

**INDEX****UNIT I****Chapter 1 : Introduction to Data Warehousing****1-1 to 1-21**

**Syllabus :** Data warehousing , Difference between Operational Database system and Data warehouse, Need for data warehousing, A Multi-tiered Architecture of data warehousing, Data warehouse Models : Enterprise Warehouse, Data Mart, and Virtual Warehouse, Extraction, Transformation and Loading, Metadata Repository, Benefits of Data warehousing.

|         |  |      |
|---------|--|------|
| 1.1     | Data Warehousing, Difference Between Operational Database System and Data Warehouse.....                                 | 1-1  |
| 1.1.1   | Data Warehouse .....   | 1-1  |
| 1.1.2   | Features of a Data Warehouse .....   | 1-1  |
| 1.1.3   | Increasing Demand for Strategic Information .....  | 1-3  |
| 1.1.4   | Why are Operational Systems not Suitable for Providing Strategic Information?.....                                       | 1-3  |
| 1.1.5   | Differences between the Traditional Operational Systems and the Newer Informational Systems that Need to be Built... 1-3 | 1-3  |
| 1.1.6   | Operational v/s Decisional Support System.....   | 1-4  |
| 1.2     | Need For Data Warehousing .....  | 1-4  |
| 1.2.1   | Failures of Past Decision Support Systems .....  | 1-4  |
| 1.2.2   | Benefits of Data Warehousing .....   | 1-5  |
| 1.3     | Data Warehouse Architecture .....  | 1-5  |
| 1.3.1   | Different Types of Architecture.....   | 1-8  |
| 1.3.1.1 | Three Tier/ Multi-tier Data Warehouse Architecture .....   | 1-8  |
| 1.4     | Data Warehouse Models : Enterprise Warehouse, Data Mart and Virtual Warehouse.....                                       | 1-9  |
| 1.4.1   | Enterprise Warehouse .....   | 1-9  |
| 1.4.2   | Data Mart.....   | 1-9  |
| 1.4.3   | Virtual Warehouse.....   | 1-10 |
| 1.5     | Extraction, Transformation and Loading .....   | 1-10 |
| 1.5.1   | Data Extraction.....   | 1-11 |
| 1.5.1.1 | Identification of Data Sources .....   | 1-11 |
| 1.5.1.2 | Data in Operational systems .....  | 1-12 |
| 1.5.1.3 | Immediate Data Extraction.....   | 1-13 |
| 1.5.1.4 | Deferred Data Extraction.....  | 1-14 |
| 1.5.2   | Data Transformation : Tasks Involved in Data Transformation.....   | 1-15 |
| 1.5.2.1 | The Set of Basic Tasks are .....   | 1-16 |
| 1.5.2.2 | Data Integration and Consolidation.....  | 1-16 |
| 1.5.3   | Data Loading : Techniques of Data Loading.....   | 1-16 |
| 1.5.3.1 | Loading the Dimension Tables.....  | 1-17 |
| 1.5.3.2 | Loading the Fact tables: History and Incremental Loads.....  | 1-18 |
| 1.6     | Metadata Repository .....  | 1-18 |
| 1.6.1   | Role of Metadata .....   | 1-18 |
| 1.6.2   | Describe Metadata of a Book Store .....  | 1-18 |
| 1.6.3   | Classification of Metadata or Types of Metadata in Data Warehouse.....   | 1-19 |
| 1.6.4   | How can Data Warehousing Metadata be Managed ? .....   | 1-20 |

**UNIT II****Chapter 2 : Data Warehouse Modeling and Online Analytical Processing I** **2-1 to 2-39**

**Syllabus** : Data warehouse modeling : Data Cube and OLAP, Data Cube : A multidimensional Data Model. Stars, Snowflakes and fact Constellations. OLAP : Need of OLAP, OLAP Guidelines, Typical OLAP Operations

|          |   |      |
|----------|---|------|
| 2.1      | Data Warehouse Modeling .....   | 2-1  |
| 2.1.1    | What is Dimensional Modeling? .....   | 2-1  |
| 2.1.2    | Difference between Data Warehouse Modeling and Operational Database Modeling .....    | 2-1  |
| 2.1.3    | Comparison of Database and Data Warehouse .....                                       | 2-1  |
| 2.1.4    | Dimensional Model v/s ER Model .....  | 2-2  |
| 2.1.5    | Why ER is Not Suitable for Data Warehouses? .....                                     | 2-3  |
| 2.1.6    | Features of a Good Dimensional Model .....  | 2-3  |
| 2.2      | Data Cube and OLAP .....  | 2-3  |
| 2.2.1    | OLTP V/s OLAP .....   | 2-4  |
| 2.3      | Data Cube : A Multidimensional Data Model .....                                       | 2-5  |
| 2.3.1    | Data Cube or Hypercube .....  | 2-5  |
| 2.4      | The Star Schema .....   | 2-8  |
| 2.4.1    | Keys in the Data Warehouse Schema : Primary Keys, Surrogate Keys & Foreign Keys ..... | 2-9  |
| 2.5      | The Snowflake Schema .....  | 2-9  |
| 2.5.1    | Differentiate between Star Schema and Snowflake Schema .....                          | 2-10 |
| 2.5.2    | Steps of Designing a Dimensional Model .....  | 2-10 |
| 2.6      | Fact Constellation Schema or Families of Star .....                                   | 2-11 |
| 2.7      | Fact Tables and Dimension Tables .....  | 2-12 |
| 2.8      | Factless Fact Table .....   | 2-13 |
| 2.9      | How Does a Query Execute ( DMQL ) ? .....   | 2-14 |
| 2.10     | Updates to Dimension Tables .....   | 2-16 |
| 2.10.1   | Slowly Changing Dimensions .....  | 2-17 |
| 2.10.1.1 | Type 1 Changes .....  | 2-17 |
| 2.10.1.2 | Type 2 Changes .....  | 2-18 |
| 2.10.1.3 | Type 3 Changes .....  | 2-18 |
| 2.10.2   | Large Dimension Tables .....  | 2-19 |
| 2.10.3   | Rapidly Changing or Large Slowly Changing Dimensions .....                            | 2-19 |
| 2.10.4   | Junk Dimensions .....   | 2-20 |
| 2.11     | OLAP : Need of OLAP and OLAP Guidelines .....   | 2-21 |
| 2.11.1   | Need for Online Analytical Processing .....   | 2-21 |
| 2.11.2   | Typical OLAP Operations .....   | 2-21 |
| 2.11.3   | Examples of OLAP .....  | 2-25 |
| 2.12     | Examples on Star Schema and Snowflake Schema .....                                    | 2-28 |

**UNIT III****Chapter 3 : Data Warehouse Designing and Online Analytical Processing II****3-1 to 3-10**

**Syllabus :** Data warehouse design and usage, A business Analysis framework for data warehouse design, data warehouse design process, Data warehouse usage for information processing. From Online analytical processing to multi-dimensional data mining, Data Warehouse Implementation- Efficient Data Cube Computation :

An Overview, Indexing OLAP Data: Bitmap Index and Join Index, Efficient Processing of OLAP Queries, OLAP Server Architectures : ROLAP Versus MOLAP Versus HOLAP

|       |  |      |
|-------|--|------|
| 3.1   | Data Warehouse Design and Usage.....   | 3-1  |
| 3.1.1 | A Business Analysis Framework for Data Warehouse Design .....                    | 3-1  |
| 3.1.2 | Data Warehouse Design Process .....  | 3-2  |
| 3.1.3 | Data Warehouse Usage for Information Processing.....                             | 3-3  |
| 3.1.4 | From Online Analytical Processing to Multidimensional Data Mining.....           | 3-3  |
| 3.2   | Data Warehouse Implementation Efficient Data Cube Computation: An Overview ..... | 3-4  |
| 3.2.1 | Partial Materialization : Selected Computation of Cuboids.....                   | 3-5  |
| 3.3   | Indexing OLAP Data : Bitmap Index and Join Index.....                            | 3-5  |
| 3.3.1 | Bitmap Indexing.....   | 3-5  |
| 3.3.2 | Join Indexing .....  | 3-6  |
| 3.4   | Efficient Processing of OLAP Queries .....                                       | 3-7  |
| 3.5   | OLAP Server Architecture : ROLAP Versus MOLAP Versus HOLAP .....                 | 3-7  |
| 3.5.1 | MOLAP .....  | 3-8  |
| 3.5.2 | ROLAP .....  | 3-8  |
| 3.5.3 | HOLAP .....  | 3-9  |
| 3.5.4 | DOLAP .....  | 3-10 |

**UNIT IV****Chapter 4 : Introduction to Data Mining****4-1 to 4-26**

**Syllabus :** Introduction to Data Mining : Mining steps in the process of knowledge discovery of Database(KDD), What kind of data can be mined ? Major issues in data mining, Data Objects and Attribute types, Data Preprocessing : Why preprocess the data? Major tasks in Data preprocessing, Data Cleaning : Missing values, Noisy data, data cleaning as a process.

|       |  |      |
|-------|--|------|
| 4.1   | What is Data Mining? .....                         | 4-1  |
| 4.2   | Knowledge Discovery in Database (KDD) .....        | 4-2  |
| 4.2.1 | Architecture of a Typical Data Mining System ..... | 4-3  |
| 4.3   | What Kind of Data can be Mined? .....              | 4-4  |
| 4.3.1 | Data Mining Functionality .....                    | 4-5  |
| 4.4   | Major Issues in Data Mining .....                  | 4-6  |
| 4.4.1 | Applications of Data Mining.....                   | 4-7  |
| 4.5   | Data Objects and Attributes Types .....            | 4-7  |
| 4.6   | Data Processing.....                               | 4-9  |
| 4.7   | Form of Data Pre-processing .....                  | 4-9  |
| 4.7.1 | Why Pre-processing is Required?.....               | 4-9  |
| 4.7.2 | Different Forms of Data Pre-processing.....        | 4-10 |
| 4.8   | Data Cleaning : As a process.....                  | 4-10 |



|          |  |      |
|----------|--|------|
| 4.8.1    | Reasons for “Dirty” Data .....                     | 4-10 |
| 4.8.2    | Why Data Cleaning or Cleansing is Required?.....   | 4-10 |
| 4.8.3    | Missing Values .....                               | 4-11 |
| 4.8.4    | Noisy Data.....                                    | 4-12 |
| 4.8.4.1  | How to Handle Noisy Data? .....                    | 4-13 |
| 4.8.5    | Inconsistent Data.....                             | 4-15 |
| 4.9      | Data Integration.....                              | 4-16 |
| 4.9.1    | Introduction to Data Integration.....              | 4-16 |
| 4.9.1.1  | Entity Identification Problem.....                 | 4-16 |
| 4.9.1.2  | Redundancy and Correlation Analysis .....          | 4-16 |
| 4.9.1.3  | Tuple Duplication.....                             | 4-17 |
| 4.9.1.4  | Data Value Conflict Detection and Resolution ..... | 4-18 |
| 4.10     | Data Reduction.....                                | 4-18 |
| 4.10.1   | Need for Data Reduction.....                       | 4-18 |
| 4.10.2   | Data Reduction Technique.....                      | 4-19 |
| 4.10.2.1 | Data Cube Aggregation.....                         | 4-19 |
| 4.10.2.2 | Dimensionality Reduction.....                      | 4-19 |
| 4.10.2.3 | Data Compression.....                              | 4-20 |
| 4.10.2.4 | Numerosity Reduction.....                          | 4-21 |
| 4.11     | Data Transformation and Data Discretization .....  | 4-23 |
| 4.11.1   | Data Transformation.....                           | 4-23 |
| 4.11.2   | Data Discretization .....                          | 4-24 |
| 4.11.3   | Data Transformation by Normalization.....          | 4-24 |
| 4.11.4   | Discretization by Binning.....                     | 4-25 |
| 4.11.5   | Discretization by Histogram Analysis .....         | 4-25 |
| 4.11.6   | Concept Hierarchies.....                           | 4-25 |

UNIT V

---

**Chapter 5 : Mining Frequent Patterns and Cluster Analysis**

5-1 to 5-53

**Syllabus** : Mining frequent patterns : Basic concepts : Market basket analysis, frequent itemsets, closed itemsets and association rules  
 Frequent itemsets mining methods : The apriori algorithm, finding frequent itemsets using candidate generation  
 Generating association rules from frequent itemsets, What is cluster analysis? Requirements for cluster analysis, Overview of Basic clustering methods, General applications of clustering.

|         |   |     |
|---------|---|-----|
| 5.1     | Market Basket Analysis.....                                   | 5-1 |
| 5.1.1   | What is Market Basket Analysis?.....                          | 5-1 |
| 5.1.2   | How is it used? .....   | 5-1 |
| 5.1.3   | Applications of Market Basket Analysis .....                  | 5-2 |
| 5.2     | Frequent Itemsets, Closed Itemsets and Association Rules..... | 5-2 |
| 5.2.1   | Frequent Itemsets .....                                       | 5-2 |
| 5.2.2   | Closed Itemsets.....  | 5-2 |
| 5.2.3   | Association Rules.....  | 5-3 |
| 5.2.3.1 | Large Itemsets.....   | 5-4 |
| 5.3     | Efficient and Scalable Frequent Itemset Mining Methods.....   | 5-4 |



---

|       |  |      |
|-------|--|------|
| 5.3.1 | Apriori Algorithm for Finding Frequent Itemsets using Candidate Generation ..... | 5-4  |
| 5.3.2 | Advantages and Disadvantages of Apriori Algorithm.....                           | 5-6  |
| 5.3.3 | Solved Examples on Apriori Algorithm.....  | 5-6  |
| 5.3.4 | Improving the Efficiency of Apriori.....   | 5-22 |
| 5.4   | Pattern Growth Approach for Mining Frequent Itemsets (FP-Growth).....            | 5-23 |
| 5.4.1 | Definition of FP-Tree .....  | 5-23 |
| 5.4.2 | FP-Tree Algorithm .....  | 5-23 |
| 5.4.3 | FP-Tree Size .....   | 5-24 |
| 5.4.4 | Examples on FP Tree.....   | 5-24 |
| 5.4.5 | Mining Frequent Patterns from FP Tree.....                                       | 5-28 |
| 5.4.6 | Benefits of the FP-Tree Structure .....  | 5-33 |
| 5.5   | What is Clustering? .....  | 5-33 |
| 5.5.1 | Requirements for Cluster Analysis.....   | 5-34 |
| 5.5.2 | Overview of Basic Clustering Method .....  | 5-34 |
| 5.5.3 | General Applications of Clustering.....  | 5-35 |
| 5.5.4 | Distance Measures.....   | 5-35 |
| 5.6   | Partitioning Methods of Clustering .....   | 5-36 |
| 5.6.1 | K-means Clustering : (Centroid Based Technique) .....                            | 5-36 |
| 5.6.2 | K-Medoids (Representative Object-based Technique) .....                          | 5-49 |

---

