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

Week 1

Lecture 1 – 21/07/2011

- Introduction to the course
- Pre-requisite
- Expectations
- Evaluation Guideline
- Term Paper and Term Project Guideline

Week 2

Lecture 2 – 25/07/2011

- What is a Data warehouse
- Properties of a Data warehouse

Lecture 3 – 28/07/2011

- Data Warehousing defined
- Data warehouse architecture
- Difference between OTLP and OLAP
- Need for a separate database

Week 3

Lecture 4 – 01/08/2011

- Multi-dimensional data model
- Data cube as a lattice of cuboids
- Number of cuboids in a data cube
- Dimension hierarchy

Lecture 5+6 – 03/08/2011

- OLAP operations
- Slicing, Dicing, Roll-up and Drill-down
- Roll-up and Drill-down in the presence of dimension hierarchy
- Term Paper team formation
- Term paper preparation explained

- Getting data from relational database tables for generating multi-dimensional view

Lecture 7 – 04/08/2011

- Roll up and drill down using relational database tables
- Tables for representing dimension hierarchy
- Materialized views.

Week 4

Lecture 8 – 08/08/2011

- Formal definition of data cube as a lattice
- Correspondence between data cube, multi-dimensional view and group by queries
- ROLAP, MOLAP and HOLAP servers

Lecture 9+10 – 10/08/2011

- Lattice in the presence of dimension hierarchy
- Greedy algorithm for view materialization
- Effect of greedy choice on optimality

Lecture 11 – 11/08/2011

- Multi-way array aggregation

Week 5

Lecture 12+13 – 24/08/2011

- Class Test held

Lecture 14 – 25/08/2011

- Dimensional modeling introduction
- Star schema
- Basic retail sales fact and dimension
- Snowflake schema
- Steps in dimensional modeling

Lecture 15 – 26/08/2011

- Retail sales business process
- Fact constellation schema

- Granularity at the level of POS transactions
- Promotion dimension
- Degenerate dimension
- Sample SQLs for getting information out of star schema

Lecture 16 – 27/08/2011

- Factless fact table – promotion coverage
- Additivity of facts – Additive, semi-additive and non-additive facts
- Discussion on date dimension
- Extensibility of star schema – addition of new dimensions without changing granularity
- Time dimension, clerk dimension and customer dimension

Week 6

Lecture 17 – 30/08/2011

- Class Test scripts shown and feedback given

Lecture 18 – 01/09/2011

- Extensibility of star schema – change in granularity, addition of new attributes in dimension tables, addition of new facts
- Inventory management business process
- Periodic snapshot schema
- Semi additive facts
- Periodic snapshot fact table vs. retail sales fact table at the same granularity level

Week 7

Lecture 19 – 05/09/2011

- Further facts in periodic snapshot schema
- Inventory transactions schema
- Similarity with banking operations

Lecture 20+21 – 07/09/2011

- Inventory accumulating snapshot
- Handling slowly changing dimensions
- Type 1, Type 2 and Type 3 changes

Lecture 22 – 08/09/2011

- Hybrid methods for handling changes
- Predictable and unpredictable changes

Week 8

Lecture 23 – 12/09/2011

- Rapidly changing dimensions
- Mini dimensions
- Dimension outriggers
- Bitmap indexing

Lecture 24+25 – 14/09/2011

- Join indexing
- Summary of topics covered under Data Warehousing
- Introduction to Data Mining
- Data Mining and SQL
- Data Mining and Machine Learning
- Introduction to Association Rule Mining
- Support and Confidence

Lecture 26 – 15/09/2011

- Support and Confidence
- Itemsets
- Properties of frequent itemsets – closure properties
- Minimum support and confidence
- A priori algorithm

Week 9

Lecture 27 – 19/09/2011

- Generating association rules from large itemsets
- Dynamic itemset counting algorithm

Lecture 28 – 21/09/2011

- Dynamic itemset counting contd..
- Generating association rules from the large itemsets obtained using DIC
- Partitioning algorithm

Week 10

Lecture 29 – 10/10/2011

- Mining frequent patterns without candidate generation
- FP-Tree Construction
- Mining FP Trees

Lecture 30+31 – 12/10/2011

- FP-Tree Construction - revisited
- Mining FP Trees – revisited
- Examples worked out from Paper and Book
- Mid-sem scripts shown and feedback given

Lecture 32 – 13/10/2011

- Sequential pattern mining
- Sequence, Maximal sequence, large customer sequence

Week 11

Lecture 33 – 17/10/2011

- Sequential pattern mining
- Phases in sequential pattern mining

Lecture 34+35 – 19/10/2011

- Sequential pattern mining
- Maximal Phase
- Example worked out by students

Lecture 36 – 20/10/2011

- Introduction to Clustering
- Hierarchical and Partitioning approaches
- Definition of centroid and error measure
- K-means clustering

Week 12

Lecture 37 – 24/10/2011

- K-mediod clustering
- PAM
- CLARA

Lecture 38 – 27/10/2011

- CLARANS

Week 13

Lecture 39 – 31/10/2011

- Problem on CLARANS worked out
- Introduction to hierarchical clustering
- Agglomerative hierarchical clustering
- Dendrogram
- Distance metrics
- Inter-cluster distance measures

Lecture 40+41 – 02/11/2011

- Agglomerative hierarchical clustering Algorithm
- CF Vector
- CF Tree

Lecture 42 – 03/10/2011

- Insertion in CF Tree
- BIRCH
- Introduction to Classification

Week 14

Lecture 43+44 – 09/11/2011

- Quantifying classifier performance
- MLP as classifier
- Back Propagation algorithm
- Introduction to Decision trees

Lecture 45+46 – 10/11/2011

- Decision tree construction algorithm
- Problems on decision tree, MLP, BIRCH

<u>**Week 15**</u>

Lecture 47 – 16/11/2011

- Clarification on all topics covered during the semester
- Summary of topics covered in the course