12 Smarter Steps
To Business Requirements

April 12, 2007

Christina Rouse
chris.rouse@silvertraininc.com
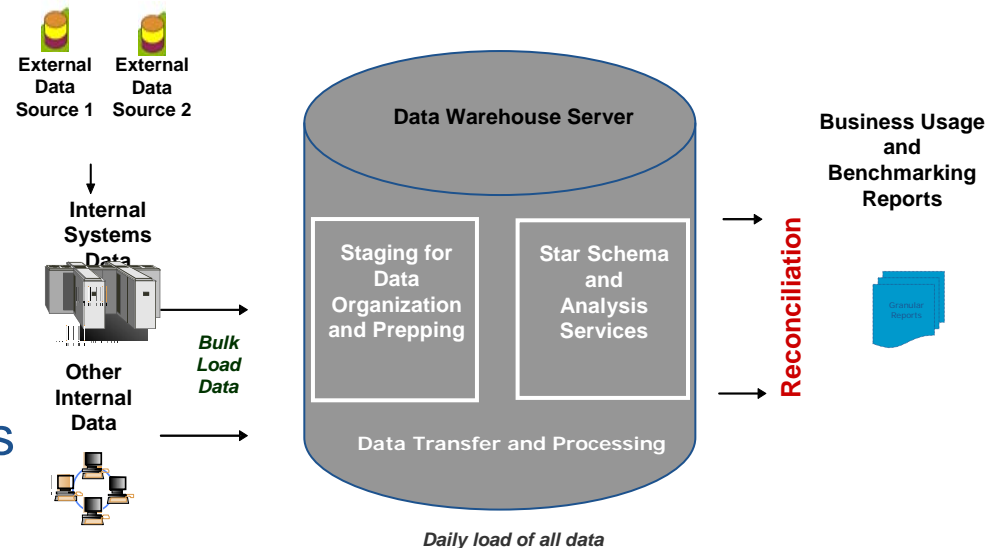www.silvertraininc.com

Why should we get smarter at business requirements?

We're very scientific at:

- Data movement
- Query performance
- Backup and recovery
- Cube compacting
- Multi-source environments
- Query governing
- Cube aggregations
- Re-use of cache

External Data Source 1 External Data Source 2

Internal Systems Data

Other Internal Data

Bulk Load Data

Data Warehouse Server

Staging for Data Organization and Prepping

Star Schema and Analysis Services

Data Transfer and Processing

Reconciliation

Business Usage and Benchmarking Reports

Granular Reports

Daily load of all data

But BOBSP isn't working anymore!

SILVERTRAIN

12 Steps To Smarter Business Requirements

1. Ask good Business <u>Use</u> Questions and document the answers

2. Complete a Source System Inventory

3. Prioritize the Source System Inventory

4. Complete the Data Element Inventory

5. Prioritize the Data Element Inventory

6. Write Data Behavior Statements

7. Identify Data Element Thresholds

8. Conduct a Data Element 'Sweep'

9. Determine the Authoritative Source for data element

10. Document the Data Provisioning

11. Determine Data Granularity and Aggregation

12. Present the Business Information Model to business users and present the physical data model to technologists

1. Business Use Questions

Ask questions that seek 'dimensionalized' answers! Document!

- What are your key business objectives, drivers and metrics and how do you currently measure your success?
- What are the key business issues you face today and what inhibits you from achieving your goals?
- What opportunities exist to dramatically improve your business management based on improved access to information? What's the financial impact?
- What information is requested from your customers and suppliers?
- What key information is required to make or support the decisions you make in the process of achieving your goals and overcoming obstacles? How do you get this information today?
- How do you identify problems and exceptions or know you're headed for trouble?
- What are your business success metrics? How do you know you're doing well? How often do you measure key success factors?

SILVERTRAIN

1. Business Use Questions
An Example

Question: Is there information, which is not available to you today, that would have significant impact on your goal achievement?

Answer:

- *Complete and fully integrated Canadian sales revenue for all 42 product categories across 8 sales districts for the past 3 years – updated daily.*

- *Raw material detail costs to each component in the product. Need standard cost, actual cost per lot and cost differential on daily basis.*

- *Customer hierarchy/rollups from customer viewpoint and the sales organization viewpoint for all active customers over past 10 years.*

- *Need customer segmentation information on 7 market spaces, 12 demand segments and 20 value segments.*

- *Customer trends and growth rates in sales revenue, gross margins, order frequency and order lot size by month for past 10 years.*

- *Need to know how current financial calculations are derived.*

SILVERTRAIN

1. Business Use Questions
Developing Dimensions from Notes

Please take a few minutes to review the example interview notes. As a group, we will draft the dimensions, facts and hierarchies.

Business User Findings

Interviewed:	Mary Barron	General Manager, A&HC Division
	Karen Konnelly	Analyst, A&HC Division
Interviewers:	Robert Pillar	Manager of Customer and Administrative Assistants
	Brad Barrett	Analyst & Finance Systems Lead
	Frank Lemons	Project Manager
	Stephen Stratton	Senior Consultant
Interview Date:	3/16/2006 at 9:00am	

A. What are your key business objectives, drivers and metrics?

Key Business Objectives
Grow key product lines
1. Gain targeted market share
2. Reduce costs via improved quality controls
3. Develop new and improved products
4. Elevate safety concerns

Metrics
1. Daily sales
2. Throughput (TP), total and per unit
3. Net income (NI)
4. Economic Value Added (EVA)
5. Returns and allowances (R&A)
6. Quality Index
7. Defects per Unit (DPU)

B. What are the key business issues you face today (i.e. What prevents you from meeting your business objectives)? What is the impact on the organization?
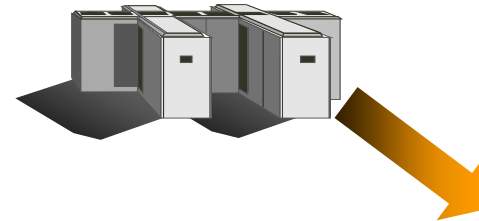
1. Over capacity in the industry.
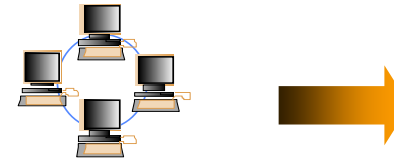
SILVERTRAIN

2. Source System Inventory

Don't forget the desktop sources, too! Document!

- Technical platform
- Application and business use
- Years of history
- Database size
- Update frequency
- Relational or not?
- Business owner(s)
- Technical owner(s)
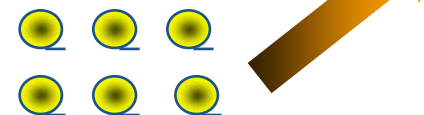- Expected life
- Percentage of total

Operational Systems Data

Operations Data

External Data

Don't worry about the data elements now!

SILVERTRAIN

2. Source System Inventory
An Example

Document source systems, owners, data volume and technical platform, etc.

Platform	Data Structure	Application & Business Function	Source Code	History	Row Counts	% of Total Data	Business Owner	Technical Owner
Mainframe	VSAM, IDMS, Non-relational	Parts BOM, Orders, Engine Configuration	Cullinet vendor package, 9 modules	1970, 30+ years	2.5 million	40%	Joe Smith	Jane Doe
Oracle	Normalized & Relational	Tables to support web applications for order tracking, "reports" and some	Home grown	1982 forward	2.5 million	30%	Dave Brown	Eli Whitaker
Lotus Notes	Lotus flat file	Work flow tracking and Engineering Change Notifications (ECN)	Home grown	2006, 6 years	300,000 est	10%	Bob Potter	Paul Smith
FoxPro	DB3 Relational	Warranty Programs	Home grown	2000, 6 years	100,000 max est	5%	Dane Henry	David Borno
MS Access	Relational	QT Warranty Tracking	Home grown	2000, 6 years	200,000 est	15%	Mary Ellis	Unknown
Mini-Tab Statistical	Matrix tables	Product Quality uses for control charts and analysis	NA	NA	NA		Roger Ganner	Unknown
						100%		

SILVERTRAIN

3. Prioritize Source System Inventory
An Example

Data volume does not equal data business significance!

Platform	Data Structure	Application & Business Function	Source Code	History	Row Counts	% of Total Data	Priority	Business Owner	Technical Owner
Mainframe	VSAM, IDMS, Non-relational	Parts BOM, Orders, Engine Configuration	Cullinet vendor package, 9 modules	1970, 30+ years	2.5 million	40%	3	Joe Smith	Jane Doe
Oracle	Normalized & Relational	Tables to support web applications for order tracking, "reports" and some	Home grown	1982 forward	2.5 million	30%	2	Dave Brown	Eli Whitaker
Lotus Notes	Lotus flat file	Work flow tracking and Engineering Change Notifications (ECN)	Home grown	2006, 6 years	300,000 est	10%	1	Bob Potter	Paul Smith
FoxPro	DB3 Relational	Warranty Programs	Home grown	2000, 6 years	100,000 max est	5%	4	Dane Henry	David Borno
MS Access	Relational	QT Warranty Tracking	Home grown	2000, 6 years	200,000 est	15%	5	Mary Ellis	Unknown
Mini-Tab Statistical	Matrix tables	Product Quality uses for control charts and analysis	NA	NA	NA			Roger Ganner	Unknown
						100%			

Prioritize the data based on value to business users.
Ask business users to weight the significance of each source.

SILVERTRAIN

4. Data Element Inventory

Make an inventory of all the data elements that users require for business decisions. The list should be from the business user perspective.

- Business name for data element (the names on the screens)
- Application(s) were used
- Screen(s) were used
- Examples of data element values
- Desired display format like $#,###.## or '01-Product Code'
- Measure or 'Report Filter' (Dimension)
- Is the 'Report Filter' flat or hierarchical?
- Update frequency
- Create user data groups from the data elements

Don't worry about the data source mappings now!

SILVERTRAIN

4. Data Element Inventory
An Example

Document data elements from the business users perspective.

Ref	Data Group	Report Filter (Dimension or Measure)	List of Values or Measure Format	Hierarchical or Flat Filter or Measure	Source (Database or Business Rule)	Siebel Screen Tab(s)	Update Frequency

1	Opportunity	Sales Stage and Sales Phase	**EXTERNAL**
			    APPOINTMENT SET
			    OPPORTUNITY
			    MANAGEMENT ASS'T
			INTERNAL
			    PROCEED STOP
			    APPROVED
				H	Siebel	Opportunities	Daily

2	Opportunity	Opportunity Revenue Type	ACTUAL SHIPPED BILLED BOOKED PROJECTED QUOTA	F	Siebel	Opportunities	Daily
3	Opportunity	Opportunity Priority	YES NO	F	Siebel	Opportunities	Daily
4	Opportunity	System Generated	YES NO	F	Siebel	Contacts Opportunities	Daily
5	Opportunity	Opportunity Created Date	MM/DD/YYYY	F	Siebel	Opportunities	Daily
6	Opportunity	Opportunity Count	###,#	M	Calculated	NA	On demand

SILVERTRAIN

4. Data Element Inventory

Exercise Screenshot #1

SILVERTRAIN

4. Data Element Inventory

Exercise Screenshot #2

SILVERTRAIN

5. Prioritize Data Element Inventory
An Example

Prioritize on Data Group because:

1. Translate nicely to dimensions in the physical model

2. Determines any battle ground early

3. Reveals any hidden data elements

Ref	Data Group	Report Filter (Dimension or Measure)	List of Values or Measure Format	Hierarchical or Flat Filter or Measure	Source (Database or Business Rule)	Siebel Screen Tab(s)	Update Frequency
1	Opportunity	Sales Stage and Sales Phase	**EXTERNAL** APPOINTMENT SET OPPORTUNITY MANAGEMENT ASS'T **INTERNAL** PROCEED STOP APPROVED	H	Siebel	Opportunities	Daily
2	Opportunity	Opportunity Revenue Type	ACTUAL SHIPPED BILLED BOOKED PROJECTED QUOTA	F	Siebel	Opportunities	Daily
3	Opportunity	Opportunity Priority	YES NO	F	Siebel	Opportunities	Daily
4	Opportunity	System Generated	YES NO	F	Siebel	Contacts Opportunities	Daily
5	Opportunity	Opportunity Created Date	MM/DD/YYYY	F	Siebel	Opportunities	Daily
6	Opportunity	Opportunity Count	###,#	M	Calculated	NA	On demand

SILVERTRAIN

6. Data Behavior Statements

Write statements that describe <u>how</u> the data relate.

- Ask business about <u>how</u> one Data Group relates to another.
- Drill from top down on one Data Group.
- Think in genealogy terms! Grandparents, parents, children and twins.
- 'Report Filters' and measures both have behaviors.
- Ask business questions about extremes cases.
- Determine if the data element is nominal, ordinal, interval or ratio type.
- Determine if a data element is slowly changing.
- Don't forget to ask about history of data behaviors. Has the data element always behaved this way?

Data behavior statements help determine the date model dimensions!

SILVERTRAIN

6. Data Behavior Statements

Data Type Determination

NOIR

• *Nominal – Red, Yellow and Blue*
  Just a name with no order or magnitude

• *Ordinal – 1st, 2nd and 3rd*
  An order, but no magnitude

• *Interval – 1", 2" and 3"*
  An order, a magnitude but absolute zero

• *Ratio – -2°, -1°, 0°, +1°, +2° C*
  An order, a magnitude and below zero possibilities

Nominal and Ordinal data type elements are dimensions and Interval and Ratio data type elements are facts in the data model.

SILVERTRAIN

Exercise

Categorize each data element as Nominal, Ordinal, Interval or Ratio

- Revenue change month-over-month
- Sales districts North, South, East and West
- Product rank by gross revenue for first quarter 2006
- Product sales group and subgroup
- Top 10 best selling sales reps
- Total units of product sold
- Shipping weight per product
- Order type and order sub-type
- Gross Profit Year-To-Date
- Birth date of client
- Account opening date
- Frequent flyer points used

SILVERTRAIN

6. Data Behavior Statements
Slowly Changing Dimensions

- Slowly Changing Dimensions:
  - (Type One), doesn't preserve history
  - (Type Two), preserve a version of history
  - (Type Three), Hybrid of Type One and Two

- Users typically want a "Type Two" methodology of SCD
- A Type Two change writes a record with the new attribute information and preserves a record of the old dimensional data.
- Type Two changes let you preserve historical data.
- Implementing Type Two changes, after the fact, will require significant analysis and development.
- Type Two changes accurately partition history across time more effectively than other types.
- Because Type Two changes add records, they can significantly increase the database's size.

SILVERTRAIN

Diagram the 3 approaches to SCD here.

Data Group	Report Filter (Dimension or Measure)	List of Values or Measure Format

Opportunity	Sales Stage and Sales Phase	**EXTERNAL**
		        APPOINTMENT SET
		        OPPORTUNITY
		        MANAGEMENT ASS'T
		INTERNAL
		        PROCEED STOP
		        APPROVED

Exercise

Diagram the 3 approaches to SCD

On Date	Opp ID	Sales Stage	Sales Phase
1/1/2007	1	External	Appointment Set
1/5/2007	1	External	Management Ass't

On Date	Opp ID	Sales Stage	Sales Phase	Begin Date	End Date	Occurrence
1/1/2007	1	External	Appointment Set	1/1/2007		1
1/5/2007	1	External	Appointment Set	1/1/2007	1/4/2007	1
	1	External	Management Ass't	1/5/2007		2

On Date	Opp ID	Sales Stage	Sales Phase	Previous Sales Stage	Previous Sales Phase
1/1/2007	1	External	Appointment Set		
1/5/2007	1	External	Management Ass't	External	Appointment Set

SCD Exercise

Best Practice Design

1. Design the table to support Order Status changes over time.

2. Order Statuses are: 'Placed', 'Credit Check', 'Packing', 'Ship Hold', 'Shipped', 'Backordered' and 'Received'.

3. Order Status can revert! Order Status can repeat.

4. Order Status can change an unlimited number of times

5. An order can be in an Order Status for an unlimited period of time

6. User want to measure "days spent in each status"
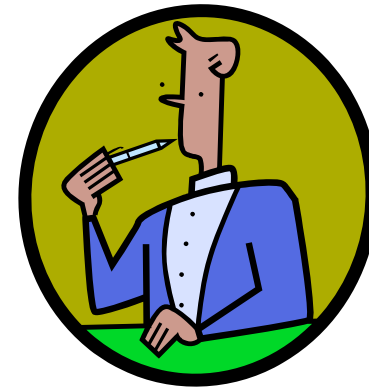
SCD Exercise
Best Practice Design

OrderID	Order_Status	Order_Status_Start_Date	Order_Status_End_Date	Status_Occurrence
1	Placed	1/1/2006	1/5/2006	1
1	Credit Check	1/6/2006	1/7/2006	2
1	Packing	1/8/2006	1/10/2006	3
1	Credit Check	1/11/2006	1/12/2006	4
1	Packing	1/13/2006		5
2	Placed	1/16/2006	1/17/2006	1
2	Shipped	1/18/2006		2
3	Placed	1/21/2006	1/22/2006	1
3	Credit Check	1/23/2006	1/25/2006	2
3	Ship Hold	1/26/2006	1/27/2006	3
3	Shipped	1/28/2006		4
4	Placed	3/1/2006		1
5	Placed	4/1/2006		1

Why have two dates?

Why a null value for Order_Status_End_Date?

Why have an occurrence counter?

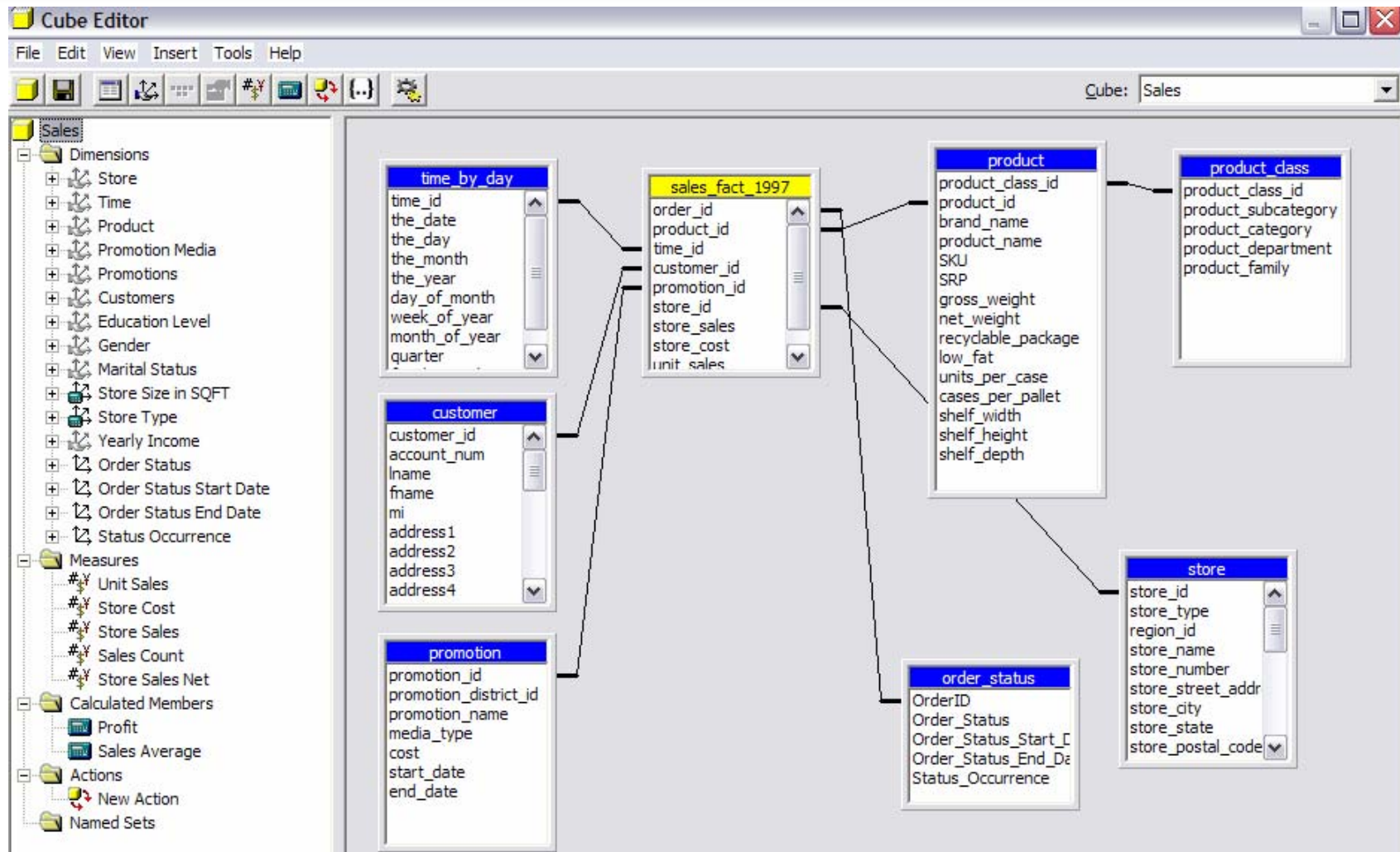SILVERTRAIN

SCD Exercise

Best Practice Design

SELECT *

FROM order_status

WHERE (((order_status.Order_Status_End_Date) Is Null));

A select MAX(status_occurrence) works on the entire column, not just on the multiple rows of each order ID.

SCD Exercise
Best Practice Design

SCD and Joins

Q. If a star schema has 3 dimensions and one fact:

- Product with 1000 rows

- Time with 365 rows

- Sales Org with 2000 rows

- Sales Order transactions fact table of 50 rows,

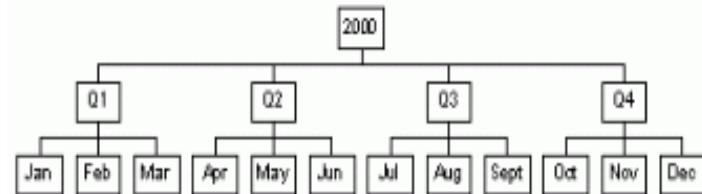How many rows are returned in a query of select * with all dimensions joined?

A. 50

SILVERTRAIN

6. Data Behavior Statements
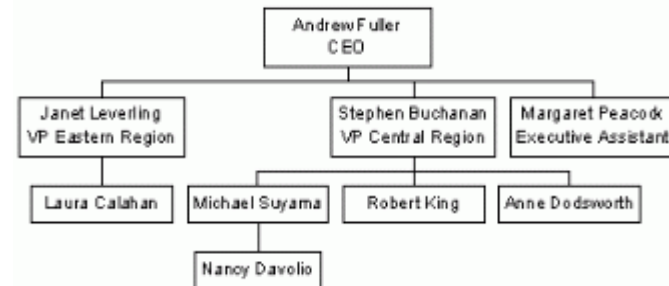## Organizational Hierarchy Types

Balanced Hierarchies

• All branches of the hierarchy descend to the same level, and each member's logical parent is the level immediately above the member.
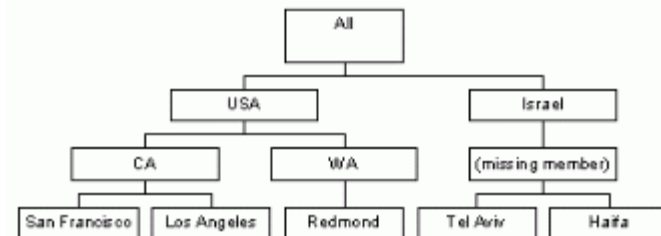

Unbalanced Hierarchies

• Branches of the hierarchy descend to different levels. For example, an Organization dimension contains a member for each employee in a company.

Ragged Hierarchies

• The logical parent member of at least one member is not in the level immediately above the member. This can cause branches of the hierarchy to descend to different levels.

SILVERTRAIN

Hierarchy Types, Table Impact

Unbalanced and Ragged Hierarchies will impact the data model columns and the OLAP tool!

Organization Table

	COLUMN 1	COLUMN 2	COLUMN 3	COLUMN 4	COLUMN 5	COLUMN 6	COLUMN 7	COLUMN 8
Balanced	X	X	X	X	X	X	X	X
	X	X	X	X	X	X	X	X
	X	X	X	X	X	X	X	X
Unbalanced	X	X		X	X		X	X
	X	X	X		X		X	X
	X	X	X		X	X	X	X
Ragged	X	X	X	X	X			
	X	X	X	X	X	X		
	X	X	X					

SILVERTRAIN

Org Hierarchy Type Exercise
Record all the organizational hierarchies

Hierarcy Name	Hierarchy Owner	Level 1 Name (Top)	Level 2 Name	Level 3 Name	Level 4 Name	Level 5 Name (Bottom)
Accounting	CFO	CEO Approver	VP Approver	Director Approver	Manager Approver	Supervisor Approver
Sales	VP Sales	VP Sales	Sales Region		Sales District	Sales Rep
Human Resources	VP HR	VP Labor	Plant General Manager	Department Foreman	Job Supervisor	

Data in Table 'ORG_HIER' in 'BI_DEMO_DATA' on '(local)'

Hierarchy_Name	Employee	Level 1 Name (Top)	Level 2 Name	Level 3 Name	Level 4 Name	Level 5 Name (Bott
Accounting	Joe	Bob	Tony	Bill	Steve	Sue
Sales	Joe	Ray	Allen	Jim	Ann	<NULL>
Human Resources	Joe	Tim	Ann	Bill	Steve	<NULL>
Accounting	Sam	Bob	Tony	Bill	Steve	Sue
Sales	Sam	Ray	Alen	Jim	Ann	<NULL>
Human Resources	Sam	Tim	Ann	Bill	Steve	<NULL>
Accounting	Fred	Bob	Tony	Bill	Steve	Sue
Sales	Fred	Ray	Allen	Jim	Ann	<NULL>
Human Resources	Fred	Tim	Ann	Bill	Steve	<NULL>
Accounting	Bob	Bob	<NULL>	<NULL>	<NULL>	<NULL>
Sales	Ray	<NULL>	<NULL>	<NULL>	<NULL>	<NULL>
Human Resources	Tim	<NULL>	<NULL>	<NULL>	<NULL>	<NULL>

SILVERTRAIN

Org Hierarchy Type Exercise
Record all the organizational hierarchies

Org Hierarchy Type Exercise
Record all the organizational hierarchies

Hierarcy Name	Hierarchy Owner	Level 1 Name (Top)	Level 2 Name	Level 3 Name	Level 4 Name	Level 5 Name (Bottom)
Accounting	CFO	CEO Approver	VP Approver	Director Approver	Manager Approver	Supervisor Approver

SILVERTRAIN
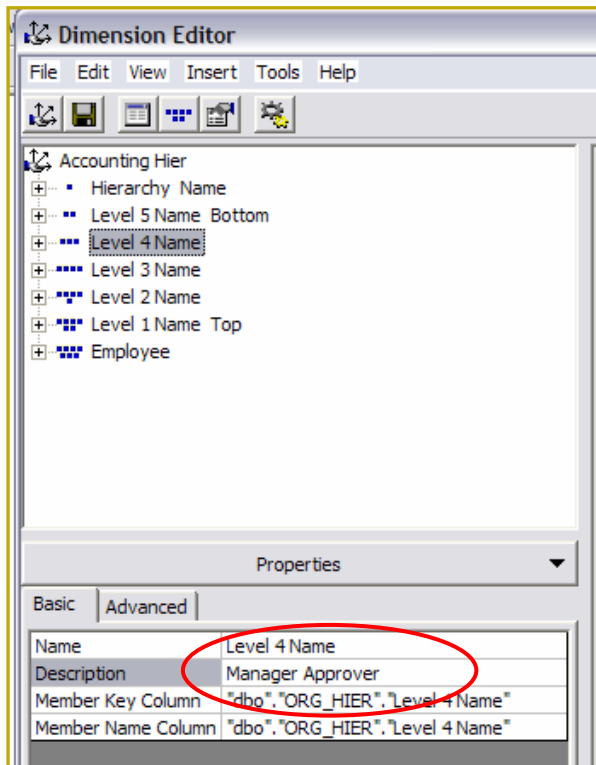
A table can translate three ways to the semantic layer of the OLAP tool:

- *One to One* - The table column translates to one object in the semantic layer.
- *Temporal Hierarchies* - A hierarchy is generated from the rows within a column, this only happens in a reporting time dimension.
- *Traditional Hierarchies* - Multiple table columns like, Sales Organization and Department and Unit, translate into 'Sales Org Hierarchy' .

C l o i s t e r e d Data?

The Primary and Foreign Keys of the physical tables are never seen within the semantic layer, and are never seen by the report consumers. Only the business keys like Customer Name are viewable the report consumers

SILVERTRAIN

Exercise
Demonstrate data element to semantic layer

OPPORTUNITY_DIM

OPPORTUNITY_KEY: float
OPPORTUNITY_NAME: float
OPPORTUNITY_PARENT_OPPORTUNITY: nvarchar(255)
OPPORTUNITY_LEAD_QUALITY: nvarchar(255)
OPPORTUNITY_REVENUE_TYPE: nvarchar(255)
OPPORTUNITY_START_DATE: smalldatetime
OPPORTUNITY_CLOSE_DATE: smalldatetime
OPPORTUNITY_END_DATE: nvarchar(255)
OPPORTUNITY_PRIORITY: nvarchar(255)
OPPORTUNITY_SOURCE: nvarchar(255)
OPPORTUNITY_CONTACT_BUSINESS_ID: nvarchar(255)
OPPORTUNITY_AOM_DISPOSITION: nvarchar(255)
OPPORTUNITY_CREATED_BY: nvarchar(255)
OPPORTUNITY_PARENT_CAMPAIGN: nvarchar(255)
OPPORTUNITY_CAMPAIGN: nvarchar(255)
OPPORTUNITY_STATUS: nvarchar(255)
OPPORTUNITY_STATUS_REASON: nvarchar(255)

OPPORTUNITY_PRIVACY: nvarchar(255)
OPPORTUNITY_OCCURANCE: float
OPPORTUNITY_STATUS_SORT: int

- Opp Name
- Lead Quality
- Revenue Type
- Opp Priority
- Opp Source
- AOM Disposition
- Logically derive opportunity status at month end by using the OPPORTUNITY_CLOSE_DATE.

SILVERTRAIN

6. Data Behavior Statements
An Example

Orders

- *Order number is unique within System A and within System B but not across System A and System B.*

- *Order number format changed in 1985 to 12 numbers from 9 numbers and letters.*

- *Orders have 1 to 999 maximum order lines in System A and 1 to 99 in System B.*

- *An order line contains only 1 product in System A. An order line in System B can have 1 to 9 products.*

- *Order line status in both systems is: 'Placed', 'Credit Check', 'Packing', 'Ship Hold', 'Shipped', 'Backordered' and 'Received'.*

- *'Backordered' order lines have a status that is updated daily.*

SILVERTRAIN

7. Threshold Identification

Ask questions that determine if a decision or action threshold exists.

- Observe a manager reviewing the daily reports and ask why she takes or does not take certain actions.
- Ask what prompts a phone call from the 'C-Level' people
- In sales, ask how the commission plan is determined.
- Review the monthly and quarterly reports and read the comments. Ask how these comments were derived. There's usually a buried threshold.
- Thresholds help users focus on what's important and what's not important.

DISCOVER THE POWER OF A LOYAL CUSTOMER

Product Dashboard

Year To Date Actuals, Budget and Forecast: March   2002

Carline	Units	Gross Sales	Net Sales	Gross Profit	Net Profit	Budget	Forecast	Net To Budget	Net To Forecast	Gross Margin	Net Margin
Product A	209,934	$26,473,479	$26,042,794	$942,362	$760,820	$23,720,338	$17,217,857	109.8%	151.3%	3.6%	2.9%
Product 123	402,808	$23,060,430	$22,657,694	$3,334,152	$3,186,012	$22,505,214	$16,348,810	100.7%	138.6%	14.5%	14.1%
Product Zx5RD	124,933	$14,814,863	$14,346,794	$843,345	$650,362	$16,461,386	$11,470,000	87.2%	125.1%	5.7%	4.5%
Product 678	58,851	$6,778,812	$6,466,643	$880,879	$727,962	$8,394,641	$5,408,333	77.0%	119.6%	13.0%	11.3%
Product RTFD3	4,234	$394,554	$384,031	$37,612	$36,055	$288,944	$216,887	132.9%	177.1%	9.5%	9.4%
Product Infinity	3,480	$389,854	$379,792	$45,626	$42,163	$414,848	$319,411	91.5%	118.9%	11.7%	11.1%
Product P-9012	5,042	$352,139	$340,913	$33,191	$27,432	$1,754,973	$341,775	19.4%	99.7%	9.4%	8.0%
Product Team-X	2,192	$334,199	$323,790	$68,174	$60,928					20.4%	18.8%
Product T-99878	28	$2,107	$2,107	$225	$225					10.7%	10.7%
Product RTF123	3	$569	$569	$219	$219					38.5%	38.5%
	8	$188	-$4	-$560	-$560					-297.7%	12557.6%
Total	985,693	$88,097,980	$85,920,556	$7,366,951	$6,574,274	$91,888,268	$63,538,555	93.5%	135.2%	8.4%	7.7%

Thresholds   >= 0.99   Between 0.9 - 0.99   Below 0.9

Thresholds can be adjusted! Build the idea of rapid and personalized threshold change into the business intelligence solution.

SILVERTRAIN

8. Data 'Sweep'

A Data 'Sweep' is all about up selling your customers on data!

- Identify the source system tables for the priority Data Groups from the Data Element Inventory.

- Calculate a column usage from the Data Element Inventory. The ORDER table in System A has 30 columns and business users requested 21 of these columns or 70%. Odds are they'll eventually want the remaining 9 columns.

- Offer the 7 columns as an up sell. It's a good way to cross check your Data Element Inventory.

Sweeping now is cheaper and easier than sweeping in the future! Don't sweep unused data under the rug!

SILVERTRAIN

9. Authoritative Source

**Understand validity and reliability. Map the data trail!**

Important!

- Validity – A data source value is true and is what it says it is. If a data element source is valid it is also reliable.
- Reliability – A data source value repeatedly returns the same value which may or may not be valid. A reliable data source is not necessarily a valid data source.

- Trace the data element from the business users decision point back to its origin (Yes, all the way back!)
- Calculate the validity of the data at each transfer point.
- Capture the values at each point and determine, with business users, the valid point – not the reliable point.

SILVERTRAIN

9. Authoritative Source
An Example

1. Trace unique customer IDs from any point of origination to any destination.
2. Calculate how many are correctly moved from system to system.
3. Discuss the results with business users.
4. Understand the errors.
5. Determine the valid source.
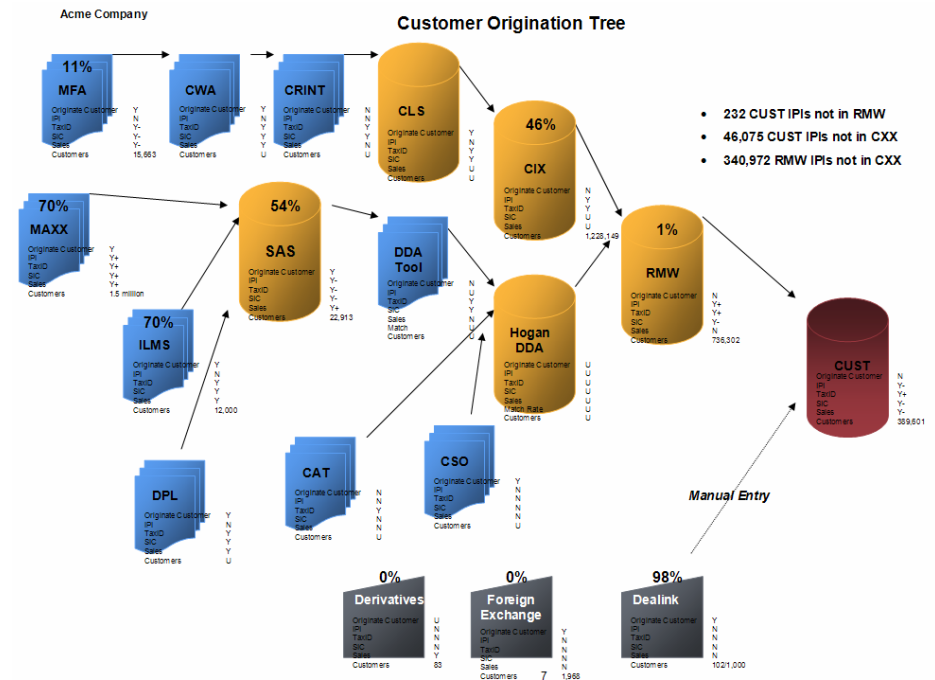
The Authoritative Source is most likely the same for all Data Elements within a Data Group.

SILVERTRAIN

10. Data Provisioning

Provision <u>after</u> you determine authoritative source!

- Once the authoritative source for each Data Group is determined, begin the individual data element mappings.
- Obtain a source system expert from business and IT.
- Charge the team with Data Group mappings.
- Divide and conquer the data element mappings. Use 3-4 two-person teams.
- Make sure you note any transformations needed.
- If a Data Element is derived make sure you note the exact calculation or logic. Provide an example whenever possible.
- Map the sources against the Data Element Inventory, don't start a new document. Your business users are accustomed to this document now.

Data Provisioning is not an IT only exercise! You need the business users involved!

SILVERTRAIN

11. Data Granularity and Aggregation

BI solution granularity may be different from the source system!

- What are the source system data update frequencies?
- What are the business requirements for the currency of the data - daily, weekly?
- Are monthly, quarterly or yearly aggregates needed?
- Will forecasting be conducted?

Rules of Thumb
1. **Store atomic level data and you will never disappointment your users.**
2. **Partition the database on time, especially if you carry many years of data. When you drop the oldest year, you drop a single partition.**
3. **Aggregate on the same dimensions as the partitions. Commonly time, sales organization and product. You may add other aggregates.**
4. **For forecasting, keep 1.5 as many years in history as you want to forecast. Example: 3 year forecast needs 4.5 years of history.**

SILVERTRAIN

11. Data Granularity and Aggregation

Database Size Estimate

- **Identify the dimensions from the Data Groups**
- **Know the source system row counts and table width**
- **Identify a growth factor**
- **Know the database vendor (Oracle, SQL, etc.)**
- **Identify an index growth**
- **Estimate the database size**
- **Estimate a daily net change volume**

Fact Table/Dimension Table Sizing Estimates
This is a estimating tool. The year parameter has been set to 1.

Assumption for calculations:			
Clients	13,000,000	Opportunities per Client	4
Contacts	2,000,000	Activities per Opportunity	2
"Active" Clients	3,600,000	Referrals per RM	3
Total # of Officers	11,000	Years of history	1
# of RMs / FAs	7,000		

TABLE	DIM/FACT	ROWS	GROWS WITH	MAX BYTES PER ROW	SIZE (IN GB)
ATOMIC LEVEL TABLES					
ACTIVITY	DIM	28,800,000	new activities	1300	34.87
CONTACTS	DIM	15,000,000	new clients/contacts	575	8.03
OPPORTUNITY	DIM	14,400,000	new opportunities	1500	20.12
REFERRALS	DIM	5,670,000	new referrals	400	2.11
TIME	DIM	1,825	static	100	0.00
SALES-ORG	DIM	11,000	static	500	0.01
SUMMATED STAR SCHEMA TABLES (AGGREGATES)					
ACTIVITY-ATTRIB	DIM (MV)	2,880,000	--	1000	2.68
CONTACTS-ATTRIB	DIM (MV)	360,000	--	500	0.17
OPPORTUNITY-ATTRIB	DIM (MV)	1,440,000	--	1000	1.34
REFERRALS-ATTRIB	DIM (MV)	567,000	--	350	0.18
RM-OPP-ACTIVITY-FACT	FACT (MV)		--		0.00
RM-REFERRALS-FACT	FACT (MV)		--		0.00
RM-GOALS-FACT	FACT (MV)		--		0.00
SALES-PROCESS-FACT	FACT (MV)		--		0.00
				TOTAL ESTIMATE SIZE (IN GB)	**69.51**

Daily transmission time can be estimated using 2GB/minute peak-time and 4GB/minute off-peak. A 40GB data stream will take 20 minutes to transfer during peak-time use.

SILVERTRAIN

12. Data Models

Public Speaking 101 – Know Your Audience! Show the business model to the business users and the physical model to IT.
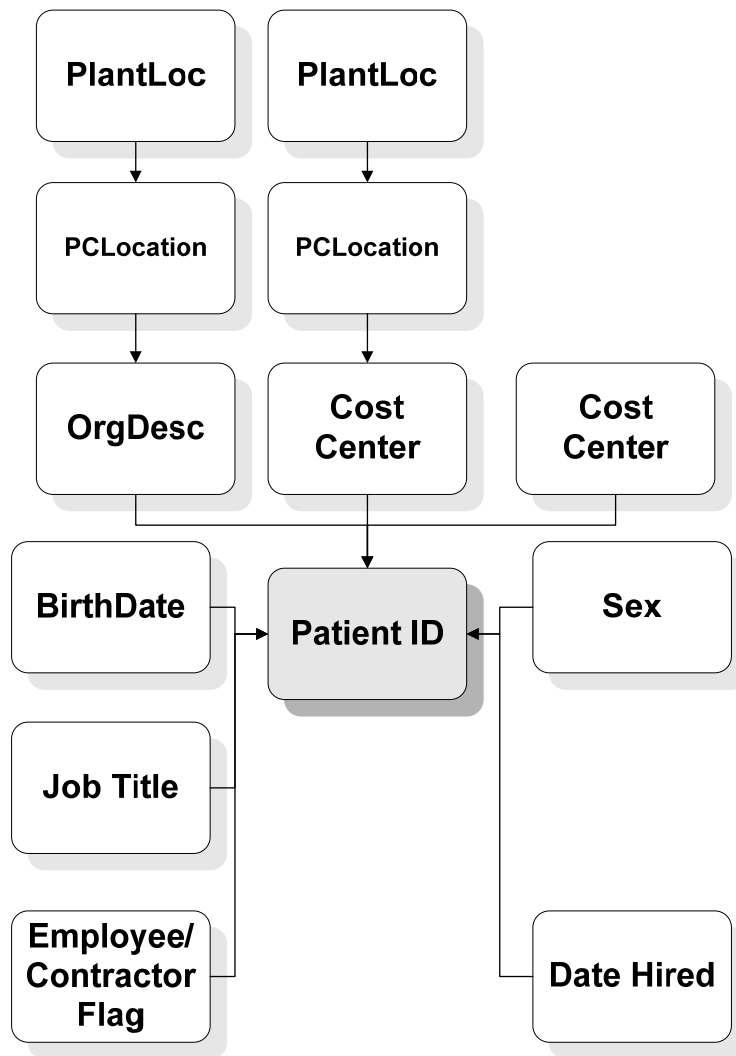
Business Users

- **No cloistered data is shown!**
- **Business keys only.**
- **Discuss relationships, not joins.**
- **Discuss 'range of values' not cardinality.**
- **Speak in measures not facts.**
- **Use summary rather than 'agg'.**
- **Show the business picture!**

IT Technologists

- **Know your surrogate keys.**
- **Know your business key uniqueness.**
- **Strive for joins from fact to dimension only. Avoid dimension to dimension joins.**
- **Know the difference between cumulative facts and non-cumulative facts. Where will you calculate each?**
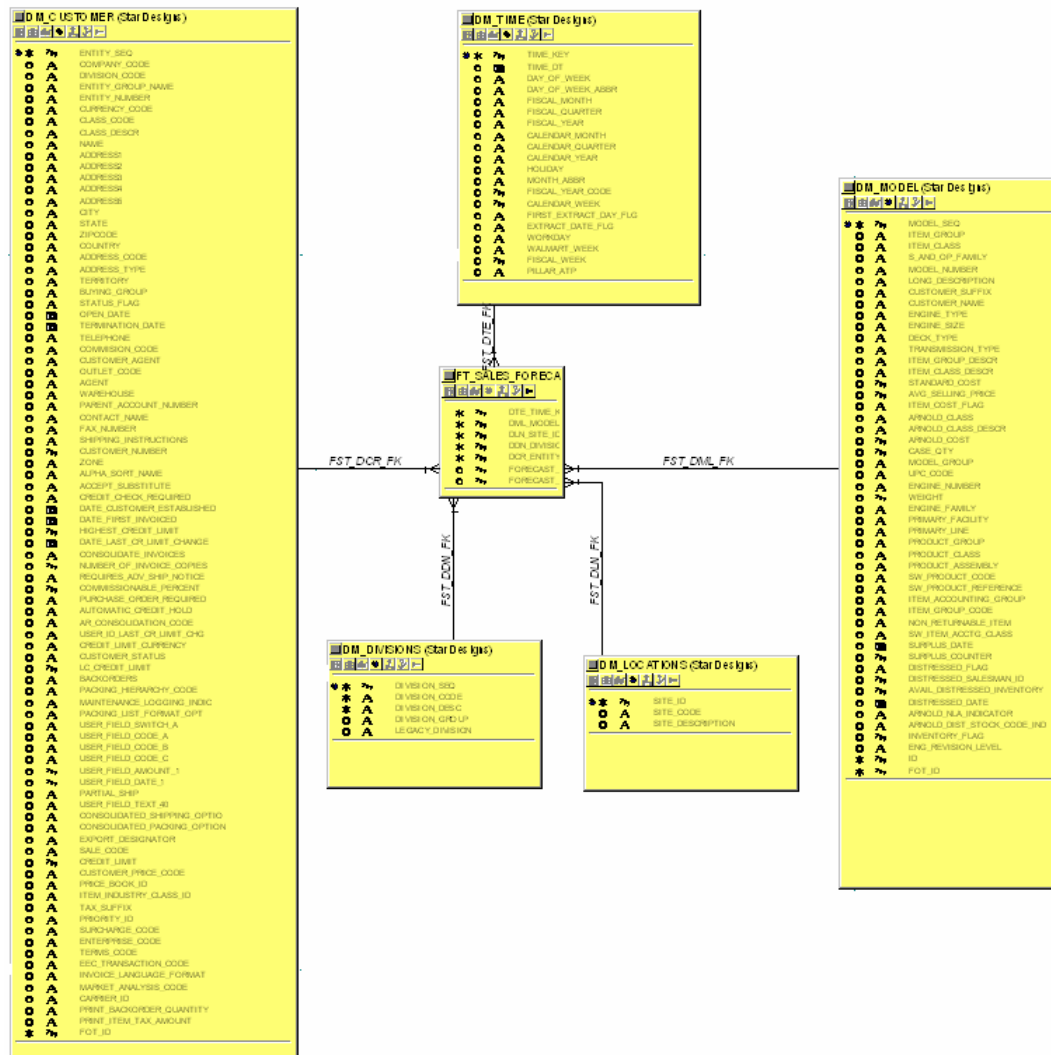- **Keep your physical model current.**

An Example for Business Users

- Business users read top-down and left-right. Present the model this way.
- Identify the primary data element (Patient ID)
- Use business data element names
- Present one dimension (Data Group) at a time

12. Data Models

An Example for Technologists

- Use a good modeling tool.
- Adopt a naming scheme for databases, tables and columns.
- Use surrogate keys.
- Include the joins.
- Keep model current.

12 Steps To Smarter Business Requirements

1. Ask good Business <u>Use</u> Questions and document the answers
2. Complete a Source System Inventory
3. Prioritize the Source System Inventory
4. Complete the Data Element Inventory
5. Prioritize the Data Element Inventory
6. Write Data Behavior Statements
7. Identify Data Element Thresholds
8. Conduct a Data Element 'Sweep'
9. Determine the Authoritative Source for data element
10. Document the Data Provisioning
11. Determine Data Granularity and Aggregation
12. Present the Business Information Model to business users and present the physical data model to technologists

Thank You!