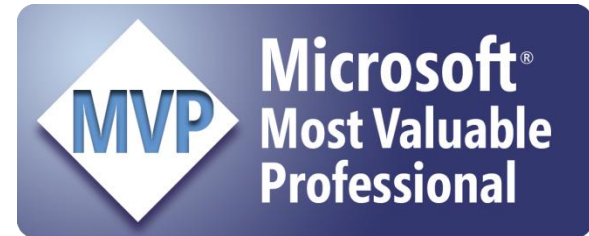


Data Warehousing Business Intelligence

An Introduction

About

- Robert C. Cain
- COMFRAME Sr. BI Architect
- 5 years as BI Dev at regional power company
- 10 years as a consultant in the B'ham Market
- Wide range of .Net applications, ASP & Win
- <http://arcanecode.com>
- <http://twitter.com/arcanecode>
- <http://altechevents.com>



Upcoming Events

- TechMixer University – October 13
 - <http://tinyurl.com/TmuSpeaker09>
- SQL Saturdays - <http://sqlsaturday.com/>
 - #26 - Oct. 3 – Redmond, WA
 - #25 - Oct. 10 – Gainesville, GA
- PASS Summit
 - November 2-5 Seattle WA
 - <http://sqlpass.org>

Why learn about Data Warehousing?

- 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.

Business Intelligence is HOT

- According to [Computerworld](#), BI is the 5th hottest IT Skill for 2009
- Dice.com over 2,800 job openings

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
- Examining the data in your warehouse to look for three main areas of interest.
- Aggregations
- Trends
- Corrolations (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

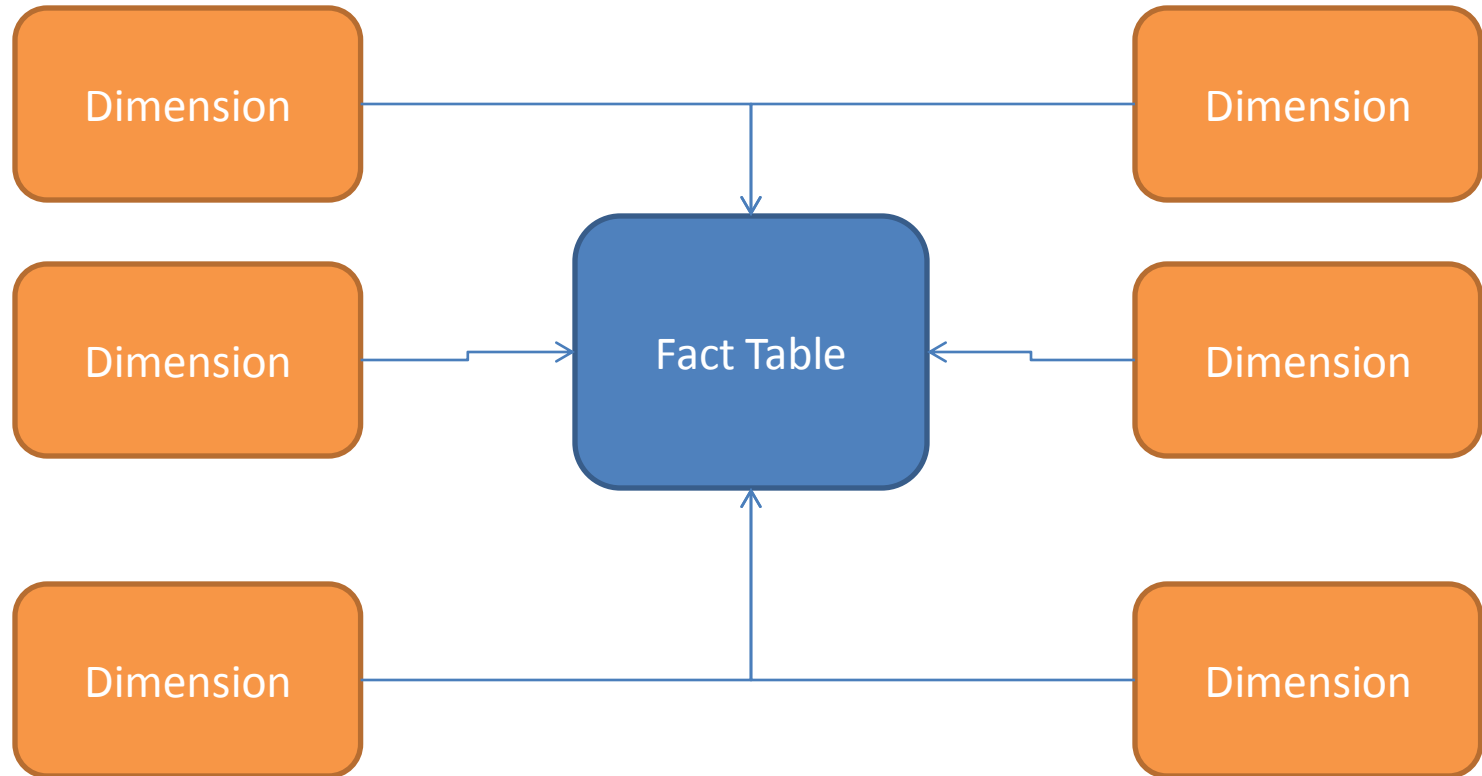
What's wrong with reporting from a Transactional System?

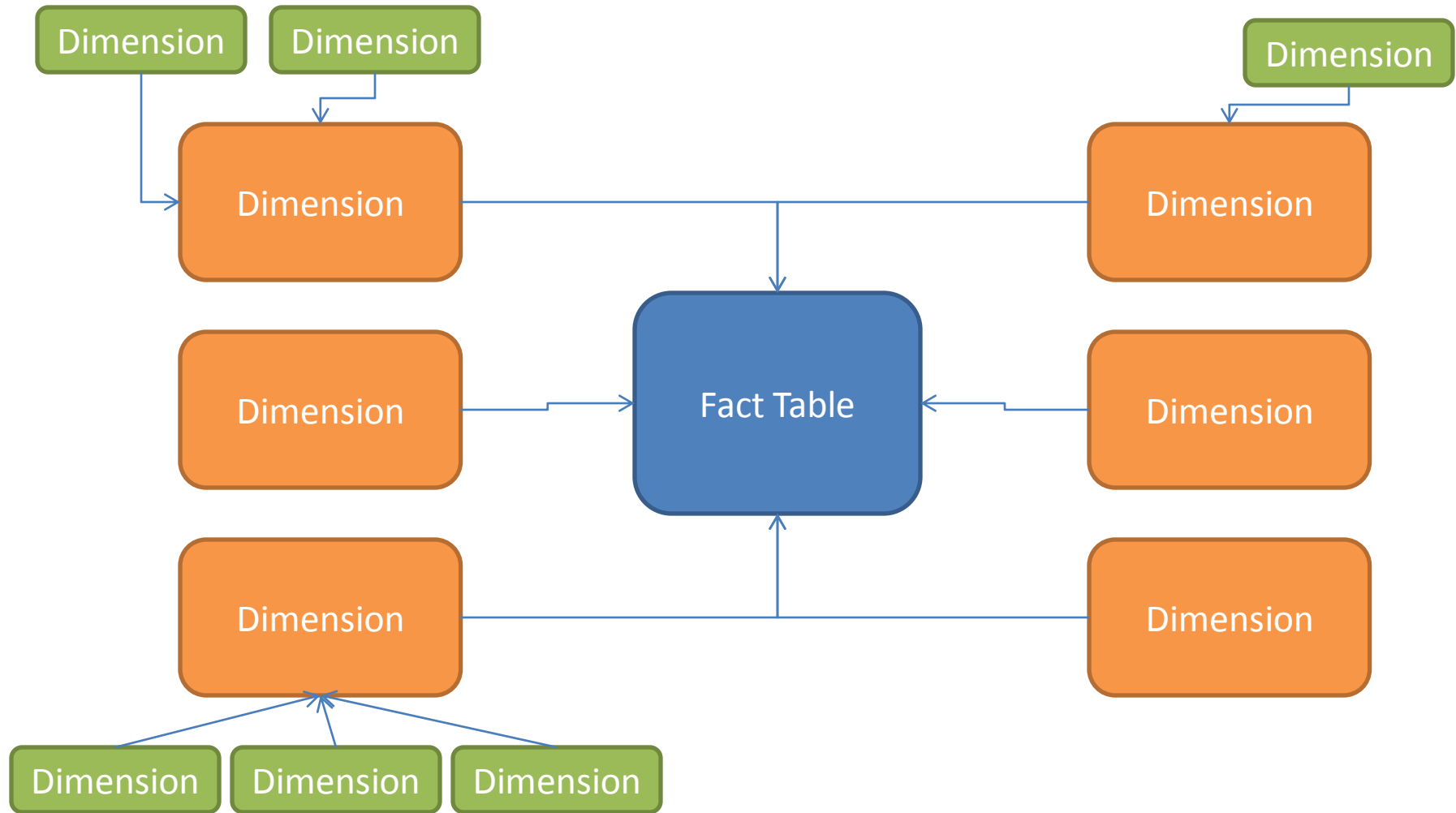
- OLTP – On Line Transaction Processing
- Designed for working with single record at a time.
- Data is highly “normalized”, i.e. duplicate values have been removed.
- Getting all data for a record can involve many table joins
- Can be quite confusing for ‘ad-hoc’ reporting
- Can also be slow, having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





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

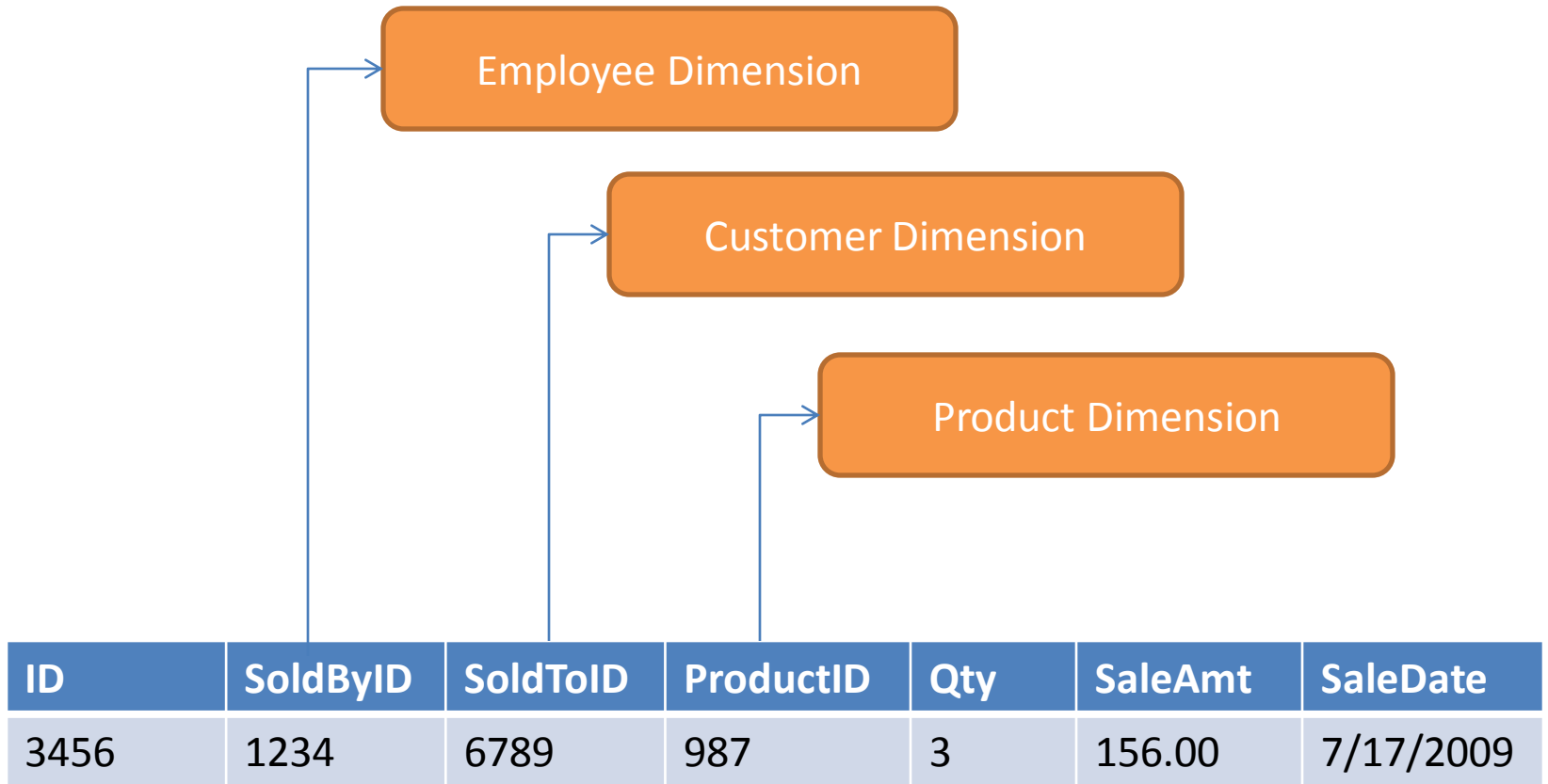
Reasons for Surrogate Keys

- 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 usually hold numeric measures and/or links to dimensions ID: SoldBy, SoldTo, Product Qty, SaleAmt, SaleDate

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
- 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 dimension's 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	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

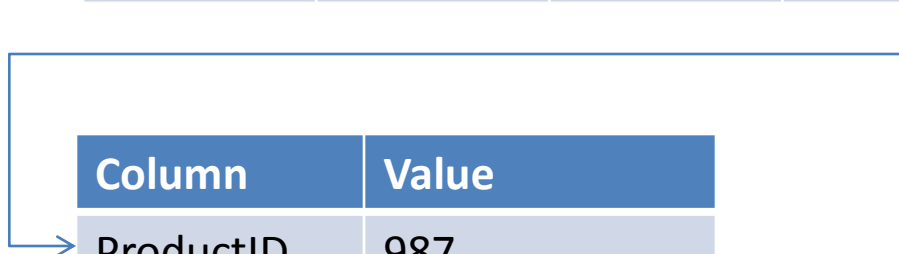
Conformed Dimensions

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

ID	Product	InventoryID	PurchasingID	WorkMgtID
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

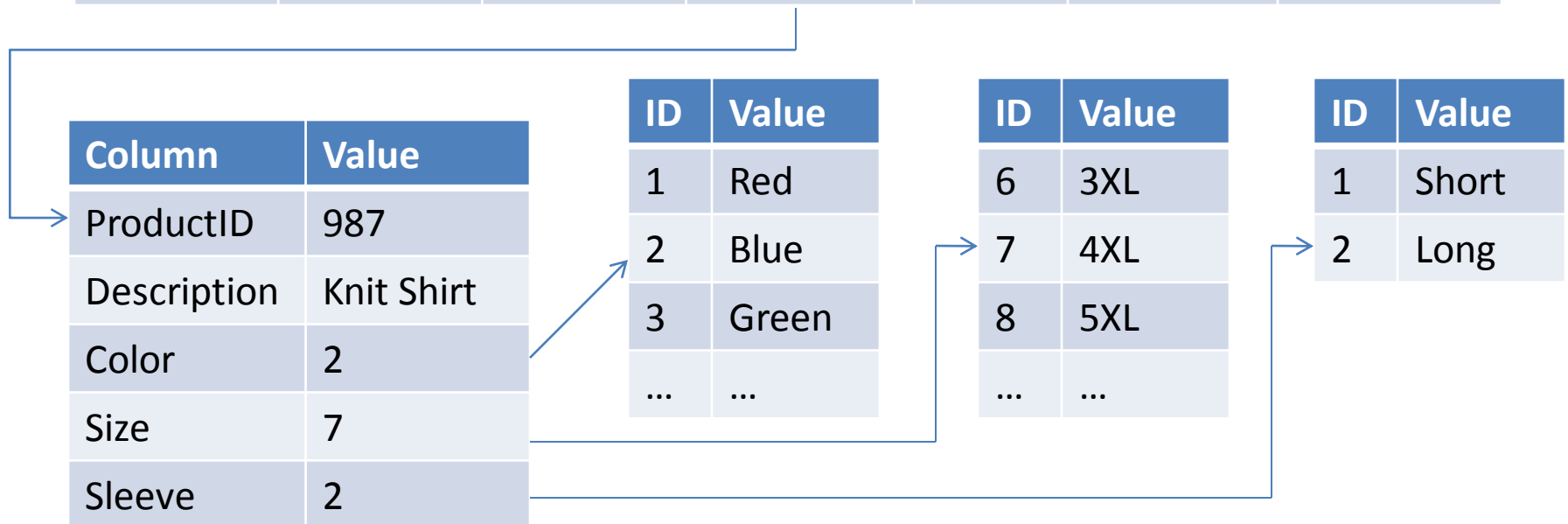


Column	Value
ProductID	987
Description	Knit Shirt
Color	Blue
Size	4XL
Sleeve	Long

- Flattened model easier to link
- Simpler for ad-hoc reporting
- Takes more database space
- Updates take more work (4XL to XXXXL requires 1 update per record)

Dimensions in a Snowflake Schema

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



- Takes less database space
- Easier to update (change 4XL to XXXXL requires 1 Update)
- More difficult as it requires many links
- More links makes it difficult for ad-hoc reporting (views can help with this)

Getting Data Into A Warehouse

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

Getting Data Out of Your Warehouse

- SSRS – SQL Server Reporting Services
- SSAS – SQL Server Analysis Services

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		

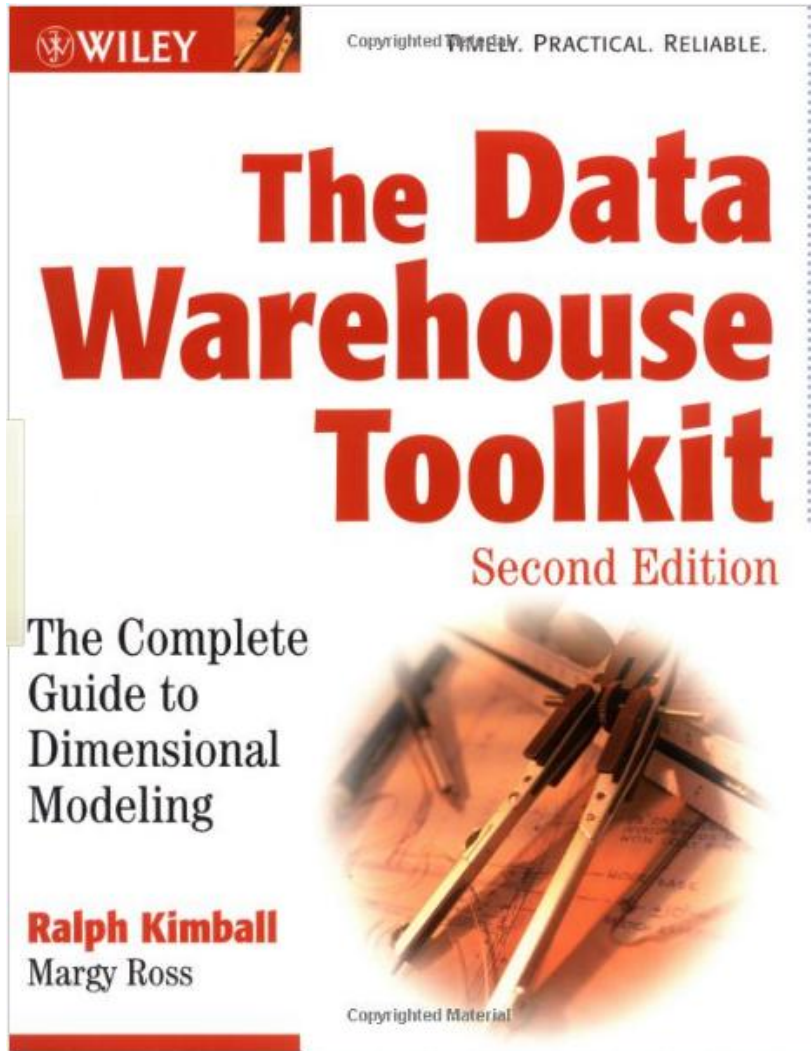
Warehousing Methodologies

- 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/>

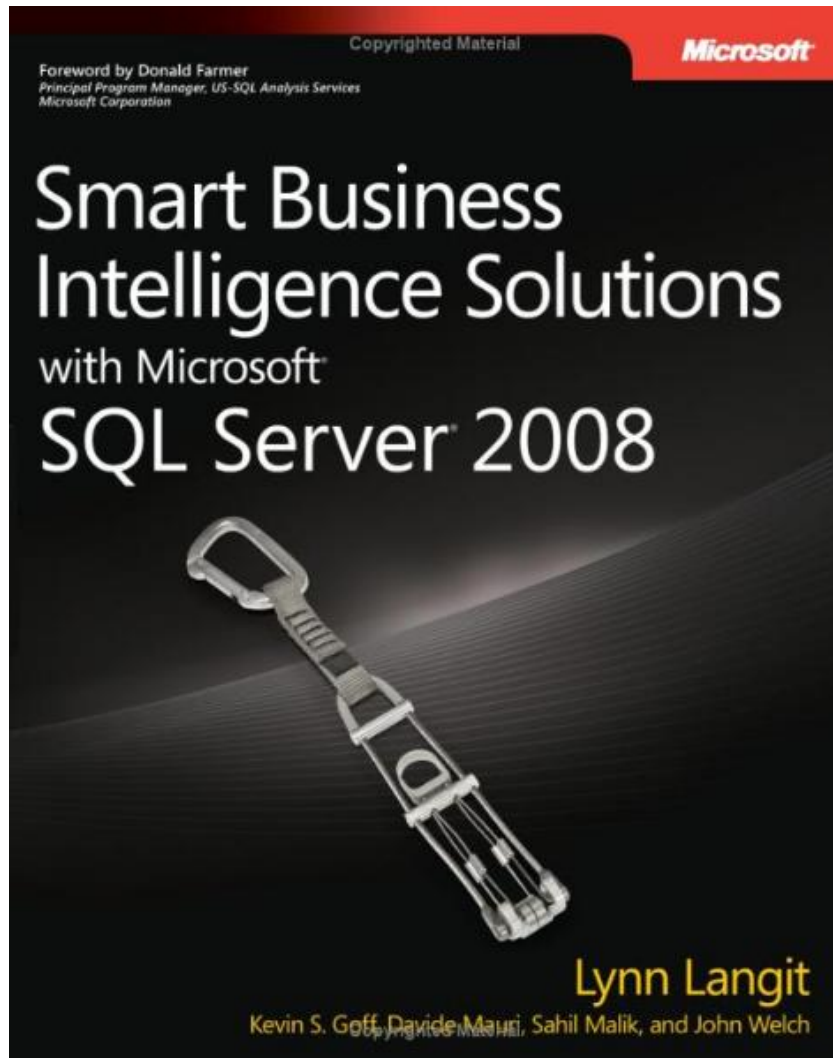
Resources



The Data Warehouse Toolkit 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

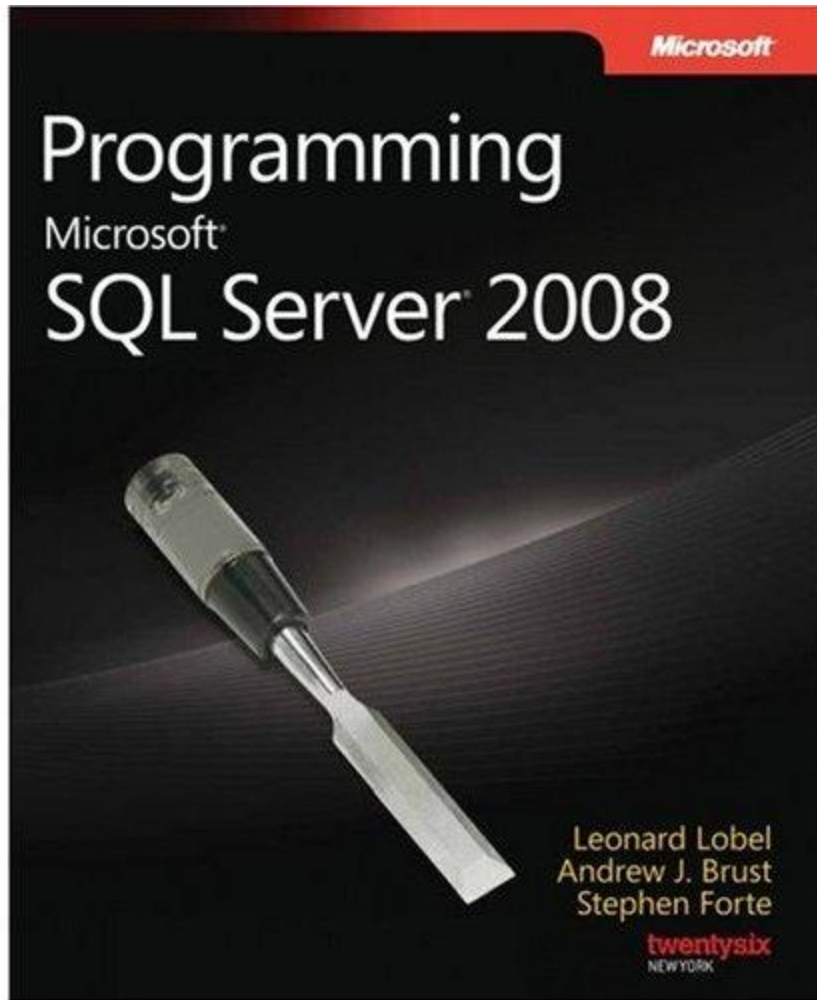
Resources



Smart Business Intelligence Solutions with Microsoft SQL Server 2008

- http://www.amazon.com/Business-Intelligence-Solutions-Microsoft%C2%AE-PRO-Developer/dp/0735625808/ref=sr_1_1?ie=UTF8&s=books&qid=1239580654&sr=1-1

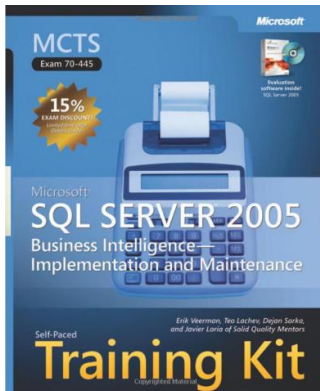
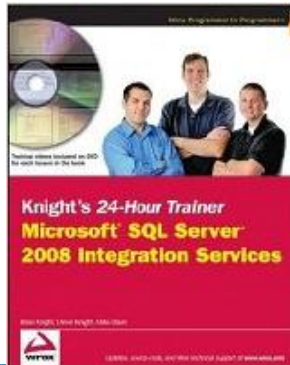
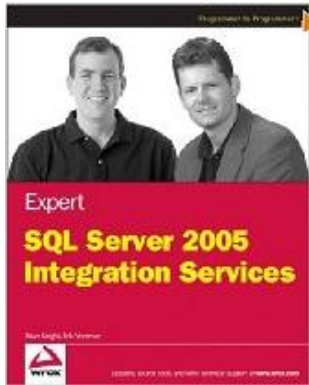
Resources



Programming Microsoft SQL Server 2008

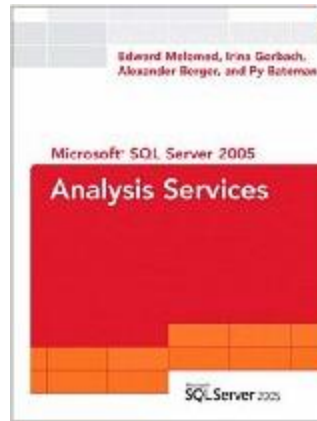
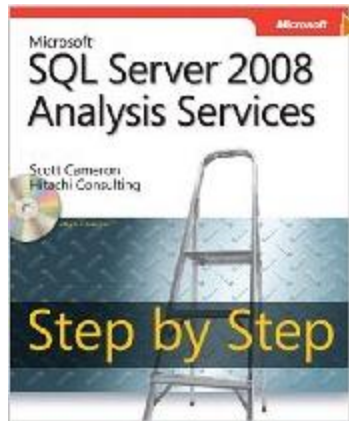
- http://www.amazon.com/Programming-Microsoft-Server-2008-PRO-Developer/dp/0735625999/ref=sr_1_1?ie=UTF8&s=books&qid=1239580376&sr=1-1

Resources - SSIS

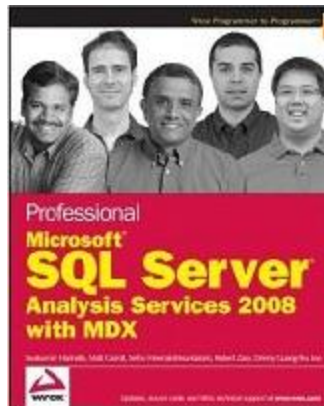


- Erik Veerman /
Brian Knight Books
 - http://www.amazon.com/Expert-Server-Integration-Services-Programmer/dp/0470134119/ref=sr_1_5?ie=UTF8&s=books&qid=1239833324&sr=8-5
 - http://www.amazon.com/Professional-Microsoft-Integration-Services-Programmer/dp/0470247959/ref=sr_1_1?ie=UTF8&s=books&qid=1239833324&sr=8-1
 - http://www.amazon.com/MCTS-Self-Paced-Training-Exam-70-445/dp/0735623414/ref=sr_1_7?ie=UTF8&s=books&qid=1239833324&sr=8-7#
 - http://www.amazon.com/MCTS-Self-Paced-Training-Exam-70-448/dp/0735626367/ref=sr_1_4?ie=UTF8&s=books&qid=1239833324&sr=8-4
 - http://www.amazon.com/reader/0470496924?encoding=UTF8&ref=sib_dp_pop_fc&page=1#reader

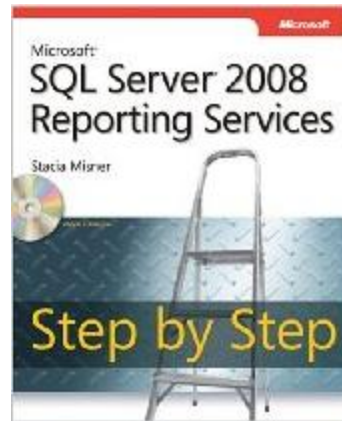
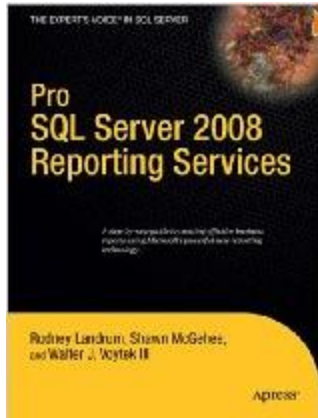
Resources - SSAS



- http://www.amazon.com/Microsoft%C2%AE-Server%C2%AE-Analysis-Services-Microsoft/dp/0735626200/ref=sr_1_5?ie=UTF8&s=books&qid=1252100419&sr=1-5
- http://www.amazon.com/Microsoft-Server-2005-Analysis-Services/dp/0672327821/ref=sr_1_2?ie=UTF8&s=books&qid=1252100419&sr=1-2
- http://www.amazon.com/Professional-Microsoft-Analysis-Services-Programmer/dp/0470247983/ref=sr_1_4?ie=UTF8&s=books&qid=1252100419&sr=1-4



Resources - SSRS



- http://www.amazon.com/Pro-Server-2008-Reporting-Services/dp/1590599926/ref=sr_1_11?ie=UTF8&s=books&qid=1252100749&sr=1-11
- http://www.amazon.com/Microsoft%C2%AE-Server%C2%AE-Reporting-Services-Microsoft/dp/0735626472/ref=sr_1_10?ie=UTF8&s=books&qid=1252100749&sr=1-10
- http://www.amazon.com/Professional-Microsoft-Reporting-Services-Programmer/dp/0470242019/ref=sr_1_15?ie=UTF8&s=books&qid=1252100793&sr=1-15



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>

JumpstartTV - <http://jumpstarttv.com/>

Forums

MSDN SSIS Forum - <http://forums.microsoft.com/MSDN/ShowForum.aspx?ForumID=80&SiteID=1>

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/>

Thanks Again!

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