****

Course Title: Data Warehouse and Mining

Course Code: IS-622

Resource Person: Usman Khalid

Department: IS Department

**HSM Vision**

HSM envisions its success in the sustainable contribution that it will make to the industry, academia and research in public and private sector. HSM will lead by providing professionally competent and ethically conscious human resources engaged in the global and local context to foster socio-economic growth and sustainability for the society. HSM envisages having faculty with high research potential and a deep desire for cutting edge research including collaboration with national and international partners.

**HSM Mission**

Being a research-oriented and student-centric business school, we emphasize research publications in impact journals as well as state-of -the-art learning methodologies.  We will prepare our students to become the future ethical business leaders and the guiding post for the society, while equipping them with the knowledge and skills required by world-class professionals.  We will be the leading choice for organizations seeking highly talented human resource. HSM will foster internationalization with key stakeholders and actively work to exchange best practices with business schools across Pakistan through collaborations, workshops, conferences and other means.

**Program Objectives**

**Course Objectives**

A data warehouse is used for the purpose of providing and analyzing information from a number of sources including inside as well as outside of the organization. The course introduces students to the terms and architecture of data warehousing, dimensional modeling, data sources, Extract-Transform-Load (ETL) processes and subsystems, and schemas. Moreover Online Analytical Processing (OLAP) and Online Transactional Processing (OLTP) will also be discussed. A few case studies will be used to help students understand the importance as well the implementation of data warehouses in a corporate environment

**Learning Objectives**

1. Introduction to the concepts and techniques of data warehousing
2. Help understand the Extract-Transform-Load process.
3. Understand how OLAP and OLTP work
4. Data mining and concepts

**Learning Outcomes**

1. Students will get familiar with the dynamics and functionality of a data warehouse as well as how it performs in the corporate environment.
2. Students will understand how the ETL process works which includes the requirements and the subsystems
3. Students will understand the dynamics and functionalities of OLAP and OLTP and the purposes they serve in the real world.
4. Students will learn about data mining concepts as well as the process, clustering, classification and association rules, as well as text and web mining

**Teaching Methodology (List methodologies used –example are given below)**

Interactive Classes

Case based teaching

Class activities

Applied Projects

Experiential Learning

**STUDENTS ARE REQUIRED TO READ AND UNDERSTAND ALL ITEMS OUTLINED IN THE PARTICIPANT HANDBOOK**

**Class Policy:-**

* Be On Time

You need to be at class at the assigned time. After 10 minutes past the assigned time, you will be marked absent.

* Mobile Policy

**TURN OFF YOUR MOBILE PHONE!** It is unprofessional to be texting or otherwise.

* Email Policy

**READ YOUR EMAILS!** You are responsible if you miss a deadline because you did not read your email.

Participants should regularly check their university emails accounts regularly and respond accordingly.

* Class Attendance Policy

A minimum of 80% attendance is required for a participant to be eligible to sit in the final examination. Being sick and going to weddings are absences and will not be counted as present. You have the opportunity to use 6 absences out of 30 classes. Participants with less than 80% of attendance in a course will be given grade ‘F’ (Fail) and will not be allowed to take end term exams. International students who will be leaving for visa during semester should not use any days off except for visa trip. Otherwise they could reach short attendance.

* Withdraw Policy

Students may withdraw from a course till the end of the 12th week of the semester. Consequently, grade W will be awarded to the student which shall have no impact on the calculation of the GPA of the student.A Student withdrawing after the 12th week shall be automatically awarded “F” grade which shall count in the GPA.

* Moodle

UMT –LMS (Moodle) is an Open Source Course Management System (CMS), also known as a learning Management System (LMS). Participants should regularly visit the course website on MOODLE Course Management system, and fully benefit from its capabilities. If you are facing any problem using moodle, visit <http://oit.umt.edu.pk/moodle>. For further query send your queries to [moodle@umt.edu.pk](mailto:moodle@umt.edu.pk)

* Harassment Policy

Sexual or any other harassment is prohibited and is constituted as punishable offence. Sexual or any other harassment of any participant will not be tolerated. All actions categorized as sexual or any other harassment when done physically or verbally would also be considered as sexual harassment when done using electronic media such as computers, mobiles, internet, emails etc.

* Use of Unfair Means/Honesty Policy

Any participant found using unfair means or assisting another participant during a class test/quiz, assignments or examination would be liable to disciplinary action.

* Plagiarism Policy  
    
  All students are required to attach a “Turnitin” report on every assignment, big or small. Any student who attempts to bypass “Turnitin” will receive “F” grade which will count towards the CGPA. The participants submit the plagiarism report to the resource person with every assignment, report, project, thesis etc. If student attempts to cheat “Turnitin”, he/she will receive a second “F” that will count towards the CGPA. There are special rules on plagiarism for final reports etc. all outlined in your handbook.
* Communication of Results

The results of quizzes, midterms and assignments are communicated to the participants during the semester and answer books are returned to them. It is the responsibility of the course instructor to keep the participants informed about his/her progress during the semester. The course instructor will inform a participant at least one week before the final examination related to his or her performance in the course.

**Course Outline**

Course code……………………………...... Course title………………………………………

|  |  |
| --- | --- |
| Program | BBIS |
| Credit Hours | 3 |
| Duration | 15 weeks |
| Prerequisites (If any) | no |
| Resource Person  Name and Email | Usman Khalid |
| Counseling Timing  (Room# ) | **Monday:** 4:00 – 6:30  **Tuesday:** 12:00 – 2:00  **Friday:** 12:00 – 2:00 |
| Contact no. | 0331-4773534 |
| Web Links:- (Face book, Linked In, Google Groups, Other platforms) | [www.facebook.com/usman](http://www.facebook.com/usman) khalid |

**Chairman/Director Programme signature………………………………….Date……………………..**

**Dean’s signature…………………………… ………………….Date………………………………………….**

**Grade Evaluation Criteria**

Following is the criteria for the distribution of marks to evaluate final grade in a semester.

**Marks Evaluation Marks in percentage**

Quizzes 10%

Assignments

Mid Term 20%

Attendance & Class Participation 25%

Term Project

Presentations 15%

Final exam 30%

Total 100

**Recommended Text Books:**

**The Data Warehouse Toolkit,** Third Edition, Ralph Kimball, Margy Ross, 2014

**Reference Books:**

1. Data Warehousing,Reema Thareja, 2009
2. Handouts

**Course: ----- -Course code: ------- ------------Book: ----------------------------------------------------------------------------------------------**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Topics to be**  **covered in the course** | **Learning Objective**  **of this topic** | **Expected Outcomes from Students** | **Teaching Method** | **Assessment Criteria** | **Deadlines and Homework** |
| 1 | Introduction, general information and key terms of the subject. | Introduction of data warehouses, and differences and limitation of Operational Databases, differences between OLAP and OLTP and schema types of Data Warehouses |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 2 | Dimensional modeling and architecture and case study | Understanding the four components of a data warehouse environment. A case study will be done to help understand fact table and dimensional modeling |  | Presentation / White Board | Assign. / Quiz / Surprise Questions. **A final project will be assigned to students at this point. This project will be due in second last week of the semester.** |  |
| 3 | Dimensional modeling and case study | Steps in dimensional design process.  Introduction of Star schema, Fact-less fact table, degenerate dimensions, snowflaking and surrogate keys. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 4 | Value chains, fact tables, and dimensions. | Implications of value chain, periodic snapshots, semi-additive facts, Bus matrix, Role playing, Junk dimensions, outrigger and mini-dimensions and grain of fact tables. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 5 | Role playing, Junk dimensions, outrigger and mini-dimensions and grain of fact tables | Understanding the concepts of Role playing, Junk dimensions outrigger and mini-dimensions. The grain of fact tables will also be discussed in-depth because of its importance with the final assignment. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 6 | Case Study | Students will complete a Case study (To better understand how to do the final project) |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 7 | ETL requirements, process steps, and subsystems. | 10 requirements for ETL, process steps and sub-systems 1-3 Extracting process, and 4-8 Cleaning and conforming process |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 8 | ETL sub-systems | Sub-systems 9-21 the cleaning and conforming process, sub-system 21-34 supporting the management |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 9 | OLAP | Purpose of OLAP, relationship between OLAP and data warehousing, representing multi-dimensional data, categories and comparisons of OLAP tools. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 10 | OLAP with extensions in SQL | Main categories of OLAP tools, OLAP extensions in SQL, |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 11 | Introduction to data mining | Data mining and its applications, decision trees, classification, assosciation, the data mining process |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 12 | Introduction to Data mining and its applications – Classification and clustering | Neural networks, supervised v unsupervised learning, k-means clustering, MLP. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 13 | What kinds of patterns can be mined? Clustering and assosciation | K-means algorithm, categories of clustering, confidence and support, the aprori algorithm, data mining myths |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 14 | Text and Web minig | Text and web mining concepts, Application area and process, NLP, Web-content and web structure mining. |  | Presentation / White Board | Assign. / Quiz / Surprise Questions |  |
| 15 | Final project presentation | Final project presentation |  | Presentation |  |  |