CS27020 Modelling Persistent Data

Study Guide 2022

Introduction

This study guide is intended to help you navigate the online resources available for the module and to help you stay on top of the learning materials. The guide is broken down into topics, you can tick off those activities completed in the following sections and then return to this page and record the completed topics.

Topic checklist

During this module you will cover the following

Week completed		Completed
1	Introduction	☐
2	Theory	☐
5	SQL	☐
6	Integrity	☐
7	Normalisation	☐
9	Persistence	☐
11	Semi-structured data	☐
11	No SQL	☐

Topic 1

Introduction

The first lectures give an overview of the module, including the timetable and assessment. It is important that you understand how you will be assessed and when work is due to avoid losing marks unnecessarily. Changes from previous years will also be communicated, which can help you avoid spending time learning topics no longer on the module, or avoid missing newly introduced topics.

Watch the first lecture part 1.

The Blackboard site for the module includes the lecture slides, lecture recordings, enrichment material, discussion board and assessment information and submission (quizzes, worksheets and exam alternative). Make sure you familiarise yourself with where everything is.

Look at the Blackboard website for the module:
https://blackboard.aber.ac.uk/webapps/blackboard/content/listContentEditable.jsp?content_id=_2186671_1&course_id=_33288_1

Discord will be used as a community space for the module, and for online help and support. It may also be used for practical lab sessions if we need to go online again.

Join the discord server https://discord.gg/eb8NaQK9Rp

In the second part of the introductory lecture you will be given an overview of the technical content of the module.

Watch the first lecture part 2.

<table>
<tr><td>

Topic

2

</td><td>

Theory

</td></tr>
</table>

The theory lectures give you the necessary background that underpins the more applied topics.

Lecture 1 on the Relational Data Model (RDM) focuses on data objects and answering the question, "how do we store data in the RDM?". Also covers terminology and the three crucial properties of relations.

> Watch the lecture on RDM and data objects

Lecture 2 on the RDM will look at answering the question, "How do we uniquely identify the data?". Provides the foundation for a lot of the work that we will look at when considering Data Integrity.

> Watch the lecture on RDM keys

Lecture 3 on the RDM examining data integrity; specifically, entity integrity and referential integrity. These aspects are important to ensure that our database correctly models the real world.

> Watch the lecture on RDM data integrity

Lecture on Manipulating Data in the RDM. Covers the key operations defined by the relational algebra along with examples of each. This lecture builds the bridge needed to cross from theory to practice.

> Watch the lecture on RDM data manipulation

The final lecture in this topic is the theory recap lecture.

> Watch the theory recap lecture

In addition to the watching the lectures there is additional information in the book chapters. Chapters 4 and 5 of Connolly and Begg (both the 6th and the 5th editions), cover The Relational Data Model and Relational Algebra (and briefly Relational Calculus which we do not cover in the lectures). Chapter 2 of Takahash and Azuma provides a superficial introduction to the relational model and relational operators.

> Read the suggested book chapters.

There are also, papers and websites in the enrichment folder on BlackBoard that can provide additional information and background.

> Read the suggested book chapters.

At the end of this topic there will be a quiz on BB, it is worth up to 2% of the module mark.

> Terminology quiz: what is what in relational database concepts.

Topic 3

SQL

The previous topic introduced the Relational Data Model (RDM). One of the most widely used implementations of the RDM is in the Structured Query Language (SQL). We will use a combination of lectures, lab sessions and enrichment material to cover SQL. One of the best ways to learn SQL is by using it interactively, which is highly recommended.

The first lecture introduces you to the basics of SQL in preparation for the lab sessions.

> Watch the introductory SQL lecture.

In order to improve and embed your SQL skills covered in the introductory lecture, the interactive lessons at codecademy are highly recommended.

> https://www.codecademy.com/en/courses/learn-sql/

In the first 2 practical sessions you will be asked to implement some SQL commands to demonstrate your understanding of database creation commands. This practical is worth 4% of your mark.

> Attend the practical sessions (weeks 2 and 3)

The second SQL lecture will build on your understanding, covering more advanced topics such as joins and NULLs.

> Watch the advanced SQL lecture.

Work through the tutorial at w3schools to prepare for the mini-quiz.

> https://www.w3schools.com/sql/

The second mini-quiz will test your understanding of SQL.

> SQL mini-quiz.

The second SQL practical will assess your understanding of some more advanced SQL concepts.

> Attend the practical sessions (weeks 4 and 5)

Topic 4

Integrity, transactions and concurrency

This topic will discuss using SQL's functions and triggers to help enforce integrity checks, timestamping, and dealing with concurrent database access issues.

The first lecture covers domain constraints (such ensuring a price is positive) and referential integrity (ensuring that foreign keys reference a valid primary key in the parent table).

Watch the introductory integrity lecture.

The second lecture looks at some difficult cases, and introduces methods for handling them, including triggers and stored procedures.

Watch the lecture on triggers and stored procedures.

The preceding lectures picks up on material covered in Chaper 7 of Connolly and Begg, 6th edition up to section 7.4.

Read Connolly and Begg chapter 7 up to section 7.4.

The next lecture introduces the concepts of transactions and concurrency.

Watch the first lecture on transactions and concurrency.

The second lecture on concurrency investigates problems and how to avoid them.

Watch the second lecture on transactions and concurrency.

The third and last lecture on concurrency revisits the ACID properties (Atomicity, Consistency, Isolation and Durability) and focusses on Isolation and Durability in the context of concurrent transactions.

> Watch the second lecture on transactions and concurrency.

The preceding three lectures covers material in Section 7.5 of Connolly and Begg, 6th edition and then the material in Chapter 22 up to section 22.3. Later sections in that chapter look at more advanced transaction models and at the facilities provided by the Oracle DBMS and may help to consolidate understanding of the earlier concepts.

> Read Connolly and Begg section 7.5 and chapter 22 up to 22.3

There will be a final recap lecture covering integrity, transactions and concurrency.

> Watch the integrity recap lecture.

There will be a Blackboard mini quiz at the end of this topic, it is worth up to 2% of the module mark.

> Take the integrity mini-quiz

Topic 5

Normalisation

Normalisation is an approach to modifying the design of a database in order to prevent redundancy and inconsistencies occurring in the stored data. It involves applying certain tests and, if those tests fail, tables are split into a new set of tables that pass the test.

The first lecture will introduce the subject, explaining the problems that normalisation is designed to solve, giving a recap on the different kinds of keys, and then working towards first normal form (1NF).

> Watch the first normalisation lecture.

Chapter 14 of the main course text (Connolly and Begg, 6th edition) covers normalisation up to third normal form (3NF). Reading it in conjunction with the lectures will help your understanding.

> Read chapter 14 from Connolly and Begg

The second lecture introduces functional dependencies and takes normalisation to second normal form (2NF).

> Watch the second normalisation lecture.

In the third normalisation lecture we are going to introduce Transitional Functional Dependencies and look at bringing relations to third normal form (3NF).

> Watch the third normalisation lecture.

Chapter 15 section 1,2 and 3, of Connolly and Begg (6th edition) covers normalisation up to Boyce-Codd normal form (BCNF). Reading it in conjunction with the lectures will help your understanding.

Read chapter 15 sections 1-3 from Connolly and Begg

You may also want to watch some videos and look at websites to help you understand BCNF. For example, watch a short video on BCNF

Watch short video on BCNF:
https://www.youtube.com/watch?v=NNjUhvvwOrk

The final lecture on normalisation covers up to BCNF and reviews the normalisation topic.

Watch the fourth and final normalisation lecture.

After the normalisation lectures, the third practical will be a practical exercise on normalisation, worth 4% of the module mark.

Attend the practical sessions (weeks 6 and 7)

<table>
<tr><td>

Topic

6

</td></tr>
</table>

Java Persistence

We often want to integrate a database into a system with other components, usually written in a high-level language, such as Java. There are various ways to do this, and we will be looking at a few in this topic.

The first lecture will introduce the topic and discuss using a Call Level Interface (CLI) to access a database by using appropriate calls to establish a connection, issue commands and receive a response. It will then move on to discussing Object DataBase Connectivity (ODBC) and specifically the Java CLI: JDBC.

Watch the first Java Persistence lecture.

The material in these lectures is covered in Section 29.7 of Connolly and Begg, 6th edition. It is presented in the context of Web applications, but the concepts are applicable to any application, with the exception of sections 29.7.4 and 29.7.7 onwards.

Section 29.8 covers Microsoft provision and provides a wider perspective on the issues.

Read sections from Connolly and Begg

The next lecture introduces the Java Persistence API and discusses entities for Object-Relational mapping.

Watch the first Java Persistence API (JPA) lecture.

The next lecture looks at the Entity manager, the Java Persistence Query Language (JPQL) and the Criteria API for writing queries.

Watch the second Java Persistence API (JPA) lecture.

The third JPA lecture looks queries and models using JPQL.

Watch the third Java Persistence API (JPA) lecture.

The enrichment material on Blackboard includes an example project and link to the Oracle's JDBC tutorial.

Explore the enrichment material on BlackBoard.

There will be a Blackboard mini-quiz on completion of this topic, it is worth up to 2% of the module marks.

Take the Blackboard mini-quiz on Java Persistence and JDBC.

Attend the practicals on Java persistence (weeks 8 and 9)

Topic 7

Semi-structured data and XML

There is a lot of data that is not stored in a well formatted database, but is in a less precisely structured storage format, such as web pages, MS Word documents, etc. It is often useful to be able to search and query such documents. This topic will provide an introduction to semi-structured data, focussing on XML, and the methods and tools for working with it.

The first lecture will introduce the eXtensible Markup Language (XML) and the general concept of semi-structured data (SSD).

> Watch the introductory lecture on XML and SSD

Work through the XML worksheet to embed and test your understanding.

> XML worksheet (unassessed)

XPath, XQuery are XSLT are the tools we'll be looking at for interacting with XML data. Xquery and XSLT depend on XPath, so we'll start by looking at XPath.

> Watch the XPath lecture

Work through the XPath worksheet to embed and test your understanding.

> XPath worksheet (unassessed)

The next lecture delves into XQuery, which is a superset of XPath and provides the ability to query XML data.

Watch the XQuery lecture

Work through the XPath worksheet to embed and test your understanding.

XQuery worksheet (unassessed)

XSLT allows you to format and transform XML data. The final lecture in this topic will look at the powerful capabilities of XSLT.

Watch the XSLT lecture

Chapter 30 of the main course text, Connolly and Begg (6th edition), covers the material in this topic.

Read chapter 30 of Connolly and Begg

There will be a Blackboard mini-quiz on this topic, which you should take after watching the lectures, working through the unassessed quizzes and reading the book chapter. It is worth up to 2% of the module mark.

Take the XML and SSD mini quiz on BlackBoard

The final practical lab session of the module will be on XML and SSD. Make sure you attend and complete the assessed work, it is worth 4% of the module mark.

Attend the practicals (weeks 10 and 11)

<table>
<tr><td>

Topic

8

</td></tr>
</table>

NoSQL

NoSQL (Not Only SQL) is not really a single thing (at all). It's more a collection of systems which have some properties in common, such as: persistence and data storage; some kind of data model (but probably not full SQL); scalability in particular, horizontal scaling and partition tolerance; Speed; and Flexibility. This topic will introduce NoSQL concepts and implementations.

The first lecture introduces the idea of NoSQL, explores some implementations, and works through a case study of MapReduce.

> Watch the NoSQL introductory lecture

The next lecture looks at Key-Value pair systems and BigTable systems.

> Watch the Key-Value and BigTable lecture

The following lecture covers graph/triplestore and document oriented systems.

> Watch the Key-Value and BigTable lecture

The final lecture slides covers tools, future directions and gives a recap of NoSQL.

> Watch the Key-Value and BigTable lecture

This topic does not really have a corresponding Connolly and Begg reference, as the Connolly and Begg approach to NoSQL is not very strong. NoSQL for Dummies is quite a clear intro to NoSQL (and written by an Aber Grad, to boot). Chapters 1, 2, 3, 4, 6, 9, 14 and 19 are all relevant to this topic.

Read (parts of) NoSQL for Dummies

There will be an BlackBoard mini-quiz at the end of this topic, it is worth up to 2% of the module mark.

Take the NoSQL mini quiz on BlackBoard

Key to activity icons

	Activity – for example practical exercises
	Discussion
	Essential reading
	Investigation – looking for an answer
	Listen – to a podcast
	Quiz
	Recommended reading
	Watch – lecture video
	Web

In addition, this Warning icon is used to highlight areas where if you encounter any difficulties you should get in touch straight away.