Lectures for the course: Data Warehousing and Data Mining (IT 60107)

Week 1

Lecture 1 – 21/07/2005

• Introduction to the course
• Expectations
• Evaluation Guideline
• Term Paper and Term Project Guideline

Week 2

Lecture 2 – 25/07/2005

• What is a Data Warehouse
• Why is a Data Warehouse required
• Data Warehouse as a separate database

Lecture 3 (A+B) – 27/07/2005

• Difference between Data Warehouse and OLTP
• Data Warehouse Architecture
• Data Marts

Lecture 4 – 28/07/2005

• Multidimensional View of Data
• OLAP operations
• Slicing

Week 3

Lecture 5 – 01/08/2005

• Dicing operations
• Roll-up and Drill-down
• Data cube as a lattice of cuboids

Lecture 6 (A+B) – 03/08/2005

• No. of cuboids in a data cube
• Drill across
• Simple Relational Implementation of data repository with single level of each dimension
• Multidimensional Implementation of data repository
• ROLAP, MOLAP and HOLAP
• Dimension hierarchy
• Schema hierarchy and set grouping hierarchy
• Roll up and Drill down using dimension hierarchies
• Simple relational implementation of data repository with multiple levels of each dimension

Lecture 7– 04/08/2005
• Introduction to the view materialization problem

Week 4

Lecture 8 – 08/08/2005
• A Greedy algorithm for deciding which set of views to materialize

Lecture 9 (A+B) – 10/08/2005
• Multiway Array Aggregation
• Chunking
• Order of visiting chunks and its impact on the amount of chunk memory required in MOLAP sub-cube computation
• Recap of the topics covered so far

Lecture 10– 11/08/2005
• Data Warehouse Schema Design
• Motivation for De-normalization

Week 5

Lecture 11 (A+B) – 17/08/2005
• Star Schema Design
• Retail Sales Schema
• Date, Product and Location Dimensions
• Fact Tables and Dimension Tables
• De-normalization of dimension tables
• Normalized fact tables
• Size estimation of fact and dimension tables
Week 6

Lecture 12 – 22/08/2005

• Four steps in Data Warehouse schema design
• Impact of granularity on schema
• Dimensions that may or may not change the granularity of data

Lecture 13 (A+B) – 24/08/2005

• Promotion Dimension
• Fact less Fact Table
• Degenerate dimension
• Star Constellation schema
• Snowflake schema

Lecture 14 – 25/08/2005

• Extensibility of Star Schema
• Effect of change of dimension attributes
• Effect of adding new dimensions
• Effect of adding new facts

Week 7

Lecture 15 – 29/08/2005

• Old topics revisited

Lecture 16 (A+B) – 31/08/2005

• Class Test 1

Lecture 17 – 01/09/2005

• Inventory Business Process
• Periodic snapshot schema

Week 8

Lecture 18 – 05/09/2005

• Class Test 1 scripts shown
Lecture 19 (A+B) – 07/09/2005

- Inventory Transactions Schema
- Rapidly Changing Dimensions

Lecture 20 – 08/09/2005

- Rapidly Changing Dimensions
- Dimension Outriggers
- Mini Dimensions

Week 9

Lecture 21 – 12/09/2005

- Rapidly Changing Dimensions revisited
- Bitmap indexing

Lecture 22 (A+B) – 14/09/2005

- Join Indexing
- Recap of topics on Data Warehousing

15/09/2005

- Classes not held due to mid-sem exam.

Week 10

Mid-sem exam was held here

Week 11

Lecture 23 – 26/09/2005

- Mid-sem scripts were shown
- Introduction to Data Mining
- Itemsets
- Support

Lecture 24 (A+B) – 28/09/2005

- Frequent Itemsets
- Association Rule Mining
- A priori algorithm
• Confidence

Lecture 25—29/09/2005
• Partitioning Approach to Association Rule Mining

Week 12

Lecture 26—03/10/2005
• Dynamic Itemset Counting

Lecture 27 (A+B)—05/10/2005
• Dynamic Itemset Counting (contd.)
• Association Rule mining without candidate generation
• FP-Tree Construction and Mining

Lecture 28—06/09/2005
• FP-Tree Construction and Mining (contd.)

Week 13

Autumn Break

Week 14

17/10/2005
• No Class—compensated later

Lecture 29 (A+B)—19/10/2005
• Sequential Pattern Mining

Lecture 30—20/10/2005
• Sequential Pattern Mining

Lecture 31 (A+B)—21/10/2005 (Compensatory Lecture)
• Sequential Pattern Mining

Week 15
Lecture 32 – 24/10/2005

- Clustering Problem Definition
- Partitioning and Hierarchical Approaches
- K-Means Clustering

Lecture 33 (A+B) – 26/10/2005

- K Medoid Algorithm - PAM
- CLARA
- CLARANS

27/10/2005

- No Class

Week 16

Lecture 34 – 31/10/2005

- Hierarchical Clustering – Agglomerative and Divisive
- Dendrogram

Week 17

Lecture 35 – 07/11/2005

- Agglomerative Hierarchical Clustering Algorithm
- CF Vector

Lecture 36 (A+B) – 09/11/2005

- BIRCH
- Introduction to Classification
- Confusion Matrix

Lecture 37 – 10/11/2005

- Classification using Multilayer Perceptron
- Back Propagation Algorithm

Week 18
Lecture 38 – 14/11/2005

- Decision Trees
- Information Gain

Lecture 39 (A+B) – 16/11/2005

- Algorithm for building Decision Trees

Lecture 40 – 17/11/2005

- Term Projects Demonstrated