

 *Database Answers*

Data Warehousing by Example



Barry and his wife Norlia on an Elephant in Malaysia

Barry Williams

barryw@databaseanswers.org

1. Why ?	2
2. The Approach	3
3. Data Warehouse for a Football Club	9
4. BI on the Beach	16
5. A Day at the Olympics	24
6. Holiday in Malaysia	41
7. Retail Sales	125
8. Retail Banks	141
9. Industry Data Models	155
10. Generic Data Models	165
11. Integrating Multiple Data Warehouses	173
12. Conclusion	178

1. Why ?

The purpose of this document is to present our 'Best Practice' approach to Data Warehouse design based on more than 15 years experience.

We are publishing it on Kindle, as cheaply as possible, in order to encourage constructive criticism so that we can improve the book.

We would be very glad to have your comments at barryw@databaseanswers.org.

After 5 years ago, a teacher emailed me to say that his students found my Data Models boring and were falling asleep in class !!!

So I began to wonder how I could make them more interesting and still easy to relate to.

My conclusion was that I could take everyday events to use as examples.

That is why I have used Football, Malaysia and the Olympics.

Of course, a holiday in Malaysia is not something we do every day, and Olympics is not an everyday event ;-0)

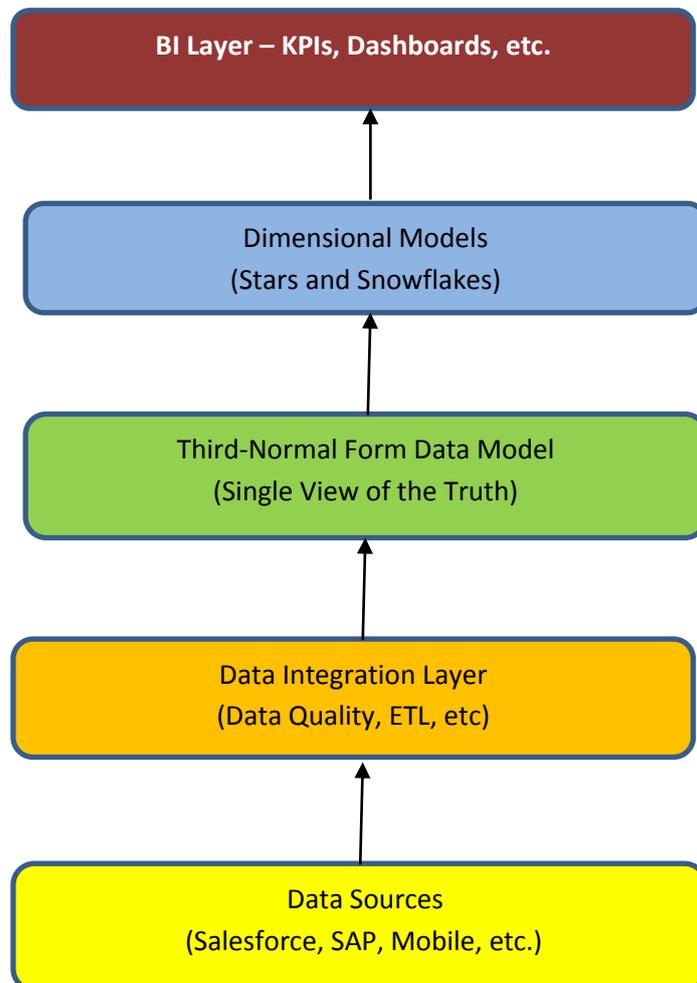
2. The Approach

In this Section we will discuss our Approach to the design of an Enterprise Data Model with associated Data Warehouses and how it applies to a Day at the Olympics and a Holiday in Malaysia.

2.1 Data Architecture

This Architecture supports Data Migration into an Enterprise Data Warehouse to meet BI requirements.

It shows the major Layers in an End-to-End Architecture for Data Migration from Data Sources, into a Data Warehouse and finally to a BI Layer to deliver data to the end-user.



2.2 Some Definitions

A Data Warehouse can be either a Third-Normal Form ('3NF') Data Model or a Dimensional Data Model, or a combination of both.

One benefit of a 3NF Data Model is that it facilitates production of 'A Single Version of the Truth'.

Multiple Data Marts will usually share common Dimensions, such as Dates, which we will call 'Conformed Dimensions'.

2.3 Steps

When all the Steps have been completed, Best Practice suggests that each item produced should be reviewed and extended or modified as appropriate.

This includes the EDM, CDM, Subject Area Models, Canonical Data Models and Canonical Entities.

In this way, the library of existing work is confirmed in its accuracy and relevance.

Step 1 : Draft the Enterprise Data Model (EDM)

Start by reviewing the list of candidates on the Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/enterprise_data_models.htm

Step 2 : Review the Subject Areas

Start by reviewing the list of candidates :-

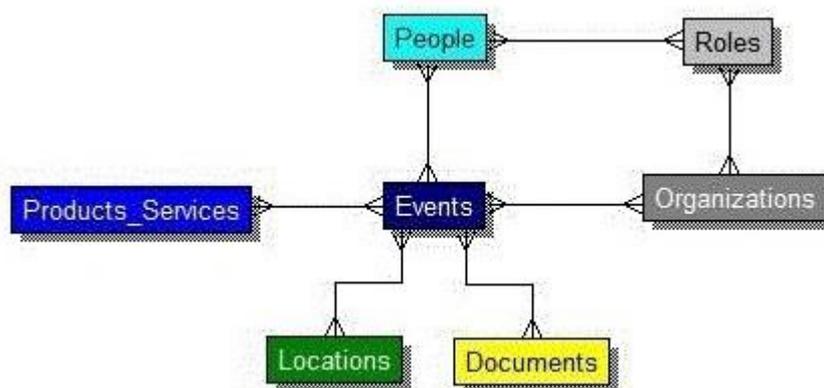
- http://www.databaseanswers.org/data_models/subject_area_data_models.htm

Step 3 : Use the Canonical Data Model (CDM)

Here is the Link for the CDM :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

This is what the CDM looks like :-



As you can see, it is centred around an 'Events' Entity.

Elephants, Olympic Judo and Data Warehouses

Typical Events could be :-

- A Customer makes a Purchase
- A Supplier makes a Delivery of Merchandise

Typical Documents could be :-

- A Sales Receipt
- A Bill of Lading
- A Contract
- A Delivery Note

People and Organisations are examples of the Roles played by Parties.

Parties are often shown in Models produced by professional Data Models.

In that case, Best Practice usually dictates that Semantic Models are produced to help business users understand how Customers and so on, are modelled as Parties and Roles.

An early example of the use of the Canonical Data Model (CDM) is to map data from Data Sources to the EDM.

This is a good opportunity to review the design of both the CDM and the EDM.

You can see examples of how this works in practice in the Chapters on 'A Day at the Olympics' and 'A Holiday in Malaysia'.

The current list of top-level Entities that feature in the CDM includes :-

- Documents, Events, Locations, People
- Products, Organisations, Roles, Services

A Party Entity is implied but does not appear at the top-level to simplify the layout of the CDM.

2.4 Development Framework

Our Framework contains two elements :-

1. The Development Approach
2. Components used in the Framework

The Approach involves these Steps :-

1. Identify the business Events
2. Define a Message for each Event
3. Map each Event to the Entities and Attributes in the CDM
4. Determine whether the CDM should be extended.
5. Create an Industry-specific CDM if appropriate.

The Components include :-

- Generic and Industry-specific Canonical Data Models (CDMs)
- Generic and Industry-specific Data Warehouse designs
- Core Entities and Core Subject Area Models – Customers, Products and Suppliers.

The rest of this Section is a Case Study for a Retail business Subject Area Models :-

- Generic (Horizontal)
- Industry-specific –
 - Insurance Common Data Model
 - http://www.databaseanswers.org/data_models/insurance_data_warehouses/common_data_model.htm

2.5 Enterprise Data Model

2.5.1 Process of Designing the Enterprise Data Model (EDM)

This shows the components used in the design of an Enterprise Data Model (EDM) with associated Subject Area Models, based on Industry-specific Models.

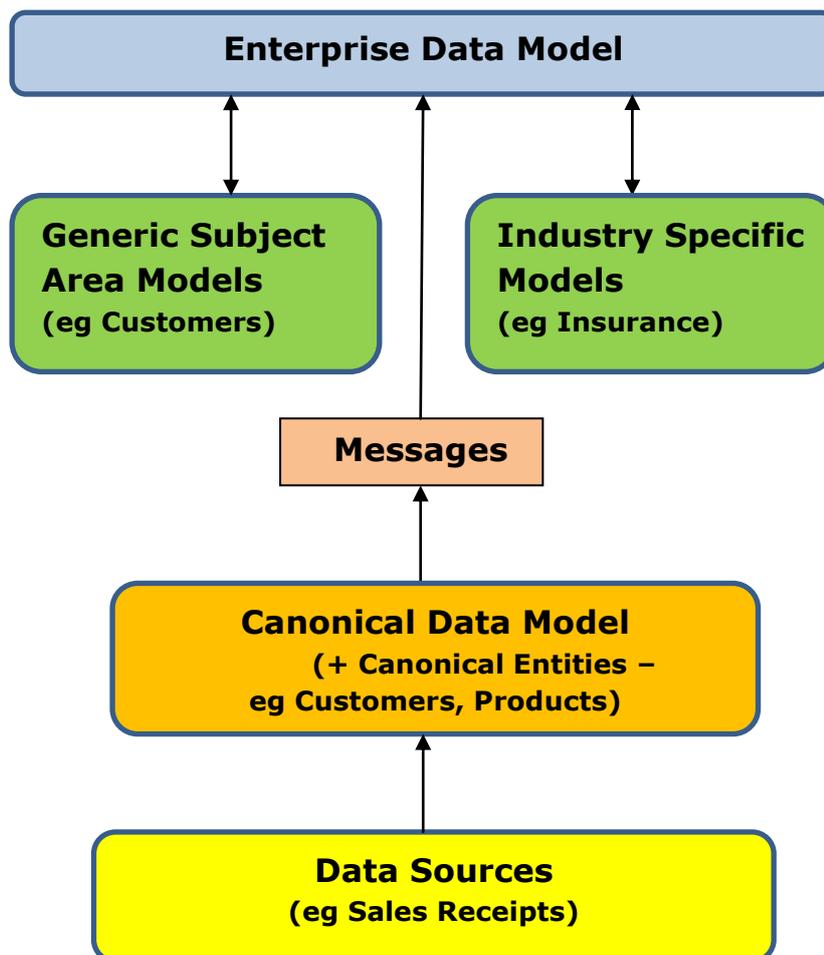
Each Data Source is reviewed against the Canonical Data Model and the appropriate Messages formats are defined. Then the data in the Message is mapped to the Enterprise and Industry-specific Models.

The current Enterprise Data Models are defined on this page of the Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/enterprise_data_models.htm

The Industry-specific Models are on this page :-

- http://www.databaseanswers.org/data_models/index.htm



2.5.2 Steps in the Process

Step 1. Create the draft EDM.

1. Review the Mission Statement in the Client's organisation
2. Talk to users and produce a Statement of Business Objectives
3. Look at the organization's Web Site
4. Establish BI Requirements, KPIs and so on.
5. Draft Life Cycles of Events, Customers, Products, others.

Compare your Draft with the off-the-Shelf Enterprise Data Models on this Page :-

- http://www.databaseanswers.org/data_models/enterprise_data_models.htm

Step 2. Identify the Subject Areas in the EDM.

Look for matches with the list on this Page :-

- http://www.databaseanswers.org/data_models/subject_area_data_models.htm

Step 3. Check Generic Data Models

Check out the list of Generic Data Models and Design Patterns on this Page :-

- http://www.databaseanswers.org/data_models/generic_data_models.htm

Step 4. Review the Canonical Data Model

Determine whether the Subject Area Models are compatible with the Canonical Data Model on this Page :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

3. Data Warehouse for a Football Club

This Chapter describes the Football Clubs Tutorial which is on this page :-

- http://www.databaseanswers.org/data_models/football_clubs_tutorial/index.htm

3.1 Starting with the Basics

We start by repeating a simple approach to designing a Database.

Then we follow with an example of a Dimensional Model and two Data Marts, sharing Conformed Dimensions.

Our simple approach consists of just these three simple rules :-

- Identify the things we are Interested in
- Decide how they are related
- Establish what else we know about them.

Rule number 1. Identify the things we are Interested in

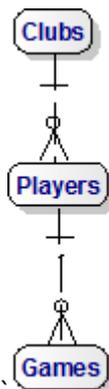
We decide these are Clubs, Players and Games and show them as Entities like this :-



Rule number 2. Decide how they are related

We decide that "Clubs employ Players and play Games against other Clubs".

So in a Data Model it begins to look like this :-



Rule Number 2b. Decide how they are related (continued)

So far so good - except that when we think about how this works in practice, we realise that not all Players play in every Game.

Therefore we need a more flexible way of identifying which players play in each Game.

Elephants, Olympic Judo and Data Warehouses

In addition , a Player will usually play in the same position.

But changes are possible and to provide for these changes, we need to identify the position each Player plays in during each Game.

When we think about Positions, we realise that there are a small number of predefined Positions and therefore, we should prefix the 'Positions' Entity name by 'Ref_' to establish that this is Reference Data.

So our Business Rules become :-

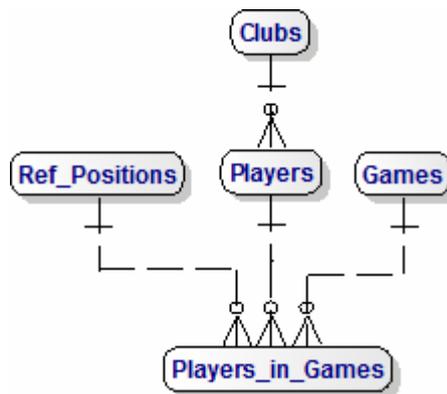
“Clubs employ Players who play in Games against other Clubs.”

The specific Players and their Position in each Game will be decided before each Game

So the ‘Things of Interest’ now look like this :-



And how they are related looks like this :-



Rule number 3. Establish what else we know about them

The additional data items that we know about the Entities is shown in this 3NF Data Model which shows all Attributes.

Let’s discuss what we know about Clubs :-

1. Because our Clubs entity refers to all Clubs, both ‘ours’ and others, we need to identify our Club. Therefore, we add a data item called ‘Our_Home_Team_YN’ which we set to either ‘Y’ or ‘N’ as appropriate.

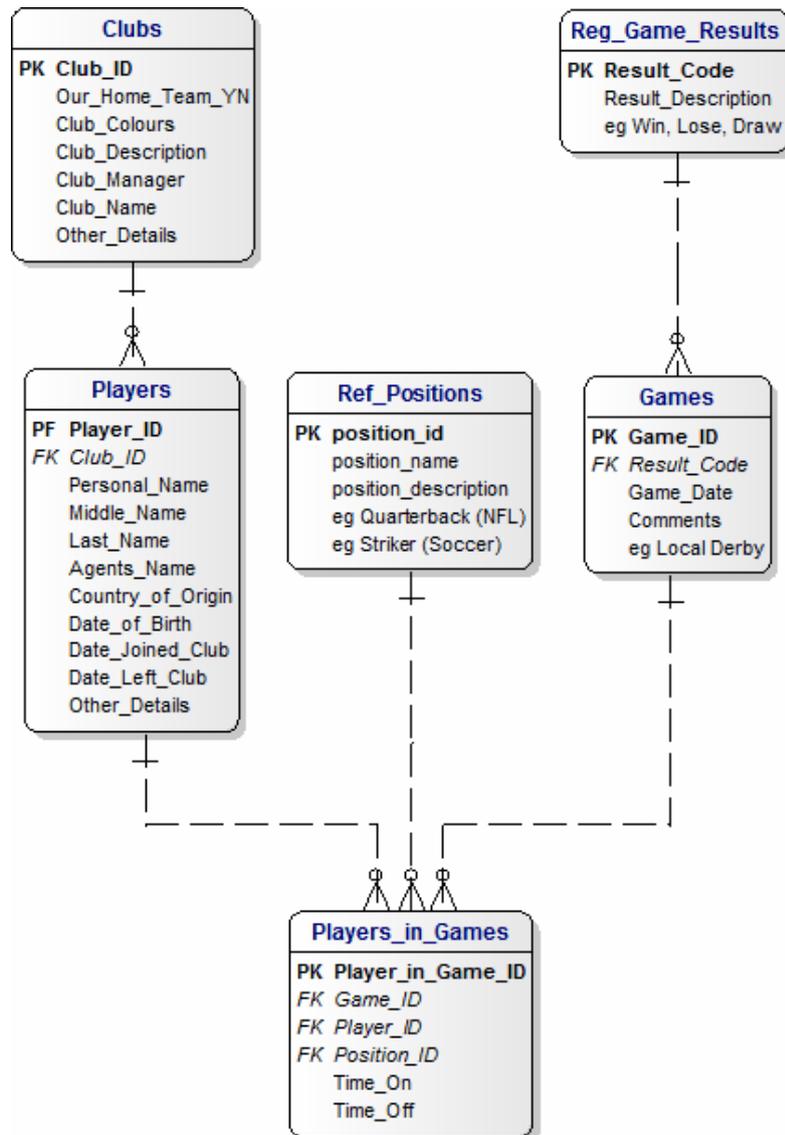
We will always want to record the Club Name, and we will usually want to know the Manager, and the Club Colours, and perhaps a description, such as the history of the Club.

To allow for future work, we include a data item called ‘Other_Details’.

In passing, let’s point out that we follow a certain order to the appearance of these items in the Entity definition so as to help us in simply having a consistent layout on all our Data Models.

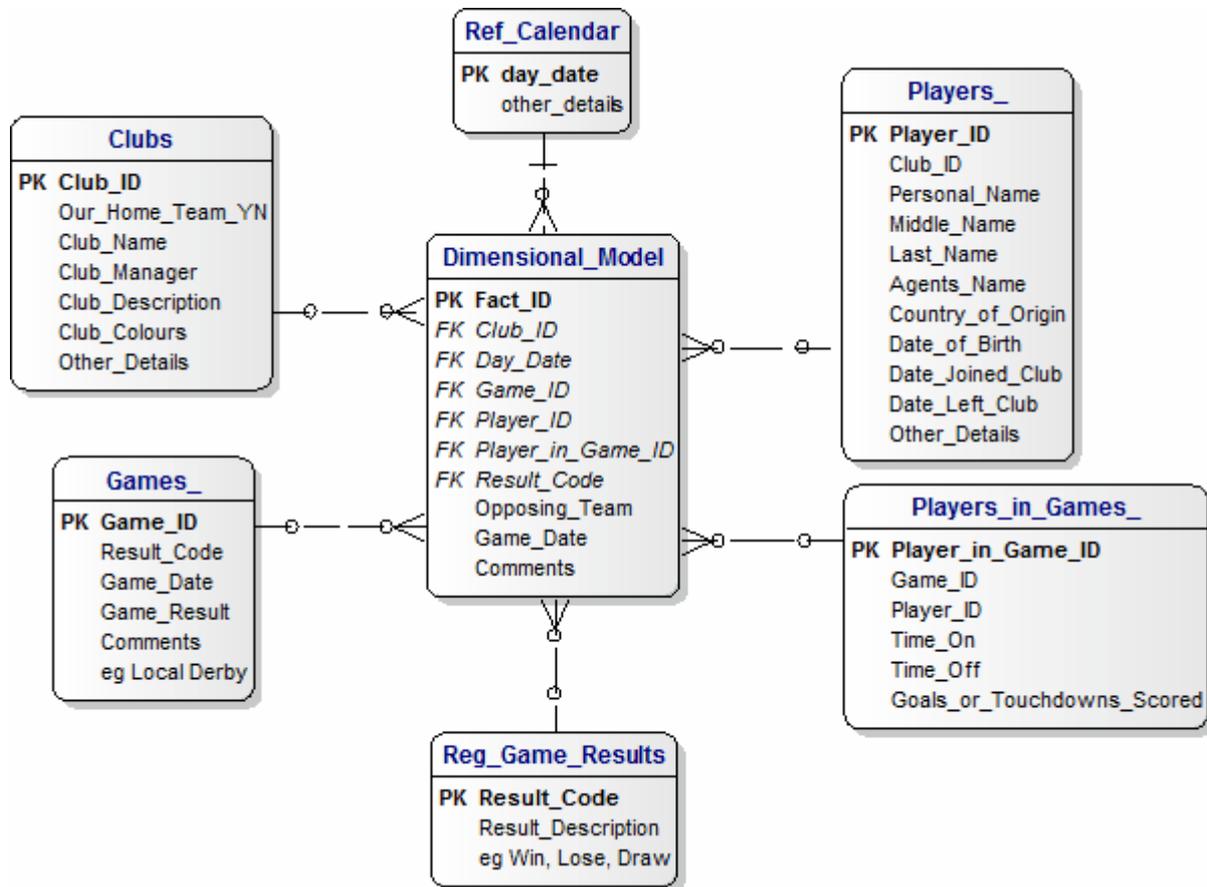
We show Keys and Flags (such as YN values) at the beginning, with all other items listed alphabetically. The additional items in the other Entities follows similar thinking.

Elephants, Olympic Judo and Data Warehouses



3.2 A Dimensional Model

We can now define our first draft of the Dimensional Model, which looks like this :-



3.3 Data Marts with Conformed Dimensions

We can also create an alternative design, based on two Data Marts with Conformed Dimensions.

In this approach we separate the data in the Dimensional Model into Games and Players and each has its own Data Mart.

Then we identify the Dimensions that they share and show them between the two Data Marts. We call these 'Conformed Dimensions' because the values have to conform to the same values in each Data Mart.

If they do not, then the values cannot be used to join Data Sets and any analysis from the two Data Marts would produce inconsistent the results.

Putting it differently, if the shared dimensions are not Conformed then our work as Data Warehouse designers and BI specialists would be completely wasted and we should take up a different line of work.

The best and simplest example of a Conformed Dimension is dates, which we show in a Calendar.

In passing, I call my Calendar entity 'Ref_Calendar' to emphasise the fact that it is Reference Data.

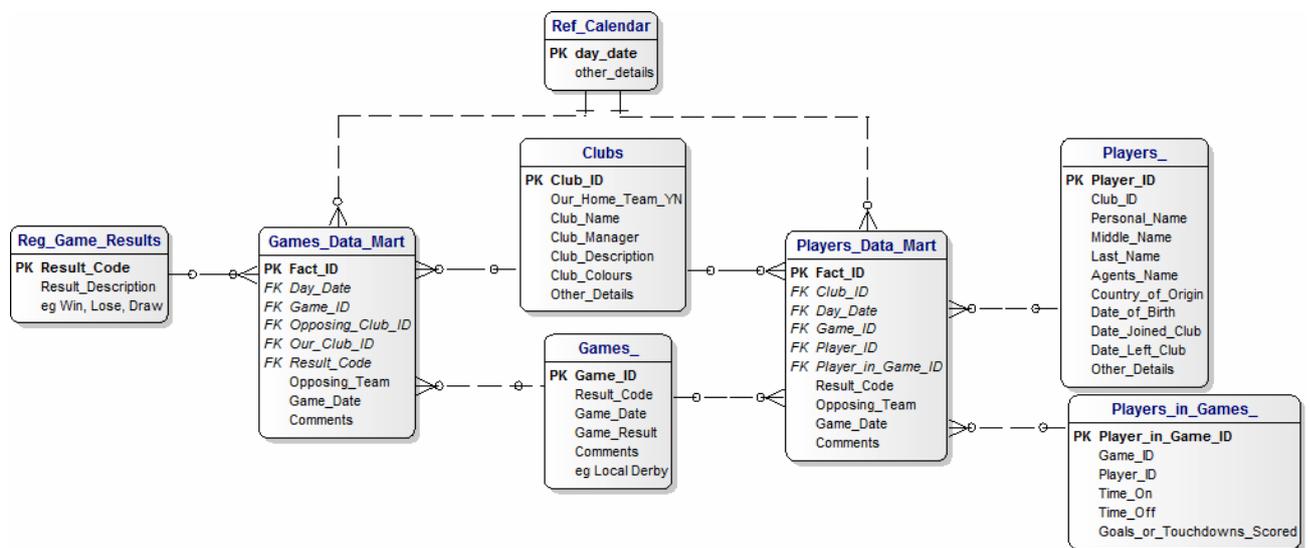
Elephants, Olympic Judo and Data Warehouses

The granularity of data involving time must be the same. For example, daily results cannot be compared from two Marts if data is not stored at daily level in both Marts.

Dates in a calendar are an example of Master Data, and by definition, Master Data will always be Conformed Dimensions because they represent one consistent value across an organisation.

In our case, we can use the Games Data Mart to obtain statistics about the Club and its standing in the League.

The Players Data Mart, of course, would be used to obtain individual statistics about our favourite players and to identify, for example, who is the most prolific scorer.



3.4 Enquiries using the Data Marts

A typical Enquiry would be :-

“What are the standings today ?”

To respond to this enquiry, we need a list of Teams with names, and shown in order of Points scored.

Points are obtained for the result of each Game :-

- A Win scores three points
- A Draw scores two points
- A Loss scores zero points

This is how it is shown on the BBC Web Site at 8 pm on Saturday, February 9th. 2013 :-

Football Tables



Show me: OR

Barclays Premier League Table																	
	Position	Team	P	Home	W	D	L	F	A	Away	W	D	L	F	A	GD	Pts
>	1	— Man Utd	25		11	0	1	32	15		9	2	2	28	16	29	62
>	2	— Man City	26		9	3	1	29	11		6	5	2	19	13	24	53
>	3	— Chelsea	26		7	4	2	30	12		7	3	3	25	16	27	49
>	4	— Tottenham	26		7	4	2	20	13		7	2	4	24	17	14	48
>	5	— Arsenal	26		7	3	3	33	18		5	5	3	17	11	21	44
>	6	— Everton	25		6	6	1	22	16		4	6	2	18	14	10	42
>	7	— Swansea	26		5	6	2	26	19		4	4	5	12	10	9	37
>	8	— Liverpool	25		6	3	3	22	10		3	6	4	22	22	12	36
>	9	— West Brom	25		7	2	4	19	12		3	2	7	15	23	-1	34
>	10	— Stoke	26		6	6	1	18	14		1	6	6	8	17	-5	33

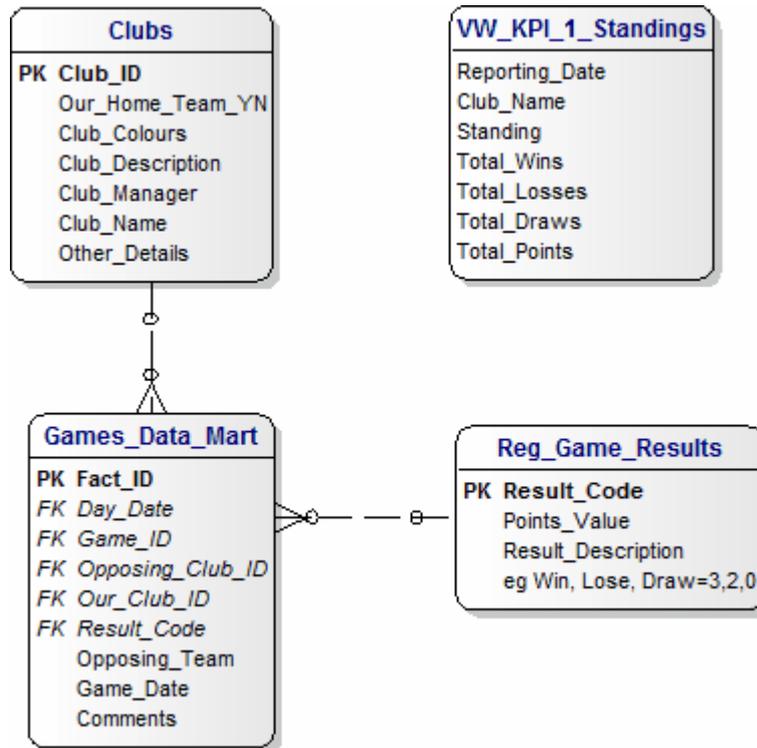
This is the data that we need to produce these statistics, listed by Number of Point, in descending order :-

- Club Name
- Where the Game Date is from the start of the season up to the present.
- Number of Points where each Win Result Code = 3, Draw=2 and Loss=0

In order to introduce an analytical Framework, we can define a Key Performance Indicator ('KPI 1') as the Standing that we have defined above.

Then we can create an SQL View as a neat way to bring the data together and populate it to a BI front-end. So we would have :-

Elephants, Olympic Judo and Data Warehouses



4. BI on the Beach

A Data Warehouse is frequently developed to meet an organisation's requirements for data to meet Business Intelligence ('BI') requirements.

One of the most common requirements is to provide Key Performance Indicators ('KPIs').

For a Retail business, a typical Financial KPI would be Sales or Profitability per staff headcount, and a Performance KPI would be On-Time-Delivery ('OTD').

At Database Answers we have designed a generic approach that we call 'BMEWS', which stands for Business Monitoring and Early Warning System.

The remainder of this Section describes how BMEWS could be applied to monitor risk of a Data Warehouse that would allow the CIO or Security Manager to relax on the beach.

This page describes this 'BI on the Beach' application :-

- http://www.databaseanswers.org/data_models/bmews_bi_on_the_beach/index.htm

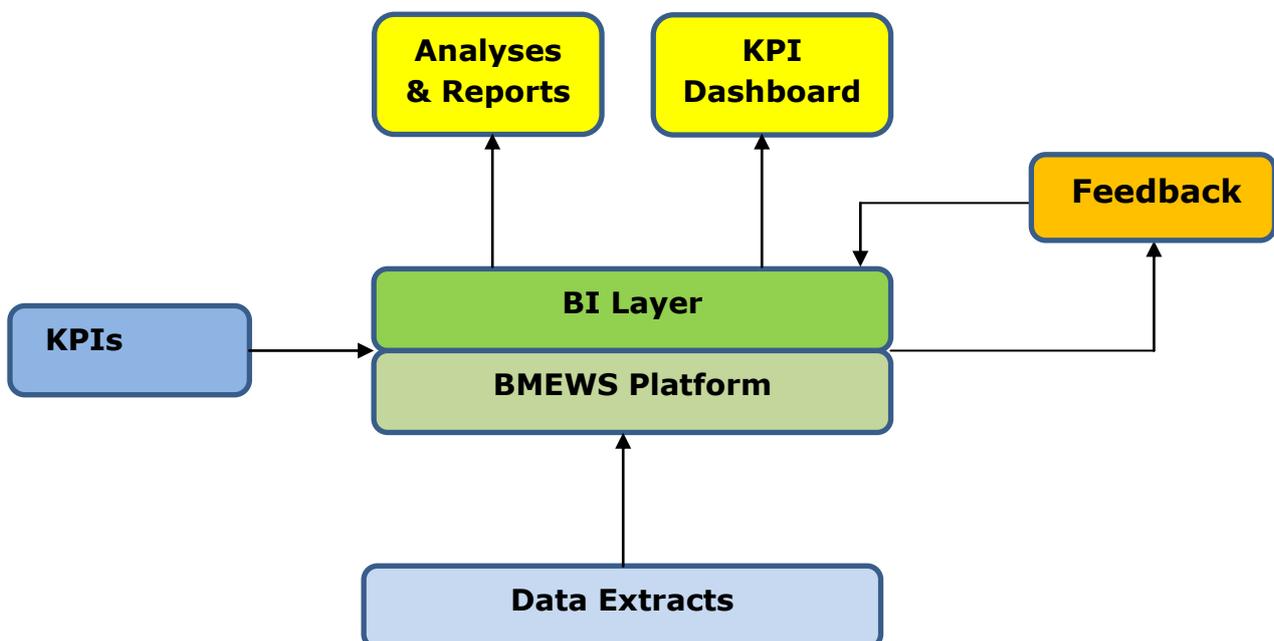
4.1 BMEWS BI KPI Monitoring System

This Section shows a generalised version of the RMS in the previous Section.

Risk Monitoring System (RMS) is set up to provide timely notification to the Security Administrator of any situation that requires his attention.

A number of Key Risk Indicators ('KRIs') will have been identified and Dashboards produced regularly by the RMS that tracks the KRIs.

The RMS allows the Security Administrator to pay attention only to the situations needing attention.



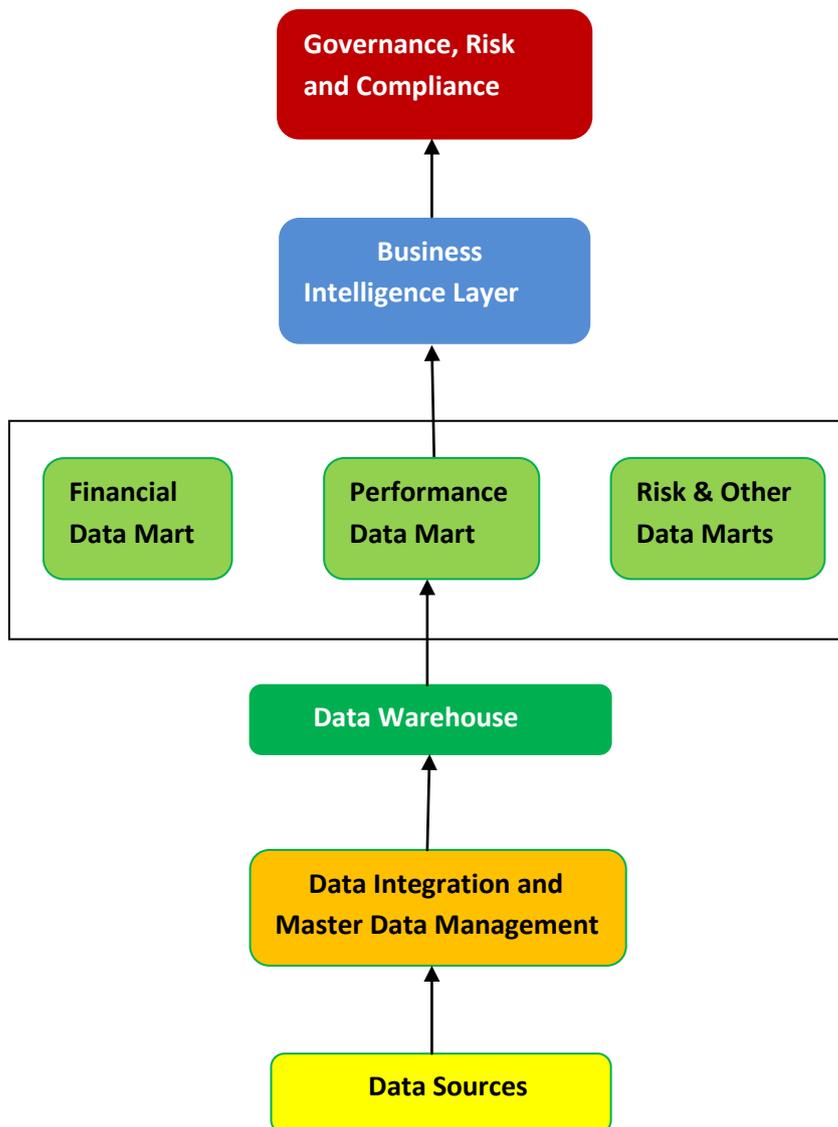
Elephants, Olympic Judo and Data Warehouses

4.1.1 A Data Migration Architecture

This Data Migration Architecture provides for a controlled transformation of data from a variety of Data Sources, through Data Integration to an integrated Data Warehouse.

This Data Warehouse provides a 'Single View of the Truth' and Data Marts are then populated for specific purposes such as Basel 3, Risk Management, Customer Sales, Profitability and so on.

Finally, the data in Data Marts is then displayed by Dashboards, KPIs, Reports and so on.

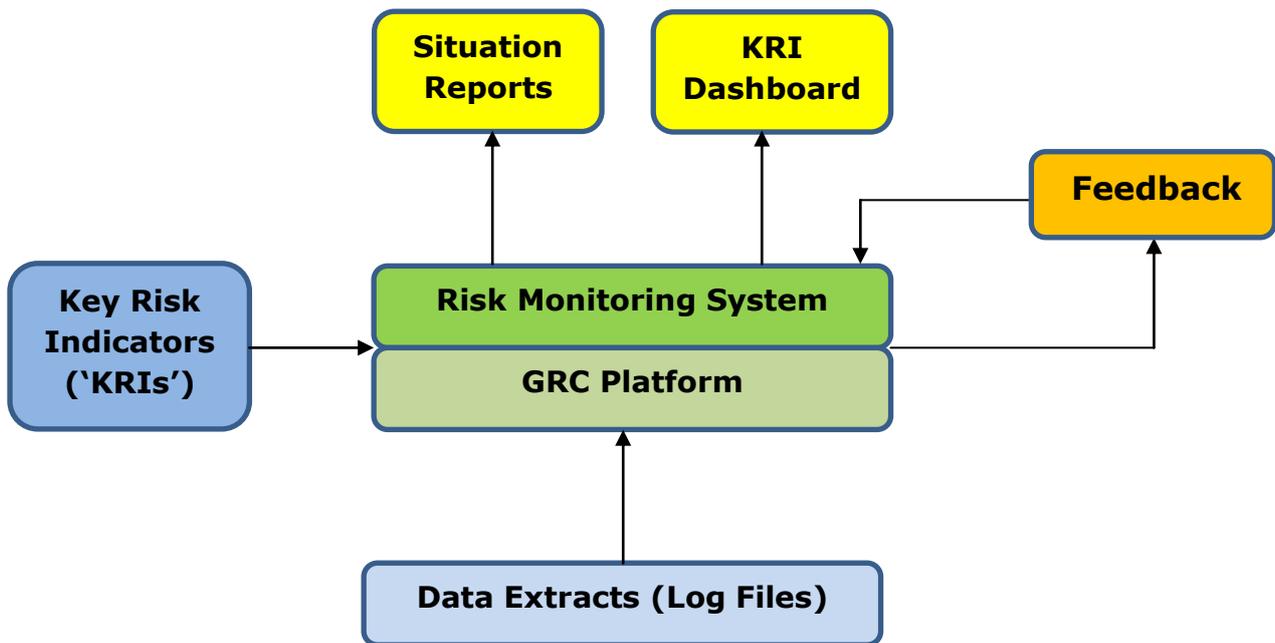


4.2 Risk Monitoring System

A Risk Monitoring System (RMS) is set up to provide timely notification to the Security Administrator of any situation that requires his attention.

A number of Key Risk Indicators ('KRIs') will have been identified and Dashboards produced regularly by the RMS that tracks the KRIs.

The RMS allows the Security Administrator to pay attention only to the situations needing attention.



4.2.1 User Scenario



- The Security Administrator is relaxing on the beach
- An Alert is sent to his Mobile phone about a situation that requires his attention.
- He drills down to the details
- He identifies the appropriate response
- He sends an Alert to the appropriate member of this staff.
- He goes back to his relaxation

4.2.2 Typical KRIs

The KRIs will be agreed with the business and will include such things as :-

- Peaks in unsuccessful attempts to login.
- Attempts to access Sensitive Data
- Compliance with Statutory regulations, such as Sarbanes-Oxley.
- Compliance with Best Practice in Data Management, such as Data Consistency checks on Master Data Management, Orphan Orders, and so on.

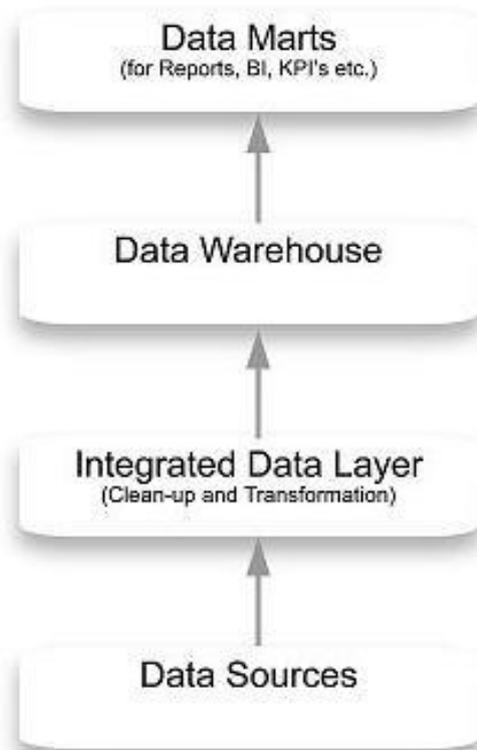
4.2.3 Risk Dashboard

The KRIs will be displayed in a Risk Dashboard looking something like this :-



4.2.4 Layered Data Architecture

This Layered Architecture appears further on in Section 6 and repeated here for convenience :-



4.2.5 Data Source

The Data Source is the Log File and looks like this :-

Log_File	
PK	Record_ID
	Message_Code
	Date_Time_Stamp
	IP_Address
	Failed_Login
	Failed_Password
	Other_Details

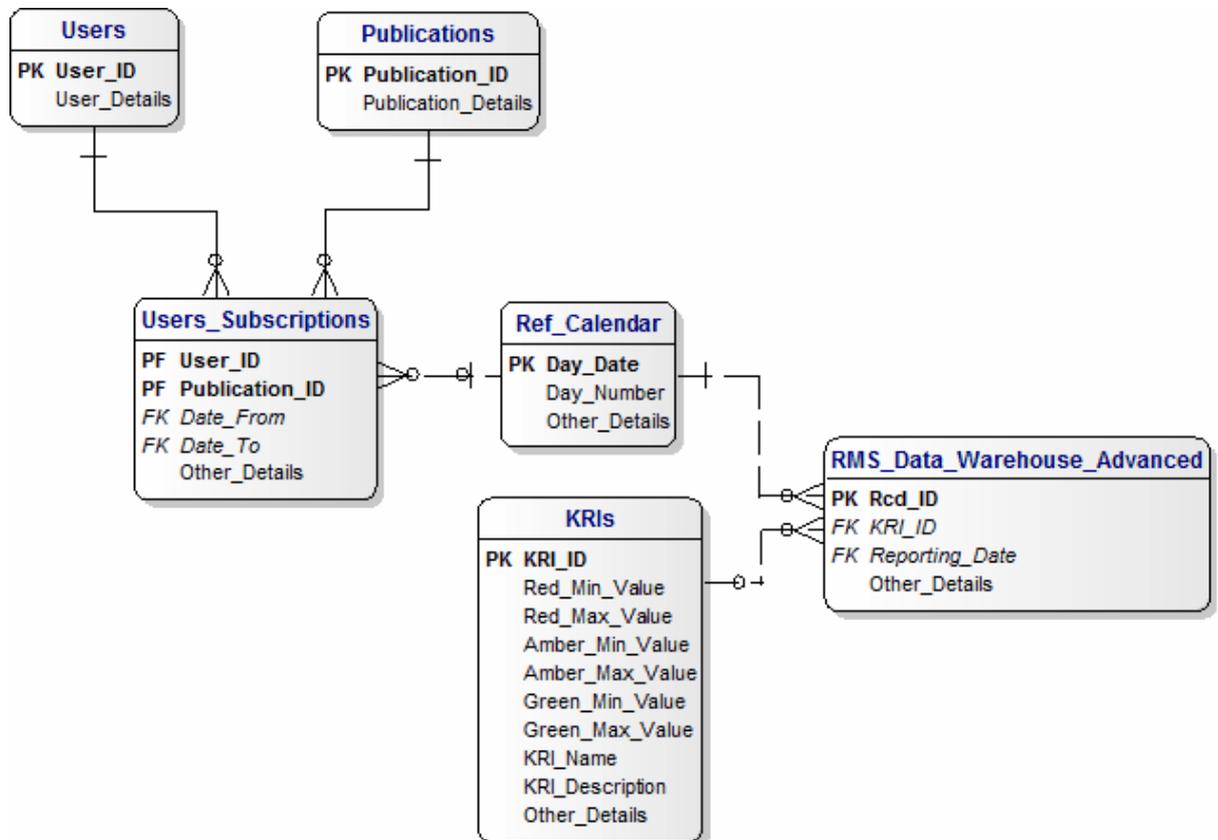
4.2.6 Integrated Data Layer

The Message Codes in the Log File will be defined by the System Supplier and at this stage, we will validate the Message Codes and translate those that we are interested in to the appropriate Event ID..

The output from this Staged will be data that has been validated and transformed ready to be loaded into the Data Warehouse.

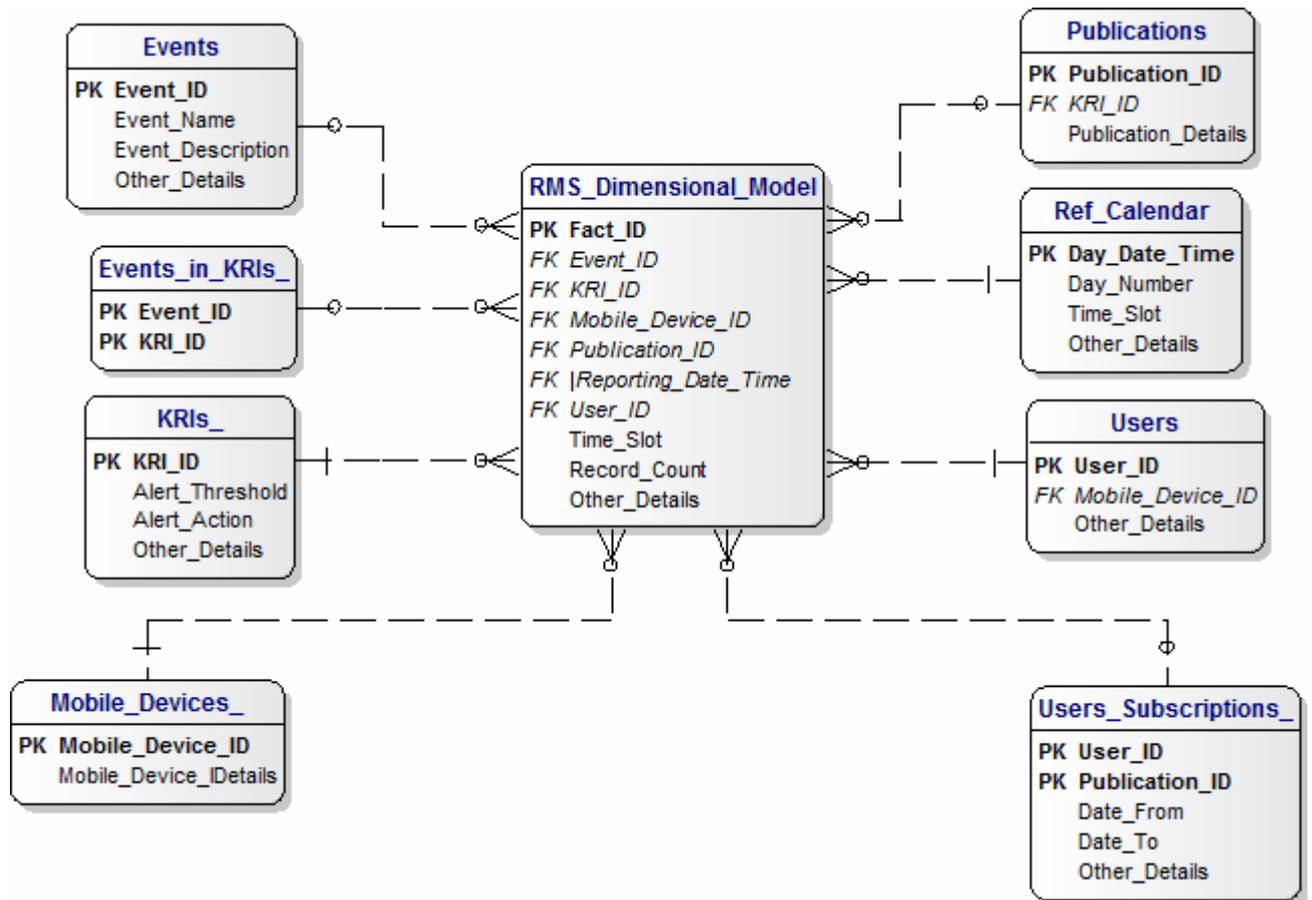
4.2.7 3NF Data Model

This help us to understand the data for Key Risk Indicators (KRIs) and how it is related.



4.2.8 Dimensional Model

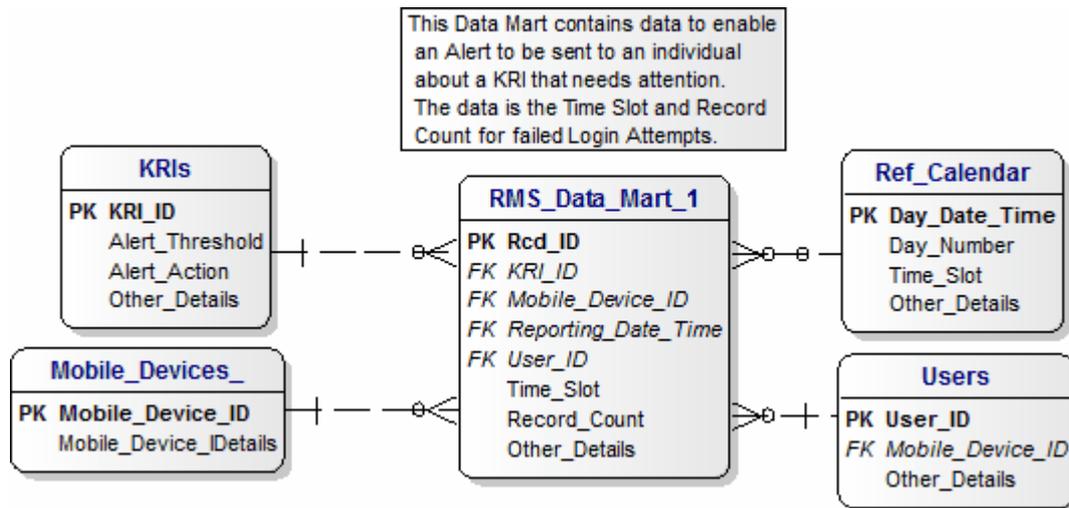
In discussion with the Users, we have identified Key Risk Indicator 1 ('KRI 1') to be the number of failed Login attempts in a 5-minute time period during the 24-hour day.



4.2.9 Data Mart 1

In discussion with the Users, we have identified Key Risk Indicator 1 ('KRI 1') to be the number of failed Login attempts in a 5-minute time period during the 24-hour day.

This Mart holds the data for KRI 1, which we use to send an Alert to all Users who have subscribed to KRI 1.



5. A Day at the Olympics

5.1 Getting Started

In this Chapter we use a day trip to watch Judo at the London Olympics Malaysia to show how we can apply our approach to a real-world situation and develop a design for a tailored Dimensional Model.

This picture shows Gemma Gibbons of the UK throws Audrey Tcheumeo of France on her way to a Silver Medal.



5.2 Introduction

In this Chapter we use a trip to the Olympics to discuss an approach to the implementation of a Reference Data Architecture and the design of a Data Warehouse.

A **Canonical Data Model** (CDM) is central to this and we discuss the benefits of using Design Patterns based on a CDM.

During the trip, I used the Ticket I had bought online, ate lunch and watched the Judo competition.

Elephants, Olympic Judo and Data Warehouses

On the way home I found myself thinking that the trip would provide a good opportunity to develop an interesting and 'User-Friendly' Tutorial on Data Warehouses.

The design of the Data Models reflects the scope and the fact that the overall aim is to provide data for Business Intelligence.

We also try to keep in mind that a well-designed Data Model should be good to look at and it should be possible to tell a story based on the Model.

5.3 The Approach

The Approach is to follow these Steps :-

Step 1 – Identify the **Events** involved

Step 2 – Define a **Design Pattern** based on the Event-driven **Canonical Data Model** from this

page on our Database Answers Web Site :

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

Step 3 - Define a **Message** Format for the data in each Event

Step 4 - Design a 3rd Normal Form **Data Warehouse** (DWH) and update it for each Event.

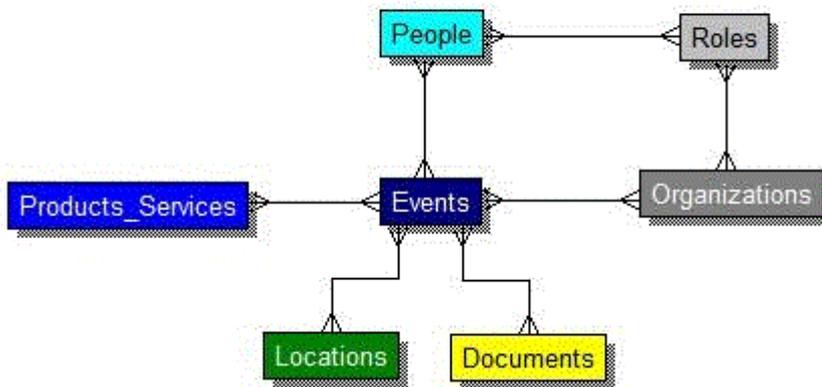
Step 5 – Define the format for **loading data** into the DWH for each Message

5.4 Canonical Data Model

Step 1 – Use the Event-driven Canonical Data Model from this page in our Web Site :

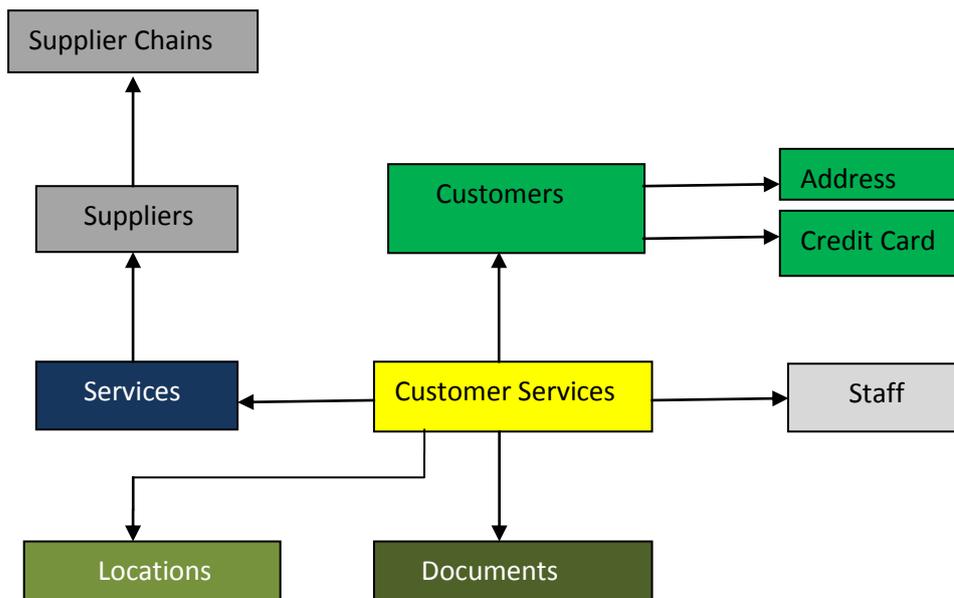
- http://www.databaseanswers.org/data_models/canonical_data_model/index.htm

Canonical Data Model May 25th
This is an ERwin Data Model
with Many-to-Many Relationships



5.5 Design Pattern

From the Canonical Data Model shown above, we can derive this Design Pattern :-



5.6 Generic Message Format

This shows the fields in the Generic Message Format :-

EVENT	DATE	LOCATION	PRICE	DETAILS

5.7 3NF Data Warehouse

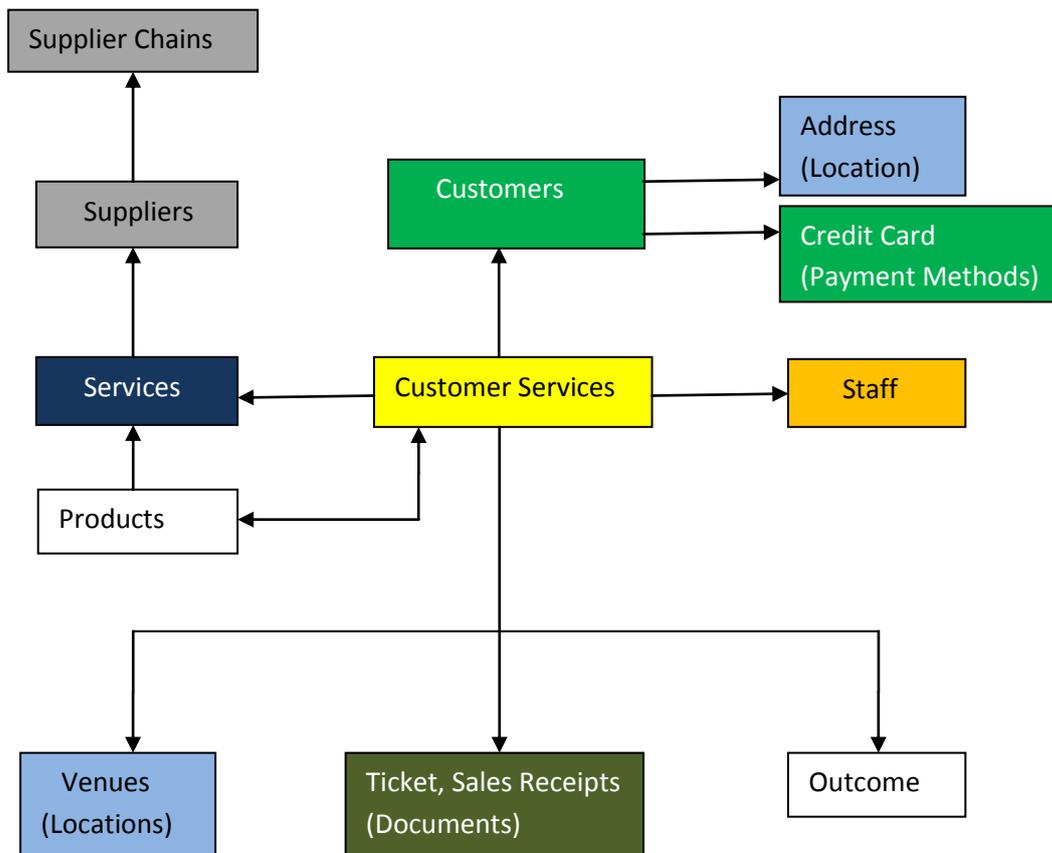
The design of this Data Warehouse is derived from the Design Patterns for the three Events that we discuss.

Every Event has an Outcome but it is not usually important and is taken for granted.

We show it here because it is very important for Judo competitions.

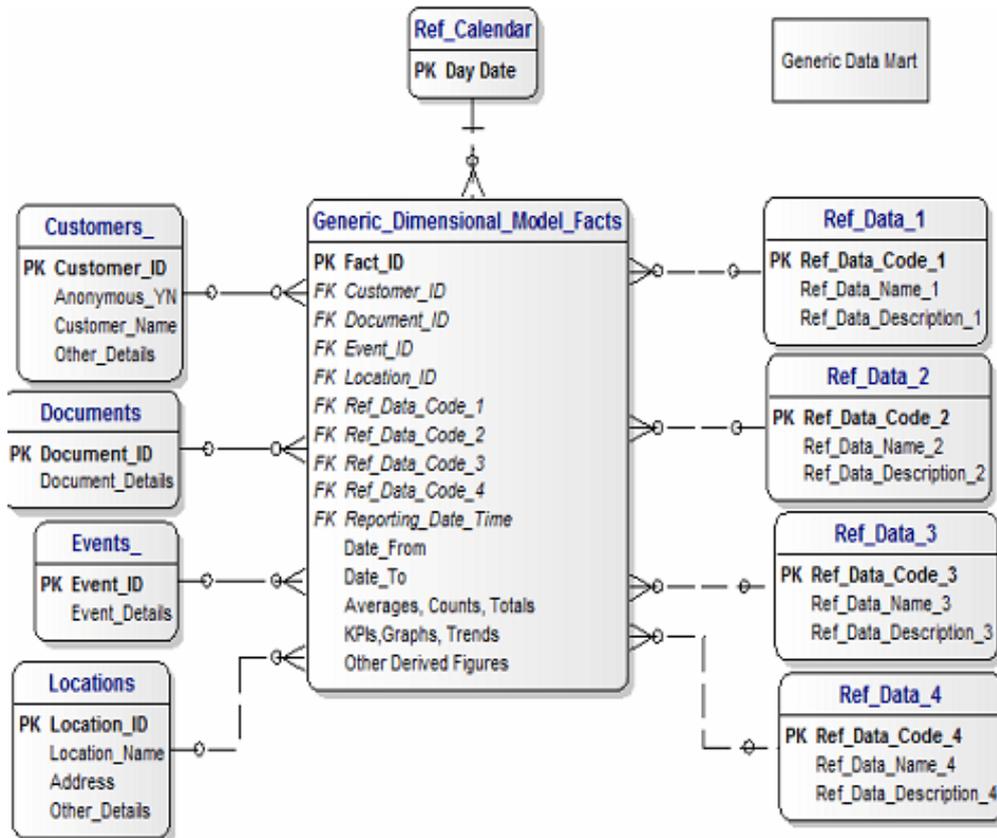
We show a many-to-many relationship between Products and Customer Services to provide for the situation where one Customer Services Order contains many Products.

In my case, this applied when I went to the Restaurant and ordered a main course, yoghurt and wine.



5.8 Generic Data Mart

This shows our starting-point

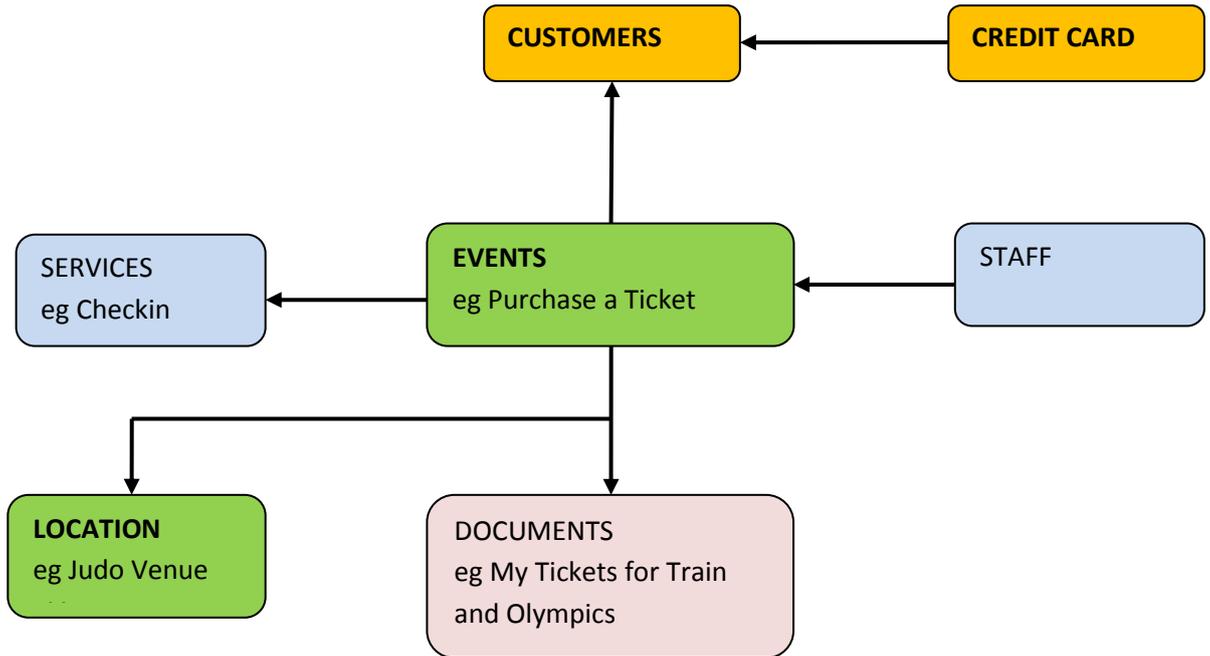


5.9 Event 1 – Buy Ticket for Judo Competition

This shows how we handle the first Event.

5.9.1 The Design Pattern

This shows how the Design Pattern applies to this Event.



5.9.2 Message Format

This shows the data items on the Ticket :-



This shows the fields in the Generic Message :-

EVENT	DATE	LOCATION	PRICE	DETAILS

This shows the fields in the Message for this Event :-

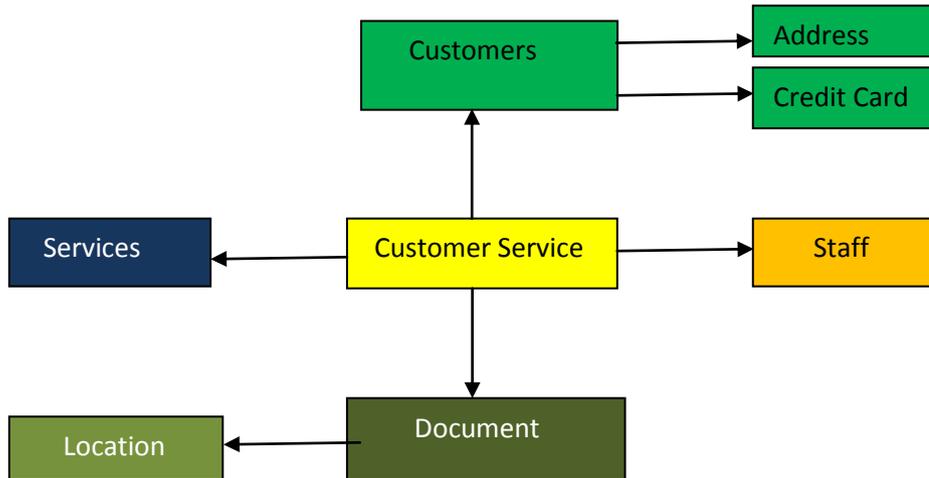
EVENT	DATE	LOCATION	PRICE	DETAILS
Purchase Ticket	Date of Purchase	Venue, Block, Row, Seat Nr	Seat Price	Date, Time of Competition

5.9.3 Data Warehouse

The benefit of adopting a Third-Normal Form ERD is that it enforces a 'Single View of the Truth'.

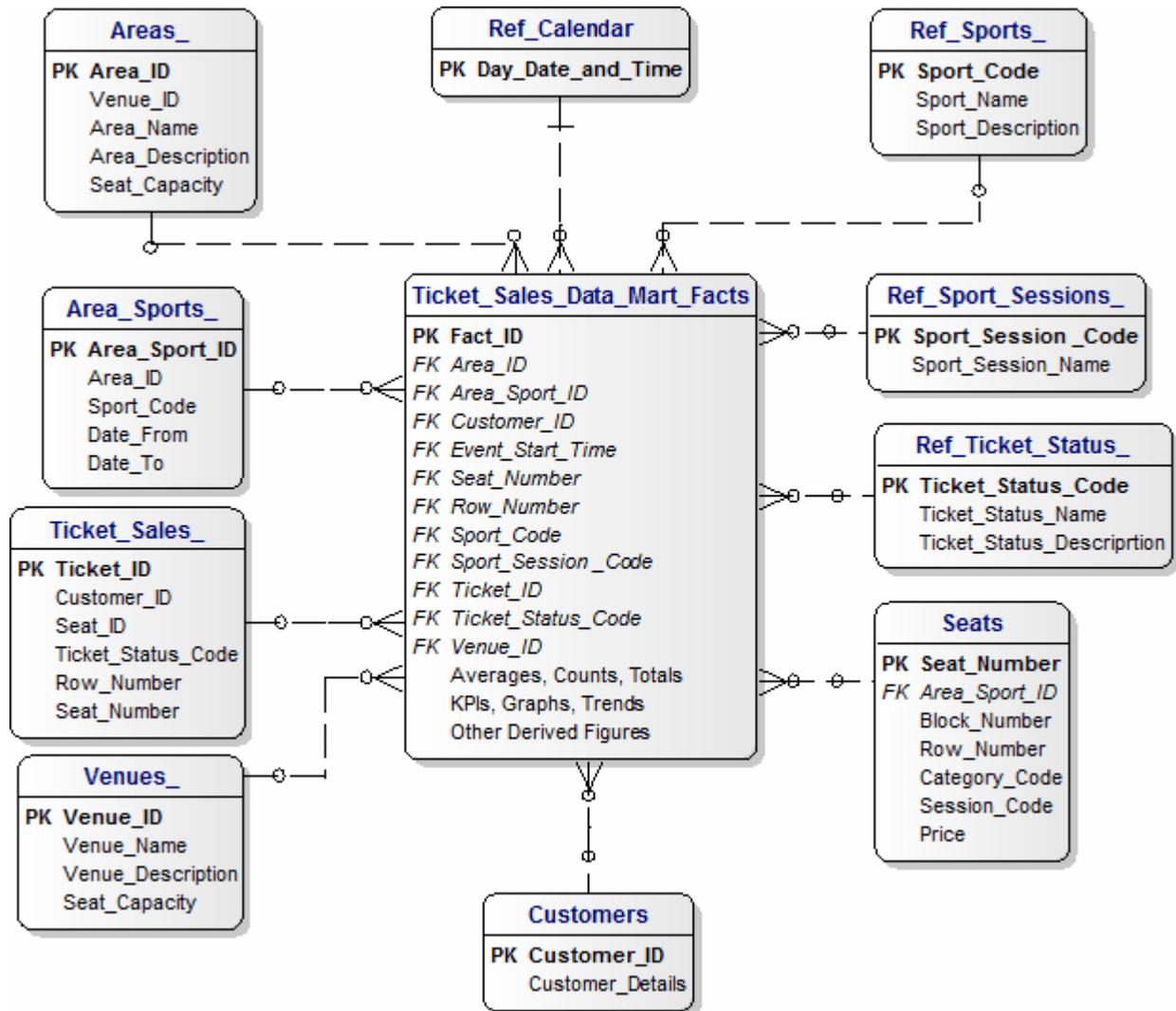
If we adopt a Dimensional Model it is not so easy to achieve this.

This shows the design of the Data Warehouse (DWH) after the first Event of Purchasing a Ticket



5.9.4 Data Mart

This shows the Data Mart for Ticket Sales.

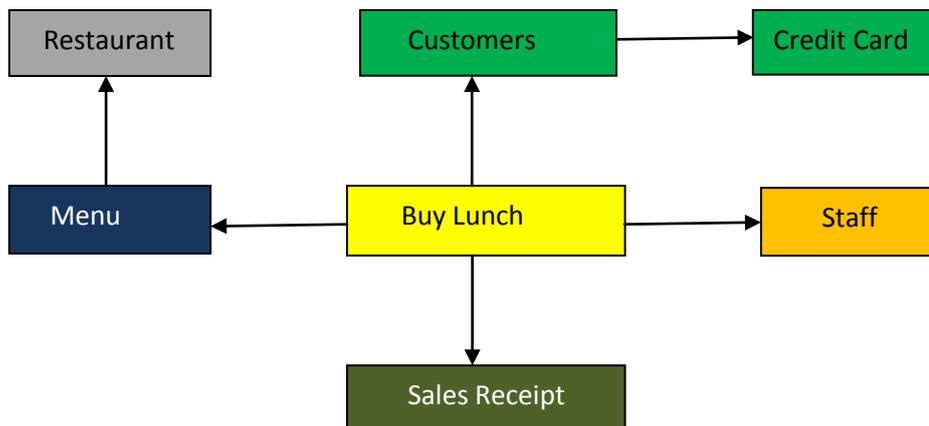


5.10 Event 2 – Get Lunch

This shows how we handle the Second Event.

5.10.1 The Design Pattern

This shows how the Design Pattern applies to this Event.



5.10.2 Message Format

This shows the fields in the Generic Message :-

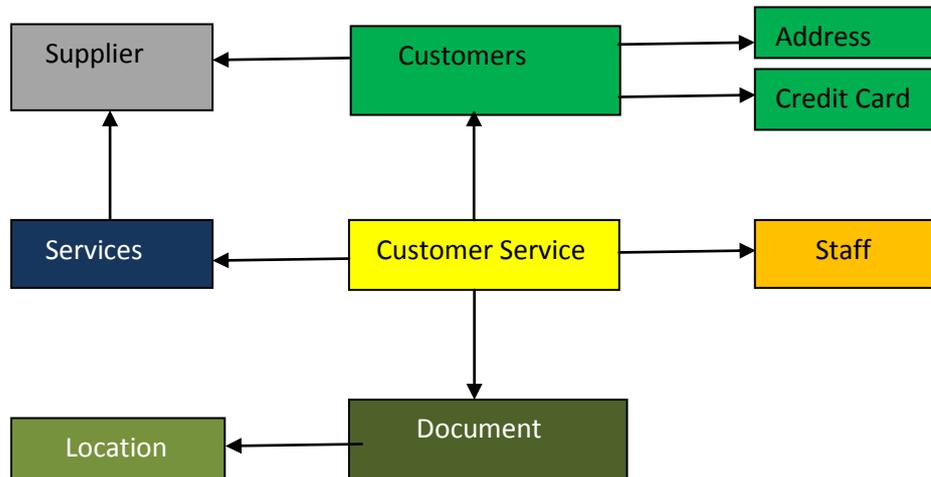
EVENT	DATE	LOCATION	PRICE	DETAILS

This shows the fields in the Message for this Event :-

EVENT	DATE	LOCATION	PRICE	DETAILS
Buy Lunch	Date & Time of Lunch	Restaurant	Total Price	Chicken and Mushroom Pie, Wine

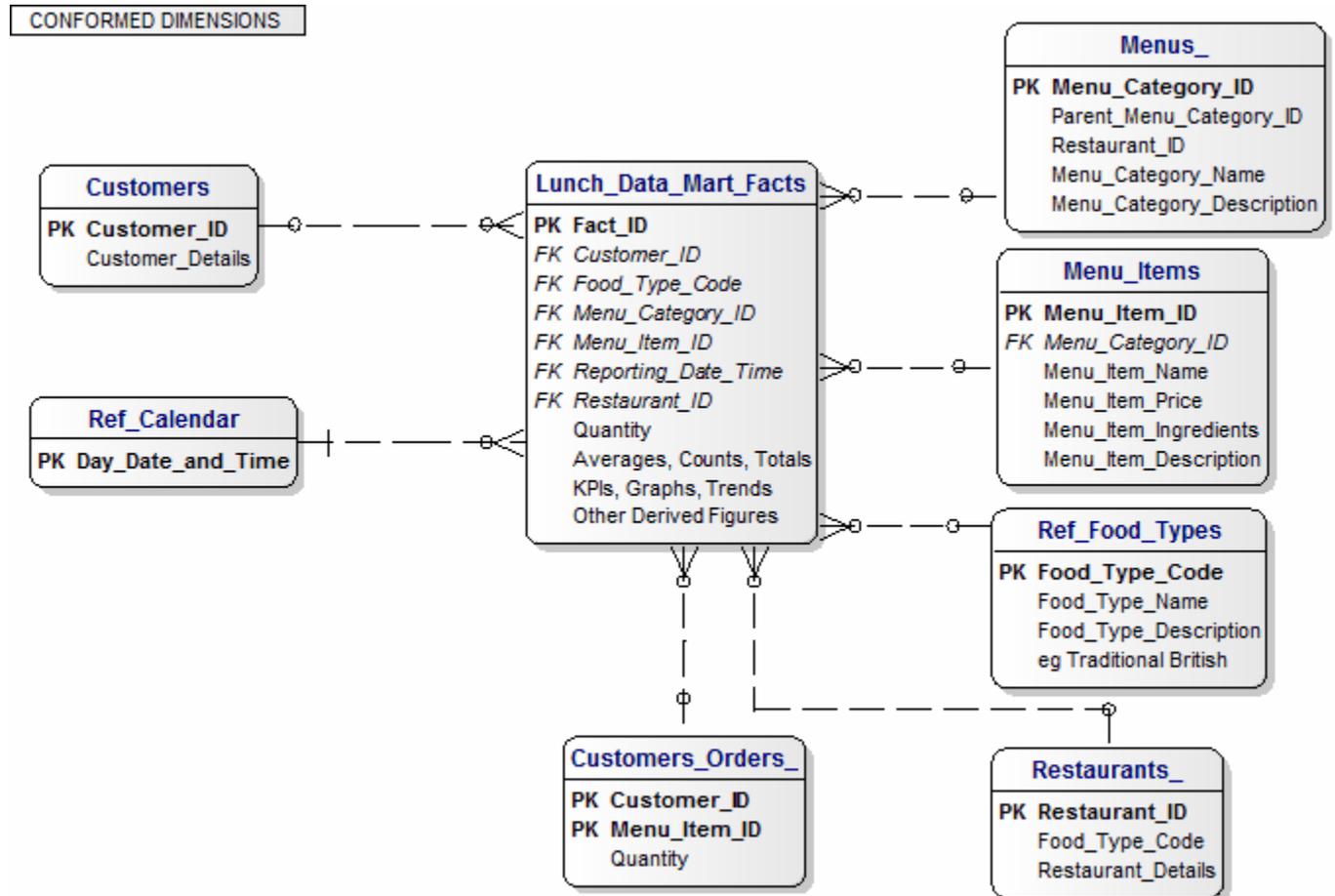
5.10.3 Data Warehouse

This shows the design of the Data Warehouse (DWH) after the second Event of Buying Lunch.



5.10.4 Data Mart

This shows the Data Mart for Restaurant data.



5.11 Event 3 – Watch the Judo Competition

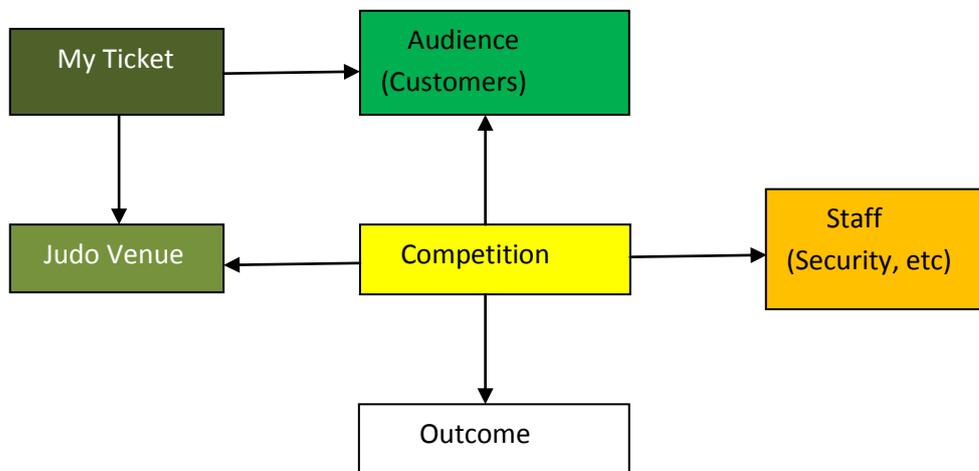
This shows how we handle the Third Event.

5.11.1 The Design Pattern

This shows how the Design Pattern applies to this Event.

In this case, the Event is the Competition between two Judo experts and the Outcome is very important

It is quite common for Event to have an Outcome, but so far, it has not been important enough to justify appearing at the top level.



5.11.2 Message Format

This shows the fields in the Generic Message :-

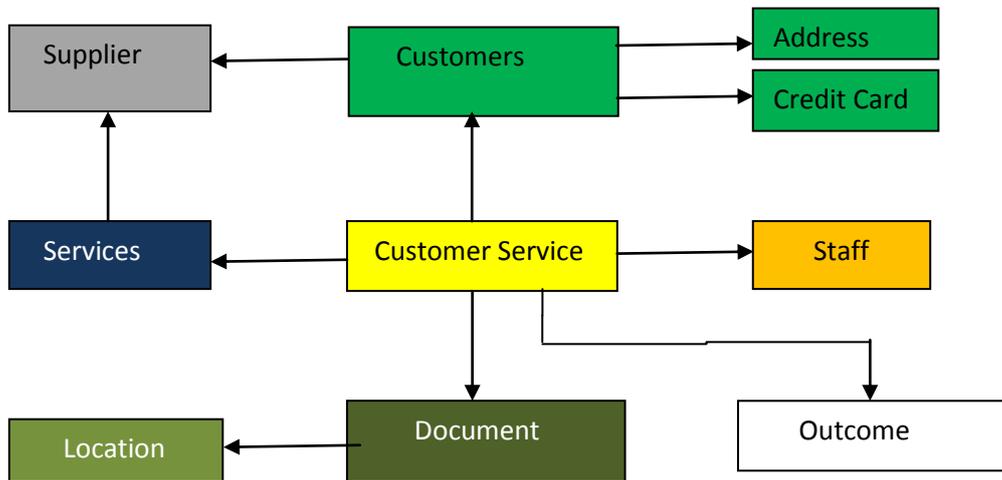
EVENT	DATE	LOCATION	PRICE	DETAILS

This shows the fields in the Message for this Event :-

EVENT	DATE	LOCATION	PRICE	DETAILS
Watch the Judo	Event Date	Judo Venue	Ticket Price	Outcome / Result

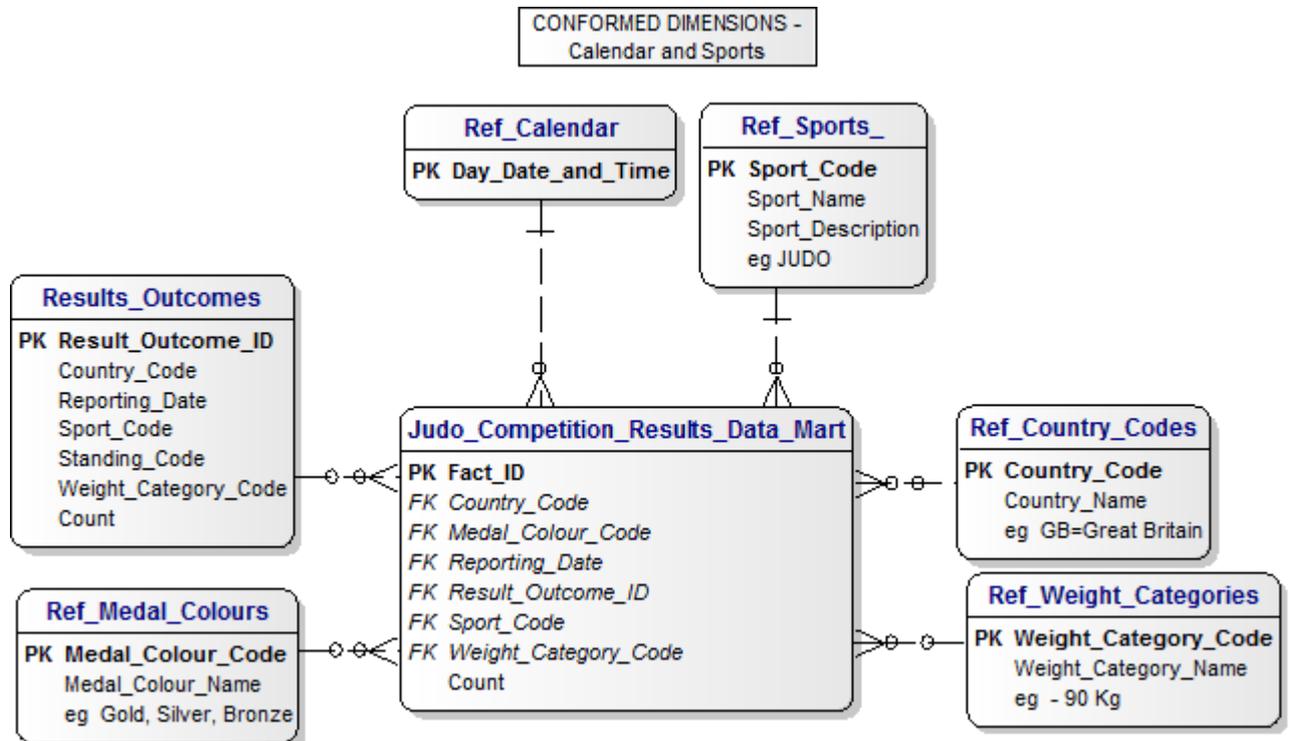
5.11.3 Data Warehouse

This shows the design of the Data Warehouse (DWH) after the third Event of watching the Judo Competition.



5.11.4 Data Mart

This shows the Data Mart for Judo Competition Results data.

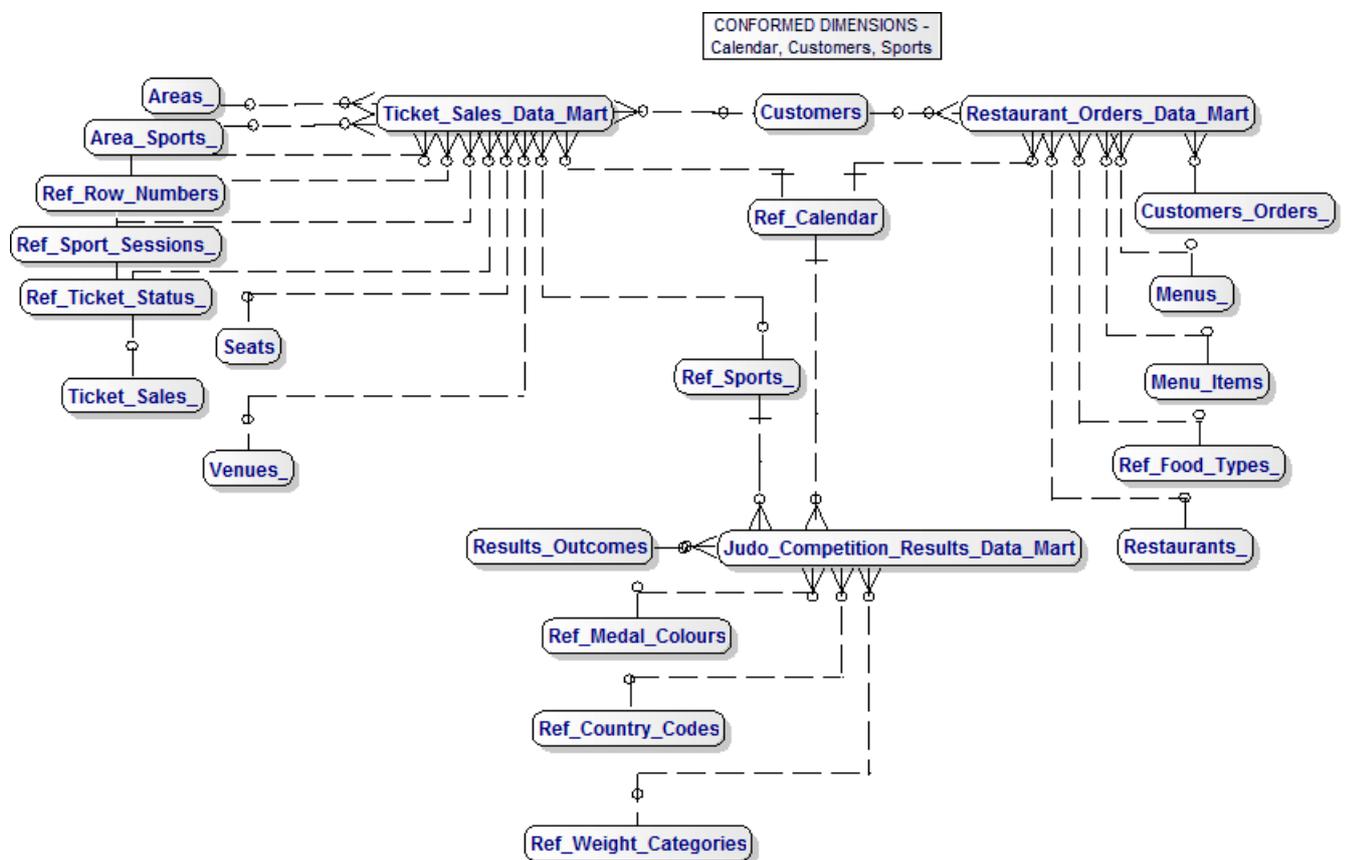


5.12 Combined Data Mart

This shows the three Data Marts :-

- Judo Competition Results
- Restaurant Orders
- Ticket Sales

These share Conformed Dimensions of the Calendar, Customers and Sports.



5.13 Business Intelligence

5.13.1 A BI Layer

The reason for all the work that we have done to get to this point is, of course, to produce Business Intelligence ('BI').

Here is a simple example to show how this works in practice.

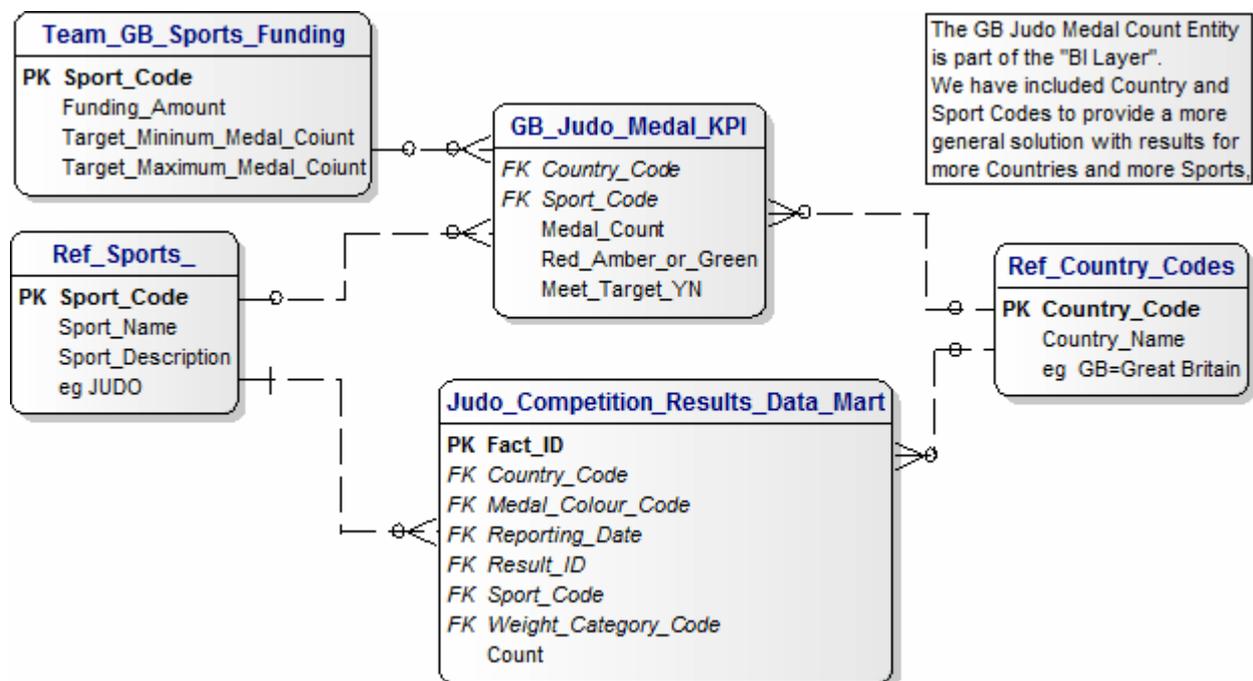
One of the key aspects of the Olympics was an analysis of the number of medals that each country won.

For Great Britain, funding was made available and targets were set for medals to be achieved by each sport.

For Judo, the funding was almost £8 million and the target was 1 or 2 medals.

In order to analyse this and produce the appropriate KPIs (Key Performance Indicators) we define a Bi Layer that takes data from the Judo Competition Results Data Mart.

This shows the Data Model for the Judo Medal Total BI Layer :-



5.13.2 BI Output

- This is a simple example of how this data could be displayed using a Green Traffic light :-

DISCIPLINE	BUDGET	TARGET	ACHIEVEMENT	TRAFFIC LIGHT (RED/AMBER/GREEN)
Archery				
Judo	£8 million	0-2 medals	2 medals	
Wrestling				

5.14 Conclusion

This Chapter has presented a Method for designing a Data Warehouse following a Canonical Data Model and Messages.

We have validated the Method by designing a Data Warehouse for a Day at the Olympics.

I would be pleased to have your comments and you can email me at barryw@databaseanswers.org.

6. Holiday in Malaysia

6.1 Management Summary

In this Chapter we use a trip to Malaysia to discuss an approach to the implementation of a Reference Data Architecture and the design of a Data Warehouse.

A **Canonical Data Model** (CDM) is central to this and we discuss the benefits of using Design Patterns based on a CDM.

During the trip, my wife and I stayed in three Hotels, hired a car and visited a number of Tourist Attractions, including an Elephant Sanctuary, a Crocodile Farm and an Underwater World in Langkawi Island in Malaysia.

After we returned to England I found myself thinking that the trip would provide a good opportunity to develop an interesting and 'User-Friendly' Tutorial on Data Warehouses.

The design of the Data Models reflects the scope and the fact that the overall aim is to provide data for Business Intelligence.

We also try to keep in mind that a well-designed Data Model should be good to look at and it should be possible to tell a story based on the Model.

6.1.1 The Approach

The Approach is to follow these Steps for each Event

Step 1 – Use the Event-driven Canonical Data Model from this page in our Web Site :

- http://www.databaseanswers.org/data_models/canonical_data_model/index.htm

Step 2 - Define a Message Format for each Event

Step 3 - Design a Data Warehouse and update it after we analyse each Event.

These Steps will be followed for these Events that occurred during our trip to Langkawi Island in Malaysia :-

- Event 1 - Check-in to a Hotel
- Event 2 - Hire a Car
- Event 3 - Go Shopping
- Event 4 - Visit Tourist Attractions
- Event 5 - Check-out of a Hotel

Elephants, Olympic Judo and Data Warehouses

For each Event, we discuss the Operational Data Source (ODS) in the form of the Message for the Event, along with the Data Warehouse and the Dimensional Model.

Our starting-point is defined by the Templates in Section 2.

6.1.2 Overview from ODS to BI

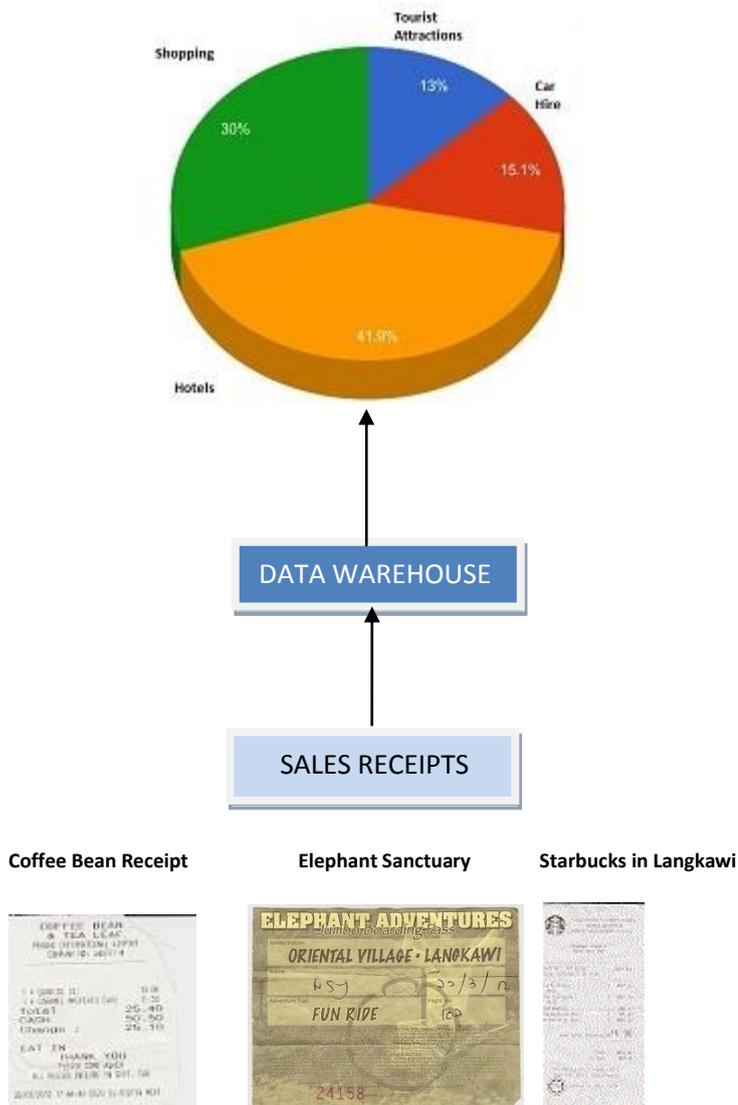
This Section presents Data Architecture to support a Data Warehouse and BI Layer.

A typical kind of display requested by users is a Pie-chart.

This example shows the expenses that we incurred during our trip to Malaysia – Car Hire, Hotel Accommodation, Shopping and Tourist Attractions

The source of the data is Tickets and Sales Receipts generated in Stores.

This data is stored in a Data Warehouse to provide a Single View of the Truth for the Pie-Charts in the BI Layer.

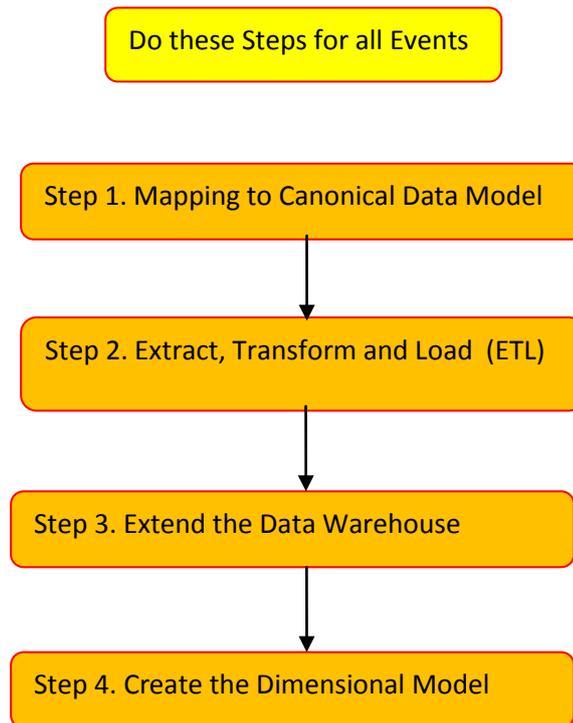


6.1.3 Slideshows

A short Slide Show has been created to give you an overview of how this approach works in practice and how the Canonical Data Model is central.

The Slide Show takes less than one minute :-

- http://www.databaseanswers.org/bmews_slideshow_book_tutorial/index.htm



6.2. The Theory

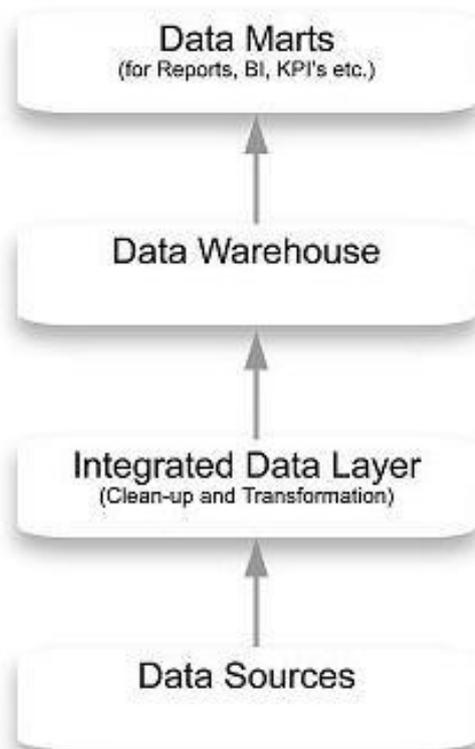
This discussion of the Theory in this Section will provide us with the background that we need to understand what follows in this book.

6.2.1 A Layered Data Architecture and SVOT

There are many different ways of defining a Data Architecture.

This diagram shows the four major Stages in a layered Architecture that is very economical but also very powerful.

The combination of a 3NF Data Warehouse with Master Data Management facilities in the Integrated Data Layer would help us in delivering a 'Single View of the Truth' ('SVOT').

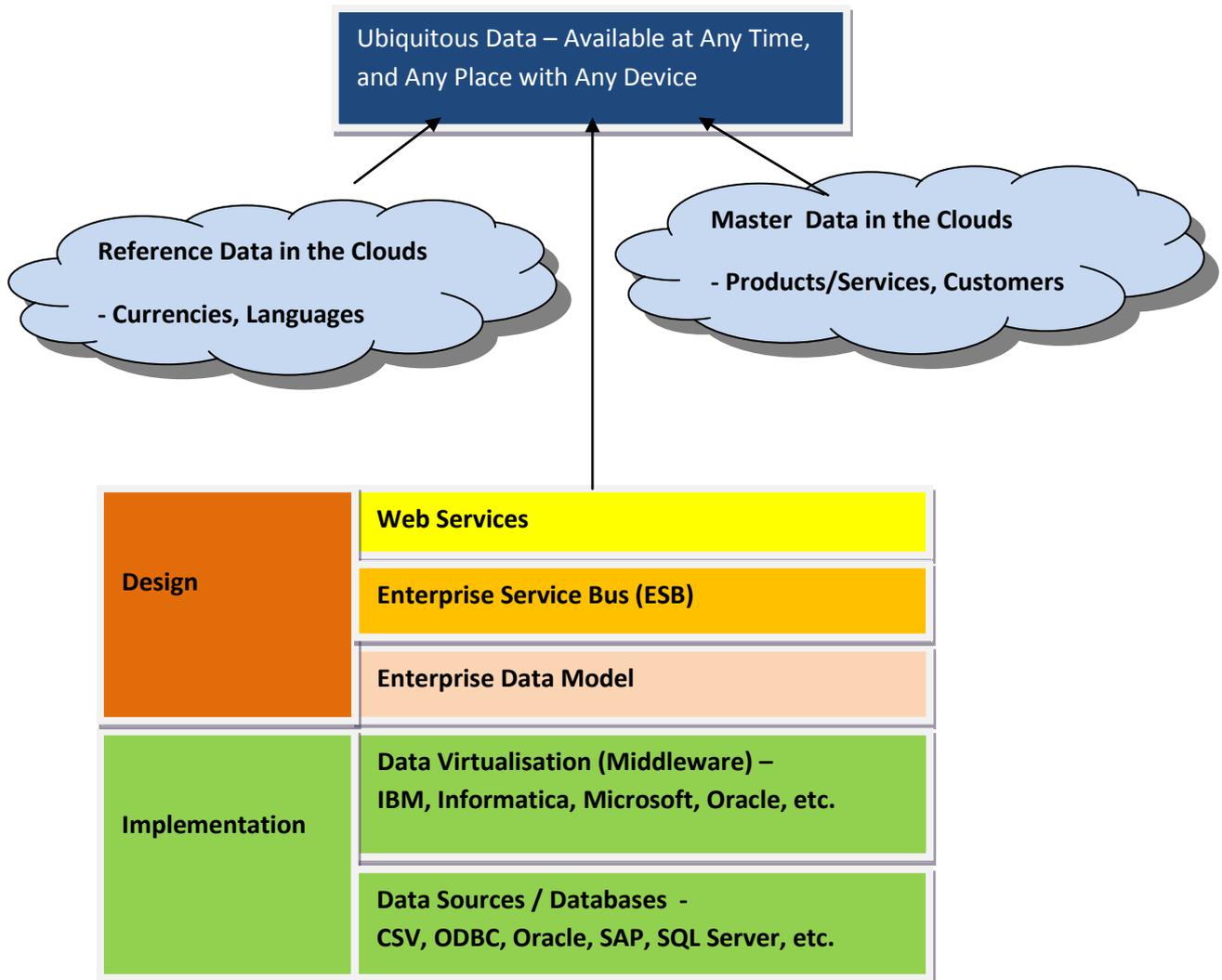


6.2.2 A Cloud-based Data Architecture for the Future

This Section describes an Architecture that can be used as a target 'To-Be' for the future.

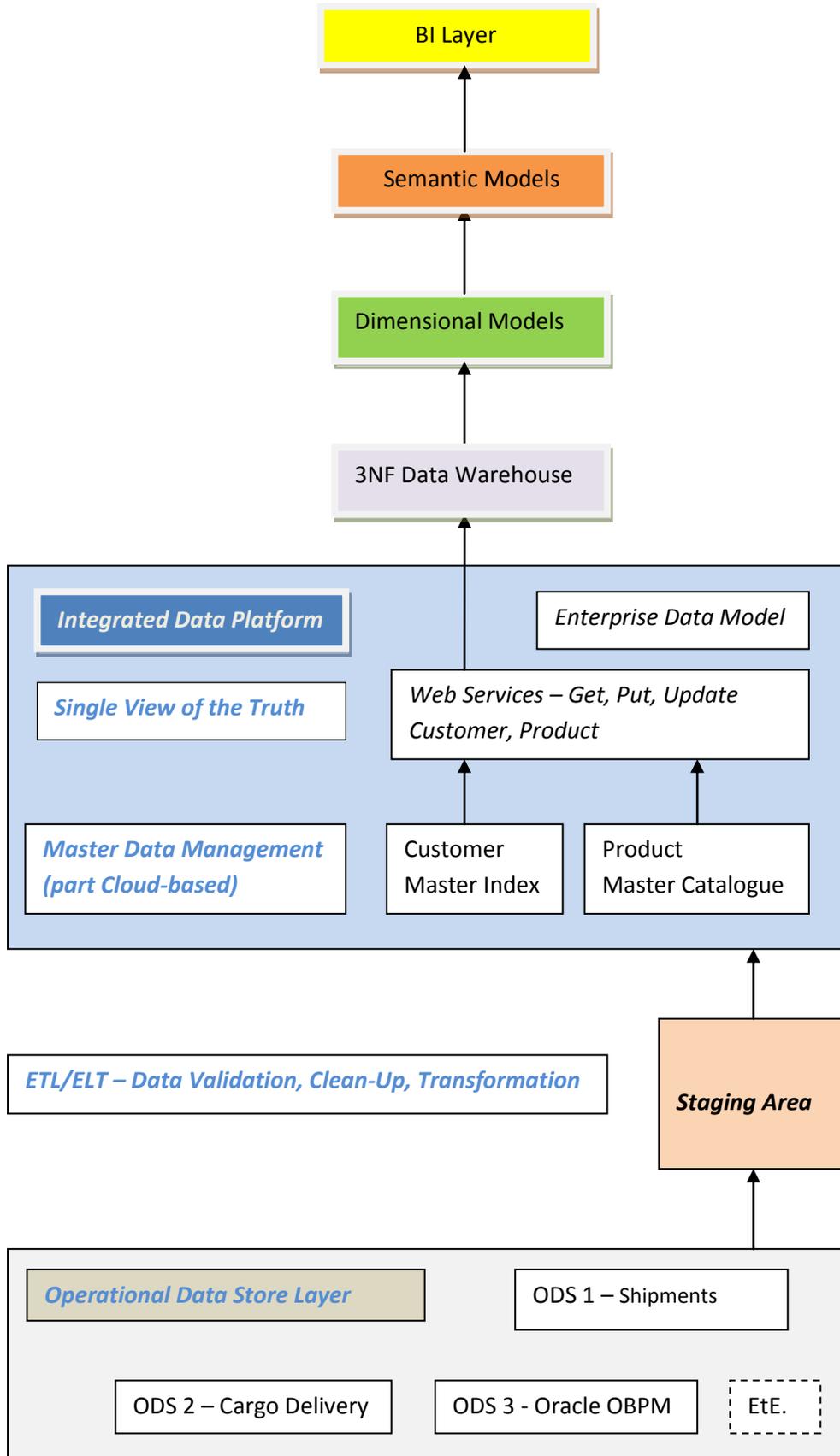
This diagram shows more clearly how Web Services would be used with an ESB providing an Abstraction Layer.

Data Virtualisation is a concept in Implementation that can be defined as 'the process of providing users with a Business View of the data in an organisation', which conceals the technical details of stored data, such as location, storage structure, APIs and storage technology.



6.2.3 Reference Data Architecture

The Integrated Data Platform is a specific example of a more general Data Virtualization Layer.



6.2.5 Canonical Data Model

The Canonical Data Model is used as a Template for a Design Pattern for an ERD for a Business Event.

This Model appears on this page on our Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

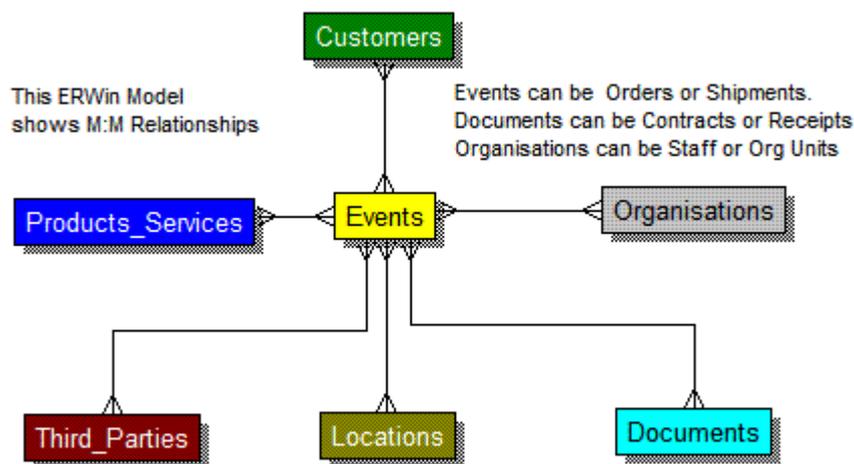
It provides a 'stripped-down' Event-oriented Model that applies to a wide range of business and everyday situations.

We use it as a standard to translate data into a common format suitable for loading into a Data Warehouse.

We have used ERWin for this Data Model.

This allows us to show Many-to-Many Relationships in a very concise and economical style.

When we come to use the CDM we will expand these into One-to-Many Relationships.



6.2.6 Mapping to the CDM

The purpose of this activity is to identify the correspondence between the Generic entities in the CDM with the specific entities for a particular Event.

This example shows how the Entities in the Data Model for our Hotel Check-In Event map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Hotel Check-In	COMMENTS
Customers	Guests	
Documents	Room Card / Key	This the nearest we get to a Document
Events	Hotel Check-In	
Locations	Hotel Address	
Organisations	Hotel Chains	
Organisations	Hotels	
Organisations	Staff in the Hotel	Staff can be involved and represent the Organisation
Products or Services	Room Reservation	
Third Parties	N/A	

6.2.7 Template Message Format

This is our Generic Template that is used in the derivation of all the specific Message Formats – for Car Hire, Shopping and so on.

Generic	Supplier	Date & Time	Customer	Products or Services	Unit Price	From Date	To Date	Total Price
Specific, eg Hotel Check-in	Eg Hotel	Date	Customer, Guest	Room,Meals, etc.				

6.2.8 Extract, Transform and Load (ETL)

At this point, we are interested in two aspects of ETL :-

1. Specification of Rules for Validation, Clean-up and Transformation
2. Definition of each item which is in common use and where a clear unique understanding is important.

Therefore, we create a table like this :-

Business Term	Validation	Validation	Clean-up	Transform
Date & Time	Cannot be in the future	Cannot be before date to be determined, eg 1950	Mark as bad data	Review with SMEs
Unit Price	Cannot be negative			
From Date	Less than or equal to 'To Date'			
To Date	Must be before date to be determined, eg 2020			
Total Price	Cannot be negative		Mark as bad data	Review with SMEs

A wide variety of software is available to meet ETL requirements.

A desirable feature is the ability to build up a library of ETL functions that can be enhanced as a result of learning in the course of a Data Quality Project.

This Library will then provide the foundation for the repetitive processing which is performed continuously until the data reaches the desired quality.

This will be determined by the business users and SMEs.

Although, as data management professionals, we might expect or want to have 100% good data, this is not always strictly necessary.

For example, in UK Local Government, it is common to find that addresses in Council Tax are of a uniform high quality, and that Dates of Birth are good in Social Services, because these two data items are critical to the respective functions.

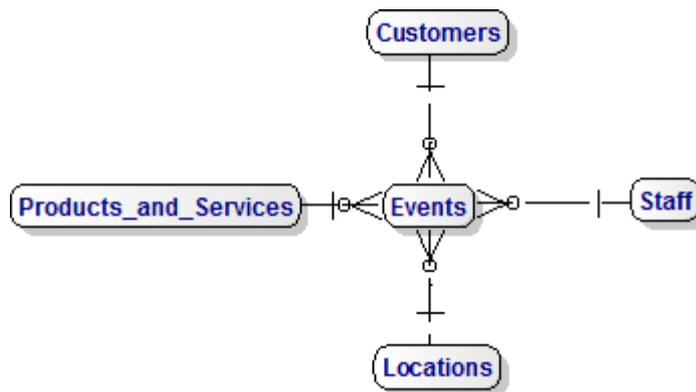
6.2.9 Template Data Warehouse

We show only the Entity names where it helps to understand the Model.

The Attribute are usually shown in the Dimensional Models.

During our analysis of the Events we did not find a place for the 'Third_Parties' Entity that is shown in our CDM in Section A.A.2.

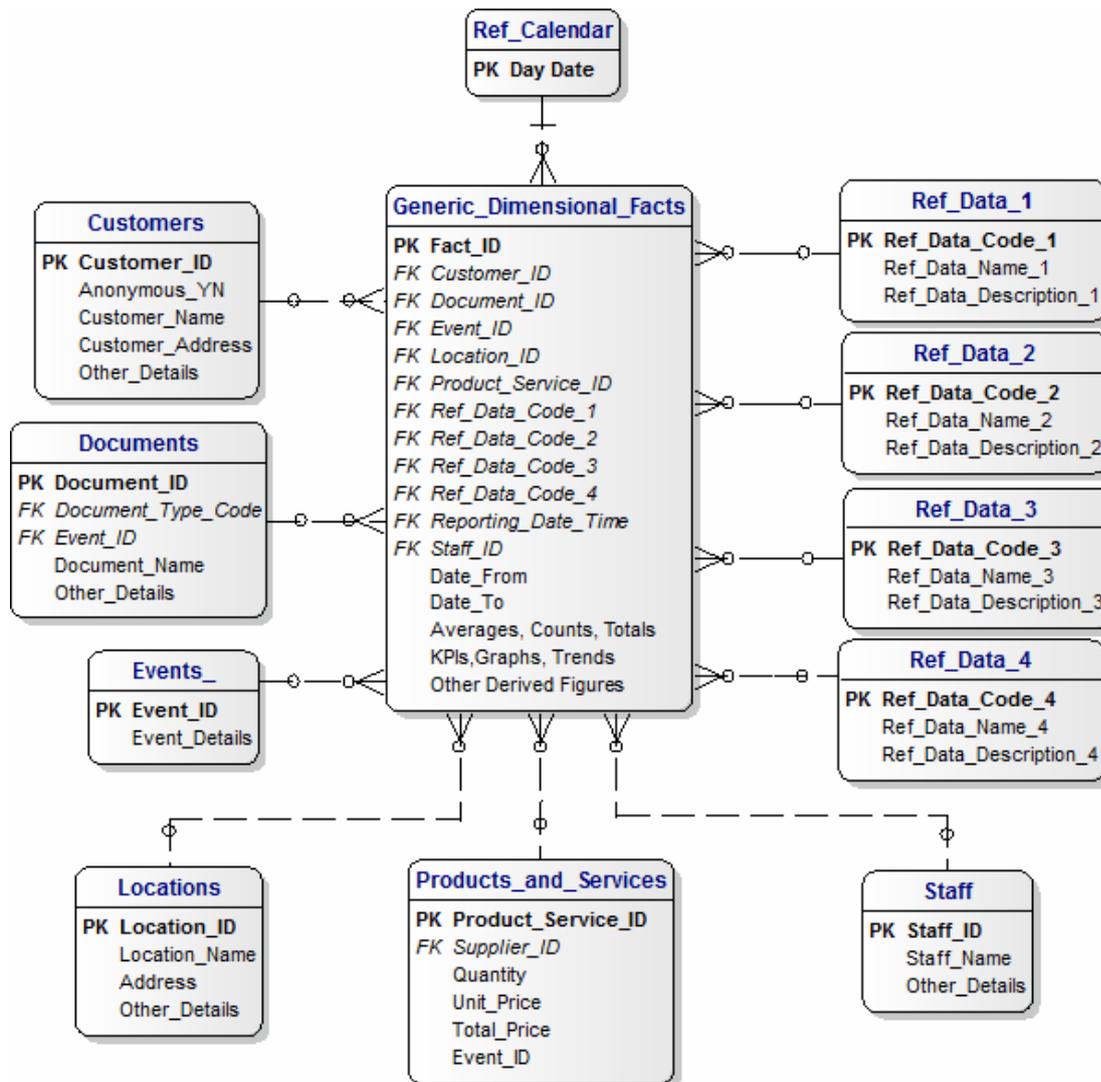
Therefore, we do not include in the future discussions in this Paper.



6.2.10 Template Dimensional Model

This is a Generic Template Dimensional Model which corresponds to our CDM shown above.

This will be an excellent starting-point for the analysis of the required Dimensional Models for the range of Events which make up a very important part of this document.



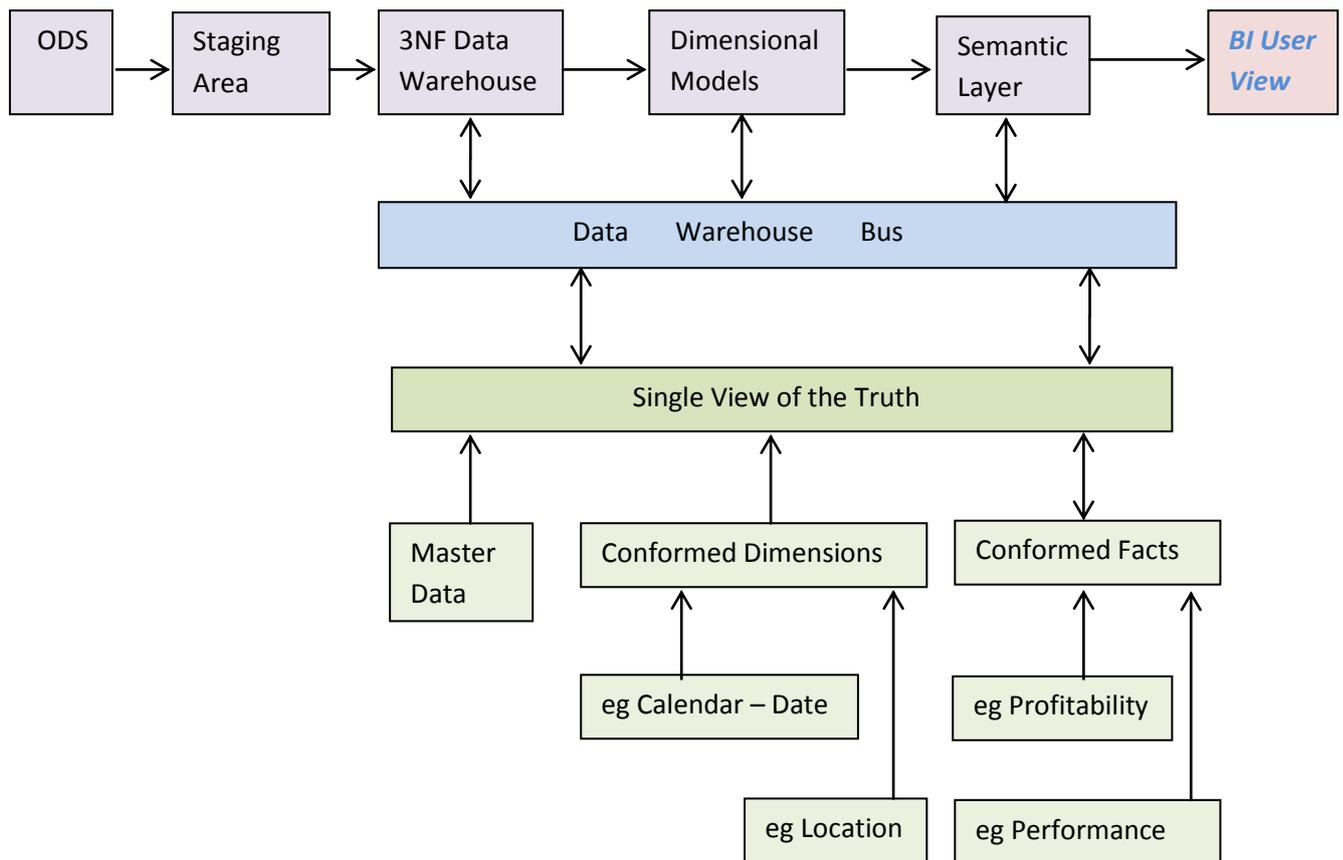
6.2.11 Data Warehouse Bus Architecture

Now, we show the complete Data Warehouse Bus Architecture

It helps us to think through how the Data Architecture will operate at run-time.

It provides a common Framework for establishing consistent standards for data and structures.

The top row shows the major Building Blocks in the migration of data from the Operational Data Stores to the Semantic Layer that defines the BI User View.



6.2.12 Master Data Management (MDM)

6.2.12.1 Introduction

The MDM Component provides a Single View of the truth for Customers and Products.

The principles behind the Single View of both the Customers and Products and Customers is the same :-

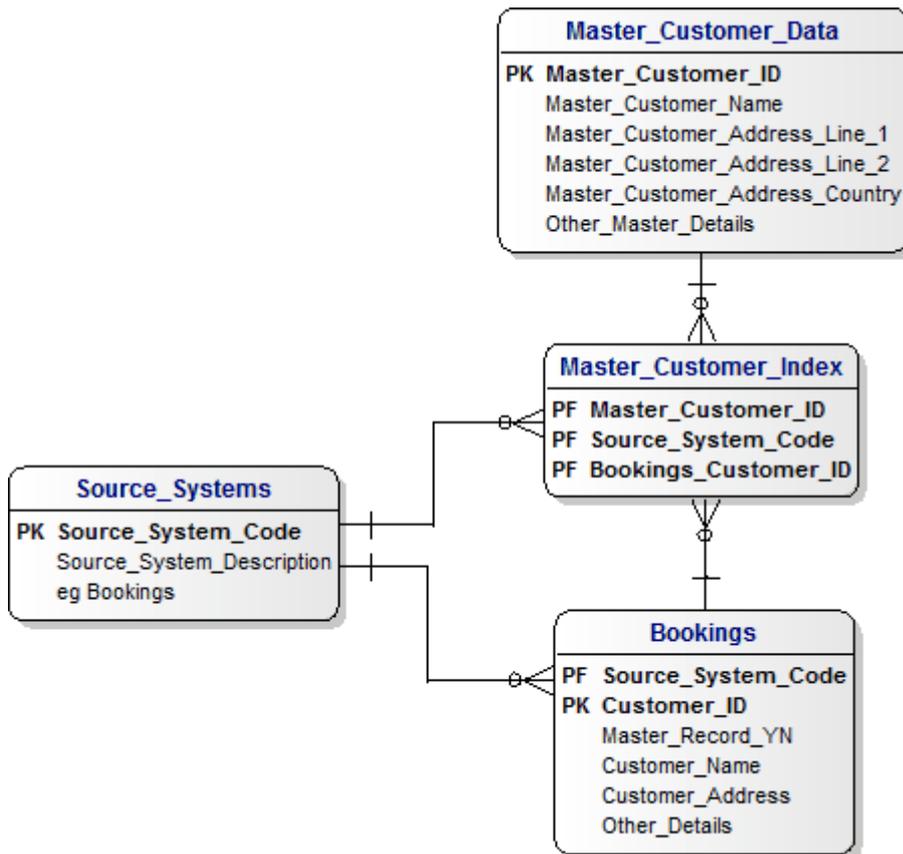
- Establish uniqueness using Business Rules
- Map from alternatives to the single record.
- If possible, define one Data Source as the 'Golden Record' or Master Record.
- For example, where Salesforce is in place, it is often provides the Master Customer record.

In this case, the 'GetCustomer' Service will try to match between the new characteristics, such as Customer name and address, and existing Customers.

6.2.12.2 Customers

A Customer Master Index (CMI) maintains a link between the single master Customer record and the Source Data Customers.

This Data Model shows how the approach above will be implemented in our Reference Data Architecture.



6.2.12.2.1 The Problem

For this Event, we must plan for loading Customer data into the Data Warehouse.

This requires that we establish a ‘Single View of the Customer’.

In the UK, the name ‘Joe Bloggs’ is used whereas in the States it would be John Doe’.

Joe, of course, is an abbreviation of Joseph.

On official documents, the name would be spelled Joseph, whereas in everyday conversation, it would normally be Joe.

Therefore we have to allow for the possibility that a ‘Joe Bloggs’ might be the same person as ‘Joseph Bloggs’

Elephants, Olympic Judo and Data Warehouses

In the States, we might find the names of John Doe, Johnny Doe and J.Doe and we would have to match them if they are the same person.

6.2.12.2 The Solution

Resolution of this Problem and the ambiguity requires us to define a set of Business Rules that can be run whenever we load a Customer who might be ambiguous.

Therefore, our solution to this problem of establishing a Single View of Customer Joe or Joseph is to have a 'Rules Engine' where we can define and execute a Rule like 'Joe is equivalent to Joseph'.

The recommended practice to implement the Customer Master Index is to use Web Services for Get, Update and Put facilities.

6.2.12.3 Products

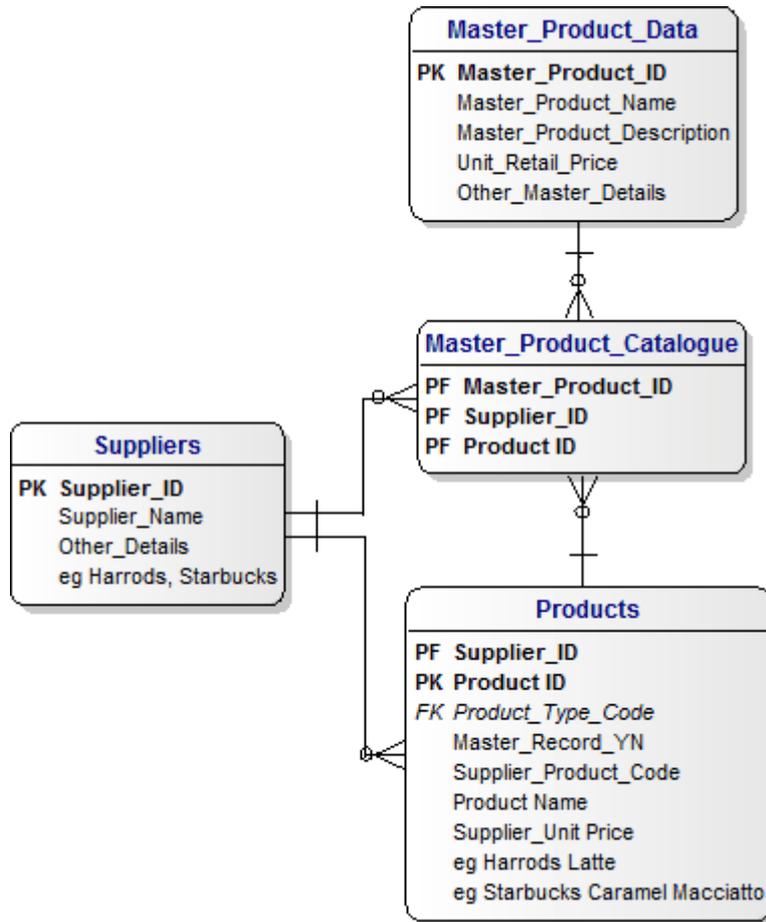
The requirements here are similar to the Master Customer Index.

We need to match similar products from different Suppliers should be considered the same Product for BI and Performance Reports. For example, Red Sweaters will have different Product Codes from different Suppliers but we will want to include them in the same Product Category that we define for our analysis purposes.

In this way, we establish a Single View of Products and Product Categories in the Data Warehouse.

This will be achieved by a combination of automated Business Rules and manual intervention by members of staff.

Elephants, Olympic Judo and Data Warehouses



6.3 The Practice

This Section discusses the Events that occurred on our holiday in Malaysia.

6.3.0 Arriving in Malaysia

We flew from London to Kuala Lumpur (KL).



Petronas Towers, KL, Malaysia

In the centre of this striking photo are the Petronas Towers in Kuala Lumpur, the capital of Malaysia, and commonly referred to simply as KL. At the time when they were built, they were the highest building in the world, and I believe they are still number two or three, with 80 floors each, and a footbridge on the 40th floor.

I walked across and it is not something I would want to do very often.

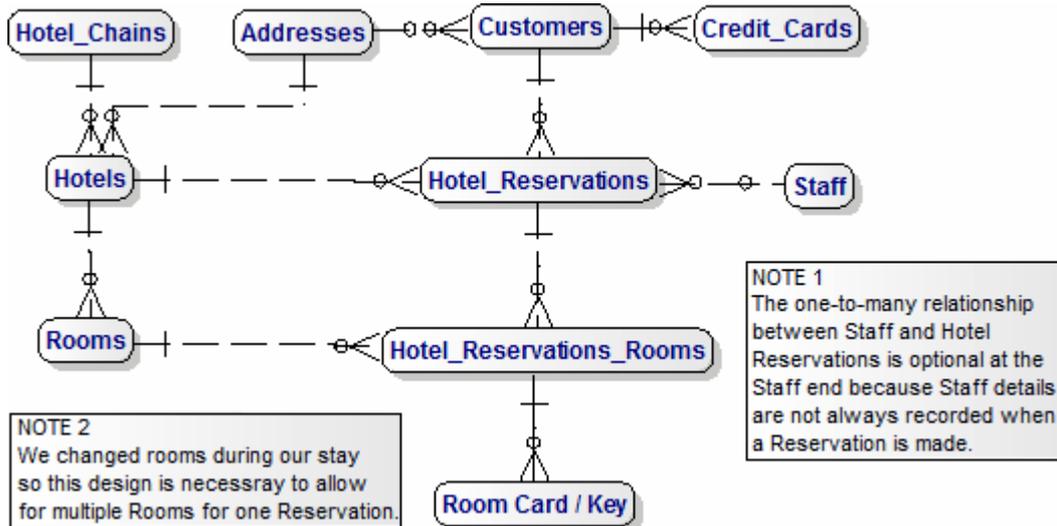
In passing, let me say that Petronas is a clever word because it is a combination of Petrol and Nasional, which is the word for National in the Malay language.

In other words, Petronas is the Malaysian equivalent to British Petroleum or BP, in the UK.

From KL we took an internal flight to Langkawi island where we checked in at the Berjaya Hotel and Resort.

For each Event, we check to see whether we can derive a Design Pattern based on our Canonical Data Model. If we can, then we have validated our CDM.

6.3.1.4 Data Warehouse for Hotel Check-In



6.4 Events 1 to 10

6.4.0 Adding new Data Sources

A very important activity that we have to plan for is adding new Data Sources to our Data Warehouse.

This is often as a direct response to a new business requirement.

The Steps that we will follow are those that we describe in the succession of Events in this Appendix.

6.4.1 Event 1 - Check-in to a Hotel

Checking-in to the Berjaya hotel in Langkawi is Event number 1.

As you can see, it is in a beautiful location and consists of over 400 wooden chalets that blend in very naturally with the forest.



6.4.1.1 The CDM Design Pattern

This Section discusses how the Canonical Data Model (CDM), shown earlier, applies to the Event of Checking-in to a Hotel. The CDM provides a Design Pattern for the Event-oriented Data Models that we need.

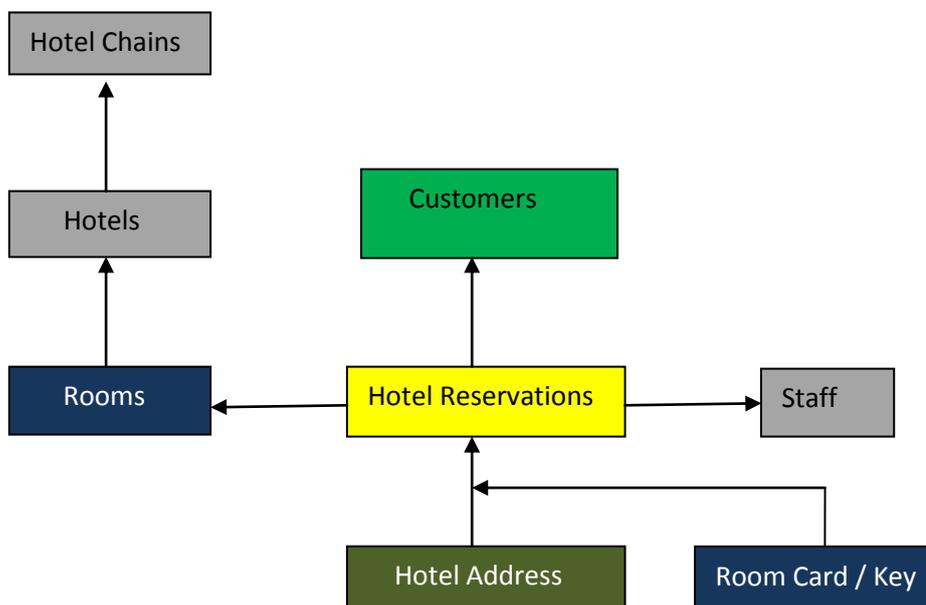
The Design Pattern based on the Hotel Check-in Event looks like this :-

This Section discusses how the Canonical Data Model (CDM) applies to the Event of Checking-in to a Hotel. The CDM provides a Design Pattern for the Event-oriented Data Models that we need.

When we think about the Check-in procedure, we realise that in a small transaction, we have obtained a rich source of material for a Data Model :-

- Customer name and Address
- Customer Credit card details
- Customer Passport details
- Hotel name and location
- Hotel Room
- Duration of Guest Stay

Using this source data, we can apply the Design Pattern to the Hotel Check-in Event and produce this Data Model as the result (arrows point from Children to Parents):-



Business Rules

It is very good practice to write out the Business Rules that define the conditions that the logic of the Model must comply with.

They can then be reviewed and agreed with a Subject Matter Expert (SME).

If it is appropriate, we can use a Business Rules Engine to automate the implementation of the Rules.

In this case, the Rules look like this :-

- A Customer has one and only one Address.
- A Customer has one or more Addresses.
- A Hotel belongs to one and only one Hotel Chain.
- A Hotel has one and only one Address.
- A Reservation is associated with one Customer.
- A Reservation is associated with one member of Staff.
- A Room belongs to one and only one Hotel.
- A Room Card or Key is associated with one and only one Room and Reservation

6.4.1.2 Mapping to the CDM

This Table shows how the Entities in our Hotel Check-In Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Hotel Check-In	COMMENTS
Customers	Guests	
Documents	Room Card / Key	This is as near as we get to a document in a Hotel Check-in
Events	Rent a Room	
Locations	Hotel Address	
Organisations	Hotel	
Organisations	Hotel Chains	
Organisations	Staff in the Office	Staff represent the Organisation and are usually involved during check-in.
Products or Services	Room for Rent	
Third Parties	N/A	

6.4.1.3 Message Format

This shows how the Generic Message Template applies to the Hotel Check-In Event.

It defines the Source Data for this Event.

TBD stands for 'To Be Determined'.

Supplier	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
Berjaya Hotel	Check-in Date & Time	Barry's Name and Credit Card Details	Room number	Price per Night	From Date	To Date	TBD

6.4.1.4 Data Warehouse for Hotel Check-In

At this point, the Data Warehouse will contain data only for the Hotel Check-in Event and therefore it will look like the CDM.

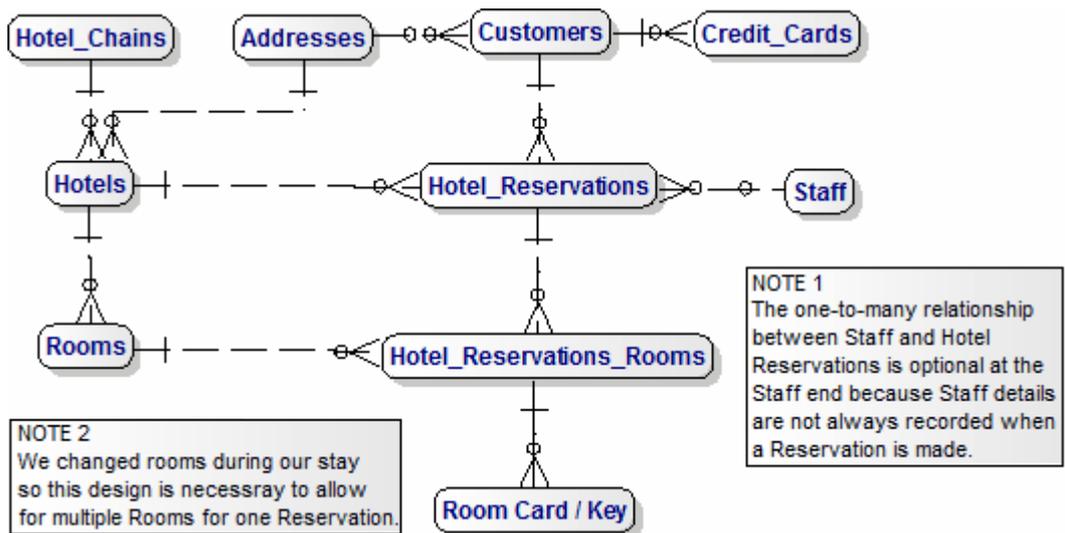
This diagram shows the Entities in the Data Model.

The attributes are shown in the Dimensional Model in the next Section.

In this diagram, we have positioned the Room Card/Key so that it corresponds to the Document Entity in the CDM.

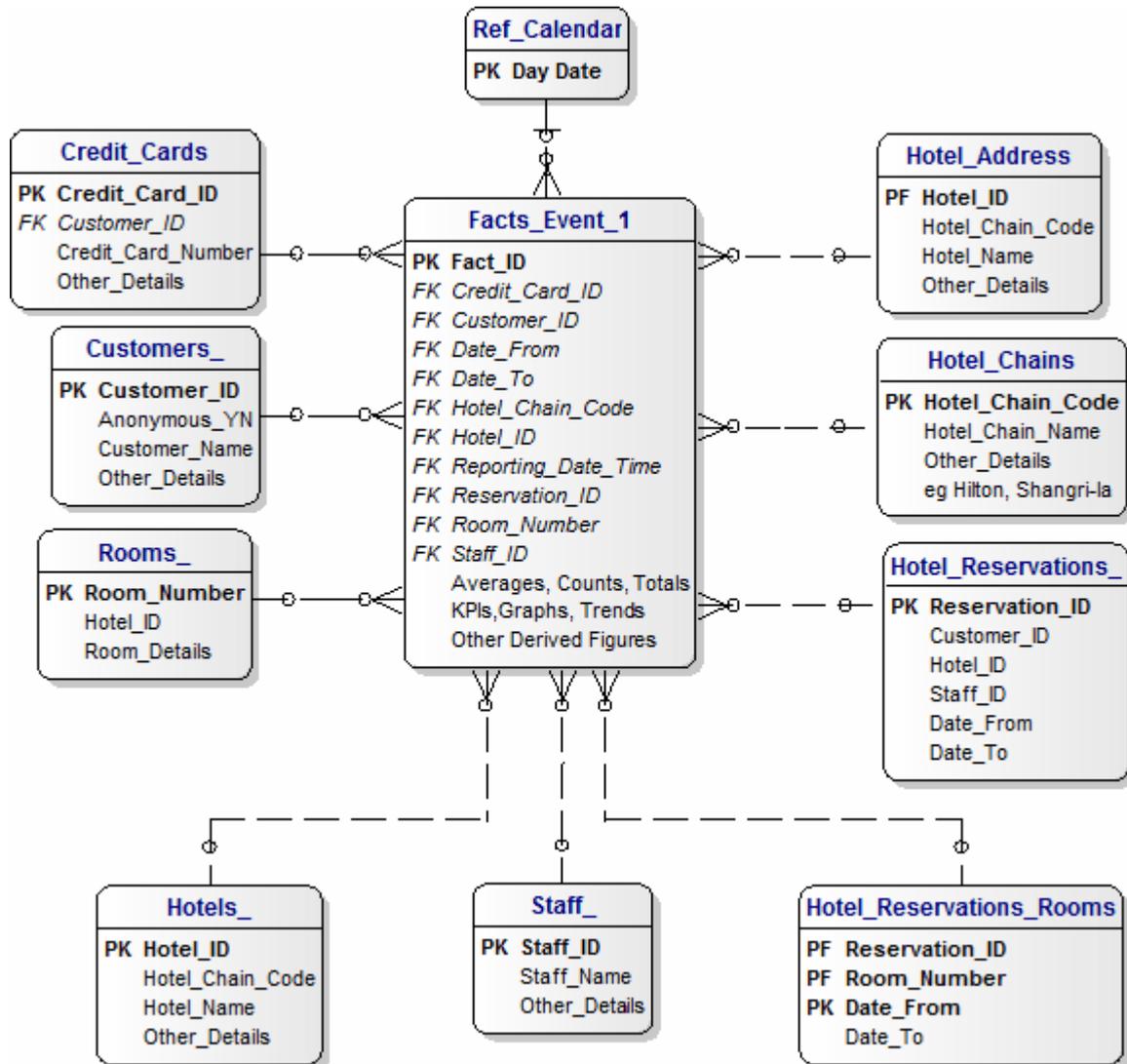
We have also shown the Hotel_Address entity with its correct relationship to the Hotel entity, whereas in the CDM, it is shown, for convenience sake, in a Many-to-Many Relationship with the Events entity.

This is the same as Third-Normal Form Data Warehouse as the one shown earlier in this Section.



6.4.1.5 Dimensional Model

Derived data must not appear in an ERD, therefore the Room Card/Key does not appear in this ERD because the data is derived from data already recorded.

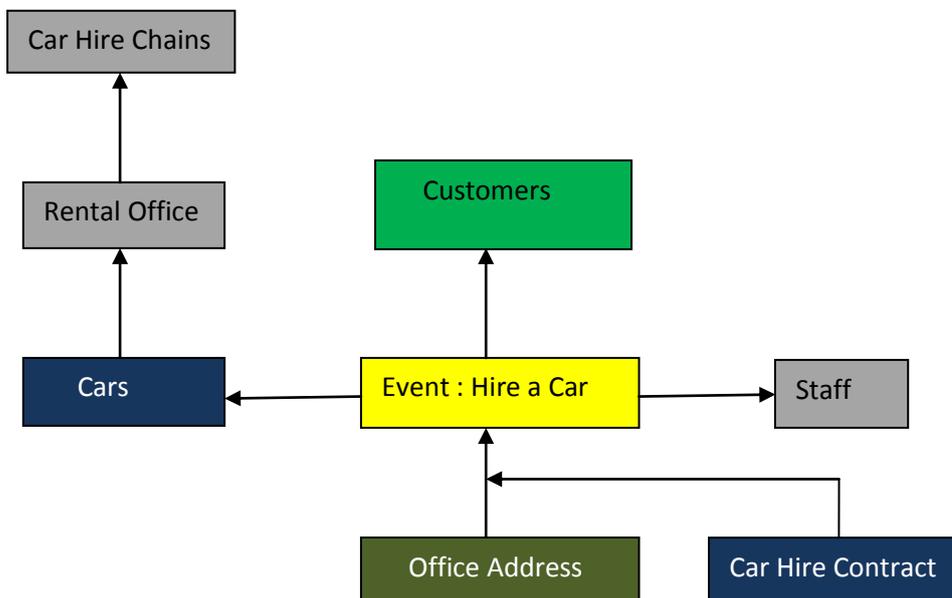


6.4.2 Event 2 - Hire a Car

6.4.2.1 The CDM Design Pattern

This Section discusses how the Canonical Data Model (CDM), shown in Section 2.1, applies to the Event of Hiring a Car. The CDM provides a Design Pattern for the Event-oriented Data Models that we need.

The Design Pattern based on the Car Hire Event looks like this :-



6.4.2.2 Mapping to the CDM

This Section discusses how the Canonical Data Model applies to the Event of Hiring a Car.

We hired a car from a local Car Hire company at the airport when we landed on Langkawi island.

We would normally create a Subject Area Model for Cars, to show details such as Car Make and Model.

This Table shows how the Entities in our Car Hire Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Hire a Car	COMMENTS

Elephants, Olympic Judo and Data Warehouses

Customers	Customers	
Documents	Car Hire Contract	
Event	Car Hire a Car	
Locations	Car Hire Office Address	
Organisations	Car Hire Company	Staff are usually involved and represent the Organisations in the CDM
Products or Services	Car Hire	
Third Parties	N/A	

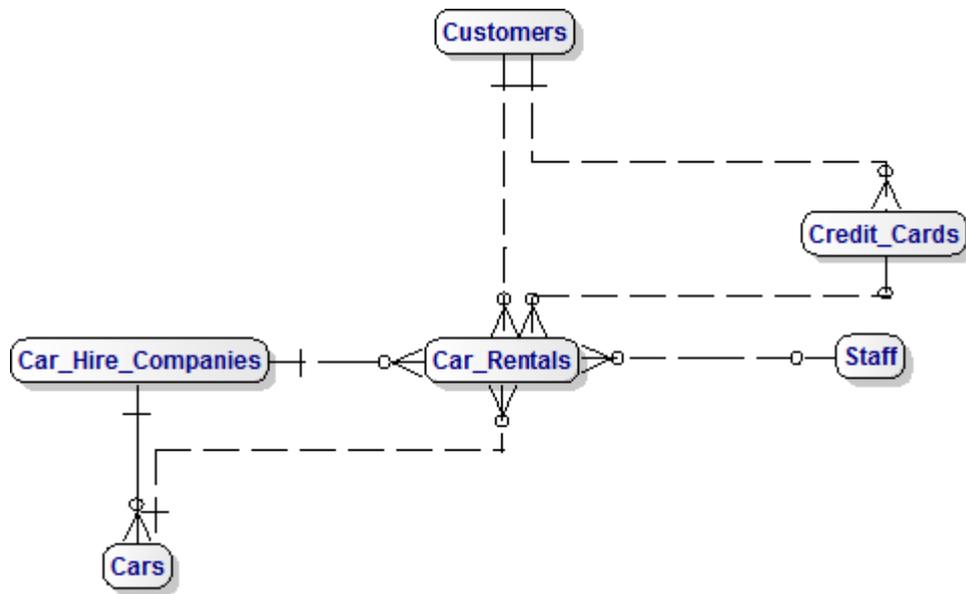
6.4.2.3 Message Format

We hired a car from a local rental company in the airport in Langkawi, which worked out very well.

Car Hire Company	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
ABC Car Hires	Check-in Date & Time	Barry's Name and Credit Card Details	Car Reg Number	Rental charge per Day	From Date	To Date	---

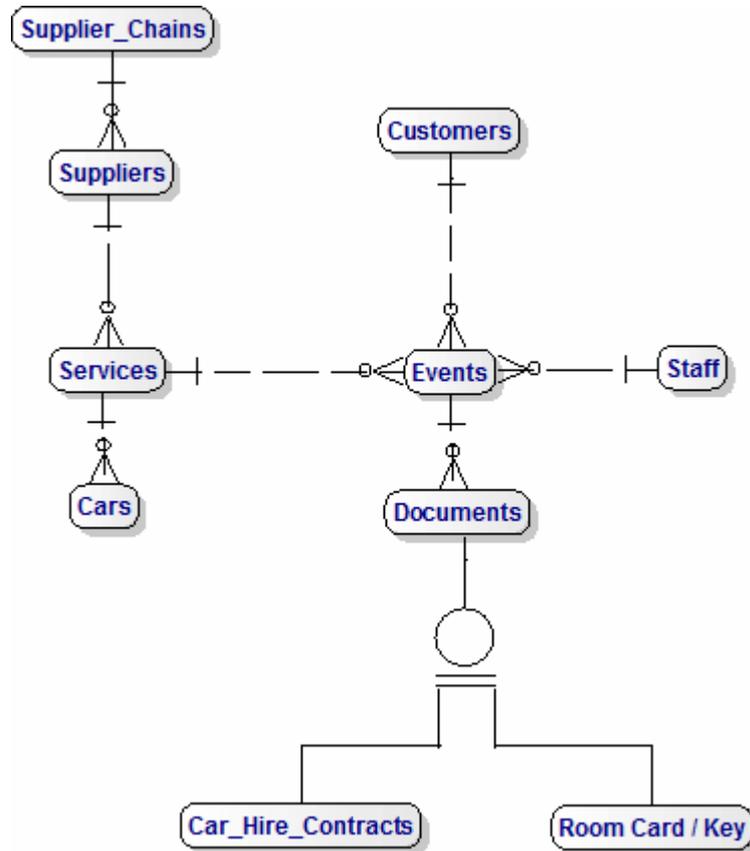
6.4.2.4 Data Warehouse for Car Hire

At this point, we add the Car Hire data to the data for the Hotel Check-in Event which is already in the Data Warehouse.



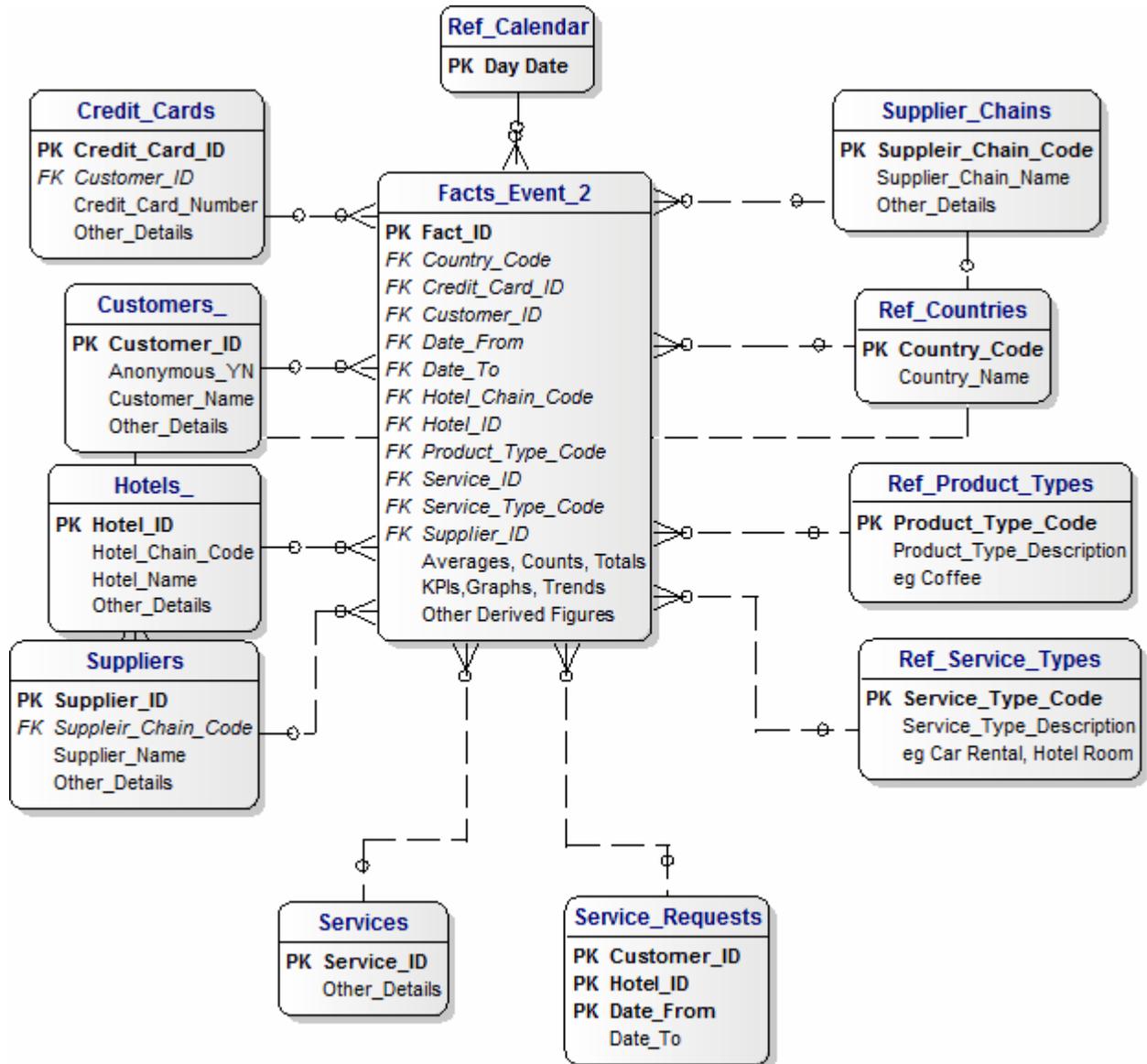
6.4.2.5 Consolidated Data Warehouse

After a little thought, we have combined Car Hire and Hotel Reservations into Suppliers and Services.



6.4.2.6 Dimensional Model

This is Event 2 where we add the Dimensions and Facts for Hotel Check-in to the existing ones for Car Hire.



6.4.3 Event 3 - Go Shopping

Harrods Store in the Airport at Kuala Lumpur, Malaysia.



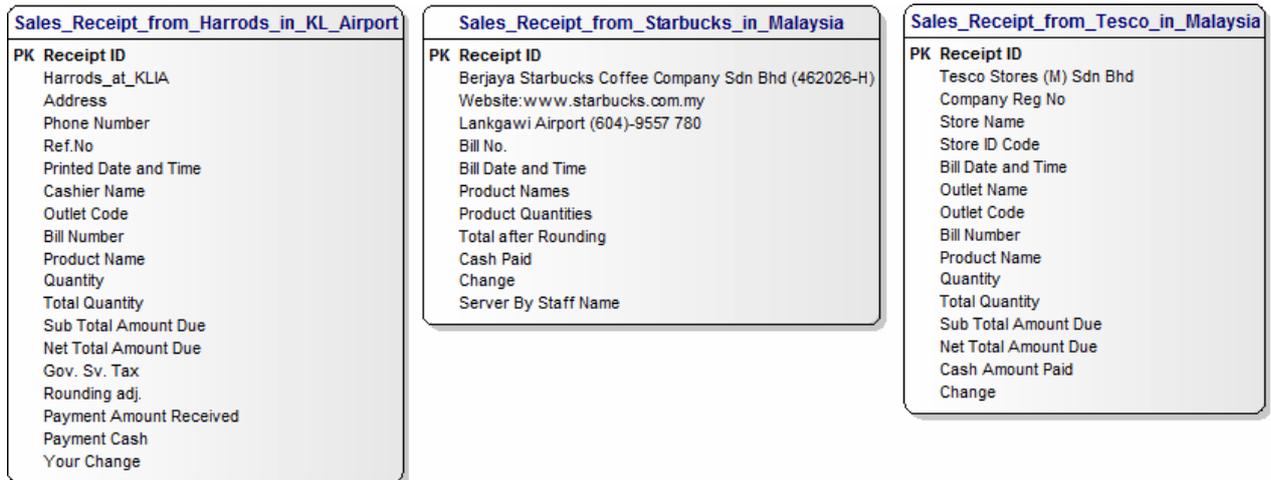
Harrods is very popular in Malaysia, as you can tell from the customers browsing in the store.

6.4.3.0 Discussion

This Section discusses Sales Receipts, both Generic and Specific.

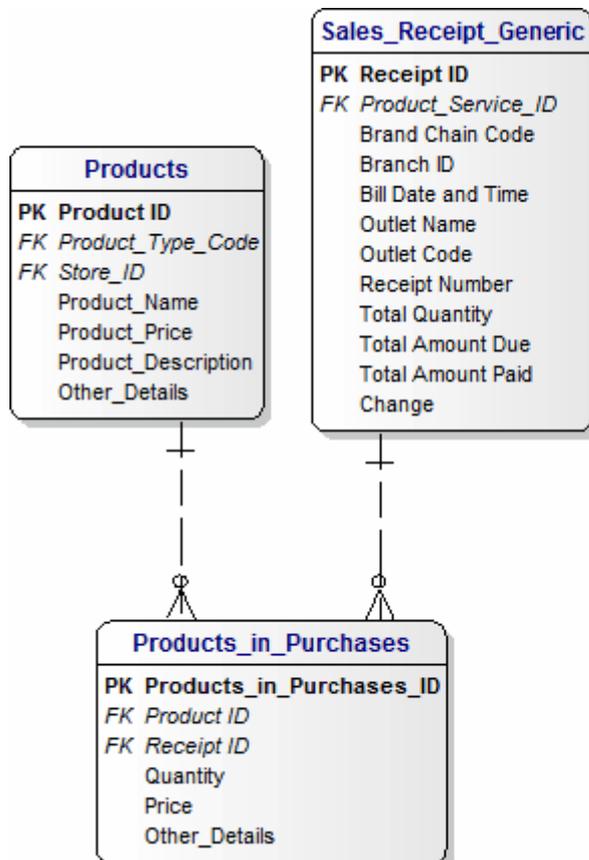
6.4.3.0.1 Specific Sales Receipts

Here we have Data Models for Receipts from Harrods, Starbucks and Tesco in Malaysia.



6.4.3.0.2 Generic Sales Receipts

This shows a consolidated Receipt that provides a generic view of the three specific examples above.



6.4.3.1 The CDM Design Pattern

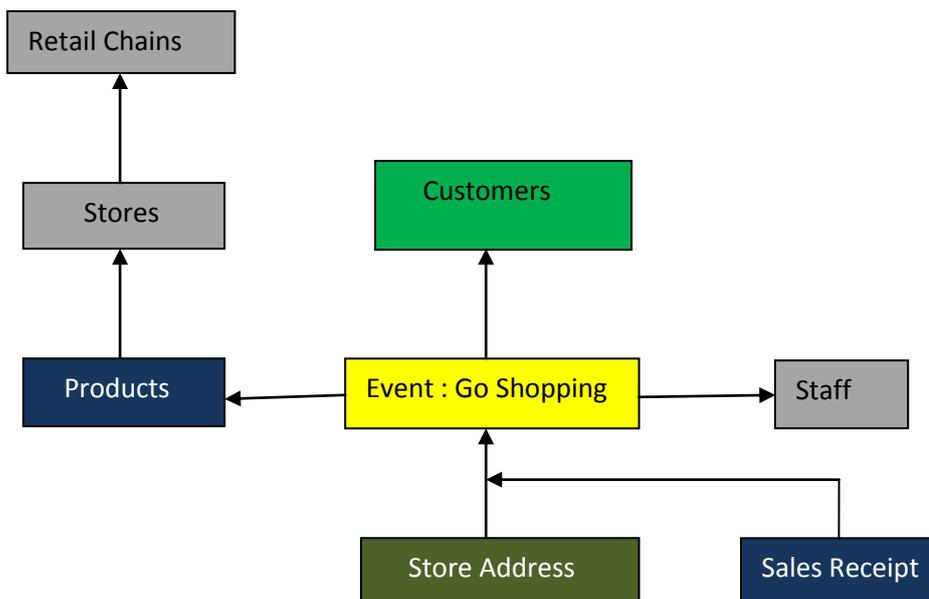
This Section discusses how the Canonical Data Model applies to the Shopping Event

We go Shopping – which is when the long-suffering husband says one of three things :-

1. “It’s a tough job but someone’s got to do it”
2. “When the going gets tough, the tough go shopping”
3. “Yes, dear”

But usually, we survive the experience ;-0)

This is how the CDM Design Pattern applies to the Shopping Event :-



6.4.3.2 Mapping to the CDM

This Section discusses how the Canonical Data Model applies to the Event of Shopping.

We went shopping at a number of stores in Malaysia.

We would normally create a Subject Area Model for Shopping.

This Table shows how the Entities in our Shopping Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Go Shopping	COMMENTS
Customers	Customers	
Documents	Sales Receipt	
Events	Go Shopping	
Locations	Stores	
Organisations	Staff, Stores and Store Chains	For example, Harrods and Tesco
Products or Services	Retail Products	
Third Parties	N/A	

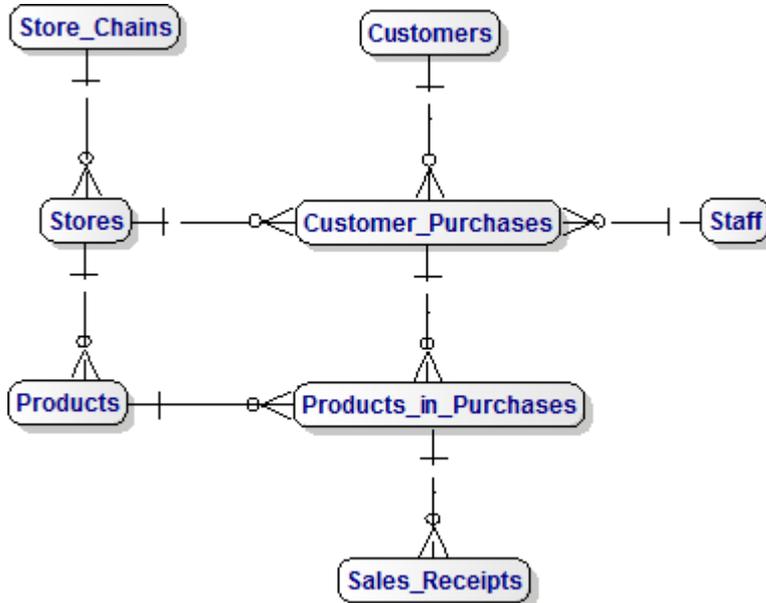
6.4.3.3 Message Format

The Message Format for this Event will resemble the Sales Receipt.

Store Name	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
Harrods	Visit Date & Time	Cash (Anonymous) or Barry's Name and Credit Card Details	One or many Products	Purchase Price	N/A	N/A	To be calculated

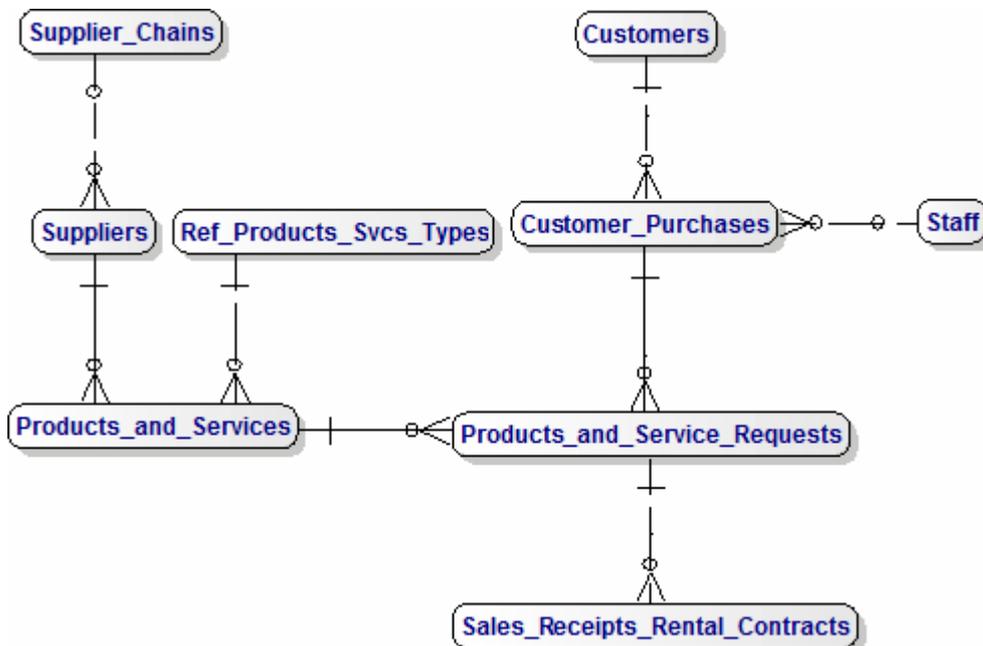
6.4.3.4 Data Warehouse for Shopping

This shows the Data for the Shopping Event :-



6.4.3.5 Consolidated Data Warehouse

At this point, we add the Shopping data for Event 3 to the data for the Car Hire and Hotel Check-in Events which is already in the Data Warehouse, so our design looks like this :-

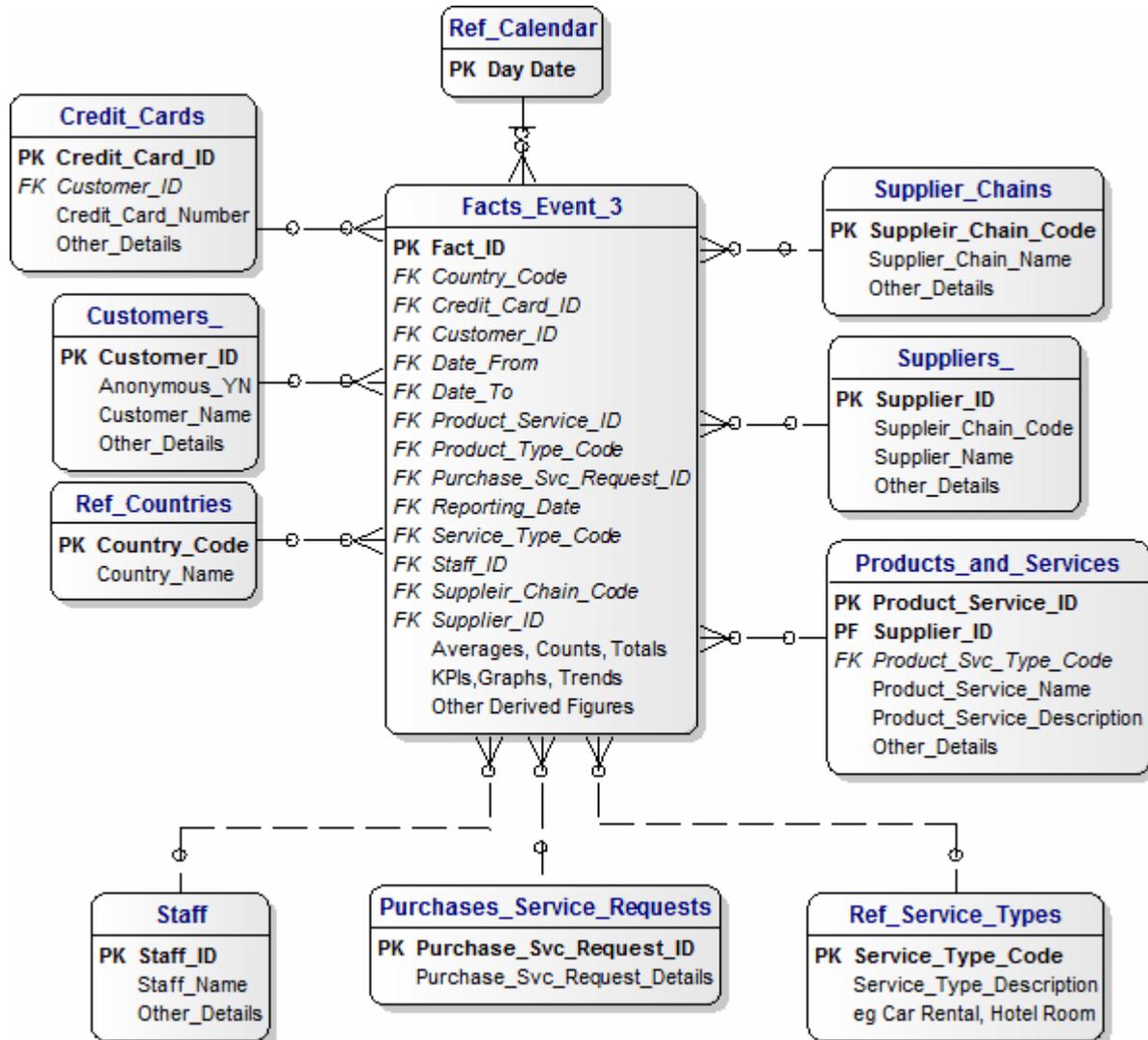


At this point, we would normally consider creating a Glossary of Terms to establish agreed definitions of the word that are in common use.

Elephants, Olympic Judo and Data Warehouses

6.4.3.6 Dimensional Model

The Dimensional Model will have data for Shopping, Car Hire and Hotel Reservations.



6.4.4 Event 4 - Visit the Elephant Sanctuary

The Elephant Sanctuary was the first Tourist Attraction that we visited.

Here we see an Elephant family in the Kuala Gandah Sanctuary.



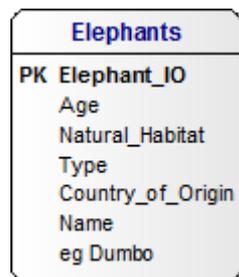
6.4.4.0 Discussion

6.4.4.0.1 Elephants

Elephants, especially in small numbers that you see in a Sanctuary or a Circus, frequently have names and we often know their ages.

However, this is not true for crocodiles.

Therefore, we store names and ages for elephants but not for crocodiles so here is the Data Model for Elephants :-



Elephants, Olympic Judo and Data Warehouses

6.4.4.1 CDM Design Pattern

This Section discusses how the Canonical Data Model applies to the Event of Visiting an Elephant Sanctuary.

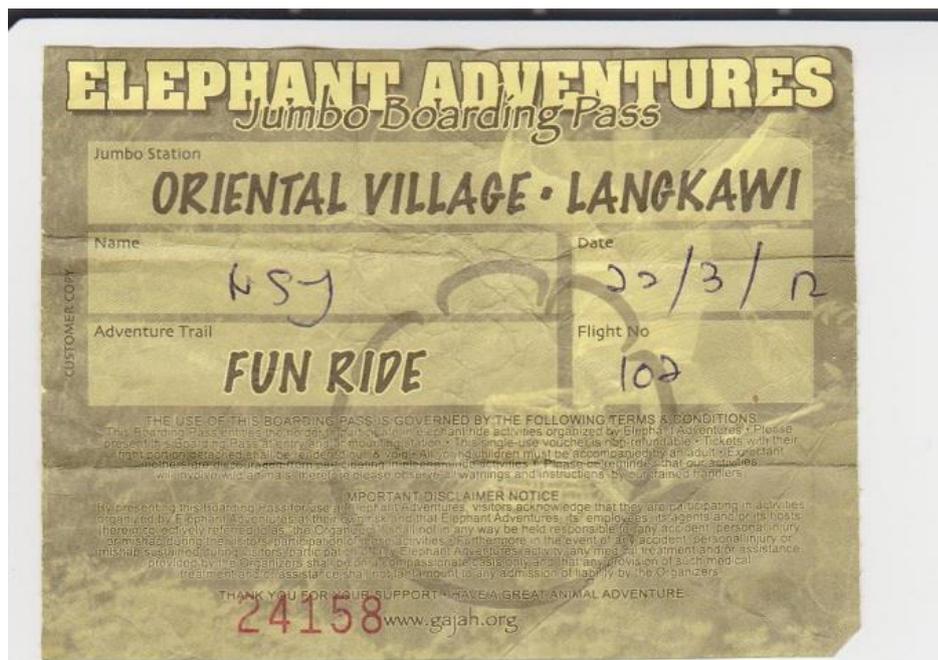
We were on Langkawi island where there are a lot of interesting things to see and do.

My wife voted for a trip to Elephant Sanctuary because she thinks baby Elephants are very cute.

So we decided on the Elephant Sanctuary, then the Crocodile Farm and finally the Underwater World.

They have an overhead aquarium and I have always wanted to see fish going over my head, and it had a number of individual Attractions, including the Fish Aquarium and the Penguin Area.

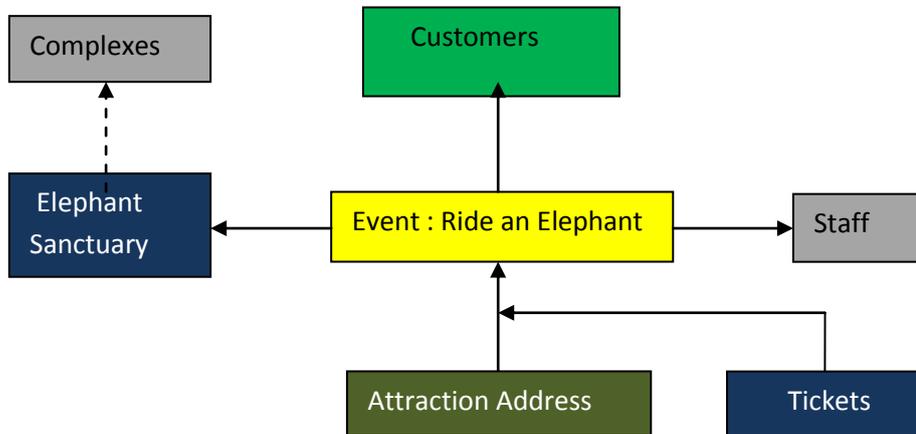
Here is our ticket for the Elephant Ride (called a “Dumbo” Boarding Pass !!!) :-



In our Canonical Data Model (“CDM”) this is an example of a Document related to an Event.

In other words, this is an example of how we are able to validate our CDM.

So we started out trip with a visit to the Elephant Sanctuary :-



6.4.4.2 Mapping to the CDM

This Section discusses how the Design Pattern for Visiting a Tourist Attraction maps to the Canonical Data Model applies to the Event of Visiting a Tourist Attraction.

If we need to include more detail we would probably create a Subject Area Model for Tourist Attraction.

This Table shows how the Entities in our Tourist Attraction Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Visit to Elephants	COMMENTS
Customers	Tourists	
Documents	Tickets	
Events	Visit to Elephant Sanctuary	An Elephant Sanctuary is an example of a Tourist Attraction
Locations	Address of Elephant Sanctuary	
Organisations	Elephant Sanctuary Owners	
Products or Services	Take a Ride on an Elephant	
Third Parties	N/A	

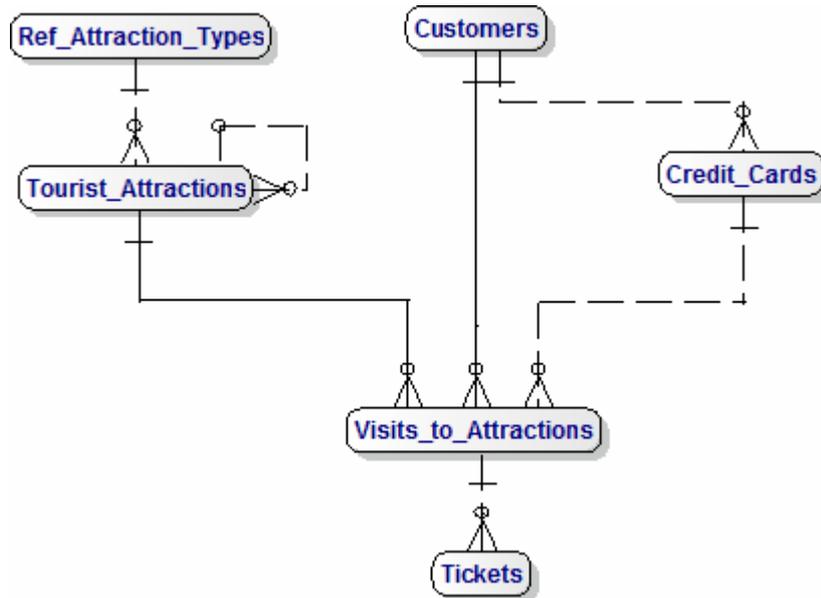
6.4.4.3 Message Format

The Message Format will resemble the Tickets.

Attraction Name	Date & Time	Customer Details	Services	Unit Price	From Date	To Date	Total Price
eg Elephant Sanctuary	Visit Date & Time	N/A	Attraction	Entry Fee	N/A	N/A	As determined

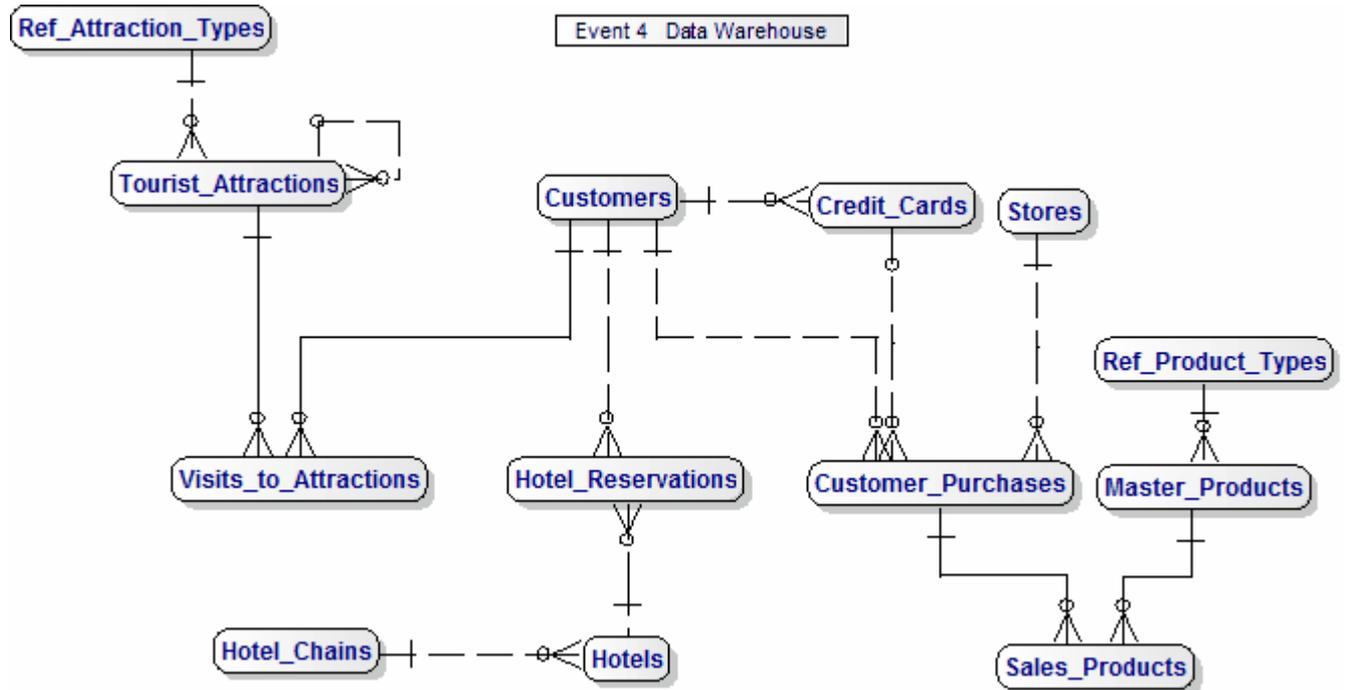
6.4.4.4 Data Warehouse for the Elephant Sanctuary

This models the Elephant Sanctuary as a Tourist Attraction.



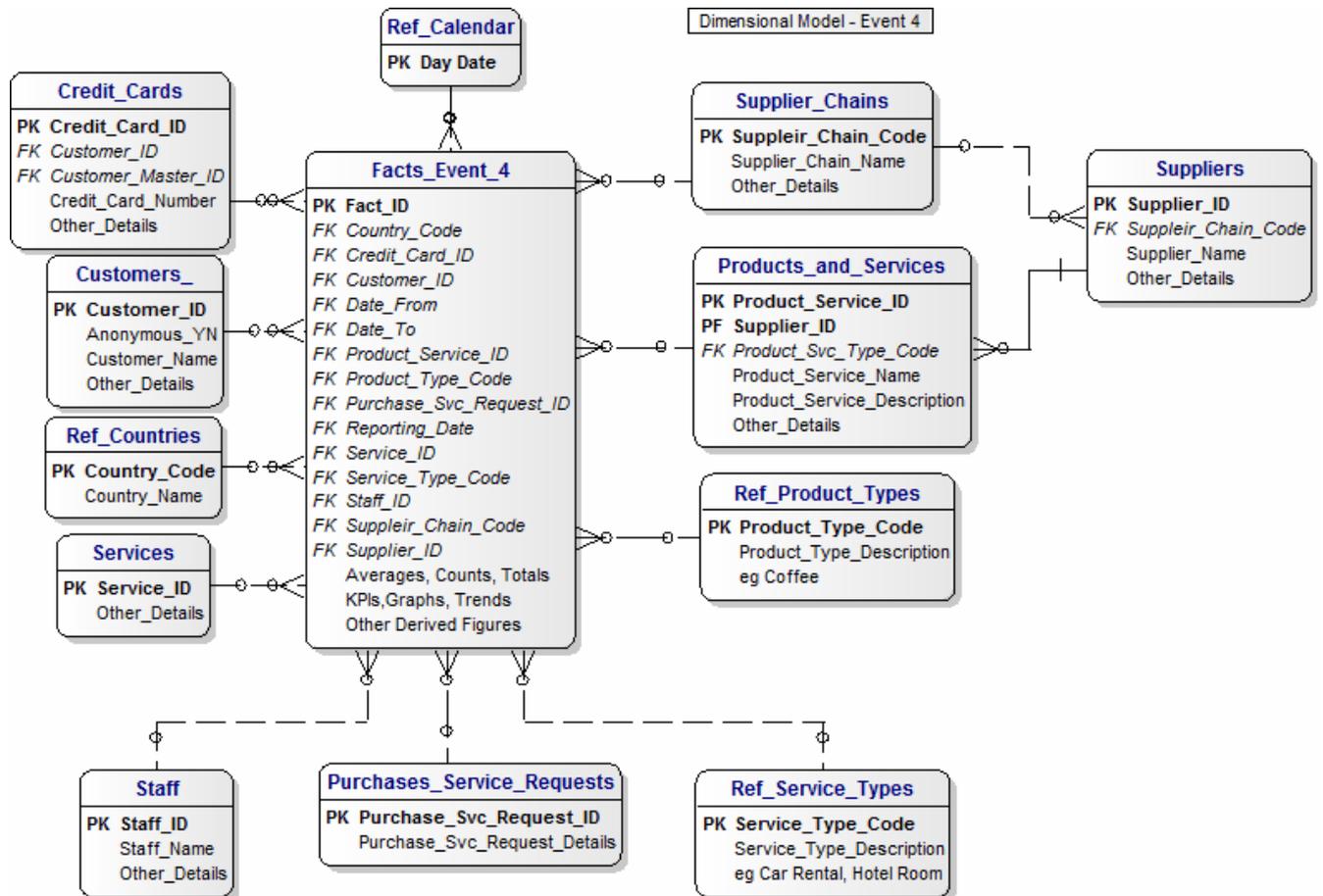
6.4.4.5 Consolidated Data Warehouse

This consolidates the Tourist Attraction Entities with the existing Entities for Car Hire, Hotel Check-in and Shopping.



6.4.4.6 Dimensional Model

At this point, data for Event 4 – Elephant Sanctuary (Tourist Attractions) – is added to data for Cars, Hotels and Shopping.



6.4.5 Event 5 - Visit the Crocodile Farm

I was keen to go to the Crocodile Farm, because I enjoyed the scene in the James Bond “Live and Let Die” movie where he (Roger Moore) escapes from a small island surrounded by crocodiles by jumping across their backs.

I have not been able to find exactly the scene that stuck in my mind, so here is the nearest, which are the crocodiles shown on a movie still :-



In Langkawi we were very impressed to see a brave guy sitting on the back of a crocodile.

Later we found that he was an employee and somehow he had trained the crocodile to let him sit on its back.

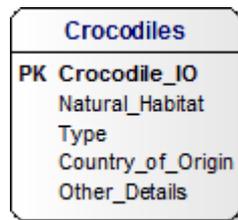


6.4.5.0 Discussion

This Section discusses some of the implications for Data Modeling of combining a Visit to the Crocodile Farm with a Visit to the Elephant Sanctuary.

6.4.5.0.1 Adding Crocodiles to Elephant Data Model

Here is the Data Model for a Crocodile.



We can see that the Crocodile Entity looks very similar to the Elephant Entity.

The only difference is that we often know the name and age for an Elephant because they are somehow more 'user-friendly' than Crocodiles.

We never know the age and name of a Crocodile !!!

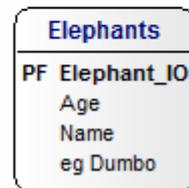
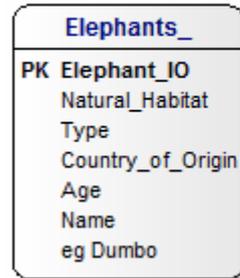
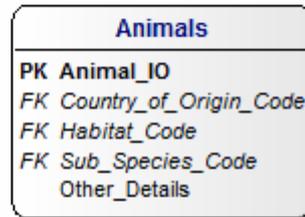
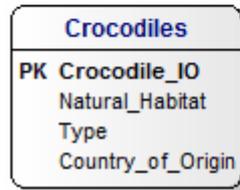
When we try to produce a combined Model for both elephants and crocodiles this is our first draft.

Elephants, Olympic Judo and Data Warehouses

Step 1 : Crocodiles, Elephants and Animals.

This is the first Step to representing a Data Modeller's view of Crocodiles and Elephants.

This shows that we have created an *Animals* Entity that has the attributes that Crocodiles and Elephants have in common. Then we have a different *Elephants* Entity that has only the attributes that are specific to an Elephant. These are Age and Name. Of course, the Age will change every year so when we develop an Entity to be used professionally, we would replace Age by Date of Birth

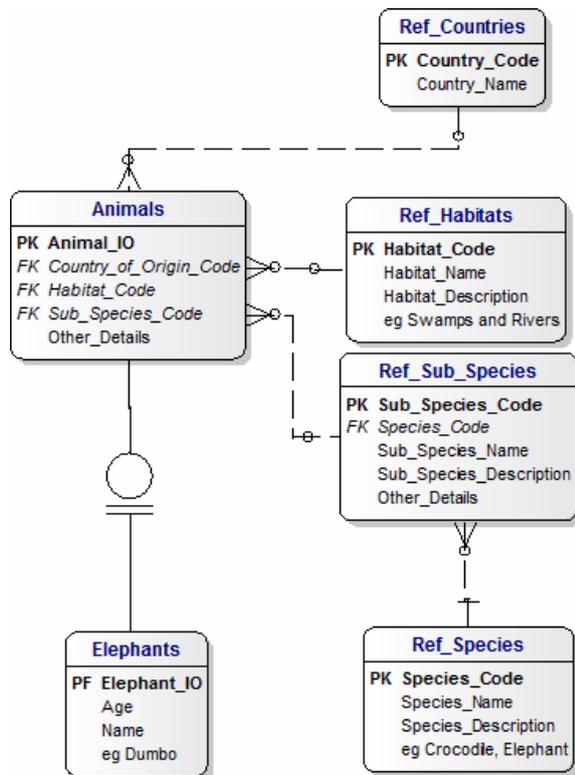


Step 2 : Crocodiles, Elephants and Animals (part 2)

In this first draft, we use the word Animals to mean both elephants and crocodiles.

At this point, we introduce the concept of Inheritance.

This is shown in our Data Model by the small circle and it means that Elephants inherit the characteristics of Animals and in addition, they have Ages and Names.



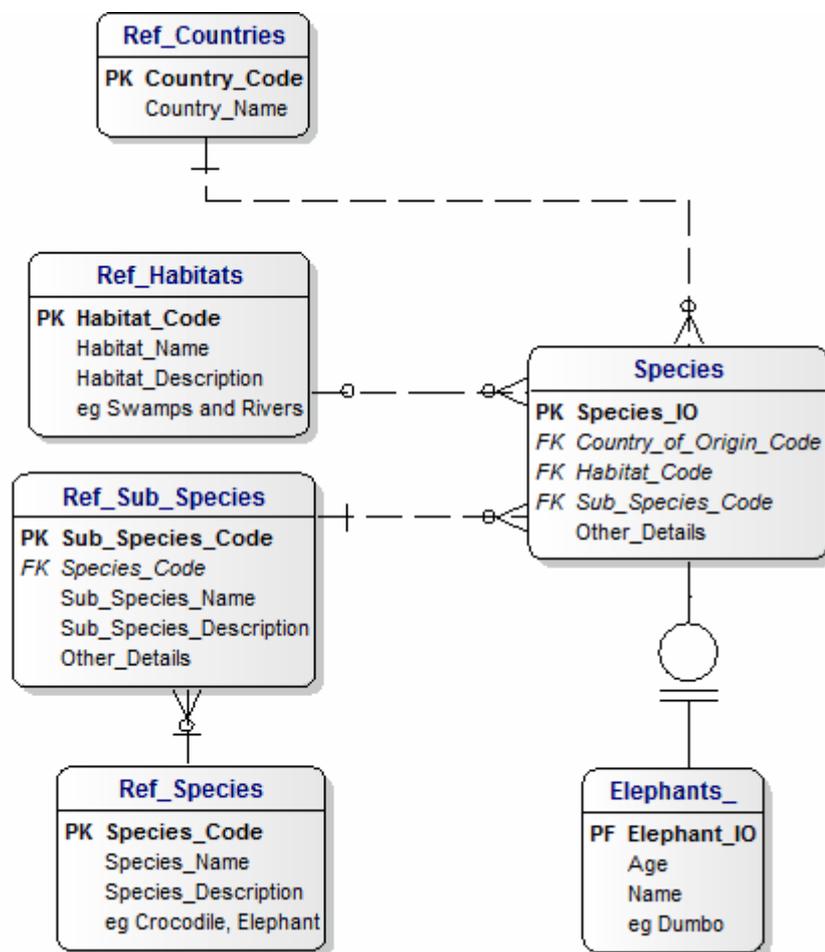
Step 3 : Crocodiles, Elephants and Species

Of course, we do not normally think of Crocodiles as Animals.

Therefore we have to find another word that can apply equally to both Crocodiles and Elephants.

Species is a word that comes to mind so we will use it, and we will use Sub-Species for specific types of Crocodiles and Elephants.

You can see that we have established that ages and names are usually available to Elephants but not for other species.



6.4.5.1 The CDM Design Pattern

This Section discusses how the Canonical Data Model applies to the Event of Visiting a Crocodile Farm.

We would expect this to be identical to a Visit to the Elephant Sanctuary.

But it is worth the effort of compiling the Mapping Analysis so that we can double-check the situation.

Sure enough, after we complete the Mapping, we can see that the logic is identical.

Therefore we do not need to change the Design Pattern or the Data Warehouse.

The Dimensional Model will simply have additional data for the Crocodile Farm.

6.4.5.2 Mapping to the CDM

This Section discusses how the Canonical Data Model applies to the Event of Visiting a Crocodile Farm.

We can see that the Data Model for Crocodile Farm is identical to that for the Elephant Sanctuary.

We can simply create a Data Model for Tourist Attraction and create Event Types of Visits to a Crocodile Farm and an Elephant Sanctuary.

Therefore, we handle Elephant Sanctuaries and Crocodile Farms as different sorts of Reference Data.

This Table shows how the Entities in our Crocodile Farm Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Visit to Crocodiles	COMMENTS
Customers	Tourists	
Documents	Tickets	
Events	Visit to Crocodile Farm	
Locations	Address of Crocodile Farm	
Organisations	Crocodile Farm Owners	
Products or Services	See the Crocodile Farm	
Third Parties	N/A	

6.4.5.3 Message Format

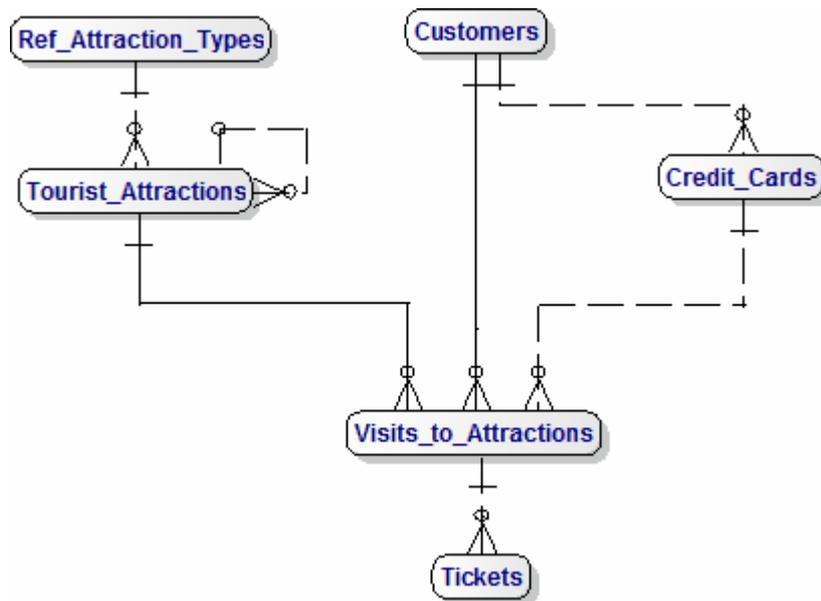
The Message Format will resemble the Tickets.

Attraction Name	Date & Time	Customer Details	Services	Unit Price	From Date	To Date	Total Price
eg Crocodile Farm	Visit Date & Time	N/A	Attraction	Entry Fee	N/A	N/A	As determined

6.4.5.4 Data Warehouse for the Elephant Sanctuary

The Data Model the Crocodile Farm is logically identical to the Elephant Sanctuary Model.

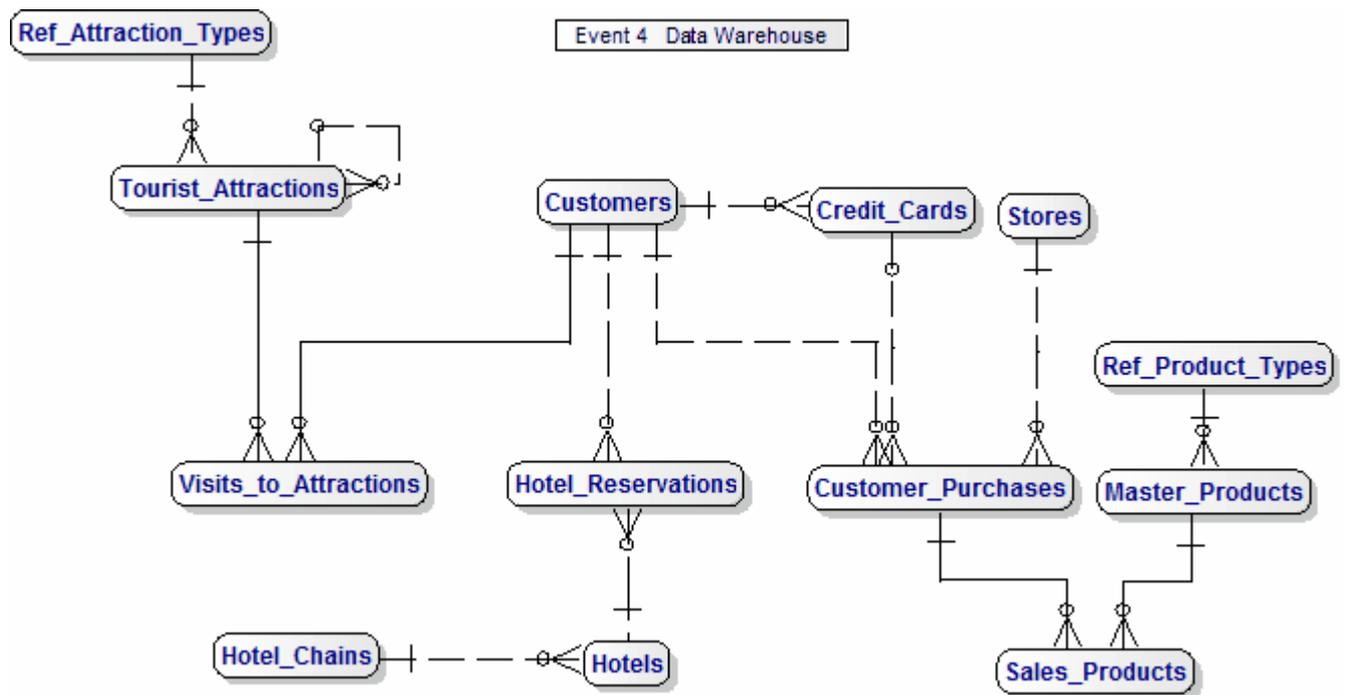
Therefore we can use the same Model for Tourist Attractions and simply add values to the Reference Data.



6.4.5.5 Consolidated Data Warehouse

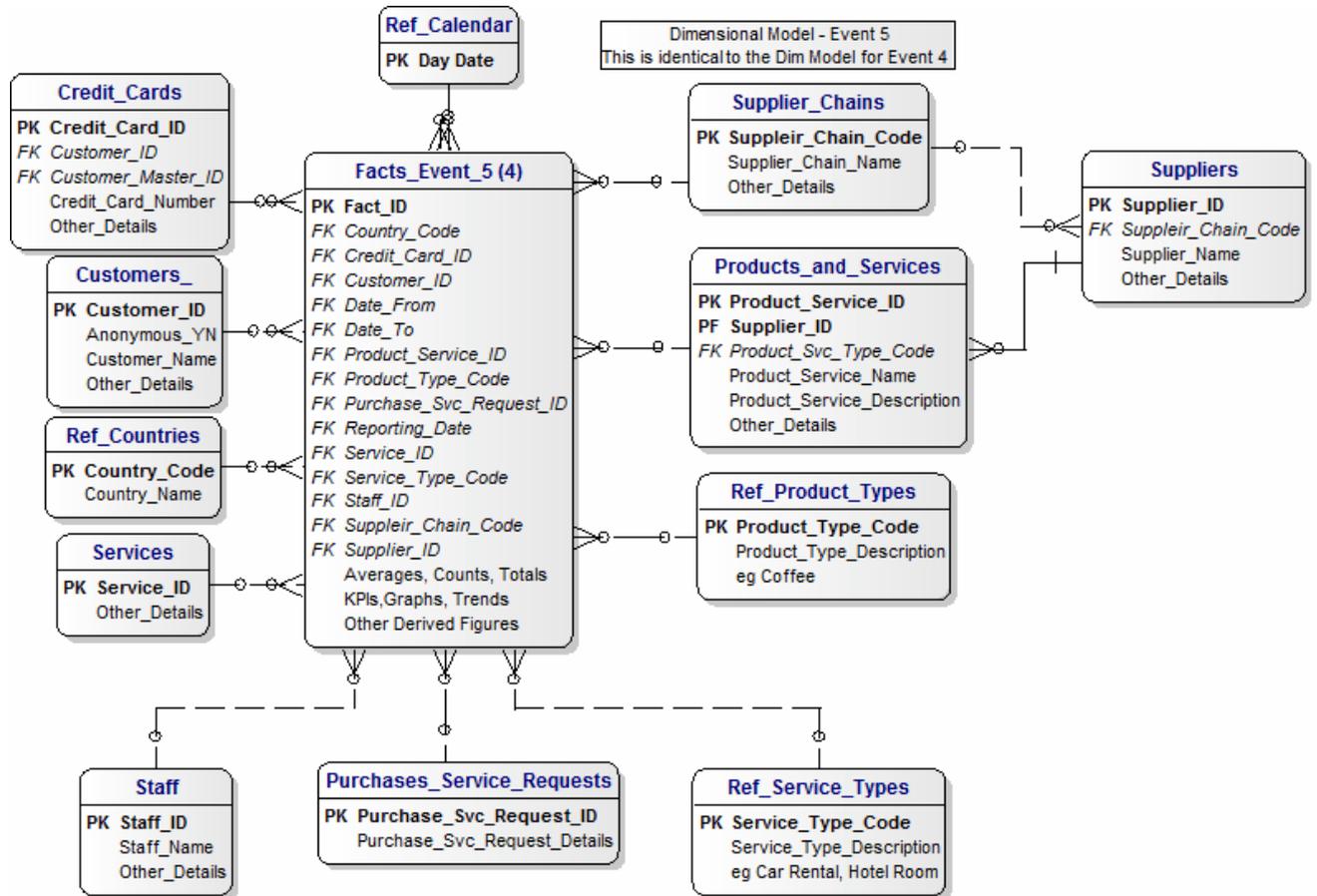
When we review this Data Model, we can see that the logic of the Elephant Sanctuary applies equally to the Crocodile Farm.

In other words, our Consolidated Data Warehouse is identical, and we simply add to the Reference Data as another kind of Tourist Attraction.



6.4.5.6 Dimensional Model

When we review the Dimensional Model, we can see that that we can accommodate Crocodile Farms by simply adding to the Reference Data.



6.4.6 Event 6 - Visit the Underwater World

The Underwater World was very difficult to find, but when we did find it, it was definitely worth it.

We paid a fee to enter but when we got inside we discovered that there many Attractions inside a very large building,

Entry to these Attractions, like the Aquarium, was free.

After a little thought, I decided to treat the Underwater World as a Tourist Attraction.

Then I decided to treat the Aquarium and Penguins as Tourist Attractions with a 'Parent' relationship to the Underwater World.

This is very easy to implement in the Model and makes it very easy to 'read' the Model.

This is important because every Data Model should tell a story and therefore a good test of a well-designed Model is that it is easy to read it like a story.

Therefore, the Aquarium and the Penguins both become Attractions with a zero entry fee and have a relationship to the Underwater World as a 'Parent'.

Here's a photo to give you a glimpse of the fascinating interior of Underwater World from this page on their Web Site :-

- <http://www.underwaterworldlangkawi.com.my/>



6.4.7 Event 7 - Visit the Aquarium

Visiting the Aquarium was great and allowed me to tick off another “once-in-a-lifetime” experience which was to walk inside an aquarium looking up at the fish swimming by over my head !!!

This photo shows the view I saw when I looked up and saw fish swimming past above my head.



6.4.7.1 Mapping to the CDM

From a Data Modelling point of view, visiting an Aquarium is identical to visiting a Crocodile Farm or Elephant Sanctuary.

Therefore we do not need a separate CDM, and we go through the process of mapping simply to confirm that it is identical.

This Table shows how the Entities in our Visit the Aquarium Event map on to our Design Pattern based on our Canonical Data Model (CDM).

CDM	EVENT : Visit to Aquarium	COMMENTS
Customers	Tourists	
Documents	Tickets	
Events	Visit to Aquarium	
Locations	Address of Aquarium	
Organisations	Aquarium Owners	
Services	See the Aquarium	
Third Parties	N/A	

Elephants, Olympic Judo and Data Warehouses

6.4.8 Event 8 - Visit the Penguins

The Penguins were also at the Underwater World and I was very lucky again because I saw live penguins which was another 'first' for me.

They were a wonderful sight.

It was very hot outside (over 90 degrees in the shade) outside in tropical Malaysia.

But inside we saw Penguins in a very comfortable (for them) sub-zero temperature.



6.4.8.0 Validation of the Species Data Model

Validation proves that our existing Species Model is correct.

Fish do not have names and their characteristics are the same as Crocodiles.

We can say the same for Penguins.

But Fish and Penguins both have Sub-Species !!!

Therefore we have validated our Species Model and it is good as it stands for Fish and Penguins, as well as Crocodiles and Elephants.

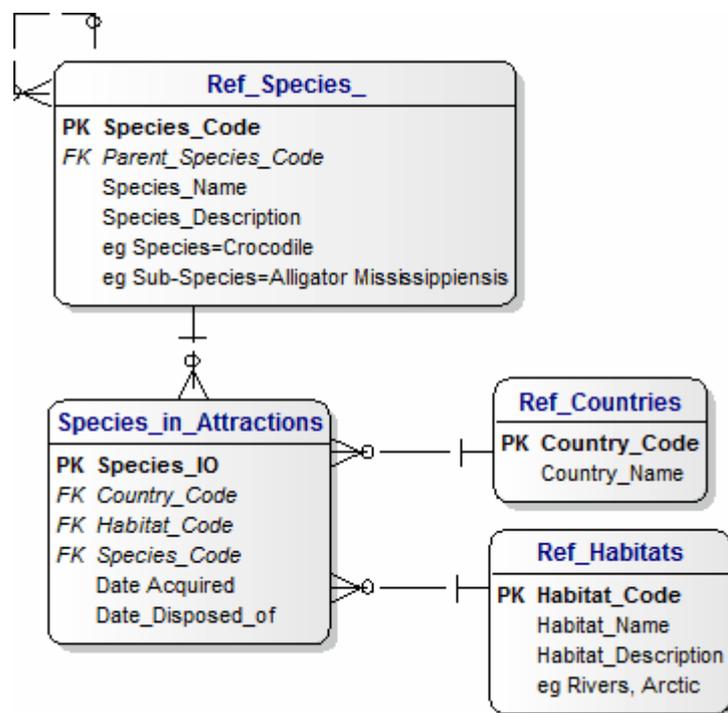
To make it clear that we are talking about the examples of Species that we see in the Underwater World we call the Entity 'Species in Attraction'.

It would be sensible for us to plan for future Species which might not be a Sub-Species.

The simplest way to do this is to add a self-referencing Relationship to the Species Entity.

This means that the Species can point to itself so that Species, Sub-Species are stored in the same table in a Database.

In HR terms, this is like having an Employee table and saying 'Each Employee reports to a Superior'.



6.4.8.1 CDM Design Pattern

This Section shows how the Design Pattern looks for Penguins.

6.4.8.2 Mapping to the CDM

This Section discusses how the CDM applies to the Penguin Area.

When we check our CDM we can see that it applies in an identical way that it does to Elephants and Crocodiles.

That is we pay for a Service and receive a Document, in the form of a ticket, that allows us to enter the Attraction.

Therefore we do not need a separate CDM, and we go through the process of mapping simply to confirm that it is identical.

This Table shows how the Entities in our Tourist Attraction Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

CDM	EVENT : Visit to Penguins	COMMENTS
Customers	Tourists	
Documents	Tickets	
Events	Visit to Penguins	
Locations	Address of Penguins Area	
Organisations	Penguin Area Owners	
Services	See the Penguins	
Third Parties	N/A	

6.4.8.3 Message Format

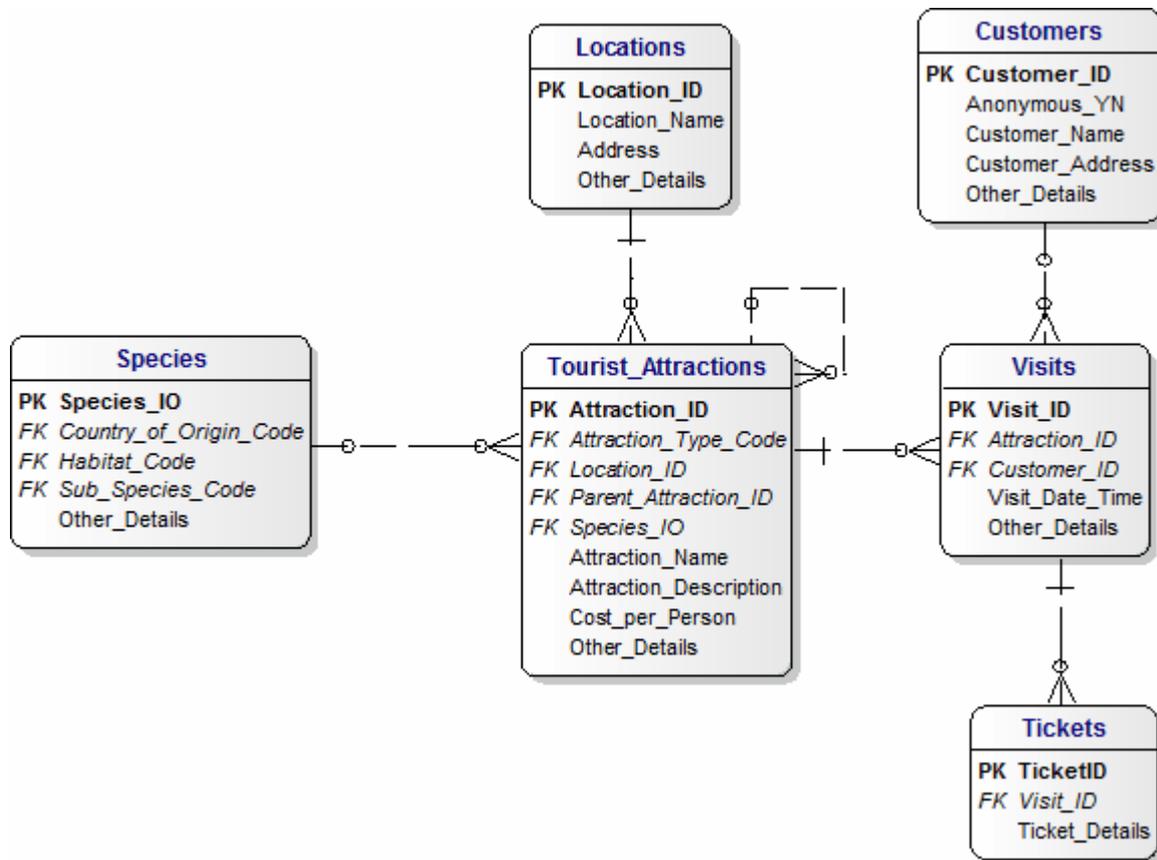
Entry to the Penguin Area is free and not monitored after you buy your ticket for the Underwater World and enter.

The Message Format will be simply the one for the Ticket to the Underwater World because there is no separate ticket for the Penguin Area.

Attraction Name	Date & Time	Customer Details	Services	Unit Price	From Date	To Date	Total Price
Underwater World	Visit Date & Time	N/A	Penguin Area	Free	N/A	N/A	Free

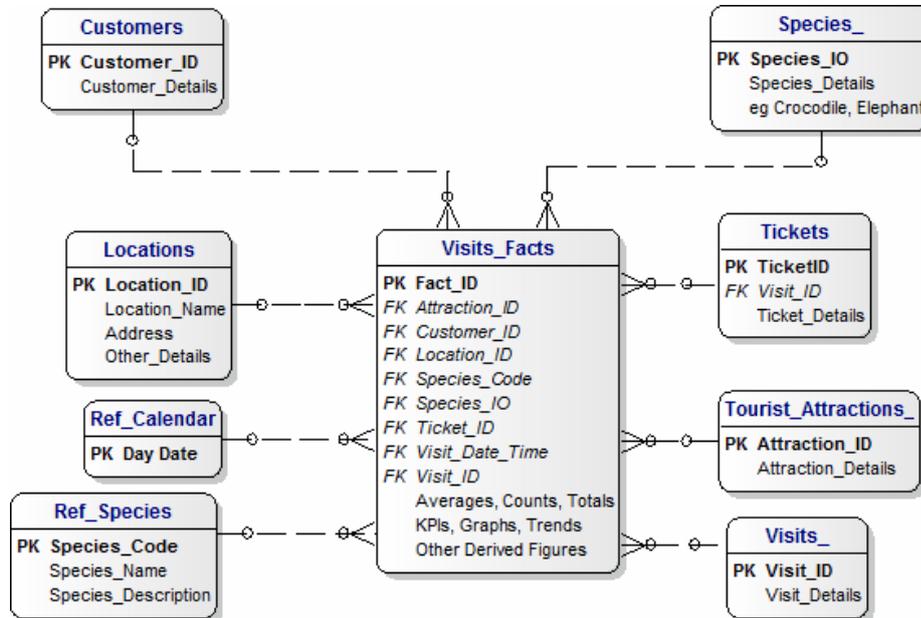
6.4.8.4 Data Warehouse for Tourist Attractions

This diagram shows the Design Pattern of the Canonical Data Model adapted for Tourist Attractions.



6.7.8.5 Dimensional Data Model

This diagram shows the Design Pattern of the Canonical Data Model adapted for Tourist Attractions.



6.4.9 Event 9 - Check-out from a Hotel

The last Hotel we stayed in was the Shangri-la in Penang.

This was the beautiful view from our room that we were very sad to see for the last time before we checked out:-

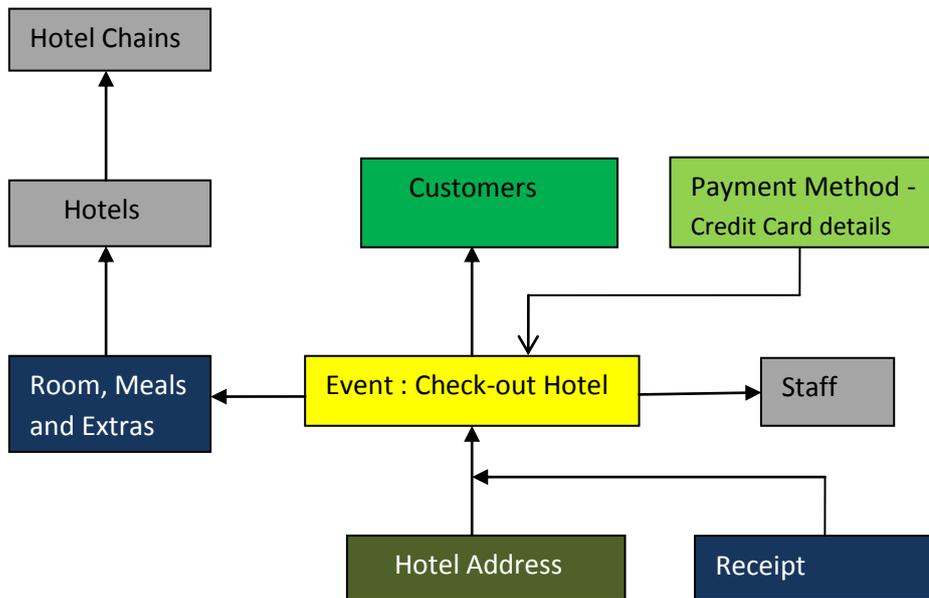


6.4.9.1 The CDM Design Pattern

The Credit Card that I use was, of course, associated with me, but it was also associated with payment of the Hotel Bill so there is a relationship between the Credit Card and the Hotel Guest (ie Customer) and between the Credit Card and the Total Hotel Bill.

This Section discusses how the Canonical Data Model (CDM), shown in Section 2.1, applies to the Event of Checking-in to a Hotel. The CDM provides a Design Pattern for the Event-oriented Data Models that we need.

The Design Pattern based on the Hotel Check-in Event looks like this :-



6.4.9.2 Mapping to the CDM

This Table shows how the Entities in our Hotel Check-out Data Model map on to our Design Pattern based on our Canonical Data Model (CDM).

We are very happy to see that it does because it helps to validate the CDM.

CDM	EVENT : Hotel Check-out	COMMENTS
Customers	Guest	
Documents	Receipts	
Events	Rent a Room, Use Restaurants, Laundry, and so on.	
Locations	Hotel Address	
Organisations	Staff in the Hotel	Staff are always involved in the course of a stay in a Hotel
Products or Services	Room, Meals and Services.	
Third Parties	N/A	

6.4.9.3 Message Format

This shows how the Generic Message Template applies to the Hotel Check-Out Event.

It defines the Source Data for this Event.

Generic	Supplier	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
Specific	Shangri-la Hotel	Check-in Date & Time	Barry's Name and Credit Card Details	Room, Meal, Services	Price per Night	From Date	To Date	Derived

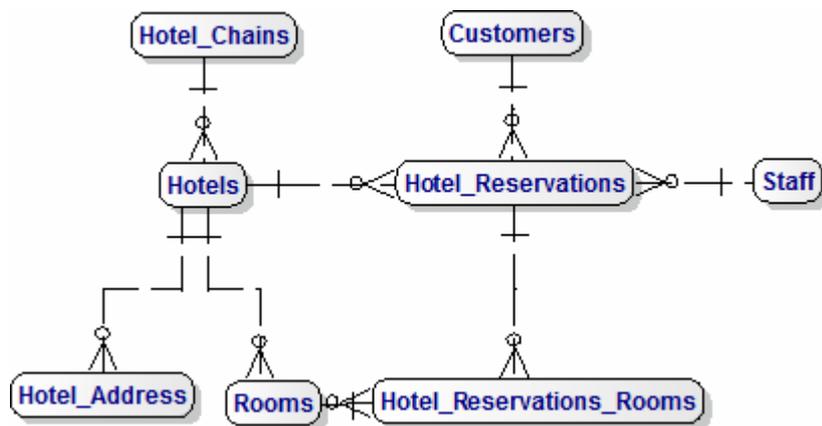
6.4.9.4 Data Warehouse for Hotel Check-Out

At this point, we add specific Check-out data to the Data Warehouse.

This diagram shows the Entities in the Data Model.

The attributes are shown in the Dimensional Model in the next Section.

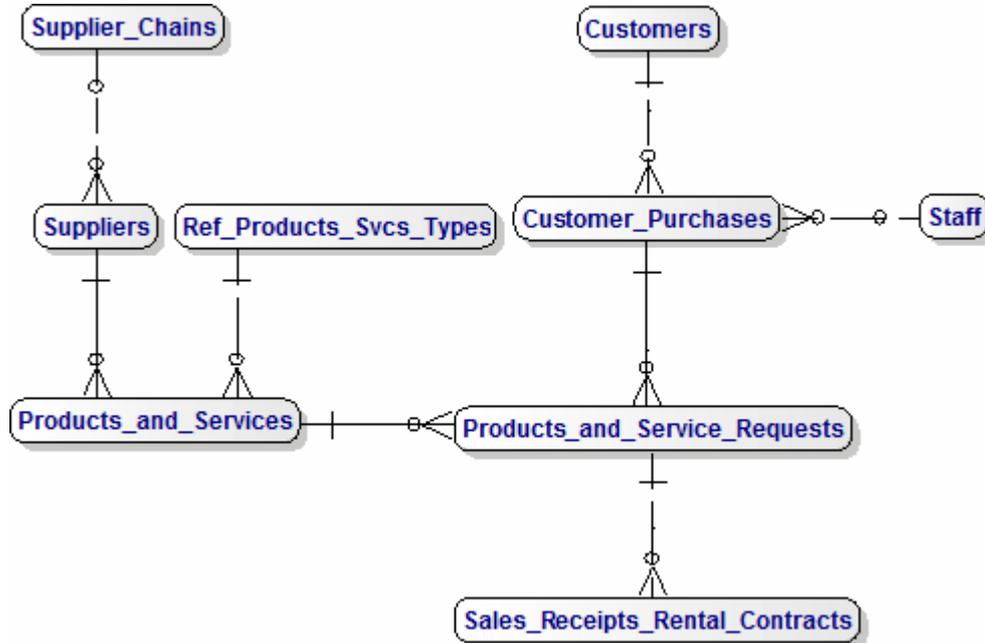
We have also shown the Hotel_Address entity with its correct relationship to the Hotel entity, whereas in the CDM, it is shown, for the sake of convenience, in a Many-to-Many Relationship with the Events entity.



Elephants, Olympic Judo and Data Warehouses

6.4.9.5 Consolidated Data Warehouse

At this point, we add the Shopping data for Event 3 to the data for the Car Hire and Hotel Check-in Events which is already in the Data Warehouse, so our design looks like this :-



Elephants, Olympic Judo and Data Warehouses

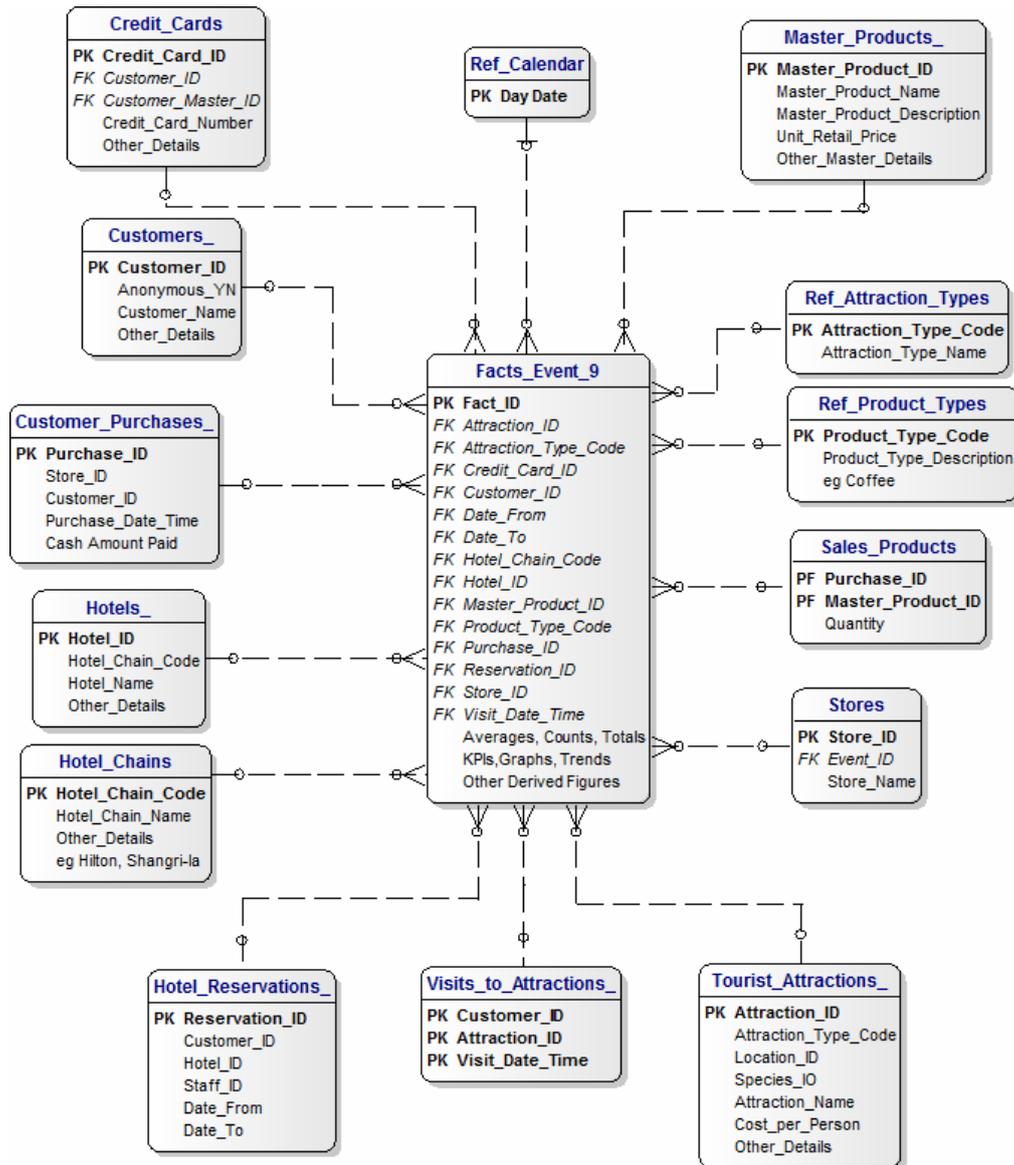
6.4.9.6 Dimensional Model

At this point, the Dimensional Model will have the complete set of Dimensions and Facts.

This Data Model shows them all consolidated into a single Dimensional Model.

An alternative design with more than one Fact Table is shown in the BI discussion.

This is normally called multiple Data Marts which require Conformed Dimensions.



6.4.10 Event 10 - Shipping a Car from the UK to Malaysia

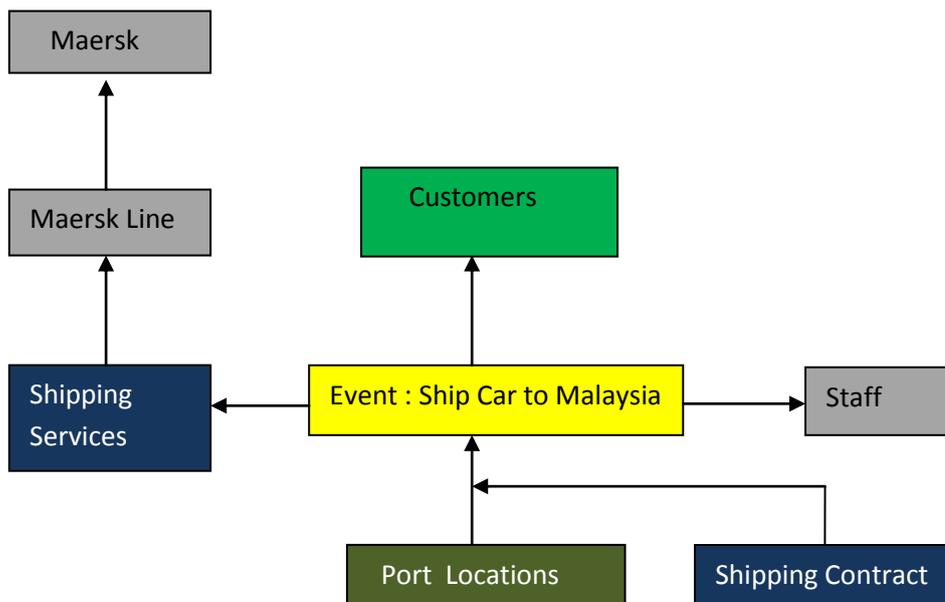
We had such a great time in Malaysia, we decided to move out there for a while. One of the things we had to arrange was to ship our car out. We decided to use Maersk to help us.

Here is the kind of Container Ship we had in mind :-



6.4.10.1 The CDM Design Pattern

This Section shows how the CDM Design Pattern applies to the Event of “Ship a Car to Malaysia”.



6.4.10.2 Mapping to the CDM

This Table is your starting-point for defining how the Entities correspond to the Entities in your 'Stopping for a Coffee' Event.

Replace the question marks by your answers.

CDM Generic Entities	EVENT : Ship a Car by Sea	COMMENTS
Customers	Customer	
Documents	Contracts	
Events	Ship a Car by Sea	
Locations	Booking Office	
Organisations	Maersk	
Organisations	Staff in the Office	Staff can be involved and represent the Organisation
Products or Services	Shipping Service	
Third Parties	N/A	

6.4.10.3 Message Format

This shows how the Generic Message Template applies to the Specific Hotel Check-In Event.

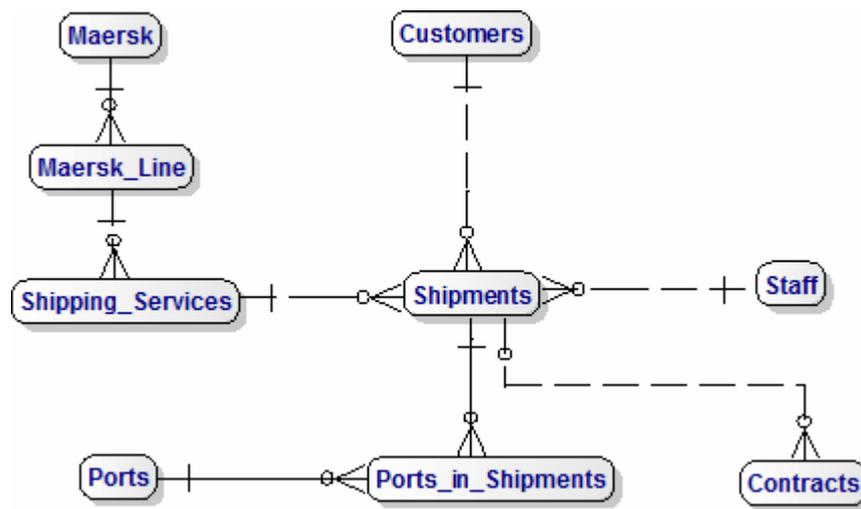
Generic	Supplier	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
Specific	Maersk Line	Check-in Date & Time	Barry's Name and Credit Card Details	Shipping Service	Quoted price to ship Cargo	From Date	To Date	---

6.4.10.4 Data Warehouse for Shipping a Car

At first, this Event seems quite different from the others, but after a little thought, we realise that it has a Start and End Date.

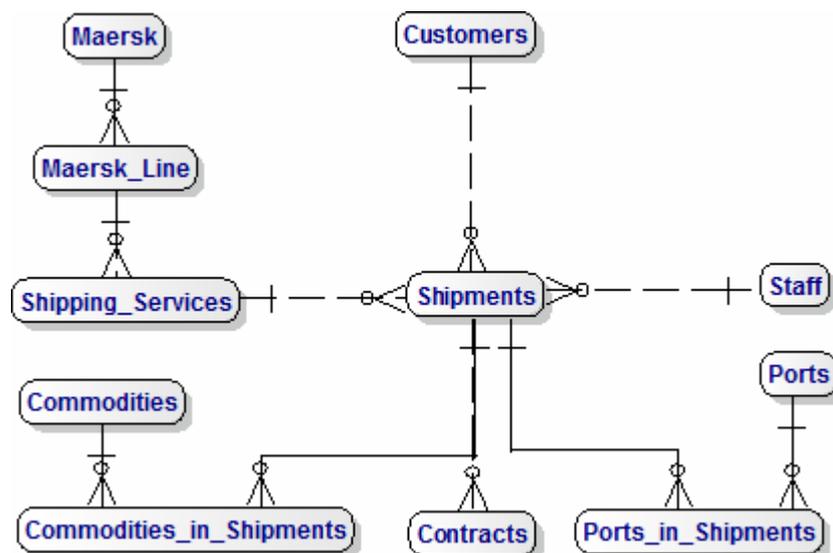
In addition, there is a Customer and we are using a Service provided by a Supplier.

Therefore Shipping a Car is therefore logically similar to a stay in a hotel and our Data Warehouse looks like this :-



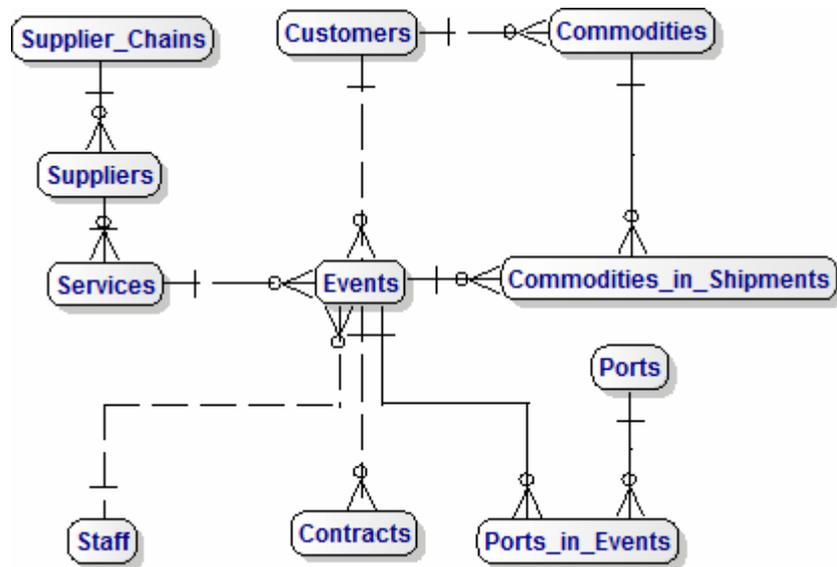
In general, we might be shipping many things and not just one thing, so our Data Model needs to provide for this. We will call these 'things' Commodities

We can do this by simply start with a many-to-many relationship between Commodities entity and the Shipments entity and resolve it to two one-many relationships, so that our Model looks like this :-



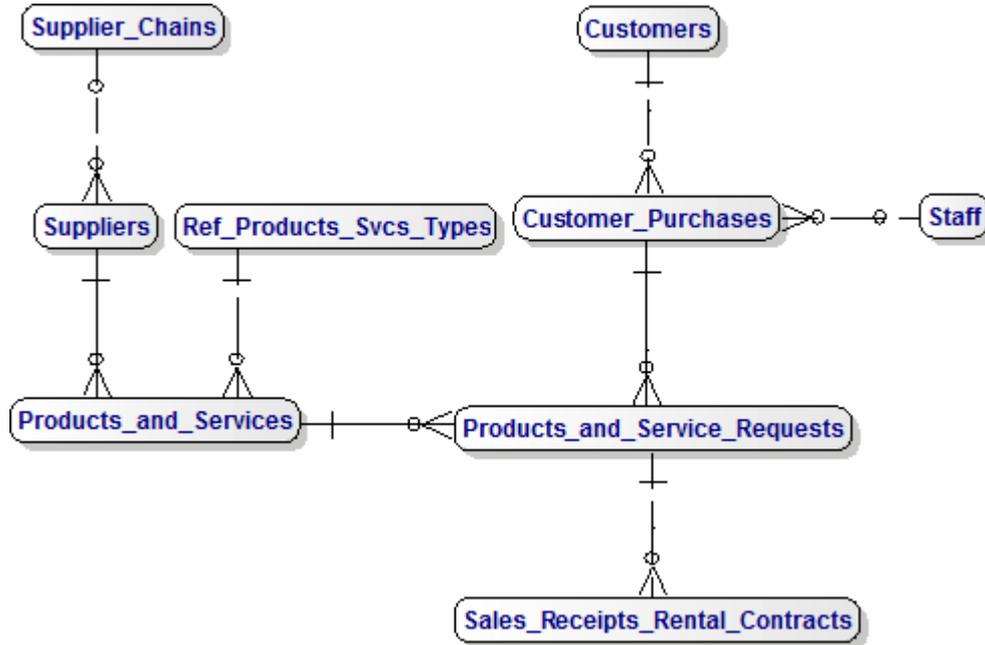
Elephants, Olympic Judo and Data Warehouses

When we think about how we would show this Data Model in a generic format, this is the result :-



6.4.10.5 Consolidated Data Warehouse

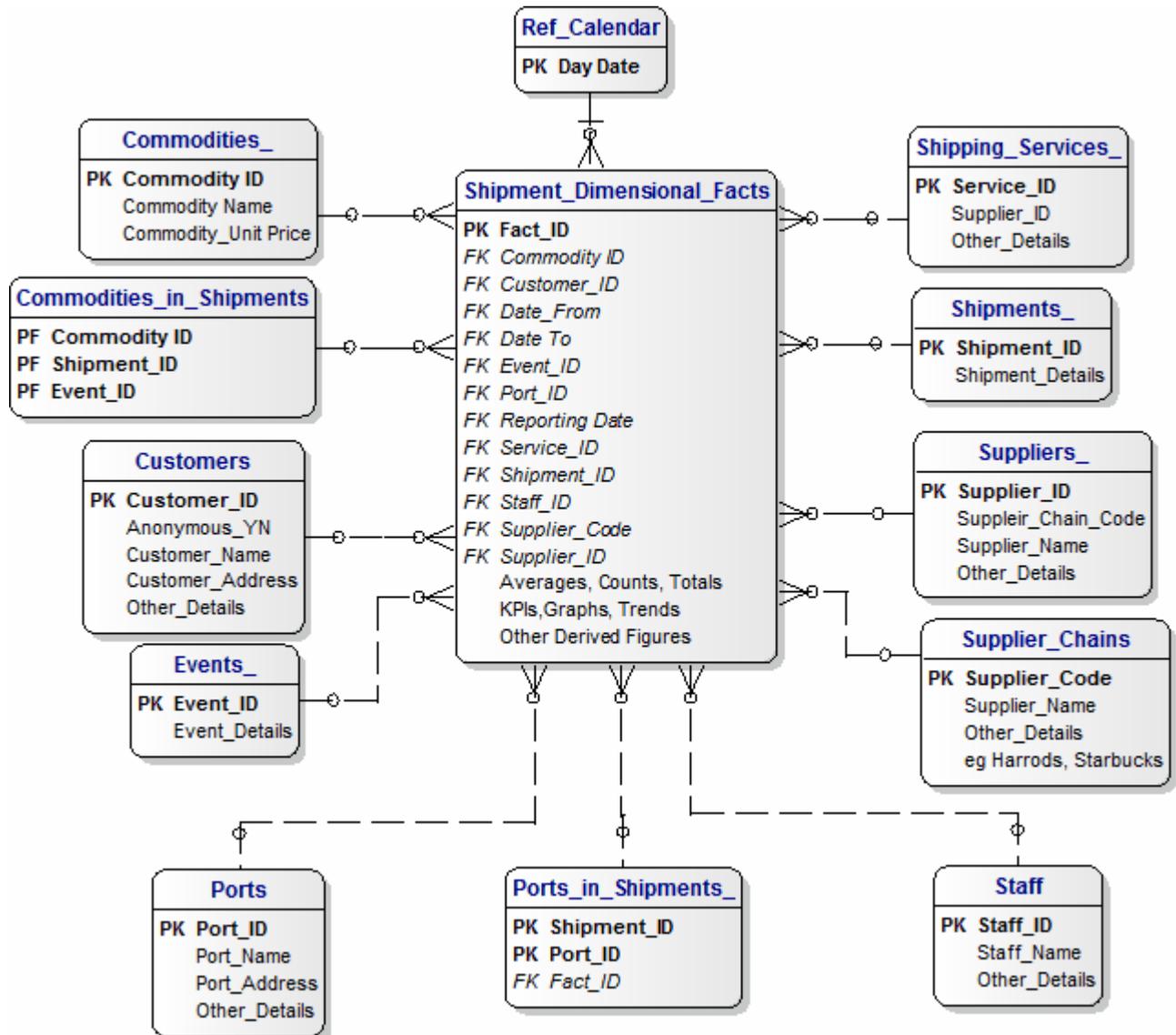
At this point, we add the Car Shipping data to the data for all the other Events which is already in the Data Warehouse, so our design looks like this :-



Elephants, Olympic Judo and Data Warehouses

6.4.10.6 Dimensional Model

This is the Dimensional Model for the Car Shipping Event.



6.5 The BI Layer

6.5.1 The Approach

Now we come to the Business Intelligence Layer – which is, of course, the reason for everything that has gone before.

Key Performance Indicators ('KPIs') provide an excellent starting-point for Business Intelligence.

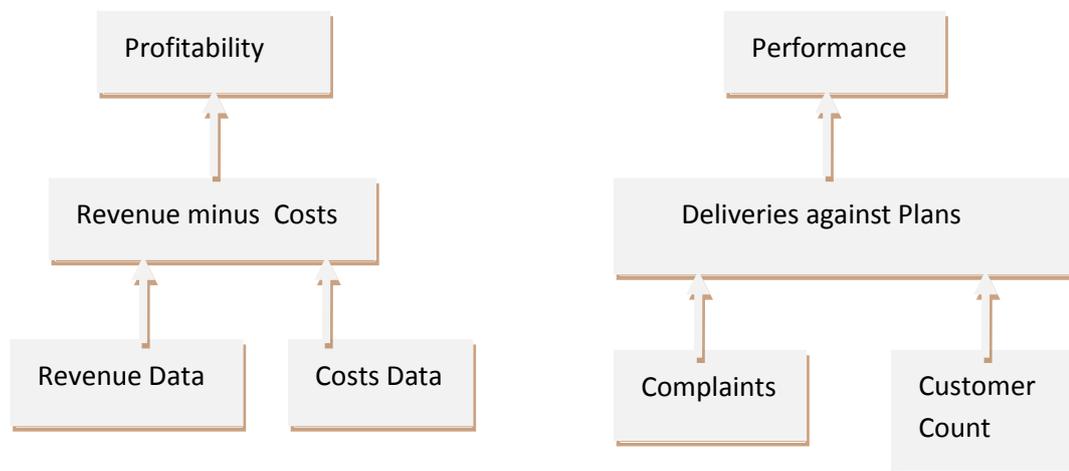
We will use some basic KPIs defined for Profitability and Operational Performance.

These can be enhanced in collaboration with the business users to establish commitment.

Profitability can initially be defined as a basic difference between Profits and Revenue.

One example of Performance KPIs is measurement of Deliveries being 'On-Time' against plans agreed in Service Level Agreements (SLAs).

In our work we do not have that kind of information available, so we will use a simple measure of the number of Customer Complaints.



The initial Data Architecture will include this basic structure of KPI and Source Data.

This will provide a simple but flexible framework which can be enhanced in a controlled manner.

We will use this as a starting-point for discussions to establish the requirements with the business users.

Elephants, Olympic Judo and Data Warehouses

6.5.2 BI Discussion

We have adopted a general approach, where we classify Car Hire companies, Hotels, Shops and Tourist Attractions as Brand Chains and specific outlets as Branches.

Therefore a particular Hotel is a Branch and the Branch Chain would be Hilton or Shangri-la.

Here are a few simple examples of the kind of questions that we can answer with this approach.

For the Finance Director of a Hotel chain :-

- Which Hilton Hotel was the most profitable last month ?
- Which Hilton had the most complaints last week as a percentage of total Customers ?

For the Marketing Director of a Car Hire Company :-

- What was the most popular type of car last week ?
- What was the most profitable type of car last week ?

For the Head of Tourism :-

- What was the most popular Tourist Attraction last month ?
- What was the most popular Tourist Attraction over the past six months ?

Elephants, Olympic Judo and Data Warehouses

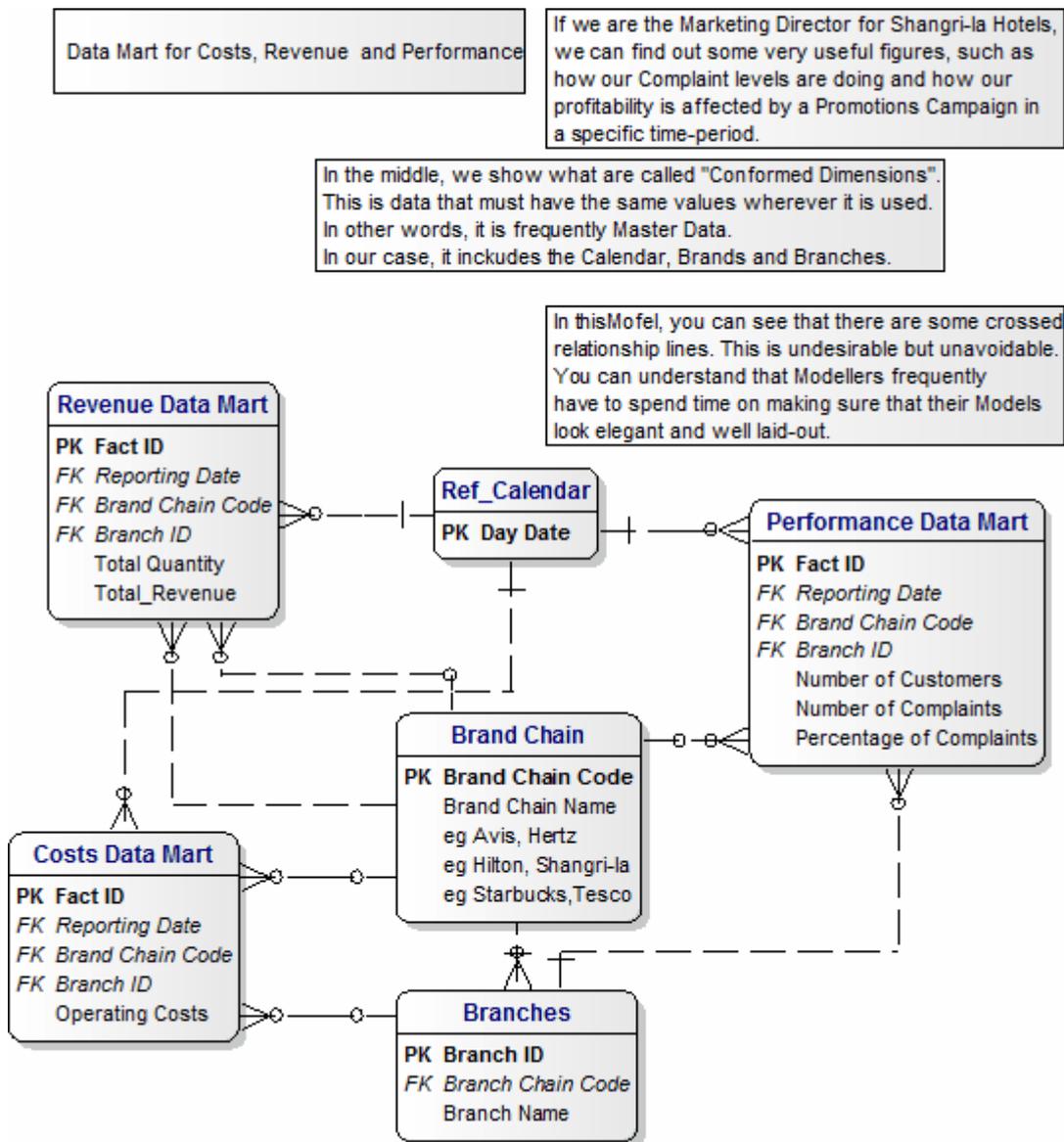
6.5.3 Data Marts

We define separate Data Marts for Revenue, Costs and Complaints.

When we consider the implementation options for the BI layer, we have quite a number of options.

These include the Microstrategy Personal Cloud, as well as some KPI-oriented Cloud-based offerings

Firstly, we design our Data Marts and then we have the basis for a selection criteria.



Data Mart for Costs, Revenue and Performance

If we are the Marketing Director for Shangri-la Hotels, we can find out some very useful figures, such as how our Complaint levels are doing and how our profitability is affected by a Promotions Campaign in a specific time-period.

In the middle, we show what are called "Conformed Dimensions". This is data that must have the same values wherever it is used. In other words, it is frequently Master Data. In our case, it includes the Calendar, Brands and Branches.

In this Model, you can see that there are some crossed relationship lines. This is undesirable but unavoidable. You can understand that Modellers frequently have to spend time on making sure that their Models look elegant and well laid-out.

6.5.4 BI Screenshots

Here we show some examples of Screenshots for currently available Cloud-based BI Products.

1) KPI Library

- <http://benchmark.kpilibrary.com/>

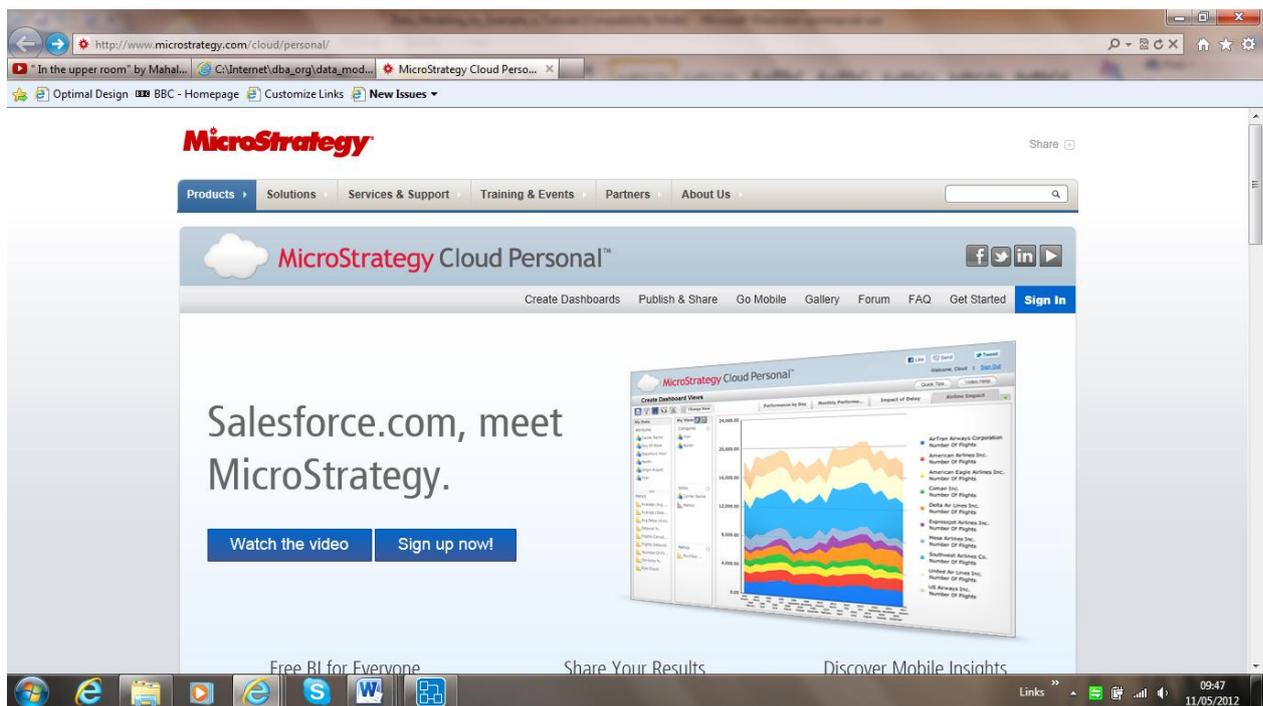
2) Smart KPIs - recommended

- http://www.smartkpis.com/i_kpi/industries/professional-services/

3) Microstrategy Cloud Personal

Here is the Web Site Link (Sign in as barryw@dba.org, with a password of m123 :-

- <http://www.microstrategy.com/cloud/personal/>



4) Stacey Barr

Stacey is a prolific writer on KPIs and Performance Measurement :-

- <http://www.staceybarr.com/products/performanceblueprint.html>

6.6 What have we Learned ?

We have learned some very valuable lessons, including :-

- The importance of a Canonical Data Model
- How to combine Entities where inheritance is involved
- How to load data into a Data Warehouse
- The Steps in loading data from Source Data into a BI Layer

When we should consider using multiple Data Marts with Conformed Dimensions to meet our BI requirements.

6.7 Here's your Chance

This Chapter offers you the chance to plan loading new data into our Data Warehouse.

You will do it by completing the model answers, which are shown below as Template documents.

If you feel like sending us your answers we will be happy to give you our comments.

You can send them to barryw@databaseanswers.org.

6.7.1 Event : Coffee in Penang Airport

After we checked out of the Shangri-la, we took a limo to Penang Airport for the first leg of our trip back home to London.

With plenty of time we got a coffee at the Coffee Bean , which was started by Herbert Hyman in California in 1963. It is now an international operation, and very popular in Malaysia.

Looking for a suitable photo, I came across an excellent one on this page :-

- http://www.airliners.net/aviation-forums/trip_reports/reae.main/126525/



Elephants, Olympic Judo and Data Warehouses

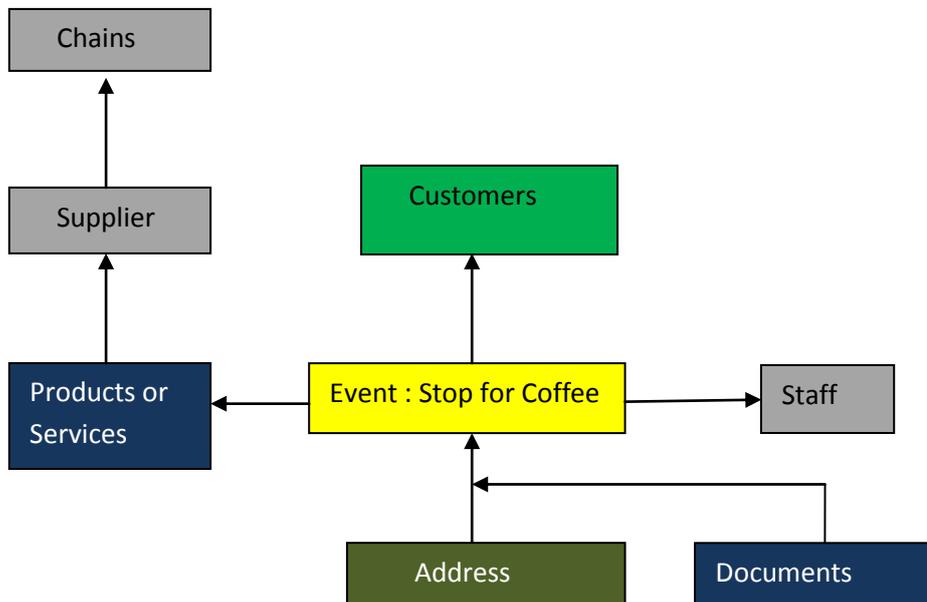
Here is our Sales Receipt from the Coffee Bean, which we were glad to see validated our CDM :-



6.7.1.1 The CDM Design Pattern

This Section shows the CDM Design Pattern.

Please think about how it applies to the Event of “Stopping for Coffee” and change it accordingly.



6.7.1.2 Mapping to the CDM

This Table is your starting-point for defining how the Entities correspond to the Entities in your ‘Stopping for a Coffee’ Event.

Replace the question marks by your answers.

CDM Generic Entities	EVENT : Stop for Coffee Entities	COMMENTS
Customers	???	
Documents	???	
Events	Stop for Coffee	
Locations	???	
Organisations	???	
Organisations	???	
Organisations	Staff in the Hotel	Staff can be involved and represent the Organisation
Products or Services	Room Reservation	
Third Parties	N/A	

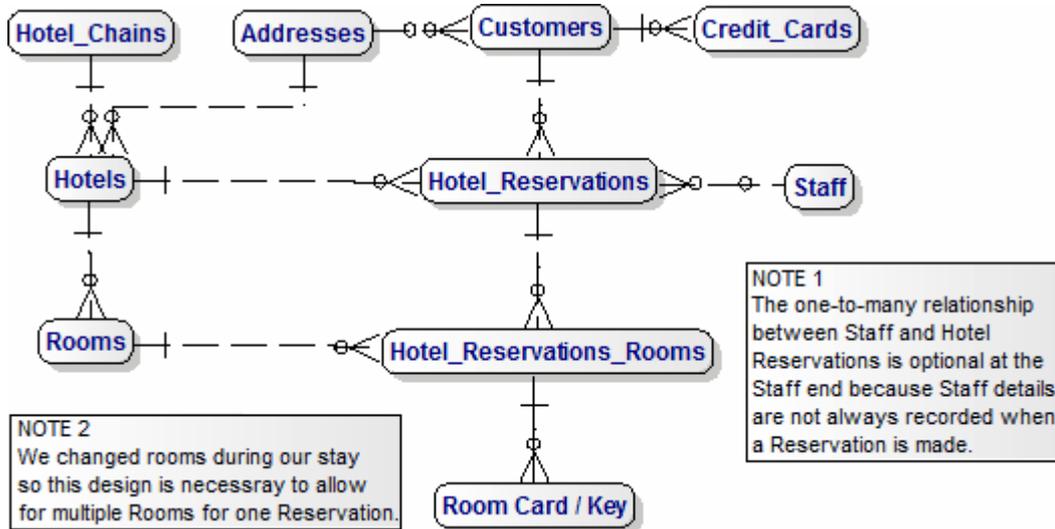
6.7.1.3 Message Format

This shows how the Generic Message Template applies to the Specific Hotel Check-In Event.

Generic	Supplier	Date & Time	Customer Details	Products or Services	Unit Price	From Date	To Date	Total Price
Specific	Shangri-la Hotel	Check-in Date & Time	Barry’s Name and Credit Card Details	Room number	Price per Night	From Date	To Date	---

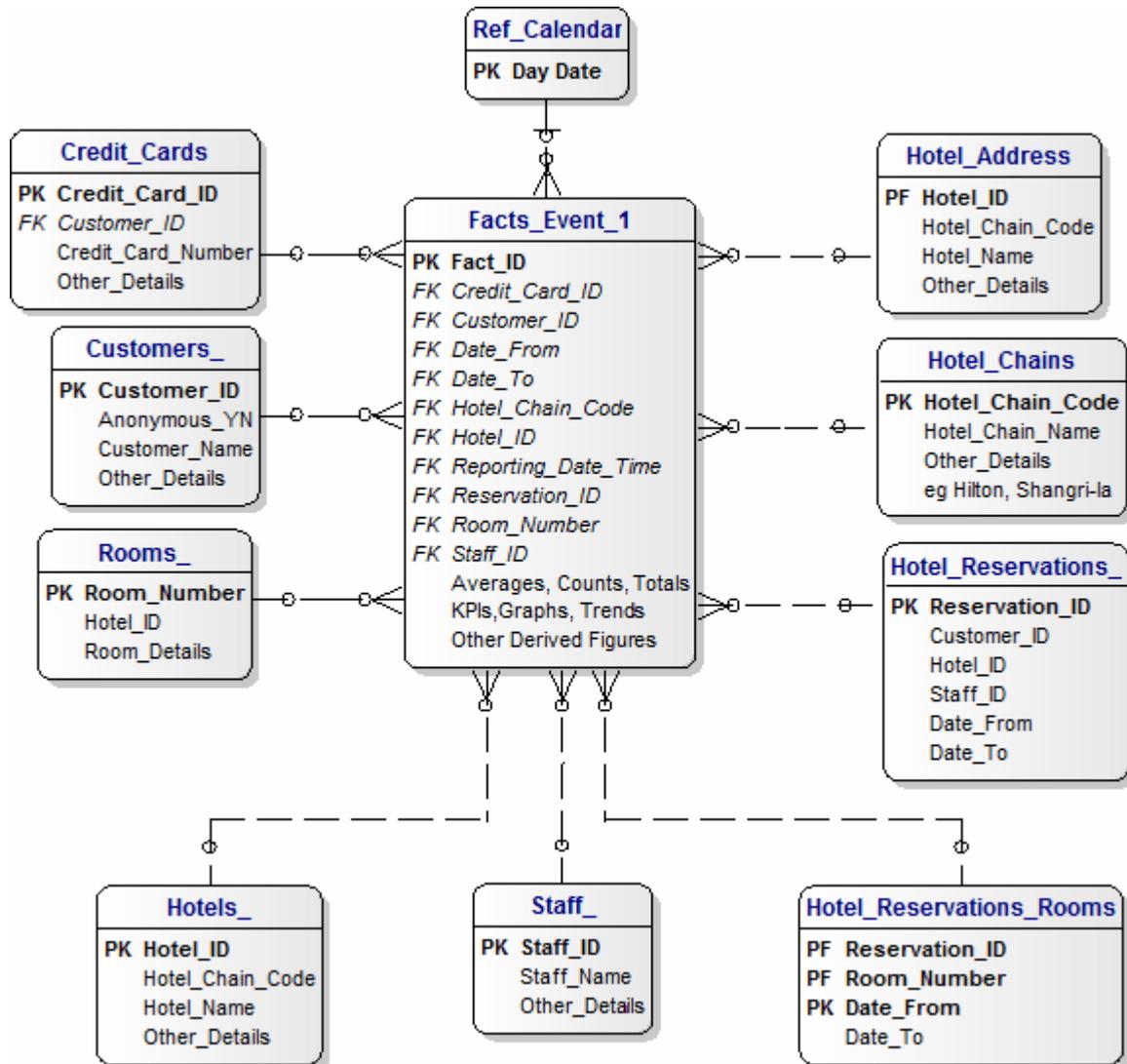
6.7.1.4 Data Warehouse for Hotel Check-In

This is the same as Third-Normal Form Data Warehouse as the one shown earlier in this Section.



6.7.1.5 Dimensional Model

This is the Dimensional Model for the Hotel Check-in Event, which you can use as your starting-point.



6.8 Data Model Review

This Section tries to describe the analysis and thought processes of a Data Modeller as he or she goes about his work.

Of course, it is subjective and these are my thoughts on the topic.

However, I have been doing this for over 15 years and I find it constantly fascinating.

The reasons I find Data Modelling so fascinating are that it is visual, creative and can be a great help when used as a source of Data Models as a vehicle for communication between business users, management and data management professionals, analysts and developers.

So here we go ;-0)

6.8.1 What are we going to do ?

We are going to step through the process of designing Data Models and then reviewing the results to see whether we can simplify the overall structure by generalising the Entities that we have produced.

It is one of the situations where it takes longer to describe it (which is boring) than doing it (which is interesting).

So (once again !!!) here we go ...

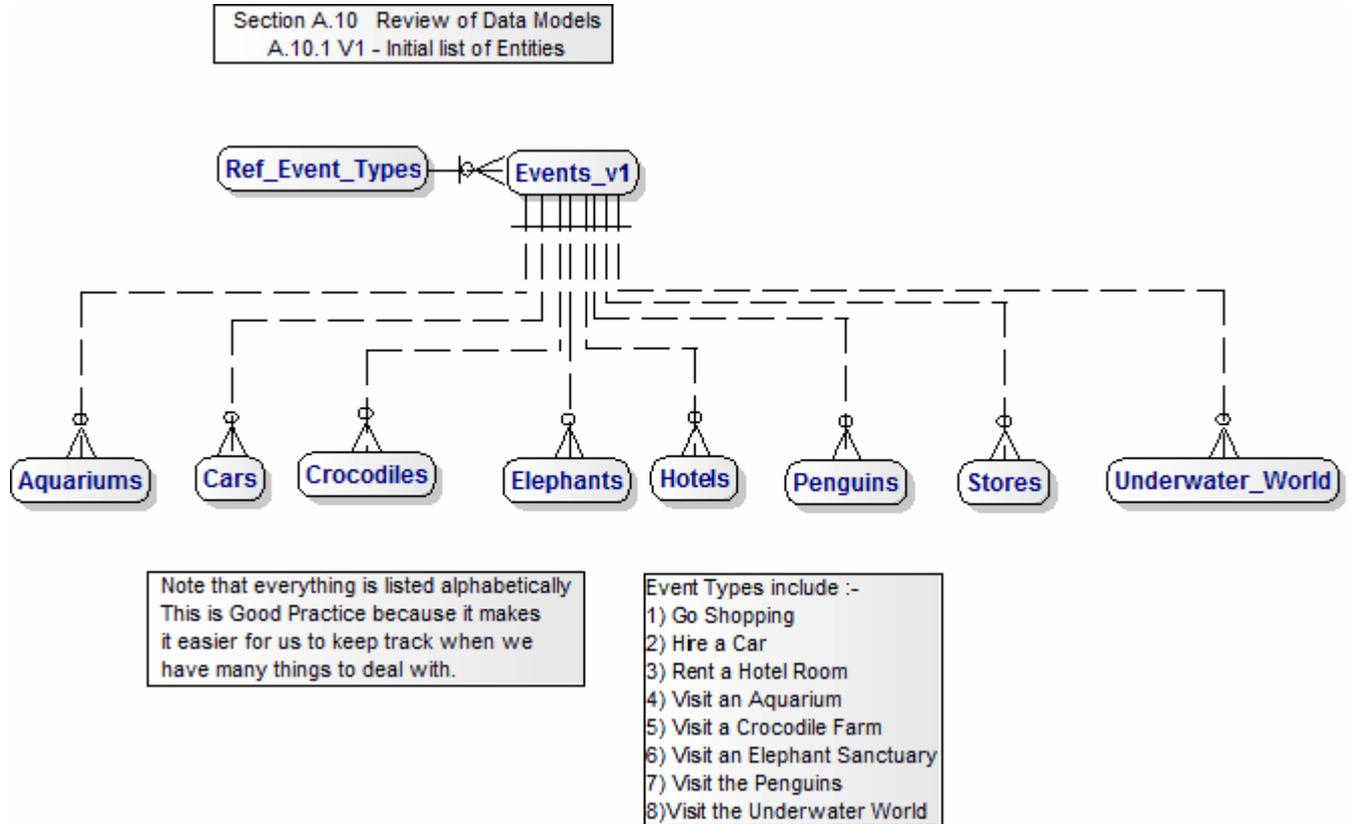
6.8.2 Where did we start ?

Our Canonical Data Model is Event-oriented and we started with nine Events :-

1. Check-in a Hotel
2. Hire a Car
3. Go Shopping
4. Visit an Elephant Sanctuary
5. Visit a Crocodile Farms
6. Visit the Underwater World
7. Visit an overhead Aquarium
8. Visit a Penguin Area
9. Check-out of a Hotel

6.8.3 First Data Model

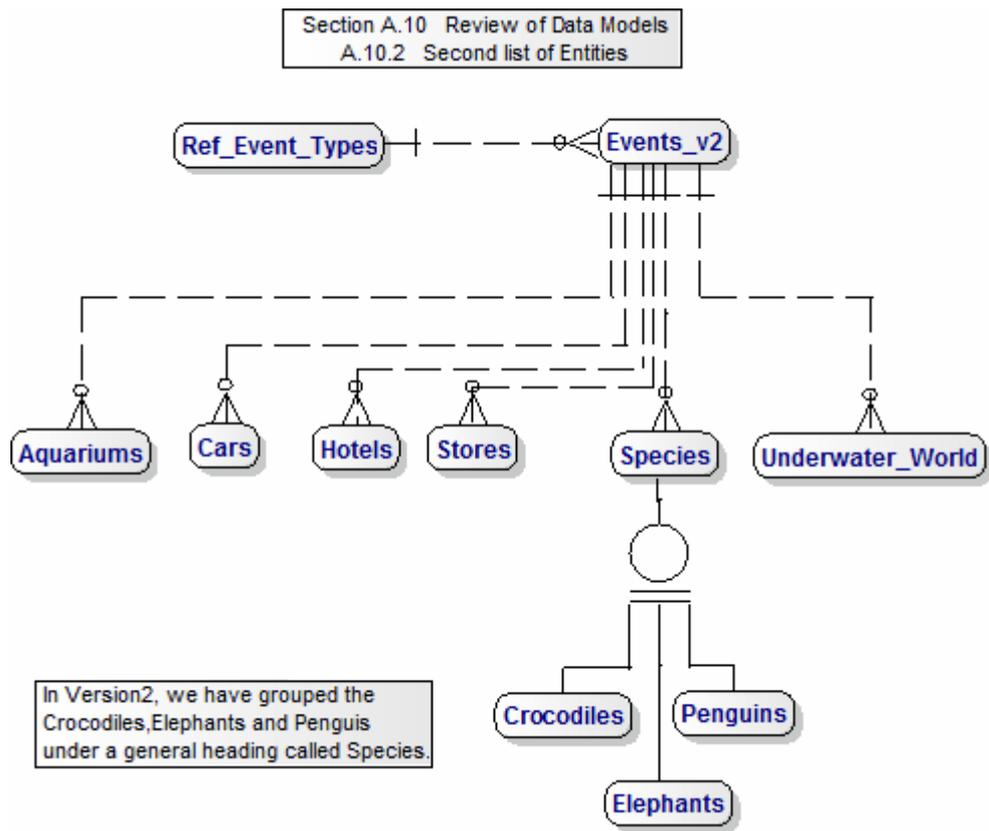
We have 9 different types of Events and our first Data Model looks like this :-



6.8.4 Second Data Model

When we think about the Entities, it seems intuitive that visiting a Crocodile Farm must be similar to visiting an Elephant Sanctuary.

Then we can review the discussion in the Section above to convince ourselves to treat them the same way.



6.8.5 Third Data Model

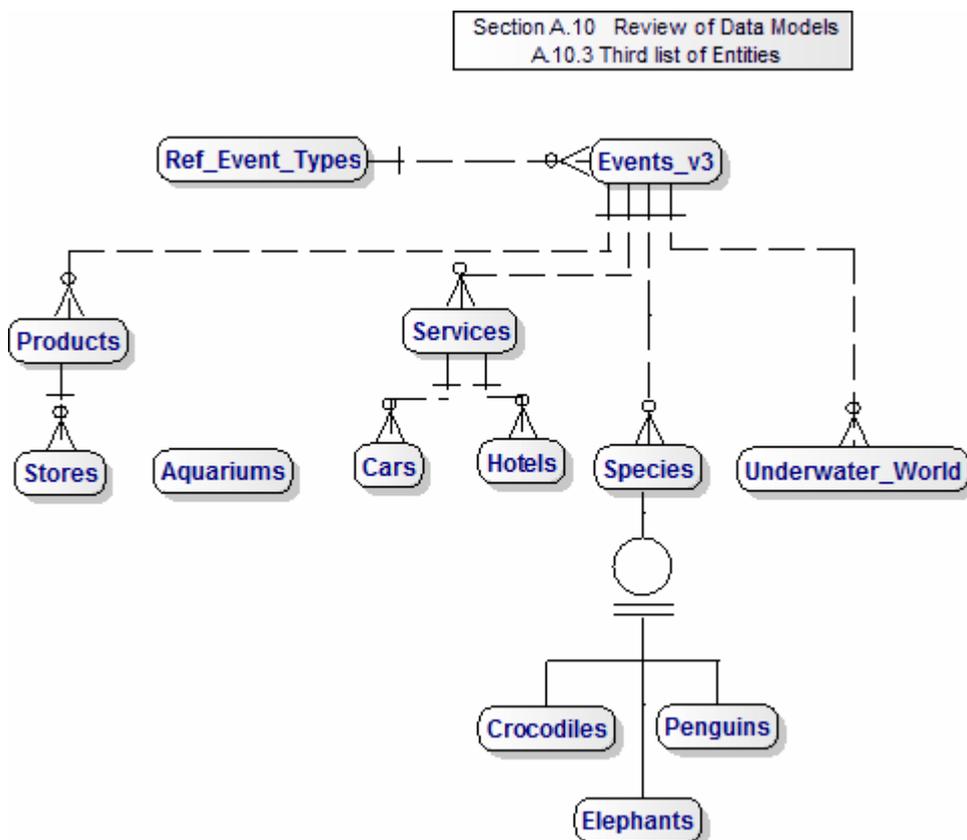
At this stage, we are thinking about the underlying relationships between the entities that we have determined are in the scope of our study.

We have started to group entities together in a way that helps us to reduce the overall size of the Data Model by identifying generic characteristics in the specific entities.

However, every Data Model should tell a story and now we have to consider how our story can be told.

For example, we would say 'Products are found in Stores' and therefore we should show a one-to-many relationship between Stores and Products.

We can see that it is the other way round in Version 3 because of the way it had evolved so now we have to change it to make it correct



6.8.6 Fourth Data Model

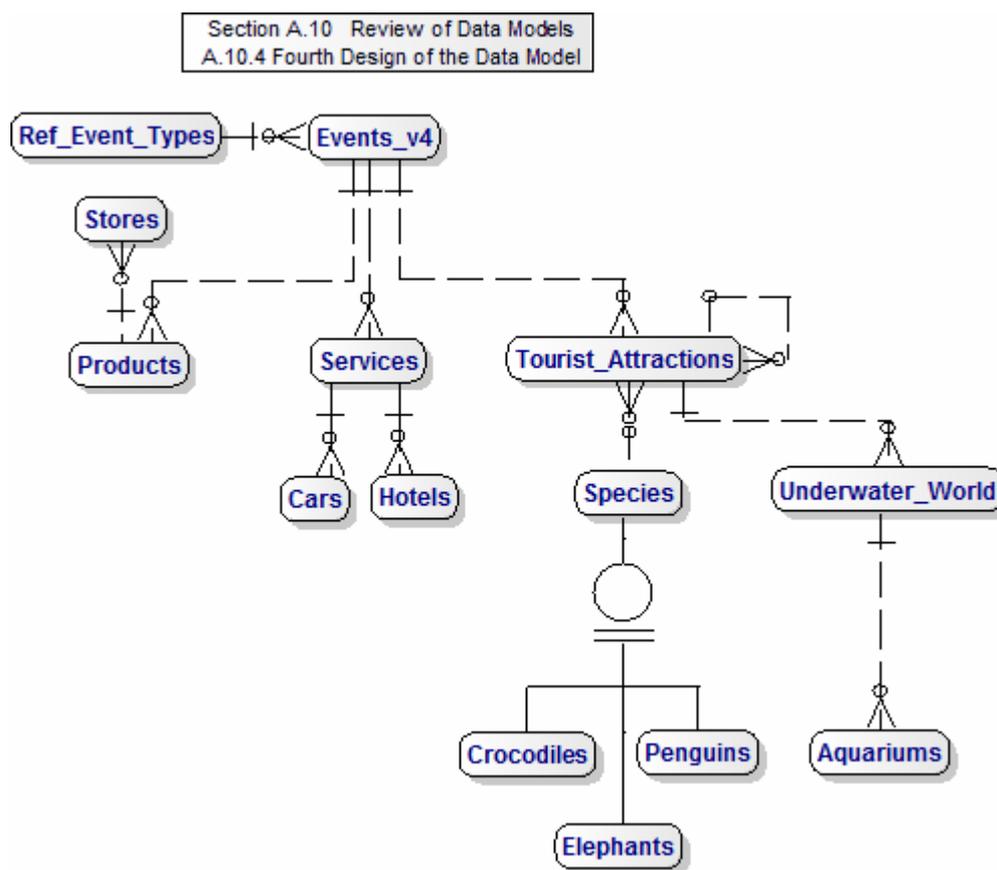
At this stage, we start thinking about the story we want to tell and how we should structure the Data Model to reflect the story.

For example, we would say ‘Products are found in Stores’ and therefore we should show a one-to-many relationship between Stores and Products.

At this point, we like the left-hand side of the Data Model and we can say :-

1. There is a ‘Go Shopping’ Event which means buying Products, rather than go to Stores.
2. This allows us to show Products at the same level as Stores which means that we can maintain consistency with the terminology we have used elsewhere in this document.
3. We have decided that Services included “Hiring a Car” and ‘Staying in a Hotel”.

So now we turn to the right-hand side of the Data Model.



6.8.7 Fifth Data Model

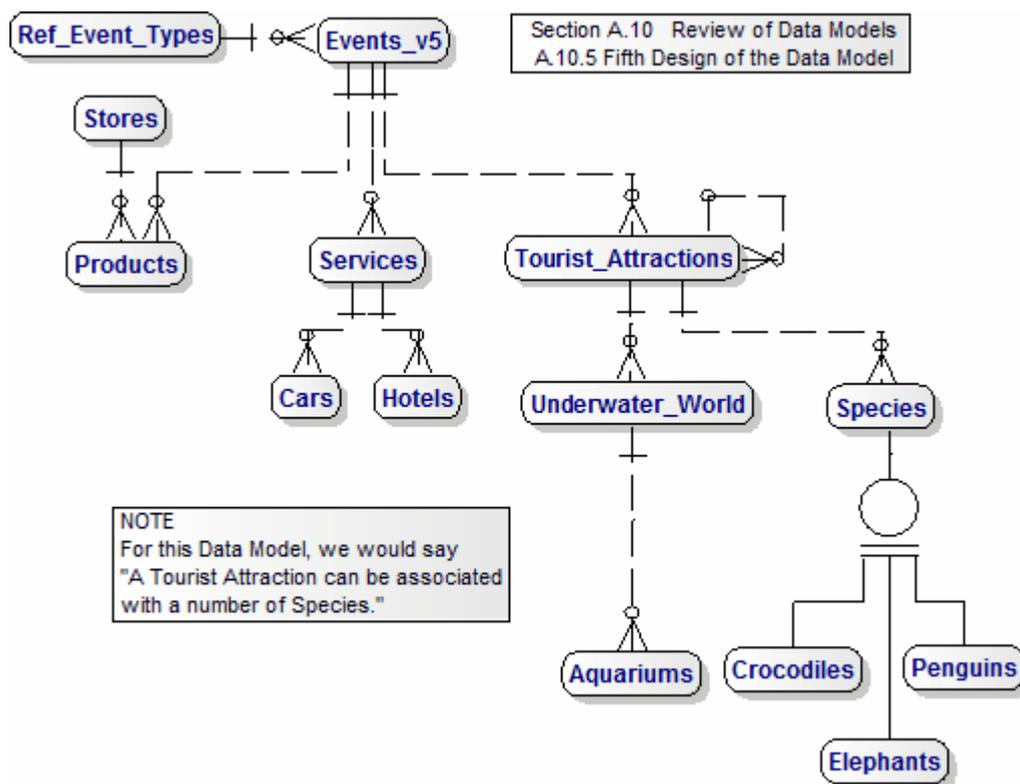
At this stage, we start thinking about the story we want to tell and how we should structure the Data

On the right-hand side, we show Species as a Tourist Attraction, whereas in fact, people say ‘Let’s go the Elephant Sanctuary’ so the entity that should be related to the Tourist Attraction is the Elephant Sanctuary, rather than the Elephant.

Of course, people can also say “Let’s go and see the Elephants” but in the physical world, the reality is that Elephants are housed in an Elephant Sanctuary and it the Sanctuary that they actually go and see.

So we conclude that it is OK to show the Tourist Attractions and their relationships.

We still have one or two questions to resolve but we are quite content because the Model looks good and the overall logic is good.



6.8.8 Conclusion

We hope you have found this discussion interesting and useful.

If you have any comments, suggestions or questions, please feel free to email us at :-

- barry@databaseanswers.org.

7. Retail Sales

7.1 Development Framework

In this Section, we discuss the Framework which contains two elements :-

3. The Development Approach
4. Components used in the Approach

The Approach involves these Steps :-

6. Identify the business Events
7. Define a Message for each Event
8. Map each Event to the Entities and Attributes in the CDM
9. Determine whether the CDM should be extended.
10. Create an Industry-specific CDM if appropriate.

The Components include :-

- Generic and Industry-specific Canonical Data Models (CDMs)
- Generic and Industry-specific Data Warehouse designs
- Core Entities and Core Subject Area Models – Customers, Products and Suppliers.

The rest of this Section is a Case Study for a Retail business.

Subject Area Models :-

- Generic (Horizontal)
- Industry-specific -
 - Insurance Common Data Model
 - http://www.databaseanswers.org/data_models/insurance_data_warehouses/common_data_model.htm

7.2 Data Architecture

This Architecture shows the major components in designing a Data Warehouse that incorporates Enterprise Data Model with associated Subject Area Models, based on Industry-specific Models.

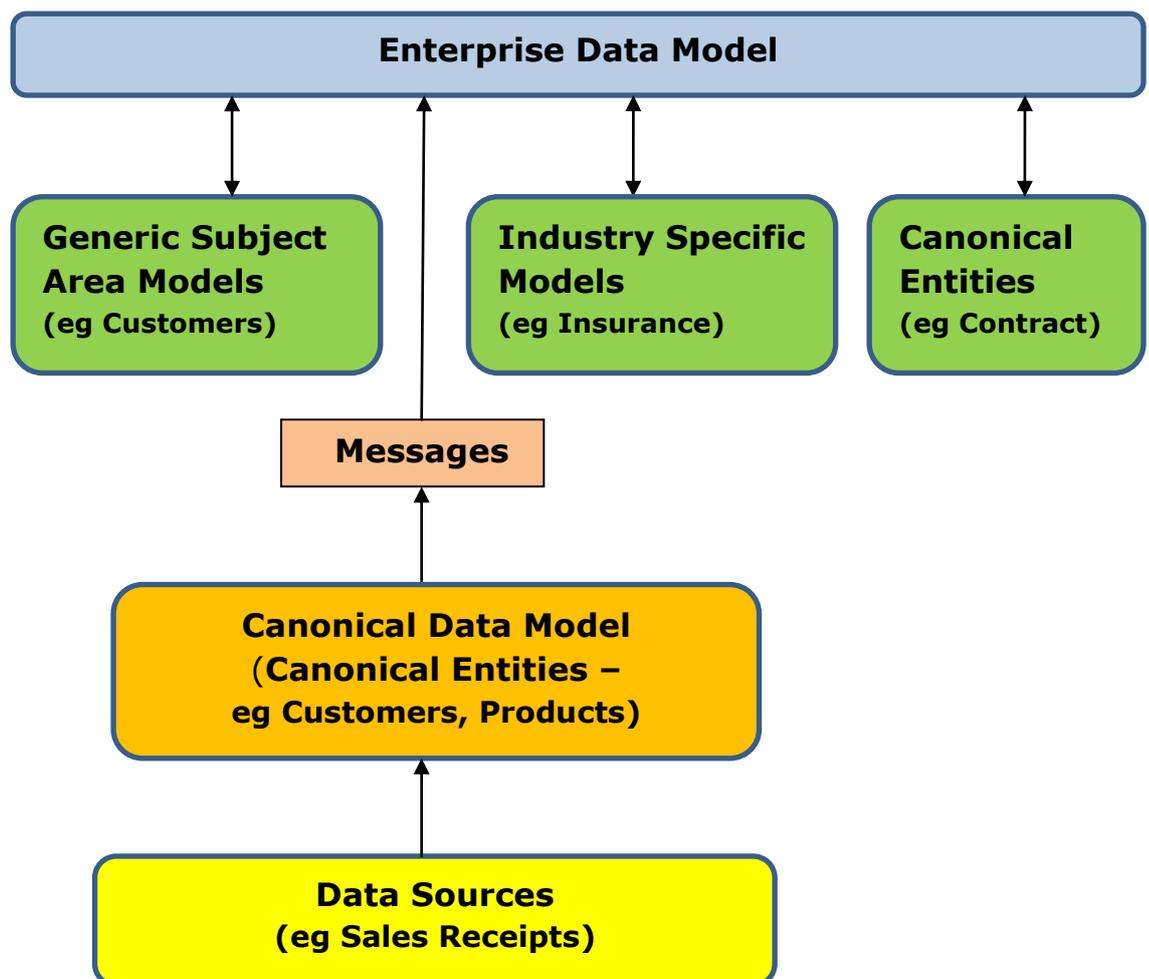
Each Data Source is reviewed against the Canonical Data Model and the appropriate Messages formats are defined. Then the data in the Message is mapped to the Enterprise and Industry-specific Models.

The current Enterprise Data Models are defined on this page of the Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/enterprise_data_models.htm

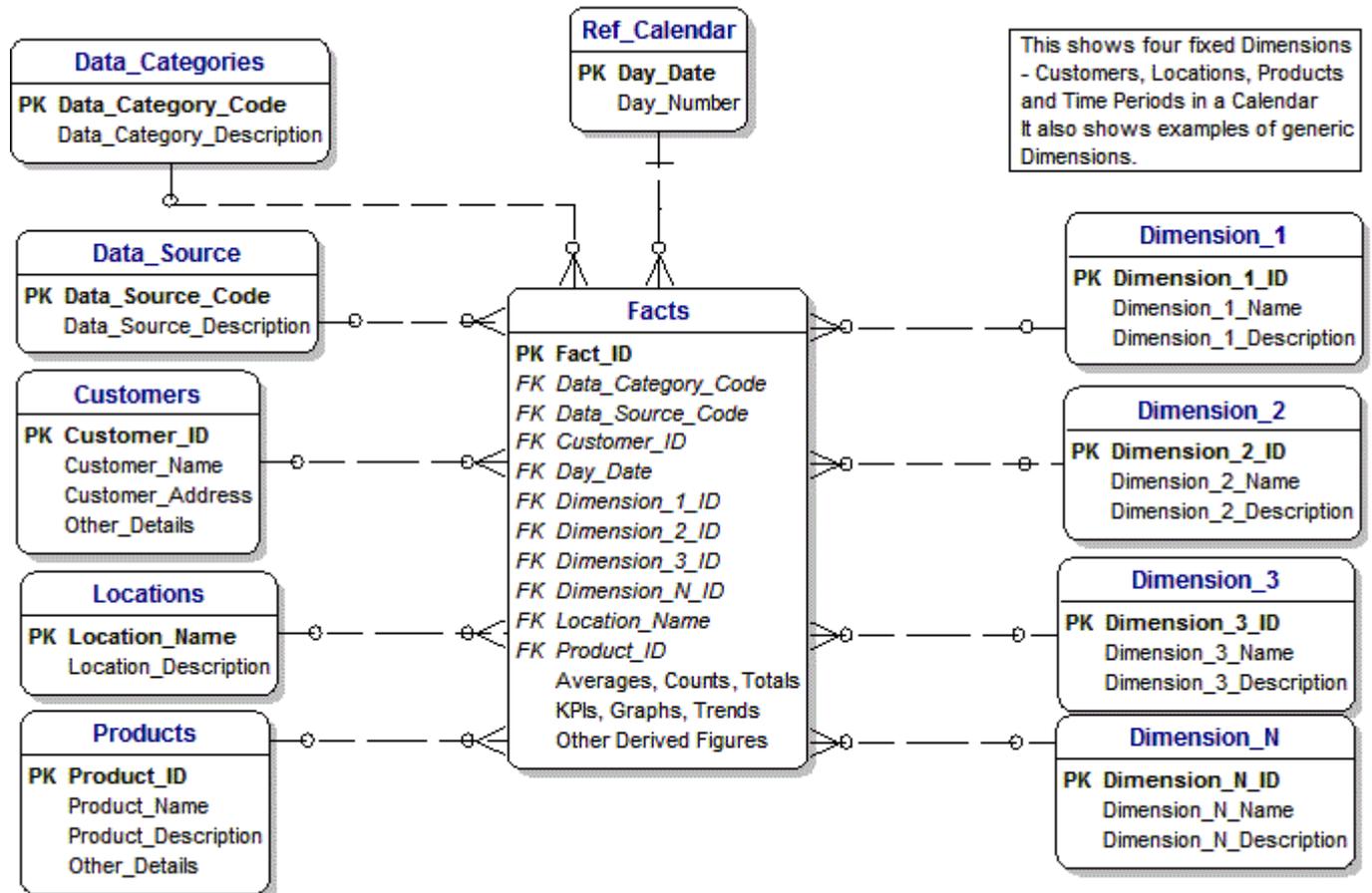
And the Industry-specific Models are on this page :-

- http://www.databaseanswers.org/data_models/industries_index.htm



7.2.1 Generic Data Mart

This should be reviewed for consistency with the Canonical Data Mart.



7.3 Canonical Data Model

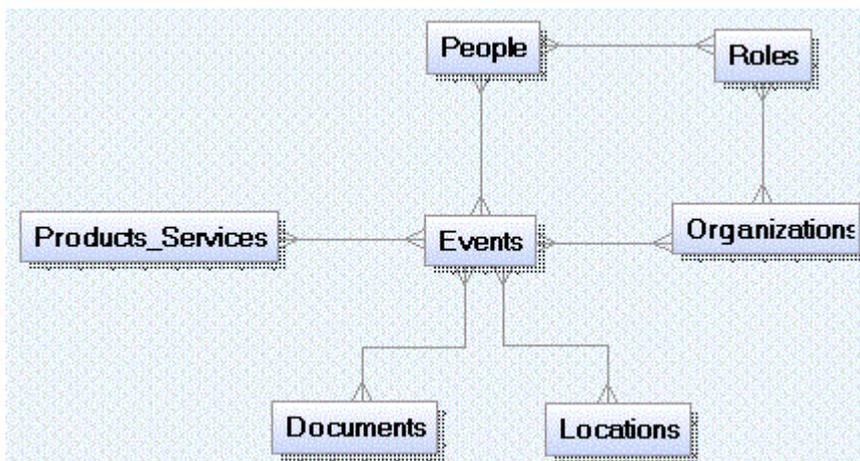
In this Section, we discuss how a Canonical Data Model (CDM) can be used to provide a frame of reference and a technique to generate standardised Messages and a 'flattening of the data ' to be loaded into a Third-Normal Form Data Warehouse and then a Dimensional Data Mart.

During this application of the CDM and determine whether we need to extend to meet the specific requirement

We will then achieve validation of the CDM which will be a very important result.

Our Model is taken from this page on the Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm



In passing, we should add that we are discussing only purchases made by Customers in a Store.

Online Purchases can be added later.

7.5 Discussion of a Sales Receipt

The next Section shows a Sales Receipt from TK Maxx, which a major UK retailer, owned and operated by TJX, a multi-national US Retailer.

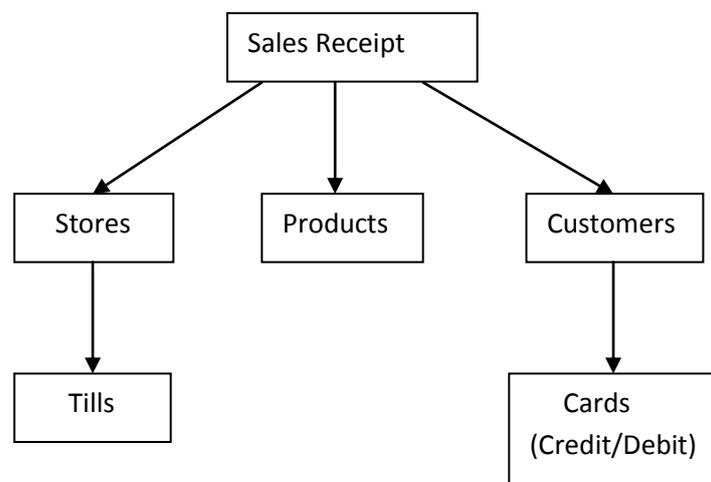
When we examine the Receipt, we can see the following item of data :-

- Credit/Debit Card –
 - Type (eg Amex, Visa) last 4 digits, Expiry Date, Transaction Authorisation Number
- Merchant ID, Terminal (Till) ID
- Product Category, Number and Prices of all the Products I purchased
- Purchase Date and Time of my Purchase
- Store Number, Name and Address and phone number

We can consider this to be our Operational Data Store, and from it we can generate details for the following 'Things of Interest' :-

- Customer Cards
- Products
- Purchases
- Stores

In other words, the structure of the data looks like this :-



Of course, when a Customer uses a Card then we know some details of name, address and so on.

Elephants, Olympic Judo and Data Warehouses

For the purpose of this book, we assume we know some basics.

If the same Customer uses more than one Card then we will need to recognise the same Customer, which can become a little complicated if he is Joe Bloggs and Joseph Bloggs or J.Bloggs on another card.

A fuller discussion of this topic can be found under the heading of Master Data Management :-

- http://www.databaseanswers.org/mdm_master_data_management.htm

If a Customer pays by cash then we know nothing about them, and we would consider this to be an 'Anonymous Customers'.

7.6 Example of a Sales Receipt

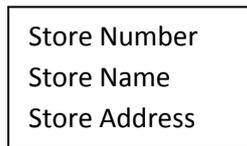


7.7 Step 1 – Set up Master Data

7.7.1 Add a Store Entity

7.7.1.1 Message

The Message for this Event includes the Store Number, Name and Address.



7.7.1.2 Enhance the Data Model

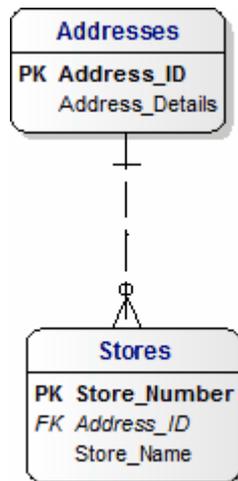
This diagram shows the new Stores Entity.

We have created an Addresses Entity to hold details of all our Addresses.

There are two benefits to this :-

1. An Address stored in a separate Table can be validated by third-party commercial software, such as QAS.
2. A single Address appears only once if it is used many times.

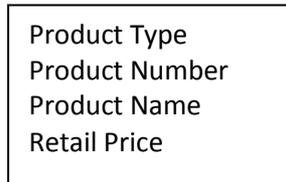
This occurs frequently for people in families, but rarely for Stores.



7.7.2 Add a Product

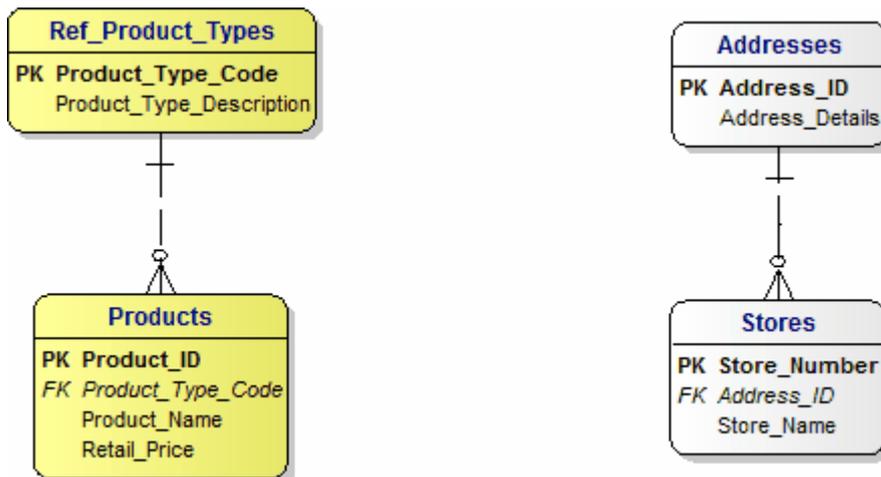
7.7.2.1 Message

The Message for this Event includes the Product Type, Number, Short Name and Retail Price.



7.7.2.2 Add Products to the Data Model

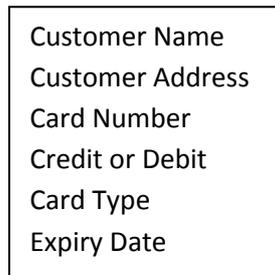
Here we add the Products Entity (shown in yellow) but, of course, there is no relationship between Products and Stores at this point.



7.8 Step 2 - Add a Customer Entity

7.8.1 Message

The Message for this Event includes the Customer Name, Address, Card Number, Merchant Name, Amount, Date and Time of Purchase.



7.8.2 Add Customer Entities to the Data Model

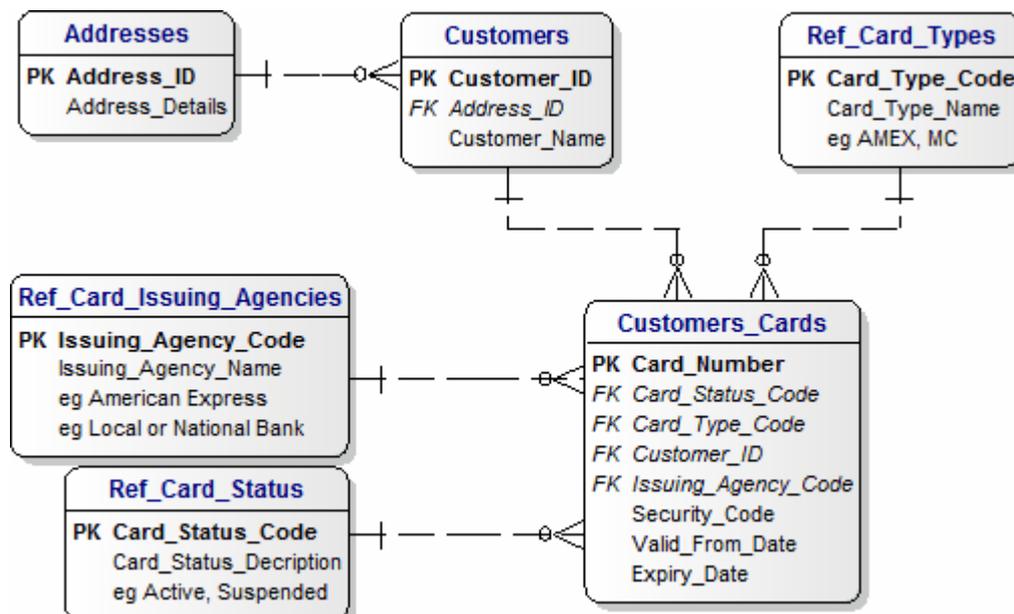
This diagram shows the new Customers and Cards Entities.

You can see that we have added the Customer's Address to the Addresses table.

This, of course, means that we are deriving increased value from the fact that we have created a separate Addresses Table.

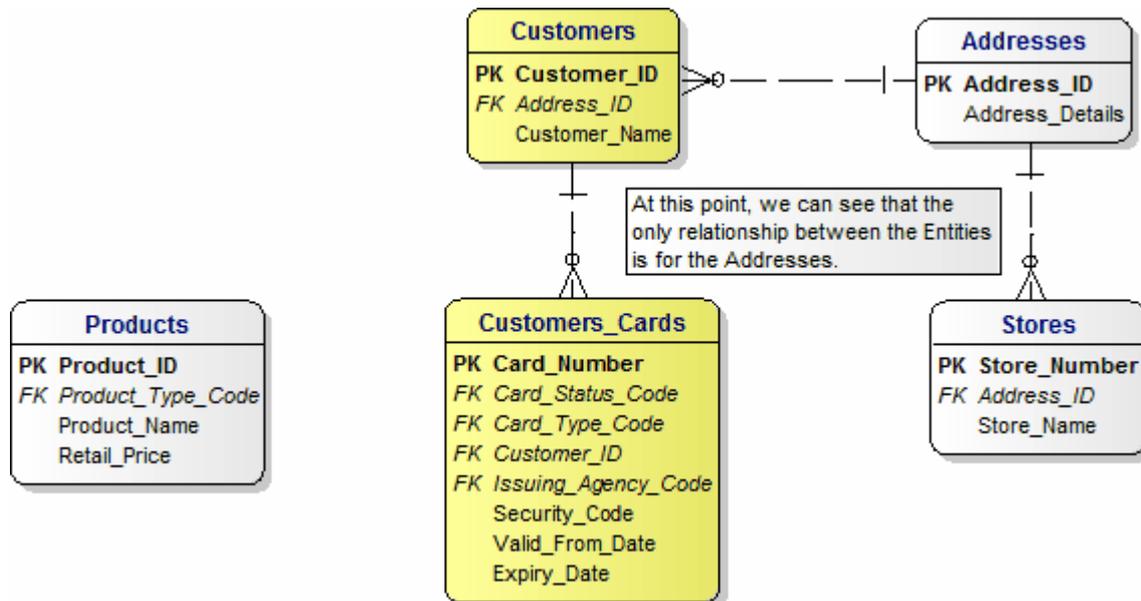
We have added a few attributes to the Customers_Cards Entity, such as the Valid_from_Date that we can see when we look at one of our physical Cards, even though it does not appear on the Sales Receipt.

This might not be in accordance with the Agile approach but it seems difficult to ignore when we think about it.



7.8.3 The Data Model without Reference data

In this diagram we have not shown the Reference Data Entities because they add too much to the detail which makes the diagrams more difficult to understand at a glance without adding significantly to the information contained.



7.9 Event 1 – A Customer makes a Purchase

At this point, we have created the Entities for Customer, Products and Stores
But, of course, they are not related.

Now we add the Event of a Customer Purchase and this will provide relationships between all the Entities.

7.9.1 Message

At this point, the Customer Purchase generates the Sales Receipt that records the details that we have examined up to this point.

Therefore we add the Customer Purchase entity that we will use to add details of Cards, Products, Purchase and Stores.

We have separated the details of the Store, Products and Customer and put them in separate Entities. Therefore, they are replaced by links in ID fields to the other Entities.

This results in a simple elegant Data Model that makes it very clear what the Entities are and how they are related.

When the time comes, it will be very straightforward for us to generate a Physical Data Model, with the SQL to create the Tables in a Database.

Elephants, Olympic Judo and Data Warehouses

The Message for this Event includes the details of the Store, Products, Customer Card, and the Date and Time of Purchase.

Sales Receipt for Customer Purchase

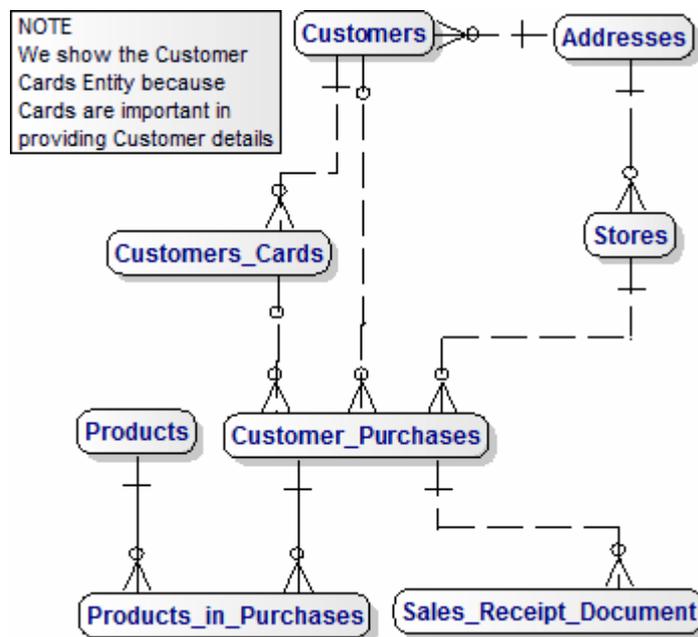
- Store Details
- Product Details
- Customer Card Details
- Date and Time Details

7.9.2 The Complete Data Model (Entity names only)

This diagram shows how the Customer Purchase is the Event that ties all the data together, and is, of course, reflects the business event that is the foundation of the business.

This Model shows only the Entity names.

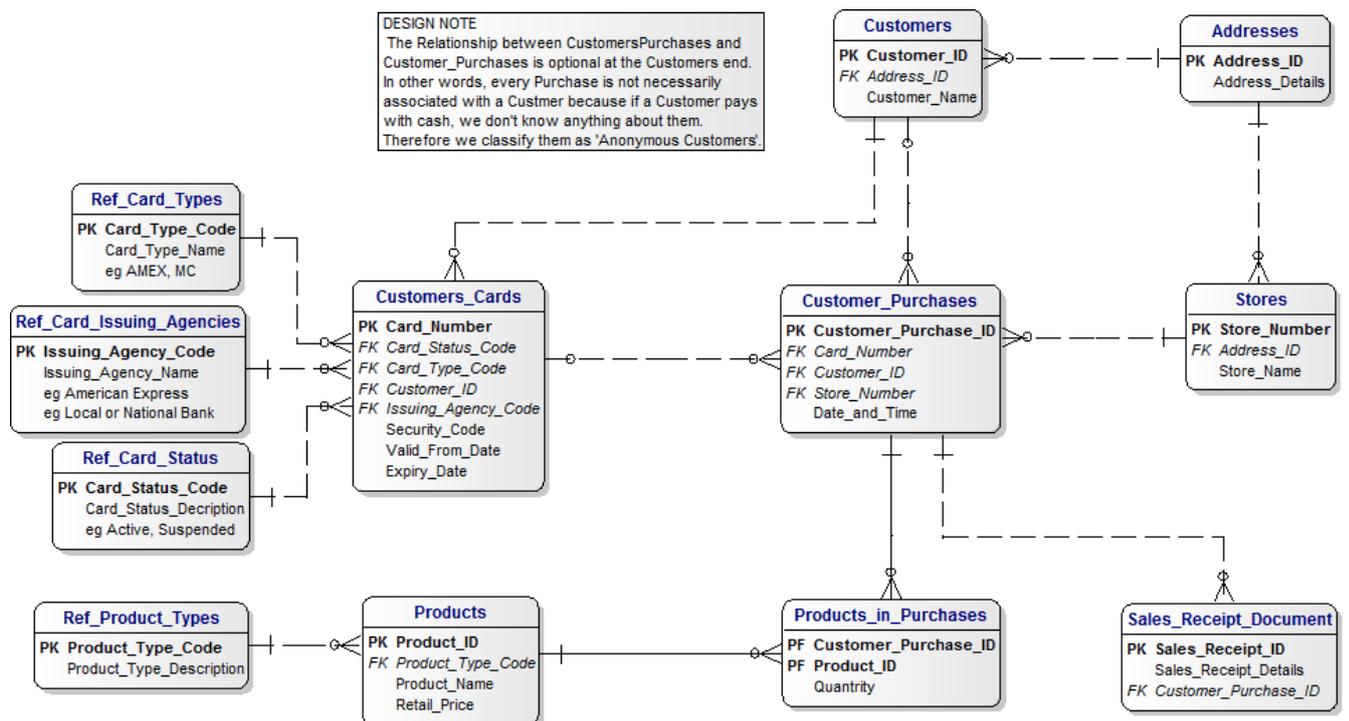
This makes it suitable for discussion with business users and Subject Matter Experts.



7.9.3 The Complete Data Model (showing Ref Data and Attributes)

This version of the Model shows the Entity names, the Attributes and the Reference Data.

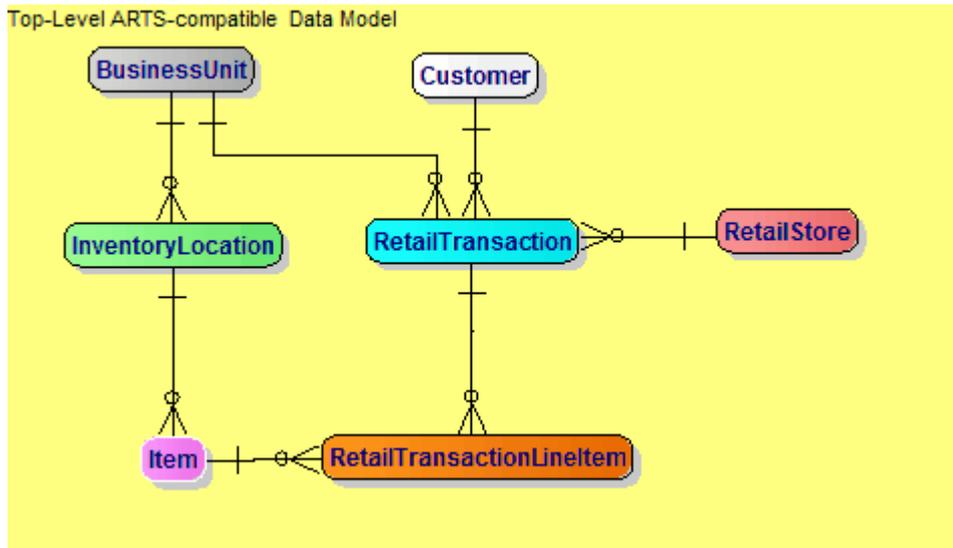
This makes it suitable for discussion with developers and anybody interested in the details, especially of the relationships between the Entities.



7.9.4 ARTS Model

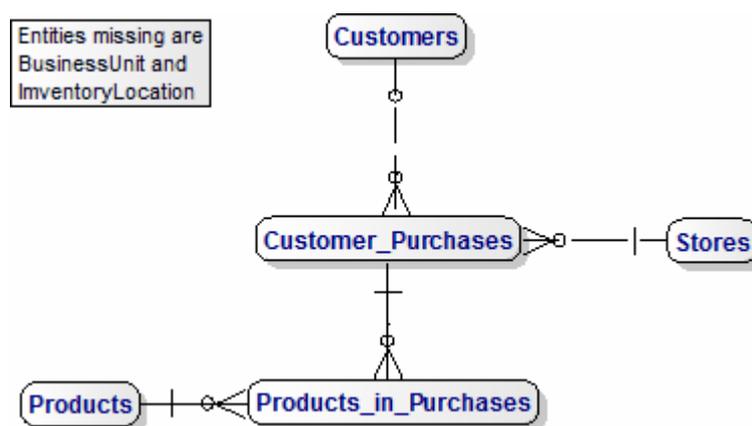
This version of the Model shows the Entity names, the Attributes and the Reference Data, and is taken from this page :-

- http://www.databaseanswers.org/data_models/arts_retail_and_can_data_model/arts_and_canonical_model.htm



7.9.5 Canonical Model aligned with ARTS Model

This version of the Model shows only the Entity names to facilitate comparison with the ARTS Model.



Elephants, Olympic Judo and Data Warehouses

7.10 What have we Learned ?

In this Tutorial we have learned how to follow a simple Step-by-Step approach to the design of a Third Normal Form Data Warehouse.

We would work with the business users and Subject Matter Experts to establish the important Events and then to agree the data items that should appear in each Message.

Our starting-point is that we know the Approach because we have used it before.

Therefore we might consider running Facilitated Workshops to guide users to a successful conclusion.

8. Retail Banks

Here we describe a simple Event-Driven Approach to Data Warehouse Design.

It also appears on our Database Answers Web Site with this Data Model for Retail Banks :-

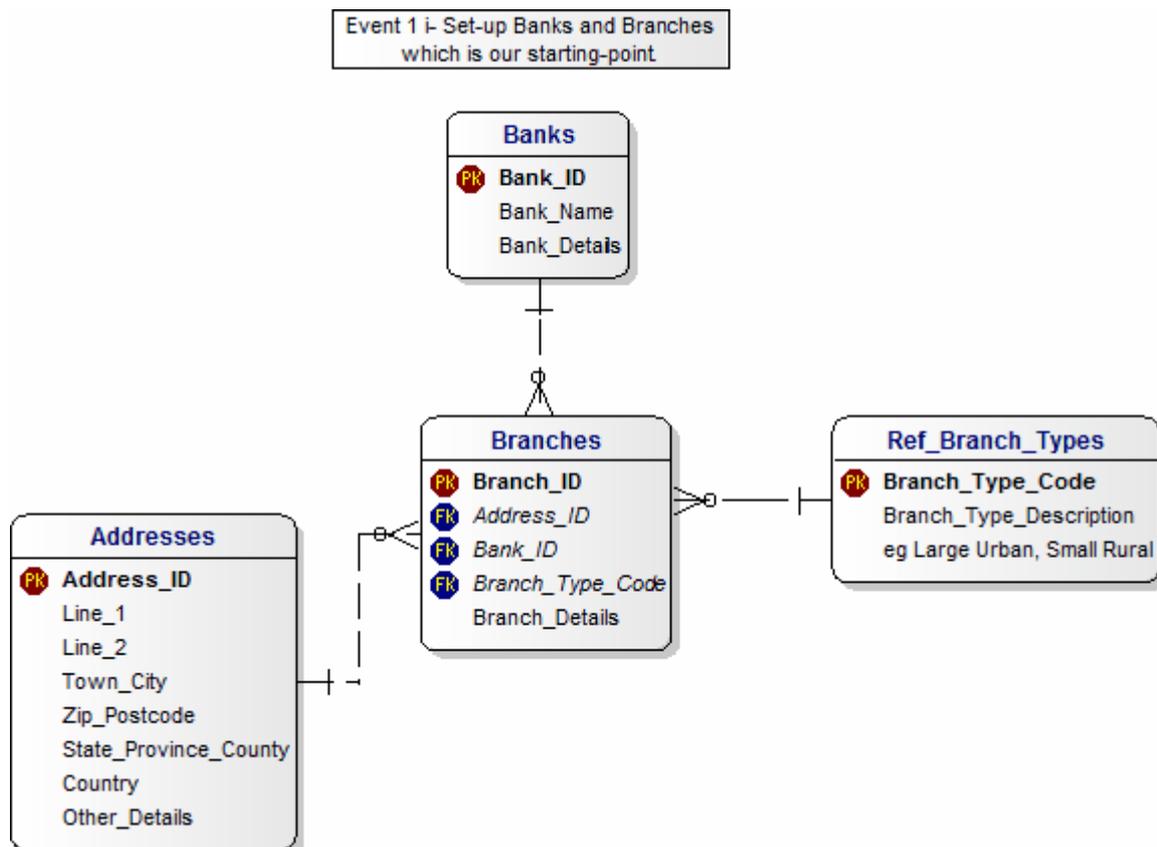
- http://www.databaseanswers.org/data_models/retail_banks/index.htm

The Steps involved are as follows :-

- Identify the Events involved in the Life Cycle of important Entities
- Define a Message with data for each Event

8.1 Event 1 - Set-up Banks and Branches

We know that the Bank has Branches so we define our starting-point accordingly.



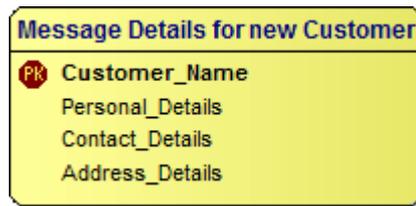
8.2 Event 2 – Create a new Customer

Now we create a Customer.

8.2.1 Message

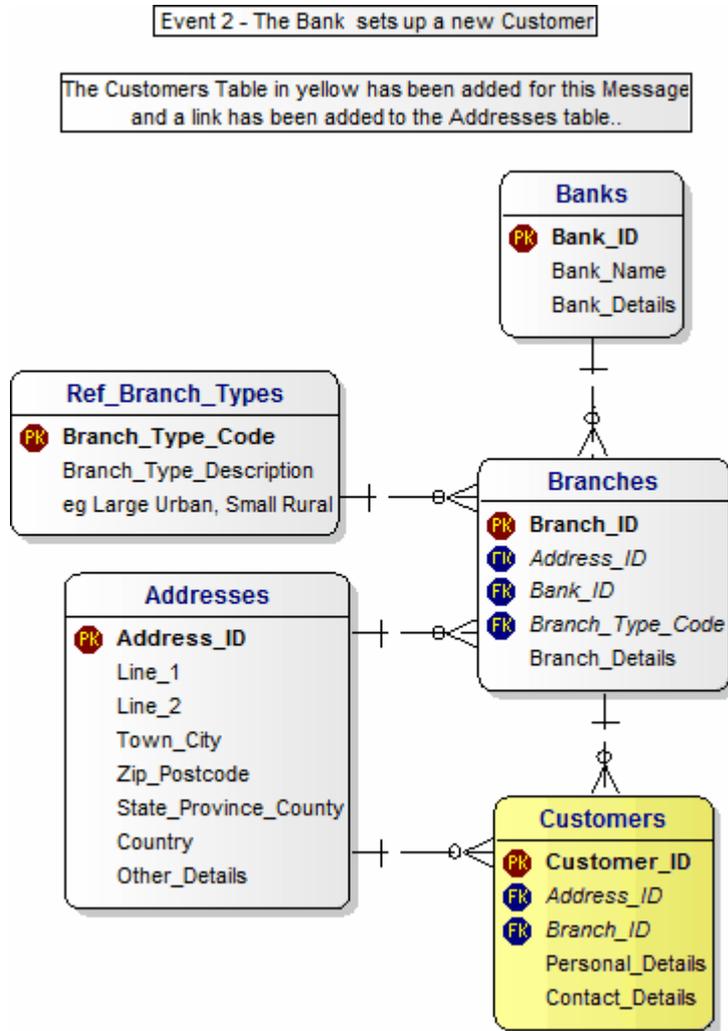
The Message for this Event includes Personal, Contact and Address Details.

First we create a new Customers Entity, then match the Address to the Address Entity and add the Address_ID field to the new Customers Entity.



8.2.2 Enhance the Data Warehouse

This diagram shows the new Customers Entity in yellow.



8.3 Event 3 – Set up a new Account

We have created a new Customer, so now we can create an Account for the new Customer.

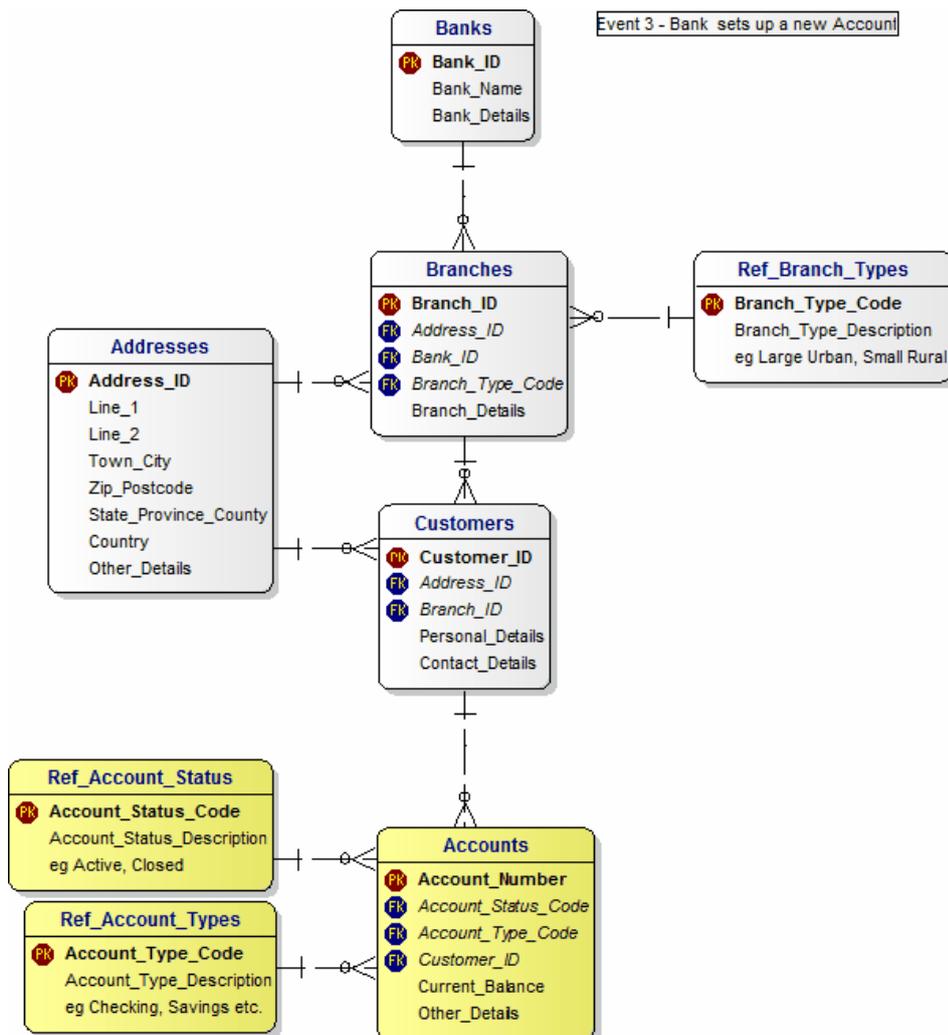
8.3.1 Message

The Message for this Event includes the data required to set up a new Account.



8.3.2 Enhance the Data Warehouse

This diagram shows the new Accounts Entity in yellow.



Elephants, Olympic Judo and Data Warehouses

8.4 Event 4 – Issue a Credit Card

The fourth Event is to issue a Credit Card.

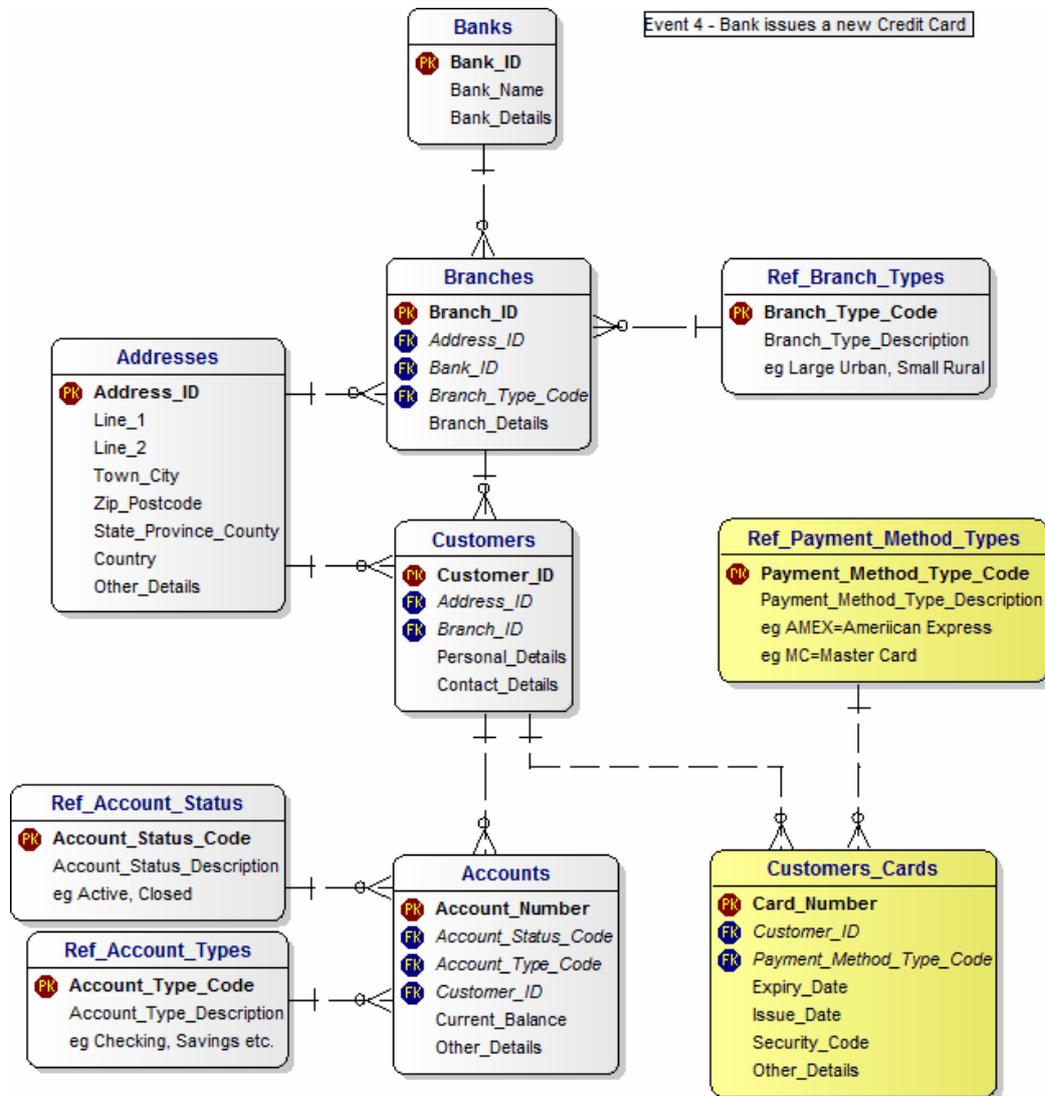
8.4.1 Message

The Message for this Event includes the Credit Card number and type and the Account Number that the Card is associated with, and the opening date and expiry dates

Message Details for a new Credit Card
Card_Number
Customer_Number
Payment_Method_Type_Code
Expiry_Date
Issue_Date
Security_Code
Other_Details

8.4.2 Enhance the Data Warehouse

This diagram shows the two new Entities in yellow.



Elephants, Olympic Judo and Data Warehouses

8.5 Event 5 – Customer deposits money in the Account

The fifth Event is for the Customer to deposit money into the new Account.

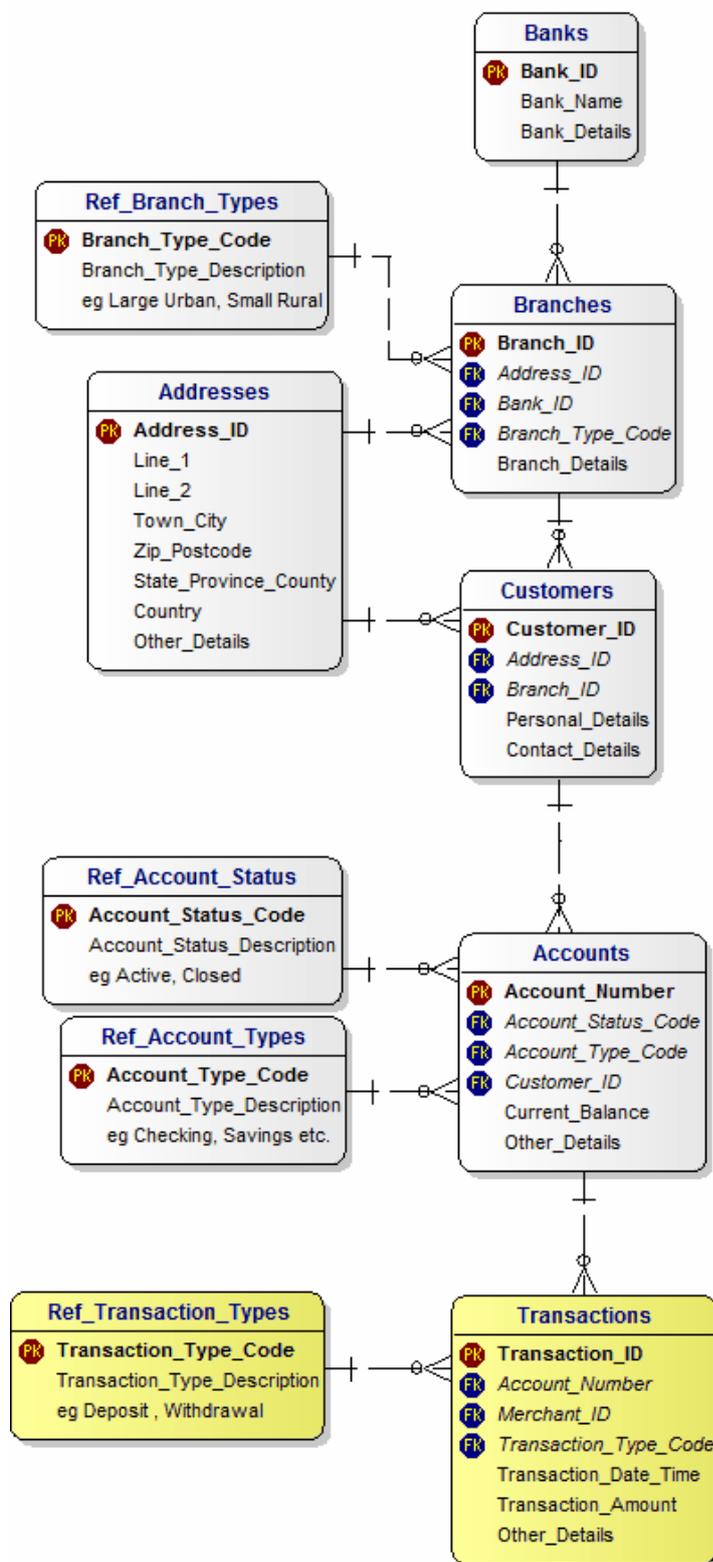
8.5.1 Message

The Message for this Event includes the Account number, the date and the amount of the deposit



8.5.2 Enhance the Data Warehouse

This diagram shows the new Transactions Entity in yellow.



Elephants, Olympic Judo and Data Warehouses

8.6 Event 6 – Customer uses new Card in a Retail Store

8.6.1 Message

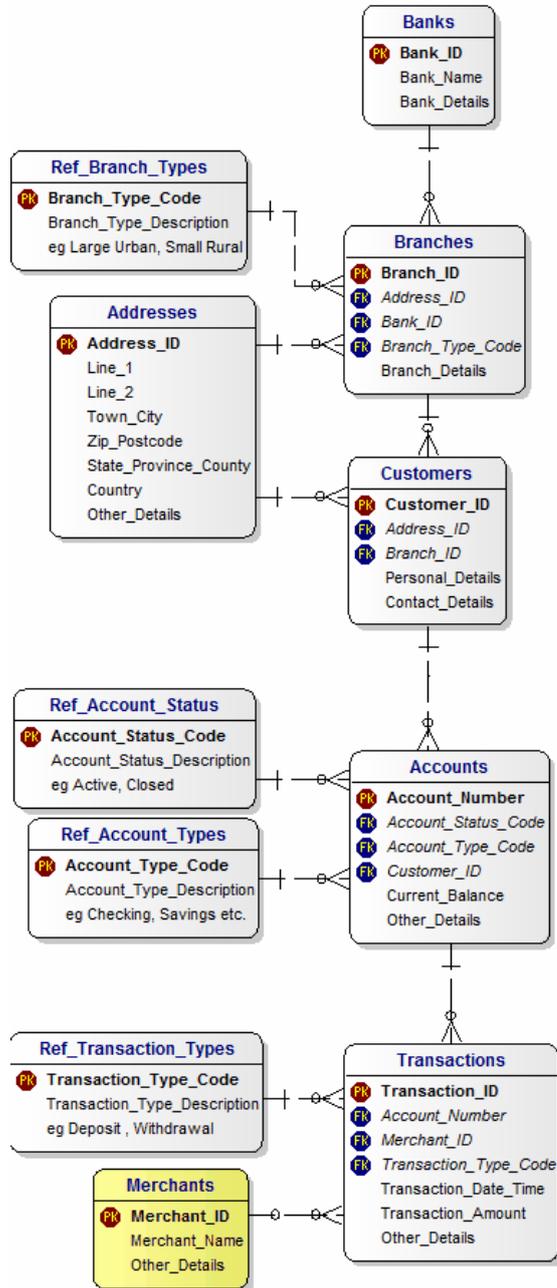
The Message for this Event show what is printed on atypical Sales Receipt, including Card Number, Merchant Name, Amount, Date and Time of Purchase.

Message Details for Sales Receipt
Receipt_ID
Merchant_Name
Purchase_Date_Time
Purchase_Amount
Card_Number_last_4_digits
Other_Details

Elephants, Olympic Judo and Data Warehouses

8.6.2 Enhance the Data Warehouse

This diagram shows the new Merchants Entity in yellow.



8.7 Event 7 – Bank issues a Monthly Statement

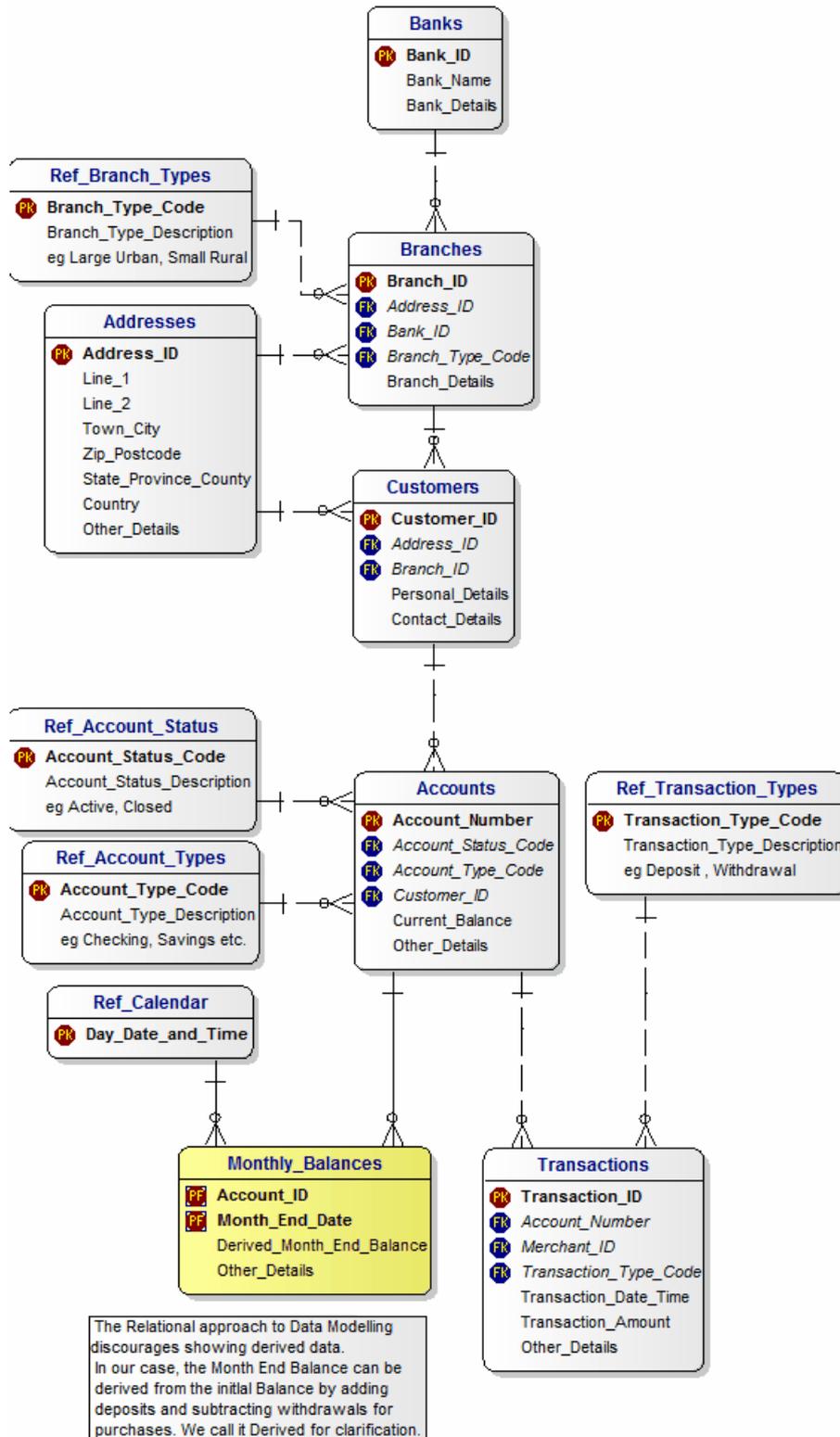
8.7.1 Message

The Message for this Event includes Customer name, Account Number, Date and Amount of Statement.

Message Details for a Monthly Statement
Customer Name
Account_Number
Month_End_Date
Derived_Month_End_Balance
Other_Details

8.7.2 Enhance the Data Warehouse

This diagram shows the new Monthly Balance Entity in yellow.



Elephants, Olympic Judo and Data Warehouses

8.8 Event 8 – Customer Closes Account

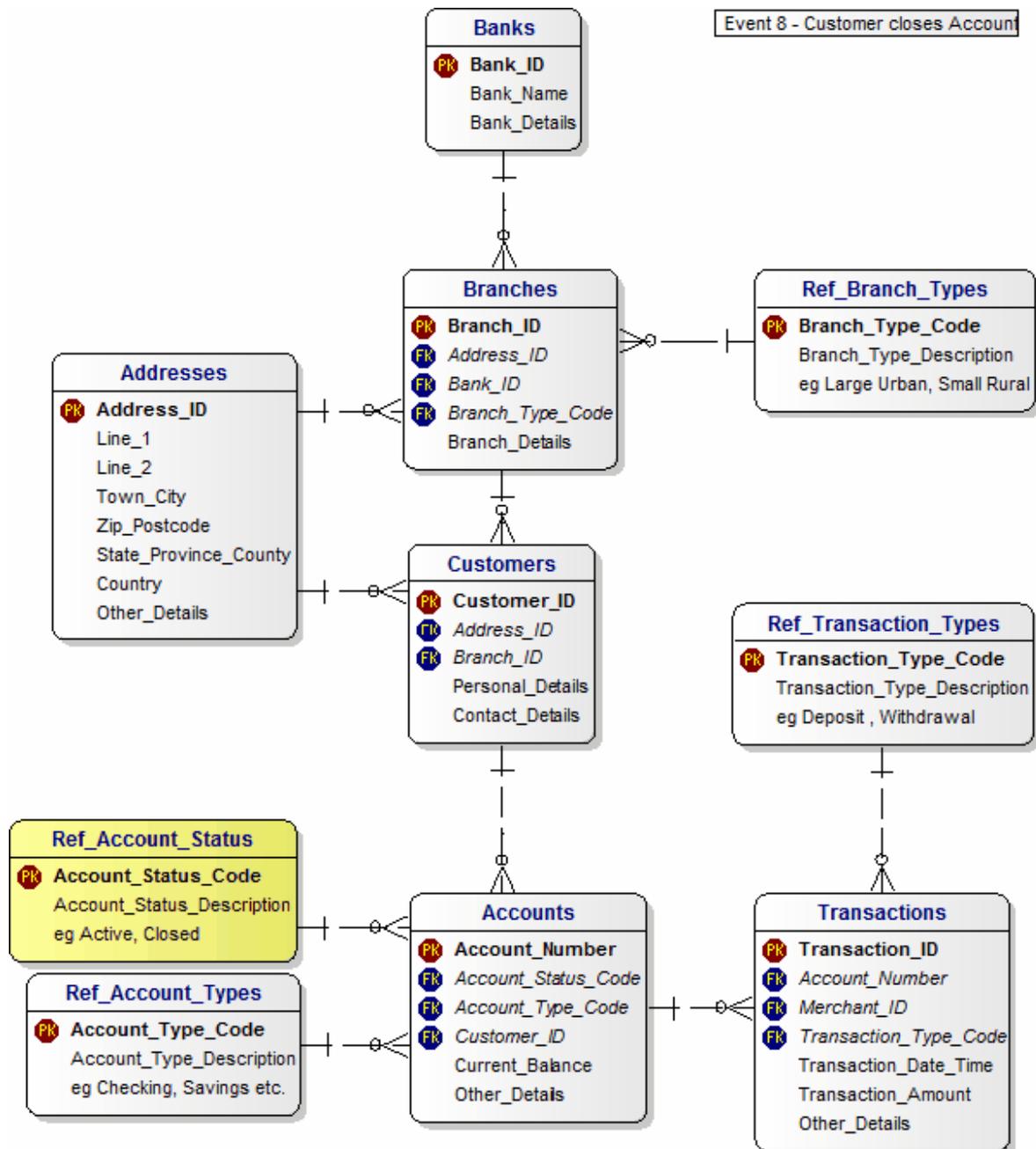
8.8.1 Message

The Message for this Event includes Account Number and Date to be Closed.

Message Details for closing an Account
Event : Change Account Status
Account Number
New Account Status Code
Date to be Closed

8.8.2 Enhance the Data Warehouse

This diagram shows the new Account Status Entity in yellow.



9. Industry Data Models

9.1 Introduction

We know from our experience that in a greenfield situation, the most workable approach is to start with a top-level summary Model and then to create Subject Area Models as appropriate to the Client's requirements.

We can call these "Kick-Start Models" that can be used as a starting-point for specific clients.

9.2 Banking

You will find this page interesting because it discusses Banking Data Warehouses – http://www.databaseanswers.org/data_models/banking_data_warehouses/index.htm

Here is a Model of Federated Data Marts for Banking - http://www.databaseanswers.org/data_models/banking_data_warehouses/federated_data_marts.htm

and this one on Banking Acquisitions - http://www.databaseanswers.org/data_models/banking_acquisitions/data_warehouse.htm

9.2.1 Investment - Links

The bottom of this page on Investment Banking Models shows three versions of a Data Warehouse as it evolved through facilitated Workshops with Client Management :

http://www.databaseanswers.org/data_models/investment_banking/index.htm

Investment Banking Enterprise Data Model with subject areas -

http://www.databaseanswers.org/data_models/enterprise_data_model_for_investment_banks/index.htm

Here's a Data Warehouse for Private Banking :-

http://www.databaseanswers.org/data_models/private_banking/data_warehouse.htm

It includes this Data Warehouse for Investment Banking

http://www.databaseanswers.org/data_models/banking_data_warehouses/investment_banking_data_warehouse_final_version.htm

9.2.2 Investment - Settlements

Settlements are somewhat like Payments but money can be exchanged both ways and the details can be more complex than conventional Payments. Therefore, we categorise Settlements as specific to Investment Banking

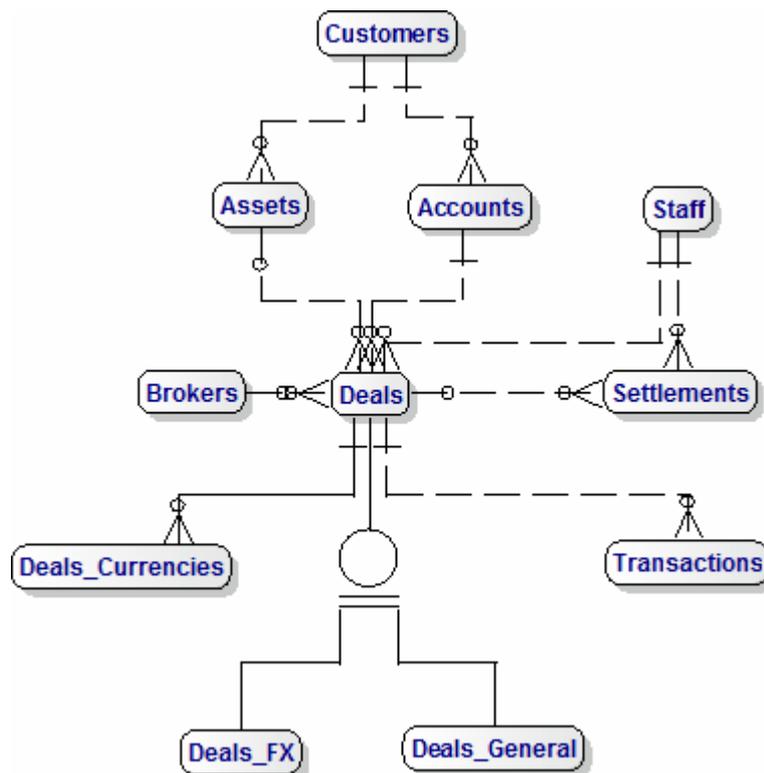
- Link to the Settlements Data Model- http://www.databaseanswers.org/data_models/investment_banking/settlements.htm

9.2.3 Investment – 3NF Model

It is rather surprising to find that the majority of Subject Area Models are Generic. However, based on our experience over ten years of working with Investment Banks, we would say that ‘the devil is in the detail’. In other words, once we start to develop more detailed Models for a specific Bank, we would find a great deal of bank-specific detail.

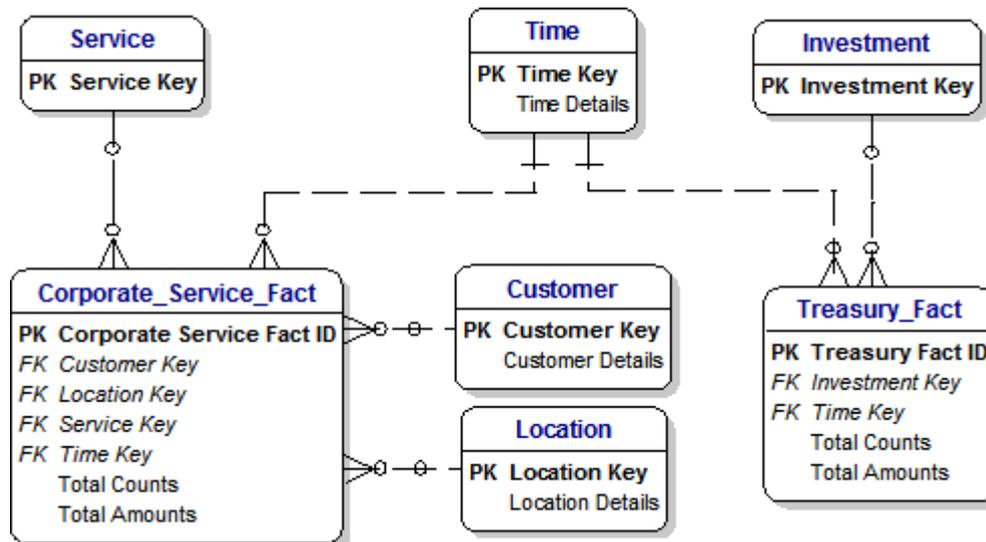
This discussion is recorded on this page –

- http://www.databaseanswers.org/data_models/enterprise_data_model_for_investment_banks/index.htm



9.2.4 Investment – Corporate Service and Treasury Data Models

This Model shows Federated Data Marts for Corporate Service and Treasury with just one shared Conformed Dimension, which is Time.



9.2.4 Link to Settlements Subject Area Model

Settlements are somewhat like Payments but money can be exchanged both ways and the details can be more complex than conventional Payments. Therefore, we categorise Settlements as specific to Investment Banking

Settlements - http://www.databaseanswers.org/data_models/investment_banking/settlements.htm

9.2.5 Retail Banks - Links

Data Warehouse for Retail Banks :-

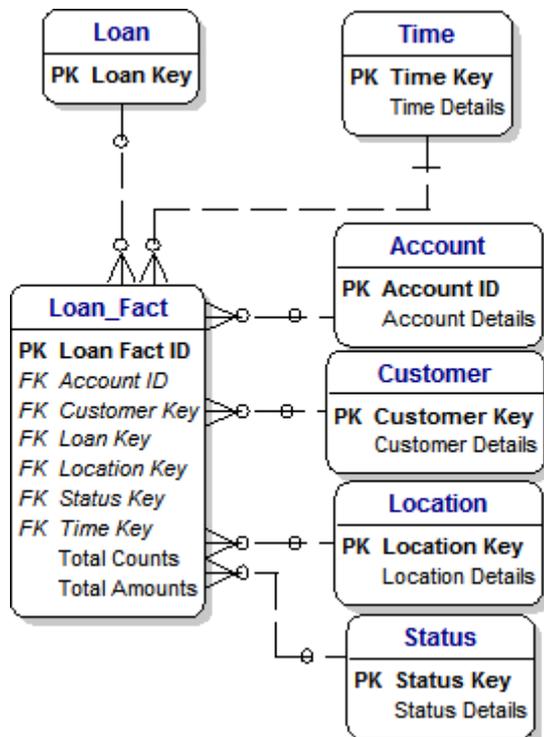
http://www.databaseanswers.org/data_models/retail_banks/data_mart.htm

3NF ERD for Retail Banks :-

http://www.databaseanswers.org/data_models/retail_banks/index.htm

9.2.5 Customer Loans

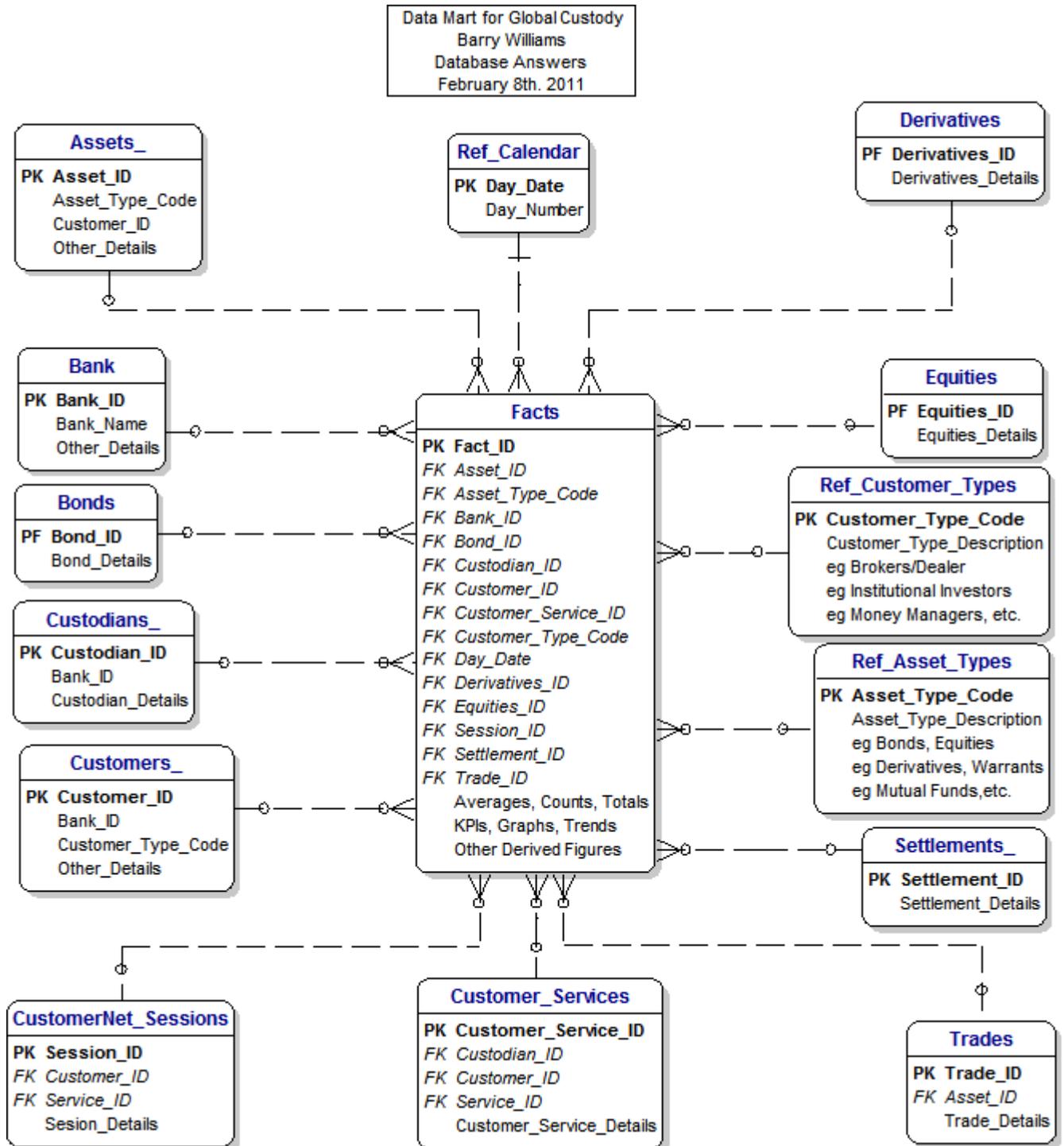
The Loan_Fact Entity shows a Dimensional Model for typical Dimensions of Account, Customer, Time and so on.



Elephants, Olympic Judo and Data Warehouses

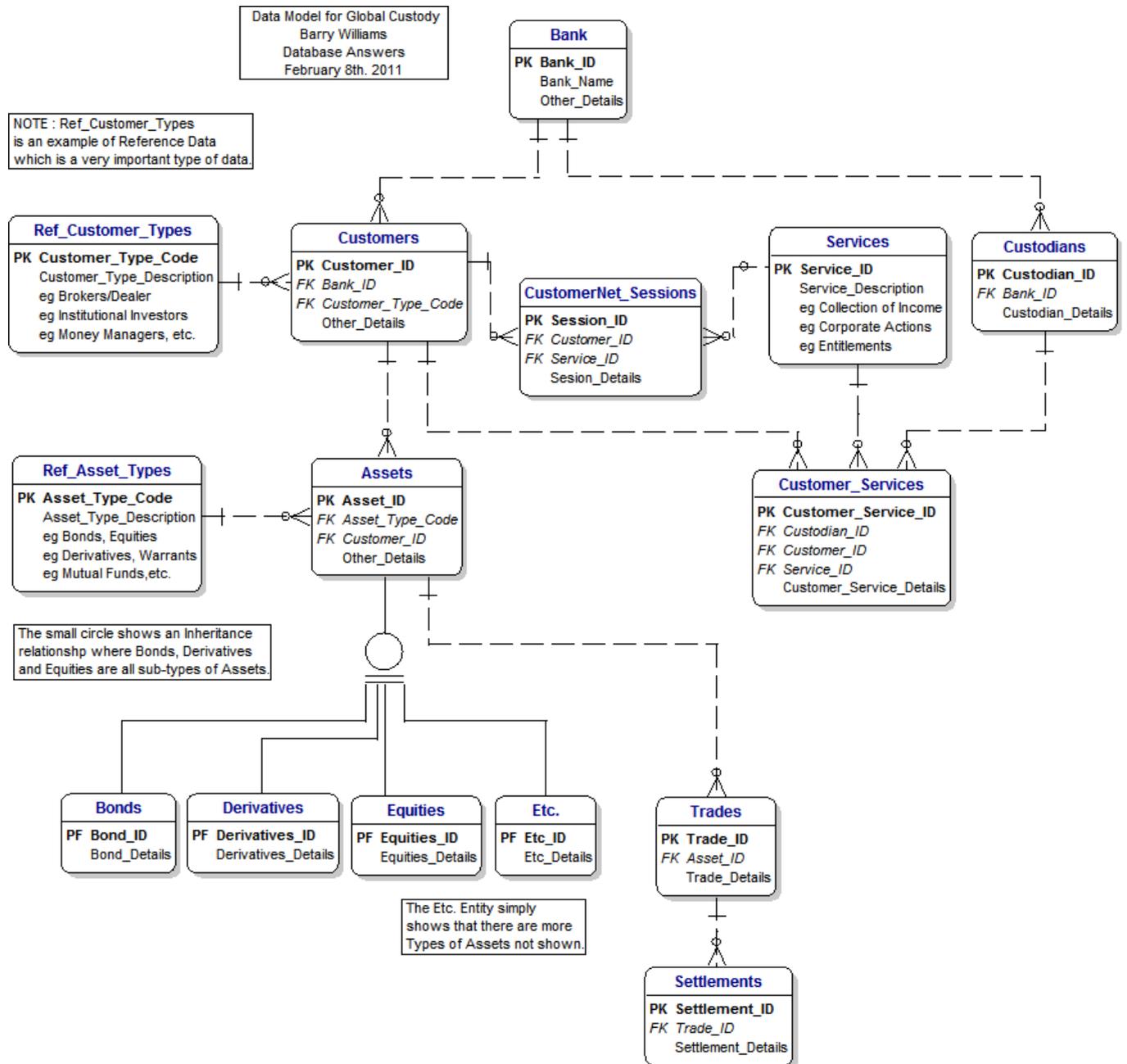
9.3 Finance

Under the heading of Finance, here's a Dimensional Data Mart for Global Custody in an Investment Bank - http://www.databaseanswers.org/data_models/global_custody/data_mart.htm



Elephants, Olympic Judo and Data Warehouses

The corresponding 3NF Data Model looks like this :-



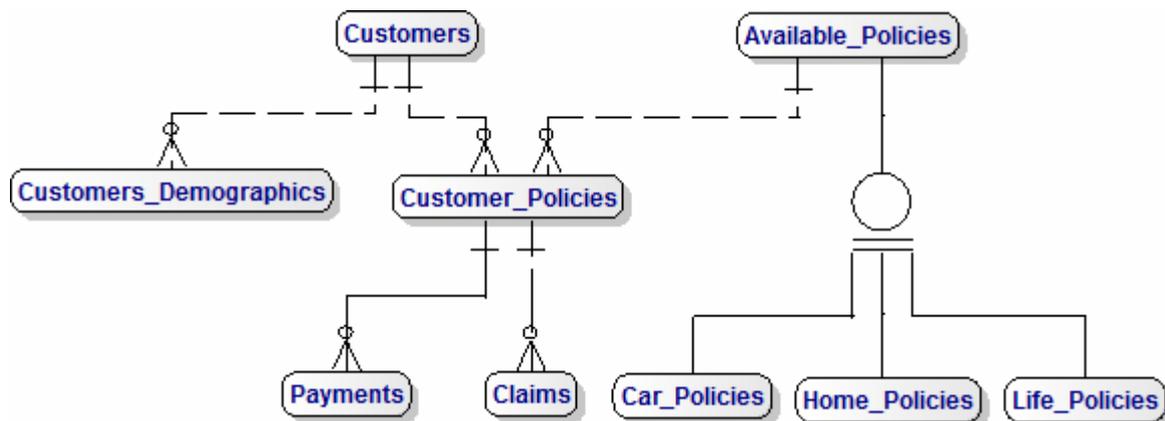
9.4 Insurance

9.4.1 Insurance Top-Level 3NF Model

This Section contains Links to our Web Site and shows the Insurance Enterprise Data Model and Common Data Model.

Here's a 3NF Insurance Data Model :-

- http://www.databaseanswers.org/data_models/enterprise_data_model_for_insurance/index.htm



9.4.2 Insurance Dimensional Model

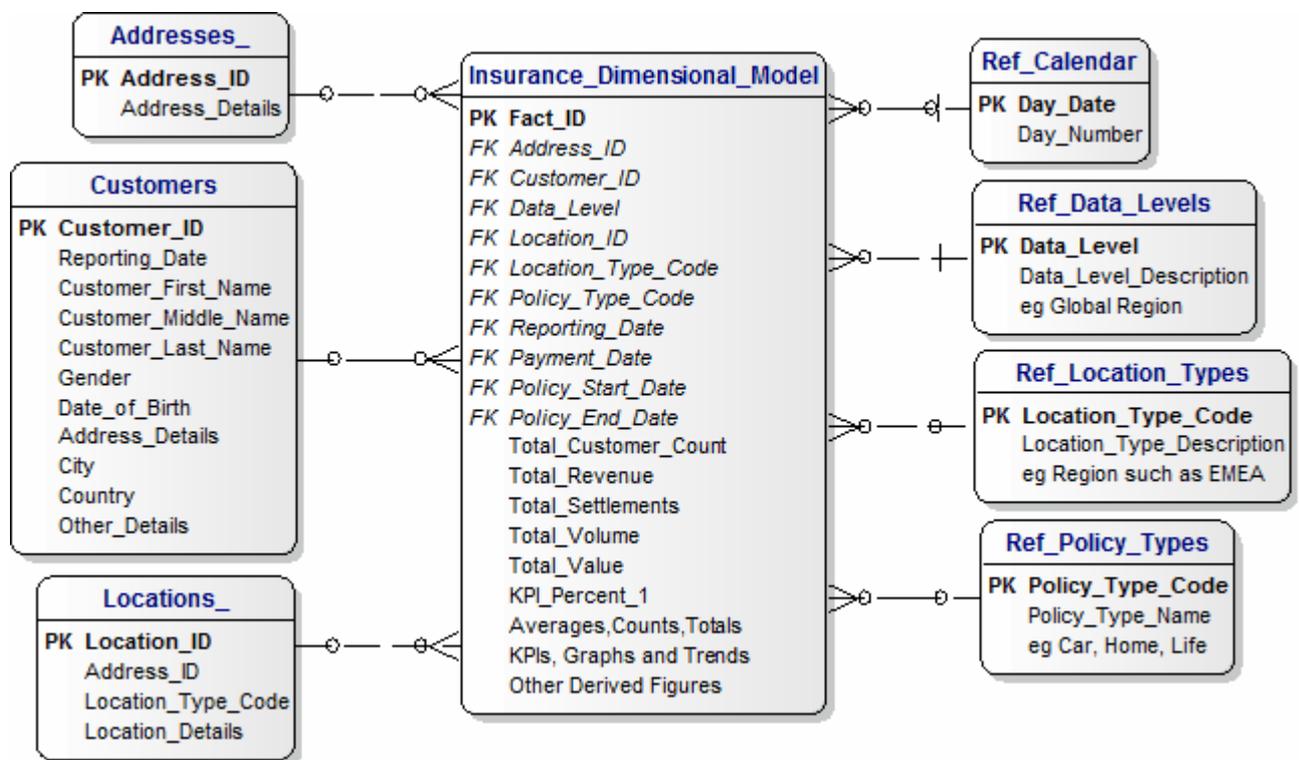
This Insurance Dimensional Model is on this page :-

- http://www.databaseanswers.org/data_models/insurance_data_warehouses/index.htm

We can see that the Dimensional Model Table will hold the Facts, where every record has a unique Fact ID as its Primary Key.

The Relationship with the Ref_Data_Levels Entity is mandatory at the Data Levels end because every record must have a Data Level.

For the Ref_Calendar Entity, we show one date as being mandatory because every analysis and report has a date field (even if it is implied).



9.4.3 Links to Insurance Subject Area Models

Customers, Policies and Brokers and Brokers –

- http://www.databaseanswers.org/data_models/insurance_brokers/index.htm

Customers, Policies and Claims (“Common Data Model”) –

- http://www.databaseanswers.org/data_models/insurance_data_warehouses/common_data_model.htm

Products (Car, Home and Personal) -

- http://www.databaseanswers.org/data_models/insurance_car_home_life/index.htm

Products (Marine and Motor Vehicles) -

- http://www.databaseanswers.org/data_models/insurance_general/index.htm

9.5 Oil Companies

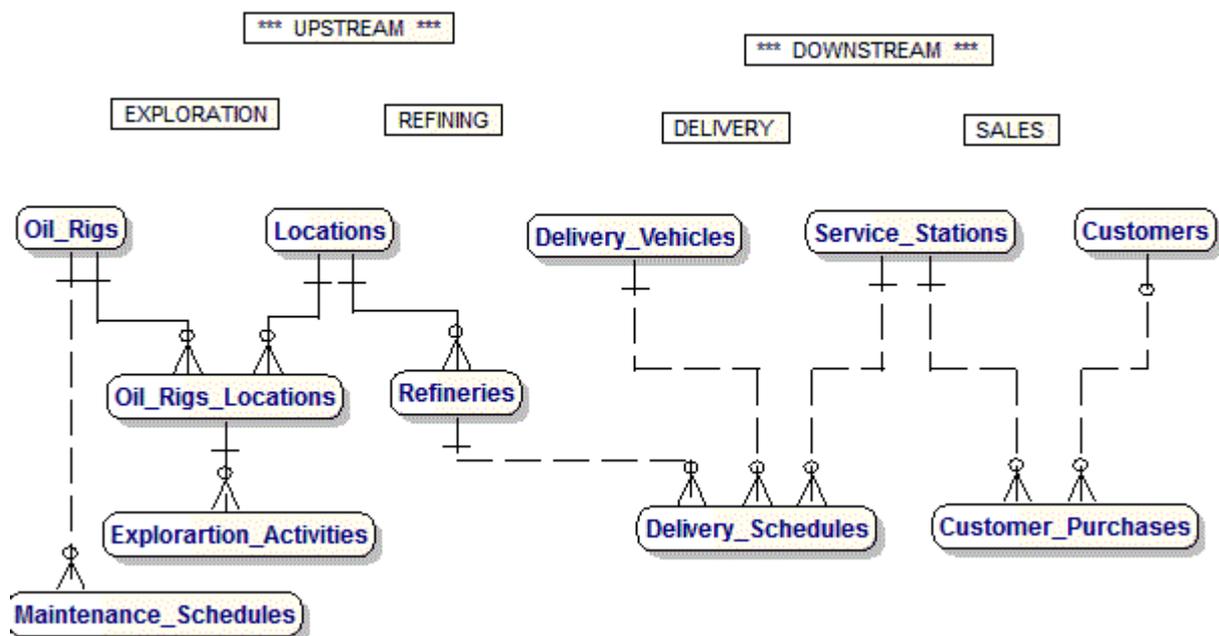
This final Enterprise Data Model for Oil Companies and is shown on this page –

- http://www.databaseanswers.org/data_models/oil_companies/index.htm

This is a draft description that we produced as we normally do.

We then discuss it with business users and SMEs and changes it so that everybody is happy with it. It follows a Product Lifecycle approach, where we follow a typical Product from source to destination.

In the case of oil, the source involves Exploration and Refining, as part of the Upstream activities. Refined Oil, in the form of petrol, is then delivered ('Downstream') in Oil Tankers, 'vehicles' to the final destination, in the shape of Service Stations, where it is purchased by Customers.



Elephants, Olympic Judo and Data Warehouses

9.6TEMP Links to Generic (Horizontal) Subject Area Models

Accounts - http://www.databaseanswers.org/data_models/investment_banking/accounts.htm

Brokers - http://www.databaseanswers.org/data_models/investment_banking/brokers.htm

Customers - http://www.databaseanswers.org/data_models/investment_banking/customers.htm

Customer Purchases -

http://www.databaseanswers.org/data_models/investment_banking/customers.htm

Deals - http://www.databaseanswers.org/data_models/investment_banking/deals_general.htm

Deliveries - http://www.databaseanswers.org/data_models/investment_banking/deals_general.htm

Staff - http://www.databaseanswers.org/data_models/investment_banking/staf6.htm

Transactions - http://www.databaseanswers.org/data_models/tracking_share_transactions/index.htm

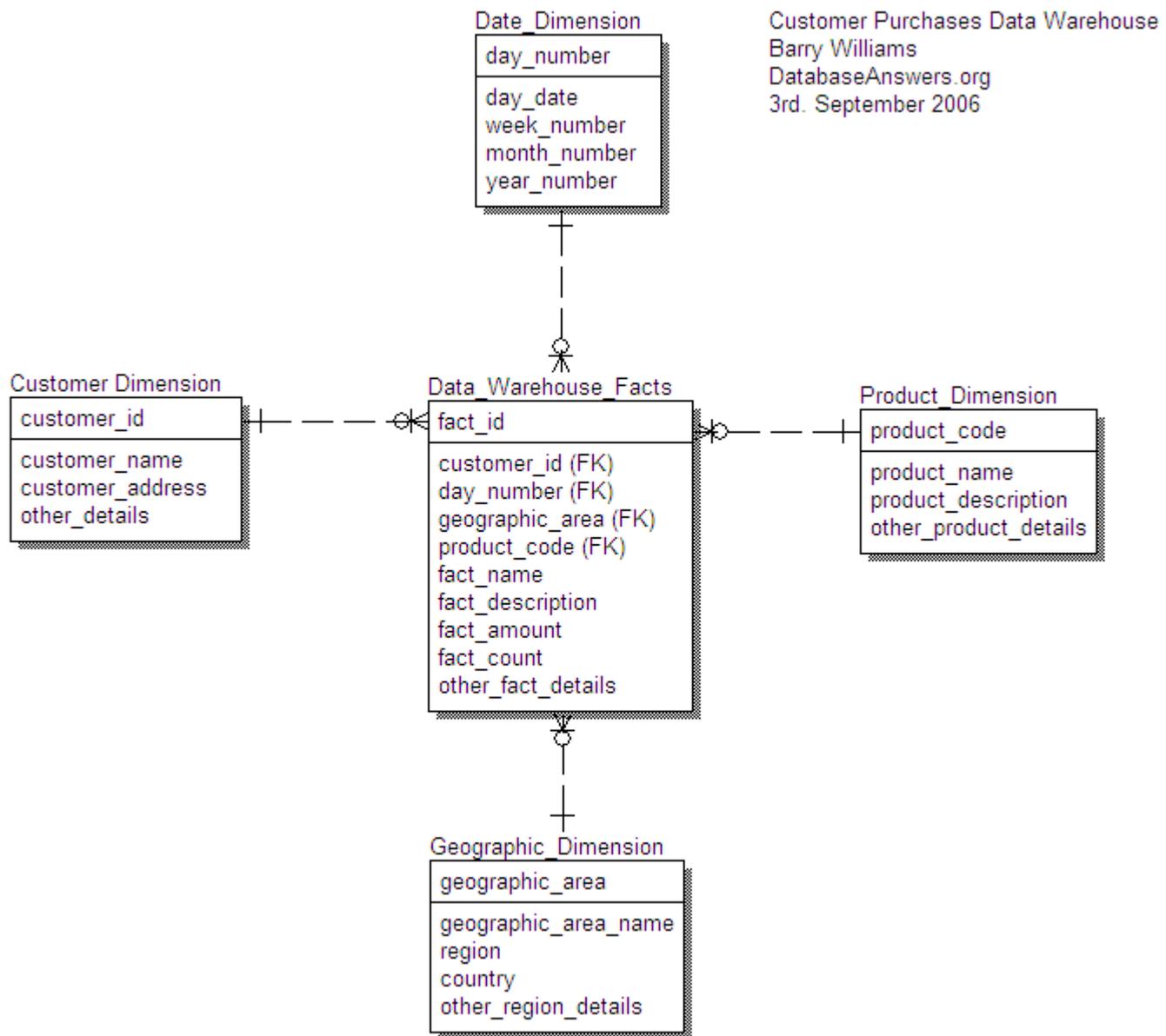
10. Generic Data Models

These Models can be used as Design Patterns.

10.1 Basic Dimension Model

This Data Model shows the basic commercial Star Schema for Customers, Dates, Geography and Products :-

- http://www.databaseanswers.org/data_models/customers_and_purchases_data_warehouse/index.htm



Elephants, Olympic Judo and Data Warehouses

10.2 Canonical Data Model

Although we are practitioners in Data Modelling Best Practice here at Database Answers, we don't often get the chance to use the word 'Canonical'.

When we do, we think of a Canonical Data Model as a 'stripped-down' Models which contains the basics for a specific application area.

This makes it possible to use it for a number of different purposes :-

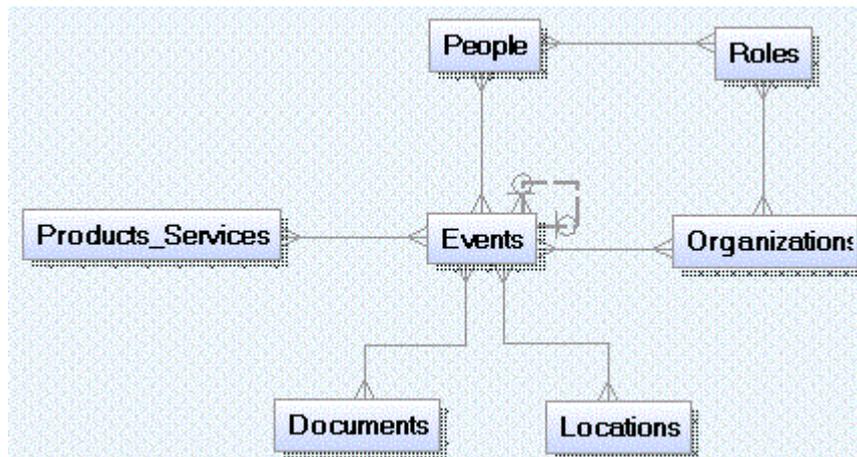
1. To be a standard for messaging in an ESB/SOA environment.
2. To provide a common point of agreement for all stakeholders.
3. To define a Kick-Start for Master Data Management.
4. To establish a starting-point for development of more detailed specific Data Models
5. To serve as a translation between different data sources.

This diagram shows our Generic Canonical Data Model which is Event-oriented and this has proved to be the best generic design in our experience.

Right now, we are using the Canonical Data Model as a general framework to analyse source data so that it can be used to design a 3NF Data Model and Dimensional Models
It occurs again further on where we refer to it in the context.

The Model appears on this page :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

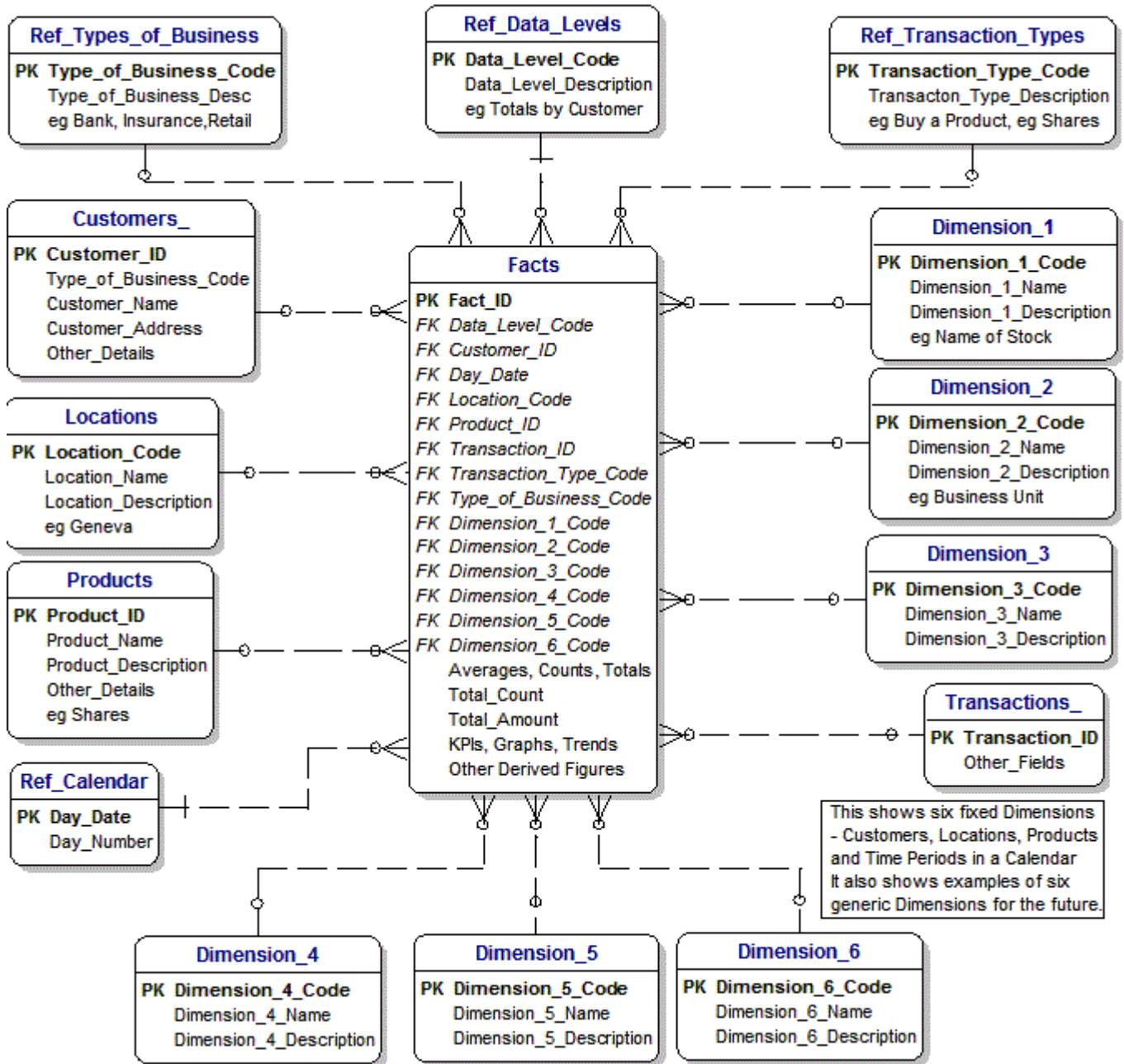


10.3 Generic Dimensional Model

The Model appears on this page :-

- http://www.databasanswers.org/data_models/banking_data_warehouses/generic_dimensional_model.htm

This shows the four most common Dimensions of Customer, Location, Product and Calendar (Time period).



10.4 Conformed Dimensions

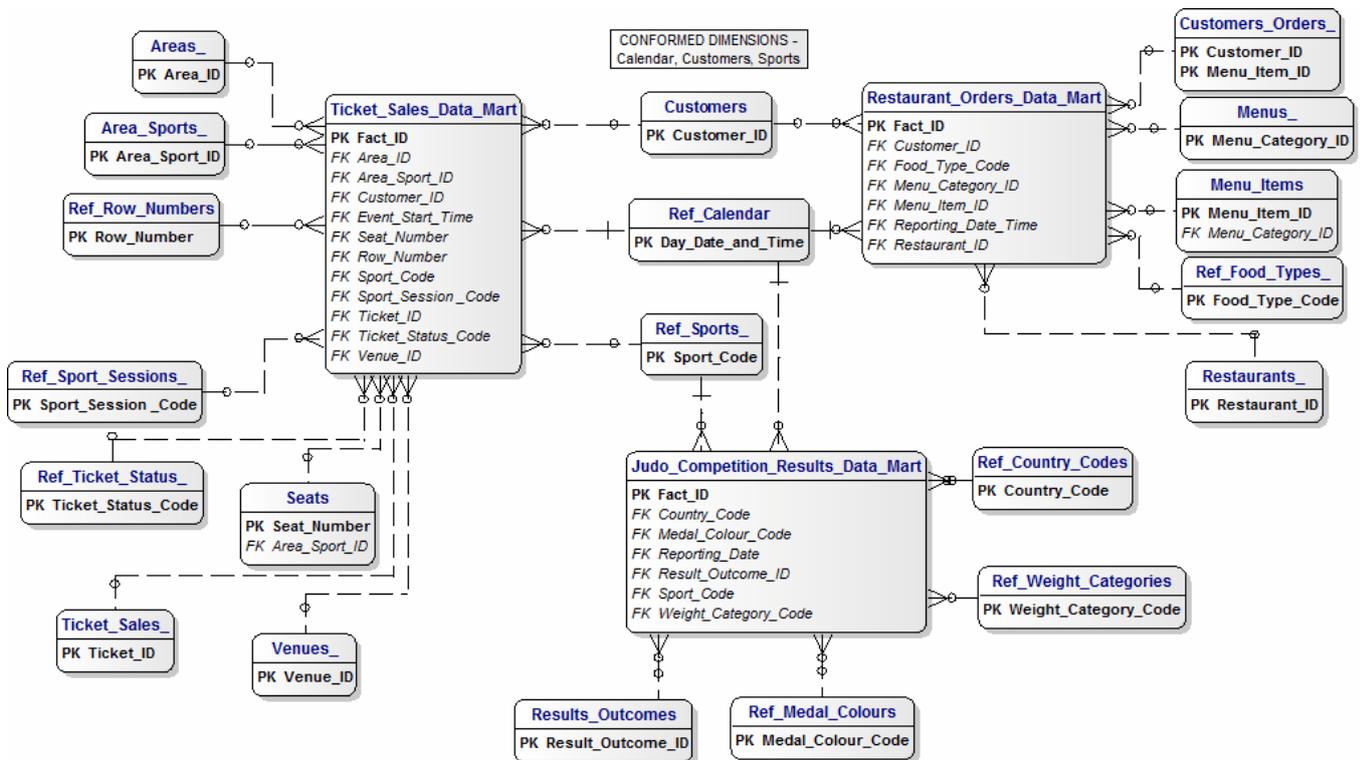
Conformed Dimensions are Dimensions can be shared between Data Marts because they share the same domains and values.

This means that SQL JOINS on these Dimensions will always find a match and return the appropriate number of values.

10.4.1 A Day at the Olympics

This shows three Data Marts with some shared Conformed Dimensions.

- http://www.databaseanswers.org/data_models/a_day_at_the_olympics/index.htm

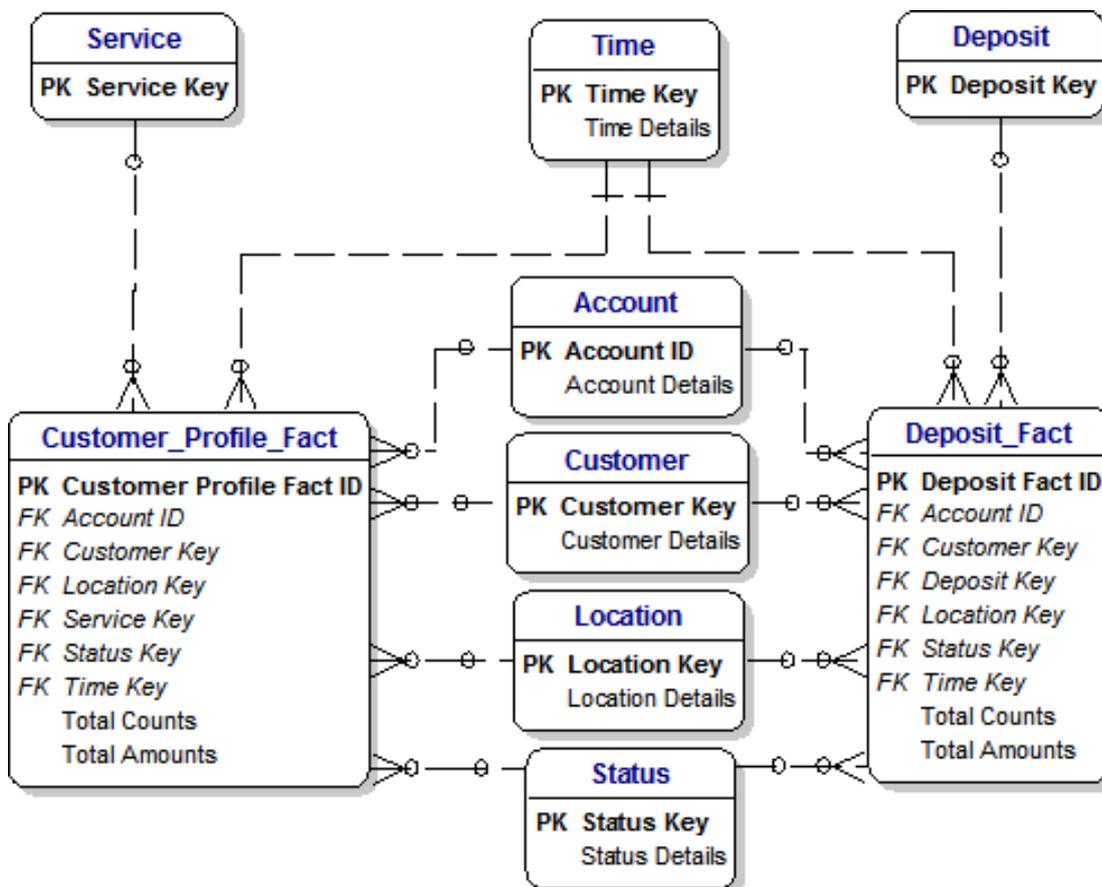


10.4.2 Customers and Deposits

This Data Model shows a Federated Data Mart for Retail Banks and is from the Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/banking_data_warehouses/federated_data_marts.htm

It shows two Data Marts related by five Conformed Dimensions of Account, Customer, Location and Status.



10.4.3 Retail

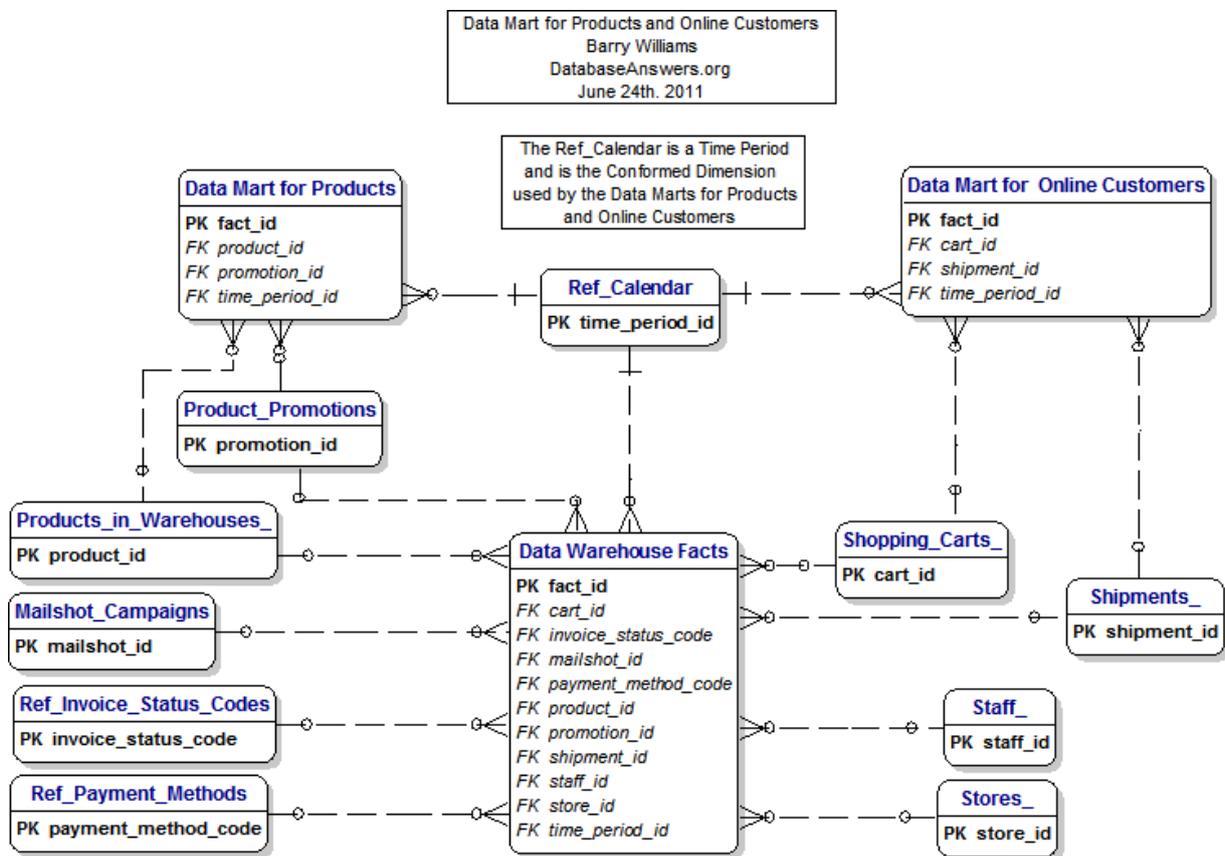
This page shows Retail Conformed Data Marts :-

- http://www.databaseanswers.org/data_models/retail_customers/retail_customers_data_mart.htm

The Calendar is a conformant Dimension for the three Data Marts.

The Data Warehouse and the Product Data Mart share the conformed dimensions of Product Promotions and Products in Warehouse.

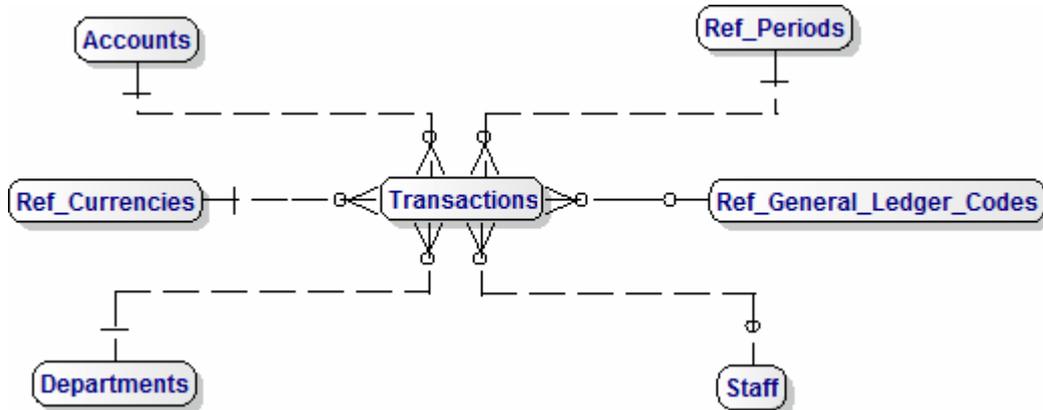
The Data Warehouse and the Online Customers Data Mart share the conformed dimensions of Shipments and Shopping Carts.



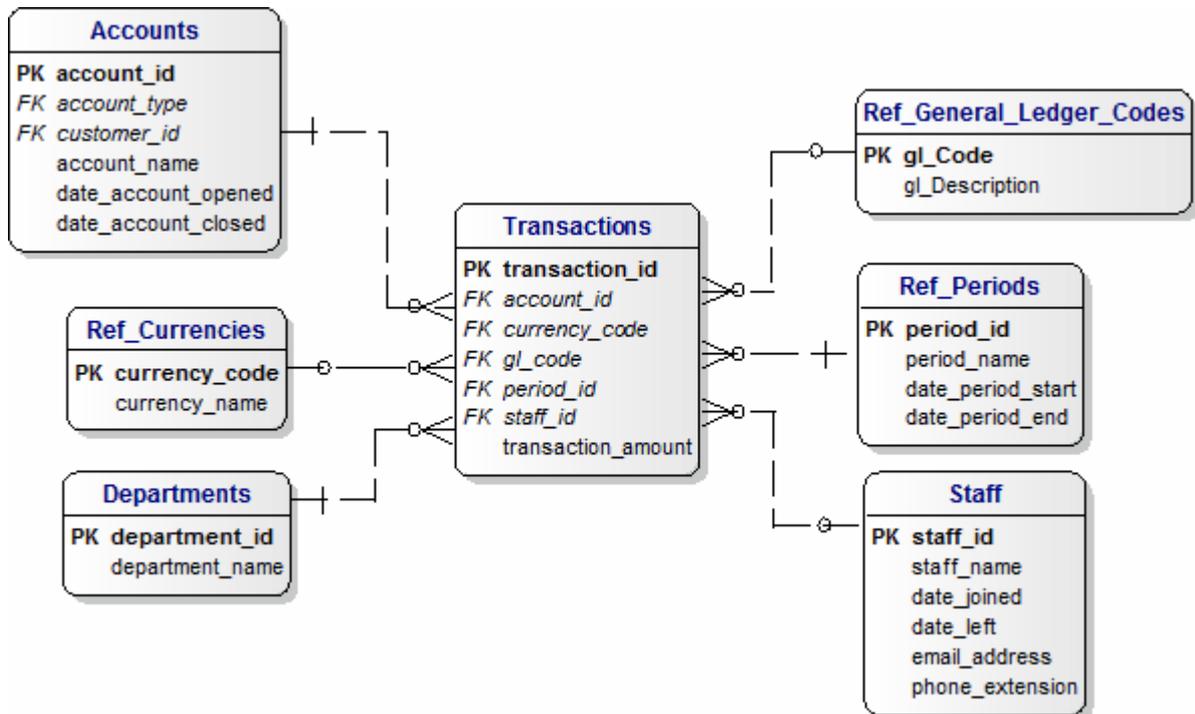
10.5 Star Schema

This Data Model shows a typical Star Schema for Customers, Accounts and Transactions. It appears on this page :-

- http://www.databaseanswers.org/data_models/star_physical.htm



This diagram shows Entities with attributes

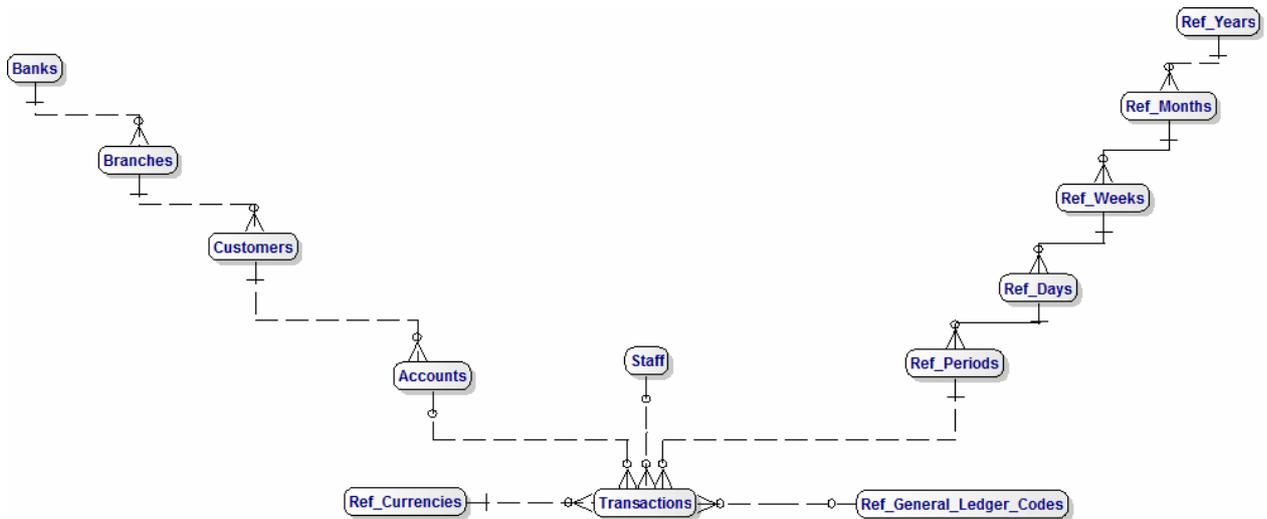


10.6 Snowflake Schema

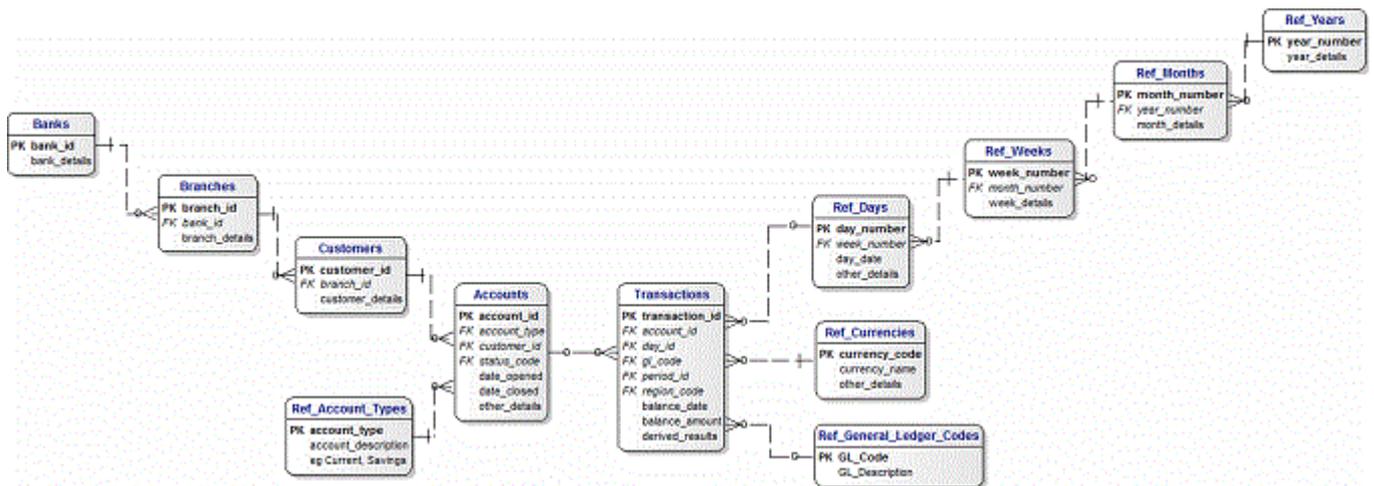
This Data Model shows a typical Snowflake Schema for Customers, Accounts and Transactions. It appears on this page :-

- http://www.databaseanswers.org/data_models/snowflake.htm

The Snowflake structure is very clear in this Data Model.



This version of the Model shows Entities with attributes -



11. Integrating Multiple Data Warehouses

A very useful reference is our short Guide to the Quality Assurance of a Data Model, which includes a Section on Data Warehouses.

It is available for download from our Web Site (currently at number 30) :-

- <http://www.databaseanswers.org/downloads.htm>

It is also available on Amazon :-

- <http://www.amazon.com/dp/B00KALS1PK>

11.1 Why ?

We have sometimes found ourselves in situations where there are multiple Data Warehouses provided from different sources that need to be integrated into one Generic design.

11.2 How ?

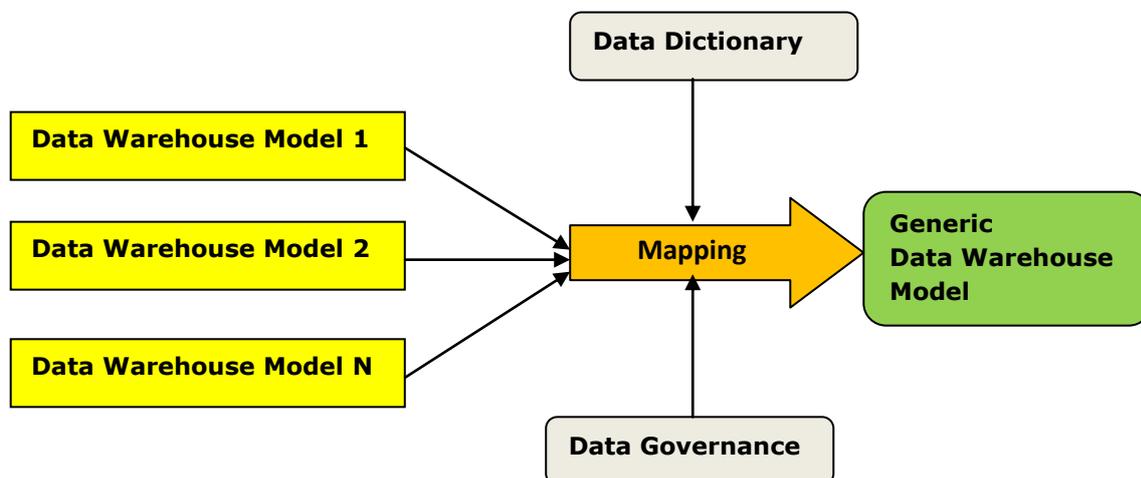
We have developed a Top-to-Bottom Approach that is based on these Steps :-

1. Establish a 'Business Mission Statement' and agree it with the business and Subject Matter Experts ('SMEs').
2. Define some Key Performance Indicators ("KPIs')
3. Use the KPIs to trace the required data down to the Generic Data Warehouse.
4. Establish Generic designs for the Conceptual Models, Semantic Layer, Data Marts and Data Warehouse
5. Create a Data Dictionary
6. Determine mapping from the Source Data Models to the Target Models.

This is summarised in the diagram shown in the next Section.

The Approach is to map the Source Data Warehouse Model to a single Generic Model.

A Data Dictionary is used to keep track of the names of all Entities and Attributes and to define the mapping between them.



11.3 Example for Telecomms

This material is available on our Web Site :-

- http://www.databaseanswers.org/data_models/telecomms/index.htm

This example shows a top-down view of the Layers in a BI Data Architecture for Telecomms :-

- Conceptual Models
- Semantic Layer
- Data Marts
- Data Warehouse

On the Database Answers Web Site, we also show the Terada Communications Logical Data Model :-

- http://www.databaseanswers.org/data_models/teradata_communications_data_model/index.htm

11.3.1 Conceptual Models

The specifications for these three was taken from the 'GameChanger' page on the Web Site for the TeliaSonera Telecoms company in Sweden :-

- <http://gamechanger.teliasonera.com/en/>

We have taken these three as examples, but without any implication that they are realistic.

That would be determined in discussion with Subject Matter Experts.

1. Location and Payment Method

- Time Period
- Location
- Payment Method
- Volume

2. Life Style of Elderly

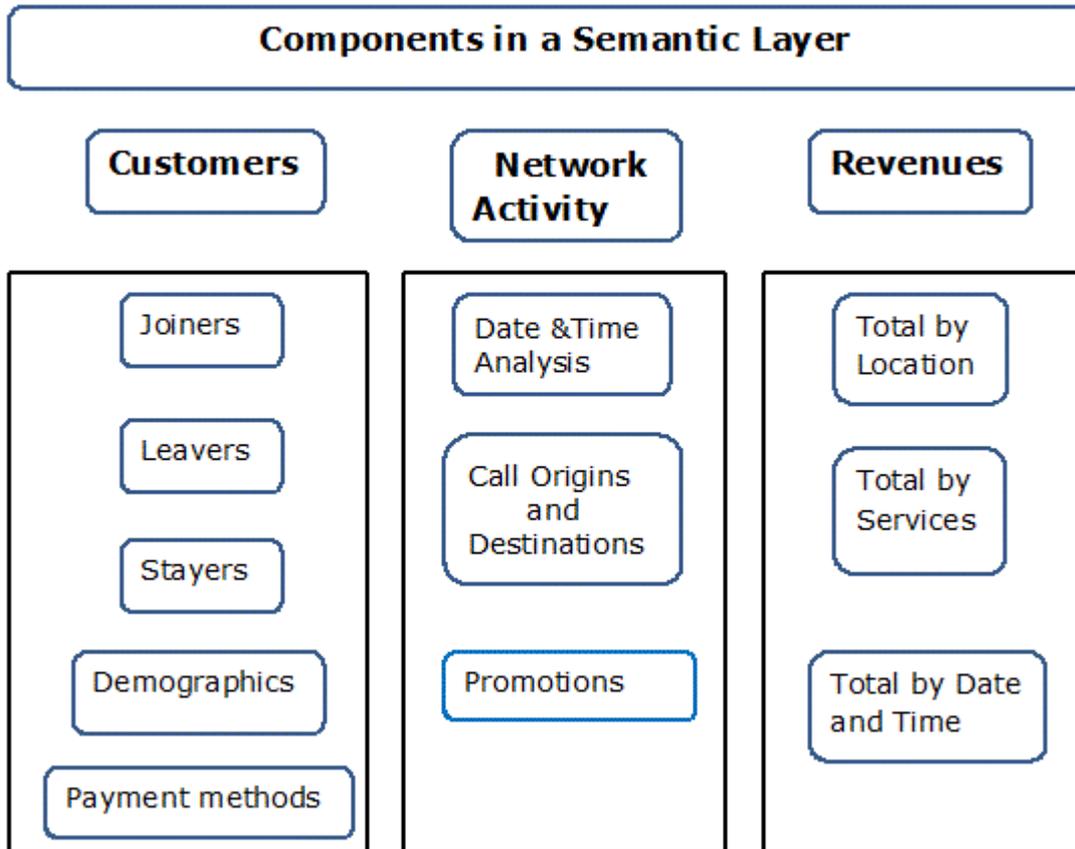
- Time Period
- Customer Age Group
- Customer Demographics
- Volume

3. Young People in Business

- Time Period
- Customer Age Group
- Customer Address

11.3.2 Semantic Layer

The purpose of a Semantic Layer is to provide business-oriented view of data. This shows an example with three Components based on the underlying Data Marts.



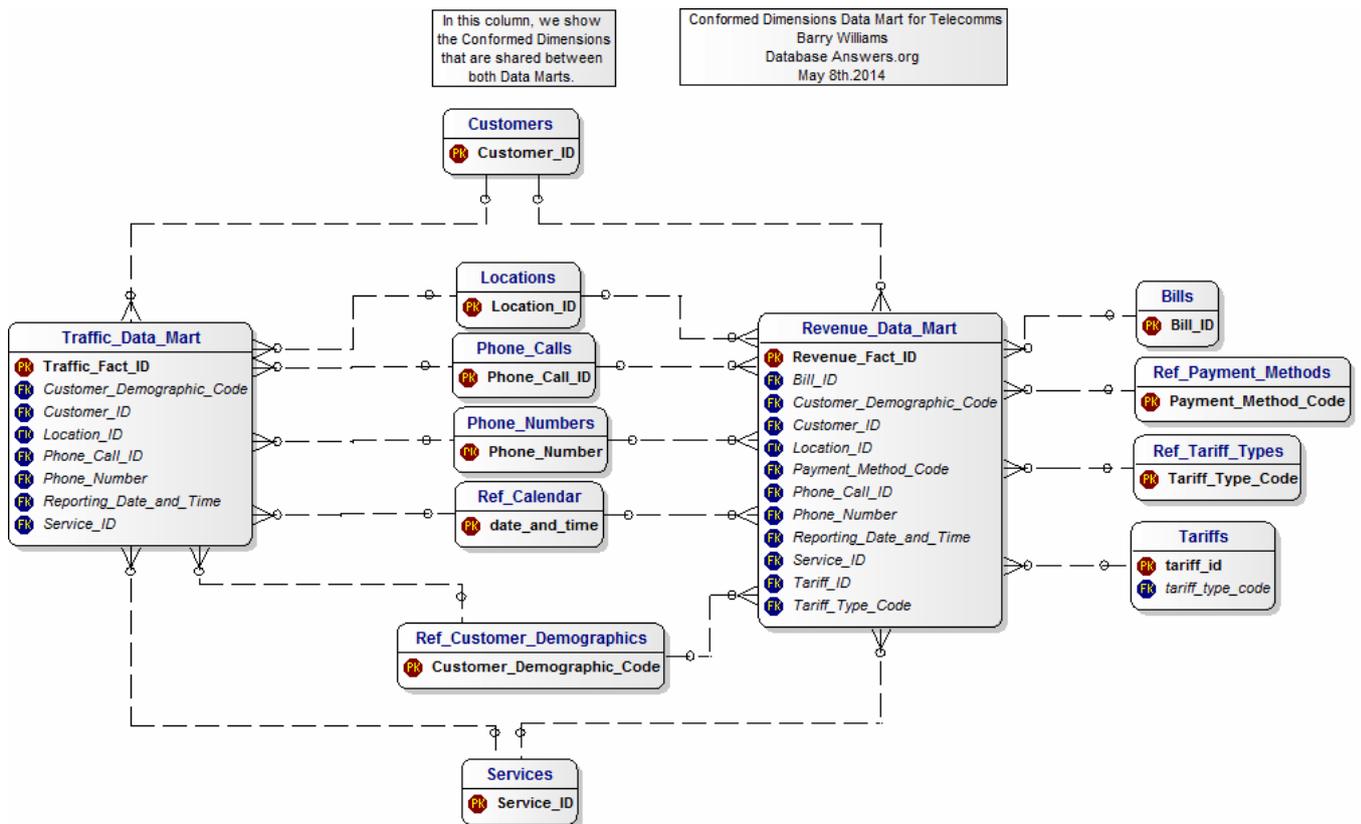
Elephants, Olympic Judo and Data Warehouses

11.3.3 Data Mart

This shows an example for Telecomms, with a e Approach is to map the Source Data Warehouse

This shows an example for Telecomms, with a e Approach is to map the Source Data Warehouse Model to a single Generic Mode.

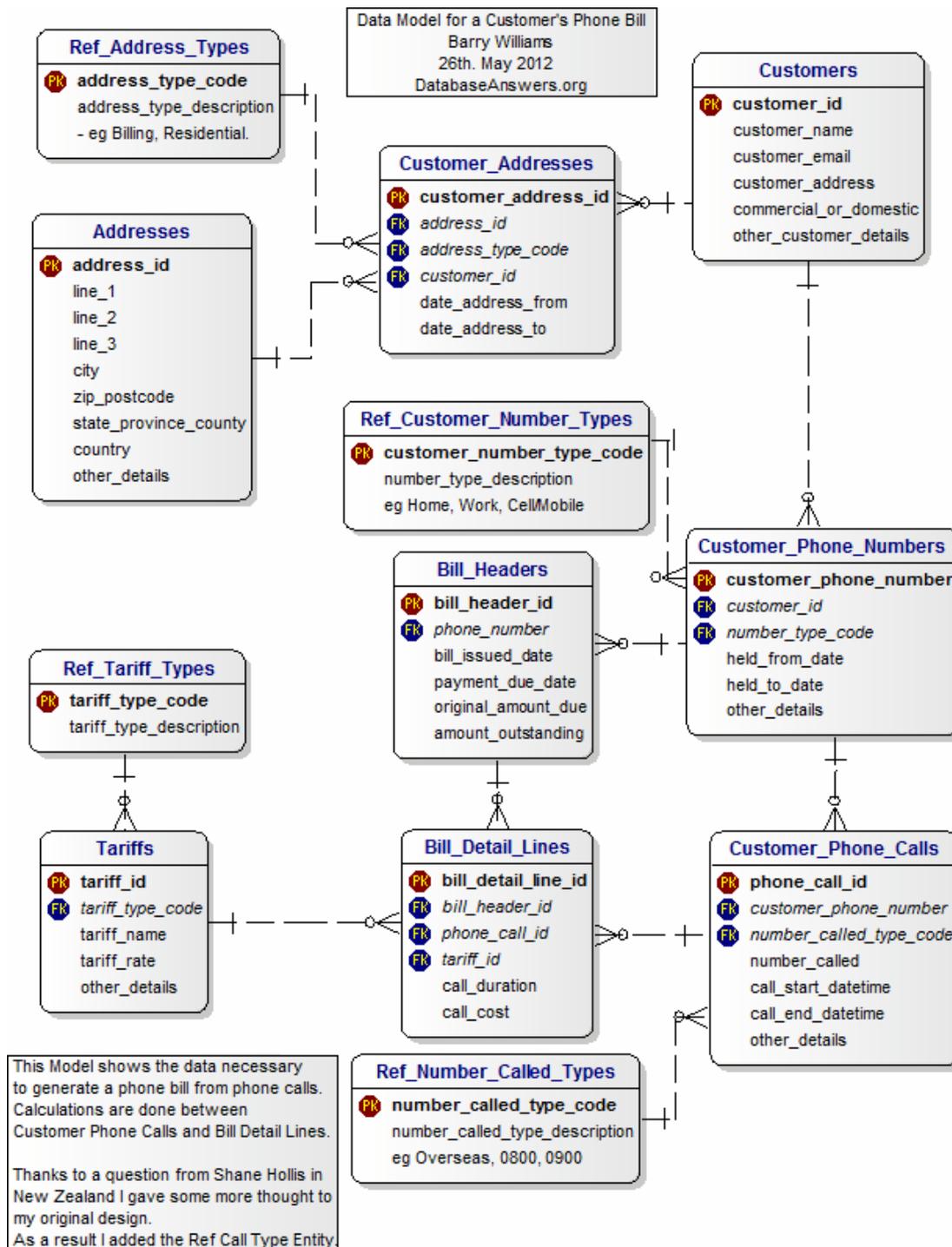
An example for Telec omms shows a Traffic Data Mart and a Revenue Data Marts, with Conformed Dimensions, which ensure common values for shared Dimensions.



11.3.4 Data Warehouse

For our Data Warehouse, we need a Third-Normal-Form Data Model, so we have taken this one as a suitable candidate :-

- http://www.databaseanswers.org/data_models/customers_and_phone_bills/index.htm



12. Conclusion

In this book we have presented an approach to Data Warehousing that we have used with great success in past assignments with a wide variety of Clients.

Our intention is to update it regularly as a Kindle e-Book and to keep it timely and to establish it as a reference book of Best Practice.

We would be very happy to hear from you if you have any comments or suggestions for improvement.

Please email us at barryw@databaseanswers.org