NAME: SHRAVANI SUBHASH PAWAR

ROLL NO: A041

CONTROL ID: 2020080174

CLASS: TYBSC IT

SUBJECT: BUSINESS INTELLIGENCE

TOPIC: DATA WAREHOUSE

A Real Time Data Warehouse Approach for Data Processing

SUMMARY:

A real-time data warehouse closes the data availability gap, allowing enterprises to focus on processing their valuable data. The effort of managing, interpreting, and maintaining raw data collected from many data sources is quite challenging and requires a lot of processing. The real-time approach enables users to process data, record data for later use, evaluate complicated structure and relationships, and retrieve data from various data sources in almost real-time.

One significant issue with data warehouse development is the capture of change data from source systems. The most common strategies in this field are log capture and triggering. We believe that log capture may be preferable to triggering, which burdens the source system. However, other than the source database, it's simple to implement. Triggers are supported by the majority of databases today, thus it would be an excellent option.

Another major difficulty for creating a real-time data warehouse is real-time segmentation. Data could not be loaded instantly into the data warehouse due to the design purposes of the data warehouses. Because your data is finished later, a staging table is required before aggregating and loading it into the data warehouse.
As a result, we employ a real-time partition. Additionally, the data in the real-time partition and data warehouse must be combined if the user asks real-time data for his study.

RESEARCH ISSUES ON DATA WAREHOUSE MAINTENANCE

SUMMARY:

There are many issues with the system because data warehousing is a new field. Data warehouse upkeep is one of the biggest issues the business is now dealing with. Data warehouses are large systems, therefore the processes of data extraction, cleaning, and loading take a lot of time. Data warehouse developers frequently encounter issues in the operational systems where data must be collected.

After being deployed, the data warehouse must be handled like a production system with service level agreements. In order to get the most out of the system, technical support for the data warehouse should regularly analyze trends in system capacity and performance. When dealing with continuing data warehouse performance monitoring, a few considerations are discussed.

The data warehouse team is responsible for making sure that the current implementations stay on course and continue to meet business needs.
Access performance for executing queries and incremental loading of snapshot changes from the source systems are the main performance challenges in data warehousing. Six concepts are mentioned that can be considered for a better performance.

Companies use extract, transform, and load (ETL) technology, which entails reading data from its source, cleaning it up and universally formatting it, and then writing it to the target repository for usage.

Classification of Data Granularity in Data Warehouse

SUMMARY:

The main challenge with data granularity is maintaining it at a level that is neither too high nor too low. Low data granularity will result in detailed data, but it will also increase storage requirements and increase query times. High granularity degrees can be searched quickly and easily, but they are limited in how much detail they can supply.

An essential component of the data warehouse, the overflow memory, which was used to store the uncommonly used data, has a significant impact on granularity. The granularity classification of the data warehouse will be directly impacted by the sort of analysis that can be done there. Less detailed processes can be carried out in the data warehouse as granularity increases.

Multiple granularities can be used to achieve a balance between the volume of data and the ability to answer queries. Data warehouse, hardware, software, and business decision analysis are all used to compromise on the classification of the data granularity.

The classification of the bank environment data granularity is made based on a thorough study of the peculiarities of banking data as well as the raised strategy and principles. Although the sample's granularity classification is rather straightforward, it can better serve the needs of banking industry analysis.
Additionally, it can be used as a model to classify the banks' other pertinent data at various granularities.

Research of Decision Support System Based on Data Warehouse Techniques

SUMMARY:

The DSS offers every type of decision information as well as answers to business-related questions as one of a kind intelligent software systems with significant application value. Data warehouse is hence suited to serve as the technology framework for the DSS's database management subsystems and may satisfy their requirements.

DSS based on data warehouses also emerged at this historic time, along with data warehouse's broad use. Starting with the DSS's actual demand, the data warehouse's characteristics and structure have been examined. The data used to make decisions is fully extracted, integrated, and synthesized by DW. Data mining and thematic-dimensional data analysis are utilized in OLAP to uncover relevant information for decisions being supported by databases and DW.

DW construction should be given priority since it is the foundation of the DSS. However, because it requires a lot of labor, physical resources, money, and time, data bazaar is one type of department advancement. DW has a limited scope, a narrow application to the department, and a low development cost. As a result, it is believed by many businesses that building DW may begin with the creation of the data bazaar and proceed in three steps: Create a couple of a fictitious data fair of departments first. Establish a data fair for each department next. Create the enterprise progression DW third.

DW is the appropriate DSS database because it offers huge data storage, quick parallel processing, preservation of history sequence data, trend analysis, data integration and synthesis, and direct user inquiry.

Application of Data Warehouse Technology in Data Center Design

SUMMARY:

The development of an E-government system is already quite popular all over the world, because it helps in the dissemination of information. The Chinese government has set ambitious goals for the implementation of e-government after completing five Golden Projects. These goals include accelerating change in government functions to meet the demands of reform, opening up, and modernization policies, enhancing the efficiency of government operations, introducing new government measures in a scientific manner, developing more effective mechanisms to monitor economic activity, and placing a greater emphasis on central coordination.

The data warehouse is now more widely understood as a technique for collecting data sources under a single conceptual and technical framework and making the information accessible for new operations or decision-support applications.

There are six basic parts that make up a data center: the platform for sharing and exchanging data, the kernel database, the platform for supporting applications, the application database, the platform for managing data centers, and the platform for managing data centers' security. The flexible and erratic access to data is the core heart of the data warehouse. Monitoring data cannot be so expensive or difficult due to its complexity, making it impossible to deploy a monitoring software when required. Additionally, the data warehouse must be able to transfer data to and from the various Nanhai city agencies.

For the sharing and exchanging of data, XML technology is employed. A data center system for Nanhai city has previously been developed using the suggested methodology with success.