for FL (~350) and TN (~700).
- Map:** A map of the United States titled "SALE\_UNITS BY CUS\_STATE AND REG\_NAME". States are color-coded by region: NE (light green), NW (light blue), SE (light orange), SW (light red), and others (light grey). A legend on the left shows the region colors and state abbreviations.
- Table:** A small table titled "CUS\_STATE NE NW SE SW" showing unit counts for different regions. The data is as follows:

CUS_STATE	NE	NW	SE	SW
CA	28	26		
GA	18			
NJ	8			
CO	7			
NY	5			

The Power BI interface includes a ribbon menu, a search bar, and a large "Fields" pane on the right containing a tree view of all data tables and columns used in the report.

# Example: SaleCoDW Data (Using PowerBI)

The screenshot displays two separate Power BI Desktop environments side-by-side, illustrating the relationship between the visualizations and the underlying data models.

**Left Report:**

- Title:** SALE\_PRICE BY P\_CODE AND REG\_NAME
- Visual:** A grouped bar chart showing sales price by product code (P\_CODE) and region (REG\_NAME). The Y-axis represents the sale price, ranging from 0 to 250. The X-axis lists product codes: 89-WRE-Q, WR3/TT3, 2232/QTY, 13-Q2/P2, 1546-QQ2, 23109-HB, 2238/QPD, 54778-2T, PVC23DRT, and SM-18277. The bars are color-coded by region: NE (teal), NW (dark grey), SE (red), and SW (yellow).
- Data Table:** A table showing the same data as the chart, with columns for P\_CODE, NE, NW, SE, and SW.
- Visual Tools:** The ribbon shows tabs for File, Home, View, Modeling, Help, Format, Data / Drill, and Visual tools.

**Right Report:**

- Title:** Untitled - Power BI Desktop
- Visual:** A bar chart titled "SALE\_PRICE by P\_CODE, V\_CODE, P\_CATEGORY and REG\_NAME". The Y-axis ranges from 0 to 120. The X-axis categories are 2232/QTY and 24288-CAT2. The bars are color-coded by region: NE (dark grey) and NW (red).
- Fields:** The Fields pane on the right lists fields from various data models, including SaleCoDWSaleCo\_Customer, SaleCoDWSaleCo\_DaySalesF..., and SaleCoDWSaleCo\_Product.
- Visual Tools:** The ribbon shows tabs for File, Home, View, Modeling, Help, Format, Data / Drill, and Visual tools.

**Bottom Navigation:**

- Windows taskbar icons for Start, Search, File Explorer, Task View, Edge, Google Chrome, and File Explorer.
- Power BI ribbon tabs: File, Home, View, Modeling, Help, Format, Data / Drill, and Visual tools.
- Page navigation: Page 1 of 1.
- System tray icons: Volume, Network, Battery, and Date/Time.

# Questions to Consider

- Do we need a tool to build dimensional modelling?
- Is star schema better than snowflake schema?
- What is the use of PowerBI and Cognos? Are both same?

