
Introduction to Databases

Assignment Project Exam Help

CSC 343 <https://powcoder.com>

Winter 2021 Add WeChat powcoder

MICHAEL LIUT (MICHAEL.LIUT@UTORONTO.CA)

ILIR DEMA (ILIR.DEMA@UTORONTO.CA)

DEPARTMENT OF MATHEMATICAL AND COMPUTATIONAL SCIENCES
UNIVERSITY OF TORONTO MISSISSAUGA



UNIVERSITY OF
TORONTO
MISSISSAUGA

Let's look at the course syllabus!

What does this course look like?!





Topics

- Relational Model
- ER Model
- SQL
- Aggregation and Joins
- Constraints and Triggers
- Relational Algebra

Assignment Project Exam Help

<https://powcoder.com>

- Database Design
- Transactions
- Concurrency

Add WeChat powcoder

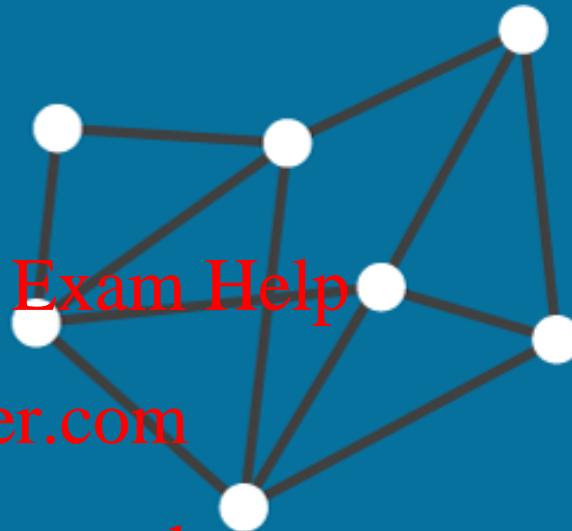
- Intro to NoSQL and MongoDB (time permitting)
- Hadoop vs. GFS vs. Cassandra (time permitting)

EVERYDAY **BIG DATA**

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Big data describes the collection of complex and large data sets such that it's difficult to capture, process, store, search and analyze using conventional data base systems. Its uses are shaping the world around us, offering more qualitative insights into our everyday lives.



GLOBAL INTERNET TRAFFIC IN 2013 WAS APPROXIMATELY

5,000,000,000,000,000,000 BYTES
(51 BILLION GIGABYTES).

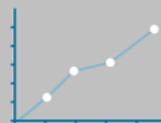
CHARACTERISTICS (V'S) OF BIG DATA

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Global internet population
GREW 14.3% BETWEEN
2011 & 2013



3 BILLION

The number of people who have
access to the internet today equals
that of the world's population in 1960



=



VELOCITY:

Refers to the increasing speed at which data is created and the speed at which it can be processed, stored and analysed.

EVERY MINUTE THERE ARE:

Assignment Project Exam Help

VARIETY:

Big data includes both structured and unstructured data, including text, version data, audio, video, click streams, log files. Technology allows the data types to be analysed together.



90%

OF DATA GENERATED IS "UNSTRUCTURED"

This includes tweets, photos, customer purchase history and even customer service call logs.



Big Data AND Trends

During a hurricane warning/prior to a hurricane occurring, Walmart found an increase in sales in:

Assignment Project Exam Help

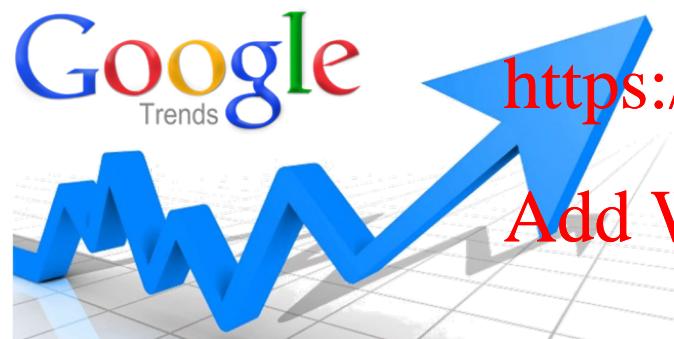
<https://powcoder.com>

Strawberry Pop-Tarts at powcoder

7 times the “norm”



Big Data AND Trends



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



UNIVERSITY OF
TORONTO
MISSISSAUGA

Google Trends

TOP QUESTIONS ON THE EUROPEAN UNION

in the UK since Brexit result officially announced

- 1 What does it mean to leave the EU?
- 2 What is the EU?
- 3 Which countries are in the EU?
- 4 What will happen now we've left the EU?
- 5 How many countries are in the EU?

google.com/trends



GoogleTrends 
@GoogleTrends

 Follow

"What is the EU?" is the second top UK question on the EU since the **#EURefResults** were officially announced

7:25 AM - 24 Jun 2016

  27,479  16,995

Data Science

Empirical Science: collect and systemize facts

Assignment Project Exam Help

Theoretical Science: formulate theories and empirically test them.

<https://powcoder.com>

Computational Science: run automatic proofs, run simulations.

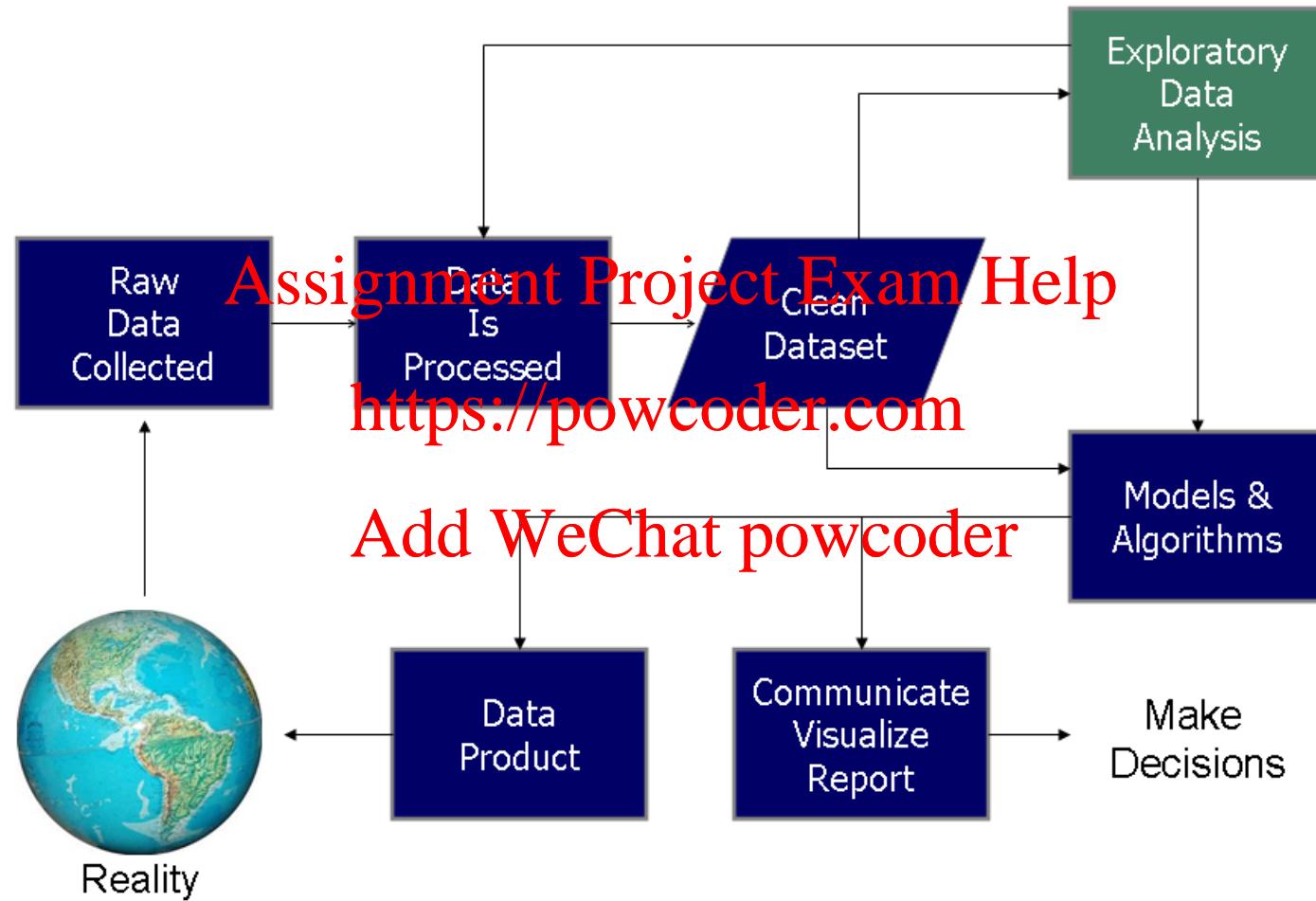
Data Science: collect data and find patterns within the data.

Add WeChat powcoder

Think statistics meets mathematics meets
computer science and programming.



Data Science Process



Question?



Assignment Project Exam Help

How do you think databases run your life?
<https://powcoder.com>

Add WeChat powcoder

How do databases run your life?

- Cloud Storage (e.g. Dropbox, Google Drive, iCloud, etc...)
 - Where is the data? How is it categorized and quickly accessible?

<https://powcoder.com>

- Online Streaming Applications (e.g. Netflix, YouTube, HBO Now, etc...)
 - Generating lists of videos based on searches and tracking users preferences. “Recommended” videos.
- Finances (e.g. Chequing/Savings Account, Stock Market, Credit Cards, etc...)
 - VISA processes an average of 150 million transactions per day.

Add WeChat powcoder

How do databases run your life?

- Social Media (e.g. Instagram, Facebook, Twitter, etc...)
 - Storing personal information and multimedia content.
 - “Suggestions For You” or “People You May Know”.

Assignment Project Exam Help
<https://powcoder.com> Huge area of Data Analytics!

- E-commerce (e.g. Amazon, eBay, Alibaba, etc...)
 - Online business that store and catalogue items.
 - Organize their product's details, pricing information, and sellers.
 - Store users' purchase history, payment information/preferences, and search history.

Add WeChat powcoder

Data Analytics

- The science of examining and interpreting raw data to find patterns and deduce conclusions.

Assignment Project Exam Help

<https://powcoder.com>

- Applying algorithms, mathematical techniques, and mechanical processes to form a conclusion about the information being analyzed.

Add WeChat powcoder

- By 2020, there will be over \$200 Billion spent annually in the 'Big Data and Business Analytics' market.



UNIVERSITY OF
TORONTO
MISSISSAUGA

What is a Database?

Assignment Project Exam Help

Can you define this?

<https://powcoder.com>

Add WeChat powcoder

What is a Database?

Assignment Project Exam Help

Naively defined as...

- a collection of information that exists over a long period of time.

<https://powcoder.com>

Add WeChat powcoder

What is a Database?

DATABASE

Assignment Project Exam Help

A very large, integrated collection of data (i.e. records or files).

Models a real-world *enterprise*.

<https://powcoder.com>

- Entities (e.g. teams, games)
- Relationships (e.g. Barack Obama received The Nobel Peace Prize)
- Constraints (e.g. at least one doctor on duty during off hours)

Add WeChat powcoder

DATABASE MANAGEMENT SYSTEM (DBMS)

A software system designed to **store**, **manage**, and **facilitate access** to data.



UNIVERSITY OF
TORONTO
MISSISSAUGA

Is the WWW a DBMS?

WWW = World Wide Web

Assignment Project Exam Help

Fairly sophisticated searches available

- Web Crawlers *index* pages
- Keyword-based search for pages

Add WeChat powcoder

Currently data is unstructured and untyped

Search ONLY

- Can't modify the data
- Can't get summaries or complex combinations of data



UNIVERSITY OF
TORONTO
MISSISSAUGA

Is the WWW a DBMS?

Few (zero) guarantees provide for:

- Freshness of data
- Consistency across data items
- Fault Tolerance

Assignment Project Exam Help

<https://powcoder.com>

Web sites (e.g. e-commerce sites such as Amazon or E-Bay, typically have a DBMS in the background to provide these functions.

Add WeChat powcoder



UNIVERSITY OF
TORONTO
MISSISSAUGA

“Search” vs. “Query”

What if you wanted to lookup all of the countries apart of the European Union (EU)?

<https://powcoder.com>

Try “countries in the eu” in a search engine
(e.g. Google)

Add WeChat powcoder

Google

countries in the eu

All Images Maps News Videos More ▾ Search tools

The EU countries are: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the UK.

Countries in the EU and EEA - GOV.UK
<https://www.gov.uk/eu-eea>

EUROPA: EU member countries
europa.eu/countries/member-countries/ ▾
When European countries started to cooperate economically in 1951, only Belgium, Germany, France, Italy, Luxembourg and the Netherlands participated. Over time, more and more countries decided to join. The union reached its current size of 28 member countries with the accession of Croatia on 1 July 2013.
Schengen Area · From 6 to 28 members · United Kingdom · Greece

Countries - Europa.eu
[europa.eu > EUROPA > About the EU](http://europa.eu/EUROPA>About the EU) ▾
Jan 25, 2016 - EU Member States, candidate countries and other European countries.

Member state of the European Union - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Member_state_of_the_European_Union ▾
The European Union (EU) comprises 28 member states. Each member state is party to the The EU's

Search

Based on keyword matching

Assignment Project Exam Help

- Our search matches countries that belong to the European Union (EU)

<https://powcoder.com>

- Results are ranked based on:
 - Popularity
 - Reputation
 - Paid Advertisements

Add WeChat powcoder

- Web documents
 - Limited structure

“Search” vs. “Query”

“Search” returns a document as is.

Assignment Project Exam Help

Countries in the EU and EEA

<https://powcoder.com>

Add WeChat powcoder



[Home](#) > [Passports, travel and living abroad](#) > [Living abroad](#)

The European Union (EU) is an economic and political union of 28 countries. It operates an internal (or single) market which allows free movement of goods, capital, services and people between member states.

EU countries

The EU countries are:

Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the UK.

Query

A request of information from a Database.

- In a DBMS, a specialized language (*Query Language*) is used.

Assignment Project Exam Help

The ease of which this information can be obtained from a database often determines its value to a user.

Add WeChat powcoder

The questions posed in a Query are generally designed for a more specific result than those in a search.



Query

Think of a University Database, some questions asked may be:

Assignment Project Exam Help

1. What is the name of the student with student ID #123456?
<https://powcoder.com>

2. How many students are enrolled in CSC 343?

Add WeChat powcoder

3. What fraction of students in CSC 343 received a grade better than B?

Is a File a DBMS?

Thought Experiment 1

Assignment Project Exam Help

You and a friend are both editing a file at the same time.

You and your friend both save the file at the exact same time.

<https://powcoder.com>

Whose change survived?

A) Yours B) Your Friends C) Both D) Neither E) Not A, B, C, or D

Add WeChat powcoder



Is a File a DBMS?

Thought Experiment 2

Assignment Project Exam Help

You and a friend are updating a file

The power goes out

<https://powcoder.com>

Whose change survived?

Add WeChat powcoder

- A) All
- B) None
- C) All Since Last Save
- D) Not A, B, or C

Q: How do you write programs over a subsystem when it promises you No Options?

A: VERY, VERY CAREFULLY!!





UNIVERSITY OF
TORONTO
MISSISSAUGA

Why Use a DBMS?

- Data independence and efficient access.

Assignment Project Exam Help

- Reduce application development time.

<https://powcoder.com>

- Data integrity and security.

Add WeChat powcoder

- Concurrent access, recovery from crashes.



Why Study Databases?

Shift from **computation** to **information**

- Always true for corporate computing
- Web made this point for personal computing
- More and more true for scientific computing

Assignment Project Exam Help

<https://powcoder.com>

Need for DMBS has exploded!

- **Corporate:** retail swipe/clickstreams, “customer relationship management”, “supply chain management”, “data warehouses”, Big Data, etc..
- **Scientific:** digital libraries, Human Genome Project, Sloan Digital Sky Survey, physical sensors, etc...

Add WeChat powcoder

DMBS encompasses much of CS is a practical discipline

- OS, languages, theory, machine learning, logic
- Yet traditional focus on real-world apps



What is Intellectual Content?

Representing Information

- data modelling

Assignment Project Exam Help

Languages and Systems for Querying Data

- complex queries with realistic semantics*
- over massive data sets

Add WeChat powcoder

Concurrency Control for Data Manipulation

- controlling concurrent access
- Ensuring transactional semantics*

Reliable Data Storage

- maintain data semantics* even if you pull the plug

*semantics: the meaning or relationship of meaning
of a sign or set of signs.





UNIVERSITY OF
TORONTO
MISSISSAUGA

Describing Data: Data Models

A data model is a collection of concepts for describing data.
Assignment Project Exam Help

A scheme is a description of a particular collection of data, using a given data model.
https://powcoder.com

The relational data model is the most widely used model today.
Add WeChat powcoder

- Main Concepts: relation, basically a table with rows and columns.
- Every relation has a schema, which describes the columns, or fields.



Data Independence

Applications insulated from how data is structures and stored.

Logical data independence Assignment Project Exam Help

- Protection from changes in logical structure of data.
 - i.e. the ability to change the conceptual (logical) schema without changing the external schema (user view).
 - e.g. addition/removal of an entity or relationship.

Add WeChat powcoder

Physical data independence

- Protection from changes in physical structure of data.
 - e.g. hardware-level considerations, system designs, etc...

Q: Why is this particularly important for DBMS?

A: Rate of change of DB applications are slow!

More Generally: $d\text{app}/dt \ll d\text{platform}/dt$

Concurrency Control

Concurrent execution of user programs key to good DBMS performance.

- Frequent disk accesses.
- Keep the CPU working on several programs concurrently.

Assignment Project Exam Help

<https://powcoder.com>

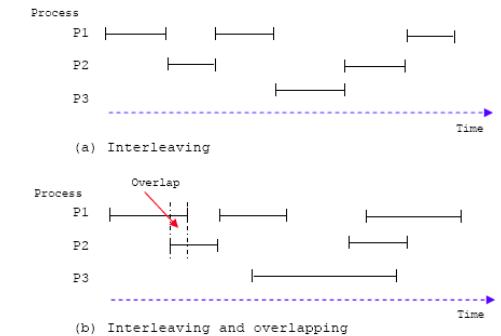
Interleaving actions of different programs: trouble!

- e.g., account-transfer and print statement at the same time.

Add WeChat powcoder

DBMS ensures such problems don't arise.

- Users/programmers can pretend they are using a single-user system (“Isolation”).
- Thank goodness! You don’t have to program “very, very carefully”.



Database Structure

Typically has a layered architecture:

Assignment Project Exam Help

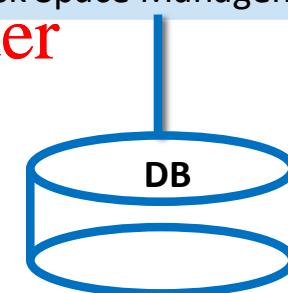
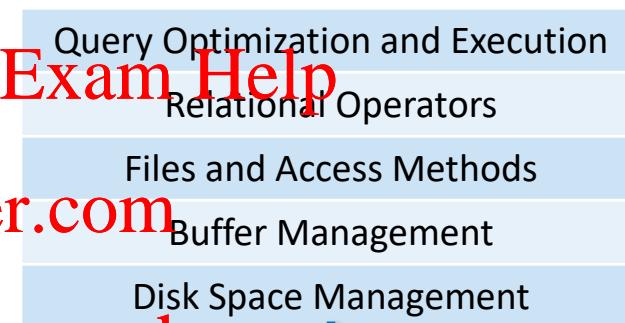
The figure doesn't show:

- Concurrency Control
- Recovery Components

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

Each system has its own variations.

These layers must consider concurrency control and recovery!



Why Don't We Always Use a DBMS?

- Assignment Project Exam Help**
1. Expensive/complicated to setup and maintain
 2. Cost and complexity must be offset by need
 3. General-purpose, not suited for special-purpose tasks (e.g. text search!)
- Add WeChat powcoder**

The ACID Approach

1. Atomicity: all changes take effect, or none do.
Assignment Project Exam Help
2. Consistency: the database is transferred from one valid state to another valid state. **<https://powcoder.com>**
3. Isolation: the results of a transaction are invisible to other transactions until the transaction is complete.
Add WeChat powcoder
4. Durability: once committed, the results of a transaction are permanent and survive future system and media failures.



UNIVERSITY OF
TORONTO
MISSISSAUGA

Databases Make These Folks Happy...

DBMS Vendors and Programmers

Assignment Project Exam Help

- Oracle, IBM, Microsoft, ...

<https://powcoder.com>

End-Users in many fields

- Business, Education, Science, ...

Add WeChat powcoder

Database Application Programmers

- Build enterprise applications on top of DBMSs
- Build web services that run off DBMSs





UNIVERSITY OF
TORONTO
MISSISSAUGA

Databases Make These Folks Happy...

Database Administrators (DBAs)

Assignment Project Exam Help

- Handle security and authorization
- Data availability and crash recovery
- Database tuning as needs evolve

<https://powcoder.com>

Data Scientists and Analysts

Add WeChat powcoder



Summary

DBMS used to maintain, query large datasets.

- Can manipulate data and exploit *semantics*

Other benefits include:

- Data Independence
- Quick application development
- Data integrity and security
- Recovery from system crashes
- Concurrent access

Add WeChat powcoder

Summary

Levels of abstraction provide data independence

Assignment Project Exam Help

- Key when $dapp/dt \ll dplatform/dt$

<https://powcoder.com>

In this course we will explore:

1. How to be a sophisticated user of DBMS technology.
2. What goes on inside the DBMS.



Citations, Images and Resources

Database Management Systems (3rd Ed.), Ramakrishnan & Gehrke

<http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/>

Assignment Project Exam Help

<https://www-01.ibm.com/software/data/big-data/big-data-dt.html>

<http://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html>

<https://www.kaggle.com/c/walmart-recruiting-sales-in-store-my-weather/forums/t/3299/predicting-strawberry-pop-tarts>

<https://powcoder.com>

http://truthaboutguns-zippykid.netdna-ssl.com/wp-content/uploads/2014/12/Strawberry_Pop_Tarts.jpg

<http://gurupk.com/wp-content/uploads/2016/03/GTSEO.png>

<http://www.npr.org/sections/alltechconsidered/2016/06/24/48074392/britain-local-searches-for-what-is-the-eu-eaa-europe-or-x-co>

<https://www.gov.uk/eu-eea>

<http://www.mjhaccountants.co.uk/wp-content/uploads/cartoon-filing-cabinet-l-e4b53be1891574f1.gif>

<http://www.cs.toronto.edu/~ryanjohn/teaching/cscc43-s12/lectures/c43-intro-v03.pdf>

http://www.tenouk.com/ModuleV_files/image002.png

<http://www.quoteslike.com/images/1480/love-girl-lyrics-and-leave-a-suggestion-at-the-bottom-of-the-page-SiPB6f-quote.jpg>

<https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>

Data science process flowchart from "Doing Data Science", Cathy O'Neil and Rachel Schutt, 2013