Chapter-5

Conclusion and Suggestions

5.1 CONCLUSION OF DATA WAREHOUSE

The thesis focused on statistical analysis of the impact of Web Data Design in Data Warehousing for a Software industry using Data Design Pattern the Efforts were been made to consider all the principles of Data warehouse design during the course of this thesis. We have shown how data can be integrated from different sources to a single repository called data warehouse which is used for delivering Integration of effective design the end users and executives. We have succeeded impact of metadata in designing how data can be made available to business for day to day activities of their businesses.

We have developed data and statistical analysis to template that user can interact with to get an immediate answer to the business question.

Our main research objectives, addressed in the scope of this thesis, have been determined by the development of comprehensive data warehousing solution for Web Data Design using integration, driven by the cost reduction and the implementation efficiency requirements. The accomplishment of these objectives demanded a thorough understanding of the various data warehousing impact of metadata related aspects, namely the positioning and
role of the data warehousing technology in the overall Web Data Design Business Intelligence framework; the definition of specific data models, which determine the storage structures of the analytical environment; the solution development methodology, guiding the effective design of web data and implementation processes; the architecture that defines the foundation of data warehousing Web Data Design Business solution development; the framework providing a set of guidelines for building impact of metadata components and describing their interaction on the analytical systems; and the actual implementation processes.

As introduction into the background of the analytical systems, we began with a presentation of their main characteristics, analyzed the historical evolution of decision support applications and discussed the similarities and differences between early Web Data Design stage systems and nowadays comprehensive technologies. data warehouse management, design and evolution based on a high level conceptual perspective, which can be linked to the actual structural and physical aspects of the data warehouse architecture. This metamodel is capable of modeling complex activities, their interrelationships, the relationship of activities with data sources and execution details. Finally, the Web Data Design metamodel complements proposed architecture and quality models in a coherent fashion, resulting in a full framework for data warehouse metamodeling. We have implemented this metamodel using the language and the metadata repository system Concept
Base. This document presented a set of results towards the effective Web Data Design modeling and management of data warehouse metadata with special treatment to data warehouse quality. The first major result that we presented was a general framework for the treatment of Web Data Design data warehouse metadata in a metadata repository. The framework requires the classification of metadata in at least two instantiation layers and three perspectives. The metamodel layer constitutes the schema of the metadata repository and the metadata layer the actual meta-information for a particular data warehouse.

We linked this framework to a well-defined approach for the architecture of the data warehouse.

Developing a DW is a complex, expensive, time consuming, and prone to fail task. Different DW models and methods have been presented during the last few years. However, none of them addresses the whole development process in an integrated manner. In this thesis, we have presented our Web Data Design DW development method, based on the UML and the UP, which addresses the analysis and design of both the DW backstage and front-end. For this task, we have extended the UML in order to accurately represent the different parts and properties of a DW. Our proposal provides a seamless method for developing DW and it is a great help when designing, implementing and deploying a DW. Following our approach, we design a DW as follows:
1. We use UML to model the data sources of the DW at the conceptual level.

2. Then, we use our Multidimensional Profile for the design of the DW at the conceptual level.

3. We apply our Data Mapping Profile for creating the mapping between the data sources and the DW at the conceptual level.

4. We use different UML extensions to model the data sources and the DW at the logical level: UML Profile for Database Design, a UML Profile for data modeling, a UML extension for the modeling of XML documents by means of Document Type Definition (DTD) and XML Schemas, etc.

5. Then, we design the ETL processes that will be responsible for the gathering, transforming and uploading data from the data sources into the DW. We use our ETL Profile for this task.

6. Finally, we use our Database Deployment Profile for taking physical design decisions.

The main advantages of our proposal are:

- The Definition of a set of UML Profile, which define an extension to the UML but keep the UML metamodel intact.
- The use of the same notation (UML) for designing the different DW models and the corresponding transformations in an integrated manner.
- The use of the UML importing mechanism, which guarantees the designer that each element is defined once, because the same element can be used in different models.

5.2 LIMITATIONS OF DATA WAREHOUSE

We encountered series of problem related to impact of metadata in Data Warehousing and Business Intelligence we were able to put too many into a
small hole without compromising the quality of the project. Due to the time and resources required to carry out further statistical analysis, we have limited our impact of metadata design and analysis of data to customer and product sales area. The topic is very wide in the real world, in the context of academic, we have been able to show our understanding in this area and future enterprise project is possible from where we stopped.

We have presented the data warehousing as a comprehensive technology used for handling the analytical environment of an enterprise, and the data warehouse as a component of the repository element and the foundation upon which Business Intelligence is built. Thus, we argued the existence of a clear differentiation between data warehousing and the data warehouse concepts, and also examined the various perspectives treated in the literature.

While data warehousing comprises a series of components and processes designed to enable the collection and integration of enterprise data from various sources, with the main goal of transforming it into strategic information, the data warehouse defines the storage component of the overall technology, the repository of integrated, subject-oriented, non-volatile and time-variant data. These collected and integrated large amounts of data are subject to various mathematical and statistical calculations and statistical analysis meant to produce valuable business insight. For this reason, the data warehouse relies on storage structures optimized for high performance
querying and analysis processes, premeditated with specific data modeling design techniques. We introduced the particularities of the various data warehousing structures (e.g. the data warehouse, the data marts), and examined the differences between them. These structures and the data of the warehouse repository are described from different perspectives by means of Impact of metadata. In order to understand their fundamental role, we presented a comprehensive study of metadata definitions, classifications and management characteristics. We focused mostly on the business and technical metadata types, highlighting the descriptive character of business metadata, essential in understanding the semantics of the business processes, and the importance of technical metadata in enabling automation in the data warehousing environment.

Considering the complexity of the data warehousing solutions, we introduced several development methodologies aimed to offer a structured and planned approach to the numerous processes involved, by providing a repeatable, consistent and reliable set of steps or guidelines. From the various methodologies analyzed, both general and data warehousing specific, we described comprehensively two reference approaches, namely the Inmon (i.e. top-down) and the Kimball (i.e. bottom-up) models. We also introduced two frameworks defined for selecting the suitable methodology for successful data warehousing implementation, and outlined our reasons for
adhering to Inmon’s data-driven spiral approach in the development of our solution proposal.

The optimized data structures of the analytical environment are calculated with specific data modeling design techniques, meant to determine and analyze the requirements expressed by business users and to produce data models capable of supporting the enterprises’ business processes. For instance, considering our comprehensive data warehousing solution approach, defined by a layer of consolidated and integrated granular data and a layer of dimensional structures, built for querying and statistical analysis performance, we described two of the most commonly used data modeling design techniques, namely RRB and PSB dimensional modeling design. The RRB dimensional modeling design technique presents characteristics adequate for modeling design normalized models with minimum redundancy, dependencies, and inconsistencies, and thus high quality data appropriate for the data warehouse granular layer, while dimensional modeling, specific to the analytical environment, produces intuitive presentations of data in the form of statistical analysis optimized models. Another important aspect covered in our thesis was the selection of the appropriate data warehouse architecture. Due to the complexity and the broad scope of the data warehousing solution, its implementation architecture is significantly different and more complex than the classical database architecture. We introduced several types of architectures
acknowledged in the conceptual literature (e.g. the independent data marts, the data mart bus architecture, the enterprise data warehouse, the centralized and the federated architectures) and presented two reference architectural implementations: the top-down (achieved by the Inmon model), also known as the Corporate Information Factory framework, and the bottom-up (achieved by the Kimball model), known as the Data Mart Bus Architecture. We also presented a framework defined for facilitating the selection of the appropriate architecture type, determined by organizational, environmental, project-related, technical and educational factors, etc. and selected as most suitable architecture for our solution development, the comprehensive enterprise data warehousing architecture recommended by the Inmon model.

Regarding the development framework proposed, we began from the idea that successful data warehousing design and implementation processes are supported by compatible architectures and methodological frameworks. Considering the numerous structures and processes which comprise the data warehousing solution, we defined an automated implementation framework and a corresponding prototype, based on the assumption that repetitive and time-consuming development activities may be carried out more efficiently and in a shorter period of time.

We justified our proposal by presenting the various benefits of automation in software development in general, and particularly in the data warehousing environment.
This document presented a set of results towards the effective web data modeling and management of data warehouse metadata with special treatment to data warehouse quality. The first major result that we presented was a general framework for the treatment of data warehouse metadata in a metadata repository. The framework requires the classification of metadata in at least two instantiation layers and three perspectives. The met model layer constitutes the schema of the metadata repository and the metadata layer the actual meta-information for a particular data warehouse.

5.2.1 END-USER SURVEY

This end-user survey in Appendix A was conducted as part of Requirement Analysis phase. The online survey questionnaire link can be found in http://www.snapsurveys.com/swh/surveylogin.asp?k=13111533830. The main goal of the questionnaire is to help us to understand the user expectation for the DW solution. It is also going to help us to understand the historical data RRB and PSB that the organization gather, we then use this information to relate requirements with actual design. It consisted of 17 questions: 07 questions were text based and the rest 10 had answer options. The options include Strongly Agree, Tend to agree, neither agree or disagree, Tend to disagree, and Strongly Agree. These answers were associated its gives the user interface and Appendix- B gives the details questions. The survey questionnaire served as supplements to the industry standards in which we followed and face–to face contact we had with
managers and business user of the DW Systems; the input from the managers and business users and industry standards are more appropriate in the actual research work as we were proposing BI systems to Crystal Entertainments.

The response acquired from the online survey are hereby analyzed in the spreadsheet attached to this work; the analysis and conclusion of the survey helped to formulate the business intelligence case/need of Crystal Entertainment Inc.

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Strongly Agree (%)</th>
<th>Tends to Agree (%)</th>
<th>Neither Agree nor Disagree (%)</th>
<th>Tends to Disagree (%)</th>
<th>Strongly Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q01</td>
<td>77.80%</td>
<td>22.20%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q02</td>
<td>77.80%</td>
<td>22.20%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q03</td>
<td>44.40%</td>
<td>50.00%</td>
<td>5.60%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q04</td>
<td>58.80%</td>
<td>17.60%</td>
<td>23.50%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q05</td>
<td>38.90%</td>
<td>44.40%</td>
<td>16.70%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q06</td>
<td>50.00%</td>
<td>38.90%</td>
<td>11.10%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q07</td>
<td>50.00%</td>
<td>38.90%</td>
<td>11.10%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q08</td>
<td>61.10%</td>
<td>33.30%</td>
<td>5.60%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q09</td>
<td>61.10%</td>
<td>38.90%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Q10</td>
<td>82.40%</td>
<td>11.80%</td>
<td>5.90%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 5.1: Showing Analysis of End-User Survey Report

From the Table 5.1; the respondent of the survey has proved that level of Data correctness in their DW is an important factor in the ability to make informed decision. 83% of the respondents support the view that since a
DW is data driven, it is important that Data imported into the DW should be validated and is as close as possible to the real life situation.

Trusted Data makes the world of Business Intelligence rosy. 78% of the respondents equally suggested that the DW should be able to integrate data from various sources is as important as the close synchronization among functional area units within the business, so that it can help managers make decision on the overall working together of the units and how they affect each other. The response to question 8 also reiterates how significant the real world events is closely related in the data, about 62% of respondents points out. Furthermore, when data in the DW becomes larger, Able to handle increase in the number of simultaneous queries without impacting system performances. As they don’t want a system that slows down their ability to make swift decision, as one of the demands placed on the system is to be able to provide data speedily with minimal impact of other users of the system.

5.3 SUGGESTIONS OF DATA WAREHOUSE

We have been able to demonstrate that the data warehouse web data design can be implemented by any organization and that the project should have a clear aim of supporting both business users and executives. We have based the Impact on Metadata on three major units within the organization to develop a data mart. It is possible to design an enterprise wide data warehouse from the
data mart. The principle we had followed can be adopted for any retail company. The project is scalable industry wise. In the future, possible integration of the business intelligence window into the collaboration tools is possible.

We have come to see that the key to data warehousing is data plan. A good data intend creates a workable data warehouse. The industry users know what data they need and how they want to use it. We focus on the users to determine what data is needed, locate sources for the data, and organize the data in a dimensional model that represents the business needs.

Data analysis and reporting are not enough to leverage the benefits of business intelligence in such a dynamic industry. The required, needed and right information should also find a way to get to the right people in ample time. But importantly, the information must be acted upon to activate the benefits. It is expedient that business intelligence should be part of the business process, within the management sector. BI must not only help to understand the past, but also work to find new opportunities and emerging trends in the future!

Obviously, this thesis work can be continued following several different research lines, because one has to draw the line somewhere and decide when to stop. The DW design can be related to other areas; actually, some research has been done from this thesis about quality metrics for DW conceptual models and DW security.
In the following, we present the future work divided into short term, medium term, and long term.

**Short Term**

- We are working on the Definition of a set of complexity metrics for the data mapping diagrams. These metrics will allow the designer to compare different design choices and select the best based on objective measures.
- We are also working on a template specification mechanism for frequently used transformations in data mapping diagrams.
- We are also studying different languages for the formal Definition of the transformations in data mapping diagrams, specifically OCL, Datalog and QVT.

**Medium Term**

- We are going to adapt our approach to UML2.0 as soon as it is accepted as a standard.
- UML2.0 will change the extension mechanism. Therefore, once UML2.0 has been accepted, we will finish the formal Definition of the different Profile we present in this work.
- Once the profiles have been defined, we are going to implement them as Rational Rose add-ins.
- We are currently working on some automatic transformations from the different models to the implementation. We are considering the implementation of the MD conceptual models on pure MD databases, object-relational databases, and OO databases. We are also thinking about the transformation of the ETL models into commercial target platforms.
Long Term

- We would like to provide a detail description of our Data Warehouse Design Engineering Process and develop in our approach.
- We plan to include new UML diagrams (sequence, collaboration, state chart, and activity diagrams) to model dynamic properties of DW.
- We would like to carry out an empirical evaluation of our proposal, in order to validate the correctness and usefulness of our approach.
- Finally, we would like to align our approach with the Multi Dimension Approach initiative.