

The Decoder Ring for Data Warehousing / Business Intelligence With SQL Server

Robert C. Cain, MVP, MCTS
Pragmatic Works

Your Presenter



- Robert C. Cain, MVP, MCTS
- Microsoft MVP SQL Server since 2008
- Senior Consultant with Pragmatic Works
- Technical Contributor to Pluralsight
- Co-Author 3 books
- Contributor to SQL Share
- <http://arcancode.com>
- <http://twitter.com/arcancode>

Why learn about DW/BI?

- **DBA**
 - Implement new Data Warehouse project
 - Install and configure BI tools (SSIS, SSRS, SSAS)
- **DB Designer / Developer**
 - Design and script a DW
 - DW significantly different from traditional database designs.
- **Software Developer**
 - Interact with warehouses for additional data or reports
 - Data mining results into your applications

What is a Data Warehouse

- A giant storehouse for your data
- ALL of your data
- Aggregation of data from multiple systems

What is Business Intelligence

- Leveraging data you already have to convert knowledge into informed actions
- Providing ways to measure the health of your business
- Examining the data in your warehouse to look for three main areas of interest.
- Aggregations
- Trends
- Correlations (Data Mining)

Why Have a Data Warehouse?

- Combine data from multiple systems and resolve inconsistencies between those systems
- Make reporting easier
- Reduce the load on production systems
- Provide for long term storage of data
- Provide consistency among system transitions

Some More Reasons for a Data Warehouse

- Make the data available for analysis
- Ability to apply advanced data mining tools
- To extract further value from the data you already own

Business Intelligence

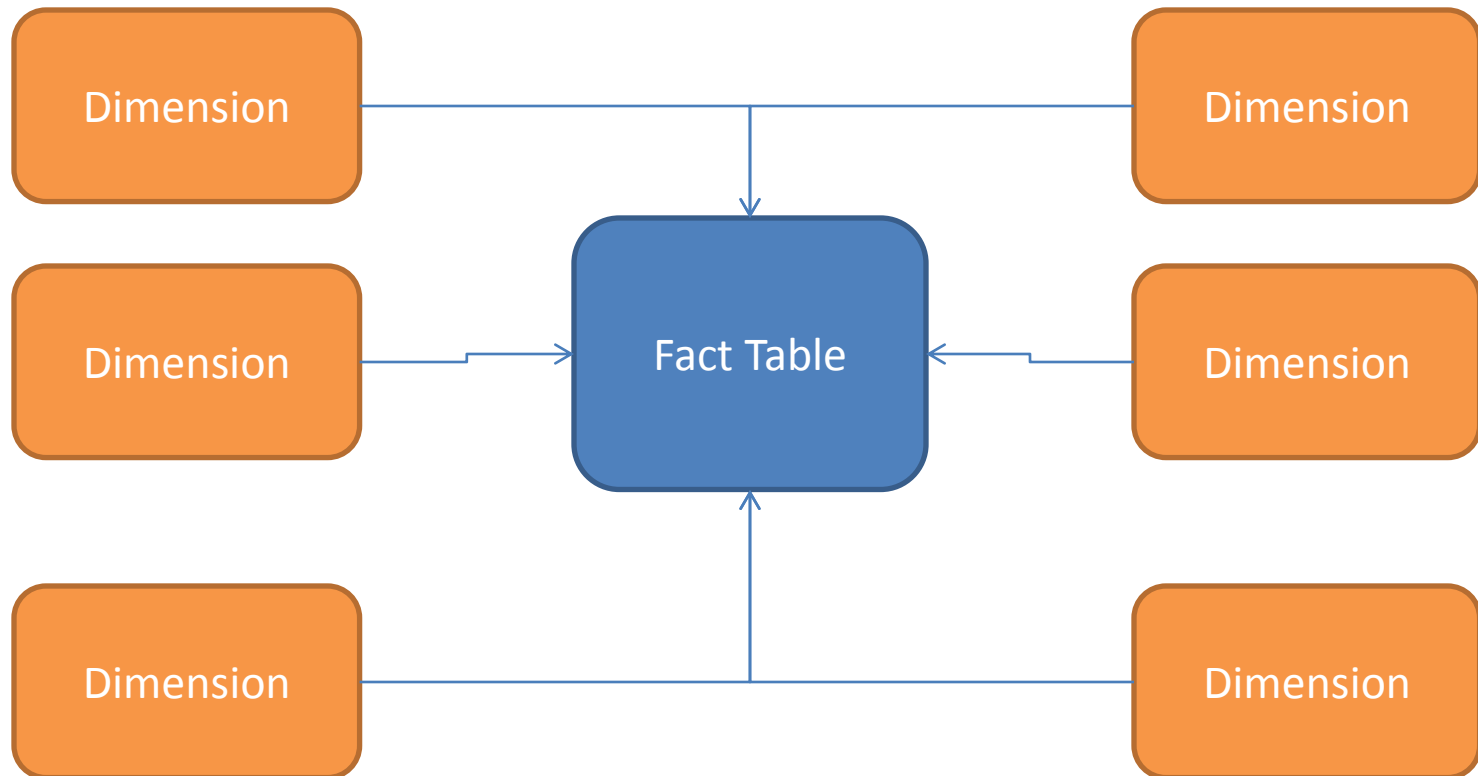
What's wrong with reporting from a Transactional System?

- OLTP – On Line Transaction Processing
- Designed for working with single record at a time.
- Data is highly “normalized”, i.e. duplicate values have been removed.
- Getting all data for a record can involve many table joins
- Can be quite confusing for ‘ad-hoc’ reporting
- Can also be slow, having an impact on the OLTP system

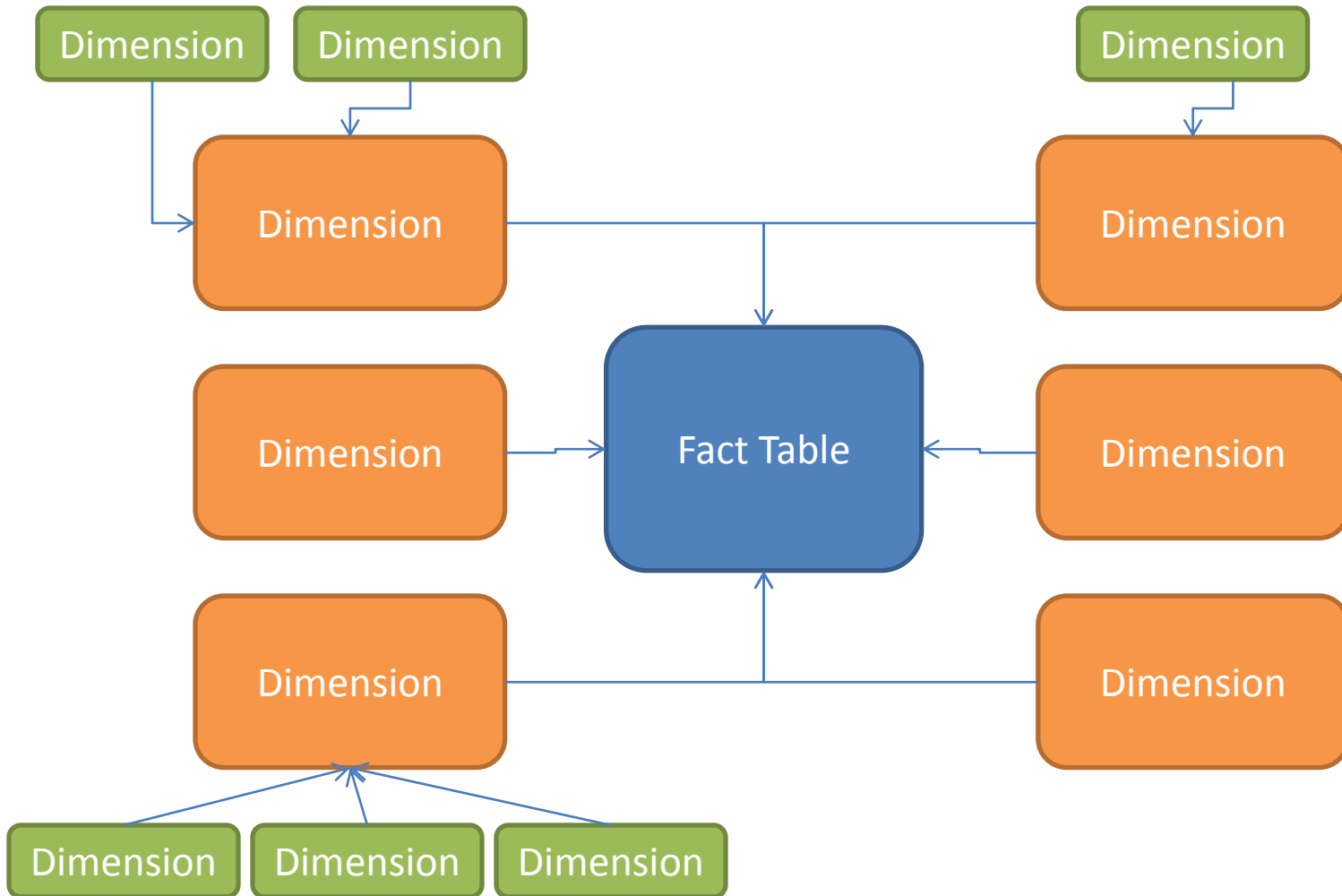
What's different about a Data Warehouse?

- Data Warehouses typically use a design called OLAP
- On-Line Analytical Processing
- Data is de-normalized into structures easier to work with.
- Number of tables are reduced, reducing number of joins and increasing simplicity
- Often a Star Schema or Snowflake Schema

Star Schema



Snowflake Schema



Types of Tables in a Warehouse

- Facts
- Dimensions
- Both require the concept of Surrogate Keys
- A new key, typically some type of INT, that is used in place of any other key as the Primary Key

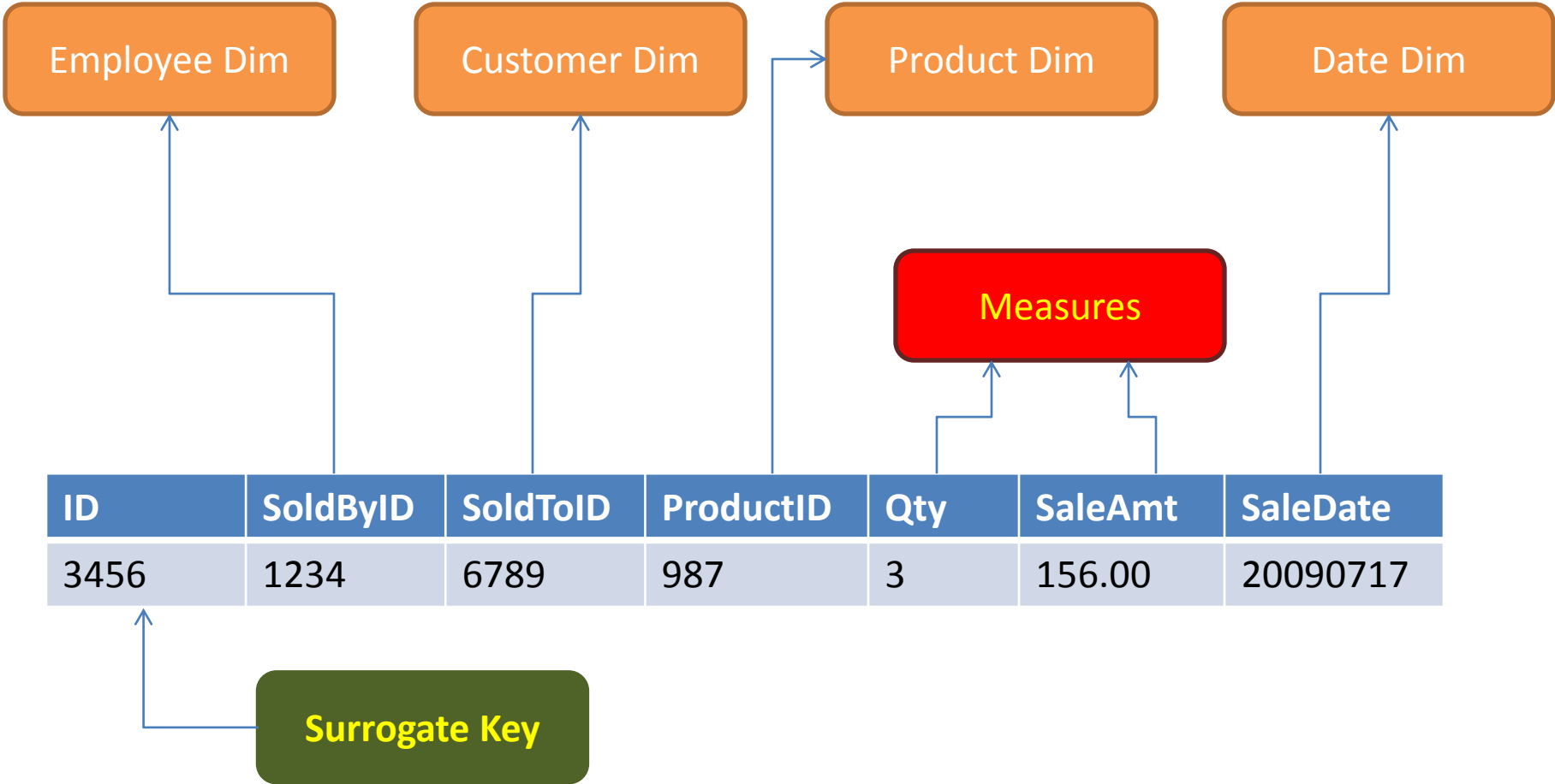
Reasons for Surrogate Keys

- Preserve data in case of source system change
- Combine data from multiple sources into a single table
- Source System keys can be multi-column and complex, slowing response time
- Often the key is not needed for many data warehousing functions such as aggregations

Fact Tables

- A Fact marks an event, a discrete happening in time
- Facts join dimensions, “who” did an action (SoldBy, SoldTo) to the “what”, what object was acted upon (Product) , date of sale (SaleDate)
- Facts also hold numeric measures to quantify the fact: Quantity, SaleAmt

Fact Table Example - Sales



Dimensions

- Dimensions hold the values that describe facts
- “Look Up Values”
- Some examples: Time, Geography, Employees, Products, Customers
- When a Dimension can change over time, it’s known as a Slowly Changing Dimension
- Dimension types typically apply to the column, not the row.
- Many types of Dimensions

Type 0 Dimension (Fixed)

- Type 0 Dimensions are also called Fixed
- For data that will not change. Ever.
- Best used for static data like colors, sizes, etc.

ID	Description
1	Blue
2	Black
3	Green
4	Yellow

Type 1 Dimension

- When a dimensions value is updated, the old one is simply overwritten

Original Value

ID	EmployeeID	Last	First
1234	PQ1894958	McGillicutty	Hortence

New Value

ID	EmployeeID	Last	First
1234	PQ1894958	Hollywoger	Hortence

Type 2 Dimension

- When a dimension is changed, a new record is inserted and old one dated

Original Value

ID	EmployeeID	Last	First	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	12/1/1998	<NULL>

New Value

ID	EmployeeID	Last	First	FromDate	ThruDate
2468	PQ1894958	Hollywoger	Hortence	7/6/2008	<NULL>
1234	PQ1894958	McGillicuty	Hortence	12/1/1998	7/5/2008

Type 3 – Just Say NO

- When a dimensions value is updated, a new column is added



Original Value

ID	EmployeeID	Last1	Last2	First
1234	PQ1894958	McGillicutty		Hortence

New Value

ID	EmployeeID	Last1	Last2	First
1234	PQ1894958	Hollywoger	McGillicutty	Hortence

- Almost never used

Type 4 Dimension

- When a dimension is changed, a old record is copied to history table and current one updated

Original Value in DimEmployee

ID	EmployeeID	Last	First
1234	PQ1894958	McGillicuty	Hortence

New Value in DimEmployee

ID	EmployeeID	Last	First
1234	PQ1894958	Hollywoger	Hortence

New Value in DimEmployee_History

ID	EmployeeID	Last	First	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	12/1/1998	7/5/2008

Type 4 Dimension (Another Way)

- When a dimension is changed, old record is updated in history table, current one copied in

New Value in DimEmployee

ID	EmployeeID	Last	First
1234	PQ1894958	Hollywoger	Hortence

New Value in DimEmployee_History

ID	EmployeeID	Last	First	FromDate	ThruDate
1234	PQ1894958	Hollywoger	Hortence	7/6/2008	<NULL>
1234	PQ1894958	McGillicuty	Hortence	12/1/1998	7/5/2008

Different Dimension Types in a Table

- Often a single row holds multiple Dimensional Types
- The business should determine what data is significant enough to track changes on.

Example

ID	EmployeeBK	Last	First	Phone	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	555-1111	12/1/1998	<NULL>

- Phone Number= Type 1
- Last Name = Type 2

Different Dimension Types in a Table

Original Value

ID	EmployeeBK	Last	First	Phone	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	555-1111	12/1/1998	<NULL>

Update to Hours Last Month (Type 1)

ID	EmployeeBK	Last	First	Phone	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	555-5678	12/1/1998	<NULL>

Update to Last Name (Type 2)

ID	EmployeeBK	Last	First	Phone	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	555-5678	12/1/1998	4/22/2010
6789	PQ1894958	Hollywoger	Hortence	555-5678	4/23/2010	<NULL>

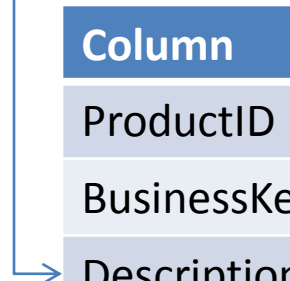
Conformed Dimensions

- When pulling in data from multiple systems, you often have to reconcile different business keys.
- This process is known as conforming your dimensions.

ID	Product	InventoryBK	PurchasingBK	WorkMgtBK
9876	Widget	459684932	Wid45968	602X56VV1

Dimensions in a Star Schema

ID	SoldByID	SoldToID	ProductID	Qty	SaleAmt	SaleDate
3456	1234	6789	987	3	156.00	7/17/2009



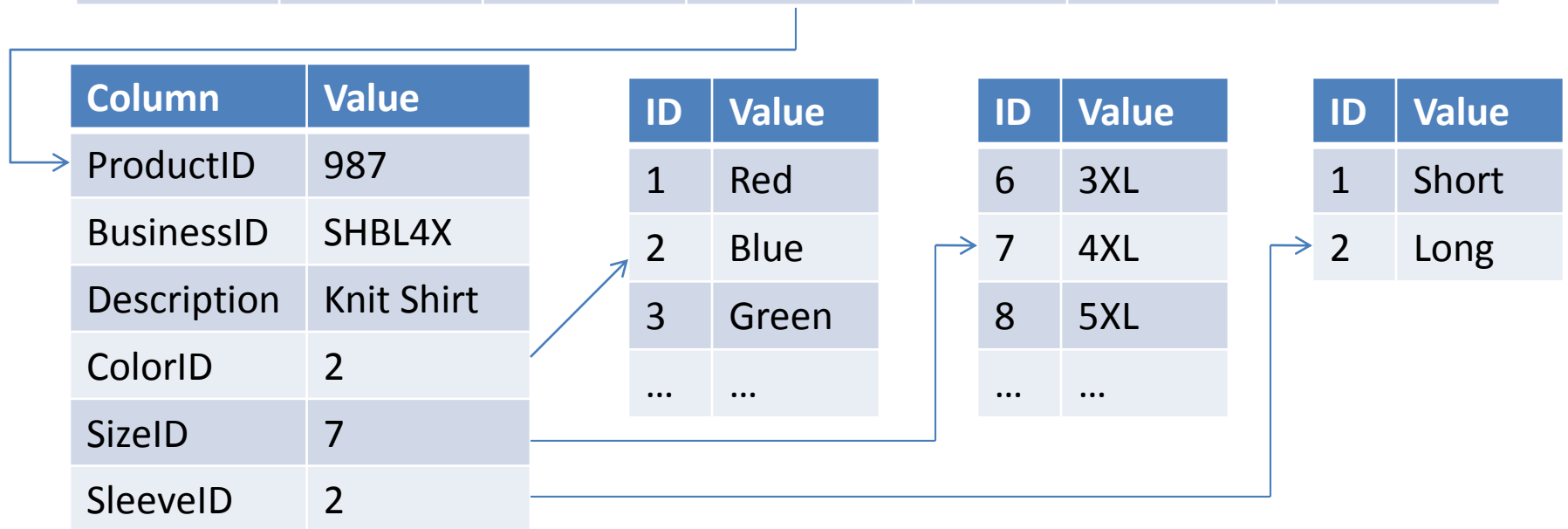
Column	Value
ProductID	987
BusinessKey	SHBL4X
Description	Knit Shirt
Color	Blue
Size	4XL
Sleeve	Long

Dimensions in a Star Schema

ProductID	BusinessKey	Description	Color	Size	Sleeve
987	SHBL4X	Knit Shirt	Blue	4XL	Long
988	SHBL3X	Knit Shirt	Blue	3XL	Long
989	SHBL2X	Knit Shirt	Blue	2XL	Long
990	SHBL1X	Knit Shirt	Blue	1XL	Long
991	SHBLLG	Knit Shirt	Blue	LG	Long
992	SHBLMD	Knit Shirt	Blue	MD	Long
993	SHBLSM	Knit Shirt	Blue	SM	Long

Dimensions in a Snowflake Schema

ID	SoldByID	SoldToID	ProductID	Qty	SaleAmt	SaleDate
3456	1234	6789	987	3	156.00	7/17/2009



Getting Data Into A Warehouse









- ETL
 - Extract
 - Transform
 - Load
- SSIS – SQL Server Integration Services

Getting Data Out of Your Warehouse

- **Data Aggregation, Trending, Correlations**
 - SSAS – SQL Server Analysis Services
 - Multi-dimensional mode
 - Tabular data mode (BISM)
- **Reporting**
 - SSRS – SQL Server Reporting Services
 - Performance Point Services (SharePoint)
 - PowerPivot
 - PowerView

KPI

- Key Performance Indicators
- Dashboards
- Quick, at a glance indicator of system health

Region	Sales (USD)	Trending	Status
US	482m		
Europe	399m		
Asia	123m		
South America	225m		

Warehousing Methodologies

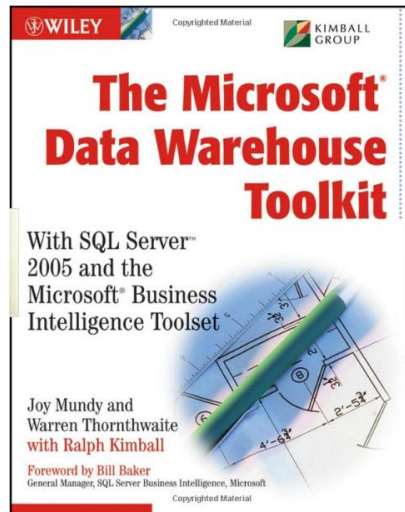
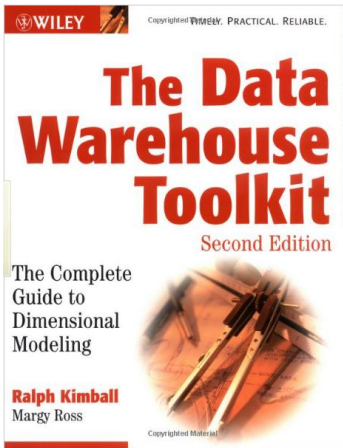
- Inmon – Bill Inmon - Top down
- Kimball – Ralph Kimball - Bottom up

Demos

- Demo Site - <http://www.codeplex.com/Wiki/View.aspx?ProjectName=SqlServerSamples>
- Sample DBs - <http://www.codeplex.com/MSFTDBProdSamples/>
- SSAS - <http://msftasprodsamples.codeplex.com/Release/ProjectReleases.aspx?ReleaseId=18652>
- SSIS - <http://www.codeplex.com/MSFTISProdSamples/>
- SSRS - <http://www.codeplex.com/MSFTRSPProdSamples/>

Resources

The Data Warehouse Toolkit and The Microsoft Data Warehouse Toolkit by the Kimball Group



- http://www.amazon.com/Data-Warehouse-Toolkit-Complete-Dimensional/dp/0471200247/ref=pd_bbs_sr_1?ie=UTF8&s=books&qid=1239580212&sr=8-1
- http://www.amazon.com/MicrosoftData-Warehouse-Toolkit-MicrosoftBusiness-Intelligence/dp/0471267155/ref=sr_1_fkmr0_1?ie=UTF8&qid=1264636802&sr=8-1-fkmr0

Resources

SSAS – <http://bit.ly/arcanessas>

SSIS – <http://bit.ly/arcanessis>

SSRS – <http://bit.ly/arcanessrs>

SharePoint BI – <http://bit.ly/arcanespbi>

PowerPivot – <http://bit.ly/arcanepivot>

PowerShell – <http://bit.ly/arcaneps>

Resources

Blogs

SSIS Junkie - <http://blogs.conchango.com/jamiethomson/default.aspx>

Brian Knight - <http://pragmaticworks.com/community/blogs/brianknight/default.aspx>

Podcast

SQL Down Under - <http://www.sqldownunder.com/PreviousShows/tabid/98/Default.aspx>

SQL Share (formerly JumpstartTV) – <http://sqlshare.com>

Forums

MSDN SSRS Forum - <http://social.msdn.microsoft.com/Forums/en-US/sqlreportingservices/threads>

MSDN SSIS Forum - <http://social.msdn.microsoft.com/Forums/en-US/sqlintegrationservices/threads>

MSDN SSAS Forum - <http://social.msdn.microsoft.com/Forums/en-US/sqlanalysisservices/threads>

Other

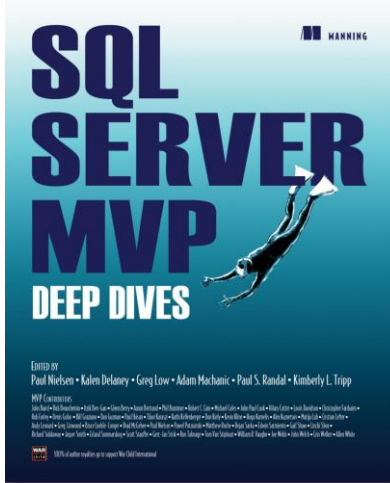
Microsoft BI Site - <http://www.microsoft.com/events/series/bi.aspx>

Wikipedia Article - http://en.wikipedia.org/wiki/Data_Warehouse

SQL Serverpedia - <http://sqlserverpedia.com/>

BIDN: <http://bidn.com>

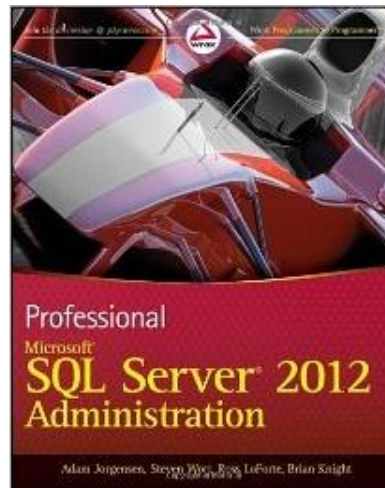
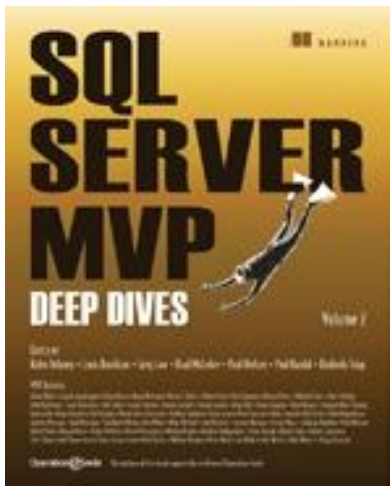
My Books



http://www.amazon.com/SQL-Server-MVP-Deep-Dives/dp/1935182048/ref=sr_1_1?ie=UTF8&qid=1333308555&sr=8-1

http://www.amazon.com/SQL-Server-Deep-Dives-Volume/dp/1617290475/ref=sr_1_2?ie=UTF8&qid=1333308555&sr=8-2

http://www.amazon.com/Professional-Microsoft-Server-Administration-ebook/dp/B007Q4OW7K/ref=kinw_dp_ke?ie=UTF8&m=AG56T WVU5XWC2



Thanks Again!

- Questions?
- All material available at <http://arcanecode.com>
- rcain@pragmaticworks.com

- SSAS – <http://bit.ly/arcanessas>
- SSIS – <http://bit.ly/arcanessis>
- SSRS – <http://bit.ly/arcanessrs>
- SharePoint BI – <http://bit.ly/arcanespbi>
- PowerPivot – <http://bit.ly/arcanepivot>
- PowerShell – <http://bit.ly/arcaneps>