

**NEW YORK CITY COLLEGE OF TECHNOLOGY  
COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT  
COURSE OUTLINE**

**CST4704 – BUSINESS INTELLIGENCE, DATA WAREHOUSING & DATA MINING**

4 Class Hours, 3 Credits

## **COURSE DESCRIPTION**

This course provides a general introduction to the design of Data Warehouses (DW). The Dimensional model and using it to model DWs for different business applications are discussed in detail. The architecture of a DW, the importance of a DW in the Business Intelligence of an enterprise, and special issues of implementation and maintenance are covered.

This course will also be an introduction to Data mining (DM). Topics will range from statistics to machine learning to database, with a focus on analysis of large data sets. The concepts of DM such as Data Preparation, Classification, Prediction, Associated Rule Mining and Clustering will be covered with focus on applications to large real world data.

## **LEARNING OUTCOMES**

After finishing the course, the students should be able to:

- Build a dimensional model of a database for different types of DWs and applications.
- Use transactional, snapshot, and accumulative snapshot models, and their combinations in building a DW value chain.
- Build efficient models of complex dimensions:
  - Large
  - Slowly changing
  - Heterogeneous
  - Minidimensions.
- Build queries for drilling a DW
- Understand basic problems of DW management and participate in resolving these problems
- Apply special approaches to optimize the size and performance of a DW
- Build Data Mining models using a DM tool (WEKA).
- Understand various concepts of DM such as Classification, Prediction, and Clustering.

## **ASSESSMENT CRITERIA**

Students will be assessed in exams, homework, case assignments, and through class participation. The major areas include:

- The understanding of the dimensional model
- Designing a dimensional model for a case assignment. Choosing the appropriate model: transactional, snapshot or accumulative snapshot depending on the analytical needs
- Designing the facts of the dimensional model of a case assignment depending on the analytical needs
- Enhancing the dimensional model for a case assignment for support of large, slowly changing and heterogeneous dimensions
- Understanding new SQL features for building analytical requests and querying the dimensional model
- Building queries for analytical requests of a case assignment
- Describing problems of support of a data warehouse and loading it with data for a case assignment
- Implementing a prototype of a data warehouse according to the design for a case assignment, loading it with data from the data sources, and querying it
- Understand how data mining is applied on real world large data

## General Education Outcomes

- **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
- **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
- **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

## PREREQUISITES

Completion of CST3604/CS 604 with the grade C or higher.

## REQUIRED TEXTBOOKS

1. "*The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*" (Second Edition) R. Kimball, M. Ross. John Wiley & Sons, 2002
2. "*Data Mining, Concepts and Techniques*" (Third Edition) Han J., Kamber M, and Pei, J. , Morgan Kaufmann. 2012

### **Attendance Policy:**

You are permitted to be absent from a class a maximum of three class sessions. This is in accordance with the college policy that sets the maximum number of permissible absences at 10% of the number of class meetings scheduled for the semester.

### **Academic Integrity Policy:**

Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

### **Tests:**

Final Exam	30%
Midterm	20%
Tests & Quizzes	15%
Case Assignments	30%
Homework	5%

### **Case Assignments**

In the first assignment students have to build a DW model for a given business; the business is described by the model of an OLTP database. Correspondence between the OLTP database and data warehouse data items must be defined, as well as the procedure of the data load from the database to the data warehouse.

In the second assignment, the model from the first assignment is developed to support slowly changing dimensions, aggregated data, hierarchical dimensions and some other special situations. Students have to build SQL queries for some analytical requests.

The above two assignments are implemented in Oracle.

In the third assignment, a data mining model will be developed using WEKA (a data mining tool). Students will be required to use classification and clustering on the given data.

### **Grading Policy:**

You cannot get a passing grade unless all case assignments are completed. The professor preserves the right to ask you to defend any of your case assignments or tests.

## **COURSE OUTLINE**

*Students have to practice the discussed aspects of data warehouse design on additional cases.*

<b>Week</b>	<b>Subject</b>	<b>Source</b>
<b>1</b>	OLTP and OLAP systems- two different worlds. General architecture of a DW. Introduction to the Dimensional Model.	Chapter 1 Handout
<b>2</b>	Grocery Store Case. General Steps in DW Design. Fact Table. Dimensions.	Chapter 2  Handout
<b>3</b>	Warehouse Case. Inventory Model of Fact Table. Database Sizing	Chapter 3

<b>4</b>	3 Basic Models of Fact Table –Transactional, Periodic Snapshot, Accumulating Snapshot. Value Chain	Chapter 3, Handout
<b>5</b>	Test #1. Procurement Case. Slowly Changing Dimensions.	Chapter 4
<b>6</b>	Order Management. Degenerate Dimensions. The Demand Value Chain. The Supply Value Chain	Chapter 5
<b>7</b>	Customer Relationship Management. Big Dimensions. Drilling up and down. Multiple Hierarchies. Slowly Changing Dimensions. Slowly Changing Minidimensions	Chapter 6, Handout
<b>8</b>	Midterm Test. Financial Services Case. Dirty Dimensions. Heterogeneous Products. Different Fact Models for Heterogeneous Products. Minidimensions	Chapters 7, 9
<b>9</b>	Factless Fact Tables. Two types of Factless Facts. Coverage Tables	Chapter 12, Handout
<b>10</b>	Querying DW. Examples of queries	Handout
<b>11</b>	General Methodology of Building a Warehouse	Chapter 16, Handout
<b>12</b>	General Methodology of Building a Warehouse	Chapter 16
<b>13</b>	Data Analysis with Data Mining: What is data mining? What makes it a new and unique discipline? Relationship between Data Warehousing, On-line Analytical Processing, and Data Mining.	Textbook 2, Chapter 1-1.3
<b>14</b>	Data Mining concepts: Classification/Prediction and Clustering	Textbook 2, Chapter 7, 8
<b>15</b>	Review. Final Test.	Handouts