**National University of Computer & Emerging Science**

**Department of Computer Science, Lahore**

**Data Warehousing & Data Mining**

**COURSE OUTLINE**

**Fall 2016**

**Course Instructor**

Muhammad Ishaq Raza, Email: [ishaq.raza@nu.edu.pk](mailto:ishaq.raza@nu.edu.pk), Office Hours: M & W: 12:30 – 2:00

**Course Description**

This course covers the concepts and techniques in the design and construction of high-performance data warehouses. The software, hardware and design factors influencing performance characteristics of the data warehouse will be emphasized. A special focus will be given to features and functions in RDBMS implementations that are appropriate in a data warehouse environment. Distinction between DSS (Decision Support System) and OLTP workloads will be made with an emphasis on performance characteristics and functionality required.

**Goals**

* Introduction to the concepts and techniques in data warehousing
* Design and construction of high performing data warehouses
* Hands-on experience with a commercial data warehousing software

**Text Book**

* Paulraj Ponniah, *Data Warehousing Fundamentals*, John Wiley & Sons, 2010.
* Handouts

**Reference Material**

* Ralph Kimball, The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses, John Wiley & Sons, 1998.
* Ralph Kimball, *The Data Warehouse Toolkit.* John Wiley & Sons, June 1996.
* W. H. Inmon, *Building the Data Warehouse* (3rd Edition), John Wiley & Sons, 2002.
* Articles

**Grading Scheme**

## Midterms (25%), Quizzes (10%), Assignments (10%), Class Participation (5%), Project (10%), Final (40%)

**Grading**

* Minimum eligibility to pass this course is to get 50% marks.
* Academic integrity is expected of all the students. Plagiarism or cheating in any assessment will result in at least an **F** grade in the course, and possibly more severe penalties.

**Lab-work:** There will be 14 Labs (3 hours/week). Attending the labs is mandatory.

**Project**

Students will design, implement, demonstrate and document a data warehouse. The project is to be done in groups of 3/4 students. Pick your partner as soon as possible. The groups are self-policing (e.g. each group is responsible for its own division of labor, scheduling, etc.).

A separate handout will be provided describing the project requirements in the 2nd week of the classes.

**Tentative Course Outline and Lecture Plan**

|  |  |  |  |
| --- | --- | --- | --- |
| Topics | Readings | Sessions | Week No. |
| Overview and Concepts:DW Fundamentals, need for a DW, decision support versus transaction processing, evolution of a DW | Ch. 1,2,3 Handout | 1-3 | 1-2 |
| Logical and Physical Data Modeling: Normalization vs. denormalization, pre-Join denormalization, column replication/ movement, pre-aggregation denormalization | Handout | 4-5 | 2-3 |
| OLAP Implementation Techniques: OLAP framework for decision support, Physical implementation techniques: MOLAP, ROLAP, HOLAP, and DOLAP,  Star schema design | Ch. 15  Handout | 6-7 | 3-4 |
| Dimensional Modeling: Principles of dimensional modeling, Physical database design for ROLAP deployment, Natural versus surrogate key design | Ch. 10,11  Handout | 8-10 | 4-5 |
| Extract, Transform, Load (ETL) Processing | Ch. 12  Handout | 11-13 | 6-7 |
| Join Techniques and Performance Evaluation for Data Warehousing: DSS vs. OLTP queries, nested loop join, sort merge join, merge join, hash join, pointer-based join, query optimization | Handout | 14-16 | 7-8 |
| Indexing Techniques for Data Warehousing: Traditional B-tree indexing, hash Indexing, primary vs. secondary indexing, single index access vs. scanning, combining multiple indexes, dynamic bitmap indexing, static bitmap indexing, composite indexing, covered indexing, cluster indexing, partial indexing | Handout | 17-20 | 9-10 |
| Advanced Physical Database Design: Horizontal and vertical partitioning, materialized views framework, materialized views for geography manipulation, advanced aggregation functions | Ch. 18  Handout | 21-24 | 11-12 |
| Data Mining and Data Visualization | Handout | 25-26 | 13 |
| Advanced Data Warehousing Concepts: Dependent versus independent data marts, integrated data marts, operational data stores | Handout | 27-28 | 14 |