SNS COLLEGE OF ENGINEE

Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna Univer

Department of Information Tech

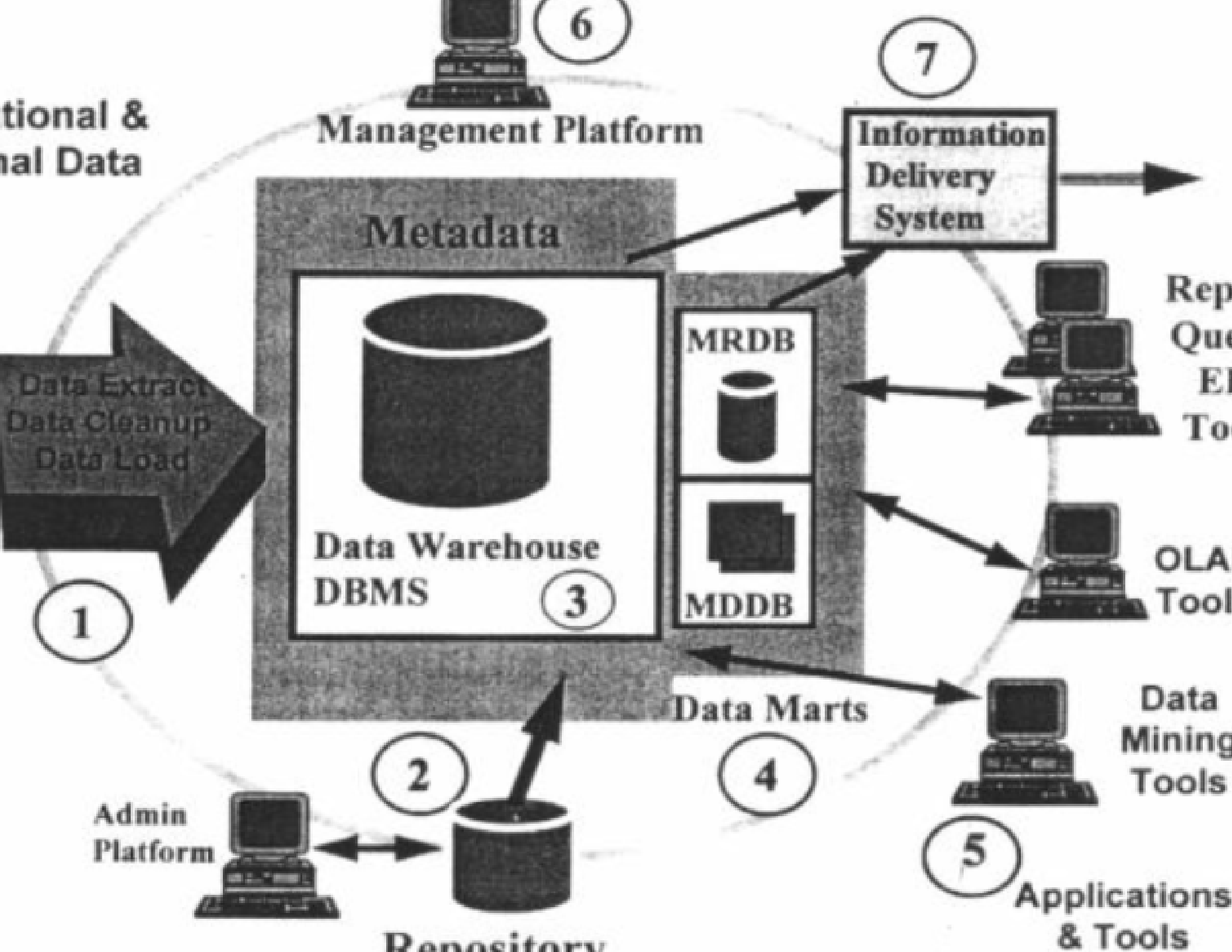
WAREHOUSE ARCHITE
OR COMPONENTS OF D

Prepared by
T.R.Lekhaa,
AP/IT

COMPONENTS OF DW

ata sourcing, cleanup, transformation, and migration tool
etadata repository
'arehouse/database technology
ata marts
ata query, reporting, analysis, and mining tools
ata warehouse administration and management
formation delivery system

Data warehouse is an environment, not a product which is
elational database management system that function
entral repository for informational data. The central
nformation is surrounded by number of key components
o make the environment is functional, manageable and a


tional &
nal Data
6
Management Platform
7
Information
Delivery
System
Metadata
MRDB
MDDB
Data Warehouse
DBMS
3
Data Extract
Data Cleanup
Data Load
1
Data Marts
4
2
Admin
Platform
Repository
5
Applications
& Tools
Data
Mining
Tools
OLA
Tool
Rep
Que
E
To

1. Data warehouse database

data source for data warehouse is coming
rational applications.

data entered into the data warehouse transforme
ntegrated structure and format.

transformation process involves conve
marization, filteration.

data warehouse must be capable of holding
aging large volumes of data as well as different stru
ata structures over the time.

a warehouse database is the central part of the
ehousing environment. This is the item number 2 i
ve arch. diagram.

a warehouse database is implemented based on R
nology.

ourcing, Acquisition, Clean up and Transformation Tools

s item number 1 in the above arch diagram.

alled as Extract, Transform and Load (ETL) To

perform conversions, summarization, es, structural changes and condensation.

data transformation is required so that nation can by used by decision support tools.

transformation produces programs, co nents, JCL code, COBOL code, UNIX scripts, DDL code etc., to move the data into house from multiple operational systems.

Sourcing, Acquisition, Clean up and Transformation Tools

functionalities of these tools are listed below:

- To remove unwanted data from operational db
- Converting to common data names and attributes
- Calculating summaries and derived data
- Establishing defaults for missing data

es to be considered while data sourcing, cleanup, extract a sformation:

abase heterogeneity. DBMSs are very different in data mode access language, data navigation, operations, concurren grity, recovery etc.

a heterogeneity. This is the difference in the way data is defin used in different models – homonyms, synonyms, u patibility (U.S. vs metric), different attributes for the sa ty and different ways of modeling the same fact.

3. Meta data

data about data. It is used for maintaining, managing and using
warehouse.

classified into two:

***cal Meta data*:** It contains information about data warehouse d
warehouse designer, administrator to carry out development a
ment tasks. It includes,

mation about data stores

formation descriptions. That is mapping methods from operatio
warehouse db

house Object and data structure definitions for target data

ules used to perform clean up, and data enhancement

mapping operations

ss authorization, backup history, archive history, info delivery histo
acquisition history, data access etc.

3. Meta data

ss Meta data: It contains information that gi
ion stored in data warehouse to users. It includes,

ject areas, and information object type including quer
orts, images, video, audio clips etc.

ernet home pages

rmation related to info delivery system

a warehouse operational info such as ownerships, audit t
,

data helps the users to understand content and f
ata. Meta data are stored in a separate data sto
is known as informational directory or Meta d
tory which helps to integrate, maintain and view
nts of the data warehouse.

3. Meta data

ollowing lists the characteristics of info
tory/ Meta data:

s the gateway to the data warehouse environment

supports easy distribution and replication of content for rformance and availability

should act as a launch platform for end user to access data alysis tools

hould support the sharing of info

hould support scheduling options for request

hould support and provide interface to other applications

should support end user monitoring of the status of the rehouse environment

4. Access tools

purpose is to provide information
iness users for decision making.

er interact with DW using front end too
ere are five categories:

- Data query and reporting tools
- Application development tools
- Executive information system tools (EIS)
- OLAP tools
- Data mining tools

4. Access tools

ery and reporting tools are used to generate query and report.

egory:

eporting tools

anaged query tools

ng tools types:

- ✓ Production reporting tool used to generate regular operational reports like calc and printing paychecks
- ✓ Desktop report writer are inexpensive desktop tools designed for end users.

ed Query tools:

used to generate SQL query.

It uses Meta layer software in between users and databases v
s a point-and-click creation of SQL statement.

This tool is a preferred choice of users to perform seg
tification, demographic analysis, territory management
aration of customer mailing lists etc.

4. Access tools

cation development tools:

is is a graphical data access environment whic
tegrates OLAP tools with data warehouse and can b
ed to access all db systems.

plication development platforms integrate well wit
pular OLAP tools, and can access all major DB system
cluding Oracle, Sybase, Informix.

amples of application development environmen
clude Visual Basic from Microsoft, PowerBuilder fro
werSoft.

4. Access tools

LAP Tools:

are used to analyze the data in multi dimensional and complex v

Business applications for these tools : product perform profitability, effectiveness of sales program or marketing campa

To enable multidimensional properties it uses (Multi Dimensional Data Base) and MRDB (Multi Relationa Base)

Data mining tools:

are used to discover knowledge from the data warehouse also can be used for data visualization and data corr purposes.

DM used to perform segmentation (group customer recor custom-tailored marketing), classification (assignment of data to a predefined class, discovery), association (discov cross-sales opportunities), preferencing (determining prefe of customer's majority)

5. Data marts

sive alternate to DW, requires less time and money to bu
dent Data Mart
rt – meaning – different things to different people.

re that is subsidiary to a DW of integrated data.
enormalized, summarized, aggregated data

→ Dependent dat
because their c
sourced from D

cted at a partition of data, that is created for the use of dedicated g
r Departmental subsets that focus on selected subjects.
rt is used in the following situation:
emely urgent user requirement
absence of a budget for a full scale data warehouse strategy
decentralization of business needs
attraction of easy to use tools and mind sized project
dent Data mart presents two problems:
ility: A small data mart can grow quickly in multi dimensions. So that while des
nization has to pay more attention on system scalability, consistency and mana

tegration

6. Data warehouse administration and management

e management of data warehouse includes,

Security and priority management

Monitoring updates from multiple sources

Data quality checks

Managing and updating meta data

Auditing and reporting data warehouse usage and statu

Purging data

Replicating, sub setting and distributing data

Backup and recovery

Data warehouse storage management which inclu capacity planning, hierarchical storage management purging of aged data etc.,

nformation delivery system

used to enable the process of subscribing fo
warehouse information.

ery to one or more destinations according t
fied scheduling algorithm.

ner words, It distributes warehouse-stored dat
other information objects to other dat
houses and end-user products like spreadshee
ocal databases