Professional Association for SQL Server

Multidimensional Thinking

Stacia Misner
Data Inspirations

Sponsored by:

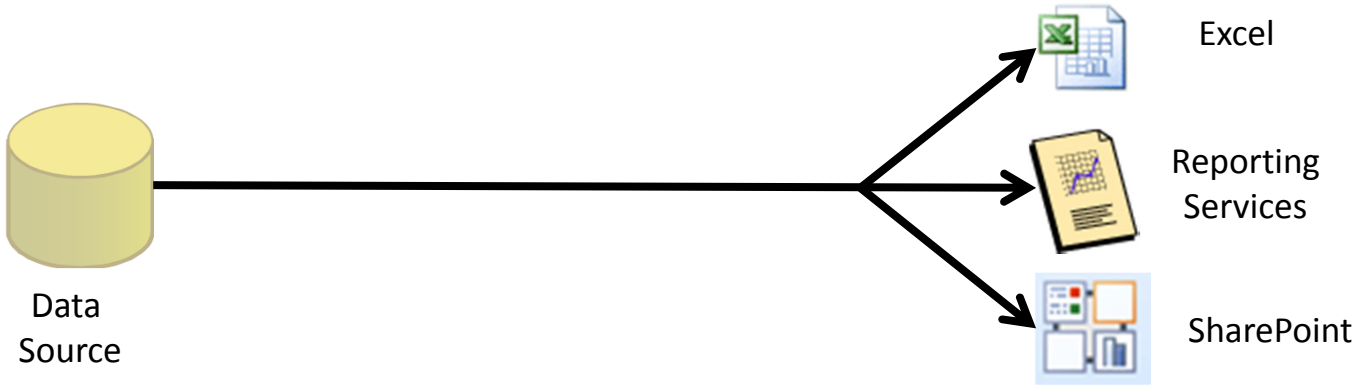
Overview

- Understanding the Problem
- Translating Requirements
- Designing the Multidimensional Model

Understanding the Problem

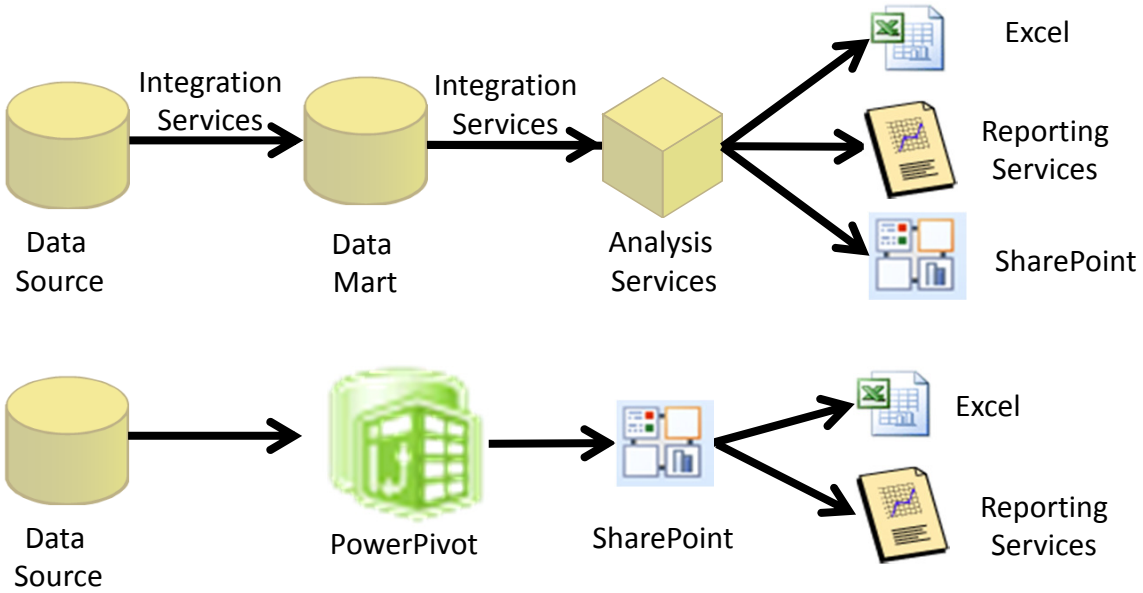
Reporting on Operational Data

- Resource contention
- Unreliable access to historical data
- Inconsistent application of business rules
- Data structure results in slower, more complex queries



A Data Mart Solution

- Separate data source reduces resource contention
- Easy access to historical data
- Consistent application of business rules
- Data structure results in faster, simpler queries



Translating Requirements

Requirements Gathering

- What information do reports include that's useful?
- What information is missing?

- What are typical questions you ask day-to-day?
- What processes are delaying access to information?

Typical Questions from Decision Makers

- Summarized data for specified period of time
  - What were total sales for last 3 years?
  - What was profit margin last year?
  - How many items were sold this month?
- Comparative data for multiple categories or time periods
  - Did sales increase this year as compared to last year?
  - What was profit margin by product category last year?
  - Is the average sale by territory increasing or decreasing?
- Consolidated data from various source systems
  - How do sales volumes for top 10 customers compare to call center volumes for the same customers?

Prioritization and Decomposition

❶

Did sales increase this year as compared to last year?

❷

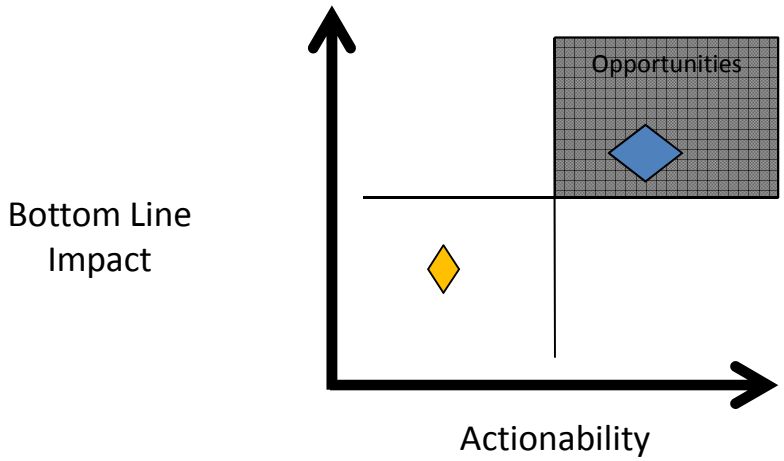
How many items were sold this month?

❸

How do sales volumes for this year's top 10 customers compare to call counts for the same customers?

The Matrix

	A	B	C	D	E	F	G	H
1					Dimensions			
2			Date	Customer	Product	Sales Territory	Call Type	Service Rep
3	Note	Measure						
4	1	sales amount	month	customer	product	region		
5	2	item count	month	customer	product	region		
6	4	margin	month	customer	category	region		
7	6	call count	day	customer		region	tier 1	CSR
8	5	call length	day	customer		region	tier 1	CSR

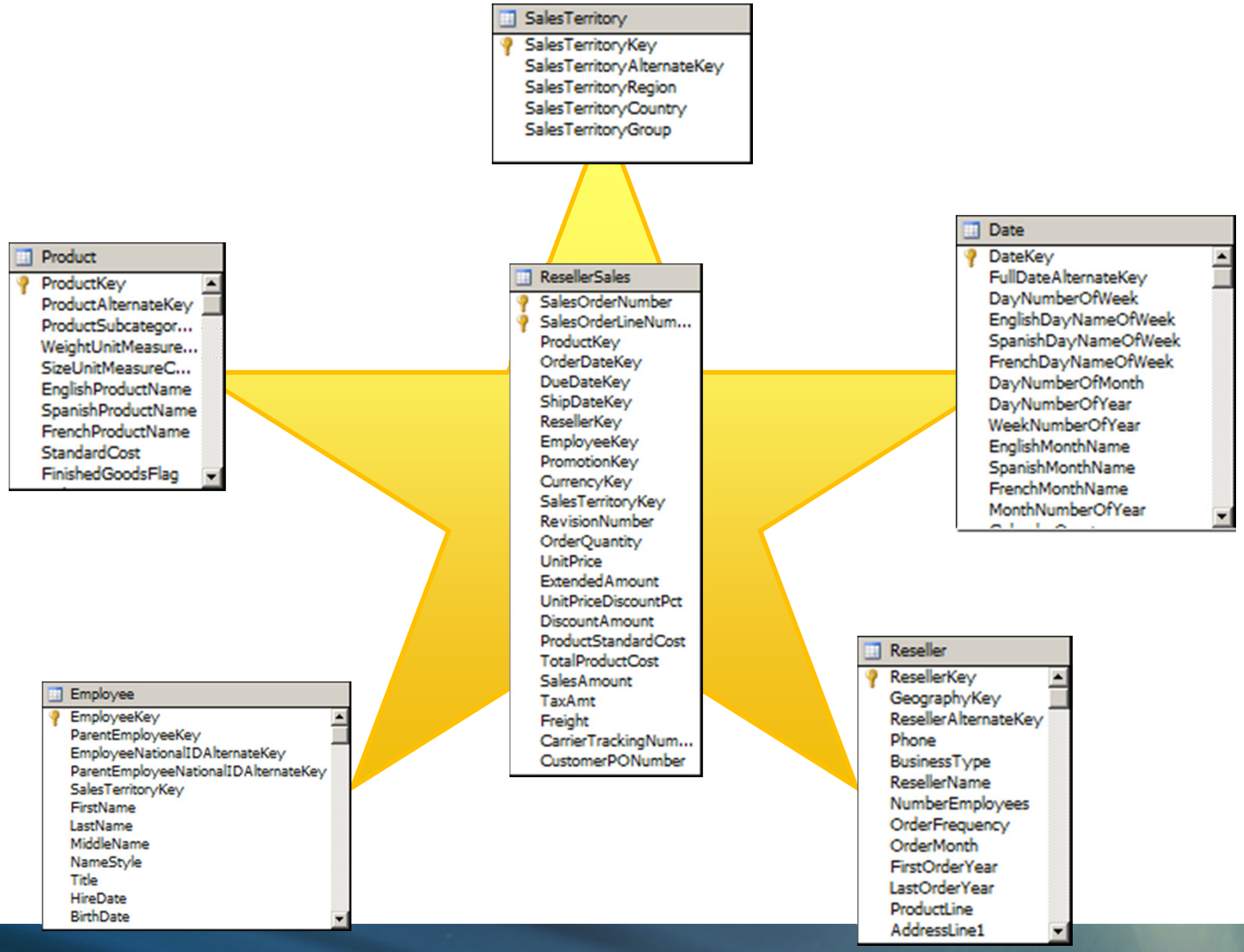


Building a Matrix

- Create spreadsheet to cross-reference source document to dimension and measures
- List measures on rows and dimensions on columns
- Identify the lowest level of detail for each dimension by measure
- Sort data to group sources by common dimensions

Designing the Multidimensional Model

Star Schema


SalesTerritory
SalesTerritoryKey
SalesTerritoryAlternateKey
SalesTerritoryRegion
SalesTerritoryCountry
SalesTerritoryGroup
Product
ProductKey
ProductAlternateKey
ProductSubcategor...
WeightUnitMeasure...
SizeUnitMeasureC...
EnglishProductName
SpanishProductName
FrenchProductName
StandardCost
FinishedGoodsFlag
ResellerSales
SalesOrderNumber
SalesOrderLineNum...
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
ResellerKey
EmployeeKey
PromotionKey
CurrencyKey
SalesTerritoryKey
RevisionNumber
OrderQuantity
UnitPrice
ExtendedAmount
UnitPriceDiscountPct
DiscountAmount
ProductStandardCost
TotalProductCost
SalesAmount
TaxAmt
Freight
CarrierTrackingNum...
CustomerPONumber
Date
DateKey
FullDateAlternateKey
DayNumberOfWeek
EnglishDayNameOfWeek
SpanishDayNameOfWeek
FrenchDayNameOfWeek
DayNumberOfMonth
DayNumberOfYear
WeekNumberOfYear
EnglishMonthName
SpanishMonthName
FrenchMonthName
MonthNumberOfYear
Employee
EmployeeKey
ParentEmployeeKey
EmployeeNationalIDAlternateKey
ParentEmployeeNationalIDAlternateKey
SalesTerritoryKey
FirstName
LastName
MiddleName
NameStyle
Title
HireDate
BirthDate
Reseller
ResellerKey
GeographyKey
ResellerAlternateKey
Phone
BusinessType
ResellerName
NumberEmployees
OrderFrequency
OrderMonth
FirstOrderYear
LastOrderYear
ProductLine
AddressLine1

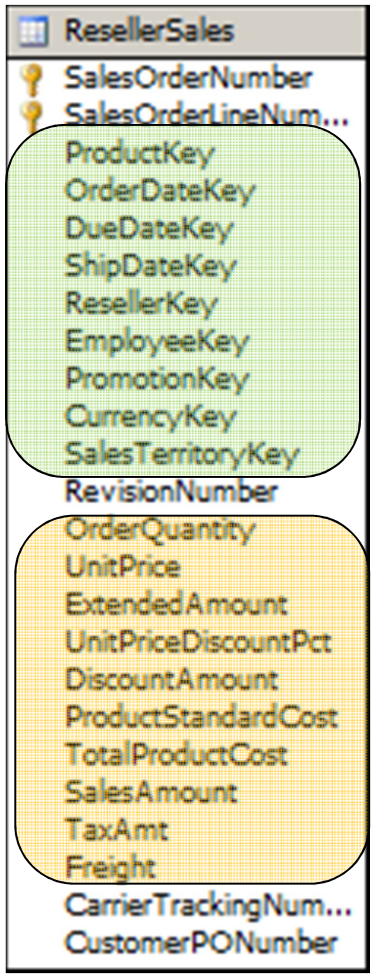
Fact Discovery

- Summarized data for specified period of time
  - What were total sales for last 3 years?
  - What was profit margin last year?
  - How many items were sold this month?
- Comparative data for multiple categories or time periods
  - Did sales increase this year as compared to last year?
  - What was profit margin by product category last year?
  - Is the average sale by territory increasing or decreasing?

	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice	iceDiscount	LineTotal	rowguid
187	514	LL Mountain Seat A...	SA-M198	1	0	NULL	500	375	98.77	133.34		2024.994000	B207C96D-D9E6-402B-84
188	515	ML Mountain Seat ...	SA-M237	1	0	NULL	500	375	108.99	147.14		6074.982000	7ABB600D-1E77-41BE-
189	516	HL Mountain Seat A...	SA-M687	1	0	NULL	500	375	145.87	196.9		2024.994000	475CF8C6-49F6-486E-B
		43659			4911-403C-98	1	50	1	2024.994	0.00		2039.994000	04C4DE91-5815-45D6-
4		43659	4		4911-403C-98	1	771	1	2039.994	0.00		2039.994000	5A74C7D2-E641-438E
5		43659	5		4911-403C-98	1	772	1	2039.994	0.00		2039.994000	

Fact Tables

- "Verbs" of the model
- Measurement of business processes
- High volume of records
- Two column types
  - Dimension keys
  - Measures
- Consistent grain

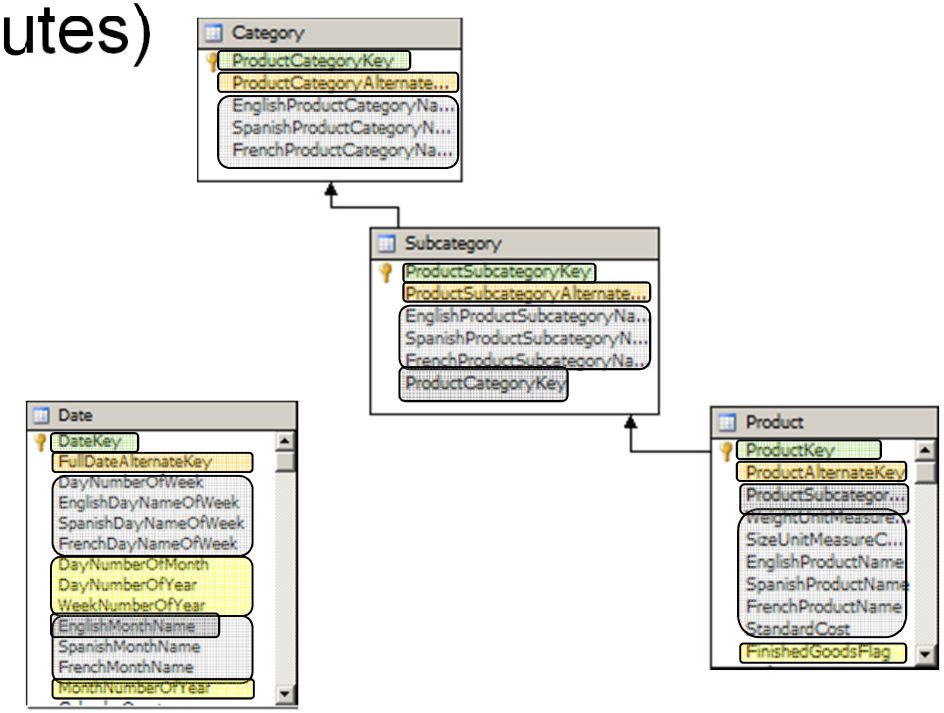


Fact Table Design

- Declare the grain
- Identify the dimensions applicable to each row
- Identify measures and atomicity
- Break down percentages and ratios into numerators and denominators

Dimension Tables

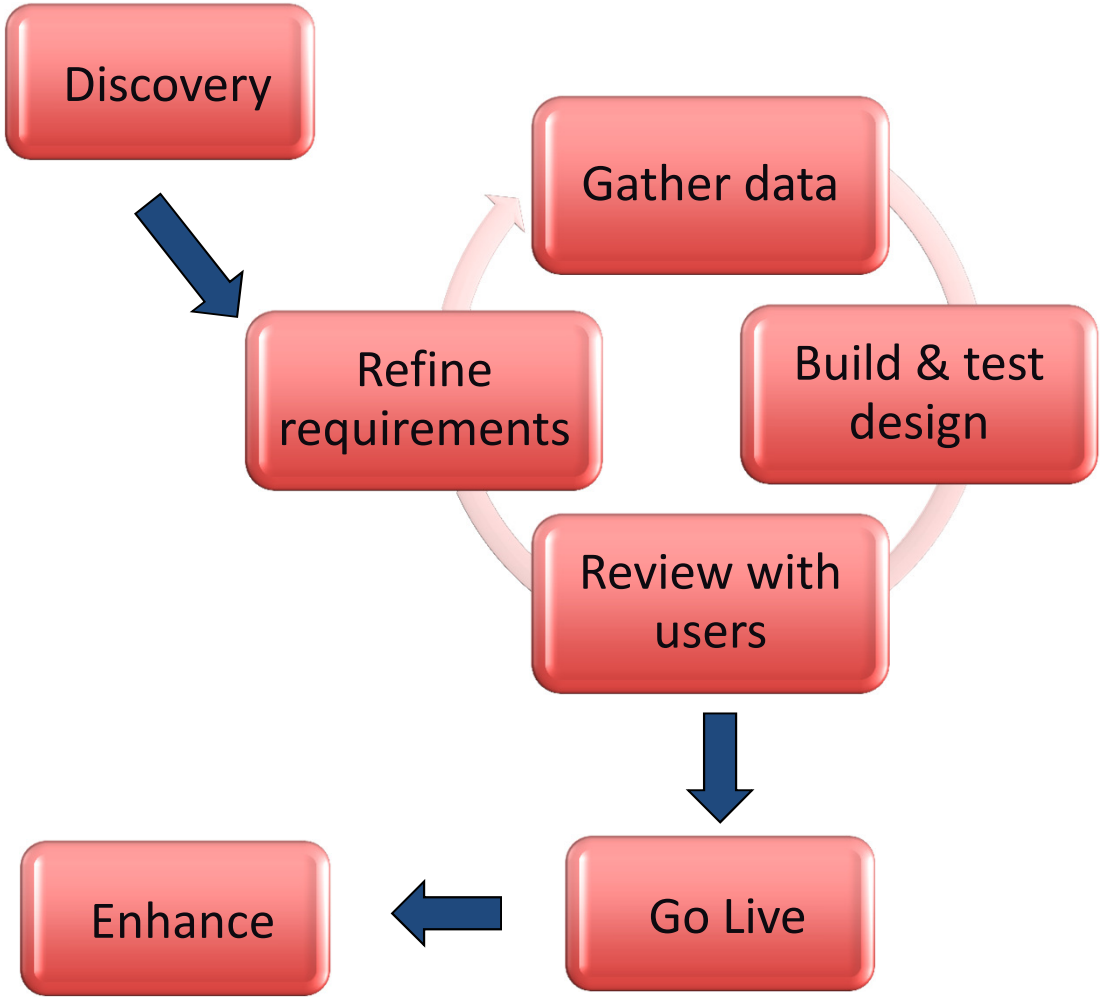
- "Nouns" of the model
- Description of people, place, things, and time
- Multiple columns (attributes)
  - Surrogate key
  - Business key
  - Descriptive data
  - Grouping, sorting, or filtering data
  - Hierarchical data



Dimension Table Design

- Define attributes
  - Descriptive
  - Grouping
  - Filtering
  - Sorting
  - Hierarchical navigation
  - Slowly changing dimension management
- Decide whether to use snowflake
- Define attributes for the date dimension

Model Confirmation and Deployment



Professional Association for SQL Server

Thank you to our sponsor

May 11-13, Orlando, FL

Email: smisner@datainspirations.com
Twitter: @StaciaMisner
Blog: blog.datainspirations.com

Save 25%: Register by April 12th

www.sqlpass.org/sqlrally

Oct 11-14, Seattle, WA

Register by March31st: save 40% and have the chance to win a cruise to Alaska!

"24HR11" code gets you $100 off

www.sqlpass.org/summit