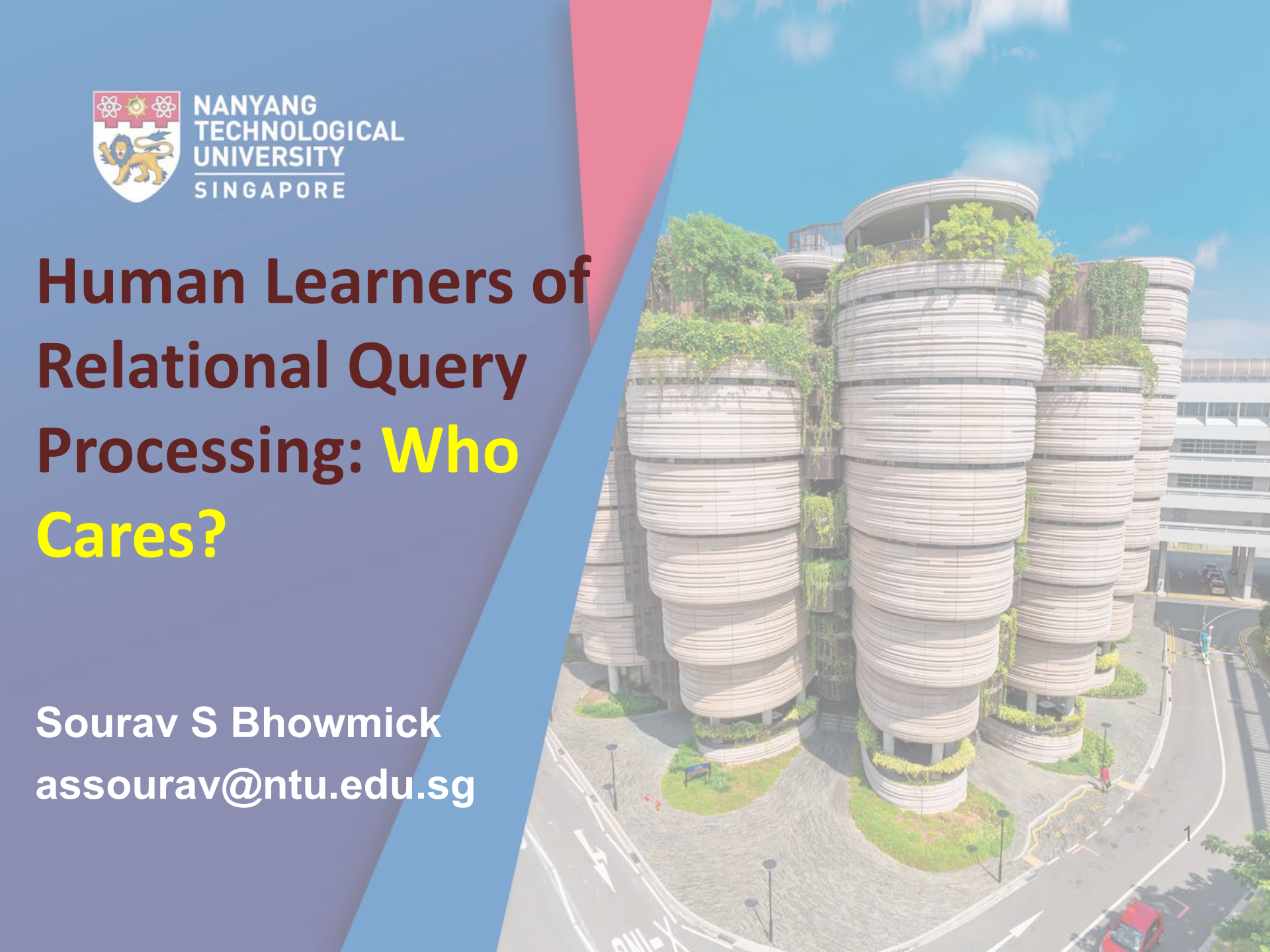
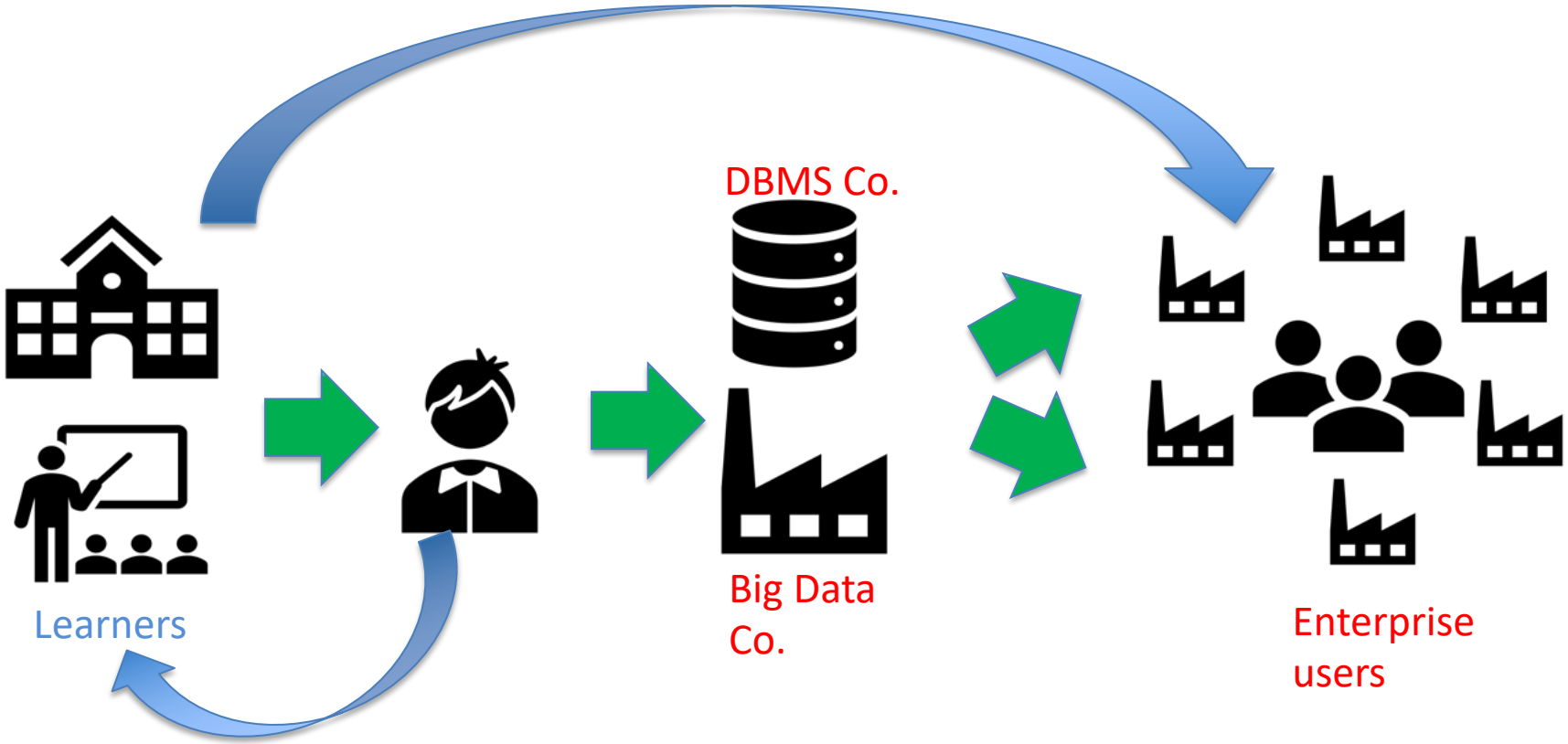


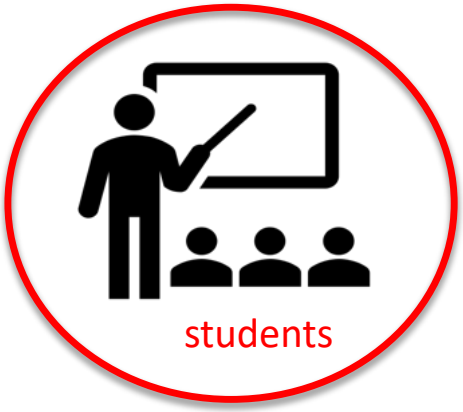
Human Learners of Relational Query Processing: Who Cares?

Sourav S Bhowmick
assourav@ntu.edu.sg

DB for Enterprise Users & Developers



Beyond Enterprise Users?



students

biologist

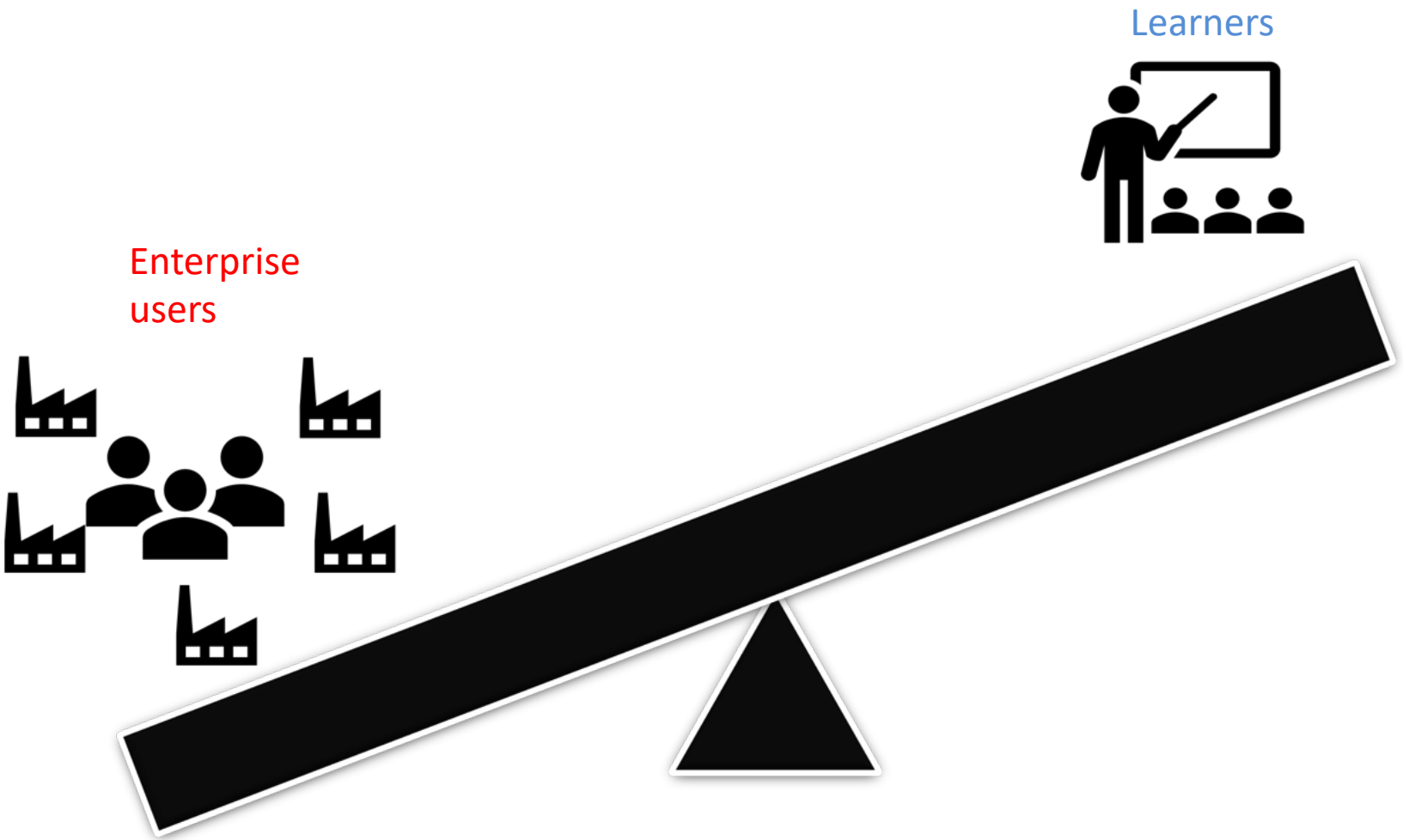
Pharmacist/chemist

ecologist

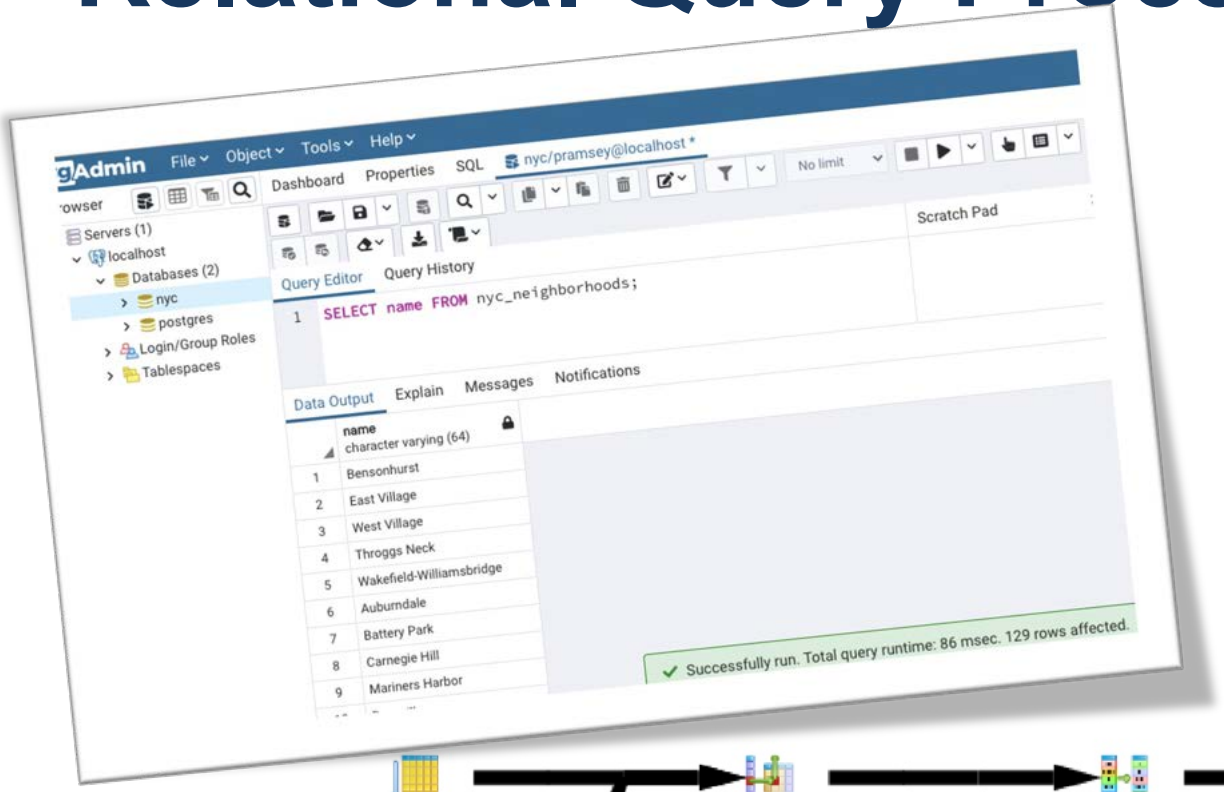
journalist

social scientist

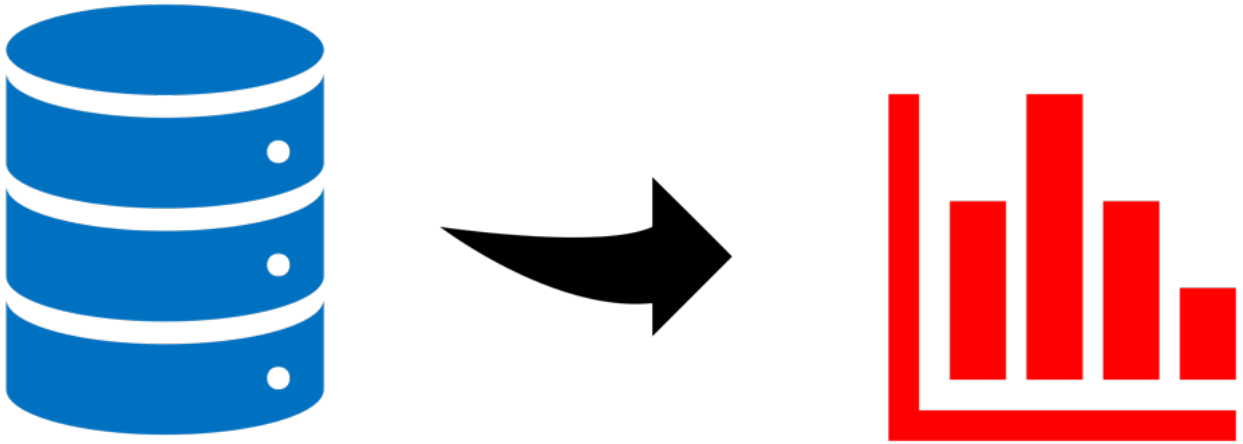
Where Our Attention Lies?



Scant Support For Learning Relational Query Processing



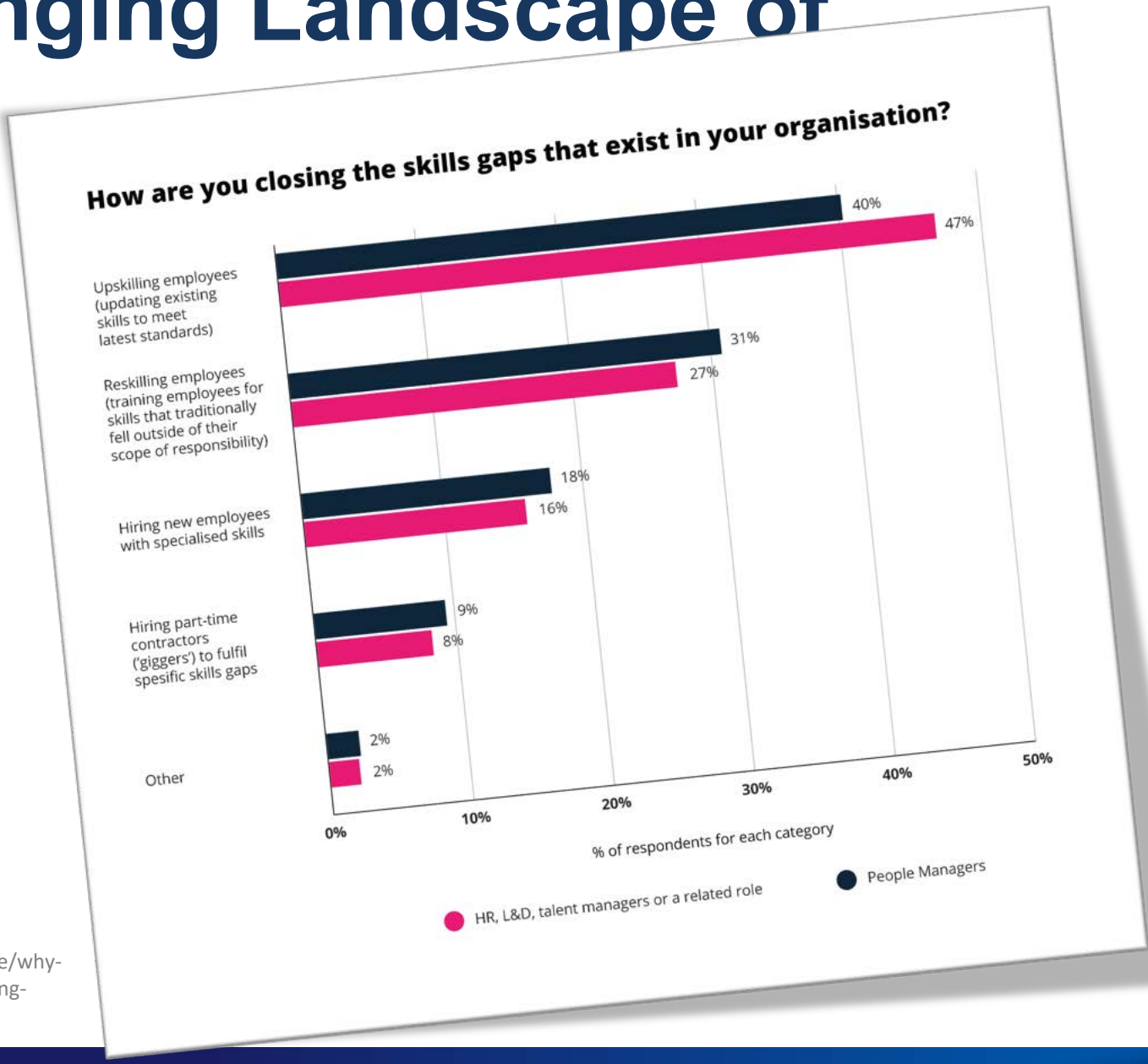
The Popularity of Data Science & AI



Data

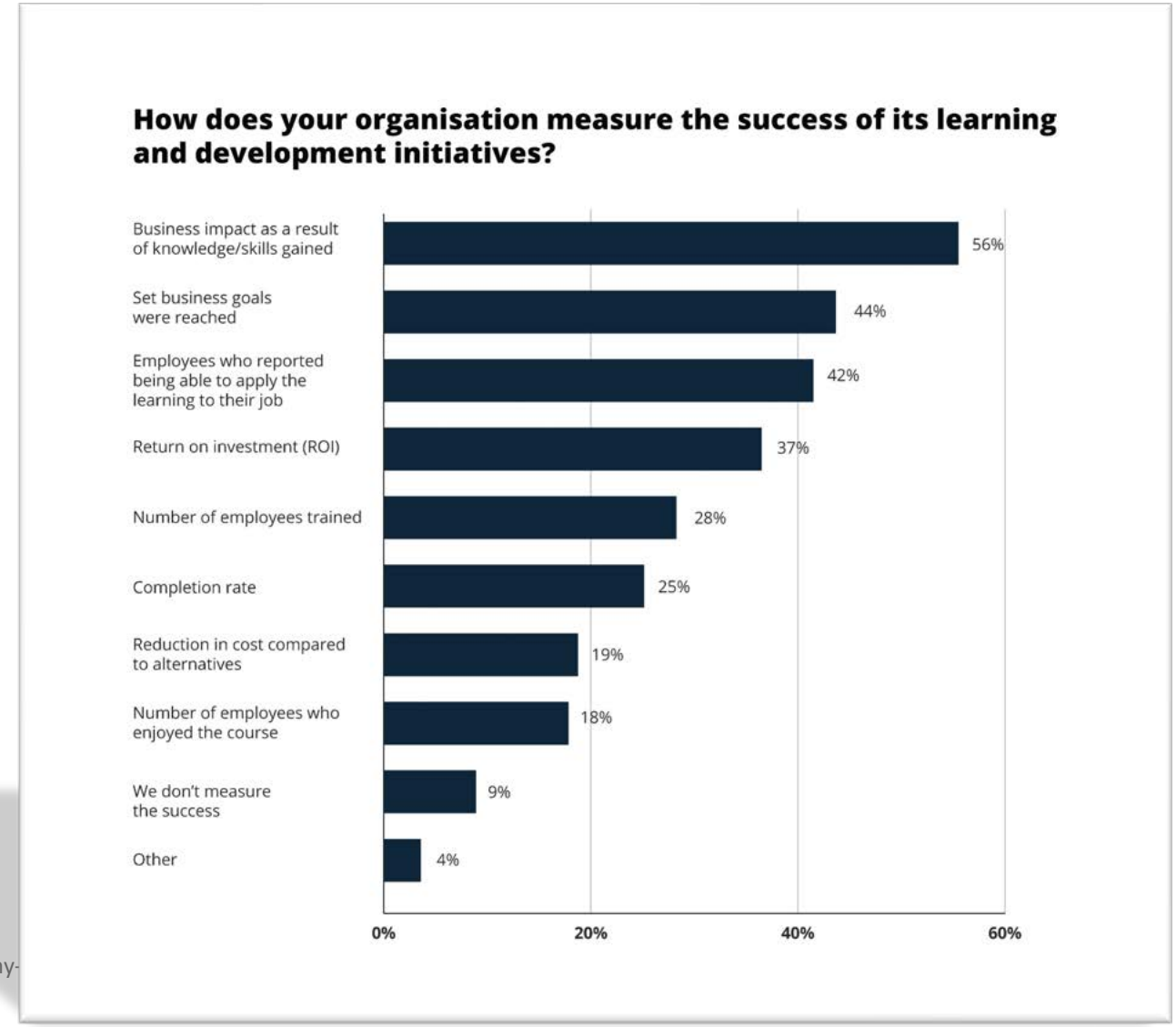
Value

The Changing Landscape of Learning



https://www.getsmarter.com/blog/career-advice/why-its-important-for-corporates-to-encourage-lifelong-learning/

The Changing Landscape of Learning



https://www.getsmarter.com/blog/career-advice/why-its-important-for-corporates-to-encourage-lifelong-learning/

Question 1

What are the observed challenges brought by the traditional modes of learning of relational query processing?

Core Topics of Relational Query Processing

Topics

- Set of physical operators
- Query processing models
- Selection of query execution plans
- Cost estimation of a query plan

Modes of Learning

- Seminars, lectures
- Textbooks, online resources
- Off-the-shelf RDBMS

Is Understanding Query Execution Plans (QEP) a Challenge?

Sem Y:

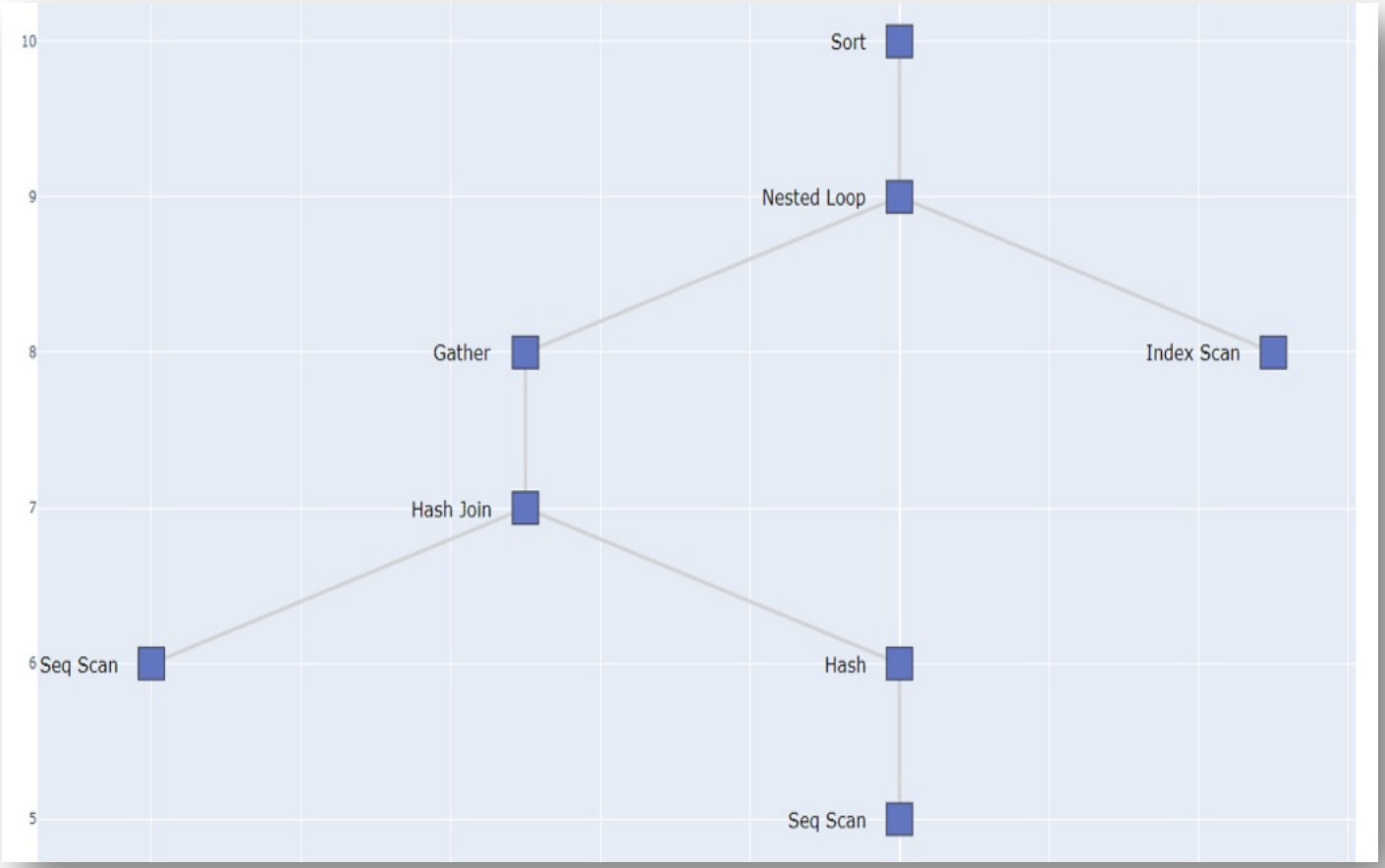
Class size: 162

Avg score: 7.4/10

Sem Y+1:

Class size: 359

Avg score: 7.1/10

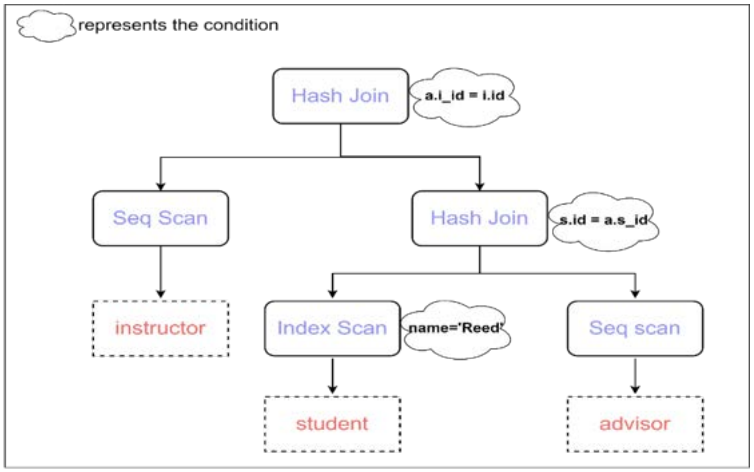


Is Understanding QEPs a Challenge to Learners?

Common Mistakes

- Incorrect ordering of steps
- Use relational algebra
- Writing SQL query
- Lumping several steps into single step
- Exclude filtering conditions in scan
- Missing intermediate relations
- Unclear specification of physical operators
- Attempt to describe implementation of various operators

Understanding Alternative Plan Choices Made By DBMS



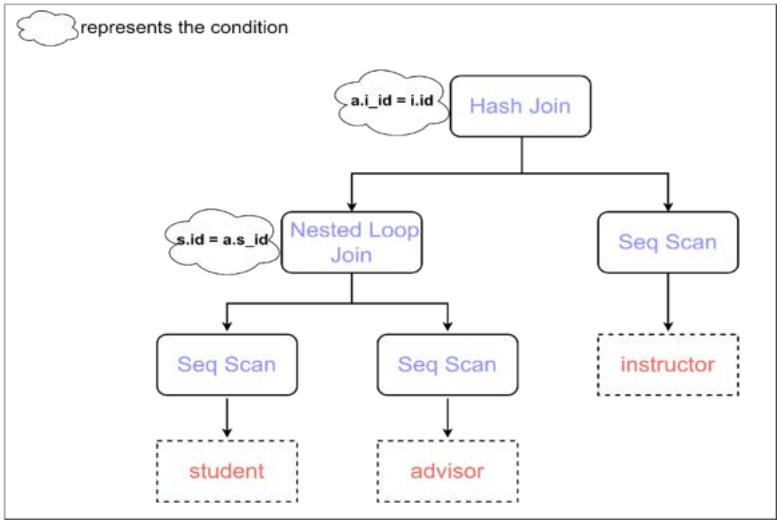
Plan A

Sem Z:

Class size: 188

Avg score: 5.1/15

Only 7.4% score 8 and above

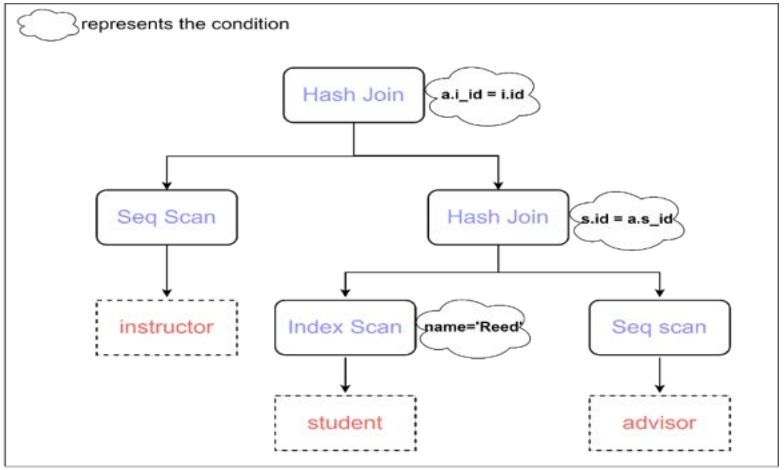


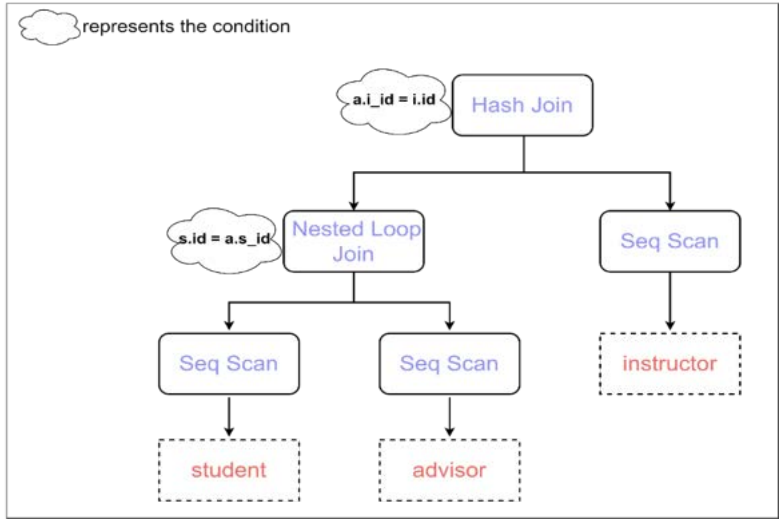
Plan B

Understanding Alternative Plan Choices Made by DBMS

Common Mistakes

- Missing role of index scan
- Missing the possible impact of join ordering
- Explanation missing
- Incorrect justification w.r.t. the type of join operator

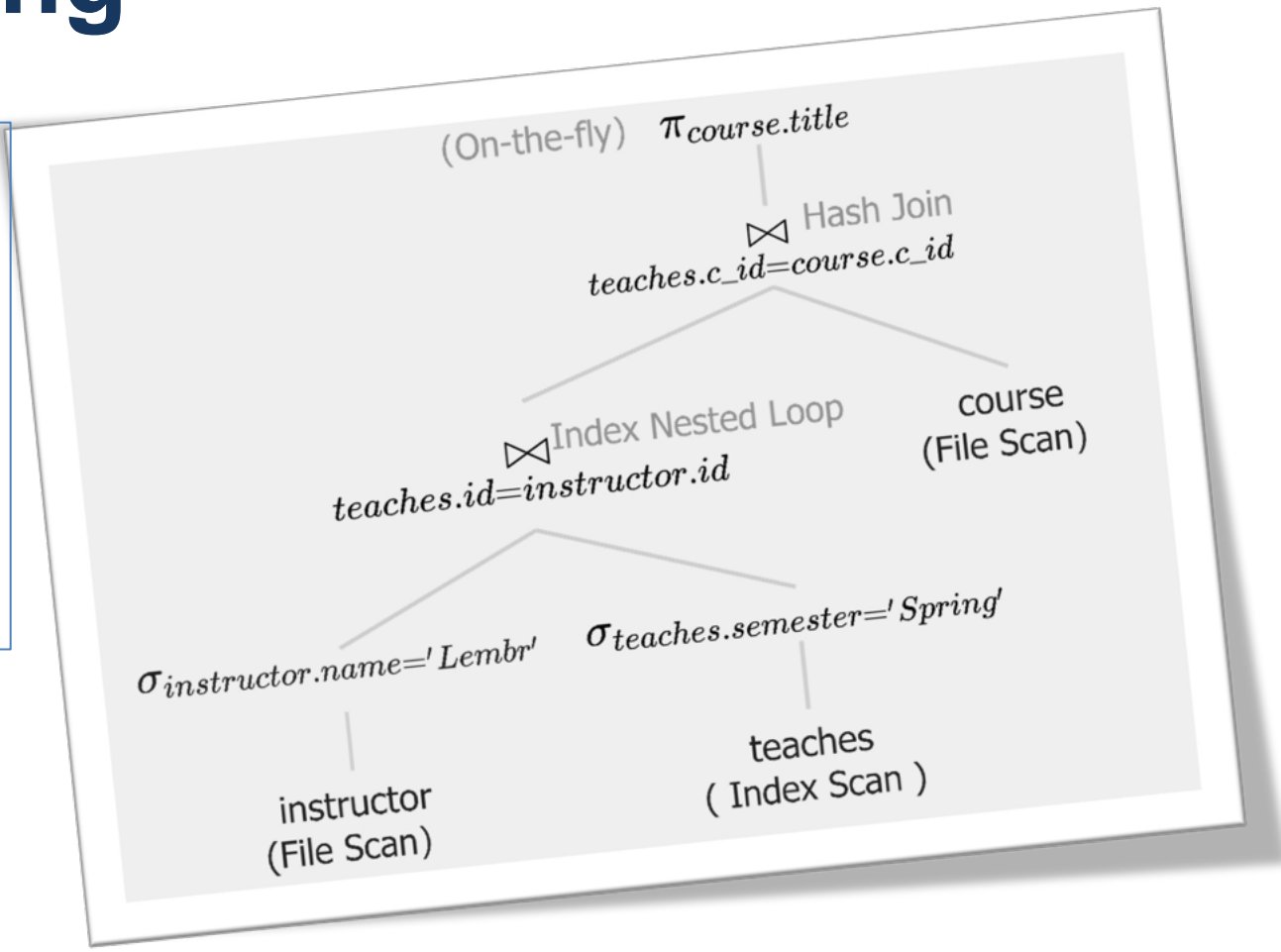




The Cost Estimation Challenge in Learning

Sem Y+1:

- 55% students scored less than 6/10
- 2/359 students got the cost estimation correct



The Cost Estimation Challenge In Learning

Common Mistakes

- Incorrect cardinality estimation of intermediate results
- Incorrect I/O cost of certain operations
- Inclusion of main memory cost

DBMS	Learners
Incorrect cardinality estimation	Incorrect cardinality estimation
Deep learning-based techniques	How to facilitate “deep” learning?

Limitations of Learning Modes

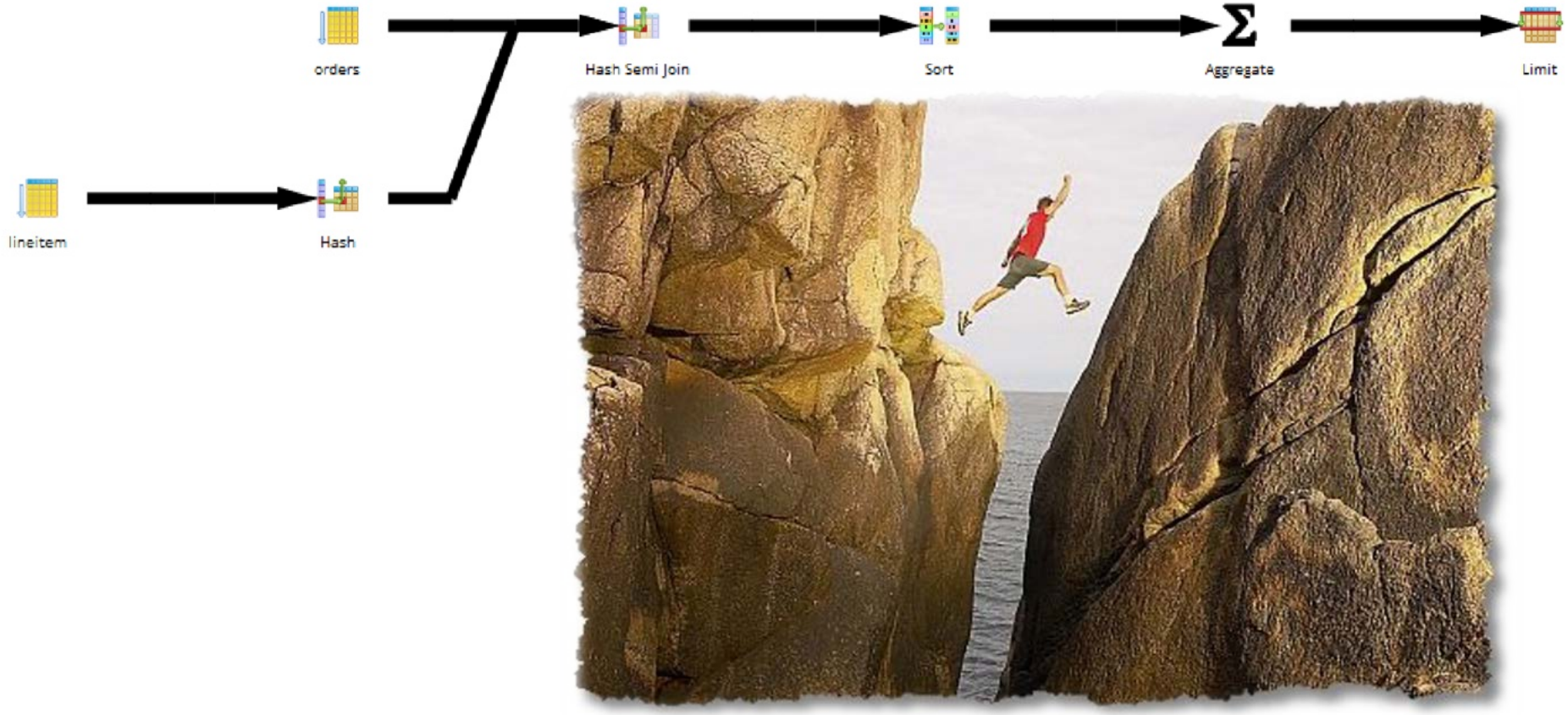
Textbook, lectures

- Limited, hard-coded example problems
- Typically deals with simple SQL to illustrate concepts
- Not interactive
- You cannot learn about any SQL queries!

Off-the-shelf RDBMS

- For enterprise users
- Not designed for pedagogical support

All We Can Get from an RDBMS (Easily)….



Observations About Learners in Traditional Learning Environment

Learners in traditional settings

- Largely extrinsically motivated (e.g., getting good grades).
- Learn-by-example.
- Avoid exhaustive online search for resources and examples.
- Prefer slide decks and videos over textbooks.
- Limited by time due to concurrent courses.
- Massing vs spacing.

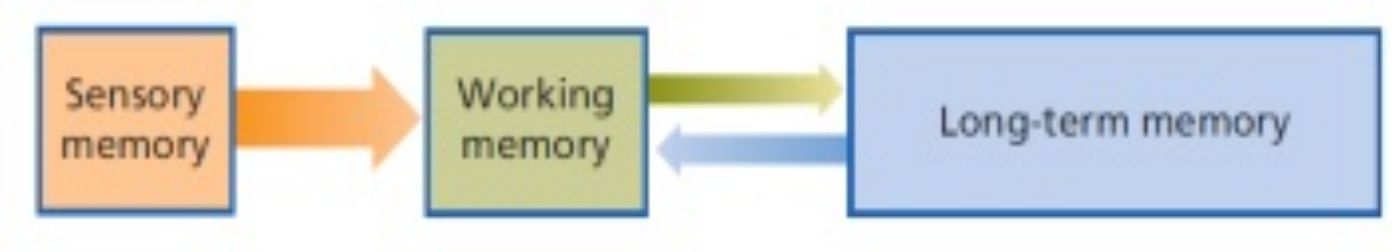
Question 2

Why learners face learning challenges in traditional settings?

Memory

Memory

- Takes meaningless sensory information (e.g., sound of professor's voice) as input
- Changes it into meaningful patterns (words, sentences, concepts) you can store and use later.



Memory is generally thought to be divided into three stages of processing

Basic Tasks of Memory

Encoding

- Memories for concepts usually require deliberate encoding effort to establish a usable memory
- Elaboration

Storage

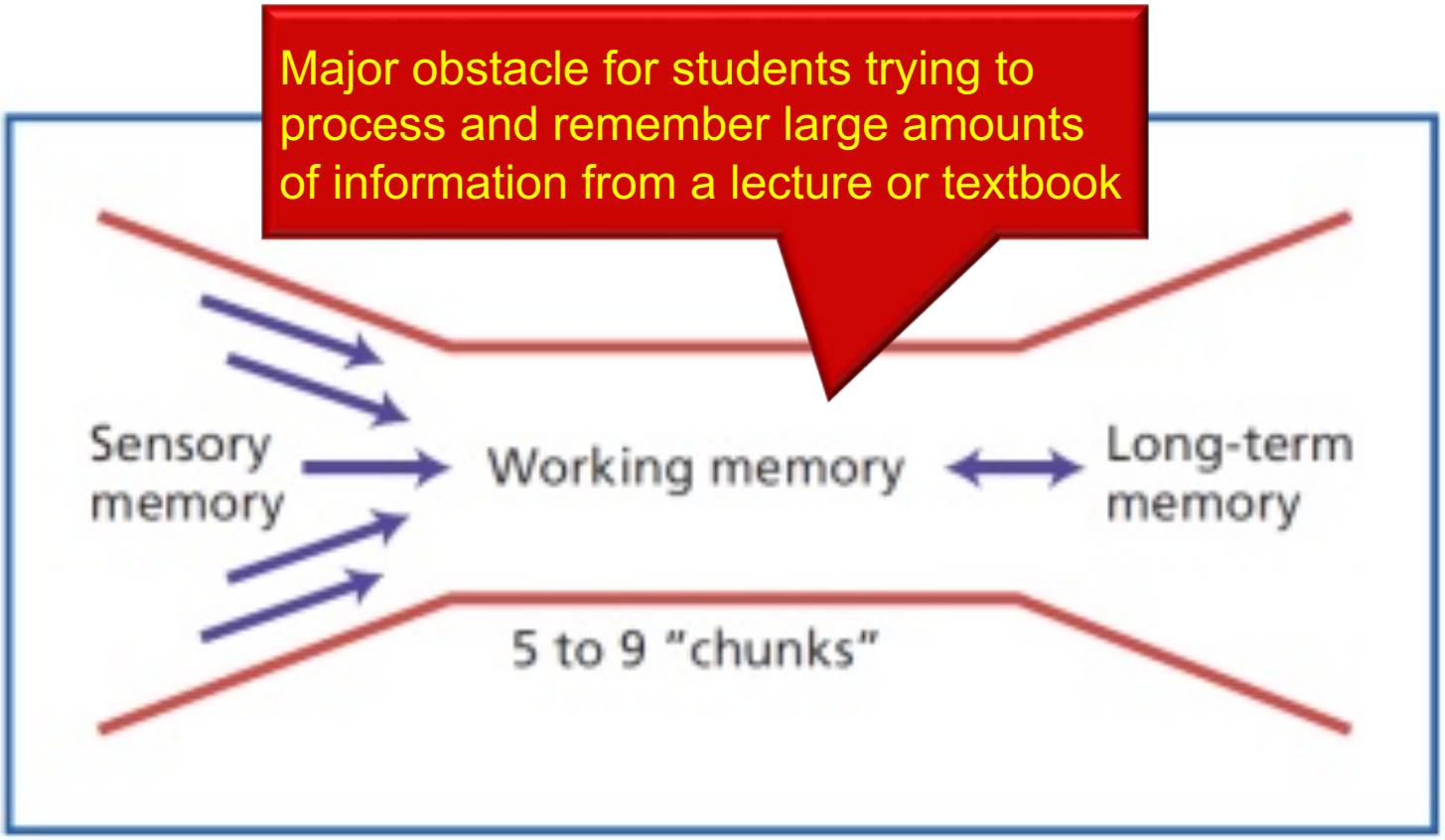
- Retention of encoded material over time.

Retrieval

- Retrieve encoded memory accurately by exploiting good cue to access the information

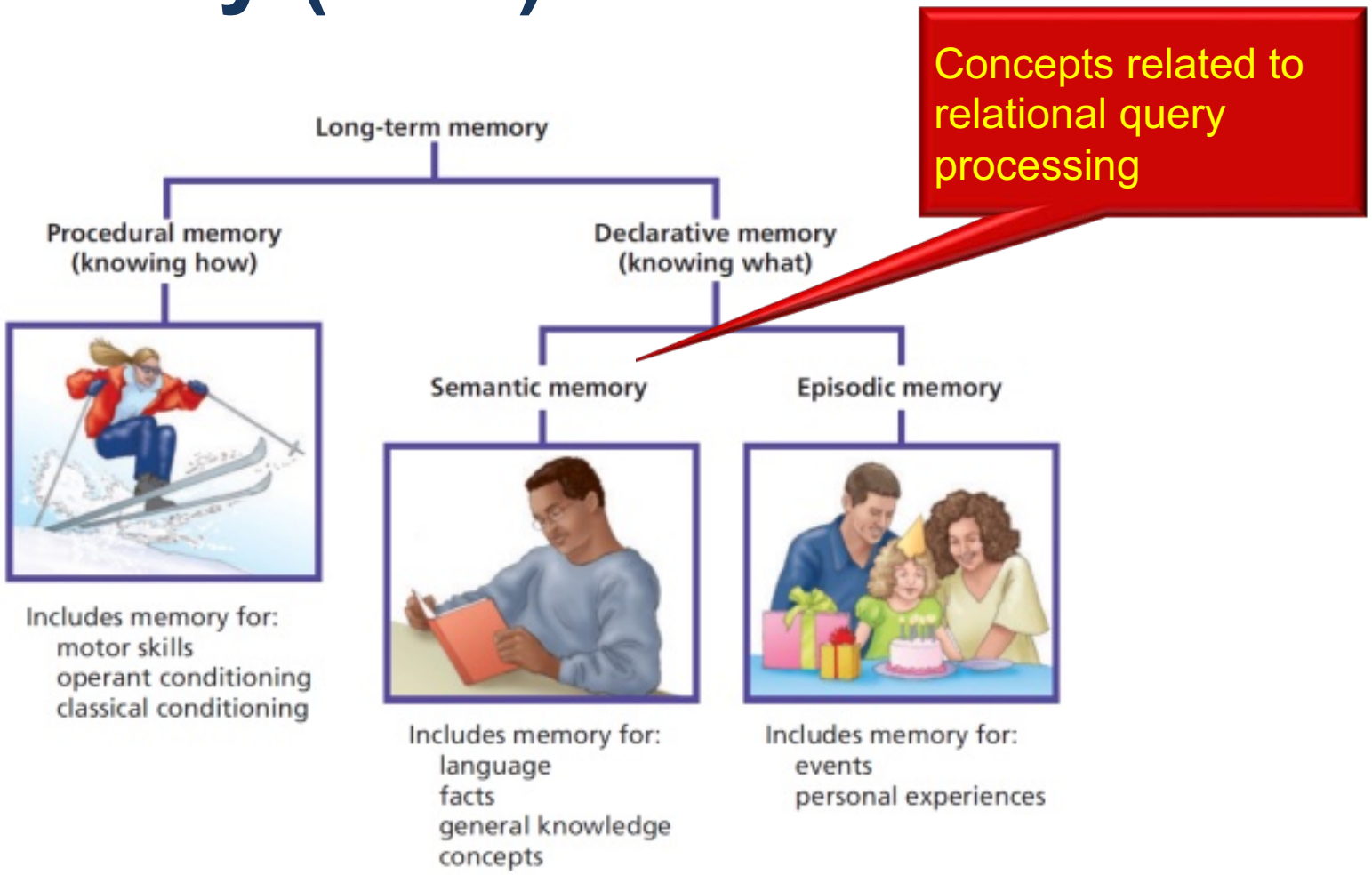
Successful retrieval depends on how they were encoded and how they are cued.

How Do We Form Memory?



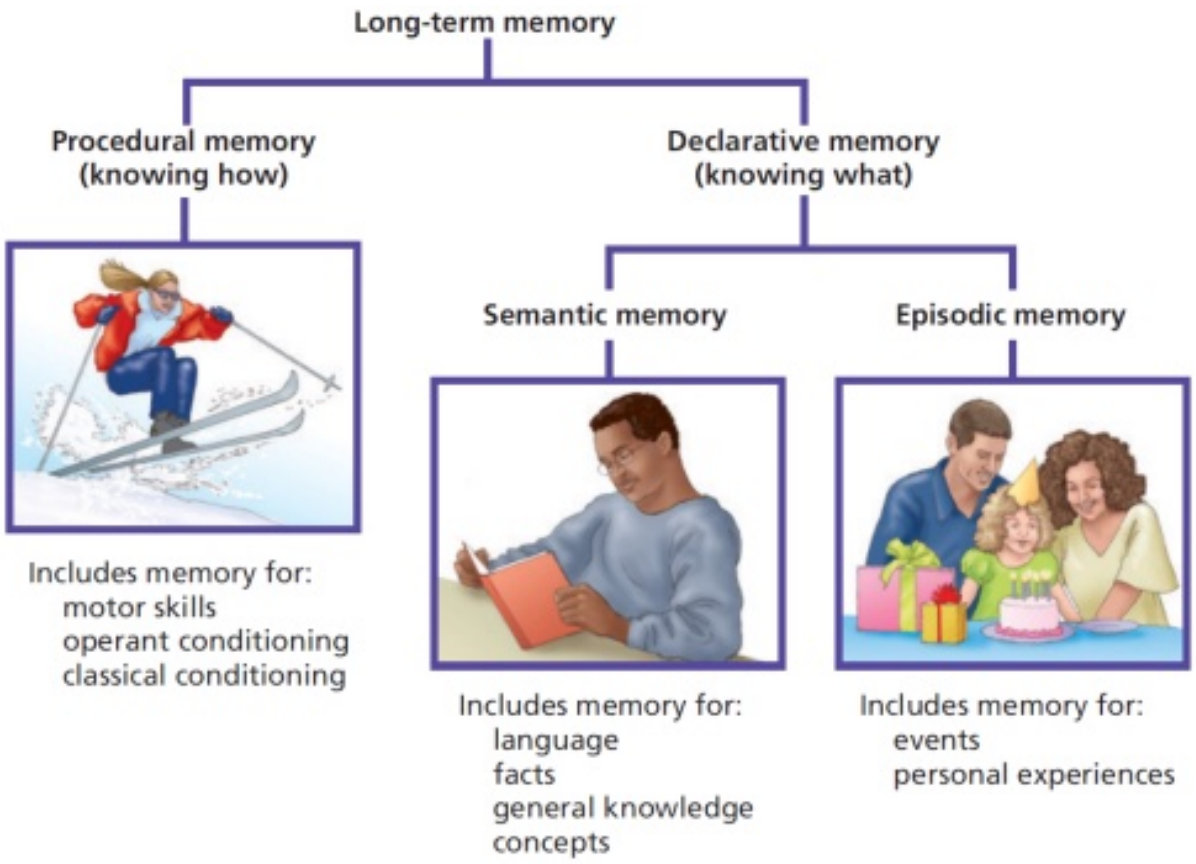
P. Zimbardo, R. Johnson, V. McCann. Psychology Core Concepts. Pearson Education, Inc., 8th Edition, 2016.

Components of Long-Term Memory (LTM)



P. Zimbardo, R. Johnson, V. McCann. Psychology Core Concepts. Pearson Education, Inc., 8th Edition, 2016.

Biological Basis of LTM



Hippocampus and amygdala are crucial to laying down new declarative memories

- Memories gradually become more permanent with the help of the hippocampus.
- Memory consolidation.

“Seven Sins” of Memory

Sin	Description	Example
Transience	Decreasing accessibility of memory over time	Simple forgetting of long-past events
Absent-mindedness	Lapses of attention that result in forgetting	Forgetting location of car keys
Blocking	Information is present but temporarily accessible	Tip-of-the-tongue
Misattribution	Memories are attributed to an incorrect source	Confusing a dream for a memory
Suggestibility	Implanted memories about things that never occurred	Leading questions produce false memories
Bias	Current knowledge and beliefs distort our memories of the past	Recalling past attitudes in line with current attitudes
Persistence	Unwanted recollections that we can never forget	Traumatic war memories

P. Zimbardo, R. Johnson, V. McCann. Psychology Core Concepts. Pearson Education, Inc., 8th Edition, 2016.

Level-of-processing Theory

Lessons from Psychology

- The more connection you can make in working memory between new information and knowledge you already have, the more likely you are to remember it later

Level-of-processing Theory

- Craik and Lockhart (1972)
- “Deeper” processing establish more connections with LTM making new information more meaningful and memorable

P. Zimbardo, R. Johnson, V. McCann. Psychology Core Concepts. Pearson Education, Inc., 8th Edition, 2016.

How Do We Facilitate Deeper Processing?

Elaborative Rehearsal

- Putting concepts into your own words
- Adding examples that illustrate the concept – a type of elaborative rehearsal

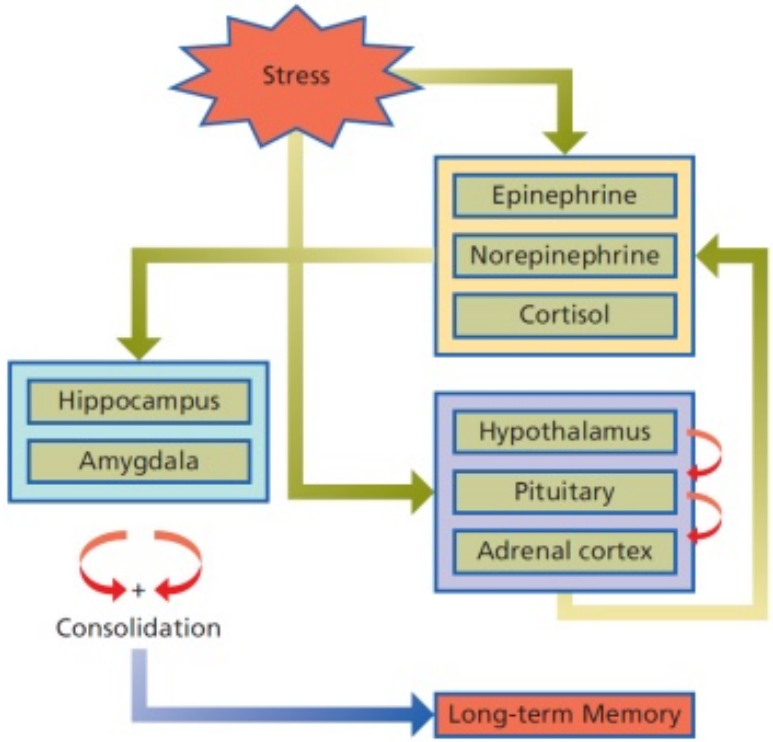
Multi-modal interactions

- Multiple modes of interactions with the course material help to build a greater web of associations into which a memory is embedded.
- Learning theories: Learners learn better when the same content can be approached in multiple ways - both visual and verbal, as well as through hands-on learning.

Memory Consolidation – Biological Basis

Biological Explanation
• New experiences consolidate much more rapidly (through hippocampus) if they are associated with existing memory schemas • Why elaborate rehearsal and depth of processing help us to create more lasting memories

The Role of Stress



P. Zimbardo, R. Johnson, V. McCann. Psychology Core Concepts. Pearson Education, Inc., 8th Edition, 2016.

Release stress hormones in the brain, which act on the amygdala and hippocampus to strengthen the emotional memory of the event.

Remember how stressful DB course is instead of the concepts ☺ ☺

Expectancy-Value Theory

If a task is too difficult or easy to complete then one may not engage with it.

Flow Theory

A psychological state where a learner is intrinsically motivated to learn. Task is neither too easy or too difficult.

Limitations of Traditional Modes

- Limited hard-coded examples
- Limited modes of approaching a content
- Processing a large amount of lecture and textbook content
- Stress due to massing, difficulty of accessing content

→

- Lesser web of association forming in long-term memory.
- Bottleneck of working memory.
- Impacts encoding and storage.
- Impacts memory consolidation in LTM.
- Expectancy Value Theory
- Flow Theory

Question 3: Towards Technology-Enabled Learning

Can we build technologies to supplement learning of relational query processing?

Broad Goals

Interactive

- Multi-modal, interactive mode
- Unlimited on-demand examples to facilitate elaborative rehearsal
- Create more web of connections

Easy

- Reduce difficulty in accessing information and content
- Facilitate operant conditioning through negative reinforcement
- Facilitate Expectancy value and flow theories
- Reduce stress

Understand

- Understand learning of learners through interaction data
- Feedback loop to improve pedagogy

Enhancing Learning Through Involvement

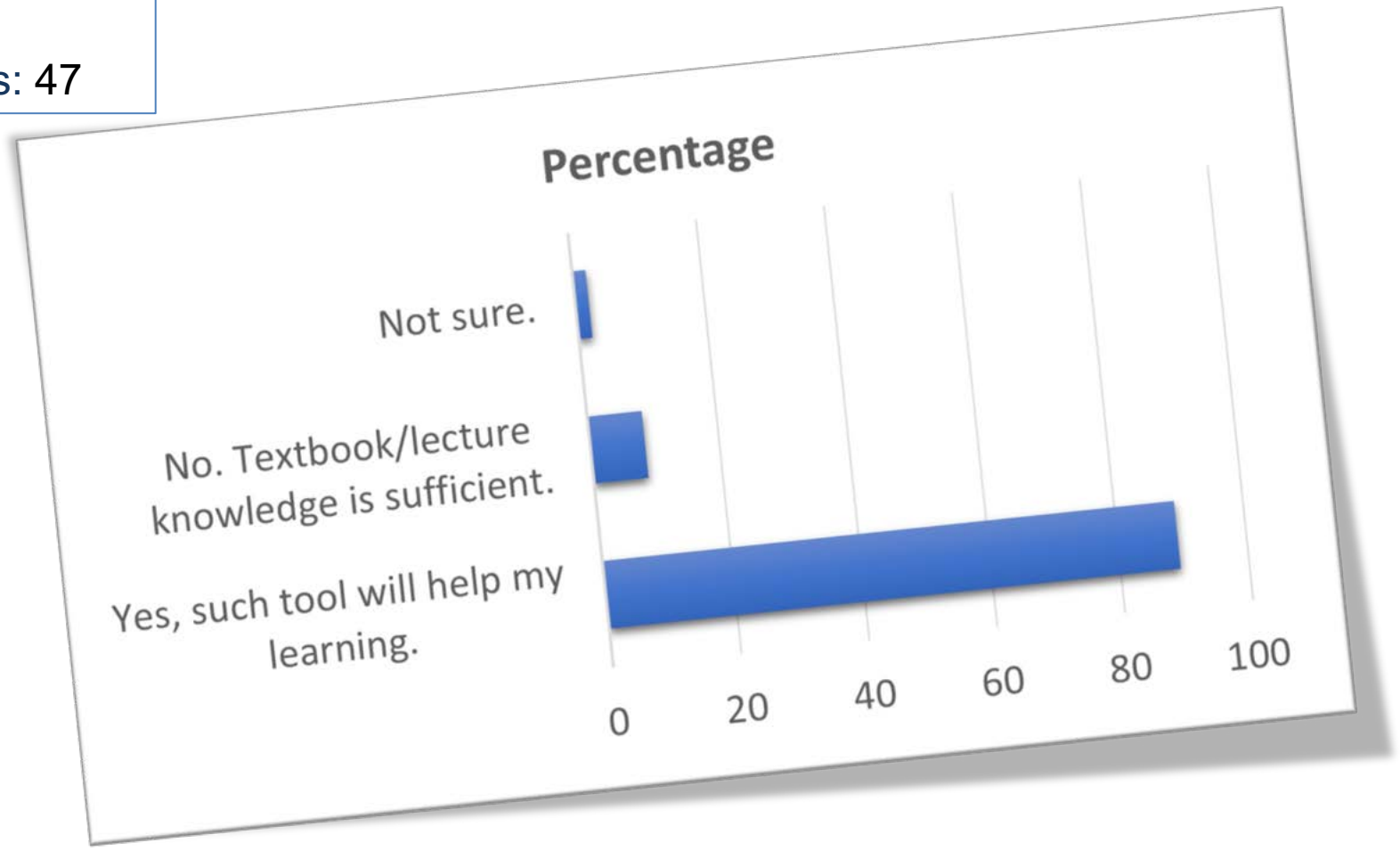
"Tell Me and I Forget, Teach Me and I May Remember, Involve Me and I Learn."

Benjamin Franklin

What Learners Think Of Tools to Augment Learning?

Sem X:

Respondents: 47



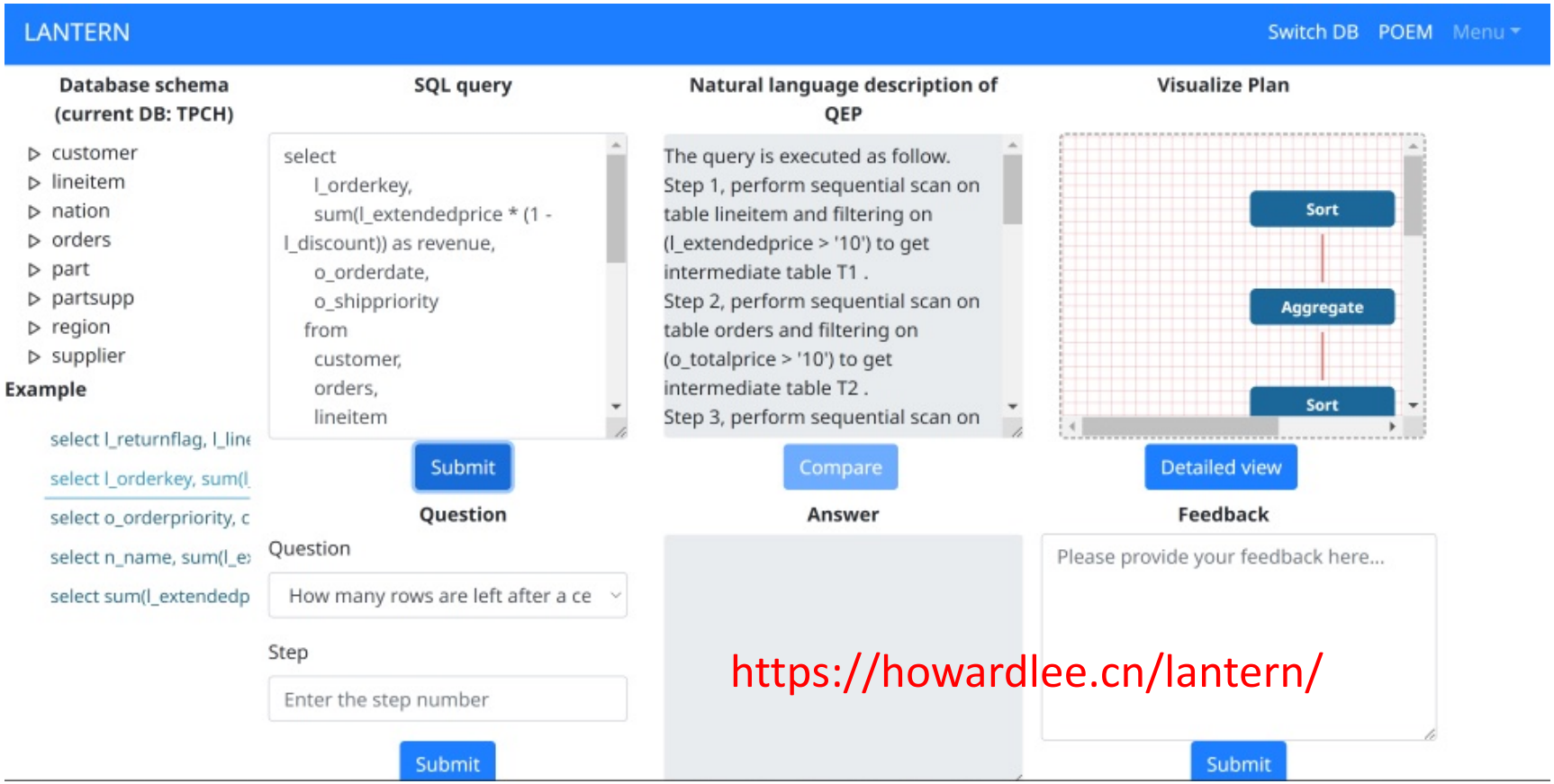
Understanding QEPs

Can we describe a QEP using visual and natural language to encourage involvement-based learning?

Challenges

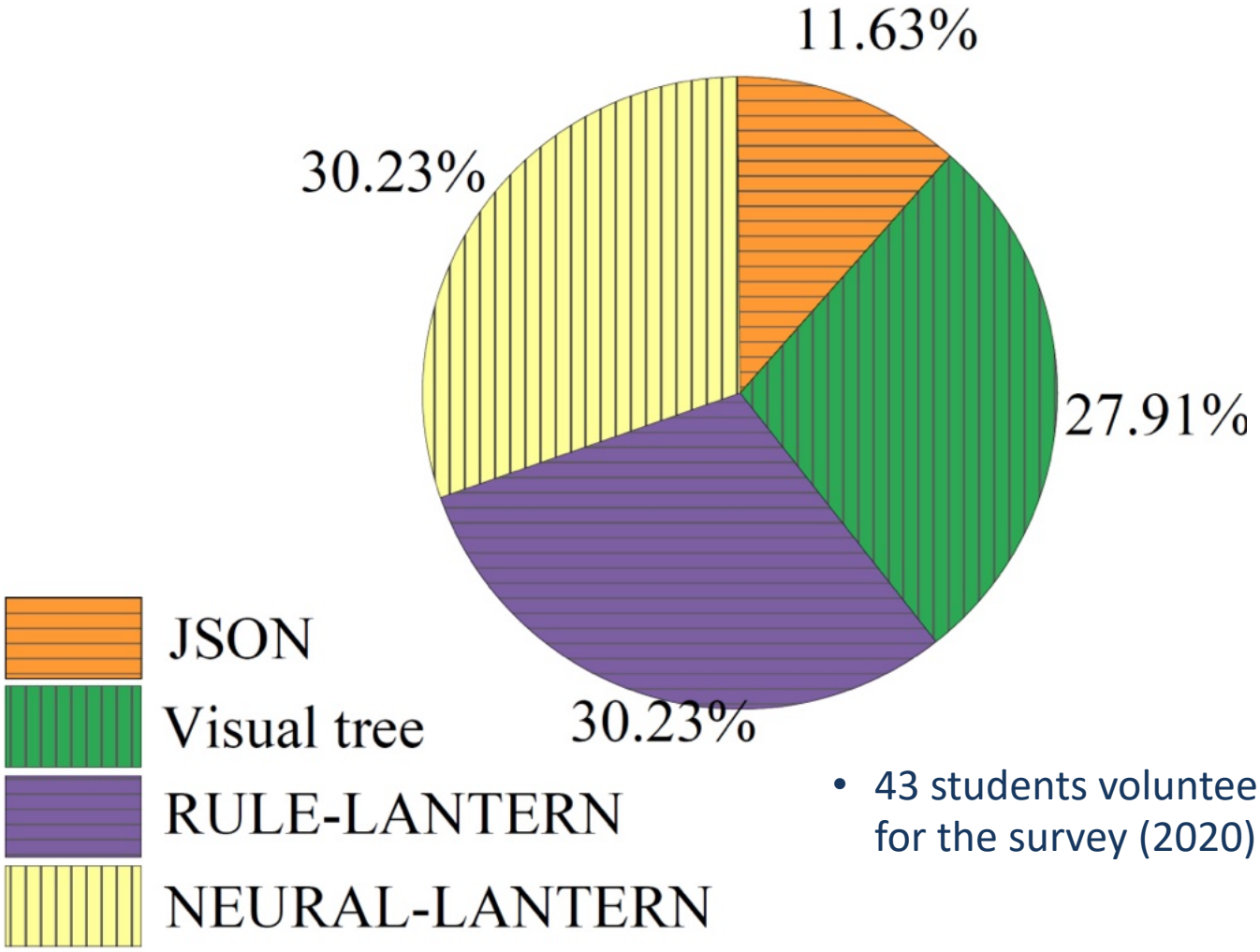
- Large QEP->NL training data is infeasible
- Rule-based generation may create boredom
- Should be generalizable w.r.t. application domain and RDBMS

NEURON & LANTERN



- **NEURON: Query Execution Plan Meets Natural Language Processing For Augmenting DB Education.** Siyuan Liu, Sourav S Bhowmick, Wanlu Zhang, Shu Wang, Wanyi Huang, Shafiq Joty. In SIGMOD, 2019
- **Towards Enhancing Database Education: Natural Language Generation Meets Query Execution Plans.** Weiguo Wang, Sourav S Bhowmick, Hui Li, Siyuan Li, Shafiq Joty, Peng Chen. In SIGMOD, 2021.
- **LANTERN: Boredom-conscious Natural Language Description Generation of Query Execution Plans for Database Education.** Peng Chen, Hui Li, Sourav S Bhowmick, Shafiq R Joty, Weiguo Wang. In SIGMOD, 2022.

User Feedback: Which query plan format is most preferred?



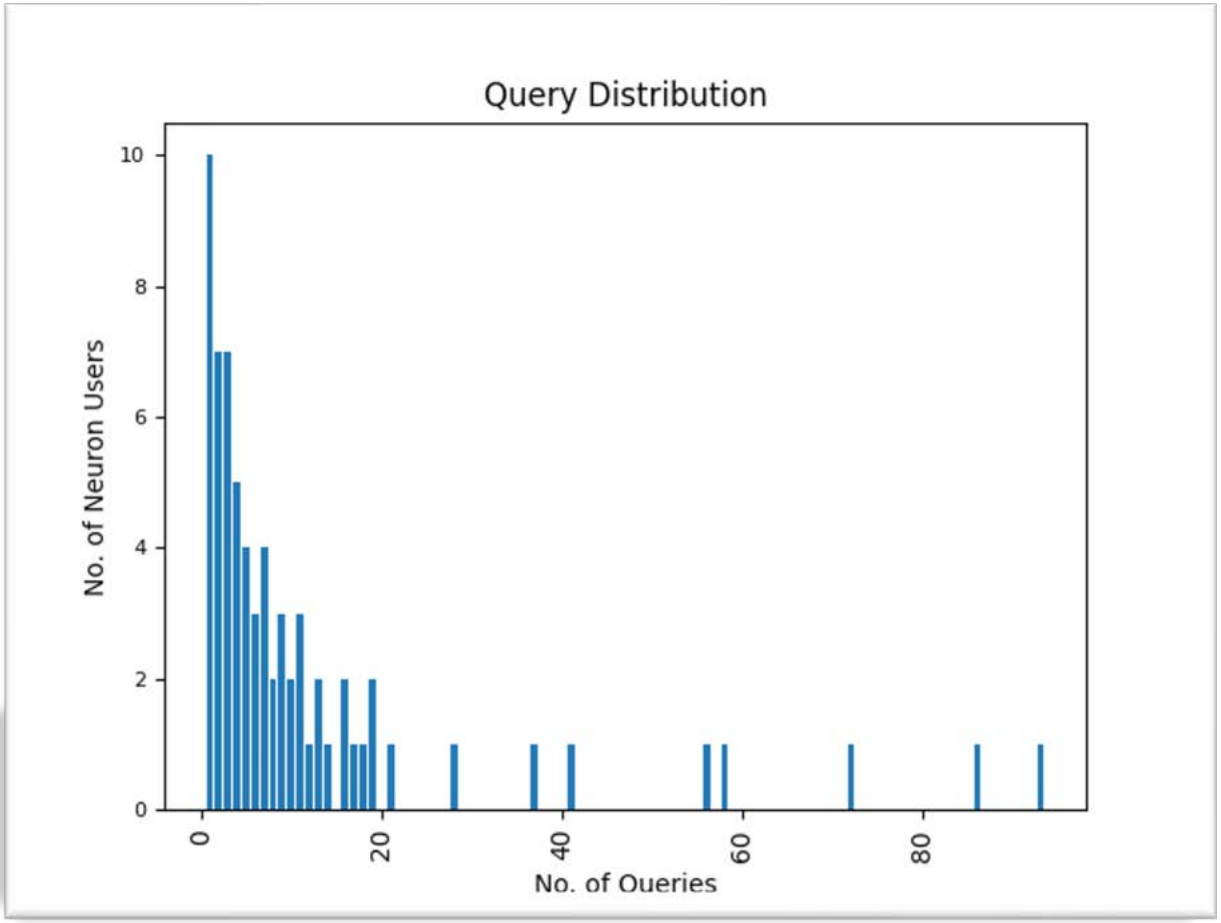
- 43 students volunteered for the survey (2020)

Impact of NEURON: When the Rubber Meets The Road

Sem X:

Class size: 166

- 41.5% used it
- No. of queries vs the number of distinct learners who posed that number of queries
- More than 85% of them posed more than one query



Sourav S Bhowmick, Hui Li. Towards Technology-Enabled Learning of Relational Query Processing. IEEE Data Engineering Bulletin, IEEE CS, September 2022

Test Performance

No. of students: 162

NEURON users

Avg: 8.43

Max: 10

Min: 6.5

Median: 8

NEURON non-users

Avg: 7.07

Max: 10

Min: 0

Median: 7.5

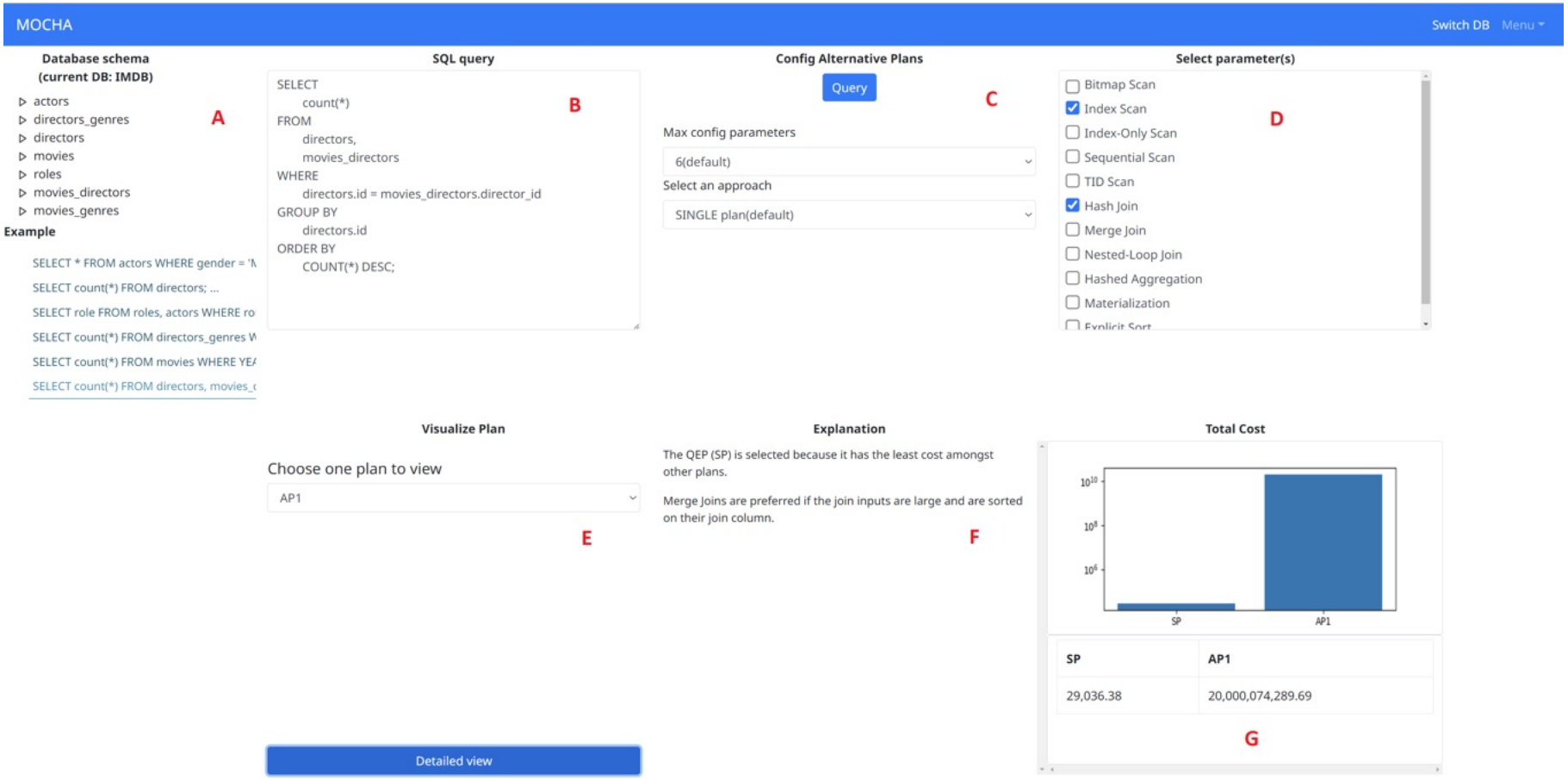
What-If Queries on QEPs

Students' Questions

- What is the impact on cost if operator A (e.g., hash join) is replaced by operator B (e.g., nested-loop join)?
- What will be the impact on cost if a specific join ordering is changed?
- What if the plan uses the index/sequential scan operator?

MOCHA

https://howardlee.cn/mocha/



MOCHA: A Tool for Visualizing Impact of Operator Choices in Query Execution Plans for Database Education. Jess Tan, Desmond Yeo, Rachael Neoh, Huey Eng Chua, Sourav S Bhowmick. In VLDB, 2022.

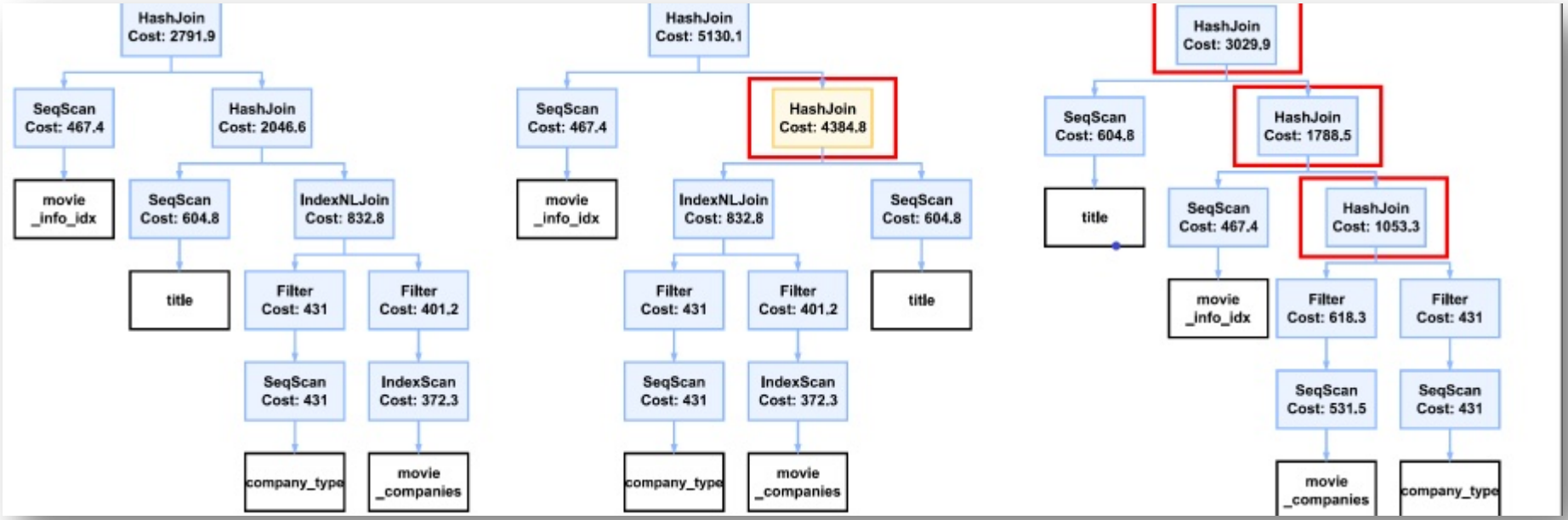
Exploring Alternative Query Plan Space

Can we explore alternative query plans in a learner-friendly manner ?

Alternative Query Plans (AQPs)

Alternative Query Plans

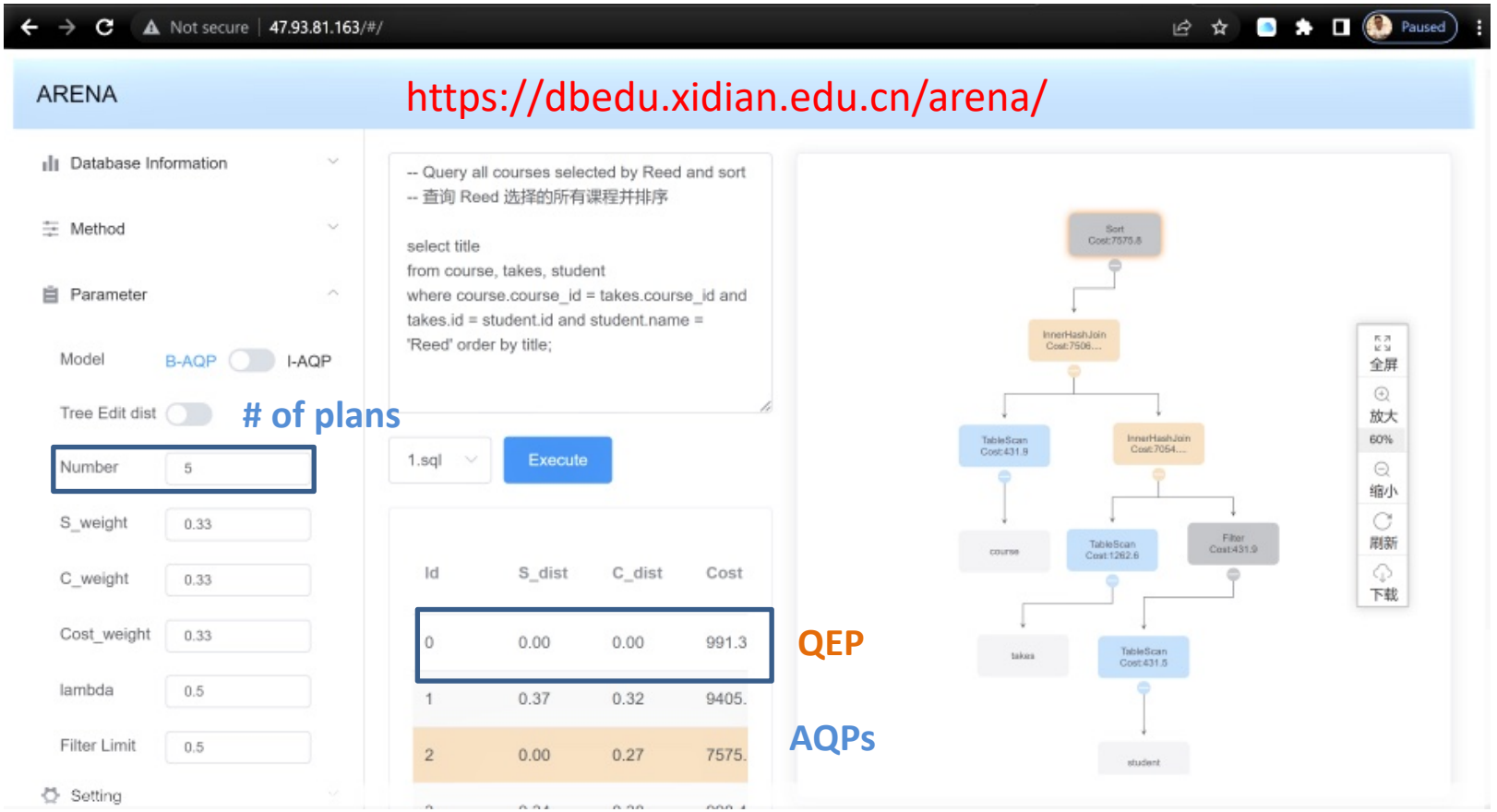
- Given an SQL query, there are many different query plans, other than the QEP, for executing it.
- Alternative query plans (AQP)



Challenges

- Which plans are informative to learners?
- How do we compute plan informativeness?
- How do we design efficient algorithms?

ARENA



ARENA: Alternative Relational Query Plan Exploration for Database Education. Hu Wang, Hui Li, Sourav S Bhowmick, Baochao Xu. In SIGMOD, June 2023

Test Performance

No. of students: 50

ARENA users (Gp 2)

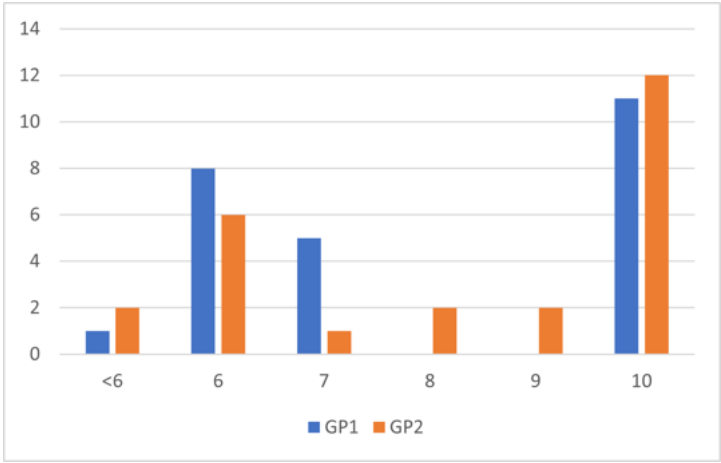
ARENA non-users (Gp 1)

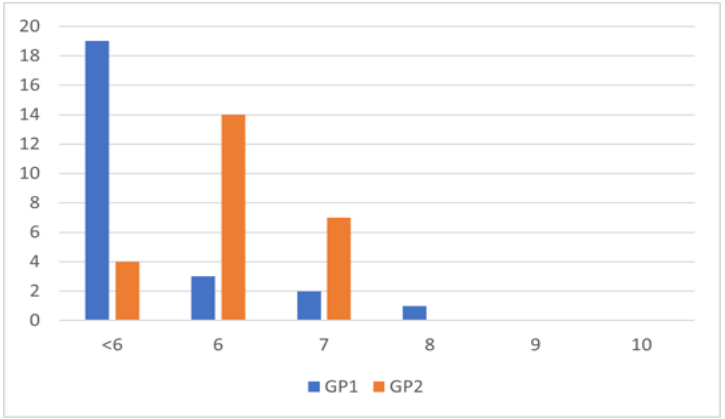
Avg: 8.24

Avg: 6.04

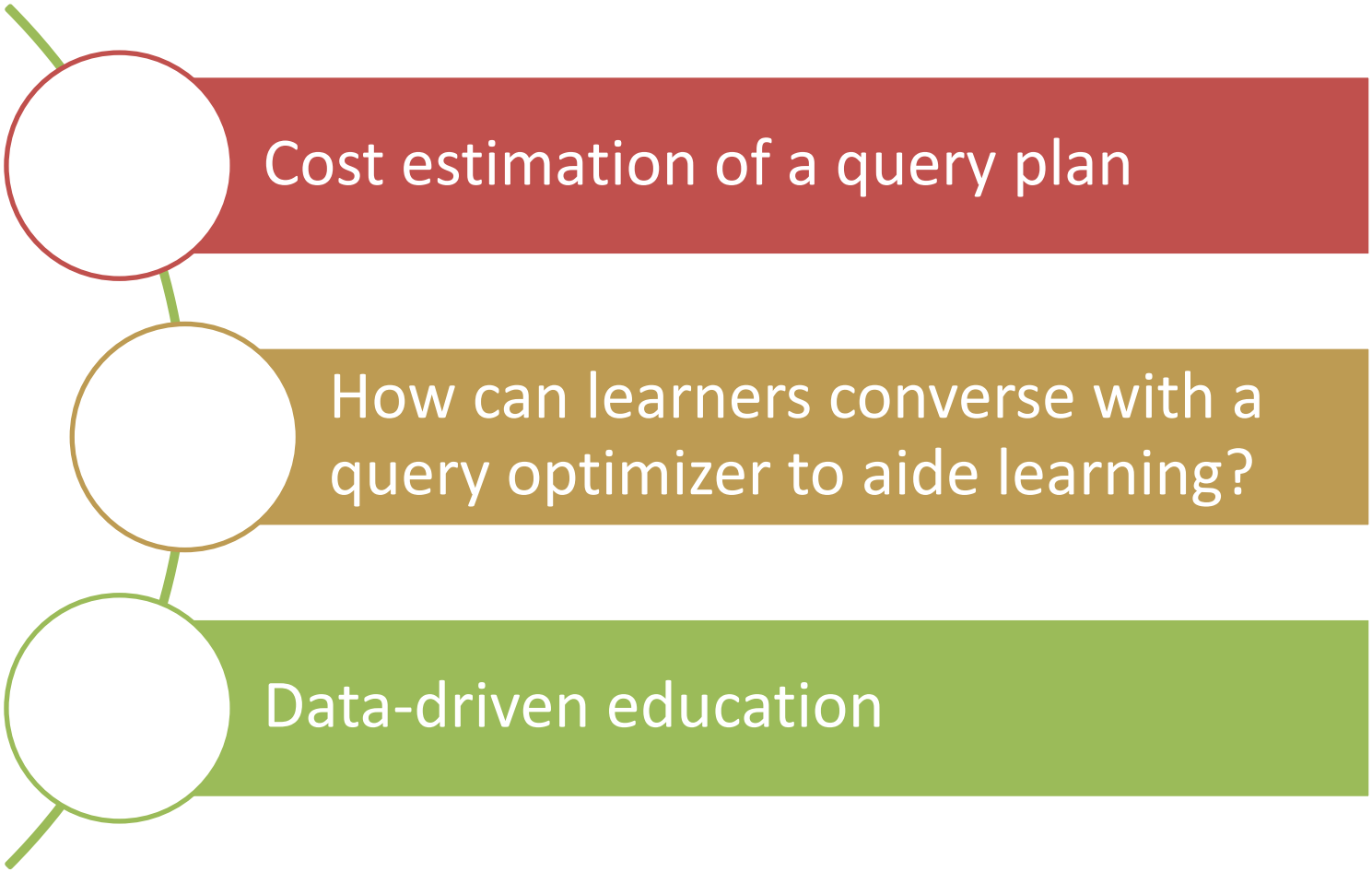
Avg: 7.76

Avg: 4.12





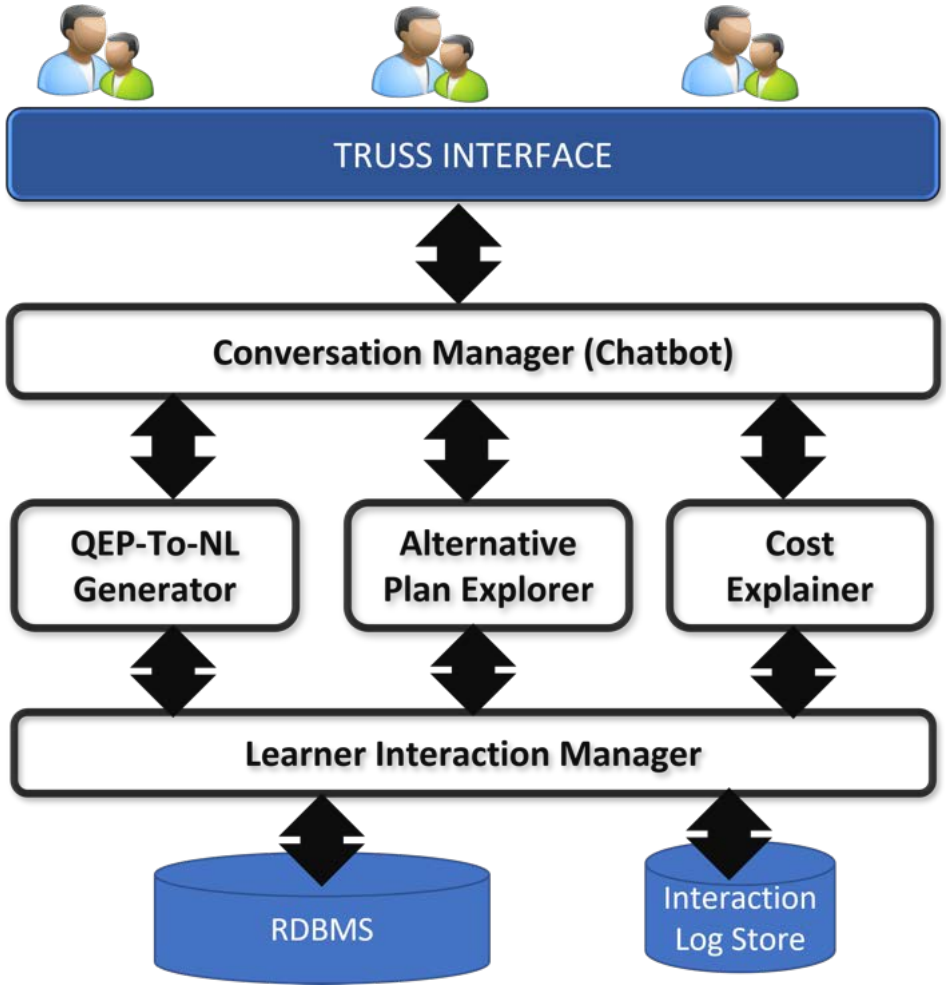
What's Next?


Cost estimation of a query plan
How can learners converse with a query optimizer to aide learning?
Data-driven education

Learning Cost Estimation - Challenges

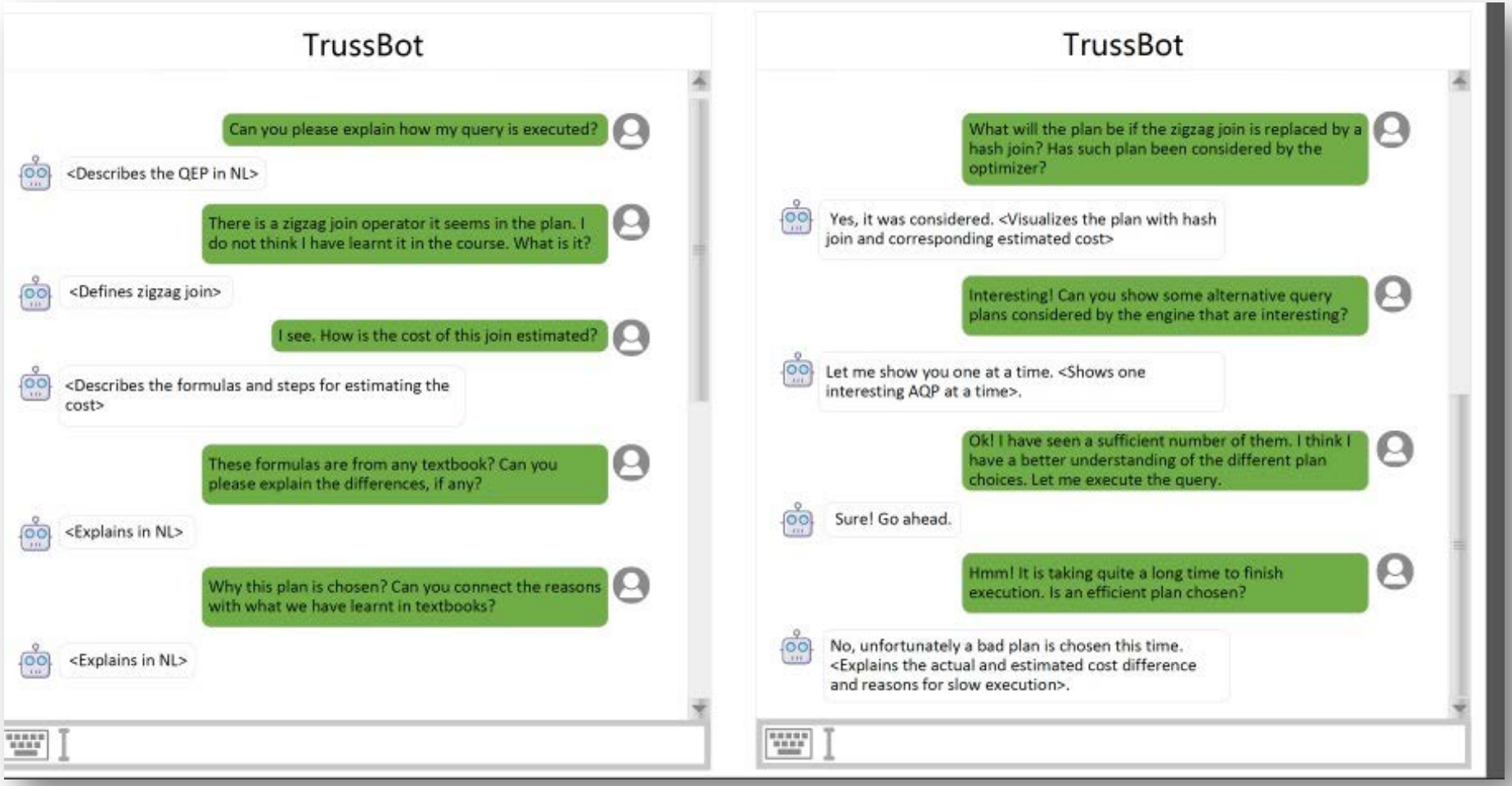
- Extract cost computation formula for subtrees
- Connecting with textbook knowledge
- Explainable and generalized approach

TRUSS System



Sourav S Bhowmick, Hui Li. Towards Technology-Enabled Learning of Relational Query Processing. IEEE Data Engineering Bulletin, IEEE CS, September 2022

Conversation with TrussBot



Towards Data-driven Education

Interaction Log

- Access time, duration, queries, interactions
- Analyze and correlate with academic outcomes

Insights

- What challenges they are facing?
- Learning preferences
- Massing vs spacing

Effectiveness

- Elaborate rehearsal and level-of-processing theory
- Any correlation between engagement of a platform and performance?

Do We Care About Disabled Learners?

DBMS for Whom?

- Primarily for able-bodied end users
- No systematic research on designing data management products for disabled users

DEI Matters!

- Diverse learners in lifelong learning environment
- How can we facilitate learning for disabled learners?

https://www.henkel.com/company/diversity-and-inclusion

How Do They Learn?

ASD/ADHD

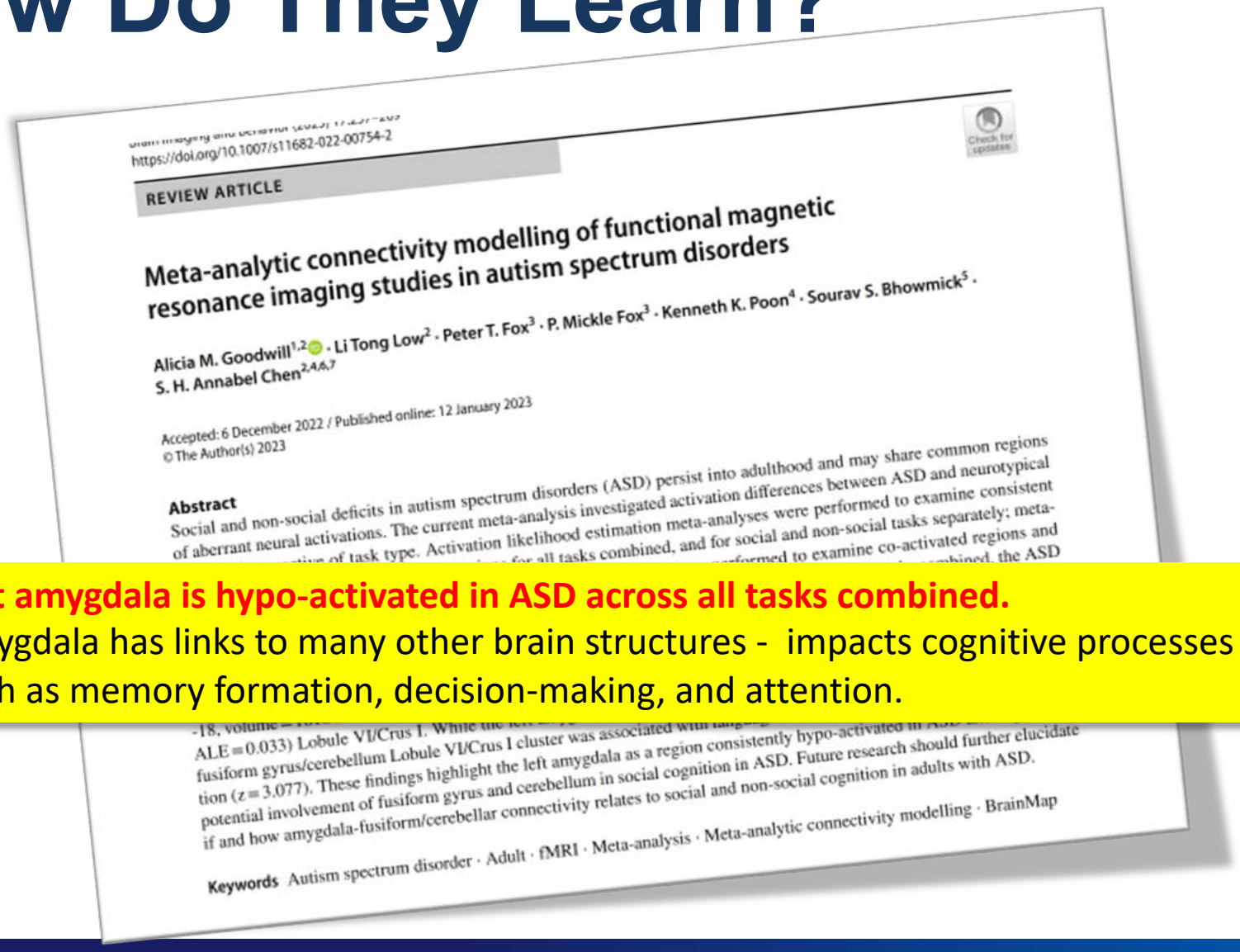
- Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD)
- Face cognitive and behavioral challenges
- Classical mode of learning may be inadequate
- Visual thinkers and learners are common among ASD

Understanding Learning

- Differential understanding of adult brain network and how it impacts learning.

Visual Tools That Can Help People With Autism Learn and Thrive. https://www.verywellhealth.com/visual-thinking-and-autism-5119992, 2021.

How Do They Learn?



https://doi.org/10.1007/s11682-022-00754-2

REVIEW ARTICLE

Meta-analytic connectivity modelling of functional magnetic resonance imaging studies in autism spectrum disorders

Alicia M. Goodwill[1,2] · Li Tong Low[2] · Peter T. Fox[3] · P. Mickle Fox[3] · Kenneth K. Poon[4] · Sourav S. Bhowmick[5] · S. H. Annabel Chen[2,4,6,7]

Accepted: 6 December 2022 / Published online: 12 January 2023

Abstract
Social and non-social deficits in autism spectrum disorders (ASD) persist into adulthood and may share common regions of aberrant neural activations. The current meta-analysis investigated activation differences between ASD and neurotypical ... of task type. Activation likelihood estimation meta-analyses were performed to examine consistent ... for all tasks combined, and for social and non-social tasks separately; meta- ... performed to examine co-activated regions and ... combined, the ASD ... -18, volume = ... ALE = 0.033) Lobule VI/Crus I. While the left ... was associated with ... fusiform gyrus/cerebellum Lobule VI/Crus I cluster was associated with ... in ASD ... tion ($z = 3.077$). These findings highlight the left amygdala as a region consistently hypo-activated in ASD ... potential involvement of fusiform gyrus and cerebellum in social cognition in ASD. Future research should further elucidate if and how amygdala-fusiform/cerebellar connectivity relates to social and non-social cognition in adults with ASD.

Keywords Autism spectrum disorder · Adult · fMRI · Meta-analysis · Meta-analytic connectivity modelling · BrainMap

- **Left amygdala is hypo-activated in ASD across all tasks combined.**
- Amygdala has links to many other brain structures - impacts cognitive processes such as memory formation, decision-making, and attention.

DB Researchers Needs to Break Out from the Enterprise Jar

How can technology supplement learning of database technology?

Beyond enterprise users

Implications to data science and data management education

Efforts on Query Visualization

Principles of Query Visualization

Wolfgang Gatterbauer
Northeastern University
w.gatterbauer@northeastern.edu

Cody Dunne
Northeastern University
c.dunne@northeastern.edu

H.V. Jagadish
University of Michigan
jag@umich.edu

Mirek Riedewald
Northeastern University
m.riedewald@northeastern.edu

Abstract

Query Visualization (QV) is the problem of transforming a given query into a graphical representation that helps humans understand its meaning. This task is notably different from designing a Visual Query Language (VQL) that helps a user compose a query. This article discusses the principles of relational query visualization and its potential for simplifying user interactions with relational data.

SQLVis: Visual Query Representations for Supporting SQL Learners

Daphne Miedema, George Fletcher
Department of Mathematics and Computer Science
Eindhoven University of Technology
{d.e.miedema, g.h.l.fletcher}@tue.nl

Abstract—SQL is a typical query language for performing data analytics. Although its usage is ubiquitous, learners experience that query formulation in SQL is error-prone and time-consuming. Prior research has shown that this is due to low expressive ease, extensive training requirements and high cognitive load, all of which present a significant burden for SQL learners. Visual representations can assist learners to significantly lower this burden. The current dominant paradigm aims to facilitate SQL querying by helping users to avoid the syntax of SQL. Such Visual Querying Systems (VQS), however, are not effective for SQL learners as they hide the syntax of the language during query formulation, rather than assisting learners to write correct queries in SQL. Furthermore, training with VQSs is system specific, which leads to system dependency for learners. We argue that novices need support from Visual Query Representation (VQR) solutions which, instead, help them in learning how to write correct and portable SQL queries. In this paper we present SQLVis, a VQR to support novice SQL users in query writing. Our system represents the query as written in SQL by the user, which can improve the SQL writing proficiency of its users. Results of an in depth empirical study demonstrate the significant value of SQLVis for learners.

Index Terms—Query languages, Visualization techniques and methodologies, Computer Science education

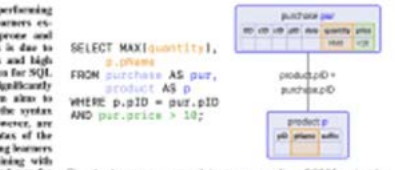


Fig. 1: A user query and its corresponding SQLVis visualization. Tables are encoded as nodes, and constraints that link tables together as edges. The renamed attributes are highlighted in orange, column based constraints in green.

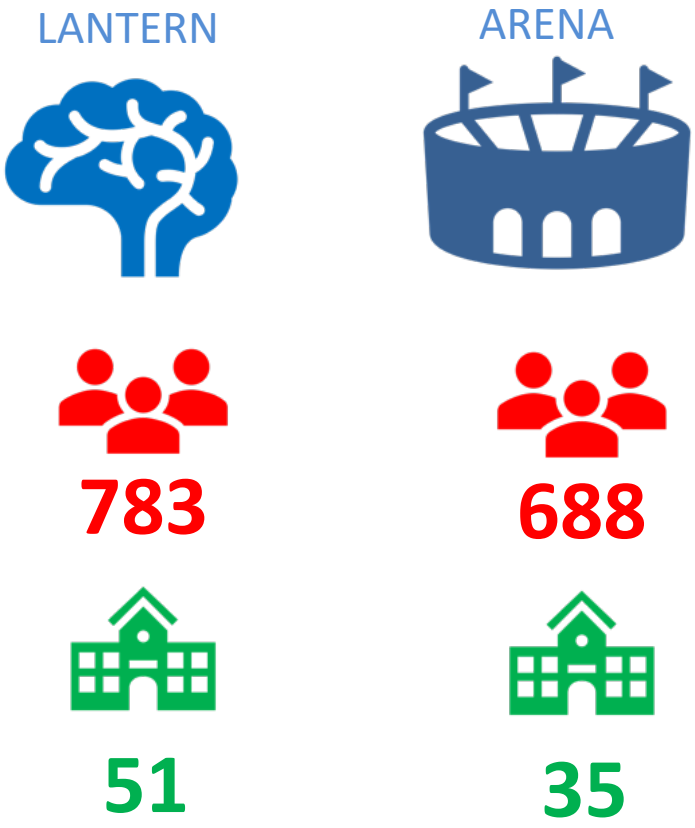
leads to increased information throughput [8]. However, they also increase system dependency, as a VQS obfuscates syntax instead of supporting the learning of SQL.

Another approach to using visualizations to support query formulation is to generate a representation from a written

Acknowledgements

Hui Li, Xidian Univ, China

- Siyuan Liu, NTU
- Weiguo Wang, Xidian
- Peng Chen, Xidian
- Zheng Li, Xidian
- Hu Wang, Xidian



* Since January 2022