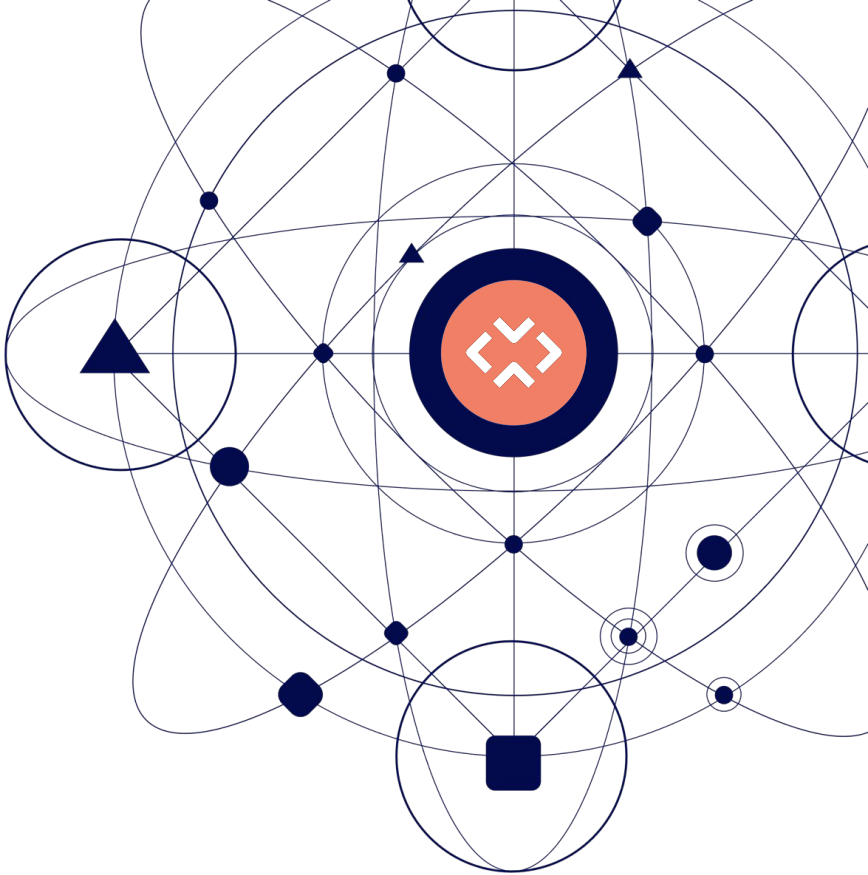
RelationalAI

In Theory in Practice

Hung Q. Ngo

RelationalAI

Outline



Dan and Floris (or ChatGPT) asked

- Difficulties when developing a research idea from theory to practice or vice versa
- The story of a project that transitioned from theory to practice or vice versa
- Their recipe for theory to practice transition or vice versa
- How to identify research problems that (do not) have potential for theory and practice.

In Theory

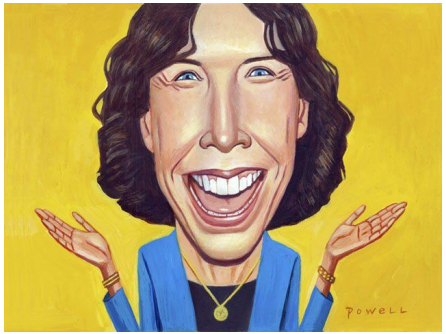
"*Marge, I agree with you – in theory. In theory, communism works.*"

— Homer Simpson

In Practice

“The trouble with the rat race is that even if you win, you’re still a rat.”

— Lily Tomlin



In Theory

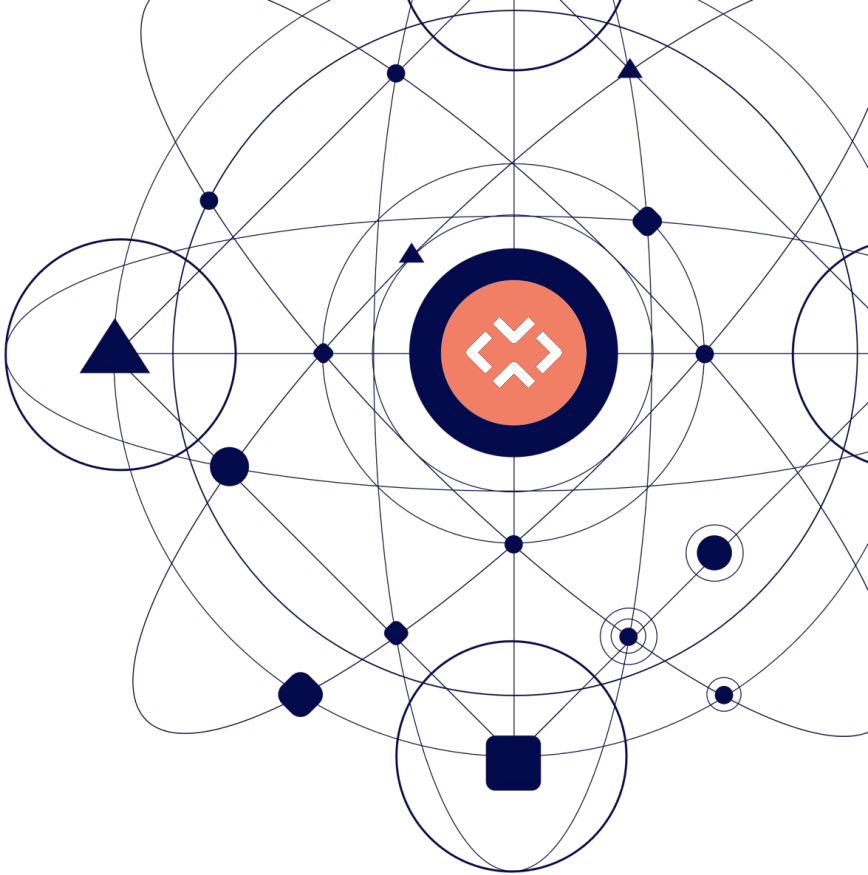
- I have been working on a production-grade query optimizer for years
- I should be able to tell you what works and what doesn’t

In Practice

- Both my English and my Vietnamese have gotten worse

RelationalAI

Theory to Practice



In Theory

Image generated by ChatGPT

In Practice

Image generated by ChatGPT

In Theory

- Engineers working for the same company should be able to communicate effectively to solve problems

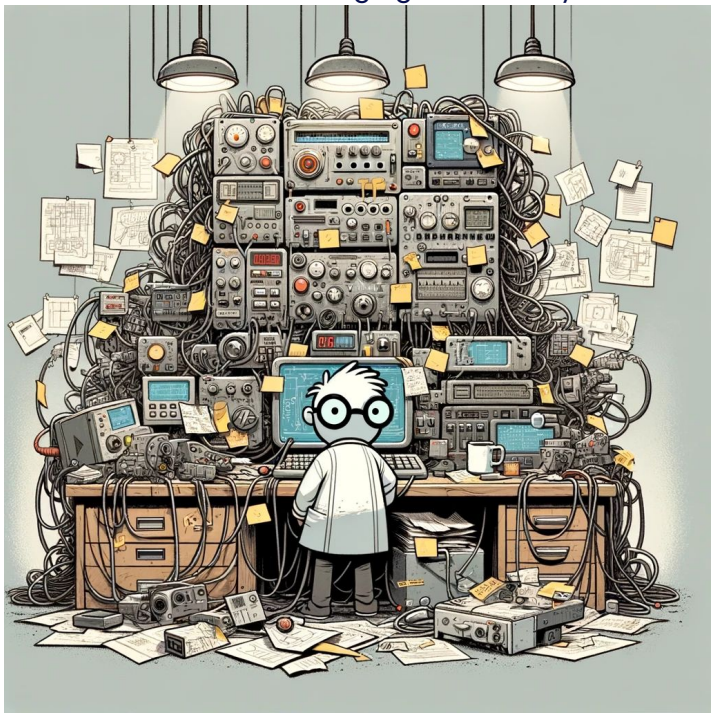
In Practice

- People have wildly different educational backgrounds
- Examples :
  - Theorists to engineers
    - Cardinality estimation based on optimizing a convex objective over polymatroidal functions
    - Query planning based on tree decomposition and submodular and fractional hypertree widths
  - Engineers to theorists
    - Optimize LLVM creation, page sizing
    - Branch prediction, frequent instruction prediction, cache locality
    - Disk seek/read speed, CPU processing cost

In Theory

Emphasis:
Mathematical depth & difficulty

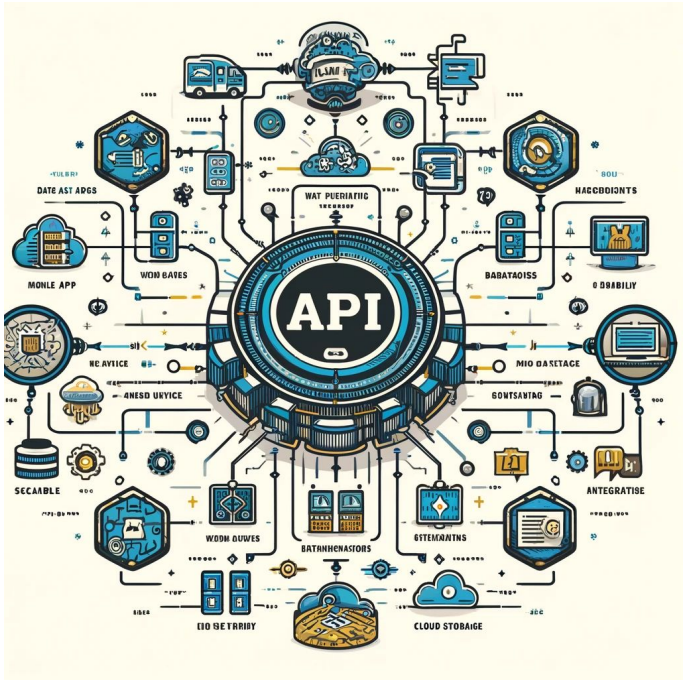
Image generated by ChatGPT



In Practice

Emphasis:
API & Robustness

Image generated by ChatGPT



In Theory

- When we can't solve a problem

  we change … the problem

- Over time, the problem is unrecognizable

In Practice

- When we can't solve a problem

  we change … the solution

- Over time, the solution is unrecognizable

In (Database Management) Theory

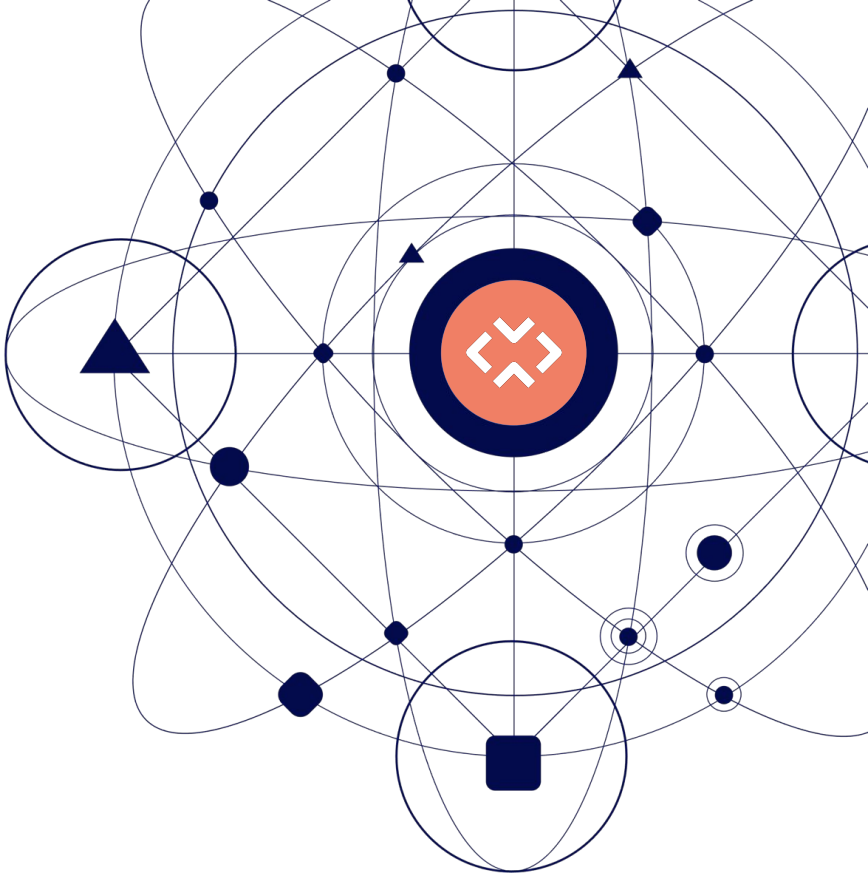
- We tend to like showing what is not possible
  - P vs NP, Complexity Classes
    - Single input parameter N
  - Abstractions are a little too crude to be immediately useful
    - Conjunctive queries
    - Datalog
    - Magic Set Transform

In (Database Management) Practice

- We need to answer all queries, including computationally hard ones
  - Beyond-worst case scenarios
    - Instance optimality
    - Multiple input parameters
  - Abstractions require great care
    - Queries with deep ASTs, with AND, OR, NOT, AGG arbitrarily nested
    - Datalog with NOT, AGG
      - How to handle errors?
    - Magic set over much more complex languages
      - How do you cost a SIPS strategy?

RelationalAI

How to get there?



Things that have been helpful to my theory

- Buy-in from management
- Collaborate with awesome colleagues (theorists and practitioners)
- Get your feet wet
- Iterate with conviction
- Pay attention to the API

Practicality = Simplicity + Robustness

simple != simplistic

A Sample Journey

- NPRR algorithm, worst-case optimal join
  - Cost estimation, new way to think about query opt & eval
- Minesweeper: memoized dynamic programming
  - Functional aggregate queries
- Output size bounds, information theory
  - Degree constraints
  - PANDA algorithm
- Recursion optimization
  - Magic set transform
  - Datalogo

Theory and Practice

- Practice tells us the what

this is wisdom

- Theory tells us the how

this is intelligence

- Good theory tends to produce awesome APIs
- Good practice tends to produce awesome questions

In Theory

"... one of the first things to strike me when I came to Brazil was to see elementary school kids in bookstores, buying physics books. There are so many kids learning physics in Brazil, beginning much earlier than kids do in the United States …"

– "Surely You Must be Joking, Mr. Feynman!" Richard Feynman

In Practice

"There are no experimental results mentioned anywhere in this book, except in one place where there is a ball, rolling down an inclined plane, in which it says how far the ball got after one second, two seconds, three seconds, and so on …

The trouble is, when you calculate the value of the acceleration constant from these values, you get the right answer. But a ball rolling down an inclined plane, if it is actually done, has an inertia to get it to turn, and will, if you do the experiment, produce five-sevenths of the right answer, because of the extra energy needed to go into the rotation of the ball. Therefore this single example of experimental 'results' is obtained from a fake experiment. Nobody had rolled such a ball, or they would never have gotten those results!"

– "Surely You Must be Joking, Mr. Feynman!" Richard Feynman

RelationalAI

Thank You!

For listening (in theory)