

Computing
Faculty of Arts, Science, and Technology
BSc (Hons) Computing, BSc (Hons) Computer Science

Level: 6
Module: COM641 Distributed Data & Data Analytics
Assignment: 1
Issue Date: 23rd September 2025
Review Date: Tutorial Sessions
Submission Dates: xxth November 2025
Estimated Completion time: 72 Hours
Lecturer: Russell Lo

Verified by:

To be completed by student:

I certify that, other than where collaboration has been explicitly permitted, this work is the result of my individual effort and that all sources for materials have been acknowledged. I also confirm that I have read and understood the codes of practice on plagiarism contained within the Glyndŵr Academic Regulations and that, by signing this printed form or typing my name on an electronically submitted version, I am agreeing to be dealt with accordingly in any case of suspected unfair practice. I also certify that my attendance for the module has been at least 70%

Name:

Student Number:

Date Submitted: xxth November 2025....

Student Signature:

Are extenuating circumstances being claimed? YES / NO

If YES, give reference number:

To be completed by lecturer

Comments:

Grade / Mark
(Indicative: may change when moderated)

COM641: Assignment 1

The assignment will entail reading a range of academic papers, journals and books. You will be writing reports and designing Distributed Database Systems and Data Warehouse Systems.

Tasks

Task 1: Distributed Database Design (30%)

Happy Cruise Lines is a cruise company with several ships and a variety of cruise itineraries, each involving several ports of call. The company headquarters is in New York and has regional offices in the cruise port cities of Miami, Houston, and Los Angeles.

The company maintains information on ships, cruises, ports, passengers and voyages.

CRUISE is a particular sailing of a ship on a particular date. VISITS records the ships visits to the ports during the cruise. VOYAGE records the passengers for a cruise. Given below is the snapshot of Happy Cruise Lines' database which lists the attributes for each entity along with the primary key and *foreign keys*.

SHIPS: (SHIPNUM, SHIPNAME, BUILDER, LAUNCHDATE, WEIGHT)

PASSENGER: (PASSENGERNUM, PASSENGERNAME, SOCIALSECNUM, STATE, COUNTRY)

PORT: (PORTNUM, COUNTRY, NUMDOCKS, MANAGER)

CRUISE: (CRUISENUM, STARTDATE, ENDDATE, DIRECTOR, SHIPNUM)

VISIT: (CRUISENUM, PORTNUM, COUNTRY, ARRDATE, DEPTDATE)

VOYAGE: (PASSENGERNUM, CRUISENUM, ROOMNUM, FARE)

SHIPS: The Company has 20 ships and the details of these ships are used in every office.

CRUISE: There are around 4000 cruise records; cruise records are used most heavily in the cities from which the cruise described in the record began.

PORT: There are 42 records in the port table. The records describe Atlantic Ocean ports that are used most heavily in New York and Miami. The records that describe Caribbean Seaports are used most heavily in Houston and Miami. The records that describe Pacific Ocean ports are used most heavily in Los Angeles.

VISIT: Consists of 15,000 records and is primarily used in the New York (headquarters) and Los Angeles.

PASSENGER: Consists of 230,000 records and is primarily used in New York (headquarters) and Los Angeles offices.

VOYAGE: Consists of 720,000 records and is used in all four offices.

- a. Write a report **evaluating** various **data distribution** options for Happy Cruise Lines over a Distributed Databases System. Your report should critically select an **appropriate data distribution solution** for Happy Cruise Lines with justifiable reasons.

20% (Approx. 750 words)

- b. Design a data distribution scheme in line with your solution

10% (Design Diagram)

Task 2: Data Warehouse Design (20%)

Millennium College offers several courses in different subject (Major) areas. Courses are managed by professors who are employed in different departments. The college records the course details, student details and their results. The college decides to implement a data warehouse. One of the requirements is to record grades for courses completed by students. Initial analysis identified the following dimensions:

CourseSection: CourseID, CourseName, Units, SectionNumber, RoomID, RoomCapacity. (During a given Semester the college offers an average of 500 course sections).

Professor: ProfID, ProfName, Title, DepartmentID, DepartmentName. (There are around 200 professors)

Student: StudentID, StudentName, Major. (Each course section has an average of 40 students, each student may take up to 5 course sections per period).

Period: SemesterID, Year (the College follows a trimester system, and it has records for the last 10 years.)

- a. Design a Star Schema for modelling the data warehouse for Millennium College. Ensure every dimension has an appropriate Primary Key. You must identify the FACT table with appropriate details (facts) and keys for the dimensions. Assign an appropriate name for the FACT table.
- b. Using the principles of normalisation convert the Star Schema to a Snowflake Schema.

Task 3: Procedure, Functions Triggers and Active Rules in Oracle 50%

Recently you have been appointed as a Database Administrator (DBA) in a university. As part of your job, you have to manage, monitor and log the database activities in their Student Management Database. The University needs to record the details of the students, courses, enrolment, grades, staff and sessions (sections). You must restrict data updates to working

hours to protect the data.

You may have to add more sample data to test the functionalities of your procedures and to produce meaningful test reports.

Task 3a: Grade Report:

The University needs to produce different reports related to the course, students and their performance, results etc.

- I. University has to produce the Final Mark sheet for students. This should display student details, description of the courses that a student has enrolled on, and Final Mark they have received for that course. Final marks must be calculated considering different types of assessment (quiz, examination, mid-term etc.), the number of assessments of the given type (3 quizzes) and their percentages towards the final grade.

10%

- II. The University likes to get statistical reports on courses which include the description of the course, sections start date, name of the instructor who manages the section, number of students enrolled for section, the highest mark, lowest mark, average mark, standard deviation, number of 1st (Final mark: 80 -100), 2nd (Final mark: 70 – 79), 3rd (Final mark: 60 – 69) number of failures (Final mark: 1 -59) etc.

10%

- III. University likes to perform location-based analysis on recruitment, student progress etc.

- a. Using the data from relevant tables, produce a report listing the number of students enrolled on specific courses, by state, city and postcode. You should be able to list popular courses by your preferred location details (i.e., by State, City or PostCode)
 - b. Produce a report detailing the results that the students achieved, display the results by state, city and postcode, You should be able to provide the highest performing course by state, city and postcode area as requested by the University.

(a & b - 20%)

Create appropriate PL/SQL procedures to produce the above-listed reports. Please

ensure that all necessary details are included and displayed in the report.

Task 3b: Active Rules Using Triggers

- I. The University student data centre that is responsible for entering student details and enrolment details works from Monday 9.00 am – Friday 5.00 pm. The system must be protected (don't allow any changes on appropriate tables) and only allow minimum operations from Friday 5.00 pm to Monday 8:00 am. For example, students' details or enrolment details are only added or amended during office hours. Database activities are monitored all the time, create an active rule to stop such activities and display appropriate messages.

5%

- II. University can only accommodate a limited number of students for each section. Create an active rule to display an appropriate message on the screen to warn the DBA while the enrolment is nearing its capacity and also when it has reached the capacity limit.

5%

Guidance

Students will get assistance to complete the tasks through the tutorial sessions. **Drafts will be reviewed and formative feedback will be given in the tutorial sessions.** So you are much less likely to obtain a good grade if you don't attend the tutorial sessions.

Support your work with relevant references from books, journals and other quality information resources using the IEEE Referencing system.

Failure to complete any of the above tasks or sub-tasks in time will result in a loss of marks.

Submission

The assignment should be word-processed. This specification document should be filed at the front of the assignment, with the front sheet (with your Name, Student Number, Date and Signature) visible at the front.

Work must be word-processed and submitted as a word file consisting of all Tasks via Turnitin link provided through VLE (Moodle) by 30th November 2025. The Glyndwr policy on assignment submission will be rigidly adhered to (see your Student Handbook).

Task 1 & 2:

Report and diagram – Word processed

Task 3:

- PL/SQL Scripts for procedures and Triggers
- Test scripts (execution blocks for Procedures, SQL statements for Triggers) to test the procedures and triggers.
- Test results, take screenshots where necessary.

Also ensure the required tables, data, procedures, functions, cursors and/or triggers are all saved into a separated file (database script file) and handed into Moodle together with your word file. (Use your student ID to create database schema in your script)

Learning Outcomes

1. Critically assess some of the more advanced developments in database technology.
e.g. Stored Procedures and Functions
2. Evaluate the current issues associated with theory to practical implementations in database research.
3. Explore advanced aspects of data warehousing, distributed data, data science and data analytics encompassing the principles, research results and commercial application of the technologies.

Key skills for employability:

1. Written, oral and media communication skills
2. Leadership, team working and networking skills
3. Opportunity, creativity and problem-solving skills
4. Information technology skills and digital literacy
5. Information management skills
6. Research skills
7. Intercultural and sustainability skills
8. Career management skills
9. Learning to learn (managing personal and professional development, self-management)
10. Numeracy

Assessment Criteria

In order to achieve an **A** grade, the work must be excellent in almost all respects, with only very minor limitations.

In order to achieve a **B+** grade, the work should show strength in most respects, but perhaps has limitations in one or two areas. A good piece of work nevertheless.

In order to achieve a **B** grade, the work should be of a satisfactory standard, showing strength in some areas, but perhaps let down by poor presentation, poor practical work, or poor written explanations where required.

In order to achieve a **C** grade, the work should be of a satisfactory standard but may have significant shortcomings in some areas. Nevertheless shows at least a basic understanding of the concepts and a basic practical ability.

A **Refer** grade will be given to work that is just unsatisfactory