

**Chapter 1 : Data Warehousing Fundamentals****1-1 to 1-78**

1.1	Introduction to Data Warehouse.....	1-1	1.3.1(E)	A Practical Approach.....	1-15
1.1.1	Need for Strategic Information.....	1-1	1.4	Metadata.....	1-16
1.1.2	Desired Characteristics of Strategic Information.....	1-2	1.4.1	Definition.....	1-16
1.1.3	Operational v/s Decisional Support System.....	1-2	1.4.2	Describe Metadata of a Book Store.....	1-16
1.1.4	Definition of Data Warehouse.....	1-3	1.4.3	Data Warehouse Metadata.....	1-17
1.1.5	Benefits of Data Warehousing.....	1-3	1.4.4	Classification of Metadata or Types of Metadata in Data Warehouse.....	1-18
1.1.6	Features of a Data Warehouse.....	1-3	1.5	E-R Modelling versus Dimensional Modelling.....	1-19
1.1.7	Relationship between Data Warehousing and Data Replication.....	1-5	1.5.1	What is Dimensional Modeling ?.....	1-19
1.2	Data Warehouse Architecture.....	1-5	1.5.2	Difference between Data Warehouse Modeling and Operational Database Modeling.....	1-19
1.2.1	The Information Flow Mechanism.....	1-5	1.5.3	Comparison Database and Data Warehouse Database.....	1-19
1.2.2	Architecture of typical Data warehouse.....	1-7	1.5.4	Comparison between Dimensional Model and ER model.....	1-20
1.2.3	Three Tier/ Multi-tier Data Warehouse Architecture.....	1-10	1.6	Information Package Diagram.....	1-20
1.3	Data Warehouses versus Data Marts.....	1-11	1.7	Data Warehouse Schemas : Star Schema	1-21
1.3.1	Data Warehousing Design Strategies or Approaches for Building a Data Warehouse.....	1-12	1.7.1	STAR schema Keys.....	1-23
1.3.1(A)	The Top Down Approach : The Dependent Data Mart Structure.....	1-12	1.8	The Snowflake Schema.....	1-23
1.3.1(B)	The Bottom-Up Approach : The Data Warehouse Bus Structure.....	1-13	1.8.1	Differentiate between Star Schema and Snowflake Schema.....	1-25
1.3.1(C)	Hybrid Approach.....	1-14	1.9	Factless Fact Tables.....	1-25
1.3.1(D)	Federated Approach.....	1-15	1.9.1	Fact Tables and Dimension Tables.....	1-25
			1.9.2	Factless Fact Table.....	1-26
			1.10	Fact Constellation Schema or Families of Star.....	1-27



1.11	Steps of Designing a Dimensional Model.....	1-29	1.15.9(B)	Loading the Fact tables: History and Incremental Loads.....	1-59
1.12	Update to the Dimension Tables.....	1-29	1.15.10	Data Quality : Issues in Data Cleansing....	1-59
1.12.1	Slowly Changing Dimensions .....	1-30	1.15.10(A)	Reasons for “Dirty” Data .....	1-59
1.12.2	Large Dimension Tables .....	1-32	1.15.10(B)	Data Cleansing.....	1-60
1.12.3	Rapidly Changing or Large Slowly Changing Dimensions.....	1-32	1.15.11	Sample ETL Tools.....	1-61
1.12.4	Junk Dimensions .....	1-33	1.16	OLTP versus OLAP.....	1-62
1.13	Aggregate Fact Tables .....	1-34	1.16.1	Hypercube .....	1-63
1.14	Examples on Star Schema and Snowflake Schema .....	1-34	1.17	OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot .....	1-65
1.15	Major steps in ETL process .....	1-51	1.18	OLAP Models: MOLAP, ROLAP, HOLAP,DOLAP .....	1-69
1.15.1	What is ETL Tool?.....	1-51	1.18.1	MOLAP .....	1-69
1.15.2	Desired Features.....	1-51	1.18.2	ROLAP .....	1-70
1.15.3	Major Steps in ETL Process.....	1-52	1.18.3	HOLAP .....	1-71
1.15.4	Data Extraction.....	1-52	1.18.4	DOLAP .....	1-71
1.15.5	Identification of Data Sources.....	1-52	1.19	Examples of OLAP .....	1-71
1.15.6	Data in Operational Systems .....	1-53	<b>Chapter 2 : Introduction to Data Mining, Data Exploration and Data Pre-processing</b>		
1.15.6(A)	Immediate Data Extraction .....	1-54	<b>2-1 to 2-54</b>		
1.15.6(B)	Deferred Data Extraction .....	1-55	2.1	Data Mining Task Primitives .....	2-1
1.15.7	Data Transformation : Tasks Involved in Data Transformation .....	1-56	2.1.1	What is Data Mining ?.....	2-1
1.15.7(A)	The Set of Basic Tasks .....	1-57	2.1.2	Data Mining Primitives .....	2-2
1.15.8	Data Integration and Consolidation .....	1-58	2.2	Architecture.....	2-4
1.15.9	Data Loading: Techniques of Data Loading.....	1-58	2.3	KDD (Knowledge Discovery in Database).2-5	
1.15.9(A)	Loading the Dimension Tables .....	1-59	2.4	Major Issues in Data Mining .....	2-6
			2.5	Applications of Data Mining .....	2-7



2.6	Data Exploration : Types of Attributes.....2-7	2.12.2(B)	Dimensionality Reduction..... 2-40
2.7	Statistical Description of Data.....2-10	2.12.2(C)	Data Compression..... 2-41
2.7.1	Central Tendency.....2-10	2.12.2(D)	Numerosity Reduction..... 2-43
2.7.2	Dispersion of Data.....2-12	2.13	Data Transformation and Data Discretization ..... 2-44
2.7.3	Graphic Displays of Basic Statistical Descriptions of Data .....2-15	2.13.1	Data Transformation ..... 2-44
2.8	Data Visualization .....2-19	2.13.2	Data Discretization ..... 2-45
2.9	Data Preprocessing : Descriptive Data Summarization .....2-25	2.13.3	Data Transformation by Normalization.. 2-45
2.9.1	Form of Data Pre-processing.....2-25	2.13.4	Discretization by Binning ..... 2-49
2.10	Data Cleaning .....2-26	2.13.5	Discretization by Histogram Analysis..... 2-49
2.10.1	Reasons for “Dirty” Data.....2-26	2.14	Concept Hierarchy Generation..... 2-49
2.10.2	Steps in Data Cleansing.....2-26	2.15	Concept Description : Attribute Oriented Induction for Data Characterization ..... 2-50
2.10.3	Missing Values .....2-27	2.16	Data Generalization and Summarization- based Characterization..... 2-51
2.10.4	Noisy Data.....2-28	2.16.1	Data Generalization..... 2-51
2.10.5	Inconsistent Data.....2-35	2.16.2	How Attribute-Oriented Induction is Performed?..... 2-52
2.11	Data Integration .....2-35	2.16.2(A)	Data Generalization..... 2-52
2.11.1	Entity Identification Problem.....2-35	2.16.2(B)	Attribute Generalization Control..... 2-52
2.11.2	Redundancy and Correlation Analysis....2-35	2.16.2(C)	Example of Attribute Oriented Induction ..... 2-53
2.11.3	Tuple Duplication.....2-38	<hr/>	
2.11.4	Data Value Conflict Detection and Resolution.....2-39	<b>Chapter 3 : Classification 3-1 to 3-64</b>	
2.12	Data Reduction .....2-39	3.1	Basic Concept : Classification..... 3-1
2.12.1	Need for Data Reduction .....2-39	3.1.1	Classification Problem .....3-1
2.12.2	Data Reduction Technique .....2-40	3.1.2	Classification Example..... 3-2
2.12.2(A)	Data Cube Aggregation.....2-40	3.1.3	Classification is a Two Step Process .....3-2



3.1.4	Difference between Classification and Prediction.....3-4	<b>Chapter 4 : Clustering</b>	<b>4-1 to 4-62</b>
3.2	Decision Tree Induction.....3-4	4.1	Basics of Clustering..... 4-1
3.2.1	Appropriate Problems for Decision Tree Learning.....3-4	4.1.1	What is Clustering ?..... 4-1
3.2.2	Decision Tree Representation.....3-5	4.1.2	Categories of Clustering Methods ..... 4-2
3.2.3	Attribute Selection Measure .....3-5	4.1.3	Difference between Classification and Clustering..... 4-3
3.2.4	Algorithm for Inducing a Decision Tree.....3-8	4.2	Types of Data in Cluster analysis ..... 4-3
3.2.5	Tree Pruning.....3-10	4.2.1	Interval-Scaled Variables..... 4-4
3.2.6	Examples of ID3 .....3-11	4.2.2	Binary Variable ..... 4-5
3.3	Naïve Bayesian Classification.....3-46	4.2.3	Nominal, Ordinal, and Ratio Variables..... 4-6
3.3.1	Bayes Theorem.....3-46	4.2.4	Variable of Mixed Types..... 4-8
3.3.1(A)	Basics of Bayesian Classification .....3-46	4.3	Distance Measures..... 4-8
3.3.2	Naive Bayes Classifier : Example .....3-47	4.4	Partitioning Methods (K-Means, K-Medoids) ..... 4-9
3.3.3	Other Classification Methods .....3-60	4.4.1	K-means Clustering : (Centroid based Technique)..... 4-9
3.4	Accuracy and Error measures.....3-60	4.4.2	Examples of K-means ..... 4-11
3.5	Evaluating the Accuracy of a Classifier : Holdout & Random Subsampling, Cross Validation, Bootstrap..... 3-62	4.4.3	Strength and Weakness od K-means ..... 4-24
3.5.1	Holdout ..... 3-62	4.4.4	K-Medoids (Representative Object-based Technique) ..... 4-24
3.5.2	Random Subsampling.....3-63	4.4.5	Example of K-Medoids ..... 4-26
3.5.3	Cross-Validation (CV).....3-63	4.4.6	Sampling Based Method..... 4-29
3.5.4	Bootstrapping .....3-64		



<p>4.5 Hierarchical Methods (Agglomerative, Divisive)..... 4-29</p> <p>4.5.1 Agglomerative Hierarchical Clustering.... 4-31</p> <p>4.5.2 Examples of Agglomerative Clustering.... 4-32</p> <p>4.5.3 Comparison of the Three Methods (Based on Distance Formula)..... 4-60</p> <p>4.5.4 Agglomerative Algorithm given by Margaret H. Dunham..... 4-61</p> <p>4.5.5 Divisive Hierarchical Clustering ..... 4-61</p> <p>4.5.6 Difference between Agglomerative and Divisive..... 4-62</p> <p>4.5.7 Advantages and Disadvantages of Hierarchical Clustering ..... 4-62</p> <hr/> <p><b>Chapter 5 : Mining Frequent Patterns and Associations</b> <b>5-1 to 5-56</b></p> <hr/> <p>5.1 Market Basket Analysis ..... 5-1</p> <p>5.1.1 What is Market Basket Analysis?..... 5-1</p> <p>5.1.2 How is it Used ? ..... 5-1</p> <p>5.1.3 Applications of Market Basket Analysis ..... 5-2</p> <p>5.2 Frequent Item Sets, Closed Item Sets and Association Rule..... 5-2</p> <p>5.2.1 Frequent Itemsets ..... 5-2</p> <p>5.2.2 Closed Itemsets ..... 5-3</p> <p>5.2.3 Association Rules..... 5-3</p>	<p>5.3 Frequent Pattern Mining..... 5-5</p> <p>5.4 Apriori Algorithm..... 5-5</p> <p>5.4.1 Apriori Algorithm given by Jiawei Han et al..... 5-5</p> <p>5.4.2 Advantages and Disadvantages of Apriori Algorithm..... 5-7</p> <p>5.4.3 Solved Examples on Apriori Algorithm ..... 5-7</p> <p>5.5 Association Rule Generation ..... 5-37</p> <p>5.6 Improving the Efficiency of Apriori..... 5-37</p> <p>5.7 Mining Frequent Itemsets without Candidate Generation : FP Growth ..... 5-38</p> <p>5.7.1 Definition of FP-tree..... 5-38</p> <p>5.7.2 FP-Tree Algorithm ..... 5-38</p> <p>5.7.3 FP-Tree Size..... 5-39</p> <p>5.7.4 Example of FP Tree..... 5-40</p> <p>5.7.5 Mining Frequent Patterns from FP Tree..... 5-43</p> <p>5.7.6 Benefits of the FP-Tree Structure..... 5-50</p> <p>5.8 Mining Frequent Itemsets using Vertical Data Formats ..... 5-50</p> <p>5.9 Introduction to Mining Multilevel Association Rules ..... 5-52</p> <p>5.10 Mining Multidimensional (MD) Association Rules ..... 5-53</p>
---	--

<b>Chapter 6 : Web Mining</b>		<b>6-1 to 6-16</b>	
6.1	Introduction to Web Mining .....	6-1	
6.1.1	How Web Mining is Different from Classical DM ?.....	6-1	
6.1.2	Benefits of Web Data Mining .....	6-2	
6.2	Web Content Mining.....	6-2	
6.2.1	Introduction to Web Content Mining.....	6-2	
6.2.2	Web Crawlers .....	6-2	
6.2.3	Harvest System.....	6-3	
6.2.4	Virtual Web View.....	6-4	
6.2.5	Personalization.....	6-4	
6.3	Web Structure Mining.....	6-4	
6.3.1	Introduction to Web Structure Mining.....	6-4	
6.3.2	Techniques of Web Structure Mining.....	6-5	
6.3.2(A)	PageRank .....	6-5	
6.3.2(B)	CLEVER Technique.....	6-8	
6.4	Web Usage Mining .....	6-11	
6.4.1	What is Web Usage Mining ? .....	6-11	
6.4.2	Purpose of Web Usage Mining.....	6-12	
6.4.3	Web Usage Mining Activities.....	6-12	
6.4.4	Web Server Log .....	6-13	
6.4.4(A)	Structure of Web Log.....	6-13	
6.4.4(B)	Web Server Log - An Example .....	6-13	
6.5	Applications of Web Mining.....	6-16	