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

Week 1

Lecture 1 – 29/07/2004
- Introduction to the course
- Expectations
- Evaluation Guideline
- Term Paper and Term Project Guideline

Week 2

Lecture 2 – 02/08/2004
- What is a data warehouse?
- Why a data warehouse is required
- Difference between a Data Warehouse and an OLTP database

Lecture 3 (a) + (b) – 04/08/2004
- Components of a Data Warehouse
- Data Warehouse architecture
- OLAP Server
- ROLAP, MOLAP and HOLAP
- Multidimensional Data Model
- Data Cube – 2D, 3D, 4D
- OLAP operations – Slicing, Dicing, Roll-up and Drill-down

Lecture 4 – 05/08/2004
- Data Cube as a lattice of cuboids
- Dimension hierarchy – Schema hierarchy and Set-grouping hierarchy
- Total order and partial order in the hierarchy

Week 3

Lecture 5 – 09/08/2004
- View Materialization
- Class Test Date was announced – 01/09/2004 from 08:00 AM – 09:00 AM.
- Why view materialization is required.
- Full, None and Partial materialization
• Data Cube structure for dimensions with multiple hierarchy levels

Lecture 6 (a) + (b) – 11/08/2004
• View Materialization problem – A Greedy Algorithm proposed by Harinarayan et al.

Lecture 7 – 12/08/2004
• Sub-cube computation in MOLAP - Chunking
• Order of visiting a 3D cube for computing 2D cube and its effect on memory requirement

Week 4

Lecture 8 – 16/08/2004
• Recap of topics covered so far
• ERD and normalized table design
• Motivation for de-normalization
• Lecture on 26th to be compensated by a lecture on the 27th from 5:30 PM – 6:30 PM

18/08/2004 – Classes Off (Foundation Day)

Lecture 9 – 19/08/2004
• Data Warehouse Table Design
• Dimensional Model
• Star Schema
• Fact and Dimension Tables

Week 5

Lecture 10 – 23/08/2004
• Snowflake Schema
• De-Normalization Advantages and Disadvantages
• Retail Sales Schema
• Date, Product and Location Dimensions
• Impact of Changing the Granularity of Fact Table
• Degenerate Dimension
• Multidimensional Cube Generation from the Dimensional Model
Lecture 11 (a) + (b) – 25/08/2004

- Steps in Data Warehouse Design
- Promotion Dimension
- Fact less Fact Table
- Data Mart
- Sharing of Dimension Tables
- Fact Constellation Schema
- Additive, Semi-Additive, Non-Additive Facts


- Additivity of Facts revisited
- Different Types of Measures – Distributive, Algebraic, Holistic
- Effect of Changes in Dimension table and Fact table attributes on Star schema

Week 6

Lecture 13 – 30/08/2004

- Inventory Business Process Data Warehouse Design
- Periodic Snapshot Schema
- Inventory Transactions
- Accumulating Snapshot Schema

Lecture 14 (a) + (b) – 01/09/2004

- C. Test 1 was held here

Lecture 15 – 02/09/2004

- C. Test Question Paper discussed and scripts shown
- Multi-way array aggregation revisited

Week 7

Lecture 16 – 06/09/2004

- Data Marts
- Conformed Dimensions

Lecture 17 (a) + (b) – 08/09/2004

- Effect of change in dimension tables
• Slowly Changing Dimensions
• Rapidly changing dimensions

Lecture 18 – 09/09/2004

• Slowly and Rapidly Changing Dimensions revisited
• Indexing of data warehouse tables
• Join Indexes

Week 8

Lecture 19 – 13/09/2004

• Bitmapped Indexes
• Summary of lectures on Data Warehousing

15/09/2004

• Classes off due to Mid-sem exams

16/09/2004 – 24/09/2004 – Mid Sem Exams were held here
27/09/2004 -- 1/10/2004 – Classes not held as the faculty was out of station

Week 9

Lecture 20 – 04/10/2004

• Mid-sem scripts were shown
• Introduction to Data Mining
• KDD and Data Mining
• SQL and Data Mining
• Items and Itemsets
• Association Rule
• Support and Confidence
• Frequent Itemsets

Lecture 21 – 05/10/2004 (Compensatory Lecture)

• A-priori Algorithm for Association Rule Mining

Lecture 22 (a) + (b) – 06/10/2004

• Partitioning algorithm and Dynamic Itemset Counting Algorithm
Lecture 23 – 07/10/2004
  • FP tree – Creation of FP Tree

Week 10
Lecture 24 – 11/10/2004
  • Mining FP Tree

Lecture 25 – 12/10/2004 (Compensatory Lecture)
  • Introduction to Clustering
  • Partitioning Techniques and Hierarchical Techniques
  • Dendrogram
  • K-Means Clustering

Lecture 26 (a) + (b) – 13/10/2004
  • K-Medoid Clustering
  • PAM
  • CLARA

Lecture 27 – 14/10/2004
  • CLARANS

Puja Vacation

Week 11
Lecture 28 – 25/10/2004
  • Agglomerative Hierarchical Clustering (Algorithm available at the end of the “Related Papers” Page)

Lecture 29 – 28/10/2004
  • CF Vectors
  • BIRCH

Week 12
Lecture 30 – 01/11/2004

- Introduction to Classification
- Confusion Matrix
- Decision Tree

Lecture 31 (a) + (b) – 03/11/2004

- Decision Tree – Continued
- ID3

Lecture 32 – 04/11/2004

- Classification using MLP
- Back Propagation Algorithm

Week 13

Lecture 33 – 08/11/2004

- Complex Data Mining Examples – Spatial Data Mining, Multimedia Data Mining, Time Series Analysis, Text and Web Mining
- Text Mining
  - Recall and Precision
  - Term Document Matrix
  - Term Frequency and Relative Term Frequency

Lecture 34 (a) + (b) – 10/11/2004

- Latent Semantic Indexing and SVD
- Web mining
- Page Rank
- HITS


- Summary and Feedback

Week 14

Lecture 36 – 13/11/2004

- Term Project Demo