

COSC2406/2407 Database Systems

Assignment Project Exam Help

Xiangmin (Emily) Zhou

<https://powcoder.com>

RMIT University

Room : 14.11.04

Online Consultation via Collaborate Ultra(no appointment required): 10:20-11.20am

Thursdays

Email : xiangmin.zhou@rmit.edu.au

Add WeChat powcoder

Course Overview

Acknowledgement of Country



Assignment Project Exam Help

<https://powcoder.com>

Ngarara Place, RMIT City campus

RMIT University acknowledges the Kulin Nations as the Traditional Custodians of the land on which the University stands. The University respectfully recognises Elders both past and present. RMIT also acknowledges the Traditional Custodians of lands across Australia where it conducts its business, their Elders, Ancestors, cultures and heritage.

Emergency Procedure Information

Before starting our session today, RMIT wishes to make you aware of the Emergency Procedures that are in place for your safety. Please look around and familiarize yourself with the emergency exits in this room and *do this for each new venue in which you attend classes.* Evacuation plans are located in all corridors.

- In the unlikely event of an evacuation, a “beep, beep” tone will sound.
On this alert signal you should prepare yourself to evacuate the building.
- If the “whoop, whoop” tone sounds, please evacuate in an orderly fashion to the assembly area indicated on the evacuation plans located in the foyer of this building. *Do not use lifts when evacuating the building.*
- Please also obey any instructions provided by fire wardens in attendance, by RMIT Security or by emergency services. At the end of the evacuation, you will be advised when it is safe to return to the building.
- RMIT thanks you for your cooperation.

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

- Databases are still a hot field.
- DBMS developers are in demand and there are many challenging unsolved problems in data management and processing.
- If you are good enough to write code for a DBMS, then you can write code on almost anything else.

<https://powcoder.com>

Pre-requisite Courses, Assumed Knowledge and Capabilities

- Fundamental knowledge of relational database concepts, including relational modeling and design, relational algebra, SQL, and Conceptual modelling, equivalent to ISYS1055/1057 Database Concepts.

- Extensive programming skills in Java language and advanced knowledge of data structures and algorithms, equivalent to COSC1285/2123 Algorithms and Analysis.

Add WeChat powcoder
Warning: If you have NOT passed both Database Concepts AND Algorithms and Analysis (or the equivalent elsewhere in previous studies) then you should WITHDRAW from this course **NOW!**

The assignment work in this course assumes knowledge of database concepts and will require advanced programming skills to implement complex data structures and algorithms.

Assignment Project Exam Help

Important dates for course change in Semester 1, 2021 are:

- 8 March, 2021 - last day to add a course/s.
- 31 March, 2021 - census date. This is the last day to drop course/s without financial penalty and closing date for refunds for cancelled enrolments.
- 15 March, 2021- myTimetable closes for allocation/adjustment - Last day to allocate or adjust classes (by 5pm)
- 30 April 2021 - last day to drop course/s without academic penalty (fail grade). Financial penalty still applies.

Add WeChat powcoder

Assignment Project Exam Help

After completing COSC2406/2407 you should be able to:

- Explain the implementation of the internals of database systems
- Evaluate and compare alternative designs for data models, file structures, index schemes, and query evaluation
- Analyse techniques used for transactions, concurrency control, and recovery
- Design and implement significant software components of a database system (such as file structures and index schemes)

<https://powcoder.com>

Add WeChat powcoder

Course Overview

Lecture Topic

- ① Introduction to course. Structured, semi-structured & unstructured data
- ② Disks and files
- ③ Files, pages and records
- ④ File organisations and indexing
- ⑤ Tree index structures
- ⑥ Hashed indexes
- ⑦ Index construction and sorting
- ⑧ Query evaluation
- ⑨ Query optimisation
- ⑩ Transactions, concurrency and recovery
- ⑪ Information retrieval
- ⑫ Revision and review

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Classes each week

- lectorials start in week 1

- tutorials and laboratory classes start in week 2

Assignment Project Exam Help

- 1 hour online Lectorial
- 1 hour online Tutorial (sign up using the online myTimetable): it is assumed that you will have attempted the tutorial questions before each workshop, and that you will be prepared to discuss your answers.
- 1 hour Laboratory class (sign up using the online myTimetable)
- 1 hour online consultation.
- *myTimetable*
<https://mytimetable.rmit.edu.au/even/student>
- If you have problems with your timetable, contact RMIT Connect
<https://rmit.service-now.com/connect/> online, by phone.

Add WeChat powcoder

From Week 2 onwards we will be using Linux machines in the Amazon AWS cloud as our main teaching environment

Assignment Project Exam Help

- Every student in the course will be assigned an individual machine on which to install database systems and implement database components in Java
- You will be given a key that will grant you root access to your specific machine
- You will be responsible for securely maintaining your own key and your machine
- Keys MUST NOT be shared between people

<https://powcoder.com>

Add WeChat powcoder

Assessment Tasks:

- Weekly tasks - Formative weekly online exercises (15%) to be completed by 11:59pm on Tuesday in weeks 3 to 10 (6 of 8 best results will be counted for final marks.)
- Assignment 1 (20%).
 - on a linux VM you will evaluate and compare alternative implementations of database systems and use Java to implement a Heap File
 - released in early week 2 (Monday)
 - final submission due 11.59pm on Sunday 4 April 2021
- Assignment 2 (45%).
 - building on your solution to Assignment 1: design, implement, critically analyse and repair of the indexing component of database systems
 - released in week 6
 - final submission due 11.59pm on Sunday 23 May 2021
 - demo and code walkthrough in week 12
- Assignment 3 (take-home 24 hour exercise in Week 14) (20%) covering all aspects of the course

Academic integrity – avoiding plagiarism

The following resources are available online:

- # Assignment Project Exam Help
- *Study skills and plagiarism resource booklet* – advice for those beginning a program of study at the School of Computer Science and IT
 - *Avoiding plagiarism for code* – Examples of how to reference computer code correctly
 - *Academic integrity workshop slides*

<http://www1.rmit.edu.au/browse;ID=r7h4nkq7a9pg>

Add WeChat powcoder

Read more about academic integrity here:

<https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity>

Assignment Project Exam Help

- See someone immediately if you have any difficulties that may prevent you from completing the course
- Be aware of RMIT special consideration process:
<https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/assessment/special-consideration>
- Let the lecturer know if you may be delayed in submitting an assignment or other assessable material (we will be much more sympathetic if given warning—expect a cold reception if you contact us a day or two before the due date)
- Give us feedback about tutors/lab assistants, lab sessions, tutorials and lectures

Add WeChat powcoder

Available through canvas: <https://rmit.instructure.com/>

- Learning Resources
 - Readings for each week
A preliminary version of the complete course notes prepared prior to the start of the semester
 - Course notes for each individual lecture (which may include some revisions) will be made available weekly
 - Tutorial questions
 - Laboratory exercises

- Assessment Tasks (to appear) - Formative weekly online exercises available each Friday (starting Friday in week 3), and assignment specifications
- Announcements
- Discussion Forum - be sure to read these forums regularly and ask questions and contribute to the discussion (but do not post solutions to assessment tasks).
- Actions from student surveys (to appear)

- Weekly (pre-lecture) reading material will be available online, you should read this before attempting the weekly formative weekly online exercises

Assignment Project Exam Help

- The following textbooks covers many of the topics in this course:
 - R. Ramakrishnan and J. Gehrke, *Database Management Systems*, McGraw-Hill, 2003 (3rd Edition).
 - H. Garcia-Molina, J. Ullman, and J. Widom, *Database Systems: The Complete Book*, Prentice Hall, 2008 (2nd edition).
 - R. Elmasri and S.B. Navathe, *Database Systems*, Pearson, 2011 (6th edition).
 - P.J. Sadalage and M. Fowler, *NoSQL Distilled: A Brief Guide to the Emerging World of Polygot Persistence*, Addison-Wesley, 2013.
- References will be given to each of these books, though we will be mostly following Ramakrishnan and Gehrke.

- **Lecturer** (Xiangmin Zhou)

Building 14, Level 11, Room 04

- Consultation: 10:20-11.20am Thursdays

- email: xiangmin.zhou@rmit.edu.au

- **RMIT Connect** is where you access student administration and support, plus work and study opportunities.

• **Online** <https://rmit.service-now.com/connect/>

• **Phone** 9925 5000 (Monday to Thursday: 9am-5pm and Friday: 10am-5pm)

• Visit Building 10, Level 4

- **College Academic Student Services (Science team)** support the student lifecycle

- Building 10, Level 9

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

• Online <https://rmit.service-now.com/serviceandsupport/>, if you have a problem with your AWS Linux VM make sure you include in the Subject field:

"COSC2406 - Database Systems - Student VM Issue - <VM ID>"

- Call ITS on +61 3 9925 8888 (8am to 8pm Mon-Fri, 8.30am to 4.30pm)
- Visit a walk up support area on campus, including:
 - Swanston library Building 10 Level 6 - Monday to Friday 10.00 am to 6.00 pm
 - Swanston Academic Building (SAB) - Building 80 Level 3 - Monday to Friday 8.00 am to 8.00 pm

Add WeChat powcoder



Assignment Project Exam Help

Does it sometimes feel like you're
the only girl in your classes?

We know what that's like!

Join the RMIT Society for Women in Information TeCHnology (SWITCH) and become part of a supportive and welcoming community of female students studying and interested in all things tech!

Add WeChat powcoder
email RMITSWITCH@gmail.com to join

All female and genderqueer students including international, domestic, undergrad and postgrad students are welcome to join!



Acknowledgements

Material in this lecture series is collected from various sources, including:

Assignment Project Exam Help

- Lecture materials from: Falk Scholer, James Thom, Hugh Williams, and Justin Zobel, and minor changed by Xiangmin (Emily) Zhou.
- Additional material from: Bodo von Billerbeck, Steffen Heinz, M.V. Ramakrishna and Santha Sumanasekara, Audrey Tam.
- Supplementary material from: Raghu Ramakrishnan and Johannes Gehrke.
- Input from other staff.

<https://powcoder.com>
Add WeChat powcoder

COSC2406/2407 Database Systems

NoSQL databases

Structured, semi-structured and unstructured data

Assignment Project Exam Help

Xiangmin (Emily) Zhou

<https://powcoder.com>

RMIT University

Room : 14.11.04

Online Consultation via Collaborate Ultra(no appointment required): 10:20-11.20am

Thursdays

Email : xiangmin.zhou@rmit.edu.au

Add WeChat powcoder

Lecture 1

Reference: Pramod J. Sadalage and Martin Fowler,

NoSQL: A Brief Guide to the Emerging World of Polygot Persistence,

Addison-Wesley, 2013. Chapters 1, 2, 3, and 9.



Assignment Project Exam Help

Many different data models for database systems, including:

- network
- hierarchical
- relational
- deductive
- nested relational
- object-oriented

<https://powcoder.com>

Add WeChat powcoder

From the 1980s, relational data model became the dominant model.

- Persistence: Database systems are all about managing persistence and using a relational database makes it easy to get at and manage small bits of persistent data within large amount of persistent data (compared with using a file system)
- Concurrency: difficult to program, relational databases provide transactions which makes it easier to get correct
- Integration: allows multiple application programs to easily share the same data
- Standard model: (well mostly standard) data model and query language (SQL)

Add WeChat powcoder

Impedance Mismatch

Impedance mismatch refers to difference between

- relational model
- in-memory record structures

Assignment Project Exam Help

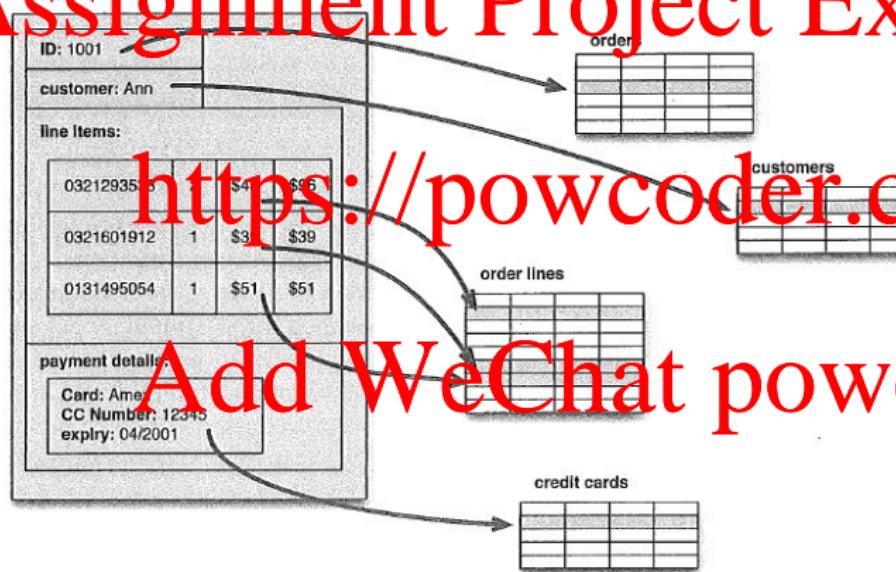


Figure 1.1 An order, which looks like a single aggregate structure in the UI, is split into many rows from many tables in a relational database

Assignment Project Exam Help

With rise of web services - integration and interoperability shifts from database to web services (“service oriented architecture”), and from SQL to HTTP (either XML or JSON).

- integration database - database acts as integration mechanism for multiple applications (integrity must be built-in to database)
- application database - database accessed only by a single application (integrity can be maintained by application programs)

Add WeChat powcoder

Assignment Project Exam Help

With clusters relational databases:

- not designed to run on clusters, even though some products such as Oracle RAC or Microsoft SQL Server provide cluster solutions - but with single point of failure
- could split data with sharding, but problems with querying, referential integrity, transactions and consistency across shards

Alternative to use clusters to deal with large volumes of data and traffic:

Add WeChat powcoder

- e.g. Amazon and Google

Big Data (3 key Vs):

Assignment Project Exam Help

- Volume — amount of data
- Velocity — new data/bandwidth required

Other Vs:

- Validity — only draw correct inferences from data (correlation vs causation)
- Veracity — how to assess accuracy of (say) a Twitter feed
- Value — from a business perspective What are costs and benefits
- Visibility — technology to make data from disparate sources visible

<http://rob-livingstone.com/2013/06/big-data-or-black-hole/>

Assignment Project Exam Help

- NoSQL = "not only SQL" (one common interpretation)
- refers to one current direction of databases and database technology

<https://powcoder.com>

Two main reasons for *considering* NoSQL instead of relational database

- scale of data requires clusters
- reduce effort mapping between database records and in-memory data structures

Add WeChat powcoder

Assignment Project Exam Help

Polyglot Persistence:

- relational databases are not the only storage option
- can consider alternative storage options provided by different types of NoSQL databases that include:

<https://powcoder.com>

- Document Store
- Key-value Store
- Column-family Store
- Graph Store

Add WeChat powcoder

Assignment Project Exam Help

Used in Document Stores, Key-value Stores, and Column-family Store

- relational model stores records corresponding to rows in relational table
- aggregate-orientation recognizes need to operate on data with complex structure

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

JSON example, from page 16, Sadalage and Fowler:

```
// in customers
{
  "id": 1,
  "name": "Martin",
  "billingAddress": [{"city": "Chicago"}]
}
```

<https://powcoder.com>
Add WeChat powcoder

Aggregate-oriented data models

JSON example continued, from page 16, Sadalage and Fowler:

```
//in orders
[
  {
    "id": 99,
    "customerId": 1,
    "orderItems": [
      {
        "productId": 27,
        "price": 32.45,
        "productName": "NoSQL Distilled"
      }
    ],
    ...
  }
]
```

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Consequences (See Section 2.1.2)

- relational database are aggregate-ignorant
- aggregates good for running on clusters
- but don't support ACID transaction
- but do support atomic consistency of a single aggregate

<https://powcoder.com>
Add WeChat powcoder

Use ID to get to aggregate data, but can also query and index content of aggregate.

Assignment Project Exam Help

Use Cases:

- event logging
- content management systems, blogging platforms
- web analytics
- e-commerce applications

<https://powcoder.com>

When not to use:

Add WeChat powcoder

- complex transactions spanning different structures
- queries against varying aggregate structures

Pairs of keys (usually strings) and values (which can be strings or primitives such as integers).

Access to aggregate value is via key, but content of aggregate is "opaque" to database.

Assignment Project Exam Help

Use Cases:

- storing session information
- user profiles, preferences
- shopping carts

When not to use:

Add WeChat powcoder

- relationships among data
- multi operation transactions
- query by data
- set operations

Column-family stores

Can store tabular data (especially sparse tables), e.g. Google Bigtable, Apache HBase, Cassandra.

A Bigtable cluster serves a set of tables, which are sparse, distributed persistent data with dimensions:

- Rows (unit of “transaction” consistency)
- Columns (organised in column families, usually of same type, also unit of access control)
- Time (each cell can have multiple versions, index by timestamp)

Use Cases:

- event logging
- content management systems, blogging platforms
- webcounters

When not to use:

- ACID transactions
- aggregate functions

- Support data represented as graph (e.g. social networking)
- SPARQL
- e.g. IBM RDF GraphStore

Assignment Project Exam Help

See Section 3.2

Use Cases:

<https://powcoder.com>

- connected data
- routing, dispatch, location-based services
- recommender systems

Add WeChat powcoder

When not to use:

- when need to update many items
- large scale

Assignment Project Exam Help

NoSQL databases are schemaless

- database changes are more flexible
- easier to deal with non-uniform data
- but still need some form of implicit schema to write programs to access the data

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Compare terminology of MongoDB (with Oracle)

- MongoDB instance (= database instance)
- database (= schema)
- collection (= table)
- document (=row)
- _id (=rowid)
- DBPerf (= ,dir)

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

- Pramod J. Sadalage and Martin Fowler, *NoSQL: A Brief Guide to the Emerging World of Polygot Persistence*, Addison-Wesley, 2013.
- NoSQL - Wikipedia
- Chang et al. "Bigtable: A distributed storage system for structured data", *ACM Transactions on Computer Systems*, Vol 26, No 2, June 2008.

Add WeChat powcoder