Data Warehousing

by Sudheer Sharma - Monday, December 15, 2008

http://dwhnotes.com/data-warehousing

What do you think of this post?

Awesome (5) Interesting (3) Useful (4)

Overview

A Data Warehouse has been used to refer to a database that contains very large stores of historical data. The data is stored as a series of snapshots, in which each record represents data at a specific time. This data snapshot allows a user to reconstruct history and to make accurate comparisons between different time periods.

Typical relational databases are designed for on-line transactional processing (OLTP) and do not meet the requirements for effective on-line analytical processing (OLAP). As a result, data warehouses are designed differently than traditional relational databases.

Typical uses

A data warehouse integrates and transforms the data that it retrieves before it is loaded into the warehouse. A primary advantage of a data warehouse is that it provides easy access to and analysis of vast stores of information.

DEFINITION, ARCHITECTURE AND CONCEPTS

- Enterprise Data Model
- Operational vs. Historical Data
- Extract Transform and Load (ETL)
- Extract Load and Transform (ELT)
- ETL vs. ELT
- Metadata
- DataMart
- Data Warehouse vs. Data Mart
- Data Integration
- Data Mining
- Operational Data Store (ODS)
- OLAP vs. OLTP
- Logical design vs. Physical design

- Normalization vs. De-normalization
- Referential Constraints
- Repository
- Summary Tables
- Fundamental Stages of DWH
- Different Methods of loading Data
- RealTime DW

DATA MODELING OPTIONS

- Data Modeling Overview
- Entity Model
- Star Schema
- Snowflake Schema

IMPLEMENT OPTIONS

- Overview
- Top Down
- Bottom up
- Sizing
- Cleaning
- Populating the data warehouse

EXTRAT, TRANSFORM, LOAD (ETL) TERMS AND CONCEPTS

- Options
- Extraction Options
- Transformation Options
- Loading Options
- Changed Data Capture (CDC) and Publishing
- Staging Areas

DIMENSIONAL MODELLING DEVELOPMENT LIFE CYCLE

- Requirements Analysis
- Requirements Gathering
- Requirements Validation
- Requirements Modeling
- Schema Design
- Project Definition
- Warehouse Design

- Implementation
- Follow-up and Review

DIMENSIONAL MODLEING DESIGN

- Overview
- Metadata Properties
- Star Schema
- Snowflake Schema
- Cubes
- Measures and Facts
- Attributes and Relationships
- Dimensions
- Hierarchies
- Joins
- Summary Tables and Aggregations

DATA CLEANING AND CONFORMING

- Data Quality Criteria
- Design Methods and Alternatives
- Cleaning deliverables
- Conforming Dimensions
- Conforming Facts

DIMENSION TABLE DELIVERY

- Dimension Table Structure
- Surrogate Key Generation
- Dimension Table Grain
- Flat(denormalized)or snowflake
- Date and Time Dimensions
- 'Big' vs 'Small' Dimensions
- Dimensional roles
- Dimensions as Sub-Dimensions
- Degenerate Dimensions
- Junk Dimensions

SLOWLY CHANGING DIMENSIONS (SCD)

- Type1, Type2, Type3, Type4
- Late arrivals

SLOWLY CHANGING FACTS (SCF)

- Slowly Changing Facts

MULTIVALUED DIMENSIONS

- Definition
- Bridge Tables

FACT TABLE DELIVERY

- Fact Table Structure
- Referential Integrity
- Surrogatekey Derivation & Flow
- Fundamental grain
- Transcation Fact Tables
- Factless Fact Tables
- Periodic Snapshots
- Accumalating Snapshots

FACT TABLE LOAD CONSIDERATIONS

- Index Management
- Partition Management
- Updates, Deletes and Inserts
- Recovery
- Summary Tables
- Parallelism

DATA WAREHOUSE PERFORMANCE DESIGN

- Materialized View
- Large Concurrent Reports
- Short Running Queries
- Long Running Queries
- Random Queries
- Occasional updates
- On-Line utilities
- Index Options
- Partitioning and Parallelism

INTRODUCTION TO STATISTICS, ANALYTIC AND OLAP SQL QUERIES

- AVG
- CORRELATION
- COUNT
- COUNT_BIG
- MAX
- MIN
- RAND

- STDDEV
- SUM
- VARIANCE
- REGRESSION FUNCTIONS
- GROUPING, ROLLUP AND CUBE

GENERAL CONCEPTS

- What is HIPAA?
- Data Warehouse in the Insurance Industry

References:

1. http://www.sqlservercentral.com/

2. http://www.programmerinterview.com/index.php

What do you think of this post?

Awesome (5) Interesting (3) Useful (4)
