F A S T   S c h o o l   o f   C o m p u t i n g

DS3003 – Data Warehousing & Business Intelligence

FALL 2024

Instructor Name: Muhammad Ishaq Raza

Email address: ishaq.raza@nu.edu.pk

Office Location/Number: NB-F-017

Office Hours: Mon, Wed 1:30 - 3:00 PM

TA Name (BDS-5A): Alaiba Nawaz

Email: l215650@lhr.nu.edu.pk

TA Name (BDS-5B): Afnan Hussain

Email: l215693@lhr.nu.edu.pk

TA Name (BDS-5C): Sarah Aslam

Email: l201390@lhr.nu.edu.pk

Course Information

Program: DS   **Credit Hours: 3**   **Type:** Core

Pre-requisites: DS2001 – Introduction to Data Science

Program Learning Outcomes (PLOs)

This course covers the following PLOs:

PLO#	PLO Name	PLO Description
PLO 2	Knowledge for Solving Computing Problems	Apply knowledge of computing fundamentals, knowledge of a computing specialization, and mathematics, science, and domain knowledge appropriate for the computing specialization to the 16 abstraction and conceptualization of computing models from defined problems and requirements.
PLO 5	Modern Tool Usage	Create, select, adapt and apply appropriate techniques, resources, and modern computing tools to complex computing activities, with an understanding of the limitations

Course Learning Outcomes (CLOs)

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. By the end of the course, students should be able to achieve the following CLOs:

CLO#	CLO Description	Taxonomy Level	PLO #
CLO 1	**Demonstrate** an appreciation of the role that Data Warehouses and Business Intelligence play in enhancing the decision-making process	**C3** (Applying)	**PLO 2**
CLO 2	**Demonstrate** an understanding of the fundamental concepts of the Star and the Snowflake Schema; learn how to design the schema of a DW based on these two models	**C3** (Applying)	**PLO 4**
CLO 3	**Understand** the architecture of DW Systems and be able to specify the advantages and potential problem areas	**C2** (Understanding)	**PLO 4**
CLO 4	**Use** Analytic SQL to aggregate, analyze, report, and model data	**C3** (Applying)	**PLO 4**

Course Textbook

- Paulraj Ponniah, _Data Warehousing Fundamentals_, John Wiley & Sons, 2010.
- Handouts

Reference Books

- 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 (Absolute)

Midterms (30%), Quizzes/Class Participation (10%), Assignments (10%), Final (50%)

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.

Tentative Course Outline and Lecture Plan

Week	Topics to be covered	Readings (Textbook)	No of Lectures	Asst.
1-2	1. **Overview and Concepts:** DW Fundamentals, need for a DW and BI, decision support versus transaction processing, evolution of a DW	Ch. 1,2,3 Handout	3	A1
2-3	2. **Logical and Physical Data Modeling:** Normalization vs. denormalization, pre-Join denormalization, column replication/ movement, pre-aggregation denormalization	Handout	2	
3-4	3. **OLAP Implementation Techniques:** OLAP framework for decision support, Physical implementation techniques: MOLAP, ROLAP, HOLAP, and DOLAP, Star schema design	Ch. 15 Handout	2	
4-5	4. **Dimensional Modeling:** Principles of dimensional modeling, Physical database design for ROLAP deployment, Natural versus surrogate key design	Ch. 10,11 Handout	3	A2
6	5. **Extract, Transform, Load (ETL) Processing**	Ch. 12 Handout	2	
7	6. **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	2	A3
8-9	7. **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	4	A4
10-11	8. **Advanced Physical Database Design:** Horizontal and vertical partitioning, materialized views framework, materialized views for geography manipulation, advanced aggregation functions	Ch. 18 Handout	3	
11-12	9. **Data Mining and Data Visualization**	Handout	3	
13-14	10. **Advanced Data Warehousing & BI Concepts**	Handout	4	