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

Week 1

Lecture 1+2 – 25/07/2012

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

Lecture 3 – 26/07/2012

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

Week 2

Lecture 4 – 30/07/2012

- Difference between OTLP and OLAP
- Need for a separate database
- Data warehouse architecture
- Basics of normalization

Lecture 5+6 – 01/08/2012

- OLAP operations
- Slicing, dicing, roll-up, drill down
- Change in the number of dimensions as various OLAP operations are performed
- Relational database design for representing OLAP cube data
- SQL queries to retrieve data for populating the OLAP cubes at different levels of detail.

Lecture 7 – 02/08/2012

- Defining relations on queries
- Lattice of cuboids
- Views for generating results
- Multidimension analysis – how it is related to queries and cuboids in the lattice
- Dimension hierarchy
- Slicing, dicing, roll-up and drill-down in presence of dimension hierarchy

- Table structure and join requirements in presence of dimension hierarchy
- Number of cuboids in the lattice in presence of dimension hierarchy

Week 3

Lecture 8 – 06/08/2012

- Schema hierarchy and set grouping hierarchy
- Determining lattice size in presence of hierarchy - examples
- Lattice definition revisited
- Drawing lattices in presence of dimension hierarchy
- Materialized view - definition
- Query execution cost definition

Lecture 9+10 – 08/08/2012

- View materialization problem
- Harinarayan's paper
- ROLAP, MOLAP, HOLAP servers
- Term paper group formation

Lecture 11 – 09/08/2012

- Multi-way array aggregation

Week 4

Lecture 12 – 13/08/2012

- Data warehouse schema design
- Multidimensional modeling
- Fact and dimension tables
- Denormalized design
- Alternative normalized design

Lecture 13 – 16/08/2012

- Star schema and snowflake schema
- Size estimation for fact and dimension tables
- Effects of de-normalized design
- Why de-normalization may not affect consistency
- Steps in data warehouse design

Week 5

Lecture 14+15 – 22/08/2012

- Class Test 1 held

Lecture 16 – 23/08/2012

- Granularity at the level of POS transaction
- Promotion dimension
- Degenerate dimension
- Promotion coverage factless fact table
- Time dimension example with city, temperature, pressure, date and time
- Extensibility of star schema

Week 6

Lecture 17 – 27/08/2012

- Time dimension
- OLTP application and its handling of time
- Timestamp column in master and detail tables – examples from retail sales and passenger reservation
- Impact of putting time key in POS fact table
- Factless fact table handling time

Lecture 18+19 – 29/08/2012

- Extensibility of star schema
- Inventory management business process
- Inventory periodic snapshot schema
- Similarity with account balance
- Additivity of facts – additive, semi-additive, non-additive facts
- Introduction to inventory transactions
- Class test 1 scripts shown and feedback given

Lecture 20 – 30/08/2012

- Inventory transaction schema
- Inventory accumulating snapshot schema
- Fact constellation/galaxy schema
- Data marts and data warehouse bus architecture

Week 7

Lecture 21 – 03/09/2012

- Slowly changing dimensions
- Type 1, 2, 3 changes
- Hybrid changes

Lecture 22+23 – 05/09/2012

- Customer dimension
- City and state dimension outriggers
- Aggregate attributes
- Rapidly changing dimensions
- Customer demographic mini dimension
- Current customer demographic mini dimension key in customer dimension table

Lecture 24 – 06/09/2012

- Distributive, algebraic and holistic measures
- Bitmap indexing
- Join indexing

Week 8

Lecture 25 – 10/09/2012

- Summary of topics covered under Data warehousing

Lecture 26+27 – 12/09/2012

- Introduction to data mining
- Data mining and machine learning
- Data mining and KDD
- Broad topics to be covered under data mining
- Association rule mining
- Itemsets and transactions
- Support and confidence
- Non-monotone property of frequent itemsets
- A priori algorithm
- Candidate sets and frequent itemsets

Lecture 28 – 13/09/2012

- Example on A priori algorithm worked out
- Students given problem to work out

Week 9

Lecture 29 – 17/09/2012

- Generating meaningful association rules from frequent itemsets using minimum confidence
- Improving the performance of the frequent itemset generation step
- Marking transactions as those not containing any frequent itemsets
- Sampling
- Partitioning

Lecture 30+31 – 19/09/2012

- Steps in partitioning approach
- Problems on partitioning approach worked out
- Dynamic itemset counting algorithm

Lecture 32 – 20/09/2012

- Problems on dynamic itemset counting algorithm
- Observation on the performance of dynamic itemset counting based on dataset
- Difficulty in generating a large number of candidate itemsets

Week 10

Mid sem exam held

Week 11

Lecture 33 – 04/10/2012

- FP Tree construction
- Mid-sem scripts shown and feedback given

Week 12

Lecture 34 – 08/10/2012

- Mining FP tree for frequent itemset generation
- Examples solved

Lecture 35+36 – 10/10/2012

- FP Tree mining examples solved
- Quantitative, multi-level and multi-dimensional association rule mining
- Mining sequential patterns
- Sequences, maximal sequences, customer sequences, large sequences

Lecture 37 – 11/10/2012

- Steps in sequential pattern mining
- Algorithms presented and examples solved

Week 13

Lecture 38 – 15/10/2012

- Introduction to clustering
- Partitioning and hierarchical approaches
- K-means clustering
- Introduction to K-medoid algorithm

Lecture 39+40 – 17/10/2012

- PAM
- CLARA
- CLARANS

Lecture 41 – 18/10/2012

- Problems on clustering worked out

Week 14

Lecture 42 – 29/10/2012

- Hierarchical clustering
- Agglomerative and divisive
- Intra-cluster and inter-cluster distance measures
- Algorithm for agglomerative hierarchical clustering

Lecture 43+44 – 31/10/2012

- CF vector
- CF tree
- BIRCH

Lecture 45 – 01/11/2012

- Problems on CF tree and BIRCH worked out

Week 15

Lecture 46 – 05/11/2012

- Introduction to classification
- Steps in classification
- Measuring classification performance
- Multi-layer perceptron
- Weight learning by back propagation

Lecture 47+48 – 07/11/2012

- Weight learning by back propagation
- Possible variations in neural network architecture
- Decision trees
- Decision tree building algorithm

Lecture 49 – 08/11/2012

- Information gain
- Problems on decision tree worked out

Week 16

Lecture 50 – 12/11/2012

- Alternative measures for attribute selection
- Scalable decision tree construction – SPRINT and SLIQ
- Problem on decision worked out
- Problem on MLP worked out

Lecture 51 – 14/11/2012

- Summary and conclusions