DS 603 DATA MODELING
TERM II (2019-20)

GENERAL COURSE INFORMATION

Course Name	DS/SE 603 Data Modeling
Instructors	Prof. Chandrashekar R Office No. 116 rc@iiitb.ac.in
Course credits	4
Pre-requisite	a) Database Systems b) Knowledge of Java or any web application development c) A liking for data!

COURSE OVERVIEW

This course will provide an in-depth understanding of data modeling both from a theoretical as well as a practical point of view. Building on the concepts introduced in any Database Systems core course, this course aims to provide a thorough understanding of advanced data modeling concepts. In particular, the course will cover in detail multiple data models including advanced relational database concepts, XML, object-oriented databases and Data Warehouse concepts. The emphasis is not only on the logical aspects of data modeling but also the physical/implementation aspects of these data models using freely available software.

COURSE CONTENTS

Introduction to Data Models and the basics
* The 'what', 'why' and 'who' of data models
* The Layers: External Schema, Conceptual Schema, Internal Schema
* The Big Picture – Enterprise Architecture

Review of Traditional Data Models
* Hierarchical Data Model
* Network (CODASYL) Data Model

Data Model Stack for Structured Data
* ER Data model
* Object oriented Data Model
* Relational Data Model
* OR Mapping

More Relational Data Model
* Codd's Rules
* Review of Normal Forms (3NF,BCNF)
* Other normal forms (4NF, 5NF)

XML Data Model
* XML Well-formedness
* XML Validation using XML Schema
* XML Querying using XQuery

Data Models for Data Warehousing
* Dimensional data modeling
  o Introduction to data warehousing, OLAP, etc.
  o Concepts: Fact tables, Dimension tables, Snowflaking, Surrogate Keys, Conformed Dimensions, Conformed Facts, Slowly changing dimensions, degenerate dimensions, factless fact tables
* Online Analytic Processing (OLAP)
  o Review the major features and functions of OLAP

- Dr. Codd's OLAP guidelines
- hypercubes, drill-down and roll-up, and slice-and-dice
- Examine the different OLAP models (ROLAP, MOLAP, etc.)
- OLAP implementation by studying the steps and the tools
- OLAP through SQL, Visual Analytics, R

Semantic Web Concepts
* The vision
* Describing Web Resources using RDF
* Web Ontology Language: OWL

Data models for BigData

* What is Big Data
* "BigData" versus "big data"
* Data models for BigData
* NoSQL concepts
* BigData architectures and programming models
* Data Lake Architecture

Other Topics
* Temporal data
* Stream data
* Model-based Software Engineering

Class projects discussions and demos

COURSE CALENDAR*

Date	Planned
1-Jan	Introduction to Data Models and the basics
3-Jan	Traditional Data Models
8-Jan	Normalization, Codd's Rules
10-Jan	Intro To Xml; XML Rules, Parsing
15-Jan	No class - Holiday
17-Jan	XML Validation using XSD
22-Jan	XML Schema Details
24-Jan	Reuse XML Schema And XML Design Styles
29-Jan	Xquery and XML Databases
31-Jan	R2X Mapping & OODBMS Architecture
5-Feb	Class Test - 1 (Quiz + XML)
7-Feb	Mapping across OO, OOPL, Relational, XML
12-Feb	Data Warehouse And Dimensional Modelling
14-Feb	Facts, Dimensions And Star Schema
19-Feb	Data Visualization
21-Feb	No class - Holiday
26-Feb	DVFA Model And OLAP Operations
28-Feb	Lab - Dimensional Modelling Using Spreadsheets
4-Mar	Mid-term exams
6-Mar	Mid-term exams
11-Mar	Advanced SQL
13-Mar	OLAP Performance Management

Date	Planned
18-Mar	Semantic Web Intro
20-Mar	Resource Description Framework (RDF)
25-Mar	No class - Holiday
27-Mar	Ontology Web Language (OWL)
1-Apr	Models for BigData
3-Apr	Class Test - 2 (Quiz + Analytic SQL)
8-Apr	CAP Theorem and Persistent Stores for Big Data
10-Apr	No class - Holiday
15-Apr	Stream Data Model
17-Apr	Temporal Data Model
22-Apr	Model Based Software Engineering
24-Apr	Buffer
29-Apr	Project Demos

* Minor adjustments to schedule can take place

GRADING

Final grade will be based on weights given below:
35%: Mid-Term Exam
25%: Tests / assignments
15%: Project
35%: End-Term Exam

REFERENCE MATERIAL

- Selected chapters from books and papers from the literature
- *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2nd Edition, R. Kimball and M. Ross, John Wiley & Sons
- Other research papers and reference material to be given as needed

CHEATING AND PLAGIARISM

This course has zero tolerance for cheating and plagiarism. Any violation may result in an F grade and further disciplinary action may be initiated as per the Institute's policies. Ignorance of what constitutes cheating and plagiarism is not an excuse! If you have any doubts, contact your instructor.

DEADLINES

Unless noted otherwise, all deadlines are due at **9:00 AM** on the date indicated

LATE POLICY

- 4 - 24 hours late submission: 25% penalty
- 24 - 48 hours late submissions: 50% penalty
- > 48 hours late submissions: 75% penalty

ANNEXURE

What is Plagiarism

Many people think of plagiarism as copying another's work, or borrowing someone else's original ideas. But terms like "copying" and "borrowing" can disguise the seriousness of the offense:

According to the *Merriam-Webster OnLine Dictionary*, to "plagiarize" means
1) to steal and pass off (the ideas or words of another) as one's own
2) to use (another's production) without crediting the source
3) to commit literary theft
4) to present as new and original an idea or product derived from an existing source.

In other words, plagiarism is an act of fraud. It involves both stealing someone else's work and lying about it afterward. But can words and ideas really be stolen?

According to U.S. law, the answer is yes. In the United States and many other countries, the expression of original ideas is considered intellectual property, and is protected by copyright laws, just like original inventions. Almost all forms of expression fall under copyright protection as long as they are recorded in some media (such as a book or a computer file).

All of the following are considered plagiarism:
- turning in someone else's work as your own
- copying words or ideas from someone else without giving credit
- failing to put a quotation in quotation marks
- giving incorrect information about the source of a quotation
- changing words but copying the sentence structure of a source without giving credit
- copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not (see our section on "fair use" rules)

Attention! Changing the words of an original source is not sufficient to prevent plagiarism. If you have retained the essential idea of an original source, and have not cited it, then no matter how drastically you may have altered its context or presentation, you have still plagiarized

Most cases of plagiarism can be avoided, however, by citing sources. Simply acknowledging that certain material has been borrowed, and providing your audience with the information necessary to find that source, is usually enough to prevent plagiarism.