Overview of Data Warehousing / Business Intelligence With SQL Server

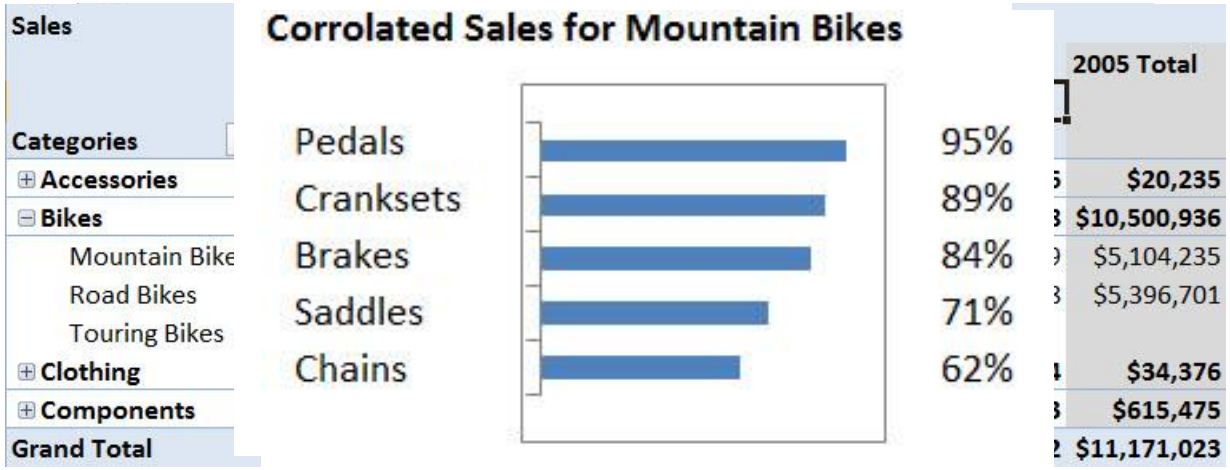
Robert C. Cain, MVP, MCTS
http://www.pluralsight.com

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
- 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 consistency among system transitions
- Provide for long term storage of data

What's wrong with reporting from Transactional Systems?

- 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
- Number of tables are reduced, reducing number of joins and increasing simplicity
- Data is de-normalized into structures easier to work with.

Normalized vs. Denormalized

Normalized – Data is broken into multiple tables

Product	
ProductID	Desc
1	Mtn Bike #778
2	Road Bike #123
3	Touring Bike #222

Color	
ColorID	Desc
1	Red
2	Black
3	Silver
4	Mauve

Product-Color	
ProductID	ColorID
1	1
1	2
2	1
2	2
2	3
3	1
3	3
3	4

Normalized vs. Denormalized

Denormalized – Data combined

Product	(denormalized)			
ProductSK	ProductID	ColorID	Desc	Color
1	1	1	Mtn Bike #778	Red
2	1	2	Mtn Bike #778	Black
3	2	1	Road Bike #123	Red
4	2	2	Road Bike #123	Black
5	2	3	Road Bike #123	Silver
6	3	1	Touring Bike #222	Red
7	3	3	Touring Bike #222	Silver
8	3	4	Touring Bike #222	Mauve

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

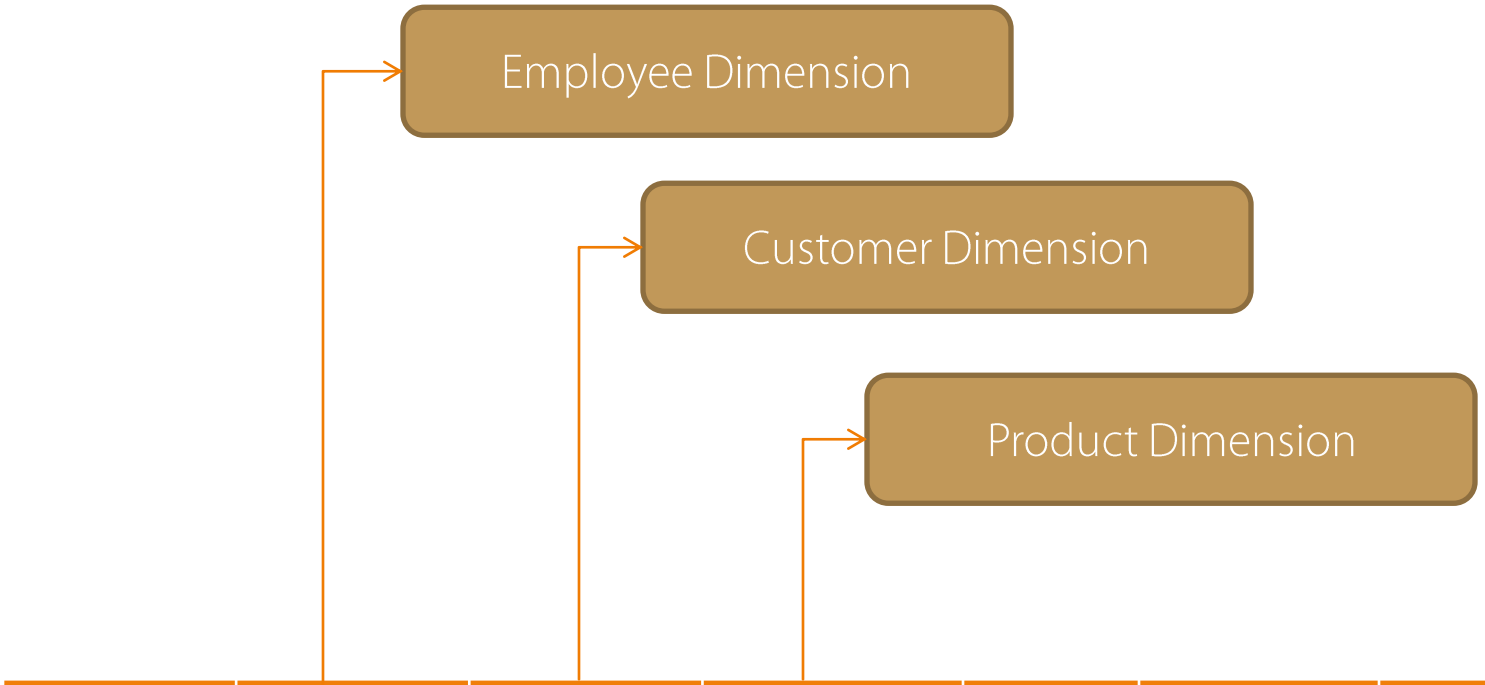
Reasons for Surrogate Keys

- 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”, “what”, “when”, and “where”
- Facts also hold numeric measures to quantify the fact, “how much”

Fact Table Example - Sales



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

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

Static Data

- For data that will not change. Ever.
- Best used for static data like colors, sizes, etc.
- Known as a Type 0 Dimension

ID	Description
1	Blue
2	Black
3	Green
4	Yellow

No history is required

- **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

- **Referred to as a Type 1 dimension**

The impact when no history is required

Sales Report

Sales Person	Month	Amount
Hortence McGillicuty	Apr-2008	$ 1,000
Hortence McGillicuty	May-2008	$ 2,300
Hortence McGillicuty	Jun-2008	$ 1,934
Hortence McGillicuty	Jul-2008	$ 232
Hortence McGillicuty	Aug-2008	$ -

The impact when no history is required

Sales Report

Sales Person	Month	Amount
Hortence McGillicuty	Apr-2008	$ 1,000
Hortence McGillicuty	May-2008	$ 2,300
Hortence McGillicuty		
Hortence McGillicuty		
Hortence McGillicuty		

Sales Report

Sales Person	Month	Amount
Hortence Hollywoger	Apr-2008	$ 1,000
Hortence Hollywoger	May-2008	$ 2,300
Hortence Hollywoger	Jun-2008	$ 1,934
Hortence Hollywoger	Jul-2008	$ 232
Hortence Hollywoger	Aug-2008	$ -

Tracking changes is important

- **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 2 dimension**

The impact of tracking changes

Sales Report

Sales Person	Month	Amount
Hortence McGillicuty	Apr-2008	$ 1,000
Hortence McGillicuty	May-2008	$ 2,300
Hortence McGillicuty	Jun-2008	$ 1,934
Hortence McGillicuty	Jul-2008	$ 232
Hortence McGillicuty	Aug-2008	$ -

The impact of tracking changes

Sales Report

Sales Person	Month	Amount
Hortence McGillicuty	Apr-2008	$ 1,000
Hortence McGillicuty	May-2008	$ 2,300
Hortence McGillicuty		
Hortence McGillicuty		
Hortence McGillicuty		

Sales Report

Sales Person	Month	Amount
Hortence McGillicuty	Apr-2008	$ 1,000
Hortence McGillicuty	May-2008	$ 2,300
Hortence McGillicuty	Jun-2008	$ 1,934
Hortence Hollywoger	Jul-2008	$ 232
Hortence Hollywoger	Aug-2008	$ -

Separating history from day to day data needs

- **When a dimension is changed, old record is updated in history table, current one copied in (type 4 dimension)**

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	DimEmpID	EmployeeID	Last	First	FromDate	ThruDate
7564	1234	PQ1894958	Hollywoger	Hortence	7/6/2008	<NULL>
8945	1234	PQ1894958	McGillicuty	Hortence	12/1/1998	7/5/2008

Different Dimension Types in a Table

- **Often a single row holds multiple Dimensional Types.**

Example

ID	EmployeeID	Last	First	HrsLastMo	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	200	12/1/1998	<NULL>

- **Hours Last Month = Type 1**
- **Last Name = Type 2**

Different Dimension Types in a Table

Original Value

ID	EmployeeID	Last	First	HrsLastMo	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	200	12/1/1998	<NULL>

Update to Hours Last Month (Type 1)

ID	EmployeeID	Last	First	HrsLastMo	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	280	12/1/1998	<NULL>

Update to Last Name (Type 2)

ID	EmployeeID	Last	First	HrsLastMo	FromDate	ThruDate
1234	PQ1894958	McGillicuty	Hortence	200	12/1/1998	4/22/2010
6789	PQ1894958	Hollywoger	Hortence	200	4/23/2010	<NULL>

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
BusinessID	SHBL4X
Description	Knit Shirt
Color	Blue
Size	4XL
Sleeve	Long

Dimensions in a Snowflake Schema

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

Column	Value
ProductID	987
BusinessID	SHBL4X
Description	Knit Shirt
Color	2
Size	7
Sleeve	2

ID	Value
1	Red
2	Blue
3	Green
…	…

ID	Value
6	3XL
7	4XL
8	5XL
…	…

ID	Value
1	Short
2	Long

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		

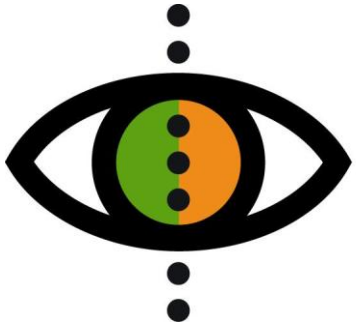
The Microsoft Toolset

- **ETL**
  - Extract – Transform - Load
  - SSIS – SQL Server Integration Services
- **Analytics**
  - Aggregation – Trending - Correlations
  - SSAS – SQL Server Analysis Services
- **Reporting**
  - SSRS – SQL Server Reporting Services
  - SharePoint Performance Point
- **PowerPivot**
  - Add-in for Microsoft Excel

Summary

- What is DW/BI
- Why use DW/BI?
- Defined many of the terms, such as facts, dimensions, and surrogate keys using concrete examples.
- When to use dimensional types
- Microsoft tools around DW/BI

For more in-depth online developer training visit



pluralsight

see what you can learn

on-demand content from authors you trust