Project Background and Overview

I Caveats

A. Disclaimer

The class project is based on a fictitious example company. All of the names, relationships, schema, and data are completely unreal and not based upon any existing company or organization. Any similarities between actual businesses or organizations (active or inactive) are purely coincidental. Names mentioned in the project are assumed to be mythical and should not be interpreted as referring to any existing companies, people, products, trademarks, etc.

B. Overview

In Milestone 1, you did (or at least should have done) the necessary preparatory activities to work as a team on the project. In Milestone 2, you will use this document in conjunction with the OLTP data models (provided to your team by the instructor after M2 is assigned, but NOT until AFTER significant progress has been made on M2) for the Source Systems to design a Star Schema with one fact table (Sales facts). In Milestone 3, your team will your team will receive a bundle of files (from the instructor) to load the OLTP models. Once these have been loaded (using simple set of batch files), your team can implement ETL to load Dimension data for a Star Schema (using a Star Schema data model and the "raw data" for the Source Systems provided by the instructor). This Star Schema will be stored in a ROLAP database for a Data Mart but you will not be required to load all the Dimension data, any of the fact data or any of the aggregates in Milestone 3. In Milestone 4, your team will receive another bundle of files (from the instructor) to completely load a Data Mart into the MS SQL Analysis Services. Your team will then create several reports running against this fully populated Data Mart.

Whenever data (or information) is loaded as part of a Milestone, you should examine the output (of any scripts / tools used) for errors and completion status. You should also perform some simple SQL queries to verify that there were no errors. For example, verify that there were no errors in the initial raw data loaded into the Source Systems and verify that there were no errors in the ETL that you implemented. Obviously, any errors (whether caused by idiosyncratic system environment details or implementation errors) in loading data or information can result in errors for the Downstream Customers. **If there are errors, you must resolve them!**

C. Consistency versus Verisimilitude

Although the information provided here (and elsewhere) as background for the project is intended to provide a "flavor of reality" to the project, it is not expected to be flawless or a complete substitute for reality. While it is intended to hold up to some scrutiny, if (and when) the sense of reality starts to break down you should **not** simply "ignore it". In other words, in a "real world situation" each team would have access to a real enterprise populated with real people who understood the business. Of course, the team members might possibly be some of these knowledgeable people themselves in the real world. However, since we do not have that situation here, when confusion and ambiguity arise, each team should direct their questions to the instructor using the per-team forums. The instructor will take the place of "the enterprise" and answer any issues raised.

The project is intended to be self-consistent and should make sense within the confines of its own data. In other words, you should accept the raw numbers provided in the data **even though** they seem "too high" or "too low" based upon current "real world situations in the RV industry", like the actual inflation rate, actual tax rates, actual prices, actual costs, etc. Once we accept the raw numbers, all the analysis should be fairly consistent. Again, if and when there is any confusion or concerns **ASK!**

D. Use This Example DO NOT Create a "Better Example"

For our purposes, I have included all you need to know about the organization and enterprise. Even though this is an academic project (**NOT A REAL** project), we can attempt to make the best project possible within the constraints and confines of the class. In other words, use this document and any other documents and data models provided by the instructor in addition to any questions, answers, discussions, or tips provided by the instructor to complete the project milestones. But **DO NOT** include any details based upon your imagination or real world experience with RV sales, dealerships, etc. While I would appreciate any suggestions to make the project better, you should **not change or augment** the descriptions, models, or documents I provided! If your project includes details with no basis in "the project's reality", points will be lost!

E. Some Example Scenarios

For example, while we may not believe a real customer would pay $285,000 for a particular RRV Make and Model in 2012, we can safely ignore that if those values are in the sample data provided. The Data provided for the OLTP data models is of course fictitious. This means that if we compared it to the real world and what a real RV costs, sells for, etc. the numbers in the project data would most likely feel fake – because they are. Having said that, there is an internal consistency to the numbers that will remain intact for all phases of the project.

1. When Sales Numbers Don't Make Sense

Suppose we run a report on our sales cube and determine we sold 2,000,000 vehicles (units) in 2012 and also report that we have only 300 customers and 10 dealerships "enterprise –wide" something is obviously wrong here. This would mean each customer bought approximately 6,667 RV units from us that year! It would also imply that each dealership **averaged** about 550 sales PER DAY (That's 7-days a week; we average 770 per dealership per day if we only have a 5-day week!) Obviously, we can't average more sales per day than the total number of customers! Clearly, something is WRONG with the data or the analysis in this situation!
If something like this were to happen for your team, you would need to address it!

2. When Production Numbers Don't Make Sense

Similarly, if our reports showed that we had only five (5) plants with a total of 350 employees but manufactured 1,000,000 vehicles in 2012 something is not quite right. That would mean on average each plant manufactured about 550 vehicles per day (which might be possible, but not when that means that on average each employee makes 7 complete vehicles per day, seven days a week all by themselves!) Clearly, something is WRONG with this situation!
If something like this were to happen for your team, you would need to address it!

3. When Sales Amounts Do Make Sense

Suppose the data set claims that we have only 500 customers for one year and that we had $100,000,000 total sales for that year – perhaps a real world business would not expect these numbers – we can ignore that here. The point is that if the total number of customers was 500 and the total sales was $100,000,000 then the average sales amount for that year per customer MUST be $200,000. Even if this is not the average MSRP, that seems fine unless ALL the vehicles are actually sold for less than $200,000 apiece!

4. When Other Analysis Does Not Make Sense

Suppose we had that same total sales amount and the same total number of customers as the previous example scenario, but also had 500 dealerships. This would mean an average of only one sale (and only one customer) per dealership per year! That does NOT sound internally consistent! While 30 sales per dealership or 300 sales per dealership might be acceptable, one seems a bit extreme! Check to see how many dealerships actually had sales in that year. If they all had one or more sale, chances are that some calculation is wrong, or we had some horrible error in loading the data. If only some dealerships had sales, while others did not, that is probably OK. In other words, **if something like this were to happen for your team, you would need to address it! You would examine the data some more, as well as the output and error logs from the tools used.** If you were unsure about the situation or unable to fix it, go back to the standard operating procedure for anything in the project, "when in doubt, please ask the instructor!"

5. Final Thoughts on Consistency versus Verisimilitude

In other words, prices are fake, so sales amounts and taxes might seem too high or too low based upon market values but the overall size and shape should be believable (albeit somewhat academic in flavor). The diversity in the data is limited (e.g. geography details, names, selected MMCC or feature details). Some details are admittedly bogus and inaccurate (I do not use real zip codes, area codes, phone numbers, people, cities, or counties). But even though the data values might seem phony compared to a real business in the real world, it should always "add up correctly" within the project.

Some of the scenarios above (one, two, and four) are obviously indicating some sort of "self-consistency" issue. Logically speaking, this must have manifested itself either in the report itself, in the OLAP and analysis, in the target database for the Data Mart / cube, in the ETL logic, in the source system database(s), or in the raw data I provided for the source systems themselves. If you encounter this situation, try to see if there is a mistake by tracing it from the "front-end" to the "back-end" and if you cannot find the problem then **PLEASE DISCUSS IT WITH THE INSTRUCTOR!** While I try to keep these problems from creeping into the source system data files, invariably the "flavor of reality" will not be perfect. I can usually help you either identify where the problem crept in or tell you if you can safely ignore it [in other words, when it is actually bad Q(data)—to use Mr. English's notation for quality of data].

II Project Background Story

Our mythical enterprise was founded by Mr. Rusty Wheels, who started **"Rusty's Recreational Vehicles Incorporated (RRV)"** in 1964. Back then, he customized trucks and vans one vehicle at a time in his at-home garage. Since that time, the company has changed dramatically. Through wise investing, venture capital, and shrewd partnerships and acquisitions, **RRV** is now one of the leading manufacturers (and seller) of recreational vehicles in the United States of America. For well over the last decade, RRV has sold two main recreational vehicle product families, though the names and details have changed over the years.

[Caveat: Ok, all of this is a little simplistic compared to reality, but for the sake of simplifying the project we will assume that the variety of makes, models, and classes listed here do not seem unusually limited. Any attempts to make the variety more realistic would require more much data, space, and time to process and comprehend without significantly increasing the value of the overall project experience.]

A. RRV Recreational Vehicle Products (MMCC)

RRV vehicles are described by Make, Model, and Class. (At the time this document was written, the Make is probably either Packrat or Behemoth.) The Model is the type of vehicle within the Make (In other words, within the Packrat or Behemoth product family / product line). For each Model, there are three Classes available. Different Makes and Models can have the same or different Classes (the list of available Classes is dependent upon the Make for current products). Further description of the Make, Model, and Class are given in the numbered sections below. RRV vehicles also come in different Colors – not just "any old color" of course, but one of a fixed set of options whose availability depends upon the Make, Model, and Class for a given Model Year. The color can be a major factor for many customers when choosing which vehicle they want to buy after they have decided upon the Make, Model and Class. Therefore, the RRV staff often refers to our products using the "MMCC" with Color being the second "C".

B. Product Features and Packages

Our customers can choose from a variety of feature options when they order their new vehicle (We **NEVER** buy or sell used vehicles and do **NOT** do trade-ins – again this is a simplifying assumption for the sake of our project.). The available features are dependent upon the Make, Model, and Class and of course the model year and purchase date. Some features are not possible in some MMC combinations. Some features are already part of the product because of the definition of that MMC. In some respects the Classes can be thought of as a preconfigured set of product features that are included "by default" when the class is selected. We offer a limited number of packages, each of which have a name, suggested price, and fixed set of particular optional features. The package definitions can change periodically throughout the year, based upon marketing and manufacturing decisions. We also potentially create new sets of packages (retiring some of the old ones) with each new model-year rollout.

C. Product Packages

Of course we also have packages that include various combinations of features from the "standard" ones to the more "deluxe" ones for a given Make, Model, and Class. These can optionally be added to customize the product beyond the limited variety of the "Class" definition. A customer would select a Class as a starting point and then add a package to that in order to provide (or omit) the desired features. These Packages have additional cost of course, but often the actual price paid for a package is different from (discounted or occasionally inflated) the package suggested price. In the end, the amount paid for any particular vehicle is similar to the price we quote the customer but rarely the same – this allows our sales staff the flexibility to "negotiate." Also, the total sales amount needs to account for any additional fees that might be incurred (if the customer orders a vehicle that needs to be transported, stored, or processed in a special way versus the standard fees) as well as the sales tax we are required to charge.

D. Physical RRV Instances

Each Physical RRV unit is called a "Vehicle Instance"; it is the combination of MMCC, and a package feature set. Each RRV is in one of three states: Unallocated (in storage at some warehouse), Allocated to a dealer (but not Sold yet), or Allocated to a dealership and then subsequently Sold to a customer. Dealers can ONLY sell Allocated vehicles (since the vehicle is not physically sent to them unless it is Allocated to them first.) Each RRV Instance has a unique serial number, which can be used to track its location and even the manufacturing details for it.

E. Sales to the Customer

When a Dealership sells a vehicle to a customer, they use an incentive plan. This describes the payment details, etc. Even when there is no "special sale" or "deal", a "standard" incentive plan is used.

For simplicity's sake, we will accept that each sale is to an **INDIVIDUAL CUSTOMER**. In other words, while the real world allows for joint-ownership we will assume that this is not necessary and all of our purchases are for only "one person". We will also accept that our business does **NOT CONSIDER USED VEHICLES OR TRADE-INS**. Obviously, this seems shortsighted or silly in the real world but for the sake of simplifying the models and the project we will accept this as not necessary. The only aspect of trade-ins we "might" entertain is to say that perhaps this is part of the behind-the-scenes details we do not capture that explains why the actual sales amount and the MSRP can be quite different for some sales. Again, I realize this is VERY IMPORTANT in the real world but including used vehicle sales and trade-in data would greatly complicate the OLTP data models as well as the MDM models with no real benefit to the project experience.

F. MMC Details

Historically, there are two well-known Makes: the Packrat and the Behemoth. Each make has several Models, which are described, briefly in this section. Caveat: Just like in the real world, the descriptions here were accurate as of the date and time that the original version of this document was written but might be slightly different from the current Makes and Models being sold and captured in the current OLTP databases.

[IMPLICITLY IN THE DESCRIPTIONS BELOW, THERE ARE EXAMPLES OF FEATURES AND PERHAPS FEATURE SETS! For example, A/C is a feature, so is the number of sleeping berths. You need to extract these details and capture them in your MDM!]

1. Packrat Models

Each Packrat can come in one of six models:

a) Packrat Hydrogen

This is a lightweight unit that sleeps one person comfortably. Perfect for the individual looking for a weekend getaway or a vagabond lifestyle. It has a limited storage capacity, but it is inexpensive and gets fairly good mileage. The features are quite limited: there is no onboard water, no kitchen, no bathroom facilities, no furnace, and no A/C (air conditioning).

b) Packrat Helium

Like the Hydrogen, this is also a lightweight unit, but it has the added feature of being able to comfortably sleep two. With a slightly larger storage capacity, and a slightly larger travel-range, it is perfect for short trips with a friend. But still, there is no onboard water, no kitchen, no bathroom facilities, no furnace, and no A/C.

c) Packrat Argon

With its added feature of a slide-out sleeping-compartment, the Argon is ideal for family camping trips. It can comfortably sleep five adults, and can optionally tow a travel trailer. There are no bathroom facilities, but it has a small onboard water system, minimal kitchen, and a furnace (mainly to prevent the water pipes from freezing and bursting in cold climates).

d) Packrat Antimony

The Antinomy is a middleweight unit that includes more features, such as a full kitchen and half-bathroom while comfortably sleeping three adults. It has a midsize onboard water system that supports both the kitchen and bathroom. Again, there is a furnace feature included for increased comfort and to prevent damage to the plumbing in cold climates.

e) Packrat Polonium

Essentially, the Polonium is equivalent to an Antimony with respect to the available features, but with some additional features. In particular, it has a slide-out sleeping compartment (allowing it to comfortably sleep five adults), and can optionally tow a travel trailer.

f) Packrat Radon

The Radon is our newest Packrat. It has better gas millage (estimated) than the Argon. It is capable of sleeping three comfortably (or four when an optional slide-out is added). The Radon has larger house-batteries for operation "off the grid" and an improved electrical system allowing it to blend modern conveniences and "green" technologies. It has three solar panels (with additional panels available as optional add-ons) to recharge the house batteries even when not attached to AC current.

2. Packrat Classes

Each Packrat can come in one of three possible classes:

a) Basic

This is the basic level, unless additional feature packages are added, the options selected for many features in this class default to the "cheapest" option available. For example, the seats are vinyl; the beds are inflated (air mattress), there is no carpeting, and there is no electric hookup.

b) Deluxe

This is one level above basic. The default options selected for features are slightly higher cost and value. For example, the seats are cloth; the beds are foam, and a 60-amp electric hookup is available.

c) Ultra

The features included here by default are more expensive and luxurious than the other classes for this Make, but there are still some features that can be changed (either selecting "higher" or "lower" options for some of the features). The standard set of feature options includes Interior Walls that are wallpapered and insulated. The seats are leather, the beds have coil spring mattresses, and there is a 100-amp electric hookup.

3. Behemoth Models

Each Behemoth can come in one of four models:

a) Behemoth Lawrencium

This is the entry level for a truly self-contained unit. It is larger and has a greater gross weight capacity than any of the Packrat models. It comfortably sleeps two, but includes a full kitchen, bath, and living room area. It has a midsize onboard water system, good furnace, and small air conditioner.

b) Behemoth Rutherfordium

The Rutherfordium is like the Lawrencium, but it can comfortably sleep four, with a slightly larger storage capacity, and a slightly larger living room. It also has a larger water capacity and larger heating and air conditioning systems.

c) Behemoth Hassium

Our biggest unit, the Hassium can comfortably sleep seven, and has a larger onboard water capacity than the other models. The Hassium has three slide-outs and a large living room, kitchen, and dining area. The furnace and air conditioner use a common air duct system and a heat pump provides efficient heat when going down the road.

d) Behemoth Tungsten

Our newest unit, the Tungsten can comfortably sleep five, with an onboard water capacity than between the Hassium and Rutherfordium. The Tungsten has two slide-outs and a large living room, and kitchen area. It can have a furnace and air conditioner (optionally using a common air duct system). It also can be equipped with a heat pump (non-standard).

4. Behemoth Classes

Each Behemoth can come in one of three possible classes:

a) Rover

Ideal for those who view the vehicle as a means of transportation and comfortable camping. Added features include a larger gasoline tank, and lighter chassis for better gas mileage.

b) Homesteader

For those who feel that the destination is more important than the journey. Features include larger water capacity, onboard generator, and entertainment package (including Televisions and Stereo Systems).

c) Boondocker

Ideal for those who camp way off the beaten trail. Features are included to enable comfortable camping in places that have no creature comforts or modern facilities. These features include a larger generator, higher quality furnace and air conditioning system, as well as solar panels.

G. RRV Sales Organization Units

All of RRV's vehicles are sold through dealers in the United States of America (actually, only in the lower 48 continental states). But in the future, we will be expanding (hopefully into Canada, Mexico, and other places too!). A description of the sales organization is given in the following paragraphs. The terminology has changed over the years, as have the details, but the overall structure is still basically the same.

1. Dealerships

Dealers are either corporately owned or independent from RRV. The independently owned Dealers that choose to become a Packrat or Behemoth dealer agree to follow RRV's corporate policies and procedures. For example, they agree to send financial statements by the first of the month, etc.

Corporate Dealerships are essentially a branch office of some corporate Department. Not all branches are dealerships however, some branches are Manufacturing Plants, and some are just normal branch offices (not plants or dealerships). Plants do not have a department over them however.

2. Sales Domains

Each Dealer sells within a single Sales Domain. Sales Domains and the other organizational units are defined by the sales organization. These definitions are based upon criteria defined (and possibly changed) by them. In other words, these things are completely within their authority, and subject to change at the whims of the head of the sales organization. That said, depending on the population density in their location, the Sales Domain is typically a group of small towns, a medium sized town with surrounding suburbs, or a metropolitan area. Vehicle allocation is controlled at the Sales Domain level, but each level above (Zone, Region, and Corporate) also shares part of the responsibility and management to ensure proper planning of inventory stock, incentive programs, etc.

3. Sales Zones

Each Sales Domain belongs to a single Sales Zone. Typically, a Sales Zone is also a larger geographic area, or at least it is possible for it to be so. Historically, we have seen Sales Zones spanning as much as two to four states, but for more populated areas, a Sales Zone may contain fewer states or even be contained within only a single state (e.g. California). It is not defined based on geography, but this description indicates some of its influence and issues.

4. Sales Regions

Each Sales Zone is within a single Sales Region. Like the other units, regions are defined based upon criteria known and understood by the Sales Organization – not simple geography. Historically, a Region typically consists of several states and one or more Sales Zone.

[The descriptions above are the business user's view of the organization—the application's view for this uses different terminology but is capturing the exact same details and the exact same organizational structure!!!]

5. Metropolitan and Micropolitan Statistical Areas

The US Census Bureau defines areas based upon population, geography and other details—these are called Metropolitan Statistical Areas (MSAs). There are smaller areas defined within each MSA called Micropolitan Statistical Areas (MICSAs). We use these definitions as another mechanism for grouping Dealerships based upon the dealership location within one of these areas.

MSA and MICSA definitions can change every 10 years or so (they are derived from the census) but we do not necessarily need to use the most current MSA / MICSA / census information—we are free to decide if and when we want to use the newer information, but we usually try to keep reasonably current.

6. Management and Ownership

Each Sales Unit mentioned above should have a Manager (e.g. a Regional Manager manages a particular Region), while a Zone Manager manages a particular Zone. From time to time, it is possible for a Sales Unit to be "without an official Manager". Sometimes this can be an "almost permanent" situation. It is also possible for the same person to manage more than one type of Sales Unit (e.g. being a Regional Manager and also a Zone Manager), and even possible for them to manage more than one unit at the same level (e.g. managing two zones). Having said that, it is not a very common occurrence. Corporate Dealerships are NOT owned by a person and therefore this concept of "Sales Unit Manager" has **NO RELEATIONSHIP** to the corporate versus independent ownership of a dealership.

H. RRV Computer Environment

RRV actively uses many systems within the organization and dealerships. Some systems date back to the early days and run mainframe applications on actual or emulated systems. Other systems were acquired during mergers and run on various flavors of mini computers or even PC's. Luckily, this "pilot project" is only concerned with data from systems that are running on flavors of (or systems compatible with) Microsoft SQL Server 2012.

I. RRV Technology Initiatives

Currently RRV Inc. is in the process of reengineering its business. Their first projects identified a critical need for capturing information about Vehicle sales from the dealerships. Therefore, support for a data warehousing initiative has begun to gain momentum. At this point, the most important Subject Areas were identified as the Vehicle and Dealer areas. Lesser areas included the Incentive Program and Sales Organization subject areas.

Although the initiative has begun, information from these areas is not currently easy to access. The executives feel that this is causing them to miss opportunities and also causing them to wonder about the overall health and direction of the enterprise. Therefore, after meetings between 12 "specially selected members" of our end-users, IT department, and data warehousing team, a set of critical needs was agreed upon.

J. Data Warehousing Requirements

For each milestone, you will be performing a subset of the tasks/activities involved in this pilot (obviously, since this is a mythical company, you will NOT be implementing a complete system and will not have the benefit of a real company's resources.)

Our project will focus on a single business process: "ONE customer buys ONE vehicle." This means that this model can safely ignore details NOT directly related to the sales process. In particular, Inventory details such as warehouse costs, Allocation status (other than sold), etc. should **NOT** be included in this model. However, dates for Dealership Allocation and Receipt should be captured for the vehicles that we sold.

The following questions were generated from the meeting. N.B. that our pilot should attempt to come as close as possible to answering **as many of these questions as possible**, but we **cannot change the scope of our pilot.** We are going to focus our pilot project on only one business process/subject area. So although these requirements might list things from multiple areas, including areas different than the one that our project will focus on (such as Inventory), we are **NOT** going to create more than one cube and **NOT** going to create dimensional attributes or facts from these other areas in our MDM.

III Questions Our Business Users "Need to Answer"

The following questions were formulated as a result of the fictitious interviews or fictitious facilitated sessions that our fictitious company did as part of the initial steps in this data warehouse initiative. In other words, these are the types of "simple questions" they want to be able to answer with our DWE. Of course, in addition to questions like these, we should actually be able to answer many more questions that are not listed here.

A. Understanding the Questions We "Need to Answer"

Each question listed in part D should be considered a "starting point" for one or more reports in the mythical DWE. For example, whenever a question indicates "Each A" it is reasonable to assume that the user might want to restrict that to a particular subset or extend it to a larger set. For example, when considering a question with "Each Dealer" in it we should look at how Dealer is modeled in our MDM. If it is a dimension then we could restrict or extend the concept of "each" by Drilling Down and Rolling Up the levels in that dimension.

Each question is worded using a sort-of "Structured English" approach. In other words, we are attempting to use the business user's vocabulary consistently, and trying to use similar grammatical forms to indicate similar relationships within the questions. This is not perfectly precise or completely foolproof (just like the real world), but we should be able to understand the "gist" of what is desired.

When a question uses the structure "By A, B, and C" it is a little vague as to the exact shape desired but the context should become clearer as we build the MDM. For example, if a question stated "By Year, Month, and Day" we would (eventually) recognize that these are different levels in the same dimension and therefore describing a Drill-Down / Roll-Up scenario. If the question had instead stated "By Customer, Dealer, and Year" we need to know more about our model to understand the relationship implied – but (eventually) I think we would all agree that these are probably separate dimensions (separate axes or "edges" in the cube) rather than levels in a single dimension. So structures like "By A, B, and C" are usually identifying levels or members in the same dimension, different dimensions, or some combination of the two.

Careful reading combined with your MDM analysis and OLTP models should help you to decide what is intended but if it is not clear, **ASK THE INSTRUCTOR!**

B. Things to Watch For

There is only one MDM. This means that different questions might specify different details about a given dimension or fact, but there is ONLY ONE VERSION of that Dimension or Fact. In other words, if one question lists concepts that we decide to model as levels in a Dimension Hierarchy for a particular Dimension, that Dimension is the SAME (has those levels and hierarchy details) for ALL QUESTIONS! Whenever a question references two concepts with a "strong cohesive relationship between them" but implies that one concept is "fixed" or "frozen" while the other concept "iterates" or "spans" multiple values, we need to consider whether these concepts are in the same dimension or separate dimensions. In other words, recall the Date and Time Example discussed in lecture. Although these can be modeled in a single dimension, if they need to be "frozen and spanned" independently (AKA "pivoted", or "sliced-and-diced" orthogonally), then they probably need to be modeled as separate dimensions.

C. What If Some Questions Are "Out Of Scope"?

Since we are focusing our project on a single business process (and a single cube), it should be obvious that we cannot answer questions about other processes with our MDM / Data Mart / Cube. So what do we do if there is a question in the list that is NOT about our process / MDM / Data Mart / Cube? Examine the question to see if it provides any useful details (levels, members, etc.) but recognize that some dimensions or facts in the question might not be a part of this project. Even though this project is NOT considering an architecture or approach that prevents stovepipes, and also not implementing more than one Data Mart, the project description attempts to be somewhat like the "real-world". We would never intentionally create a stovepipe Data Mart project. Therefore, when more than one Data Mart is affected, it is important to consider the requirements (at least superficially) for all Data Marts. In other words, if we focus on Sales for this project but a question is asked about Inventory, the details for any dimensions or facts shared in both should be compatible if not identical. In other words, we want the same definition for our Dimensions across all Data Marts that share them. (In Kimball's world, we could use his approach, techniques, and architecture to ensure this and we would call them Conformed Dimensions and Conformed Facts – but since we are not doing Kimball's approach here we can only say that we are attempting a minimal effort to prevent stovepipes.)

D. Examples of the Types of "Actual Questions" that our Business Users "Need to Answer"

1. What is the sales trend in quantity and dollar amounts sold for each Make, Model, Class, and Color (MMCC) for a specific Dealer, by each Sales Domain, Sales Zone, and Sales Region for each State / Metropolitan Statistical Area (MSA) or Micropolitan Statistical Area (MICSA)?

2. What is the average selling price of a MMCC for each Dealer, Sales Domain, Sales Zone, Sales Region, and MSA or Micropolitan Statistical Area (MICSA) broken down by purchase date for each quarter for the past two years? Which week of the quarter is the worst week for sales?

3. What is the pattern in quantity of inventory by MMCC, for each Dealer, by each Sales Domain, Sales Zone, Sales Region, and MSA or Micropolitan Statistical Area (MICSA)?

4. How does the quantity of sold vehicles by MMCC, having a particular onboard water-storage capacity by Dealer, Plant, Sales Domain, Sales Zone, and Sales Region compare with the same time frame last year and the year before?

5. What is the history (two-year comparison) of the quantity of units sold by MMCC associated dollar amounts compare to the amounts by retail versus wholesale purchase types?

6. What are the dollar sales and quantities by MMCC this year-to-date as compared to the last year for each dealer?

7. What is the year-to-date average selling price and average quantity sold for each Make and Model, by Dealer Type (corporate or independent), Sales Domain, Sales Zone, Sales Region, and MSA or Micropolitan Statistical Area (MICSA)?

8. Do the RRV's made by any specific plant outsell the same MMCC made by other plants?

9. What were the maximum and minimum total sales (quantity and amount) for each incentive plan used in 2004? How do they compare with 2003?

10. How many Customers buy RRV's manufactured in the same State that they live in? Same City? What percent of our total sales per year are these "in same state" customers?

11. What day of the week is the best for sales in January 2005? What is the worst? What is the best and worst day of the week for the entire year 2004? What is the best and worst day of the week in our history?

12. For all the dealers in a particular Sales Domain, did we have more sales to "First Time Customers" or "Returning Customers"? How many of customers of each of these types do we currently have? Which type of customers has more sales (consider both number of purchases and total sales amount dollars) within each Domain?

13. Which Sales Areas had higher number quantity of Wholesale sales than Retail? What about higher total sales dollar amounts instead of higher number of units sold?

14. What Model Year sold the best for each Make and Model, broken down by Sales Region Manager, Zone Manager and Domain Manager?

15. What was the average MSRP for the feature packages on the best selling MMCC for last month? For last year? What did the customer actually pay for the package?

16. What Day of the Week do we make the most sales on? What Day of the Week to we have the highest profit sales on? Is this the same across all Sales Areas (Regions, Domains, and Zones) as it is within each Sales Area (within each Region, Domain, or Zone)?

17. What was the total sales for vehicles manufactured this year and also sold this year broken down by MMCC, dealership, manufacturing plant state (In other words, what state was this vehicle manufactured in), and model year?

18. How many of last-year-model RRVs were manufactured this year, for each MMCC? How many of last-year-model RRVs were sold this year, for each MMCC?

19. How many vehicles were sold in the same month that the dealer received them broken down by dealership? How many were sold in the same month that they were allocated to the dealer?

20. What day of the week were most vehicles allocated to a dealership on, for a particular year, broken down by sales region?

21. What day of the week were most vehicles received at a dealership on, for a particular year, broken down by sales domain? Do we receive more on weekdays or weekend days?

22. What were the minimum, and maximum processing, handling, and transportation fees for all vehicles sold last month at a particular dealership?

23. Which new dealership (only been established as a dealership in sometime this current year) has the most sales so far? Which new dealership has the highest sales total?

24. Which plant has manufactured most of the vehicles we have sold (In other words, ranking number of vehicles sold by manufacturing plant)? Rank these results by manufacturing state and city as well.

25. Which incentive plan is the most popular (used most often) for sales this year? Which is the most popular for all time? What is the breakdown by customer city for each plan (how many customers within each of the cities that our customers live in, used each particular plan) this year?

26. Did we sell more vehicles with leather seats or with cloth seats this quarter?

27. Do most of our customers buy RVs with full bathrooms, half bathrooms or no bathrooms?

28. Look at the number of RVs sold in the past five years: do we sell more RVs with both a furnace and an air conditioner, with only one or the other, or with neither?

29. What is the average size of the gas tank in an RV sold this quarter?

30. Show the number of RVs sold this year in the Midwest, broken down by the number of sleeping berths in the RV (e.g. the number sold with a single berth, the number sold with 2 berths, the number sold with 5 berths, etc.)?