Computing on Data

Computing on Data

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Kevin C.C. Chang, Professor

Computer Science @ Illinois

Learning Objectives

By the end of this video, you will be able to:

- Explain what computing on data means.

- Explain why computing on data is necessary for a database.

- Give examples of commutation framework for data.

What is "Computation"?
Why Computing on Data?

- Computation--
  - A process that transforms one or more inputs to one or more outputs, following a well-defined "model", or "algorithm."

- Why computation?
  - It's often the very purpose we use a computer.
  - That's the subject of computer science!

- Why computing on data?
  - Because we cannot use data "as is" -- need computation to transform to what we want.
  - Because what we want from data vary in different situations.

Haven't We Learned--
How to Compute on Data?

- We learned how to compute on numbers?
  - Arithmetic algebra: for numbers.
  - Defines: addition, subtraction, multiplication and division.

- But data is not just numbers.

- What is the closest we have?

Closest to "Data": Maybe Matrices?

We can think a table as a matrix, and use linear algebra.

- It assumes "numbers".

Beers

name	brewer	alcohol
Sam Adams	Boston Beer	4.9
Goose IPA	Goose Island	5.9
Summer Ale	Boston Beer	5.3

➡️

name	brewer	alcohol
1	1	4.9
2	2	5.9
3	1	5.3

Sells

bar	beer	price
John Bar	Sam Adams	5
Green Bar	Goose IPA	6
Purple Bar	Sam Adams	4

➡️

bar	beer	price
1	1	5
2	2	6
3	1	4

- Does it help transformation we need?
  - Find beers made by Boston Beer.
  - Which bars sell Boston Beer's beers?

Computing on Data Depends on …

- What data?
  - Relations
  - Key-value pairs
  - Documents
  - Graphs

- What purpose?
  - Asking questions about the world we manage.
    - Which classes are taken by CS students?
    - Which bars sell Boston Beer's beers?

  - Processing data for preparing another datasets
    - To normalize/reorganize/reformat into another datasets.

Examples of Data Computing Framework

- Relational data
  - Relational algebra: Foundation of the SQL query language.
  - Datalog: Popular mechanism for inferencing with relational data.

- Graph data
  - Graph is well-studied data structure, before it became a database model.
  - Finding neighbors, nodes matching paths, shortest paths, etc.

- Key-value data
  - MapReduce