University of Denver

University College

ICT Program

Cathie Wilson, Instructor

ICT 4410 – Data Warehousing Design

Assignment 4

**NAME: \_\_\_Frederick McGovern\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Use the assigned material from units 7 and 8, notes, and outside sources to answer questions in this assignment.**

**SECTION 1 (25 points)**

**Directions**: The answer to each of the following questions is either True or False. Answer questions by placing a T or F in the cell immediately to the right of each question.

**Answer ALL questions**

|  |  |  |
| --- | --- | --- |
|  | **QUESTION** | **T/F** |
| 1 | The corporate data model represents a highly structured description of the elementary processes within the enterprise that identify the interrelationships among the different processes  (This definition would be for corporate process model Marakas pg 185) | F |
| 2 | One of the most significant data warehouse critical success factors is strong management support  (“Single most critical factor - Marakas pg 188 | T |
| 3 | Organizational culture can have an impact on their readiness to develop a data warehouse because data warehouse development implies cross departmental cooperation and information sharing.  (Analytic Culture of the Organization should be holistic – Marakas pg 189) | T |
| 4 | A bottom-up approach to data warehouse brings more benefit to a data warehouse project than a top-down approach  (“Wherever possible, the top-down approach brings more benefit to the project” Marakas Pg 185) | F |
| 5 | *Horizontal integration* is the means by which the application designer ensures that the software application is congruent with the stated business requirements.  (This definition would be for vertical integration Marakas pg 186) | F |
| 6 | The GA’s smallest unit is called a gene. The gene represents the smallest unit of information in the problem domain and can be thought of as the basic building block for a possible solution.  (Under Basic Component of Gas Marakas pg 140) | T |
| 7 | The corporate process model describes the entities generated by the elementary processes carried out by the enterprise.  (This definition would be for corporate data model Marakas pg 185) | F |
| 8 | Vertical integration ensures that consistent data definitions occur across the entire organization.  (This definition fits Enterprise Integration Marakas pg 186) | F |
| 9 | The long-term benefit of an enterprise data warehouse is the increased integration of business processes throughout the organization.  (Third paragraph Marakas pg 185) | T |
| 10 | One common approach to data warehouse project prioritization is to rank initiatives by impact on the information technology organization.  (“Rank the set of initiatives across two dimensions: impact and near term completion pg 191) | F |

`

**SECTION 2 (25 points)**

**Directions**: Answer each of the following questions by indicating the letter that corresponds to the correct answer. There is only one right answer for each question. Place answer letter in the cell immediately to the right of each question.

**Answer ALL questions**

|  |  |  |
| --- | --- | --- |
|  | **Question** | **Answer** |
| 1 | When the application designer ensures that the software application is in harmony with the business requirements, this is referred to as:  a. horizontal integration.  b. vertical integration. Marakas pg 186  c. enterprise integration.  d. None of the above. | B |
| 2 | Which of the following is not a data warehouse project readiness assessment factor?  a. Strong sponsorship  b. A compelling business motivation  c. An analytic organizational culture  d. All of the above are factors. Marakas pg 188 | D |
| 3 | Which of the following is considered to be a project prioritization dimension?  a. Perceived impact on the core business (along with feasibility of near term completion Marakas pg 191)  b. Economic impact on the IT organization  c. Feasibility of staffing the project internally  d. All of the above. | A |
| 4 | Which term refers to the process of developing a relatively low cost prototype to demonstrate that a data warehouse will achieve the perceived benefits?  a. Proof of prototype  b. Proof of concept (See section titled such Marakas pg 191  c. Proof of benefit  d. None of the above. | B |
| 5 | In a genetic algorithm, the smallest unit of information in the problem domain is called a:  a. chromosome.  b. gene. (See T/F question 6)  c. neuron.  d. synapse. | B |
| 6 | Which of the following is not a tangible benefit?  a. Increase in process quality  b. Reduction in employee turnover - (See table 6-5 Marakas pg 197)  c. Improving time to market  d. Decreased need for travel | B |
| 7 | Which of the following refers to a perspective from which an argument to pursue an investment can be made?  a. Faith  b. Fear  c. Facts  d. All of the above (Though Facts should be only reason investment in DW is made Marakas 194) | D |
| 8 | Which economic feasibility measure results in a return on investment expressed as an interest rate?  a. Net present value  b. Breakeven analysis  c. Internal rate of return (See Marakas pg 202)  d. All of the above. | C |
| 9 | Which of the following is not a genetic algorithm operational process?  a. Evaluation (Marakas pg 141-142)  b. Initialization (Marakas pg 140-141)  c. Mutation (Marakas pg 143)  d. None of the above. | D |
| 10 | A computer program that employs a set of adaptive processes that mimic the concept of “survival of the fittest” by regenerating recombinants of itself in response to a calculated difference between the network’s guess and the desired solution state is known as a:  a. neural network.  b. artificial life system.  c. genetic algorithm. (Type of optimization technique based on natural selection Marakas Pg 140)  d. None of the above. | C |

**SECTION C (50 points)**

**Directions:** Provide complete answers to ALL the following questions based on research and investigation of the topic. Use in-text citations to references and properly cite all references in CMS format.

1. ***Describe the two main components of an enterprise model.***

According to Marakas, the enterprise model contains numerous individual models that when combined together will provide a comprehensive overview of the entire enterprise. Marakas states that of specific interest to designers of data warehouses is the corporate process model and the corporate data model.

The corporate process model provides a “highly structured description of the elementary processes within the enterprise that identify the interrelationships among the different processes (Marakas pg 185).” Essentially, the corporate process model acts as a catalog of processes which one can use to identify a specific process and see its relationships to other processes.

The corporate data model “describes the entities generated by or required by the elementary processes carried out by the enterprise (Marakas pg 185).” Inman states that the corporate data model “reflects the integration of business at the corporate level (Inman pg 219.”, The corporate data model is the starting point in any migration plan to an architecture environment. Inman further states that the corporate data model needs to identify and define such things as major subsets of the corporation, relationships between these subsets and the identification of key and attribute groupings that represent these subsets (Inman pg 270).

***2 Define and describe three levels of enterprise integration.***

Integration occurs on three different levels.

The first level is horizontal integration. There are different definitions of horizontal integration depending on whether you are discussing an enterprise strategy or enterprise modeling. Marakas refers to the latter stating this level should be considered “the most basic form of integration (Marakas pg 186).” In this level, every application in operation and being developed is integrated within itself to ensure that there are no inconsistencies in the final product. The overall purpose is to improve product quality. This level can be achieved without the need to reference the overall enterprise model.

The second level is vertical integration. The purpose of this level is to aid the application designer in ensure that the application being built meets all documented business requirements. Marakas points out that this level is not in widespread use in organizations. However, he points out that to get to enterprise integration this needs to be completed. This level would require the need of an enterprise model at some high level.

The final level is enterprise integration. At this level, all data definitions and process definition across the enterprise will be consistent. This may require the need to engage in the refinement of naming conventions but once completed will ensure that all data and processes in the organization will have been standardized. To achieve this level of integration requires the existence of a enterprise model that has been validated and meets all compliance standards set by the organization. (Marakas pg 186)

***3. Discuss some of the benefits AND limitations associated with genetic algorithms.***

Inman, in his discussion of genetic algorithms describes a number of benefits and a few limitations. The following are a list of these benefits and limitations.

Benefits

* GA’s usually provide a solution must faster than a neural network.
* The process time of a GA is highly predictable and is based upon the number of chromosomes in the initial solution along with the number of generations that will be run.
* If the GA is applied to an optimized problem you are almost guaranteed to find a reasonable solution.
* GA’s can be used to solve problems that we have no idea how to solve.
* The GA has established concepts that are easy to understand and as a result the approach of using a GA can gain acceptance in an organization even from those without a technical background.

Limitations

* Crossovers and Mutations cannot be underutilized or overly utilized. Without crossover you wouldn’t have new structures created to be evaluated. Too high a level of crossover and the good genes disappear along with the bad genes. With mutation too high a level results in duplication of solution space members. However, you need to set some level of mutation to ensure “that the probability of searching any point within the solution is never zero. (Marakas 143)”.
* The population of initial solution chromosomes initially considered should be suitable for the problem in question and should be based on the time it will take to carry out the process and the amount of diversity that exists in the solution population.

***4. Compare and contrast local and global data warehouses***

The local data warehouse contains data that is primarily of interest only to the local level with scope set at that local site. Each site that a corporation has may contain a local data warehouse containing data that is of interest to only that location. Each local data warehouse can be based on different technology or have separate data entry processes. Local data warehouses may exist due to the company having locations in different geological regions. Data in a local warehouse is fed by its own operational systems. The data in a local data warehouse is historical and integrated solely with the singular local data warehouse. There is no coordination of data definitions or data sharing from one local data warehouse directly to another local site.

The global data warehouse on the other hand has as its scope the enterprise integrated across all business entities in the corporation. It contains historical data that is of interest to all local sites. The source of a global data warehouse is the data that is contained in the local data warehouses. This data is data that must be integrated at a corporate level. This typically would be financial data but also may be include customer or product related data.

While local data warehouses primarily support the interests of the local site, it is possible that interests may intersect with the interests of another local site. When an intersection of interest does exists, it is usually a good idea to store the common information in a global database. Since the data for a global data warehouse comes from a local data warehouse, there is typically some redundancy in the data contained in both a local and global data warehouse. Some redundancy may be called for under certain circumstances. Other redundancy can be handed through the use of staging areas in local data warehouse that would contain any data being propagated to the global data warehouse.

The mapping of the data from each local site is one of the more difficult steps that needs to be completed when building the global warehouse. Inman points out that the initial mappings will seldom be optimal the first time so iterative development of a global warehouse is strongly recommended.