

# STEPS For Design and Construction Of DW

- 1. Design of a Data Warehouse: A Business Analysis Framework  
Data Warehouse Design Process**
- 2. Multi-Tiered Architecture-Three Data Warehouse Models**
- 3. Data Warehouse Back-End Tools and Utilities**
- 4. Metadata Repository**
- 5.OLAP Server Architectures**

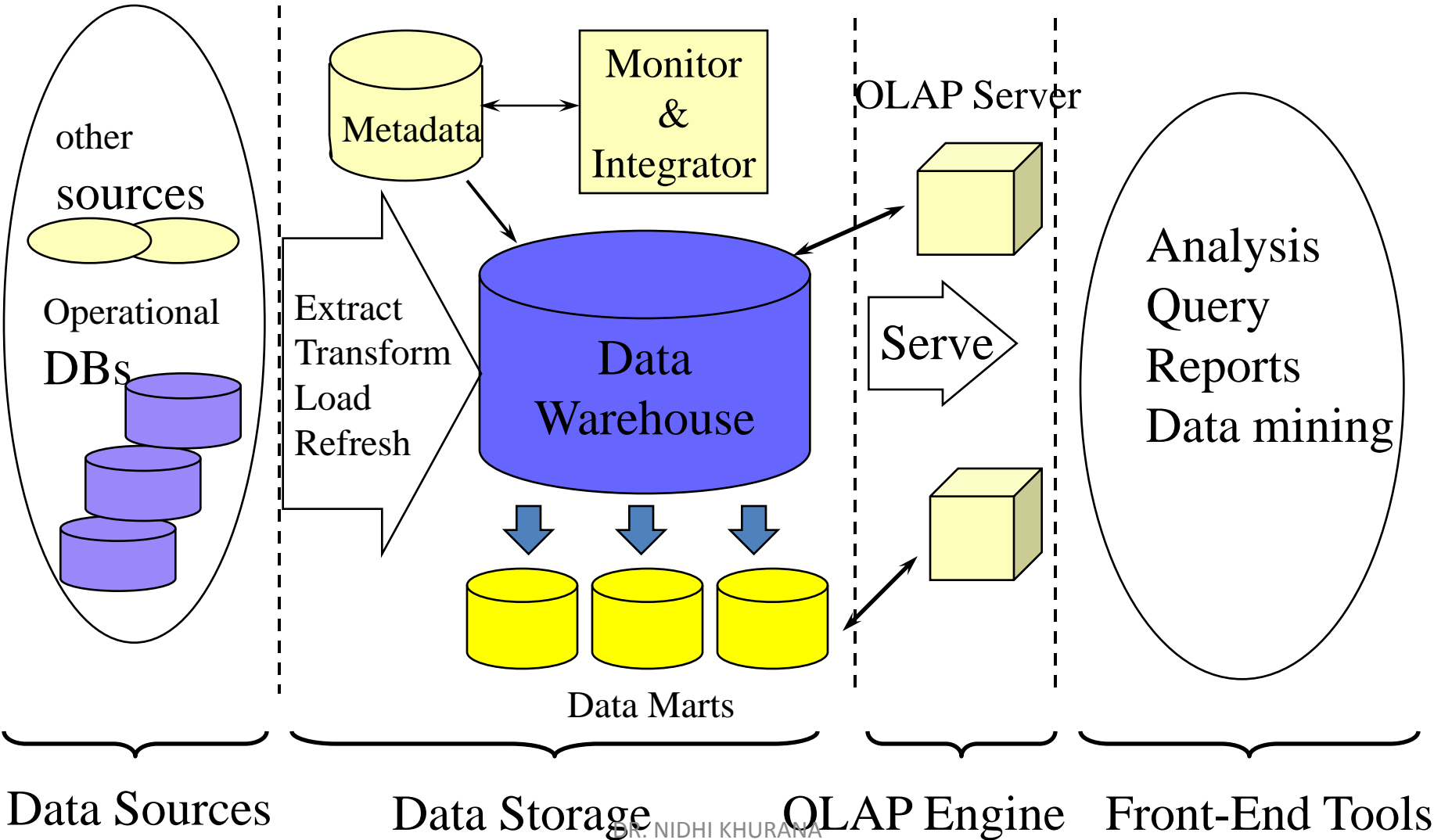
# 1.Design of a Data Warehouse: A Business Analysis Framework

- Four views regarding the design of a data warehouse
  - Top-down view
    - allows selection of the relevant information necessary for the data warehouse
  - Data source view
    - exposes the information being captured, stored, and managed by operational systems
  - Data warehouse view
    - consists of fact tables and dimension tables
  - Business query view
    - sees the perspectives of data in the warehouse from the view of end-user

# Data Warehouse Design Process

- Top-down, bottom-up approaches or a combination of both
  - Top-down: Starts with overall design and planning (mature)
  - Bottom-up: Starts with experiments and prototypes (rapid)
- From software engineering point of view
  - Waterfall: structured and systematic analysis at each step before proceeding to the next
  - Spiral: rapid generation of increasingly functional systems, short turn around time, quick turn around
- Typical data warehouse design process
  - Choose a **business process** to model, e.g., orders, invoices, etc.
  - Choose the **grain (atomic level of data)** of the business process
  - Choose the **dimensions** that will apply to each fact table record
  - Choose the **measure** that will populate each fact table record

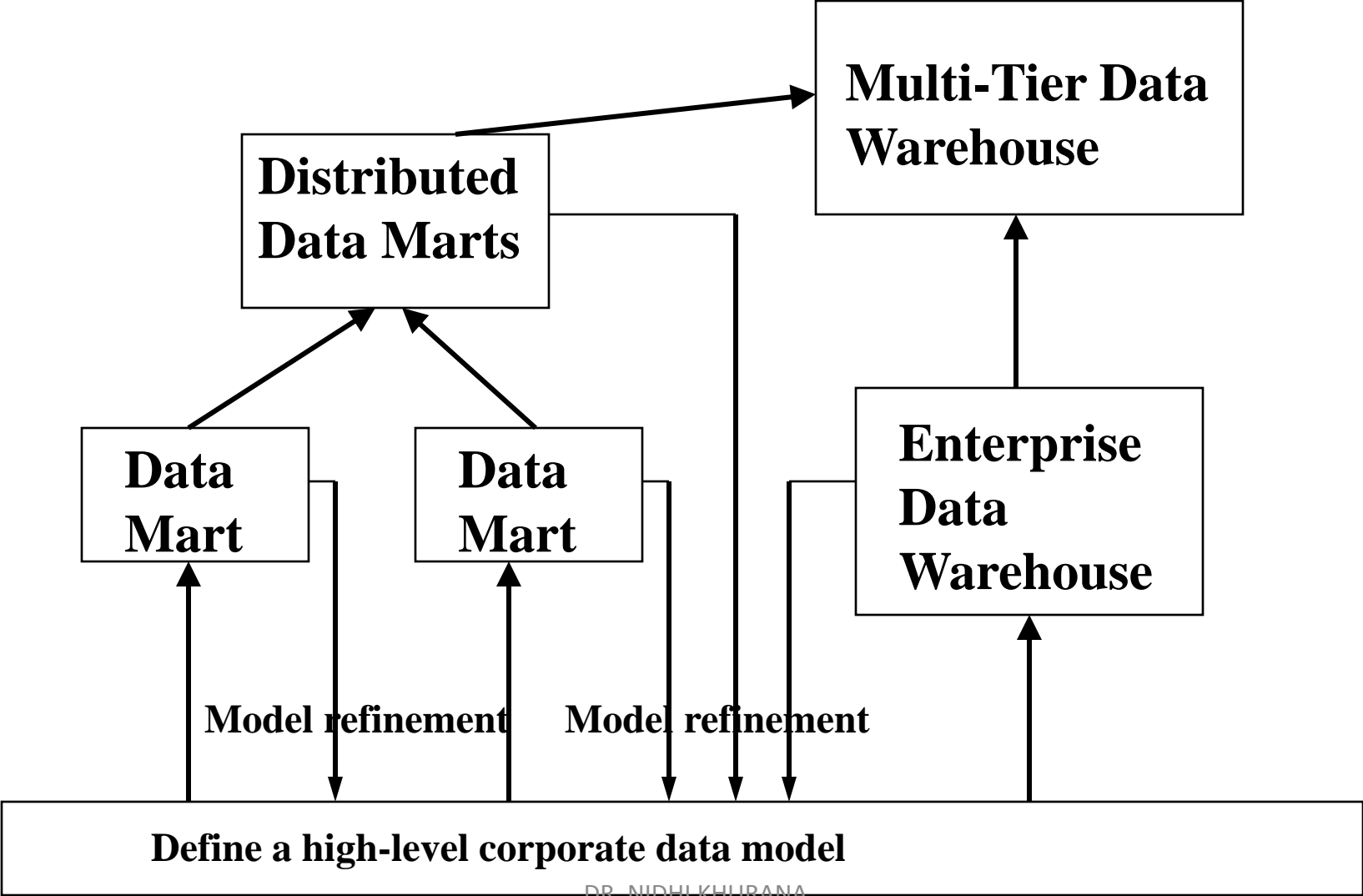
# 2. Multi-Tiered Architecture



# Three Data Warehouse Models

- Enterprise warehouse
  - collects all of the information about subjects spanning the entire organization
- Data Mart
  - a subset of corporate-wide data that is of value to a specific groups of users. Its scope is confined to specific, selected groups, such as marketing data mart
    - Independent vs. dependent (directly from warehouse) data mart
- Virtual warehouse
  - A set of views over operational databases
  - Only some of the possible summary views may be materialized

# Data Warehouse Development: A Recommended Approach



### 3. Data Warehouse Back-End Tools and Utilities

- Data extraction:
  - get data from multiple, heterogeneous, and external sources
- Data cleaning:
  - detect errors in the data and rectify them when possible
- Data transformation:
  - convert data from legacy or host format to warehouse format
- Load:
  - sort, summarize, consolidate, compute views, check integrity, and build indices and partitions
- Refresh
  - propagate the updates from the data sources to the warehouse

# 4. Metadata Repository

- Meta data is the data defining warehouse objects. It has the following kinds
  - Description of the structure of the warehouse
    - schema, view, dimensions, hierarchies, derived data defn, data mart locations and contents
  - Operational meta-data
    - data lineage (history of migrated data and transformation path), currency of data (active, archived, or purged), monitoring information (warehouse usage statistics, error reports, audit trails)
  - The algorithms used for summarization
  - The mapping from operational environment to the data warehouse
  - Data related to system performance
    - warehouse schema, view and derived data definitions
  - Business data
    - business terms and definitions, ownership of data, charging policies



# 5. OLAP Server Architectures

- [Relational OLAP \(ROLAP\)](#)
  - Use relational or extended-relational DBMS to store and manage warehouse data and OLAP middle ware to support missing pieces
  - Include optimization of DBMS backend, implementation of aggregation navigation logic, and additional tools and services
  - greater scalability
- [Multidimensional OLAP \(MOLAP\)](#)
  - Array-based multidimensional storage engine (sparse matrix techniques)
  - fast indexing to pre-computed summarized data
- [Hybrid OLAP \(HOLAP\)](#)
  - User flexibility, e.g., low level: relational, high-level: array
- [Specialized SQL servers](#)
  - specialized support for SQL queries over star/snowflake schemas

# Implementation steps in OLAP System

- **Dimensional Modeling**
- **Design and building of the MDDB.**
- **Selection of the data to be moved into the OLAP system**
- **Data acquisition or extraction for the OLAP system**
- **Data Loading into the OLAP server.**
- **Computation of data aggregation and derived data**
- **Implementation of application on the desktop**
- **Provision of user Training**