Data Management

Overview

- ➢ **Data management comprises all the disciplines related to managing data as a valuable resource.**

- ➢ **Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.**

What is Data Preparation?

➢ **Data Preparation is a pre-processing step in which data from one or more sources is cleaned and transformed to improve its quality prior to its use in business analytics.**

Why perform data preparation?

The goal of data preparation is the same as other data hygiene processes: to ensure that data is consistent and of high quality. Inconsistent, low quality data can contribute to incorrect or misleading business intelligence. It can create errors and make analytics and data mining slow and unreliable. By preparing data for analysis up front, organizations can be sure they are maximizing the intelligence potential of that information.

What are the activities in data preparation?

Data preparation involves the rationalization and validation of data to make sure data is formatted consistently and that the data will be understood once removed from its source. It can involve changing the formats of dates, or deleting duplicate fields.

When is data preparation needed?

Data preparation efforts are often needed during the integration of disparate applications that occur during merger and acquisition activities, but also when siloed data systems within a single organization are brought together for the first time in a data warehouse or big data repository.

What are the benefits of data preparation?

When data is of excellent quality it can be easily processed and analyzed, leading to insights that help the organization make better decisions. High-quality data is essential to business intelligence efforts and other types of data analytics, as well as better overall operational efficiency.

What is Data Cleaning

> ➢ **Data cleaning, also called data cleansing or scrubbing, deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data.**

Data quality

High-quality data needs to pass a set of quality criteria. Those include:

➢ **Validity - The degree to which the measures conform to defined business rules or constraints.**

Data Quality (cont.)

➢ **Decleansing is detecting errors and syntactically removing them for better programming.**

➢ **Accuracy - The degree of conformity of a measure to a standard or a true value.**

Data Quality (cont.)

➢ **Completeness - The degree to which all required measures are known.**

➢ **Consistency: The degree to which a set of measures are equivalent in across systems.**

➢ **Uniformity: The degree to which a set data measures are specified using the same units of measure in all systems.**