29.10.20 15:59

CSCI 101
Connecting with Computer Science
Cloud Computing II

Jetic Gū

2020 Fall Semester (S3)

Overview

- Focus: Massive Data Solution
- Readings: R10
- Core Ideas:
  1. There is no such thing as Big Data

P1
"Big" Data

There's No Such Thing as "Big" Data

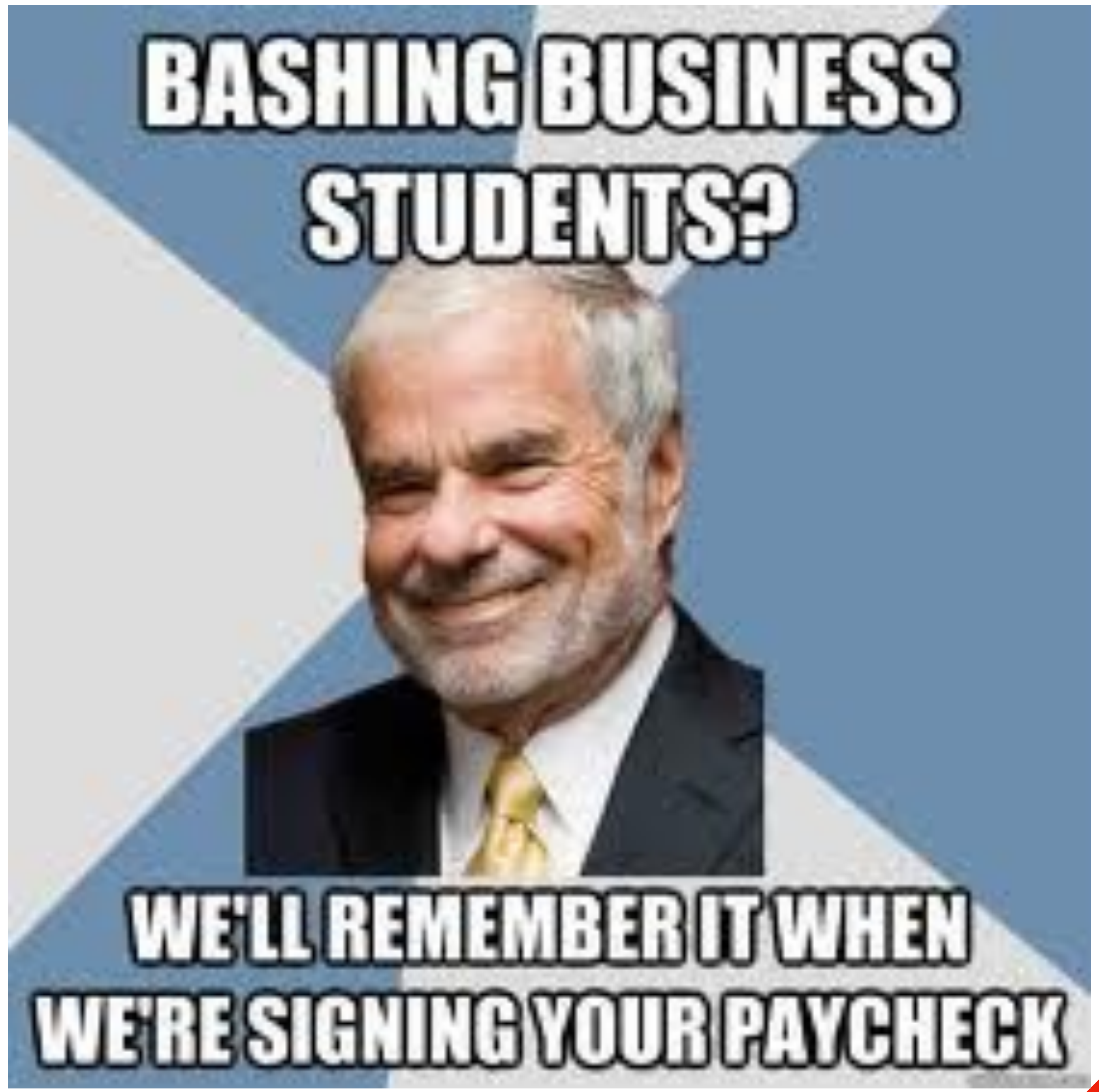
It's just data analysis

Summary

P1
"Big" Data

What is Big Data

- It is an invented concept by Business people
  - Cloud Solutions -> Massive user/market/etc. data
  - Essence of Big data -> Exploit these data


BASHING BUSINESS STUDENTS?
WE'LL REMEMBER IT WHEN WE'RE SIGNING YOUR PAYCHECK

Concept

P1
"Big" Data

What does Cloud Computing Provide?

- Volume: massive amount of information
- Variety: all formats of information
- Velocity: new information gets generated extremely rapidly

Review

The Big Data Challenge

- Volume: massive amount of information
    - You need to be able to process huge amount of data
- Variety: all formats of information
    - You need to be able to process information of various formats
- Velocity: new information gets generated extremely rapidly
    - You need to process data faster than they arrive

P1
"Big" Data

Volume

- Google: 3.5 Billion searches everyday[1]
- Amazon is responsible for 45% of US commerce spending[2]
- 98% of Facebook's revenue comes from Advertising[3]
- More than 1 billion youtube video views each day[4]

[1] https://www.oberlo.ca/blog/google-search-statistics
[2] https://www.repricerexpress.com/amazon-statistics/
[3] https://www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp
[4] https://www.youtube.com/about/press/

Stats

P1
"Big" Data

Volume

- How many ads do you see every day?
  - Estimation from 2017, average American[1]: 4,000 - 10,000
  - We are **trained** and **adapted** to **filter-out** uninteresting **ads**
    - Targeted advertisement is a huge challenge! Because as you get better, the users also get better.

Stats

[1] https://www.forbes.com/sites/forbes-personal-shopper/2020/10/27/best-tvs-for-gaming-2020/#2cec242a1d8c

Variety

- Information comes in different forms
  - Text
  - Images
  - Videos
  - Audio
- Statistics
- Tables
- Databases
- Code
- Environmental data

Example

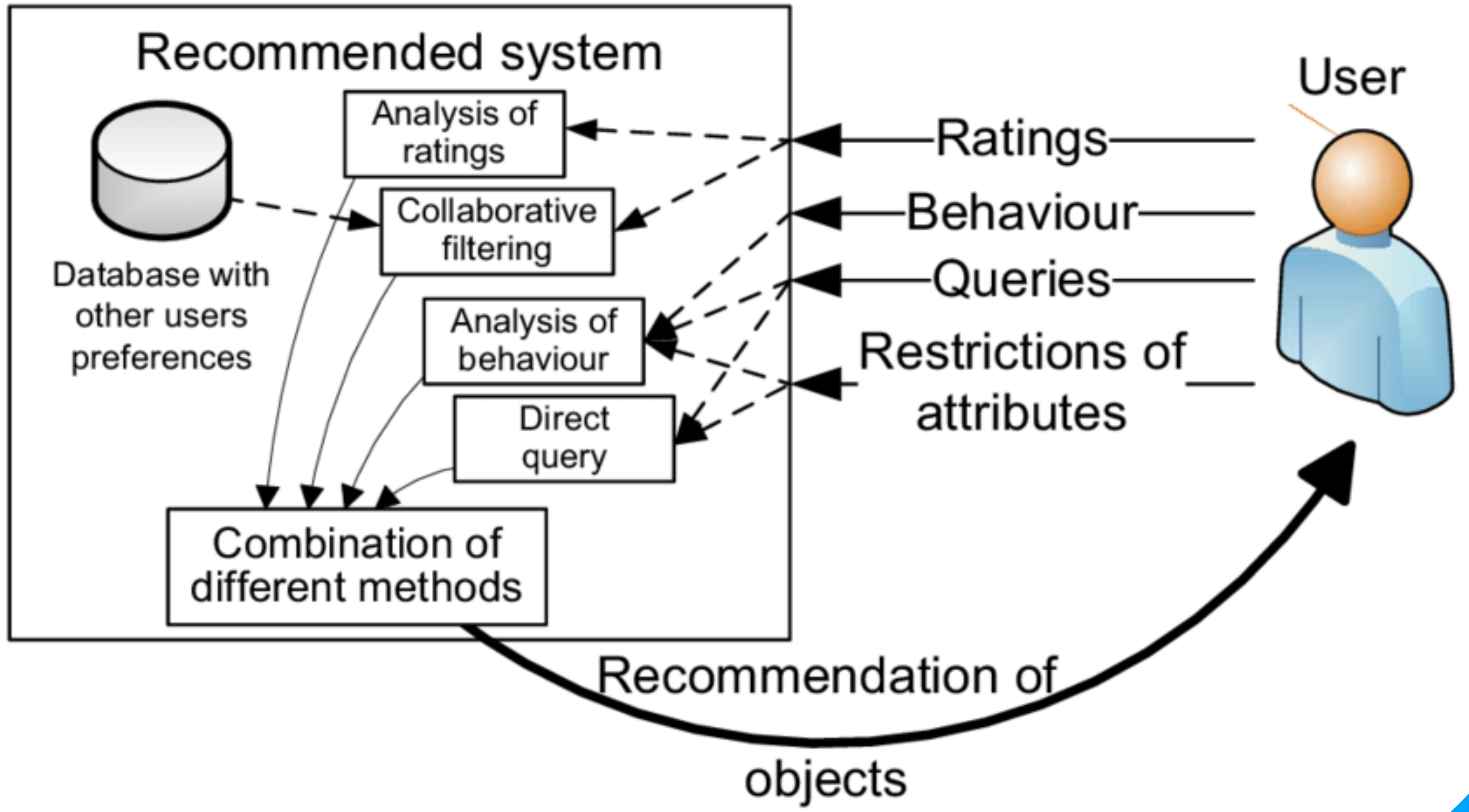
Variety

- What data can be collected from my amazon shopping?
  - Search history
  - Item selection history
  - Purchase history
  - You think that’s it? Too naive

Example

Variety

- What data can be collected from my amazon shopping?
  - Number of seconds you spend on a page
  - Mouse movement
  - Other user behaviours
  - Demographic identification


Recommended system
Analysis of ratings
Collaborative filtering
Database with other users preferences
Analysis of behaviour
Direct query
Combination of different methods
User
Ratings
Behaviour
Queries
Restrictions of attributes
Recommendation of objects

Example

Velocity

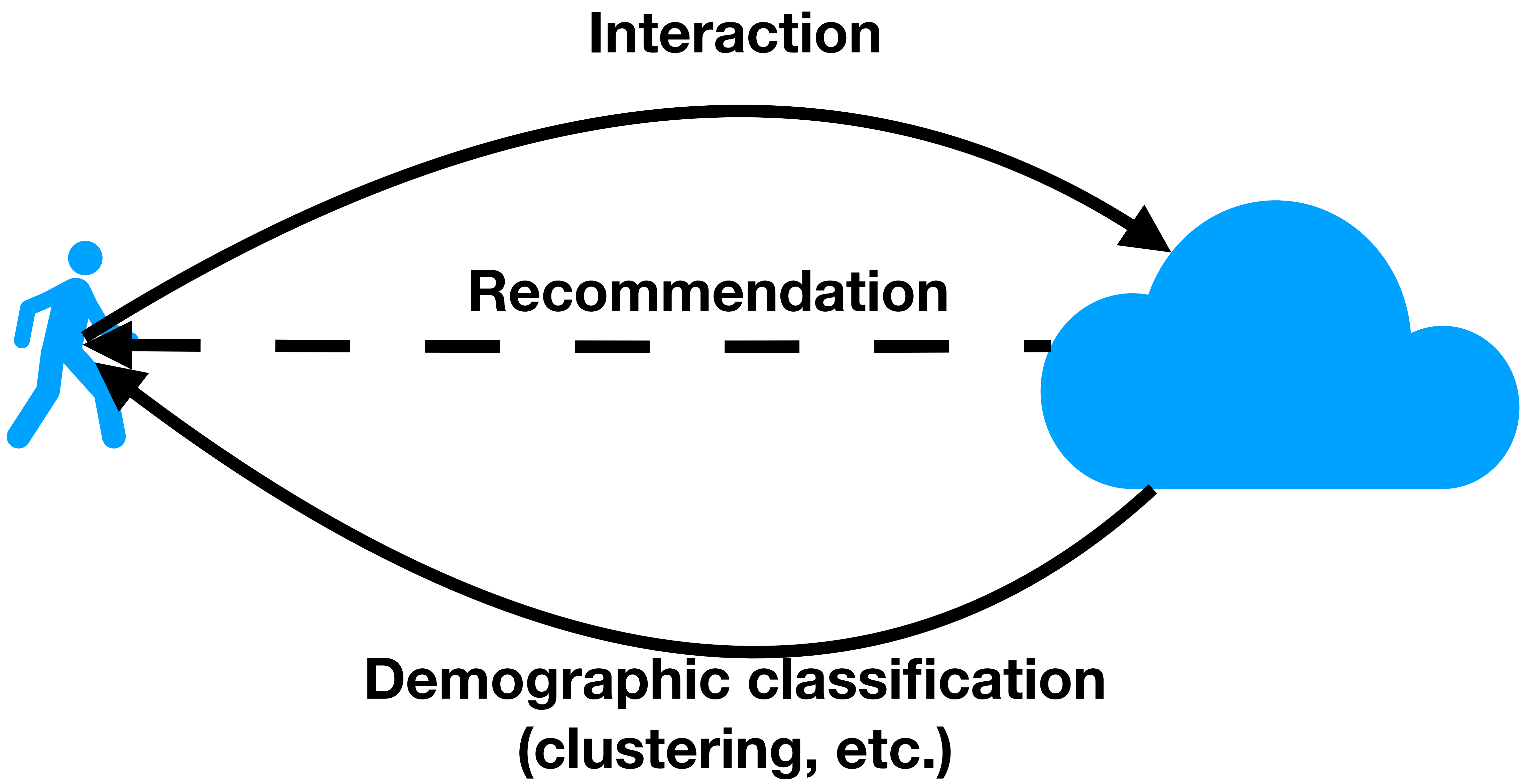
- 300 hours of video are uploaded to YouTube every minute[1]
- Daily instagram "Story" posts: 500 millions[2]
- 140 traffic cameras in BC
  - 24 hours recordings, 140 x 24 x 3600 frames (1 FPS), 12 million images to be recognised and processed everyday
  - Each frame: license plate(s) recognition (AI), speed calculation, etc., needs to be processed within 1/140 sec

1. https://merchdope.com/youtube-stats/
2. https://www.omnicoreagency.com/instagram-statistics/
3. https://www2.gov.bc.ca/gov/content/safety/public-safety/intersection-safety-cameras/where-the-cameras-are

P3
Big Data

Typical Recommendation System Pipeline

- Server collects your data
- Server find other users similar to you
- Server recommends stuff other users like



Example

P1
"Big" Data

Why Big Data?

- **The technology was not there**
    - Modern algorithms for exploiting the massive data are still relatively new
- **The data was not there**
    - Data collection: the **terms of agreement** is actually **Terms of Surrendering Your Data**
    - **Digitisation of all records**: Medical Record, Purchase History (Online Shopping), etc.
    - Social Media: Why not **share all your secrets** to your friends through someone else’s servers? (We promise we WoN’T LoOk!)

Concept

P1
"Big" Data

Business Using Big Data

- Maximise productivity
- Targeted Advertising
- Deliver better products
- Better investments: smart business (Wednesday)

Technical

P1
"Big" Data

Academics Using Big Data

- Quantitative Analysis
- Machine Learning / AI research
- Medical Research
  - Establish vital correlations between medicine and symptoms
  - Public health/rescue: determine/contain virus outbreak through big data

Technical

P1
"Big" Data

5 Stages in "Big" Data Workflow

- Preparing data:
  - Acquisition: Getting the unformatted data
  - Extraction: Extract structured data
  - Integration: Combine data from multiple sources
- Analysis
  - Modelling: Design analysis models to process the data
  - Interpretation: Draw conclusions

Concept

P2
Examples

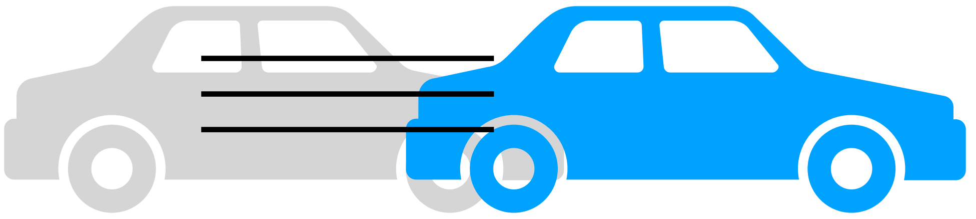
Some Cases of "Big" Data Usage

Summary

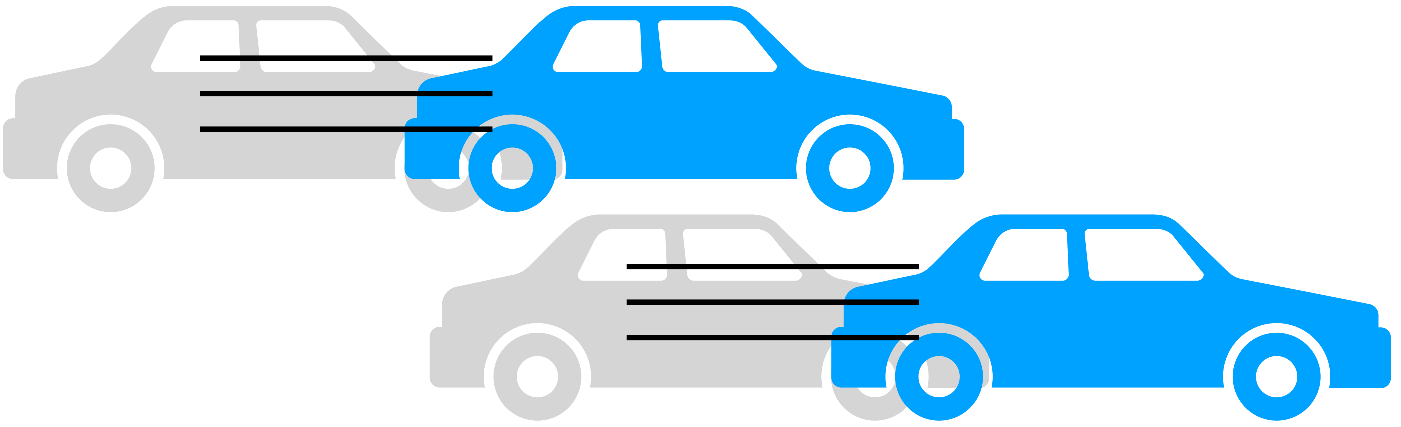
P2
Case 1

Traffic Speed Camera

1. Acquisition
   - Images in sequence



2. Extraction
   - Batch of 3 images in sequence for analysis
     (So that we can determine velocity)



3. Integration
   - Offload images from camera to the database, then send to analysis pipeline

Case Study

P2
Case 1

Traffic Speed Camera

4. Modelling

   - Use Computer Vision technology designed for license plate recognition to ID cars

   - Determine velocity by calculating distance traveled divided by time

5. Interpretation

   - Generate speeding ticket to Jetic, who is totally innocent (not really)

Case Study

P2
Case 2

Recommendation System

1. Acquisition
   - Users and their search histories in Amazon’s Database
2. Extraction
   - Extract searches that leads to order placements
3. Integration
   - Combine results from Amazon US, Amazon CA, and Amazon UK

Case Study

P2
Case 2

Recommendation System

4. Modelling
   - Grouping similar users together
5. Interpretation
   - Recommend users in a group, products that the group's other members bought

Case Study