

University of Science and Technology
Faculty of computer Science & Information Technology
Collage of Post Graduate Studies and Academic
Advancement

M.S.c in Information systems

AN IMPLEMENTATION OF
DATA INTEGRATION
SYSTEM

Thesis submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Information System

Prepared by :

Ammar kararr Mahdi mohamednor

Supervisor :

Dr. Atif Ali Mohamed

February 2014

Abstract

The goal of a data integration system is to provide a uniform interface to a multitude of data sources. Given a user query formulated in this interface, the system translates it into a set of query plans. Each plan is a query formulated over the data sources, and specifies a way to access sources and combine data to answer the user query. In practice, when the number of sources is large, a data integration system must generate and execute many query plans with significantly varying utilities. Hence, it is crucial that the system finds the best plans efficiently and executes them first, to guarantee acceptable time to and the quality of the first answers.

In this thesis we described data integration in two sections, one about overview of data integration, goals and techniques of data integration besides to some related work.

The other section is about implementing data integration in a small system searching in a multi data source in a single computer using vb.net 2008.

As a result of the implemented system we found that the system helps more in searching process with accurate result in all type of data source.

المستخلص

إن الهدف من بناء نظام متكامل للبيانات هو تصميم شاشة تتيح للمستخدمين الوصول الي انواع مختلفة من البيانات حيث يقوم نظام تكامل البيانات بتوفير شاشه واحده لتنفيذ البحث والوصول الي مختلف انواع البيانات .

يتم تصميم مثل هذه الانظمة عن طريق انشاء محتوى وسيط يستطيع الوصول الي هذه الانواع المختلفة من البيانات والموجوده في مصادر مختلفة

في هذا البحث تحدثنا عن نظام تكاملية البيانات من جانبين في الجانب الاول قمنا بتعريف تكاملية البيانات واهدافها و التقنيات المستخدمة في هذا المجال بالاضافة الي بعض الدراسات التي لها علاقة بتكاملية البيانات.

و الجانب الثاني يتحدث عن تصميم نظام صغير لتكاملية البيانات للبحث في انواع مختلفة من البيانات الموجوده في جهاز حاسوب باستخدام فجوال بيسيك دوت نت 2008.

ونتيجة لتطبيق هذا النظام فقد وجدنا انه يساعد في عملية البحث بصورة افضل ويعطي نتائج دقيقة في جميع انواع البيانات.

1.1 Introduction:

The goal of a data integration system is to provide a uniform interface to a multitude of data sources. Given a user query formulated in this interface, the system translates it into a set of query plans. Each plan is a query formulated over the data sources, and specifies a way to access sources and combine data to answer the user query. In practice, when the number of sources is large, a data integration system must generate and execute many query plans with significantly varying utilities. Hence, it is crucial that the system finds the best plans efficiently and executes them first, to guarantee acceptable time to and the quality of the first answers

When we use data integration in in large companies and firms that means we create main interface to access different data type in different source, this lead to achieve some objectives such as the first objective in Query

(The focus of most data integration systems is on querying disparate data sources. However, updating the sources is certainly of interest), the second objective in Number of sources (Data integration is already a challenge for a small number of sources (fewer than 10 and often even 2!), but the challenges are exacerbated when the number of sources grows), the third objective in Heterogeneity (A typical data integration scenario involves data sources that were developed independently of each other. As a consequence, the data sources run on different systems: some of them are databases, but others may be content management systems or simply files residing in a directory. The sources will have different schemata and references to objects, even when they model the same domains. Some sources may be completely structured (e.g., relational databases), while others may be unstructured or semi-structured (e.g., XML, text), and fourth objective in Autonomy (The sources do not necessarily belong to a single administrative entity, and even when they do, they may be run by different sub organizations. Hence, we cannot assume that we have full

access to the data in a source or that we can access the data whenever we want, and considerable care needs to be given to respecting the privacy of the data when appropriate. Furthermore, the sources can change their data formats and access patterns at any time, without having to notify any central administrative entity).

1.2 Problem Definition:

The great development in computers and networks has led to the emergence of a lot of different types of data deployed in these networks which became the resource of the data that users need to enquires with in different sources, further more to access data without complexity and very quickly.

The main problem that we discuss in this search is how to develop data integration system users to access varies type of data.

1.3 Main Objectives:

The main objectives in this search is implement data integration system enable users to access different types of data exist in a single computer by using visual basic dot net 2008.

1.4 Sub Objectives:

Data integration in helping to produce an effective official statistical system is becoming increasingly apparent. The process of bringing together information from different sources paves the way for a broader range of questions to be answered. Through integration it becomes possible to examine underlying relationships between various aspects of society, thus improving our knowledge and understanding about a particular subject

1.5 Research Structure:

The research Structured as follow:

- Chapter two contains a background for data integration and its' goals and discuss some algorithms in data integration and architectures of designing data integration.
- Chapter three discuss techniques in data integration and some algorithms in designing data integration system.
- Chapter four shows a related work on the data integration.
- Chapter five implemented data integration system for do query in single computer by using visual basic.net 2008.
- Chapter six our research conclusion and recommendation.

