But where do I keep this?!

Using SQL & NoSQL Databases to store your data

Rutgers Libraries - NB Data Science Workshop Series

Pratiksha Sharma

Nov 17, 2022

Fall 2022 Hours

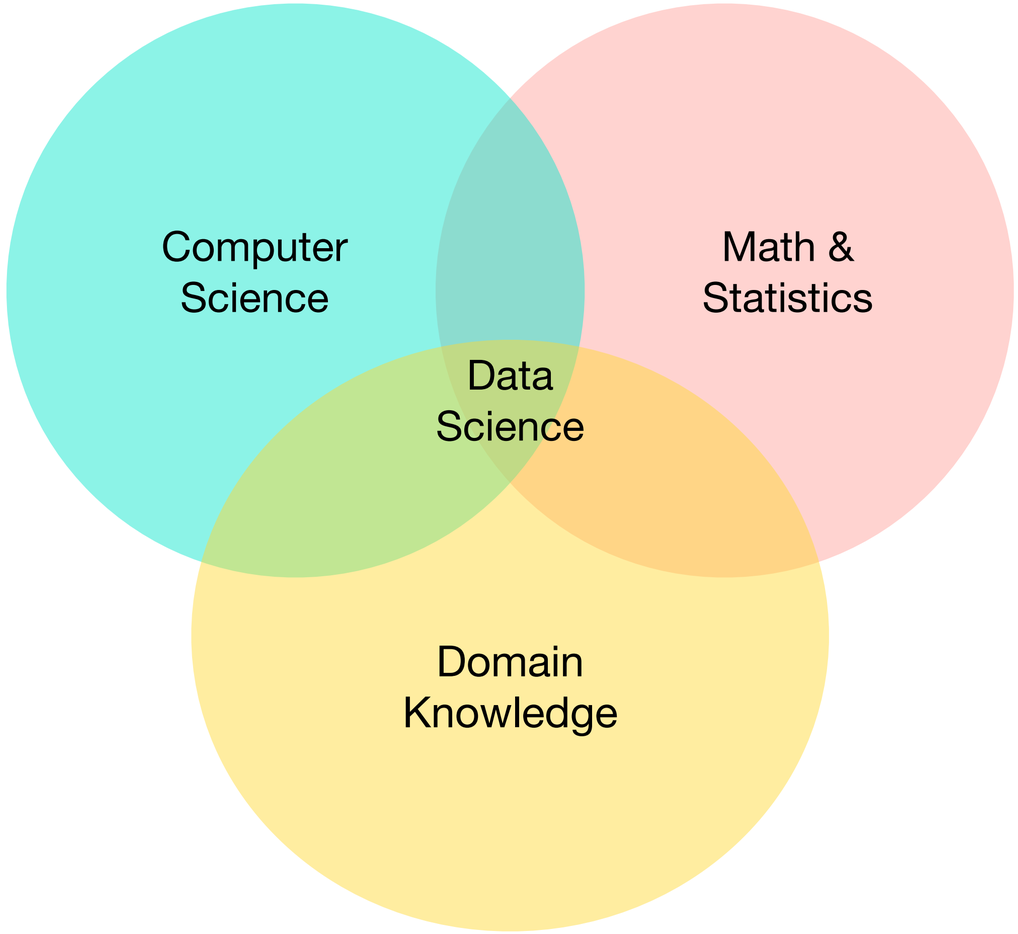
Pratiksha Sharma - Data Science Graduate Specialist

Email: pratiksha.sharma@rutgers.edu
Topics: Data Science, Tableau, Python, SQL & NoSQL Databases
Office Hours (by appointment):
Thursday 12:30 - 01:00 pm (on days when workshop ends at 12:30 pm)
Thursday 01:00 - 01:30 pm (on days when workshop ends at 01:00 pm)
General Consultation: Request an appointment via email
Location: Zoom

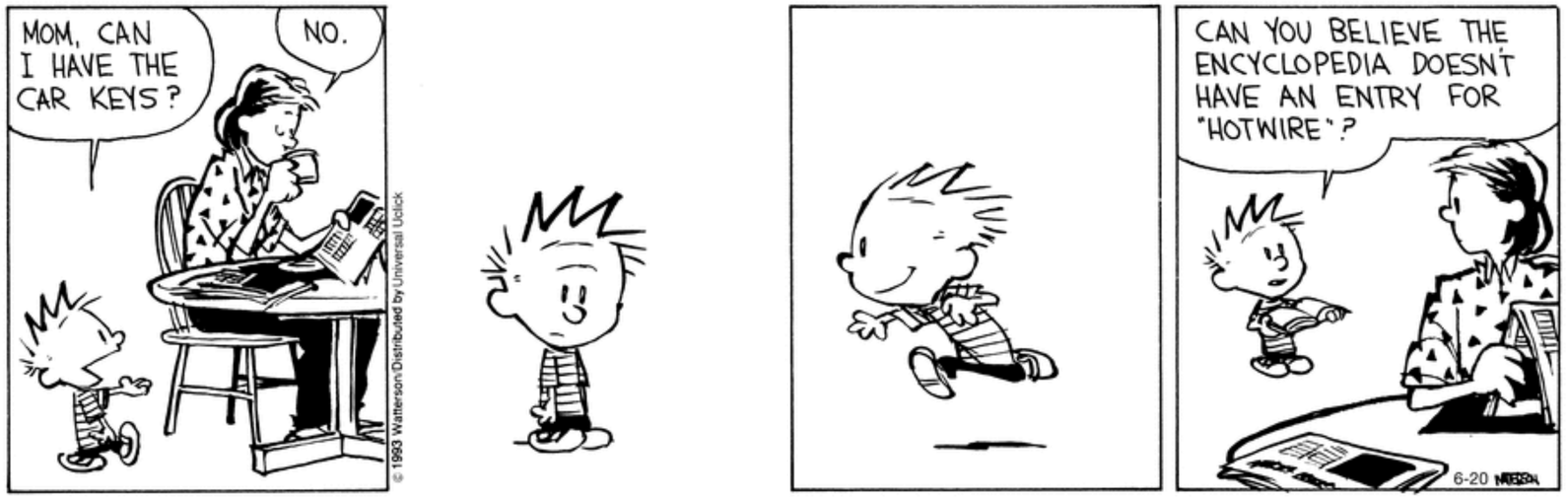
Quick Look: Data Science

What's all the fuss about?

- A combination of Maths & Statistics, Computer Science and Domain Knowledge.
- This workshop is on Programming - concentrates on the **Computer Science!**
- You don't need to be an expert - but data science is a part of everyday life!



Motivation



Source: Bill Waterson | Universal Press Syndicate

Before we begin..

- We will be using MySQL Community Server, MySQL Workbench, MongoDB Community Server & MongoDB Compass.
- The goal for today's workshop is to show you that your different data has different storage needs and how you can use SQL & NoSQL databases to store this kind of data.
- You are encouraged to participate & interact!
- Materials would be provided in the Zoom chat & will be available later on with the workshop materials on: Rutgers Libguides Data Science Workshops

Different Data has Different Storage Needs

- Relational Data:
  - Simply data you can write up in a tabular format: rows x columns
  - For eg: Store transactions, Banking Customer Details, User info etc.
- Non-relational Data:
  - Data may or may not conform to a similar structure or ‘schema’.
  - For eg: Media files, guest transactions (think airline ticket transactions), web-activity data etc.

Different Data has Different Storage Needs

- Relational Data:
  - Requires a storage service that provides organized, customizable, ‘tabular’ storage with an ability to retrieve data easily.
  - Retrieve data via queries.
- Non-relational Data:
  - Requires a storage service that provides an easy, no-structure, no-schema, flexible storage with an ability to retrieve data easily.
  - Retrieve data via queries or GUI.

Different Data has Different Storage Needs

- Relational Data:
  - MySQL, OracleDB, Postgres, AWS Aurora
- Non-relational Data:
  - MongoDB, Cassandra, Google Bigtable

Different Data has Different Storage Needs

- Relational Data:
  - MySQL - Structured Query Language

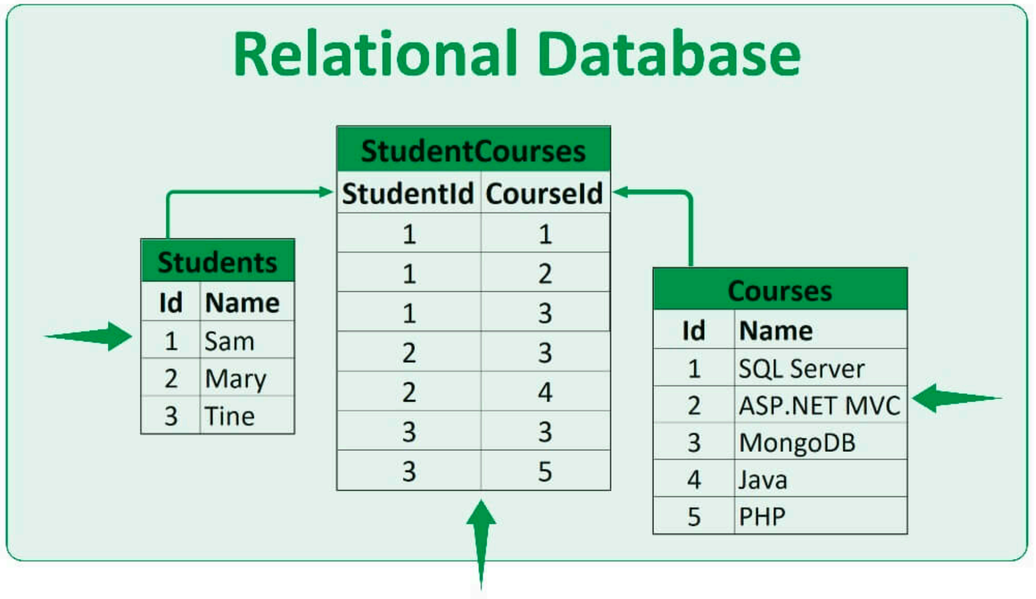


Image Source: https://www.pragimtech.com/blog/mongodb-tutorial/relational-and-non-relational-databases/

Different Data has Different Storage Needs

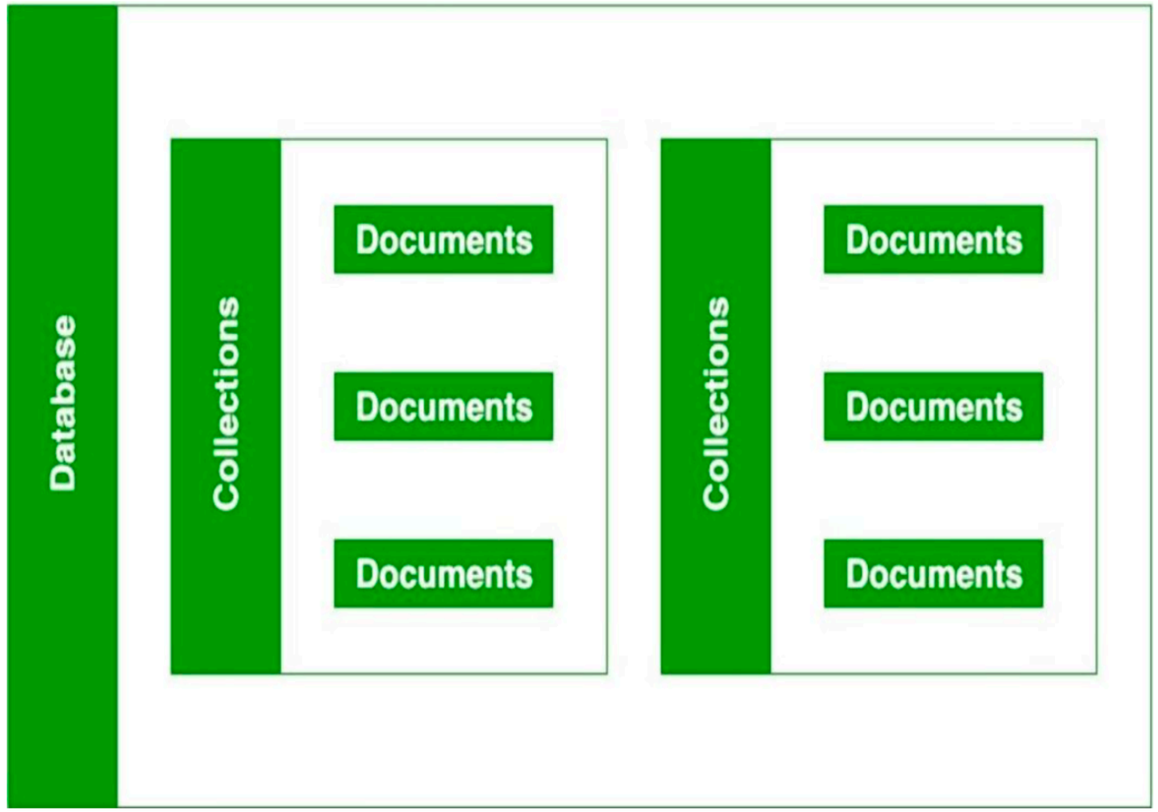
- Non-Relational Data:
  - MongoDB - Mongo's own query language

Key	Document
1001	{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }
1002	{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }

Source: Microsoft Azure

MongoDB Architecture

RDBMS		MongoDB
Database	⟺	Database
Table	⟺	Collection
Row	⟺	Document
Column	⟺	Field
Join	⟺	Embedded Document
Foreign Key	⟺	Reference



Takeaway:

Focus on the Industry Requirements!

Upcoming Workshops

https://libcal.rutgers.edu/nblworkshops

- “Powerful & Simple?!" Perform Analytics using MySQL & MongoDB : Dec 01
- A Peek Behind the Curtain: An end-to-end data science project : Dec 08

Feedback Form

https://rutgers.libwizard.com/f/graduate_specialist_feedback