

UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF GRADUATE STUDIES AND
ACADEMIC ADVANCEMENT

Design a Data Quality Framework For
The Distributed Database System
(Case Study The Mena Factory Iron Production)

A Thesis

Submitted to the College of graduate Studies and Academic
Advancements
in Partial Fulfillment of the Requirement for the Degree of Master of
Computer Science and Information Technology

Prepared By:

Abdelrhman Ismail Mohamed

Supervisor By:

Dr: Atif Ali Mohamed

Nov2016

Abstract

Database is a set of data that describes the activities of one or more of the organizations linked to with clear structure parameters. The database management control (DBMS) through the maintenance and use of large data sets.

Distributed database is a collection of data that can be stored in different locations across the computer network. Distributed systems require the amount of independent hardware and these independent devices work with each other or operating independently but share the source.

The problem of the research institutions are needed in all areas for accurate data in order to be utilized in the form required of them in making the right decisions using the data quality techniques to allow institutions in all areas of ability to focus on good information in the database.

So all institutions need information to help companies and institutions to take decisions and actions to increase knowledge and identify opportunities and focus on competitive threats.

If this is the need for information The information should be a high degree of precision, power and integration and also distributed databases are distributed geographical distribution even be a need to organize the information to be achieved through the implementation of transactions on the database and implemented to determine the key objective is to build a policy or work environment to ensure the implementation of distributed databases and organize data on the degree of integrity of information and overall quality.

المستخلص

قاعدة البيانات عبارة عن مجموعة من البيانات التي تصف أنشطة أو أكثر ذات صلة بالمنظمات ذات بيئة واضحة المعالم . يتم التحكم بقواعد البيانات (DBMS) من خلال صيانة واستخدام مجموعات كبيرة من البيانات. قواعد البيانات الموزعة هي عبارة عن مجموعة من البيانات التي يمكن تخزينها في مواقع مختلفة عبر شبكة الحاسب الالى . النظم الموزعة تتطلب كمية من الأجهزة المستقل وهذه الأجهزة المستقلة تعمل مع بعضها البعض أو تعمل بشكل مستقل ولكنها تشترك في المصدر.

أن مشكلة البحث هي حوجة المؤسسات في جميع المجالات لبيانات دقيقة لكي تتم الاستفادة منها بالشكل المطلوب في اتخاذ القرارات الصحيحة بأستخدام تقنيات جودة البيانات لتمكين المؤسسات في جميع المجالات القدرة علي التركيز علي المعلومات الجيدة في قاعدة البيانات.

لذلك كل المؤسسات تحتاج إلى المعلومات لمساعدة الشركات والمؤسسات في اتخاذ القرارات والأعمال لزيادة المعرفة وتحديد الفرص والتركيز على التهديدات التنافسية.

إذا كانت هذه هي الحاجة إلى المعلومات يجب أن تكون المعلومات على درجة عالية من الدقة والقدرة والتكامل وبما ان قواعد البيانات الموزعة يتم توزيعها الجغرافي حتى يكون في حوجة في تنظيم المعلومات المراد تحقيقها من خلال تنفيذ المعاملات على قاعدة البيانات وتنفيذها لتحديد هدف رئيسي هو بناء سياسة او بيئة عمل لضمان تنفيذ قواعد البيانات الموزعة و تنظيم البيانات على درجة من السلامة والجودة الشاملة للمعلومات.

1.1. Introduction:

We live in an era of unprecedented data abundance and aggregation. The sheer variety of new information available on the Internet, in databases, and from other sources has changed the way we conduct business, undertake research, and communicate. Most of the changes are positive. Yet, increased reliance upon networked data has also introduced new challenges. One serious problem we need to address is that of incorrect data missing or inaccurate information that resides in (and, indeed, frequently results from) the abundance and aggregation of data in our lives today. Dirty data can have several pernicious effects. In particular, it:

1. Impacts the quality of care.
2. Introduces privacy and other civil liberty concerns.
3. Increases costs and inefficiencies.
4. Creates liability risks and undermines the reliability and benefits of information technology (IT) investments, including the potential to streamline service delivery, accounting, and billing. Wrong data to system which will effect in business, prevent system itself from unauthorized access.

In current research comes as an attempt to address the subject through the side of a theoretical addressed of intellectual the foundations for the integrity of the data by the field and taking design a framework for applying data quality within the Mena factory Steel Products.

1.2. Problem Statement:

the data when it is distributed in distributed databases and implement by some of the processes can cause some of the problems such as the complexity, cost, the Security, difficulty of ensuring the validity of data and synchronous and that you need to implement some of the techniques to solve these problems, controls the enterprises need to design a framework to ensures quality data to help organizations in making the right decision.

1.3. Research Objective:

Objectives of the research is to design a framework for building data quality and technology policy and to ensure that similar data after the implementation of the processes of the deletion, and modification added ... etc application on a base

the Mena factory data Steel Products in distributed database to be the data on the a high level of quality and integrity.

1.4. Research Methodology:

In this research design has been adopted the framework on the previous studies and research undertaken for the purpose of to find solutions and challenges of data quality and the application of comprehensive quality of the policies of data in a distributed database .

1.5. Research Structure:

Contains five chapters: the first chapter is introduction. While the second chapter is related work chapter, which discussed distributed database and related work. And the third chapter is literature review. The fourth chapter is design framework . While the fifth chapter is the last chapter, defines conclusions and recommendations.

