



■ ■ ■ ■ **Tiber Solutions**

**Optimal Design Solutions For
Successful BI Applications**

Jim Hadley

Tiber Solutions

- Founded in 2005 to provide Business Intelligence / Data Warehousing thought leadership to corporations and government agencies.
- Deeply skilled in all facets of BI/DW solutions – star schema, ETL, BI, data architecture, information architecture, BI agile development methodology, and MDM/governance.
- Provide hands-on coaching within IT organizations from the CIO down to the developers.
- Partner with business executives to co-invent optimal information-delivery applications to dramatically improve their business.

Tiber Solutions

- Amtrak
- Business Objects
- ▶ Census Bureau
- ▶ Department of the Treasury
- Fannie Mae
- ▶ Federal Depository Insurance Corporation
- Freddie Mac
- ▶ Graduate Management Admission Council
- Internal Revenue Service
- Military Health System
- ▶ Occupational Safety and Health Administration
- Securities and Exchange Commission

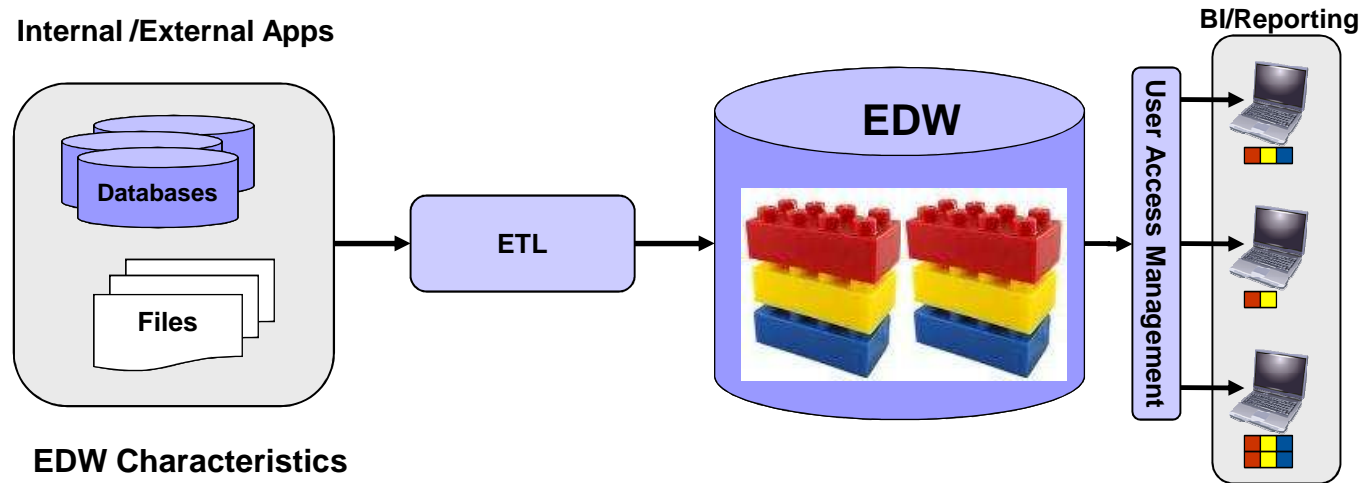
Agenda

- The Hardest Problem
- The Foundation – Architecture and Methodology
- Collecting Requirements to Fail
- Collecting Requirements to Succeed
- Information Delivery – Universe Design Best Practices
- Information Delivery – Report/Dashboard Design Best Practices

The Hardest Problem

- Developing a BI application is much more complicated than developing a transaction (OLTP) application. Transaction application requirements are constrained and static.
- BI application requirements are broad and dynamic.
- The presentation of information requires assimilating large volumes of useful and often disparate data into a small amount of space (e.g., one page), all while preserving clarity.
- **Hardest Problem:** *How do I accurately design and deliver an application based on dynamic, broad business requirements which requires high volumes of detailed data to be assimilated, aggregated, and visually displayed as meaningful information with clarity, precision, relevance, and quick understanding?*

The Foundation – Architecture



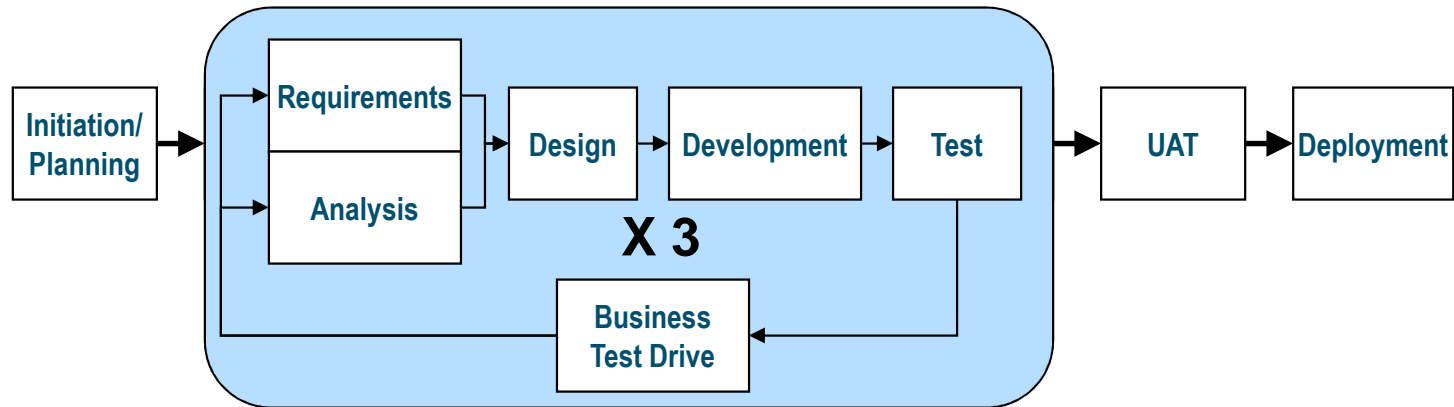
EDW Characteristics

- Star Schema Model
- Atomic (transaction) level of detail
- Developed incrementally by subject area (approx 3-4 month)
- Integrated by using conformed dimensions across subject areas
- Developed iteratively – involving end users early/often in development
- Data marts are “configured” by grouping subject areas into virtual data marts per user group

EDW Purpose

- Reporting/analytics ease of use
- High query performance
- Cross-subject area analysis
- Supports dim and fact reuse and consistency

The Foundation – Methodology



Benefits

- Business users provide feedback early and throughout the process which guides development.
- Business users see results early.
- Business users begin to understand the enabling capabilities of the tools.
- Business users begin to trust the data.
- Business users begin to invent. They begin to think outside the box.
- The solution is “proven out” and matured throughout the development process.

Collecting Requirements to Fail

Typical Requirement Session Events	Typical Outcome
Several business questions/statements are collected (How many banks are predicted to fail next quarter?).	Business questions/statements are never translated into a meaningful design and never implemented.
A set of reports that needs to be recreated is collected.	95% of the development time is spent successfully recreating the set of reports.
End users request the ability to create ad hoc queries.	Data model is hyper-designed for a specific set of reports and is very brittle. Most ad hoc queries are not supported.
End users are not involved with the project again until UAT.	End users are unfamiliar, untrusting, and unsatisfied with the application.

Collecting Requirements to Succeed

Business questions / statements	<ul style="list-style-type: none">• How successful were our sales last year?• What was the most overstocked warehouse?• Which divisions will need to hire people to replace retiring employees?
Successful Format	Fact by Dim 1, Dim 2, ... Dim N
Examples	<ul style="list-style-type: none">• Net Sales by Month, Product Name, and Sales Division• Inventory Count by Product, Warehouse, and Quarter for 2011• Employee Count by Projected Retirement Year and Division

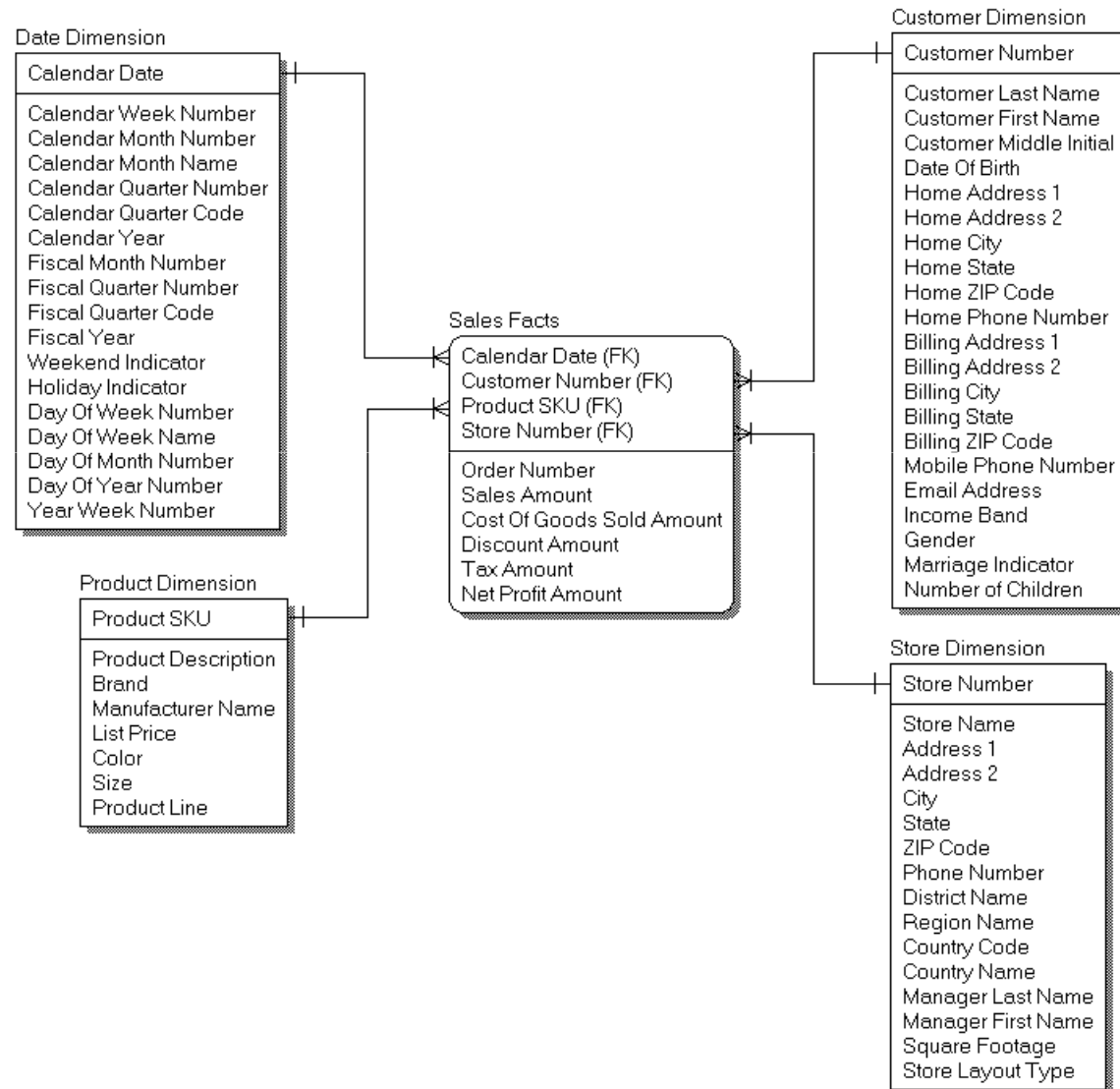
Collecting Requirements to Succeed

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH		
1		Time												Product							Customer															
2	Facts	date	day of week	day of month	day of year	calendar month	calendar quarter	calendar year	fiscal month	fiscal quarter	fiscal year	holiday flag	season	product id	sku	product name	brand name	manufacturer name	unit price	customer account	customer full name	customer last name	customer first name	customer middle initial	street address	city	state	zip code	billing street address	billing city	billing state	billing zip code	salary range	marital status		
3	net sales amount	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	cost of goods sold	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5	gross sales amount	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
6	tax amount	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
7	discount amount					x	x	x	x	x	x	x	x	x	x	x	x	x	x																	
8	promotion amount					x	x	x	x	x	x	x	x	x	x	x	x	x	x																	
9	quantity sold	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Benefits

- Full requirement coverage to support ad hoc queries.
- Common language between business and developers.
- Easy transition to logical star schema design and universe (test drives)

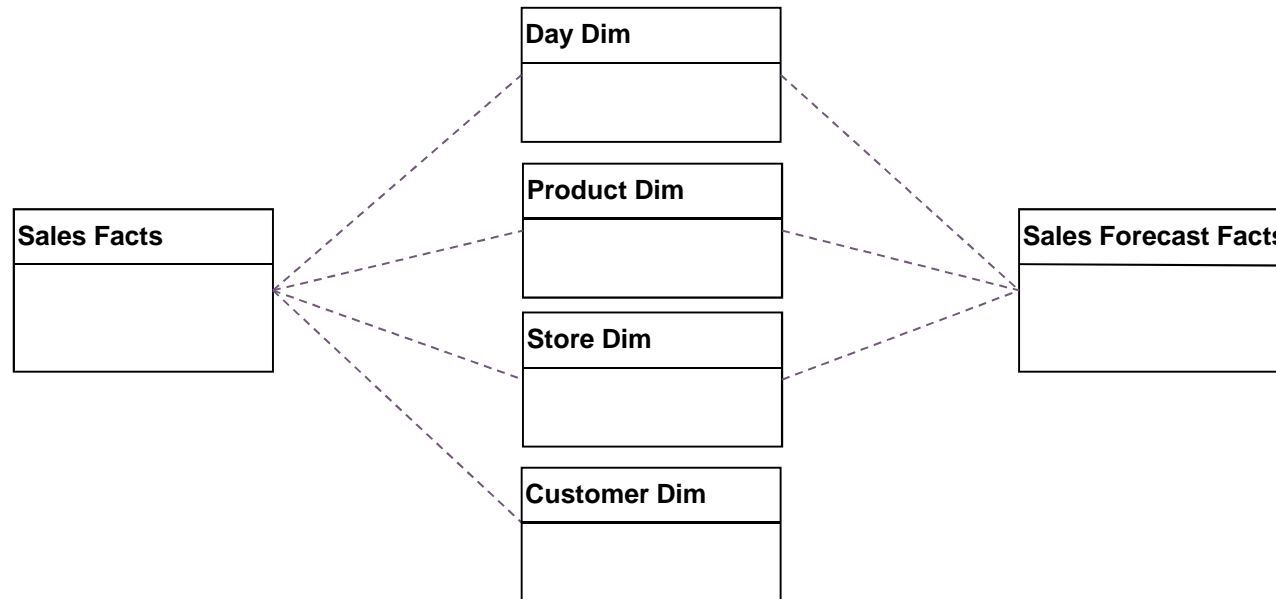
Collecting Requirements to Succeed



Information Delivery – Universe Design Best Practices

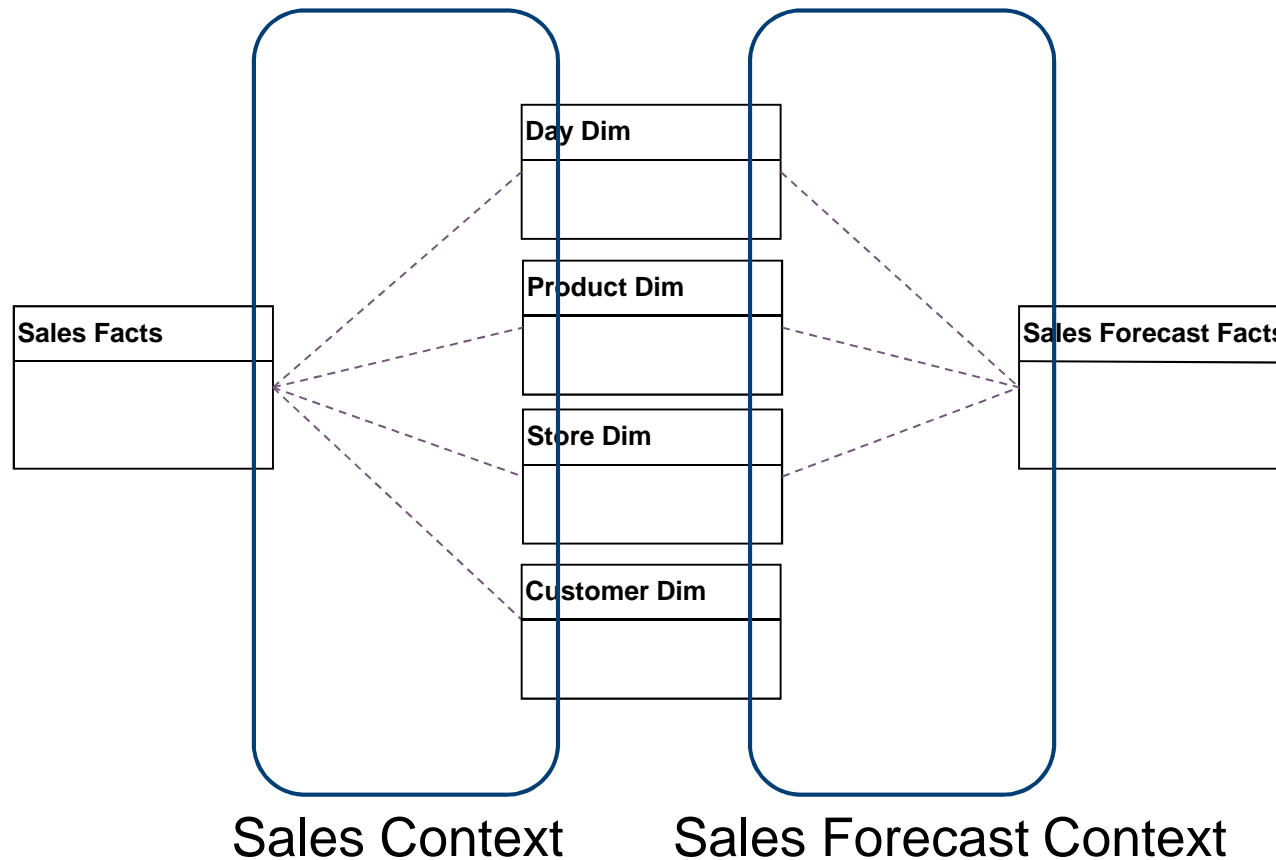
- All class, object, filter, and prompt names should use business terms that are familiar to end users.
- Class (i.e., folder) order should begin with the most common dimension classes (e.g., date-related dimension objects) to the least common dimension classes followed by the most common measure classes to the least common measure classes.
- If a fact table requires a dimension table to take on multiple “roles”, the dimension table should be aliased for each role. Each alias dimension should be joined to each foreign key in the fact table.

Information Delivery – Universe Design Best Practices



- Each fact table and its related dimension tables should be defined in a context to support multi-pass SQL.

Information Delivery – Universe Design Best Practices



- Each fact table and its related dimension tables should be defined in a context to support multi-pass SQL.

Information Delivery – Universe Design Best Practices

Database Aggregation

The screenshot shows the 'Edit Properties of Sales Quantity' dialog box with the 'Definition' tab selected. The 'Name' field contains 'Sales Quantity' and the 'Type' dropdown is set to 'Number'. The 'Description' field is empty. The 'Select' field contains the SQL expression 'SUM(SalesFact.SalesQuantity)'. The 'Where' field is empty. At the bottom, there are buttons for 'Tables...', 'Parse', 'OK', 'Cancel', 'Apply', and 'Help'.

Report Aggregation

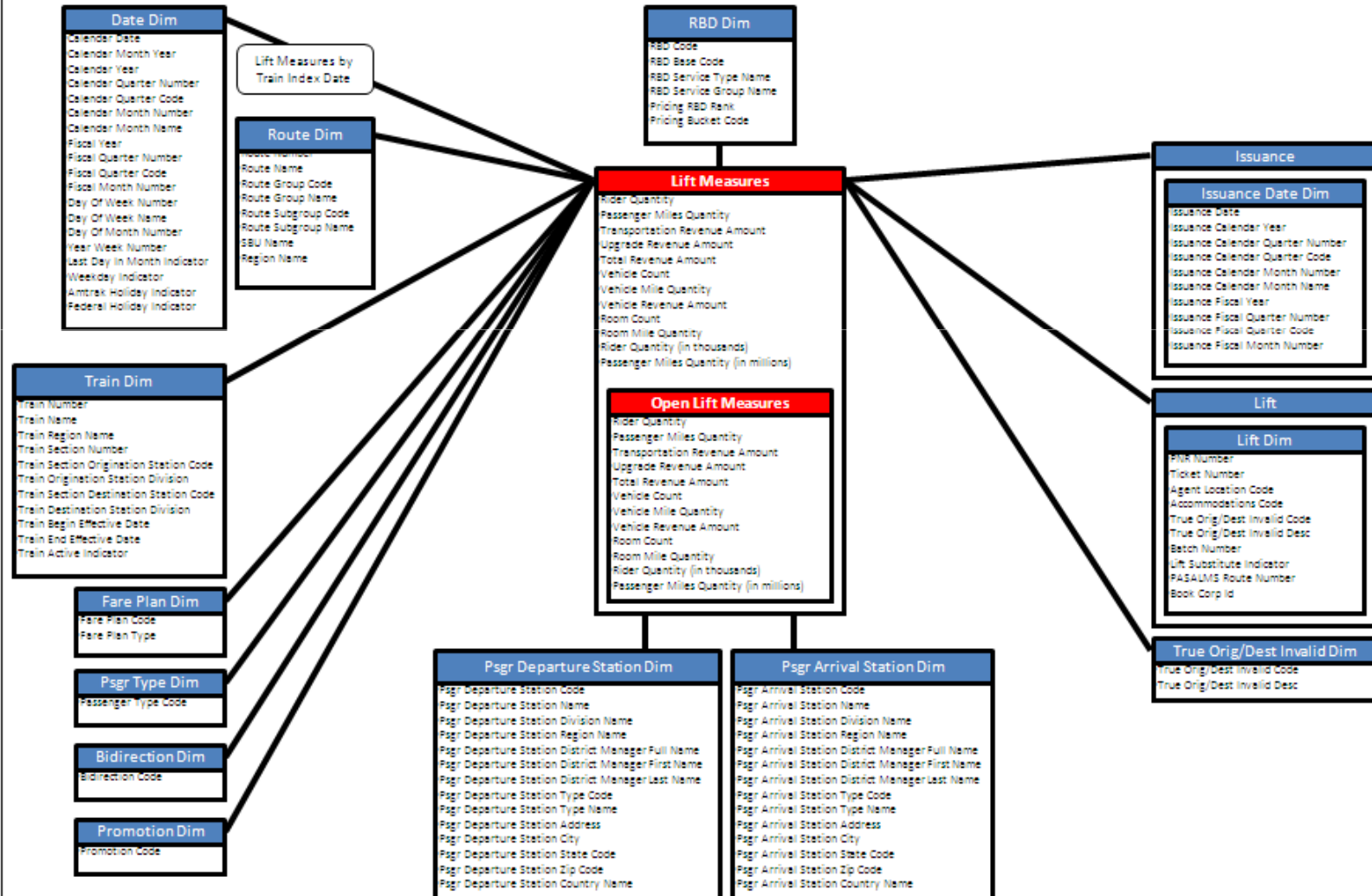
The screenshot shows the 'Edit Properties of Sales Quantity' dialog box with the 'Properties' tab selected. The 'Qualification' section is active, showing 'This object has the following qualification for multidimensional analysis:'. There are three radio button options: 'Dimension' (unselected), 'Measure' (selected), and 'Detail' (unselected). Below this, the text says 'Choose how this measure will be projected when aggregated:'. The 'Function' dropdown is set to 'Sum'. There is a checkbox for 'Associate a List of Values' which is unchecked. Below it, the 'List Name' field contains 'SALES006'. There are two checkboxes: 'Allow users to edit this list of values' (checked) and 'Automatic refresh before use' (unchecked). At the bottom, there are buttons for 'Restore Default', 'Edit...', 'Display...', 'OK', 'Cancel', 'Apply', and 'Help'.

- All measure objects should use a SQL aggregate function (e.g., SUM(), COUNT(), etc.).
- Measure objects must have the appropriate aggregate function for report-level aggregation defined in the object's properties tab. The SUM function is usually most appropriate.

Information Delivery – Universe Design Best Practices

Universe Name: Ticket Life Cycle
Subject Name: Lift

Universe Query Model

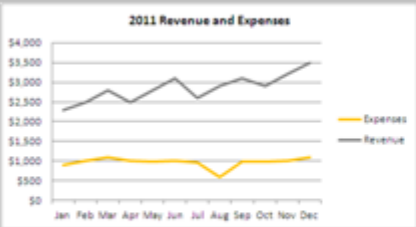
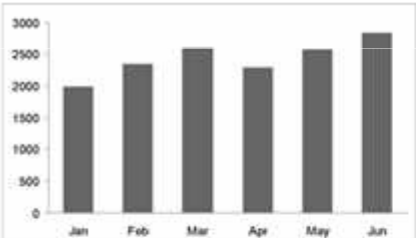



Query Format = Measure by Dim 1 and Dim 2 and ... Dim N

Information Delivery – Reports / Dashboards

- There are eight different information types:
 - Time Series
 - Ranking
 - Part-to-Whole in Percent
 - Nominal Comparison
 - Correlation
 - Frequency Distribution
 - Deviation
 - Maps
- Start by identifying which information type the data represents.
- Each information type has only two or three chart types that are appropriate for optimal information display.
- Review the possible chart types and select the most appropriate one.

Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions
Time Series Values display how a measure changed through time (yearly, monthly, etc.)	Line Chart to emphasize overall pattern		<ul style="list-style-type: none"> • Always place time on the horizontal axis. • Start the Y-axis at zero. • A combined line-vertical bar chart is useful when two measures need to be displayed. The line is associated with the left y-axis. The vertical bars are associated with the right y-axis or vice versa.
	Bar Chart to emphasize individual values		
	Line Chart with points to slightly emphasize individual values while still highlighting the overall pattern		


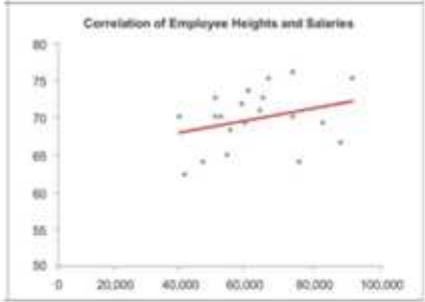

Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions
Ranking Values are ordered by size (either descending or ascending)	Horizontal Bar Chart		<ul style="list-style-type: none"> Sort in descending order to highlight high values. Sort in ascending order to highlight low values. Start the quantitative axis at zero.
	Vertical Bar Chart		

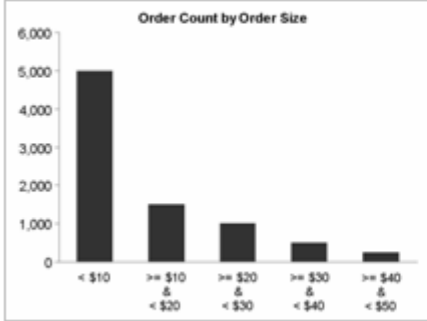
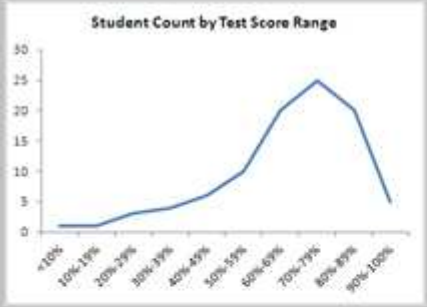
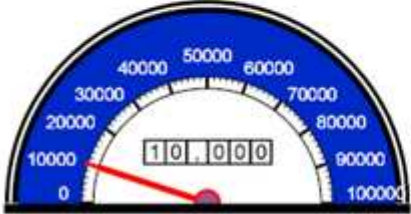
Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions										
<p>Part-to-Whole in Percent Values represent parts (ratios) of a whole</p>	<p>Vertical Bar Chart</p>	<p>2011 Regional % of Total Expenses</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>East</td> <td>40%</td> </tr> <tr> <td>West</td> <td>30%</td> </tr> <tr> <td>North</td> <td>20%</td> </tr> <tr> <td>South</td> <td>10%</td> </tr> </tbody> </table>	Region	Percentage	East	40%	West	30%	North	20%	South	10%	<ul style="list-style-type: none"> • Use stacked bars only when you must display measures of the whole as well as the parts. • Use percent values and not actual amounts in pie charts. • Make sure all percent values are visible in a pie chart. • Start the quantitative axis at 0%.
Region	Percentage												
East	40%												
West	30%												
North	20%												
South	10%												
<p>Horizontal Bar Chart</p>		<p>2011 Regional % of Total Expenses</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>East</td> <td>40%</td> </tr> <tr> <td>West</td> <td>30%</td> </tr> <tr> <td>North</td> <td>20%</td> </tr> <tr> <td>South</td> <td>10%</td> </tr> </tbody> </table>	Region	Percentage	East	40%	West	30%	North	20%	South	10%	
Region	Percentage												
East	40%												
West	30%												
North	20%												
South	10%												
<p>Pie Chart</p>		<p>2011 Regional % of Total Expenses</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>East</td> <td>40%</td> </tr> <tr> <td>West</td> <td>30%</td> </tr> <tr> <td>North</td> <td>20%</td> </tr> <tr> <td>South</td> <td>10%</td> </tr> </tbody> </table>	Region	Percentage	East	40%	West	30%	North	20%	South	10%	<p>Note: Pie charts should only be utilized to depict percentage values.</p>
Region	Percentage												
East	40%												
West	30%												
North	20%												
South	10%												

Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions										
<p>Nominal Comparison A simple comparison of <u>values</u> for a set of unordered items.</p>	Vertical Bar Chart	 <p>Q1 2003 Calls by Region</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Calls</th> </tr> </thead> <tbody> <tr> <td>North</td> <td>3,000</td> </tr> <tr> <td>East</td> <td>4,500</td> </tr> <tr> <td>South</td> <td>2,200</td> </tr> <tr> <td>West</td> <td>4,800</td> </tr> </tbody> </table>	Region	Calls	North	3,000	East	4,500	South	2,200	West	4,800	<ul style="list-style-type: none"> Start the Y-axis at zero.
Region	Calls												
North	3,000												
East	4,500												
South	2,200												
West	4,800												
<p>Correlation Comparison of two paired <u>sets</u> of values (e.g., the heights and weights of several people) to determine if there is a relationship between them.</p>	Scatter Plot	 <p>Correlation of Employee Heights and Salaries</p>	<ul style="list-style-type: none"> Use for 2 quantitative values Points and a trend line in the form of a scatter plot. Gartner Magic Quadrant is an example. 										
	Radar Chart	 <p>Success Factors Radar</p>	<ul style="list-style-type: none"> Use for more than one quantitative value. Each quantitative value is plotted on a <u>radii</u> or spoke. 										

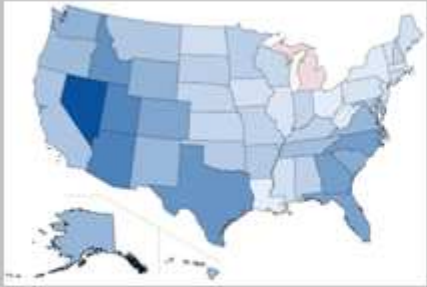


Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions
Frequency Distribution Counts of something per category range	Vertical Bar Chart to emphasize individual values.		<ul style="list-style-type: none"> • Good for showing the “bell curve” effect. • Start the Y-axis at zero.
	Line Chart to emphasize the overall pattern.		
	Gauge (i.e., speedometer, barometer, thermometer)		<ul style="list-style-type: none"> • Use for real-time metrics that need to be to remain within a defined tolerance range.

Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions
<p>Deviation The difference between two sets of values (e.g., the variance between actual expenses and budgeted expenses)</p>	Vertical Bar Charts		<ul style="list-style-type: none"> Always include a reference line to compare the measures of deviation against.
	<p>Balanced Scorecard – once an organization establishes strategic goals to impact positive change, a balanced scorecard becomes the “report card” for tracking actual values to the strategic goal values.</p>		<ul style="list-style-type: none"> Always show goal, actual, and variance values. Have one KPI per row on the scorecard. Use trending arrows for month-to-month comparisons.

Information Delivery – Reports / Dashboards

Information Type/Description	Recommended Chart/Table	Example Chart/Table	Design Suggestions
Maps	Color-Intensity Map		<ul style="list-style-type: none">For Color-Intensity Maps, use the same color varying the shade based on the quantitative amount. Low values are coded with lighter shades. High values are coded with darker shades.
	Size Map		<ul style="list-style-type: none">For Size Maps, use circles that vary in size to encode differences in value—the larger the greater.
	Heat Map		

Information Delivery – Reports / Dashboards

- Use a chart instead of a table when:
 - The message is contained in the shape of the values.
 - The document is used to reveal relationships among multiple values.
- Use a table instead of a chart when:
 - The purpose is to look up individual values.
 - The purpose is to compare specific values.
 - Precise values are required.
 - The quantitative information to communicate involves more than one or two measures.

Information Delivery – Reports / Dashboards

- A common problem with tables, charts, and dashboards is the excessive presence of visual content that doesn't represent actual data.
- Less is more! Whenever quantitative information is presented, the data itself should stand out clearly, without distraction.
- Reduce the non-data ink. Examples of non-data ink are chart gridlines or chart borders. Reduce means to eliminate or make less visible (for instance, light gray rather than black).
- Enhance the data ink. An example of data ink is the chart line or chart columns.
- Avoid 3D charts. They require non-data ink and in most cases are difficult to read.
- Avoid pie charts. It is usually difficult to interpret angles.
- Generally include the value zero in your quantitative scale, and alert your readers when you do not include zero.
- Include no more than eight data sets in a single graph.

Questions

TIBER SOLUTIONS

Jim Hadley, President

Email: jhadley@tibersolutions.com

Phone: 703.593.2833