***TEMPLATE FOR DATA WAREHOUSE PROJECT DOCUMENTATION***

***Table of contents***

***List of figures***

***List of tables***

***Acknowledgment***

***Executive summary***

**1.0 Introduction**

The concept phase is about describing what the big idea is. The business may have a concept and the IT team will be able to describe the major component and concepts of a data warehouse.

***1.1 Analysis***

The goal of the analysis phase is to identify the sources of the information required to populate the physical data models. The main goal should be to populate the repository data model as this is used as the source for all data in the data marts. This is achieved in a number of steps:

**1.1.1 Source Systems Analysis (SSA)**

The source system analysis is a high-level analysis that gathers information about available systems. Each system is a potential candidate as a source system and generates its own analysis document. Some systems may be documented and then rejected e.g. because it is a secondary source and only contains information created in another system that can be used as the source. The document covers hardware, software, network connectivity, availability and functional areas (e.g. CRM system containing customer data etc.).

**1.1.2 Source Entity Analysis (SEA)**

The source entity analysis is the detailed documentation of the sources selected because data profiling has validated these sources as being useful for the data warehouse. This includes detailed information about every table and column, including data types and data quality metrics. A source entity analysis will be produced for each system that is to be used as a source.

**2.0 Requirements**

The requirements gathering phase of any data warehouse is one of the most difficult. The objective of these templates is to give breadth and depth to the requirements. Breadth is the ability to ensure that all truly required information would be covered, whilst depth is the amount of detail that is specified in the requirements to ensure that the developers have sufficient, unambiguous, detail with which to develop. Requirements should have a programme-long life cycle. After the initial version of the requirements is developed a project can start the build, however the business moves on and therefore whilst the build phase is occurring it is important that new versions of the requirements are also being developed. A project within a programme of work should have a fixed version of the requirements; however each project may work with a different version of the requirements.

**2.1 Data Warehouse Business Requirements (WBR)**

The first template that Data Management & Warehousing use is called the Data Warehouse Business Requirements and it details the ‘soft’ requirements for business information according to a number of subject areas of interest to the business. A business requirement is something of the form: ‘Provide the average and total revenue for each product category by customer market segment for the last three years’ It is a requirement that is specified in business language and without regard for the practicalities of delivering it. These requirements should be used to get business users to underwrite the business benefit, i.e. if I could answer all of these questions then I would be able to increase margin by a given percentage for a given product.

**2.2 Data Warehouse Data Requirements (WDR)**

The second document details the ‘hard’ requirements for business information from the data perspective. This document goes a step deeper into understanding the requirements, but is still written from the business users’ perspective. This is the refinement of the business requirements in that the analysts can use the business requirement to drive out the data required to answer the questions. In the example above it is clear that both some part of the product hierarchy and of the customer hierarchy are required as well as a time dimension and information about revenue. It has also told the analyst that the minimum retention period is three years.

Consequently, the analyst would start to build data requirements that may fulfill many business requirements and to add additional attributes to help make sense of the data. The data requirements lifecycle is similar to that of the business requirements i.e. fixed for a project and variable over the lifespan of a programme.

**2.3 Data Warehouse Query Requirements (WQR)**

The third document lists a number of potential queries to which the solution should be able to provide answers. This is not an exhaustive list, but rather represents the types of queries that are being asked by the business. It is used to test the relationship between the data requirements, the data model and the business requirements. A set of queries should be able to provide the data required to answer a business requirement. At the same time the data must be available as described in the data requirements and joined in such a way as to be usable in the data model.

**2.4 Data Warehouse Technical Requirements (WTR)**

The fourth document details the functional and non-functional requirements that are expected of the solution. Again, these requirements are stated from the business perspective rather than the technical perspective. The document should include topics such:

• The functionality required of the query tools.

• The general retention requirements for data.

• The performance characteristics.

• The systems availability expectations.

**3.0 Architecture**

The architecture category contains a number of documents that describe how the system should be built, these provide a blueprint to developers on how to approach any particular problem by helping them select the appropriate tools, platforms and configurations to both meet their need and conform to the overall strategy.

**3.1 Technical Architecture**

The technical architecture describes the technical components that will be used to build the system. This will include the hardware, software and network configuration, along with specific versions where appropriate and standards as to which software product should be used for which job. It also specifies whether two level architecture is used or any other.

**4.0 Data Models**

Data models are (normally) graphical representations of the data that is required. These are normally created in special software that can also generate the DDL required to create the physical objects in the database.

**4.1 Data Modeling Standards**

This document describes the naming conventions of objects in the database, as well as any particular modeling methods (e.g., a hierarchy must always be modeled in a specific way and any exceptions noted along with a justification for the difference). This document should describe the standards for all three data models described modeling techniques for Logical and Physical models.

**4.2 Information Package**

The new methodology for determining requirements for a data warehouse system is based on business dimensions. It flows out of the need of the users to base their analysis on business dimensions. The new concept incorporates the basic measurements and the business dimensions along which the users analyze these basic measurements. Using the new methodology, you come up with the measurements and the relevant dimensions that must be captured and kept in the data warehouse. You come up with what is known as an information package for the specific subject

**4.3 Conceptual model**

The logical data model is a model that represents the true structure of data used by the business, independent of software or hardware implementation constraints. Normally the model is closely related to the information described in Data Warehouse Business Requirements.

**4.4 Logical Model**

The logical data model is a model that represents structure of data used by the business.it depends upon software or hardware constraints.

**5.0 Analysis snapshots**

Here 5 to 6 types of analysis and its results are displayed by using snapshots of the created data warehouse.

**6.0 References**

In this section, references are provided by using APA format.