Meta Data

- **Metadata** is loosely defined as data about data.
- Metadata is a concept that applies mainly to electronically archived or presented data and is used to describe the
  - a) definition,
  - b) structure and
  - c) administration of data files with all contents in context to ease the use of the captured and archived data for further use.
  - example: a web page may include metadata specifying what language it's written in, what tools were used to create it, where to go for more on the subject and so on

What is Meta data

- **Metadata** (**meta data**, or sometimes **metainformation**) is "data about other data", of any sort in any media. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items and hierarchical levels, such as a database schema. In data processing, metadata provides information about, or documentation of, other data managed within an application or environment. This commonly

  defines the structure or schema of the primary data.

  - metadata would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.). Metadata may include descriptive information about the context, quality and condition, or characteristics of the data. It may be recorded with high or low granularity

Definition:

Metadata contains information about that data or other data

Metadata is structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities

Why Metadata is important

- Assume that the project team has completed successfully the development of first data mart.But the user can have several things in mind:
  - Are the predefined queries I look at
  - What are the various elements in data warehouse
  - Is there information about unit sales and unit costs by product
  - How can I browse and see what is available
  - From where did they get the data for data warehouse? From which source system
  - How old are data warehouse
  - When is the last time fresh data was brought in
  - Are there summaries by months and product

- We can define meta data in terms data warehousing which includes:
  - Data about data
  - Table of content for data
  - Catalog for data
  - Data warehouse roadmap
  - Data warehouse directory

Applications of Metadata

- **Libraries**
- Metadata has been used in various forms as a means of cataloging archived information.
- **Photographs**
- Metadata may be written into a digital photo file that will identify who owns it, copyright & contact information, what camera created the file, along with exposure information and descriptive information such as keywords about the photo, making the file searchable on the computer and/or the Internet
- **Web pages**
- Web pages often include metadata in the form of meta tags. Description and keywords meta tags are commonly used to describe the Web page's content. Most search engines use this data when adding pages to their search index.

Classification of Metadata types

- Classification of metadata types by functional areas:
  - Data acquisition
  - Data storage
  - Information delivery

- Acquisition process:
  - Data Extraction
  - Data transformation
  - Data cleansing
  - Data Integration
  - Data staging

- Metadata Types:
  - Source system platforms
  - Source structure definition
  - Data extraction method
  - Data transformation rules
  - Data cleansing rules
  - External data sructures
  - External data definition
  - Summerization rules
  - Target physical and logical models

Data Storage

- The metadata used recorded by the process in data storage area is used for development ,administration and for user.

- User would like to see what is the last time previous data loaded.

- DBA will use the metadata for processes backup and incremental loads.

Information Delivery

- Information delivery
  - Report generation
  - Query processing
  - Complex Analysis

- Metadata types:
  - Source systems
  - Source data definitions
  - Data extraction tools
  - Query templates
  - Preformatted reports
  - OLAP content