

# STW220CT: Data and Information Retrieval Coursework

This is your 220CT coursework. There are 4 tasks in this Coursework. All Tasks are mandatory. All submissions through Moodle.

## Task 1: Database design and Development (30 Marks)

### a. Introduction

Condor Building Services Ltd is a long-established construction supplier for construction companies based in Coventry and the surrounding area. The company specialize in supplying construction equipment such as piping, timber, wiring, plumbing UPVC frames and doors and masonry. As construction demand is slowly picking up Condor has regular orders for equipment to meet the needs of its client base. To support this, an orders book, **see table below**, is maintained in which equipment orders for its clients are recorded. For each order the client's details, date, equipment, quantity, unit price and overall price of the order is logged.

Order ID.	Supplier ID.	Client Name	Client Address	Date	Equipment	Qty	Unit Price	Total Price
CON-2237	168	Coventry Building Services Ltd	Units 2-4, Binley Industrial Estate, CV3 2WL	14/12/2014	Butterfly valve	2	£5.00	£99.00
					¾" Locknut	6	£1.50	
					Sch 40 Blk Pipe	4	£20.00	
CON-3664	527	Allied Construction Ltd	34, Lythalls La Industrial Estate, CV6 6RG	16/01/2015	Thin Stranded Copper Wire	6	£6.00	£36.00
CON-2356	169	Ricoh Builds Ltd	Unit 12, Stoneleigh Park, CV8 2UV	12/02/2015	Sch 40 Blk Pipe	3	£20.00	£280.00
					4x8x3/4 Cos Plywood	2	£10.00	
					¾" EMT	2	£50.00	
					Duplex Ivy Rec	1	£100.00	
CON-1234	032	Grand Designs Ltd	32-34, Bilton Industrial Estate, CV3 5YB	16/04/2015	Sch 40 Blk Pipe	1	£20.00	£23.00
					¾" Locknut	2	£1.50	

Currently there is no database being used for managing the Order information in the table above.

This task is split up into two parts:

1. In its current form, it's a traditional DB. Keep it that way? Your call. Explain your decision. [3 Marks]
2. Design the database for the information above. Implement the DB using any method of your choice (SQL, MongoDB, Cassandra DB or Graph DB). [12 Marks]

**Note:**

- If it is a relational database, it should be normalized first.

**b. Using the data above you are required to perform the following task**

1. Display all the client details with its order. [1 mark]
2. Display all the order details with its client name and equipment associate with it. [2 marks]
3. Display all the client details whose total equipment's quantity is more than 6. [3 marks]
4. Display all the equipment ordered between 01/11/2014 and 01/03/2015. [2 marks]
5. Display all the equipment associate with the client named 'Coventry Building Services Ltd'. [2 marks]
6. Delete '¾" Locknut' from order id CON-1234 and put it in order id CON-2356 and display all the order details with its equipment associate with it and the total price. [3 Marks]
7. Delete all the order between 01/11/2014 and 20/01/2015 and display all the order details with its client name and equipment. [2 Marks]

**Evidence**

No marks will be awarded for statements that have only been word-processed. You are required to run your SQL/No-SQL statements and to include each output immediately after its statement. You should have a brief description of the SQL/No-SQL statements and the outputs.

**Task 2: Graph DB Design (10 Marks)**

**Scenario:**

Two people, John and Sally are friends since 12-01-2016. Both John and Sally have read the book Moby Dick written by Herman Melville. John's date of birth is 06-12-1990 and Sally's dob is 12-04-1991. John read the book on 2/03/2021 and Sally read the book on 2/02/2021 and John gave the rating of 5 stars and Sally gave 4 stars.

1. Identify the nodes, relationships and create the graph. [5 marks]
2. How old Sally is? [1 mark]
3. Who read Moby Dick first, Sally or John? [1 mark]
4. Remove the relationship between Sally and the book. [1 mark]
5. Update Sally's name to Salena. [1 mark]

6. Delete all the nodes and relationships except John.

[1 mark]

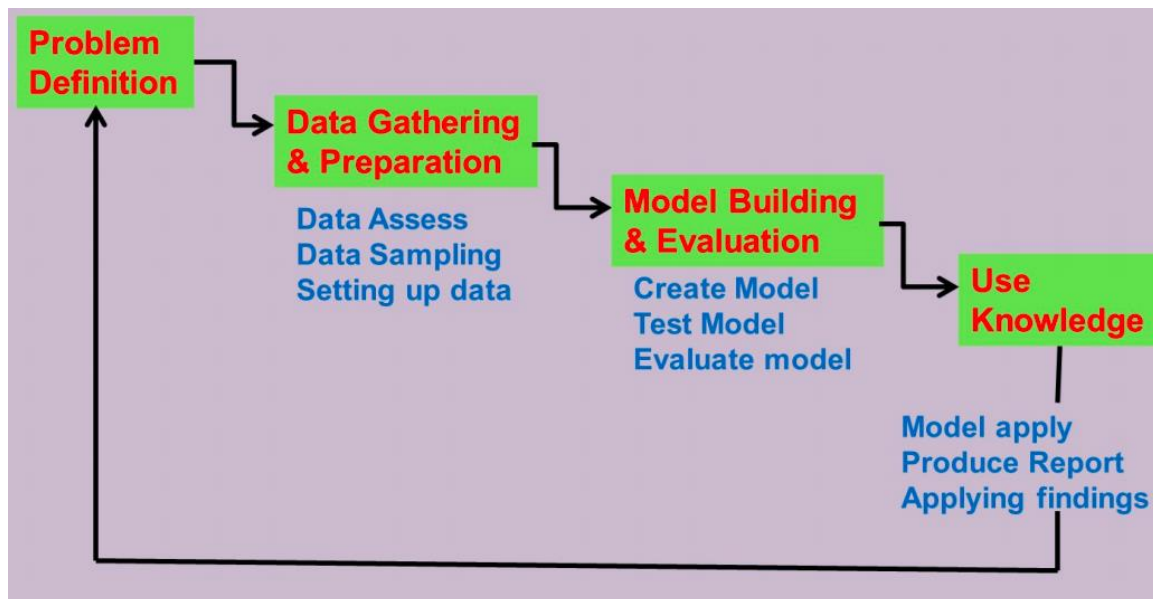
### Task 3: A data mining system for a Hospital (25 Marks)

A hospital has been collecting a great deal of data on their patients and have heard that use of data mining could improve their service. They would like you to create a brief report that includes the following:

- i. What data mining is and an appropriate data mining application for the Hospital. [5 Marks]
- ii. How you would go about creating the system using the data mining life cycle below. [5 Marks]
- iii. If the small amount of data (diabetes.arff) collected so far by the hospital is appropriate for assessing if a person has diabetes. [5 Marks]
- iv. The use of a data mining model such as a multilayer perceptron or decision tree to determine whether a person has diabetes. [10 Marks]

**Note:**

- You will need to use a data mining tool like **WEKA** to create your model and use the diabetes.arff data to train and test this model.



**Deliverable:**

Include a report section that addresses the four sections above and fulfils the marking criteria.

#### **Task 4: Recommendation System (35 Marks)**

You are asked to build a 'recommendation' system for any eCommerce website. You should tailor each recommendation to the customer, their previous purchases and other relevant information that you can use to identify the interest of this customer. Using what you've learnt in the module, propose a complete solution for this that includes Data Storage, collection and analysis and recommendation type and explanation on how the solution will be implemented.

#### **Evidence**

Include a report section that explores the issues above no more than 1200 words. You should carry out research into these areas and reference your work using the CU Harvard Style.