BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES

COURSE HANDOUT

Part A: Content Design

Course Title	Data Warehousing
Course No(s)	CSI ZG515 / SS ZG515/SE ZG515
Credit Units	5 (2: Class Room Hours; 2: Students Preparation; 1: Case Studies) *(1 credit unit translates to approximately 32 hours)*
Course Author	YASHVARDHAN SHARMA
Version No	
Date	

Course Description

Corporate decision makers require access to all the organization's data, wherever it is located. To provide comprehensive analysis of the organization, its business, its requirements and any trends, require access to not only the current data in the database but also to historical data. To facilitate this type of analysis, data warehouses have been created to contain data drawn from several sources, maintained by different departments of the organization. This course will involve an in-depth study of various concepts needed to design, develop, and maintain a data warehouse. It also provides an introduction to end user access tools like OLAP and reporting.

Course Objectives

CO1	Understand the importance of data warehouse and the business intelligence that can be gained with the help of data warehouse
CO2	Understand schema designs, information delivery techniques and architectures that are appropriate for data warehouse
CO3	Understand processes, management, and infrastructure necessary for building data warehouse
CO4	Understand evolution of data warehouse with the presence of big data and cloud environments.

Text Books

T1	Ponniah P, ***"Data Warehousing Fundamentals"***, Wiley Student Edition, 2012
T2	Kimball R, ***"The Data Warehouse Toolkit",*** 3e, John Wiley, 2013

Reference Books

R1	Anahory S, & Dennis M, "***Data Warehousing in the Real World"***, Pearson Education, 2008.
R2	Kimball R, Reeves L, Ross M, & Thornthwaite, W, ***"The Data Warehouse Lifecycle Toolkit",*** John Wiley, 2e, 2012.
R3	Jiawei Han, Micheline Kamber and Jian Pei, "***Data Mining: Concepts and Techniques***", Morgan

	Kaufmann Publishers 2012
R4	Krish Krishnan, "***Data Warehousing in the Age of Big Data****",* Morgan Kaufmann Publishers 2013
R5	William H Inmon, et al., "***DW 2.0 : The Architecture for the Next Generation of Data Warehousing***", Morgan Kaufmann 2012

Modular Structure

No	Title of the Module
M1	Introduction to Data Warehousing
M2	Introduction to Dimensional Modelling
M3	Architectural Components of a Data Warehouse
M4	Extraction, Transformation & Loading
M5	Advanced Dimensional Modelling
M6	Online Analytical Processing (OLAP) & Multidimensional Databases (MDDB)
M7	Query Performance Enhancing Techniques
M8	Metadata
M9	Support for Data Warehousing in RDBMS/SQL
M10	Real-time Data Warehousing
M11	Current Trends in Data Warehousing

Learning Outcomes:

No	Learning Outcomes
LO1	Knowledge of business intelligence that can be gained from data warehouse.
LO2	Knowledge of schema designs, information delivery techniques and architectures for data warehouse.
LO3	Knowledge of processes, management, and infrastructure for building data warehouse.
LO4	Evolution of data warehouse with the presence of big data and cloud environments.

Part B: Contact Session Plan

Academic Term	First Semester 2021-2022
Course Title	Data Warehousing
Course No	CSI ZG515 / SS ZG515/SE ZG515
Lead Instructor	G.SHOBHA

Glossary of Terms

1. Contact Hour (CH) stands for a hour long live session with students conducted either in a physical classroom or enabled through technology. In this model of instruction, instructor led sessions will be for 22 CH.
    a. Pre CH = Self Learning done prior to a given contact hour
    b. During CH = Content to be discussed during the contact hour by the course instructor
    c. Post CH = Self Learning done post the contact hour
2. Contact Hour (CS) stands for a two-hour long live session with students conducted either in a physical classroom or enabled through technology. In this model of instruction, instructor led sessions will be for 11 CS.
    a. Pre CS = Self Learning done prior to a given contact session
    b. During CS = Content to be discussed during the contact session by the course instructor
    c. Post CS = Self Learning done post the contact session
3. RL stands for Recorded Lecture or Recorded Lesson. It is presented to the student through an online portal. A given RL unfolds as a sequences of video segments interleaved with exercises
4. SS stands for Self-Study to be done as a study of relevant sections from textbooks and reference books. It could also include study of external resources.
5. LE stands for Lab Exercises
6. HW stands for Home Work.
7. M stands for module. Module is a standalone quantum of designed content. A typical course is delivered using a string of modules. M2 means module 2.

Teaching Methodology (Flipped Learning Model)

The pedagogy for this course is centered around flipped learning model in which the traditional class-room instruction is replaced with recorded lectures to be watched at home as per the student's convenience and the erstwhile home-working or tutorials become the focus of classroom contact sessions. Students are expected to finish the home works on time.

Contact Session Plan

- Each Module (M#) covers an independent topic and module may encompass more than one Recorded Lecture (RL).
- Contact Sessions **(2hrs each week)** are scheduled alternate weeks after the student watches all Recorded Lectures (RLs) of the specified Modules (listed below) during the previous week
- In the flipped learning model, Contact Sessions are meant for in-classroom discussions on cases, tutorials/exercises or responding to student's questions/clarification--- may encompass more than one Module/RLs/CS topic.
- Contact Session topics listed in course structure (numbered CSx.y) may cover several RLs; and as per the pace of instructor/students' learning, the instructor may take up more than one CS topic during each of the below sessions.

Detailed Structure

Introductory Video/Document: *<< Introducing the faculty, overview of the course, structure and organization of topics, guidance for navigating the content, and expectations from students>>*

- Each of the sub-modules of **Recorded Lectures** (RLx.y) shall delivered via **30 – 60mins videos** followed by:
- **Contact session** (CSx.y) of 2Hr each for illustrating the concepts discussed in the videos with exercises, tutorials and discussion on case-problems (wherever appropriate); contact sessions (CS) may cover more than one recorded-lecture (RL) videos.

Course Contents

Contact Session 1

M1: Introduction to Data Warehousing

Time	Type	Description	Content Reference
Pre-CS	RL1.1	Introduction to Data Warehousing - I RL1.1.1 = Evolution of Data Warehousing RL1.1.2 = Operational System Characteristics RL1.1.3 = Data Warehousing System Characteristics RL1.1.4 = Data Warehouse Architecture	
	RL1.2	Introduction to Data Warehousing - II RL1.2.1 = Introduction to ETL, Data Staging Area & Presentation Servers RL1.2.2 = Introduction to Data Warehouse Design RL1.2.3 = Continuum of Analysis	
During CS	CS1.1	CS1.1.1 = Review of concepts of DW, ETL, Information Delivery (T1, Ch 1) CS1.1.2 = Examples of Business Intelligence applications (T2, Introduction) CS1.1.3 = Challenges in designing DW (T2, Ch 1)	
	CS1.2	CS1.2.1 = Compare DW with Data Mining (R1, Appendix B) CS1.2.2 = DW, Unstructured data and Big data (T2, Ch 21)	
Post-CS	SS1.1	Self Study (T1, Chapter 1; T2, Chapter 1)	
	HW1.1	Do exercises given at the end of chapter 1 of T1	
Lab Reference			

Contact Session 2

M2: Introduction to Dimensional Modeling

Time	Type	Description	Content Reference
Pre-CS	RL2.1	RL2.1.1 = E R Modelling RL2.1.2 = Dimensional Modelling Vocabulary RL2.1.3 = ER Modelling vs. Dimensional Modelling - Qualitative Comparison	
	RL2.2	RL2.2.1 = Data Warehouse Design Steps RL2.2.2 = Grocery Store Case Study	
During CS	CS2.1	CS2.1.1 = Review Dimensional Modeling (T1, Ch 10) CS2.1.2 = Compare ER and Dimensional Modeling (T1, Ch 10)	
	CS2.2	CS2.2.1 = Explain Retail Store analysis requirements (T2, Ch 3) CS2.2.2 = Retail Store Dimensional Model (T2, Ch 3)	
Post-CS	SS2.1	Self Study (T1, Chapter 10; T2, Chapter 3)	
	HW2.1	Do exercises given at the end of chapter 10 of T1	
Lab Reference			

Contact Session 3

M3: Architectural Components of a Data Warehouse

Time	Type	Description	Content Reference
Pre-CS	RL3.1	RL3.1.1 = Data Marts RL3.1.2 = ODS RL3.1.3 = Top-down vs. Bottom up approaches to Data Warehousing	
	RL3.2	RL3.2.1 = ETL & Data Staging Area RL3.2.2 = Presentation Servers RL3.2.3 = OLAP	
During CS	CS3.1	CS3.1.1 = Review concepts of data mart (T1, Ch 2) CS3.1.2 = Explain pros and cons of taking data mart approach towards EDW (T1, Ch 2)	
	CS3.2	CS3.2.1 = Explain concepts of staging and OLAP (T1, Ch 7) CS3.2.2 = Explain desirable features for staging server and OLAP (T1, Ch 7)	
Post-CS	SS3.1	Self Study (T1, Chapter 7)	
	HW3.1	Do exercises given at the end of chapter 7 of T1	
Lab Reference			

Contact Session 4

M4: Extraction, Transformation, & Loading (ETL)

Time	Type	Description	Content Reference
Pre-CS	RL4.1	RL4.1.1 = ETL Overview RL4.1.2 = Major ETL Tasks RL4.1.3 = ETL Requirements RL4.1.4 = Data Extraction	
	RL4.2	RL4.2.1 = Data Transformation RL4.2.2 = Data Loading RL4.2.3 = Initial Load & Refresh Cycles	
	RL4.3	RL4.3.1 = Data Quality	
During CS	CS4.1	CS4.1.1 = Explain concepts of ETL (T1, Ch 12) CS4.1.2 = Why data requires pre-processing (R3, Ch 3)	
	CS4.2	CS4.2.1 = Major types of transformations (R3, Ch 3) CS4.2.2 = Methods of applying data to the warehouse (T1, Ch 12)	
Post-CS	SS4.1	Self Study (T1, Chapter 12)	
	HW4.1	Do exercises given at the end of chapter 12 of T1	
Lab Reference			

Contact Session 5

M5: Advanced Dimensional Modelling

Time	Type	Description	Content Reference
Pre-CS	RL5.1	RL5.1.1 = Changing Dimensions, Surrogate keys & Look up tables RL5.1.2 = Mini-dimensions & Outriggers	
	RL5.2	RL5.2.1 = Time Dimension RL5.2.2 = Conformed Dimensions & Facts	
	RL5.3	RL5.3.1 = Multi-valued Dimensions (Bridge & Helper Tables) RL5.3.2 = Dimension Hierarchies RL5.3.3 = Role-playing Dimensions RL5.3.4 = Factless Fact Tables	
	RL5.4	RL5.4.1 = Academic Warehouse Case Study	
	RL5.5	RL5.5.1 = Bank Data Warehouse Case Study	
During CS	CS5.1	CS5.1.1 = Illustrate problems of changing dimensions (T2, Ch 2) CS5.1.2 = Techniques to handle changing dimensions (T2, Ch 2) CS5.1.3 = Examples for mini-dimensions and outriggers (T2, Ch 2)	
	CS5.2	CS5.2.1 = Illustrate time dimension (T2, Ch 3) CS5.2.2 = Explain enterprise bus architecture (T2, Ch 4) CS5.2.3 = Illustrate bridge, role-playing dimensions, factless fact tables (T2, Ch 2)	
Post-CS	SS5.1	Self Study (T1, Chapter 11)	
	HW5.1	Do exercises given at the end of chapter 11 of T1	
Lab Reference			

Contact Session 6

M6: OLAP & Multidimensional Databases (MDDB)

Time	Type	Description	Content Reference
Pre-CS	RL6.1	RL6.1.1 = Introduction to OLAP & Multidimensional Analysis RL6.1.2 = Limitations of Spreadsheets & SQL	
	RL6.2	RL6.2.1 = Major OLAP Features & Functions RL6.2.2 = Cube Operator	
	RL6.3	RL6.3.1 = Introduction to Multidimensional Databases (MDDB) RL6.3.2 = Multidimensional Analysis & MDDBs	
	RL6.4	RL6.4.1 = OLAP operations using MDDBs	
	RL6.5	RL6.5.1 = MDDB vs. RDBMS	
	RL6.6	Cube Computation: Complexity & Optimization	
During CS	CS6.1	CS6.1.1 = Explain concepts of OLAP (CS6.1.2 = Various OLAP operations with examles (R3, Ch	

		4) CS6.1.3 = Relative strengths of MOLAP, ROLAP, and HOLAP (R3, Ch 4)	
Post-CS	SS6.1	Self Study (T1, Chapter 15)	
	HW6.1	Do exercises given at the end of chapter 15 of T1	
Lab Reference			

Contact Session 7

M7: Query Performance Enhancing Techniques

Time	Type	Description	Content Reference
Pre-CS	RL7.1	RL7.1.1 = Aggregation RL7.1.2 = Sparsity Failure	
	RL7.2	RL7.2.1 = Shrunken, Lost, & Collapsed Dimensions	
	RL7.3	RL7.3.1 = Aggregate Navigator RL7.3.2 = Aggregate Navigation Algorithm	
	RL7.4	RL7.4.1 = Partitioning RL7.4.1 = Partitioning wrt Time	
	RL7.5	RL7.5.1 = View Materialization RL7.5.2 = Selection of Views to Materialize	
	RL7.6	RL7.6.1 = View Maintenance Strategies RL7.6.2 = Incremental Maintenance Algorithms	
	RL7.7	RL7.7.1 = Bitmap Indices RL7.7.2 = Bitmap Compression Strategies	
During CS	CS7.1.	CS7.1.1 = Data Warehouse performance challenges(T1, Ch 18) CS7.1.2 = Concepts of physical design (T1, Ch 18) CS7.1.3 = Use of aggregates, partitions for performance and operations (R1, Ch 6, Ch 7)	
Post-CS	SS7.1	Self Study (T1, Chapter 18)	
	HW7.1	Do exercises given at the end of chapter 18 of T1	
Lab Reference			

Contact Session 8

M8: Metadata

Time	Type	Description	Content Reference
Pre-CS	RL8.1	RL8.1.1 = Role of Metadata RL8.1.2 = Types of Metadata	
	RL8.2	RL8.2.1 = Metadata Design & Implementation	
During CS	CS8.1	CS8.1.1 = Concepts of Metadata (T1, Ch 9) CS8.1.2 = Why Metadata becomes more important in DW (T1,Ch 9) CS8.1.3 = Examples of Metadata (T1, Ch 9)	

Post-CS	SS8.1	Self Study (T1, Chapter 9)	
	HW8.1	Do exercises given at the end of chapter 9 of T1	
Lab Reference			

Contact Session 9
M9: Support for Data Warehousing in RDBMS/SQL

Time	Type	Description	Content Reference
Pre-CS	RL9.1	RL9.1.1 = Support for Data Warehousing in RDBMSs	
	RL9.2	RL9.2.1 = New Features in SQL	
During CS	CS9.1	CS9.1.1 = Commonality between DW and RDBMS(T1, Ch 1) CS9.1.2 = SQL extensions for analytics (R4,Ch 2, Ch 3, Ch 18)	
Post-CS	SS9.1	Self Study (R4, Chapters 2, 3, 18)	
Lab Reference			

Contact Session 10
M10: Real-Time Data Warehousing (RTDWH)

Time	Type	Description	Content Reference
Pre-CS	RL10.1	RL10.1.1 = Need for Real-time Data Warehousing RL10.1.2 = Solutions for Real-time Data Warehousing	
	RL10.2	RL10.2.1 = Real-time ETL RL10.2.2 = Role of ODS in RTDWH	
During CS	CS10.1	CS10.1.1 = Concepts of real-time solutions(T2, Ch 20) CS10.1.2 = Challenges with real-time data warehousing (T2, Ch 20) CS10.1.3 = ETL vs. ELT (R6, Ch 13)	
Post-CS	SS10.1	Self Study (T1, Chapter 3; T2, Chapter 20)	
Lab Reference			

Contact Session 11
M11: Current Trends in Data Warehousing

Time	Type	Description	Content Reference
Pre-CS	RL11.1	RL11.1.1 = Introduction to World's Largest Data Warehouse	
	RL11.2	RL11.2.1 = Big Data Analytics RL11.2.2 = Extended RDBMS Architecture RL11.2.3 = MapReduce/Hadoop Architecture	
During CS	CS11.1	CS11.1.1 = Recent trends in data warehousing (T1, Ch 3) CS11.1.2 = Place of structured DW vs. Unstructured data, Big Data (T2,Ch 21)	

		CS11.1.3 = Possibilities of convergence of DW and Big Data (R5, Ch 13)	
Post-CS	SS11.1	Self Study (T1, Chapter 3)	
	HW11.1	Do exercises given at the end of chapter 3 of T1	
Lab Reference			

Refer Appendix for detailed course plan

Evaluation Scheme:
Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz-I/ Assignment-I	Online	-	5%	August 16-30, 2021
	Quiz-II	Online		5%	September 16-30, 2021
	Quiz-III/ Assignment-II	Online		5%	October 16-30, 2021
EC-2	Mid-Semester Test	Open Book	2 hours	35%	Sunday, 26/09/2021 (AN) 2 PM – 4 PM
EC-3	Comprehensive Exam	Open Book	2 hours	50%	Sunday, 14/11/2021 (AN) 2 PM – 4 PM

Note *- Evaluation components can be tailored depending on the proposed model.*

Important Information:
Syllabus for Mid-Semester Test (Open Book): Topics in CS 1-5.
Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study
Evaluation Guidelines:

1. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
2. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.

Appendix

Course Plan

Session	Pre-contact prep	Contact session	Post-contact sessions

1	RL1.1, RL1.2	CS1.1, CS1.2	SS1.1, HW1.1
2	RL2.1, RL2.2	CS2.1, CS2.2	SS2.1, HW2.1
3	RL3.1, RL3.2	CS3.1, CS3.2	SS3.1, HW3.1
4	RL4.1, RL4.2, RL4.3	CS4.1, CS4.2	SS4.1, HW4.1
5	RL5.1, 5.2, 5.3, 5.4, 5.5	CS5.1, CS 5.2, Review of all topics covered so far	SS5.1, HW 5.1
		Mid-Semester	
6	RL6.1,6.2,6.3,6.4,6.5,6.6	CS6.1	SS6.1, HW6.1
7	RL7.1,7.2,7.3,7.4,7.5,7.6,7.7	CS7.1	SS7.1, HW7.1
8	RL8.1, RL8.2	CS8.1	SS8.1, HW8.1
9	RL9.1,9.2	CS9.1	SS9.1
10	RL10.1, RL10.2, RL11.1, RL11.2	CS 10.1, CS 11.1, Review of all topics covered	SS11.1, HW11.1
		Comprehensive Exam	