

# CMPT 732 - Fall 2022

Programming for Big Data 1 - Introduction

WELCOME!



# Teaching Team - G2S3

## Instructors

Greg Baker

George Chow

## TA

Shivek Chhabra

Shrey Grover

Sambhav Rakhe



# George's Background

- ❏ Dev, Management, Leadership
- ❏ Small and large organizations
- ❏ Databases
  - ❏ Data Warehouse
  - ❏ Data Lake/house
- ❏ Closed source products, open source projects



# Greg's Background







Azure Cosmos DB



Couchbase



mongoDB®



presto 

SAP HANA



APACHE HBASE



BigQuery

APACHE   
Spark™



APACHE DRILL



HIVE



Amazon Athena

APACHE  
PHOENIX



# Course Format

Lectures (Mon 8:30PT)

Labs (Wed & Thu; 10:30/12:30)

Quizzes (some Mon)

Term Project





# Resources

## **Lecture Zoom Bridge:**

<https://sfu.zoom.us/j/69807880094?pwd=dnhhSFFSbktXODNBb2ZlbnhucjFIUT09>

## **Lab Zoom bridge:**

<https://sfu.zoom.us/j/65195601693?pwd=R2pKUnBmRVhLTGRTUjV3MDE2emRKUT09>

Teaching Team (instructor & TA): [cmpt732g1-help@sfu.ca](mailto:cmpt732g1-help@sfu.ca)

Class Email (students): [cmpt-732-g1@sfu.ca](mailto:cmpt-732-g1@sfu.ca)

Slides, assignments, quizzes, etc: <https://coursys.sfu.ca/2022fa-cmpt-732-g1/pages/>

CourSys Discussion Forum: <https://coursys.sfu.ca/2022fa-cmpt-732-g1/forum/>

## CMPT 732 G1: Discussion Forum

[\[Forum Summary\]](#)

[\[New Thread\]](#)

[#1 Linux Resources](#)

Kai Yee Chow, last activity 1 hour ago



[\[Discussion forum identities\]](#)

[\[Activity digest\]](#)

### Unanswered Questions

None

A question is considered “answered” if (1) an instructor/TA has replied, (2) an instructor/TA has reacted positively (👍, ❤️, 🥰) to a student reply, or (3) the question-asker has marked it answered or reacted positively to a reply.

### Unread Activity

None

### Search Posts

Search posts:

Search

CourSys is hosted by SFU Computing Science.

[Admission](#)  
[Programs](#)  
[Learning](#)  
[Research](#)  
[Community](#)  
[About](#)

[Maps + directions](#)  
[Library](#)  
[Academic Calendar](#)  
[Road Report](#)  
[Give to SFU](#)

[CONNECT WITH US](#)  
[Facebook](#)  
[Twitter](#)  
[YouTube](#)

[CONTACT US](#)  
Simon Fraser University  
8888 University Drive  
Burnaby, B.C.  
Canada V5A 1S6

[Terms and conditions](#)  
© Simon Fraser University

# Grading

Component	Weighing	Grading
Assignments	$10 \times 7\% = 70\%$	Individual
Quizzes	$3 \times 3\% = 9\%$	Individual
Term Project	21%	Small Group (3-4)



# Lectures

In-person at AQ3005

September lectures will be live-streamed, recorded and available

Minority of class outside of country still



# Lab Schedule (SECB 1010)

	<b>Wednesday</b>	<b>Thursday</b>
<b>10:30-12:20 (SECB 1013)</b>	Instructor: George*	Instructor: Greg*
<b>12:30-2:20 (SECB 1010)</b>	Instructor: George*	Instructor: Greg*

\* George on even Wed; Greg on even Thu; reverse for even weeks.



2020 AL271590LJ  
13-11

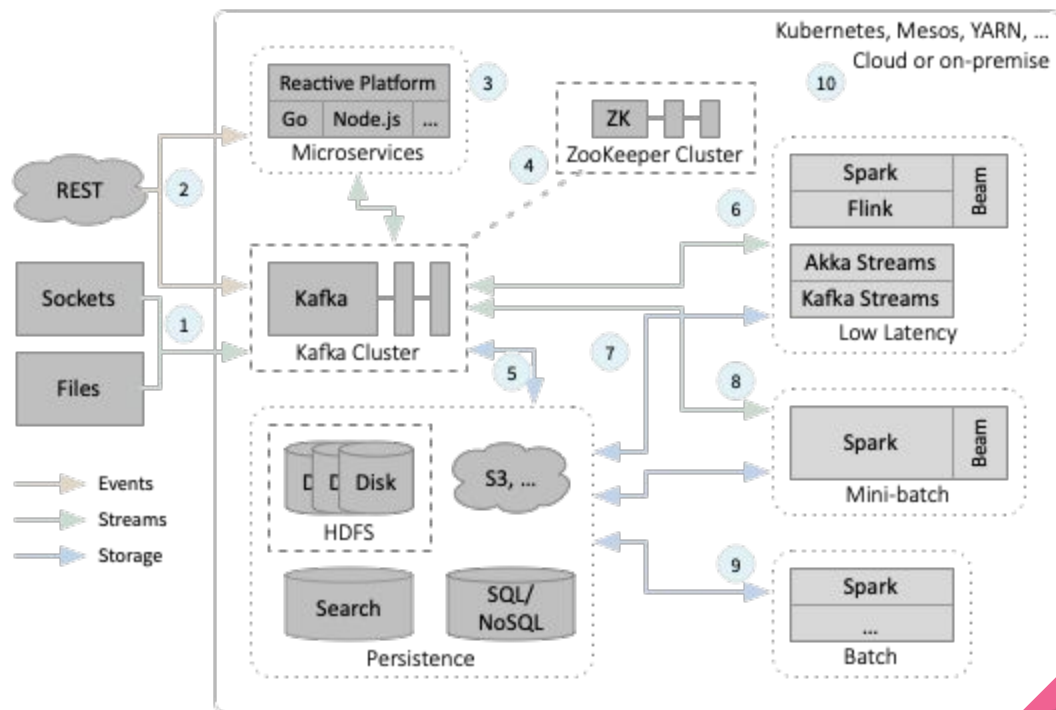
BEST BEFORE ▲ MILLER NAME

Dairyland.

2%

M.F.  
PARTLY  
SKIMMED  
MILK

LACTOSE  
FREE



This Course

# Programming With Data





Big Data

**Volume**

VS

**Variety**

VS

*Velocity*



# Volume

Transactional database, nominal record sizes:

Customer: 4KB

Transaction: 2KB

Source: <https://help.salesforce.com/s/articleView?id=000318951&type=1>



# Digital Breadcrumbs

2-3 actions for search;  
1 action for purchase

Stream of location  
records

- Search for movie schedule & buy tix
- Go to the theatre
- Search for restaurant & reviews
- Pay by credit card
- Write a review of the movie

2-3 actions

1 action for purchase

2-3 action



# Volume because of Variety (2/3)

Digital Breadcrumb tally:

$$3+1 + 10 + 3 + 1 + 3 = 21 \text{ actions}$$

Actions:

“traditional transaction”

Interaction with an app

Create new data





Big Data

**Volume**

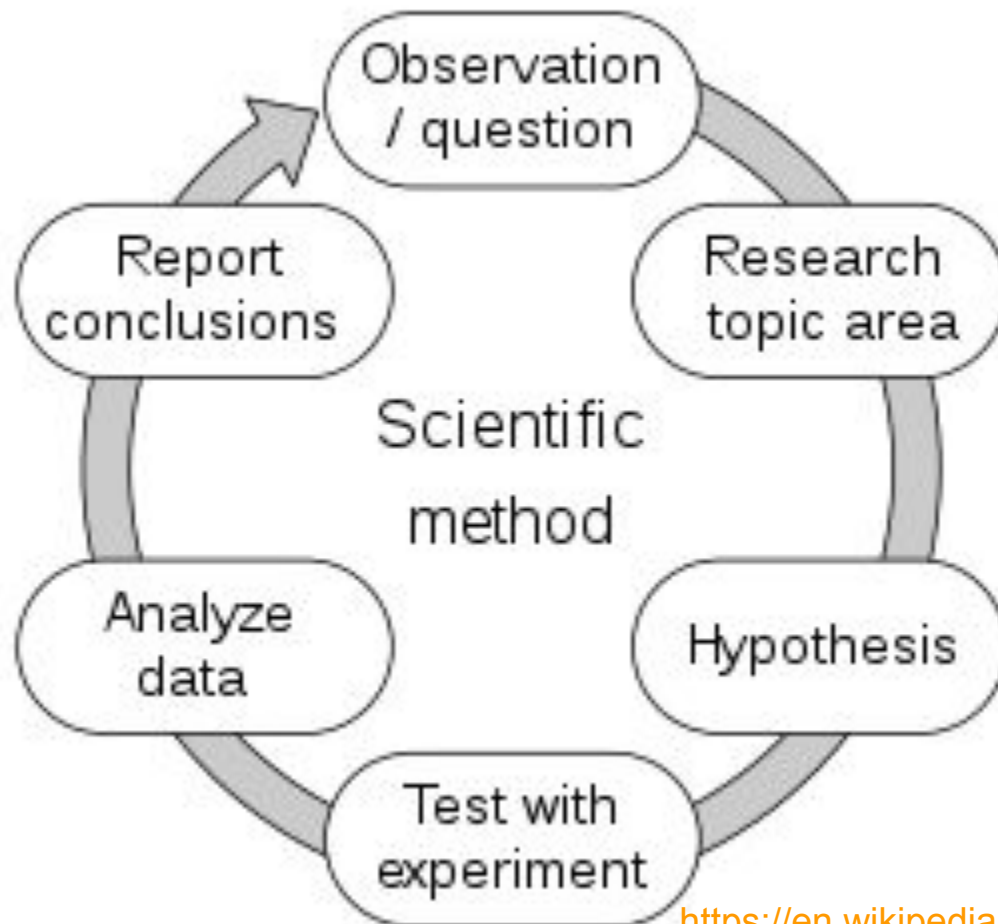
VS

**Variety**

VS

*Velocity*







**SCIENCE**



**ENGINEERING**



# Data Science & Data Engineering

## Science

- Can it be done?
- Is it reproducible?
- What are the limits/constraints?

## Engineering

- How does the system “do it”?
- Can we leverage existing systems?
- How much does it cost?
- How can we save cost/time?

