UNIT - III

META DATA, DATAMART & PARTITION STRATEGY

Meta Data - Categories of Metadata - Role of Metadata - Meta data Repository - Challenges for Meta Management - Data Mart - Need of Data Mart - Cost Effective Data Mart - Designing Data Marts - cost of Data Marts - partitioning - strategy - Vertical partition - Normalization - ROW splitting - Horizontal partition.

TOPIC NAME:- META DATA

Data about data, contains
- location & description of system components
- deliver end-user analytical tools
- update snapshots, owners, authorization
- analyze end user usage patterns
- describe, maintain, manage data Warehouse
- Easy access - tools

METADATA INTERCHANGE INITIATIVE:-

- Launched by vendors / users
- difficult manage, share, exchanging metadata
- Include API & manage access
- declare standards
- Extending & updating Industry user needs

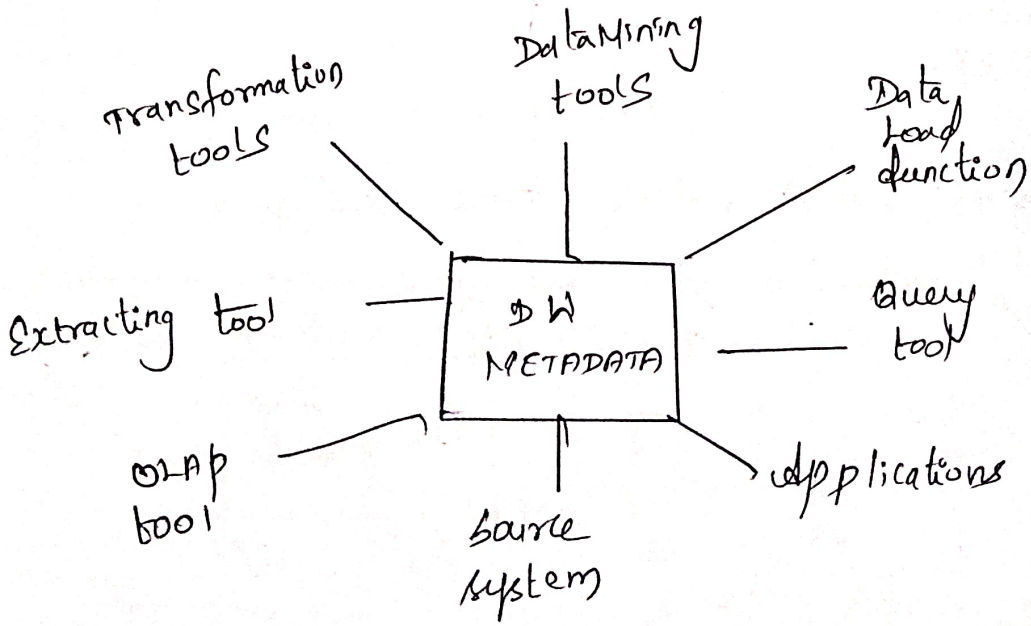
Approach - procedural / ASCII Batch / Hybrid
- procedural → Communicate API
- ASCII Batch → Access parameter
- Hybrid → Interact standard access of object needed

TOPIC NAME:-

CATEGORIES OF METADATA:-

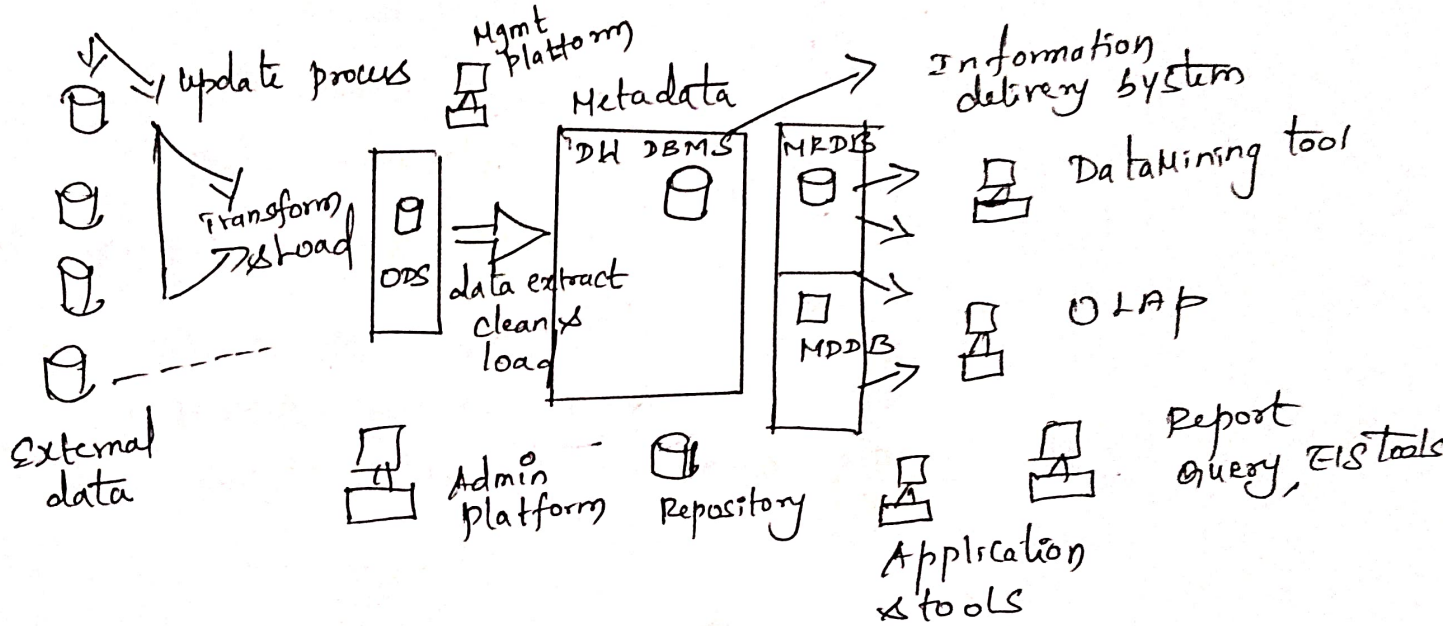
Backroom Metadata – extract /clean/load process
Front room Metadata – Direct end user reports
Process Meta data – Access end users
Data Lineage – transform source system
Business definitions – store metadata for future reference
Technical definitions – Reduce ambiguity & reuse existing table
Business Metadata – NO data manipulation
Technical Metada – Monitor db load & Backup
Operational Meta data – store data source
Source system Information – db, specification, definitions etc
ETL Job Metadata – Store jobs /table /files
Transformation Metadata – ETL used & reuse predefined transformations

TOPIC NAME :- Role of Metadata



TOPIC NAME :- METADATA REPOSITORY

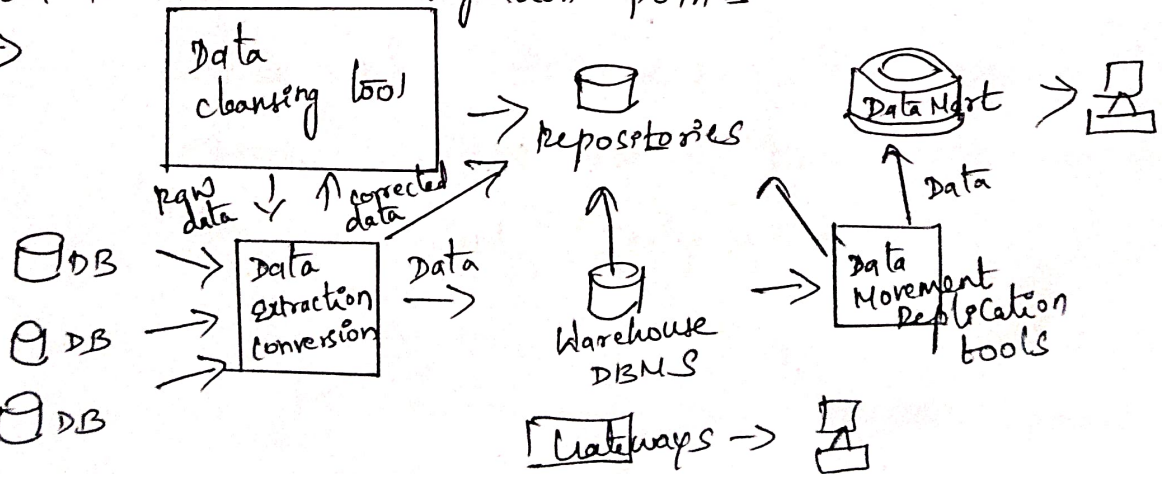
Include to map source data to target db generate code for data transformation, integrate & control workstations enables user to specify data mapping, conversion & summarization defines location of data relationship between operational database accusible by end-user tools find data definations on subject areas.



DH : ARCHITECTURE

Maintain access path, reduce & eliminate information Redundancy, inconsistency and under utilization.

fig :- Tool & Metadata integration points



→ Improve organization control & increase coordination & utilization
→ Increase flexibility / ctrl / reliability
→ utilize existing application.
→ eliminate redundancy & reuse

TOPIC NAME :- CHALLENGES FOR META MANAGEMENT :-

lack of standardization
Data quality
Data integration
Data governance
Data security

TOPIC NAME :- DATA MART :-

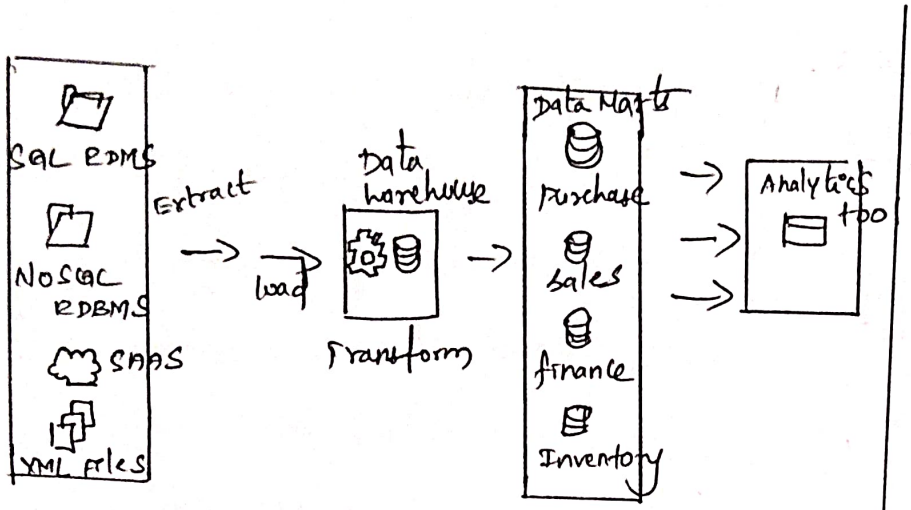
subset of information store
provide business needs
analytical record stores designed to focus business fns

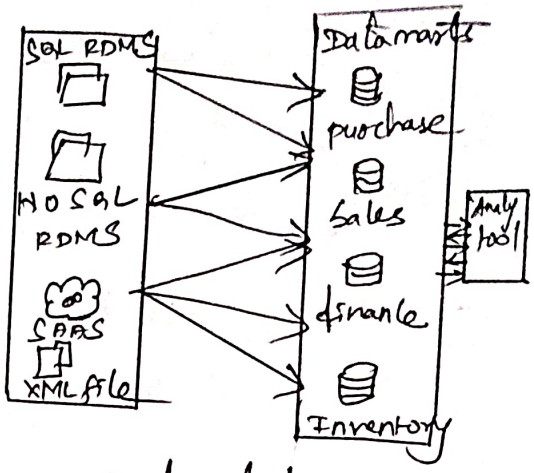
fig : Data Mart

fundamental use of data mart is Business Intelligence (BI)
used to gather, store, access & analyze record
utilize data / less expensive / easy creation / lower cost
Improve end user response time / access frequently group of users
Types — dependent / Independent

↓ Extract record — top down approach
↓ Integrate to develop datawarehouses — bottom up approach

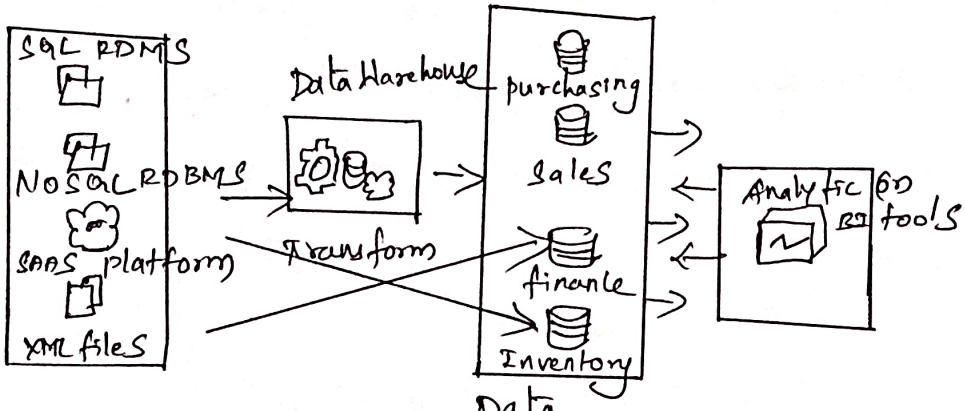
DEPENDENT DATA MART





Independent data Mart

Hybrid data Mart :-



Steps Implementing Data Mart :-

Gather data & develop physical & logical design of data mart

Gather technical requirements

Identify data sources

constructing — Access data structure

Populating — Extract data from source & modify

Accessing — Analyze report / chart / graphs

Managing — Secure access to data / better performance availability of system failures ensure

DATA WAREHOUSE	DATA MART
vast repository information	meet user group requirement
hold multiple subject	hold only one subject
detailed information	summarize data
integrate all data source	integrate source system
Centralized system	decentralized system
data oriented	project oriented

TOPIC NAME:- NEED OF DATA MART:-

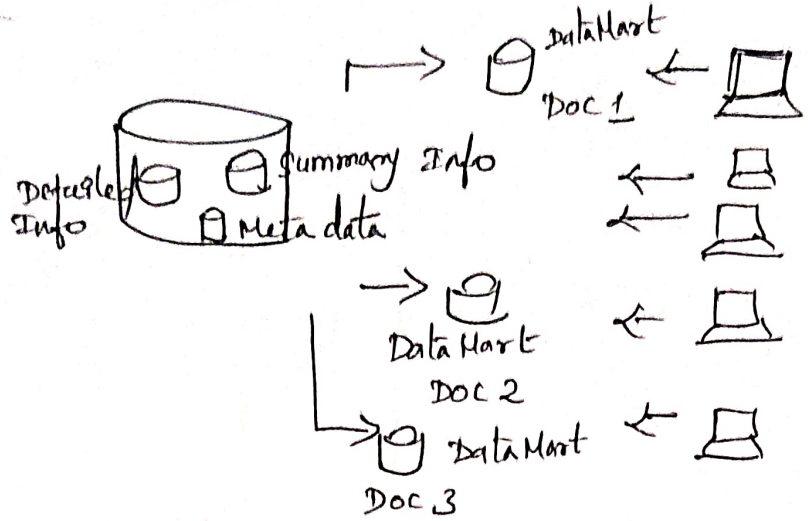
Improve user response time / access frequently / simple
Easy to implement / less expensive

TOPIC NAME:- COST EFFECTIVE DATA MART:-

Speed up queries
Suitable user access tools
Split within organization
Allow data to be loaded
data replication / cost effective
Identify user acess control

steps : Identify functional splits
Identify user access tools requirement
Identify Access control Issues.

A) Identify functional split :-
determine business structure / isolation.



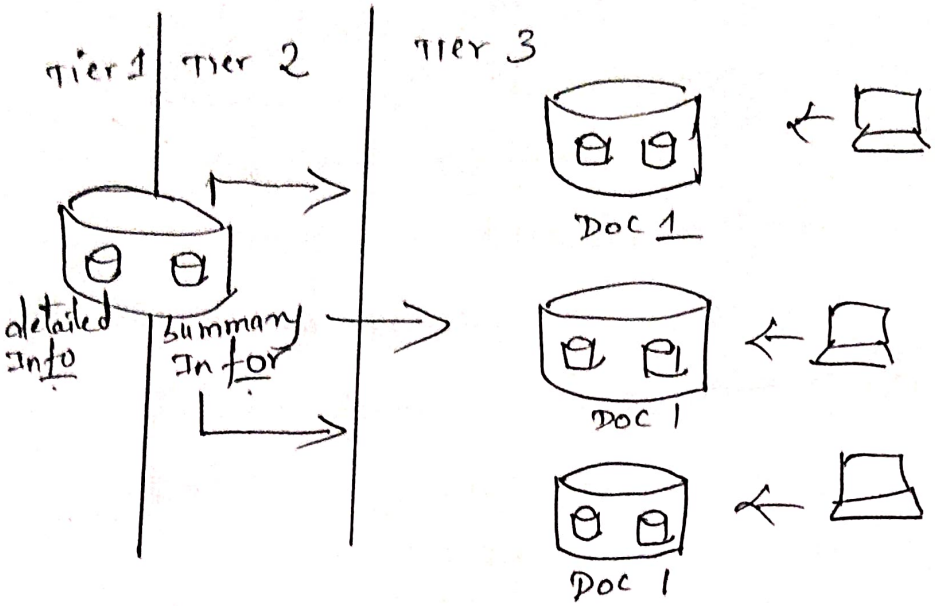
* Structure may change
* Switch form one dept to other
* analyze sales happening

B) Identify user access tool Requirements :-
Support User Access tools
Identify future
Ensure consistency of data
Each tool can have own data mart

C) Identify access control Issues :-
Ensure authorized user only
load detailed account data .

TOPIC NAME :- DESIGNING DATA MARTS :-
It should be designed as smaller version
Maintain control over database instances
utilize all dimension data in starflake schema
Identical / structured / operate same dimension of data

Different db design :-
transform one db schema to other db
less complex / effective / perform data transformation



design of data mart that create to support functional group

TOPIC NAME :- COST OF DATA MARTS :-

A) HARDWARE & SOFTWARE COST

store & manage replicated data

More expensive

B) NETWORK ACCESS :-

Capacity to handle data volumes

data mart load process

C) TIME WINDOW CONSTRAINTS :-

depend transformation of data volume

depend on nlw capacity / time / load / transformation

convert - internal structure of data mart.

TOPIC NAME:- BENEFITS OF DATA MART

Quick Access
Single source of truth
faster Insight /decision making
Simple / fast implementation
Scalable but transient analysis

TOPIC NAME:- partitioning strategy

Enhance performance/Easy mgmt of data /optimize b/w performance

Partitioning ← easy mgmt - size
Backup/Recovery - partitioning backup
Enhance performance - does not scan whole data

TOPIC NAME:- VERTICAL PARTITIONING:-

split data vertically

performed ← Normalization - Extra space used / split column as frequently access fact table
Row splitting

↓

one to one map b/w partitions
Access large table & reduce size
perform join operation b/w two partitions
data access by drill down operation.

TOPIC NAME:- HORIZONTAL PARTITIONING

Not expect each segment as same size
decide optimum solutions
Based on 1) partition by time into equal segment
2) partition by time into different size segment
3) partition on different dimension
4) partition by size of table

5) Partitioning Dimension
6) Round Robin partition

Explanation :-

1) basis of time period Eg :- Month to data value
   reuse partitioned table
   store transactions

2) Implement small partitions for current / inactive data
   detailed Information available in online
   use aggregations / reduce operating cost
   Entire history required / increase operation cost

3) statedy state basis / scan irrelevant data for query process
   entire fact table repartitioned

4) complex to manage
   partition table based on size

5) check size of dimension / affect response time

6) allow user access tool / easy to automate table
   Manage Meladata..

—x— x —x——