DATA ANALYTICS USING DEEP LEARNING

GT 8803 // FALL 2019 // JOY ARULRAJ

SPEAKING TIPS

CREATING THE NEXT®

CREDITS

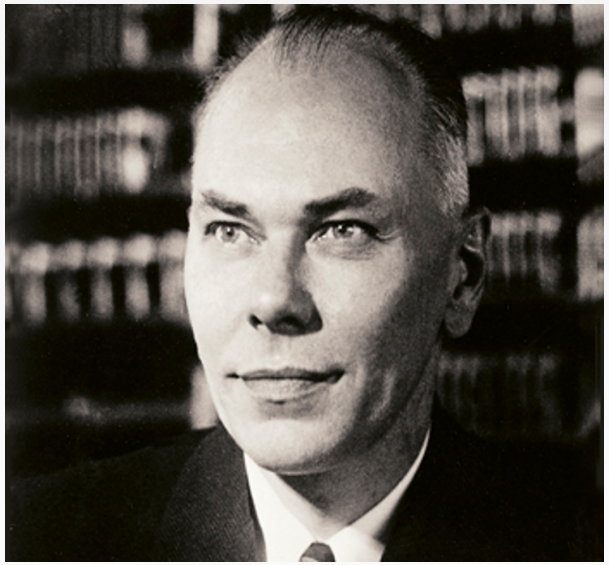
- Based on a talk given by:
  - Margaret Martonosi (Princeton)
  - Computer architect

MOTIVATION

- Communication is essential for:
  - Disseminating important results
  - Ideas don't sell themselves
  - They will lie on the shelf and gather dust unless you sell them

MOTIVATION

- Howard Aiken
  - Don't worry about people stealing an idea. If it's original, you will have to ram it down their throats.



MOTIVATION

- Communication is essential for:
  - Explaining your work to **colleagues**
  - Teaching concepts in a **class**
  - Giving talks/seminars in **industry** or academia
  - Selling your ideas to **funding agencies** (or VC firms)
  - Interviewing for **jobs**
  - Crystallizing your ideas for **research**

FORUMS FOR COMMUNICATING IDEAS

- Conference talk
- “Elevator pitch” or hallway conversation
- Poster Session
- Thesis defense or job talk

BEFORE YOU START, CONSIDER THIS…

- Who is the audience?
  - What is their background?
  - What will they know or not know?

BEFORE YOU START, CONSIDER THIS…

- What are your goals?
  - Teach them something?
  - Change their minds about something?
  - Get them to read your paper?
  - Convince someone to hire you?
- Example
  - When I talk about query execution in this class, I discuss it differently than in a research presentation.

THE FOUR QUESTIONS

- What is the problem?
- Why is it important?
- What have others done about it?
- What am I doing about it?
  - That is useful, novel, interesting, different…
- Nearly all oral and written research presentations begin from these questions

TALK OUTLINE

- Conference talk
- “Elevator pitch” or hallway conversation
- Poster Session
- Thesis defense or job talk

CONFERENCE TALKS

ORAL PRESENTATION: THE THREE MUST HAVES

- **Content**: Know your material really well
- **Design**: Organize the material and create a high-quality presentation
  - Drive home key points
  - Illustrate with figures and graphs
- **Delivery**: plan your oral presentation/what you will say along with each slide
  - practice, practice, practice

CONFERENCE TALKS

- Remember
  - There is no way you will cover every detail of a 10 page paper in 20 minutes
  - The main goal is to **get the audience interested** in your work so they go read the paper
  - The talk is that sales job (but don't overdo the selling)

A GENERAL TALK STRUCTURE (25 MINS.)

- Title/author/affiliation (1 slide)
- Motivation and problem statement (1-3 slides)
- Related work (0-1 slides)
- Main ideas and methods (7-8 slides)
- Analysis of results and key insights (3-4 slides)
- Summary (1 slide)
- Future work (0-1 slide)

A GOOD TALK IS LIKE A GOOD MUSEUM TOUR…

- Informative, easy to hear, information at the right level, just about the right length…
- Bad talks…
  - Uninformative, hard to hear, or hard to understand…
  - The tour goes on too long, so that the material stops being interesting…
  - The kidnapping: Never told where we are going or why…

THE BEGINNING…

- Tell the audience **where** we are going
- And tell the audience **why** we are going there…

OUTLINE SLIDE?

- Common to start with an outline slide, but…
  - IMHO, it's too much detail before you've told anyone what you are doing…
  - Tell the audience more about **what the destination** is, before you **detail out the route** you'll take to get there.

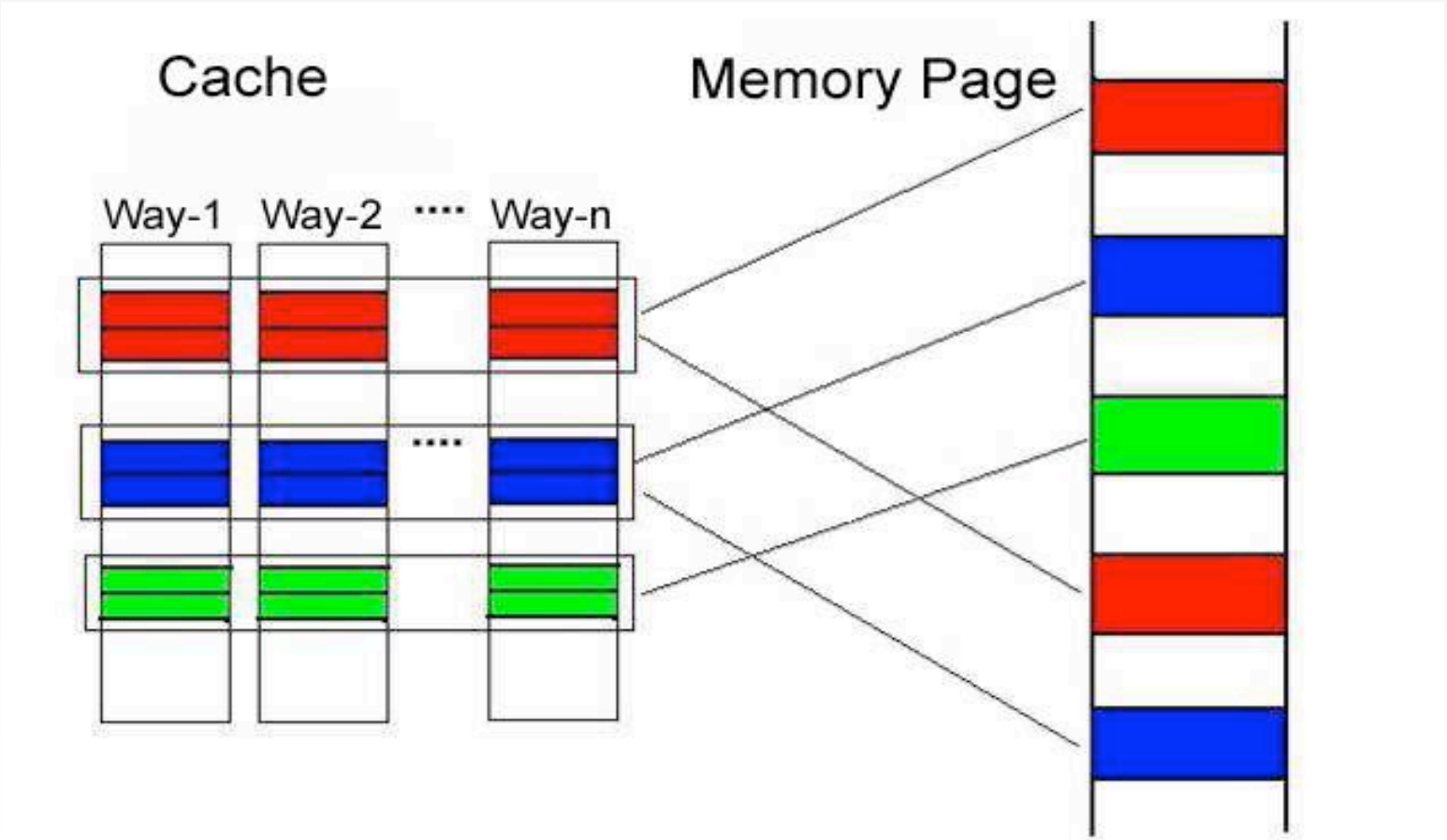
OUTLINE SLIDE?

- But if you wait too long to show the outline slide…
  - The audience starts to feel a bit lost…
  - "Where are we going?"
  - Pick a happy medium: Brief Motivation, then outline

ROADMAP

- Background
- Design
- Evaluation
- Conclusion

BACKGROUND: PAGE COLORING

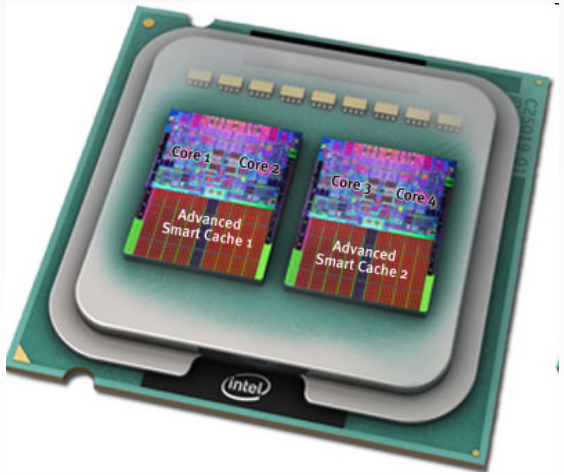


INSTEAD …

THE MULTI-CORE CHALLENGE

- Multi-core chips
  - Dominant on the market
  - Last level cache is commonly shared by sibling cores, however sharing is not well controlled

- Challenge: Performance Isolation
  - Poor performance due to conflicts
  - Unpredictable performance
  - Denial of service attacks

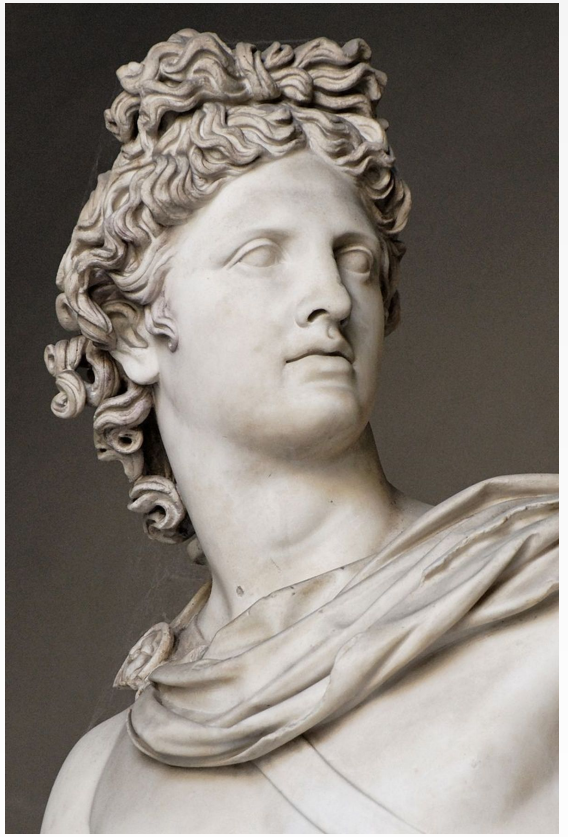


APOLLO

- Holistic toolchain for debugging database systems
  - Inspired by Jepsen

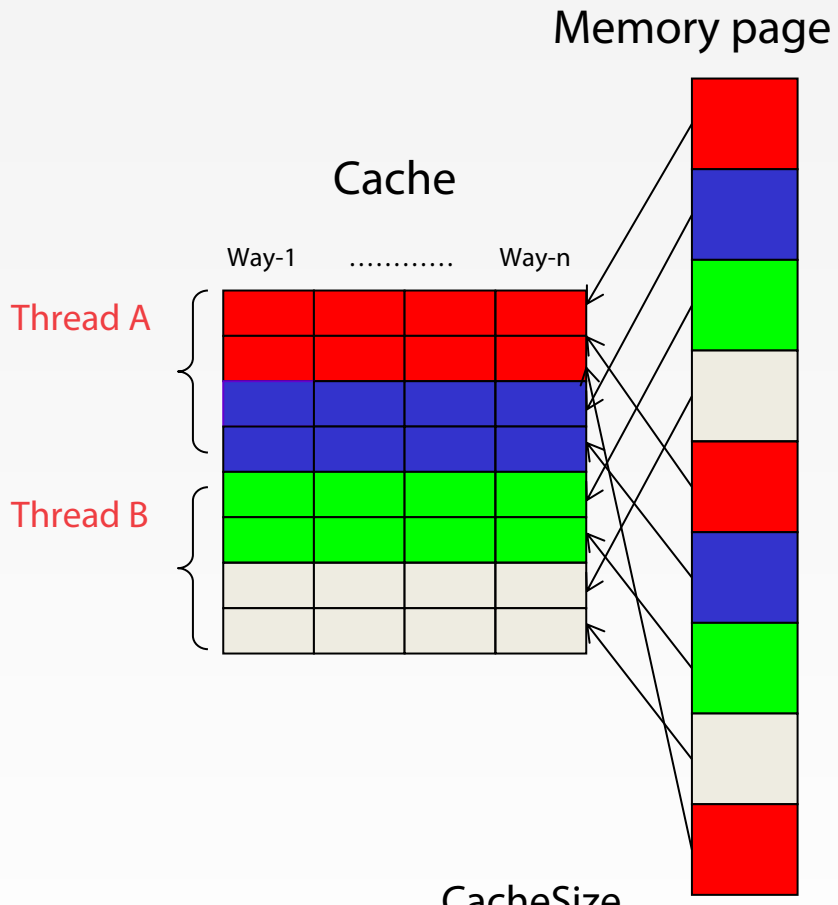
1. AUTOMATICALLY FIND SQL QUERIES EXHIBITING PERFORMANCE REGRESSIONS

2. AUTOMATICALLY DIAGNOSE THE ROOT CAUSE OF PERFORMANCE REGRESSIONS



POSSIBLE SOFTWARE SOLUTION: PAGE COLORING

- Partition cache at coarse granularity
- Page coloring: advocated by many previous works
  - [Bershad'94, Bugnion'96, Cho '06, Tam '07, Lin '08, Soares '08]
- Challenges:
  - Expensive page re-coloring
    - Re-coloring is needed due to optimization goal or co-runner change
    - Without extra support, re-coloring means memory copying
    - 3 micro-seconds per page copy, >10K pages to copy, possibly happen every time quantum
  - Artificial memory pressure
    - Cache share restriction also restricts memory share



$$\text{Color \#} = \frac{\text{CacheSize}}{\text{PageSize*CacheAssociativity}}$$

OUR WORK: HOTNESS-BASED PAGE COLORING

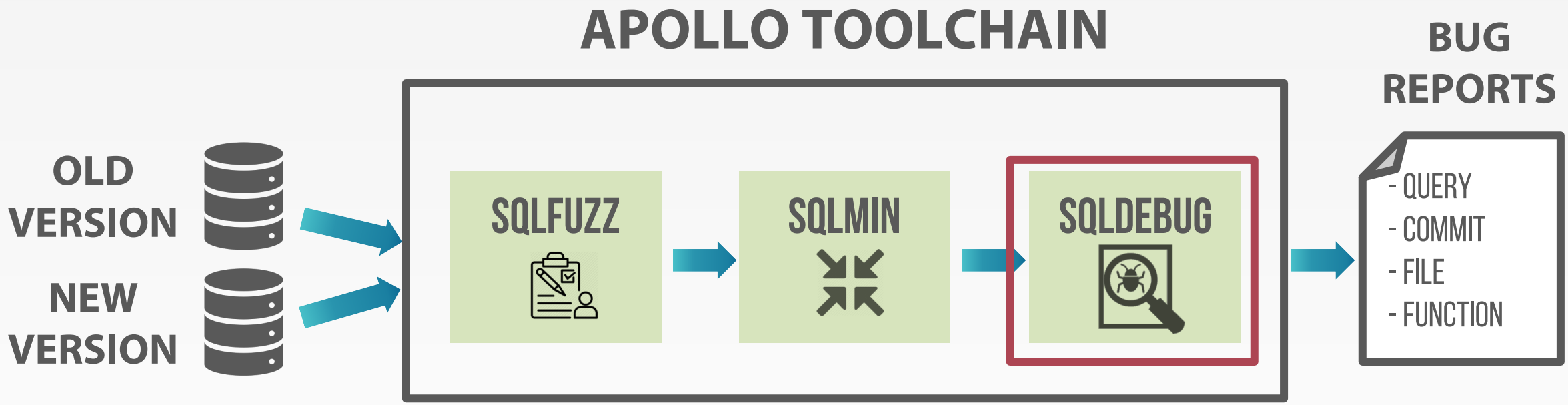
- Basic idea
  - Restrain page coloring to a small group of hot pages

- This paper's key idea:
  - How to efficiently determine hot pages

OUTLINE

- Efficient hot page identification
  - locality jumping
- Cache partition policy
  - MRC-based
- Hot page coloring

TALK OVERVIEW



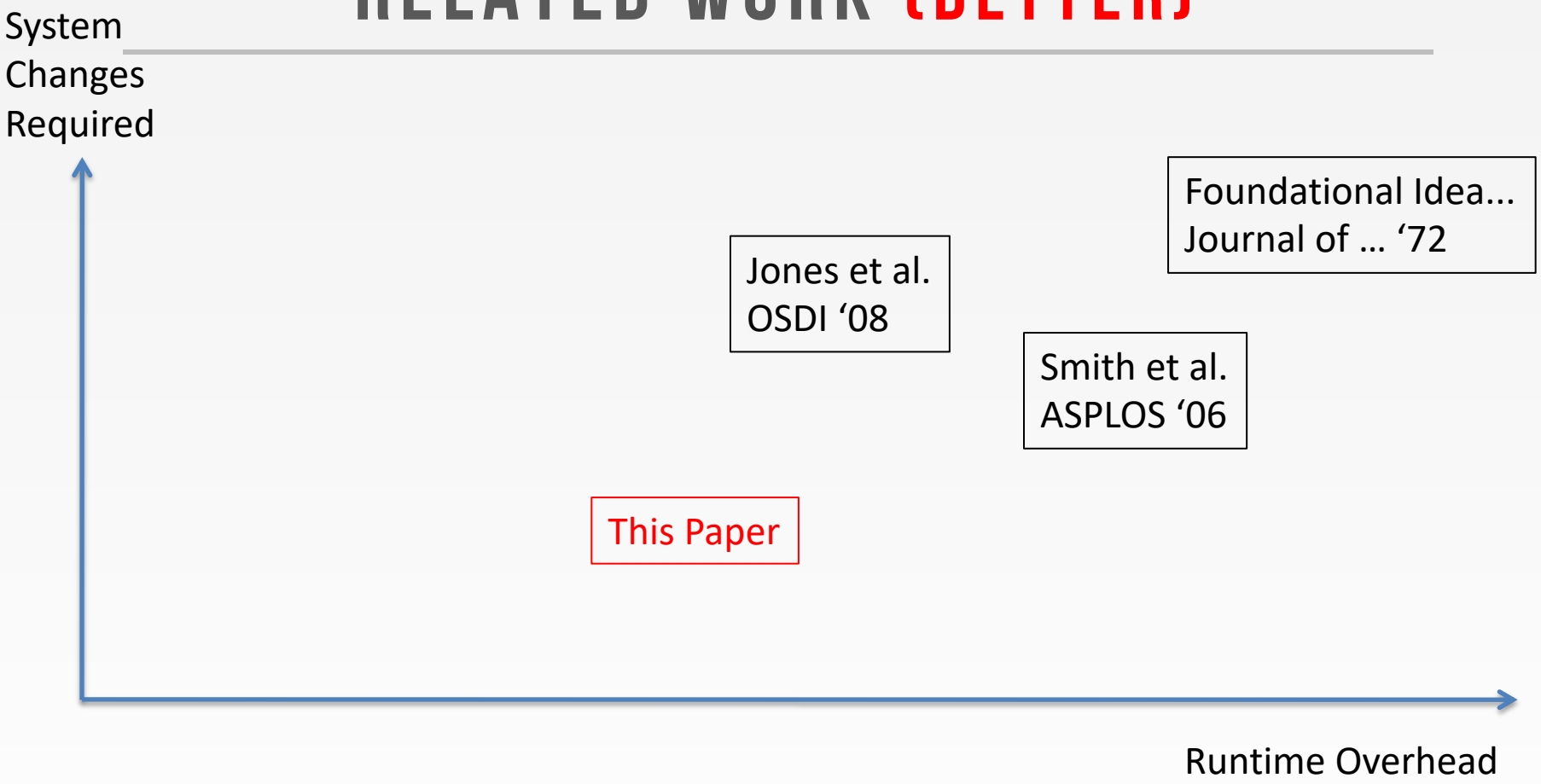
RELATED WORK

- Almost always included in a talk/paper
  - Beginning or end?
- Think about what your goal is:
  - To motivate your own work?
  - To appease the authors who are in your audience?
  - To convince the audience you are well-informed?

RELATED WORK (LESS EFFECTIVE)

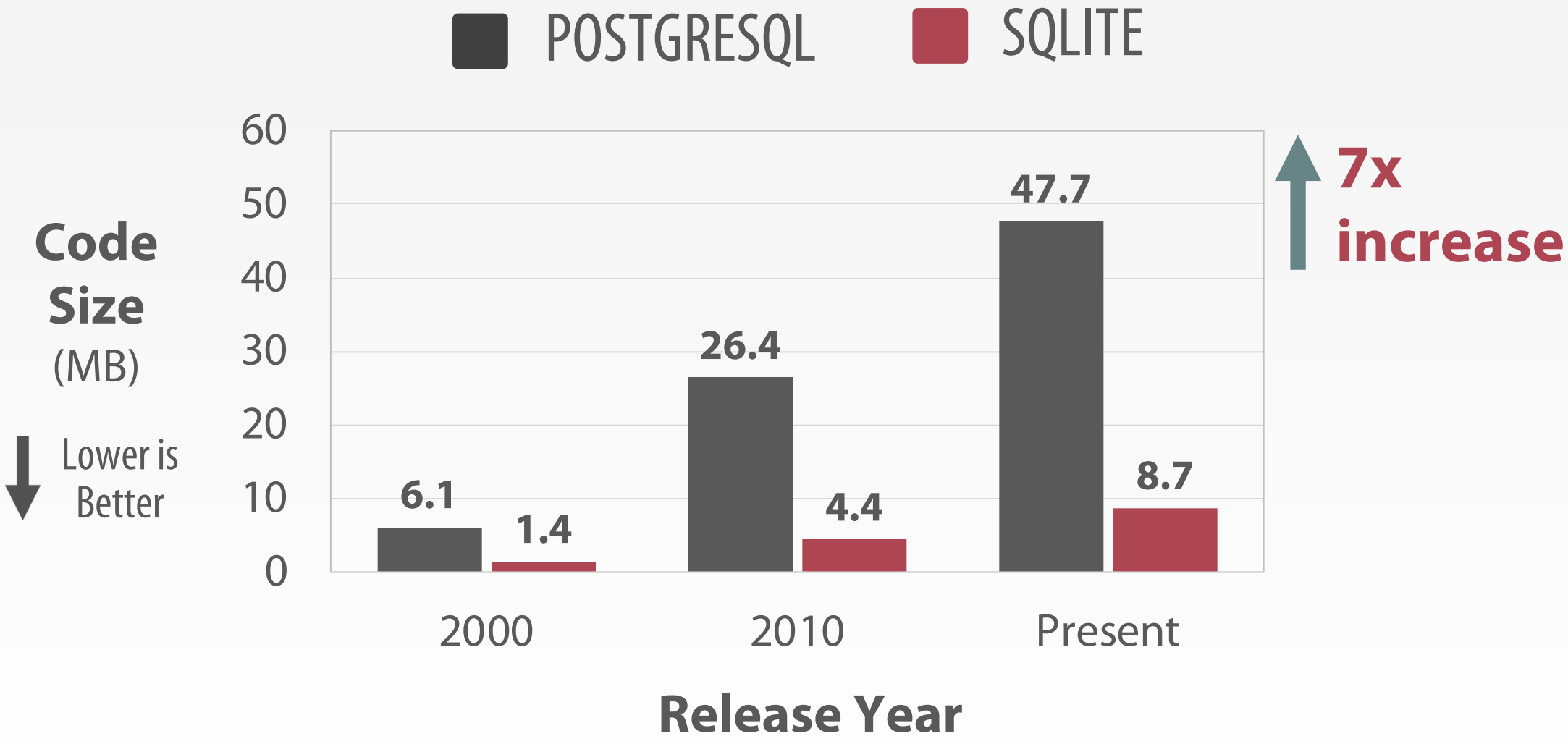
- "A reasonable approach to page coloring"
  - ASPLOS '06
- "Another page coloring idea"
  - OSDI '08
- …
- Enumerating each paper is only a bare minimum.
  - How does the work *relate* to yours? How is yours novel?
- Also be sure to consider papers > 5 years old!
- And include author names!

RELATED WORK (BETTER)



- Spatial display of design space can visually highlight what are your novel claims
- Also can you show an optimality limit and show how different prior papers approached that limit? Where will your work be?

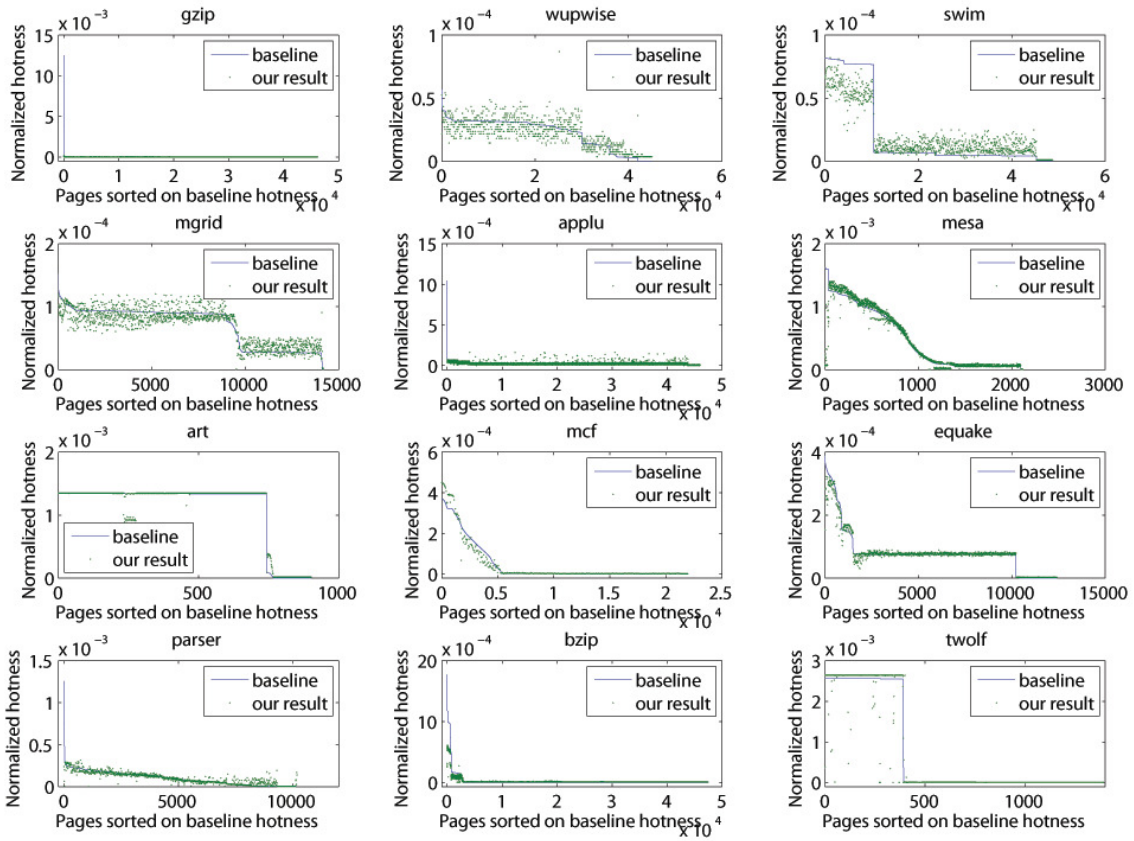
MOTIVATION: DBMS COMPLEXITY



THE MIDDLE OF THE TALK…

- Methods
  - What was most novel or creative about your approach?
  - Flowcharts and diagrams to illustrate key components
- Results
  - Show enough results to get your point across
  - Don't bludgeon the audience with endless unreadable graphs…
  - Select a subset to discuss in detail

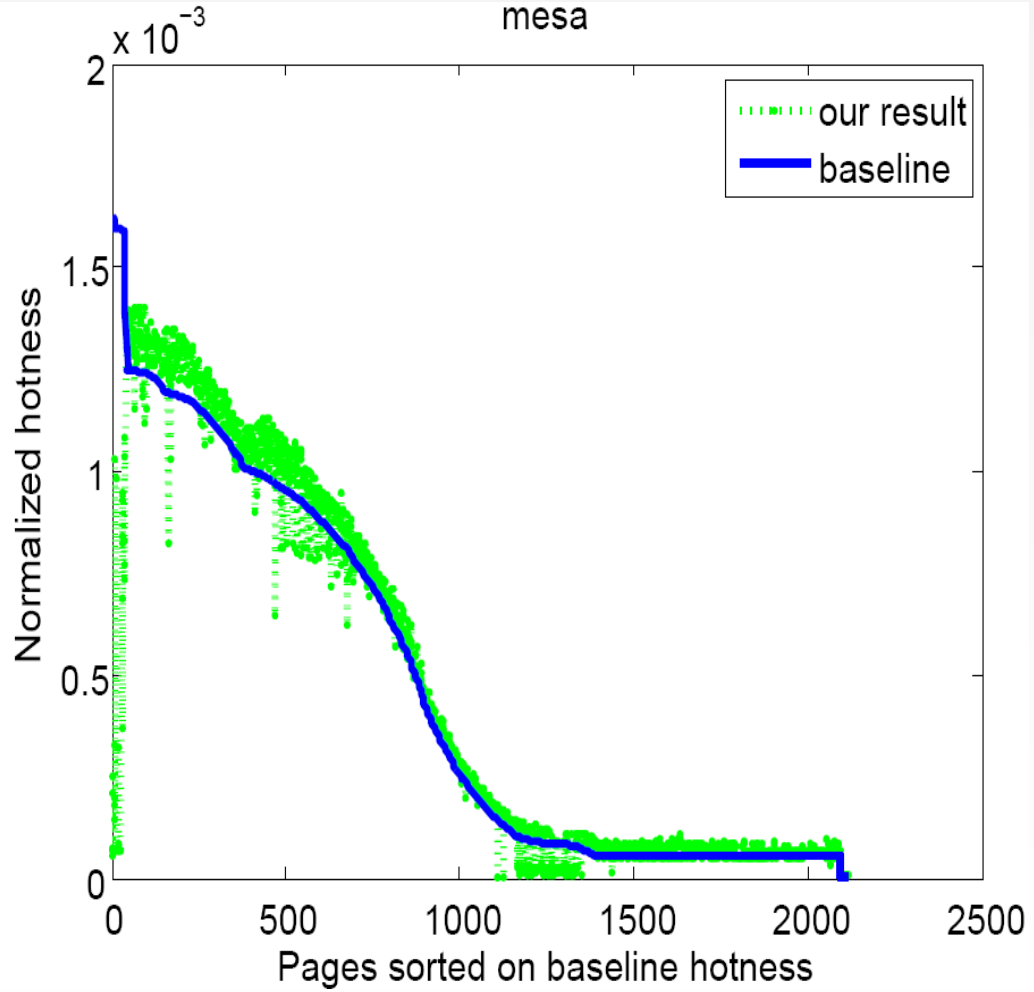
ACCURACY (BAD)



INSTEAD …

HOT PAGE IDENTIFICATION ACCURACY

- No major accuracy loss due to jumping as measured by two metrics (Jeffrey divergence & rank error rate)
- Result is accurate within 10%



EVALUATION

- Tested database systems
  - PostgreSQL, SQLite
- Instrumentation to get control flow graphs
  - DynamoRIO instrumentation tool
- Evaluation
  - Efficacy of SQLFuzz in detecting regressions?
  - Efficacy of SQLMin in reducing queries?
  - Accuracy of SQLDebug in diagnosing regressions?

#1: SQLFUZZ — DETECTING REGRESSIONS

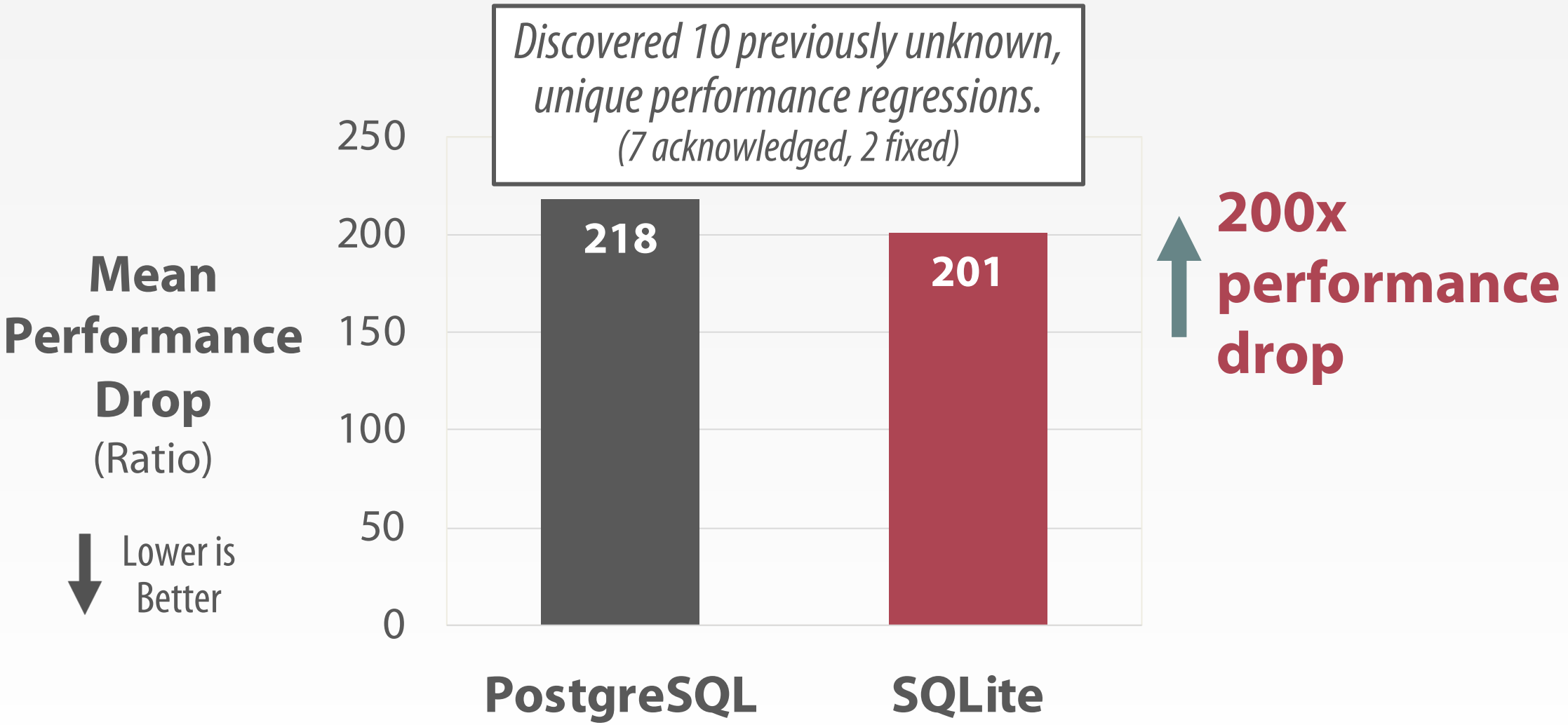


ILLUSTRATION AND COLOR

- "A picture speaks a 1000 words"
  - A 1000 words don't speak, however
  - The picture may need a little help
- Color for emphasis (when appropriate)
  - Not too much…
- Animation when appropriate
  - Not too much!

ILLUSTRATION AND COLOR

- Tip: Record yourself giving a practice talk, and look for places where you are gesturing with your hands to “draw diagrams” in mid-air.
- That’s a good hint you need another figure there!

PAGE RE-COLORING PROCEDURE

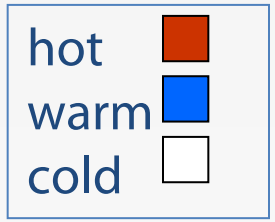
- Quick search for *K*-th hottest page's hotness
  - *Bin[i][j]* indicates # of pages in color *i* with normalized hotness in
  - *[j, j+1]* range

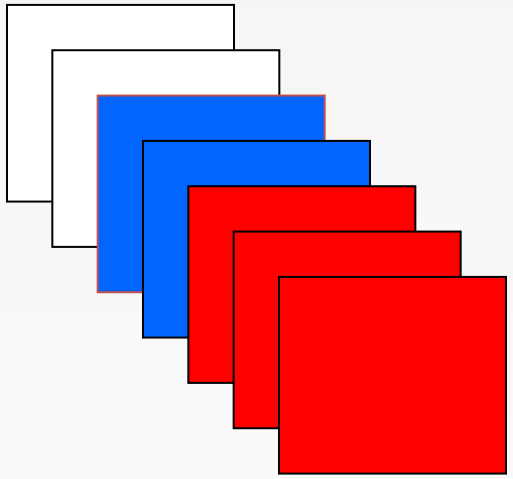
procedure *Recolor*

budget (recoloring budget)
old-colors (thread's color set under old partition)
new-colors (thread's color set under new partition)
if *new-colors* is a subset of *old-colors* **then**
subtract-colors = *old-colors* − *new-colors*.
Find the hot pages in *subtract-colors* within the *budget* limit and reallocate to *new-colors* in a round-robin fashion.
end if
if *old-colors* is a subset of *new-colors* **then**
add-colors = *new-colors* − *old-colors*.
Find the hot pages in *old-colors* within the $\frac{|new\text{-}colors|}{|add\text{-}colors|} * budget$ limit, and then move at most *budget* (*i.e.* $\frac{|add\text{-}colors|}{|new\text{-}colors|}$ proportion) of them to *add-colors*.
end if

INSTEAD …

RE-COLORING PROCEDURE(I)



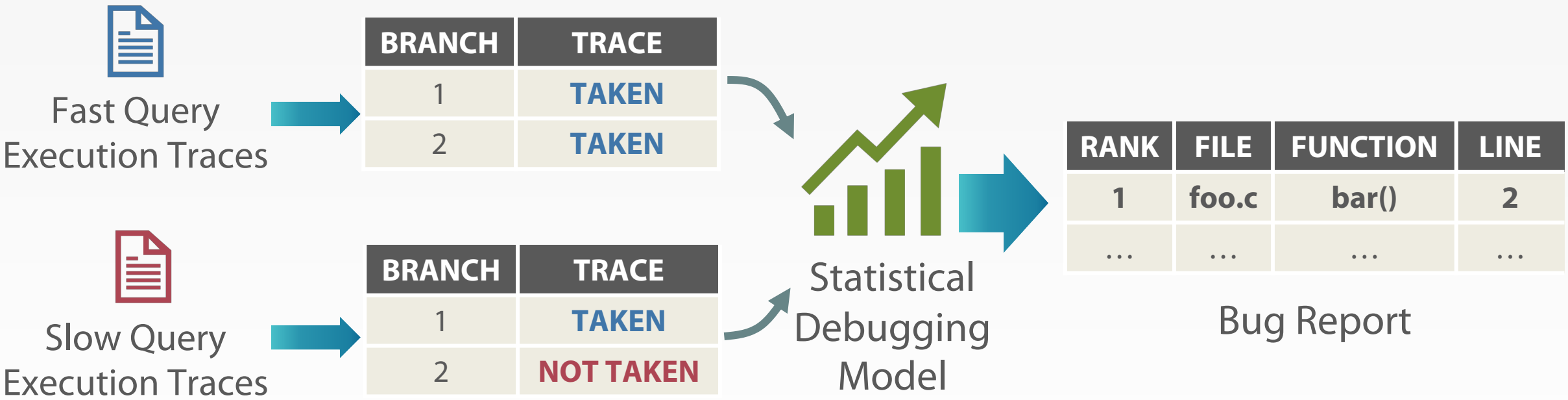


Cache share decrease

Budget = 2 pages

#3: SQLDEBUG — DIAGNOSING REGRESSIONS

4 STATISTICAL DEBUGGING: FAST AND SLOW QUERY TRACES



#2: SQLMIN — REPORTING REGRESSIONS

- Top-Down Query Reduction
  - Iteratively remove unnecessary query elements
- Bottom-Up Query Reduction
  - Extract valid sub-queries

#2: SQLMIN — REPORTING REGRESSIONS

```
SELECT S1.C2
FROM (
  SELECT
    CASE WHEN EXISTS (
       SELECT S0.C0
       FROM ORDER AS R1
       WHERE ((S0.C0 = 10) AND (S0.C1 IS NULL))
    ) THEN S0.C0 END AS C2,
  FROM (
     SELECT R0.I_PRICE AS C0, R0.I_DATA AS C1,
       (SELECT ID FROM ITEM) AS C2
    FROM ITEM AS R0
    WHERE R0.PRICE IS NOT NULL
       OR (R0.PRICE IS NOT S1.C2)
    LIMIT 1000) AS S0) AS S1;
```

#2: SQLMIN — REPORTING REGRESSIONS

```
SELECT S1.C2
FROM (
  SELECT
    CASE WHEN EXISTS (
      SELECT S0.C0
      FROM ORDER AS R1
      WHERE ((S0.C0 = 10) AND (S0.C1 IS NULL))
    ) THEN S0.C0 END AS C2,
  FROM (
    SELECT R0.I_PRICE AS C0, R0.I_DATA AS C1,
      (SELECT ID FROM ITEM) AS C2
    FROM ITEM AS R0
    WHERE R0.PRICE IS NOT NULL
      OR (R0.PRICE IS NOT S1.C2)
    LIMIT 1000) AS S0) AS S1;
```

BOTTOM-UP REDUCTION

EXTRACT SUB-QUERY

Remove dependencies

#2: SQLMIN — REPORTING REGRESSIONS

```
SELECT S1.C2
FROM (
SELECT
  CASE WHEN EXISTS (
     SELECT S0.C0
     FROM ORDER AS R1
     WHERE ((S0.C0 = 10) AND (S0.C1 IS NULL))
  ) THEN S0.C0 END AS C2,
FROM (
   SELECT R0.I_PRICE AS C0, R0.I_DATA AS C1,
     (SELECT ID FROM ITEM) AS C2
  FROM ITEM AS R0
  WHERE R0.PRICE IS NOT NULL
     OR (R0.PRICE IS NOT S1.C2)
  LIMIT 1000) AS S0) AS S1;
```

TOP-DOWN REDUCTION

REMOVE ELEMENTS

Remove conditions

Remove columns
Remove sub-queries

Remove clauses

#2: SQLMIN — REPORTING REGRESSIONS

```
SELECT S1.C2
FROM (
SELECT
  CASE WHEN EXISTS (
     SELECT S0.C0
     FROM ORDER AS R1
     WHERE ((S0.C0 = 10))
  ) THEN S0.C0 END AS C2,
FROM (
   SELECT R0.I_PRICE AS C0,
   FROM ITEM AS R0
   WHERE R0.PRICE IS NOT NULL) AS S0)
AS S1;
```

THE END OF THE TALK…

- Conclusions
  - Don't just repeat what you did.
  - Use this as a chance to broaden your scope.
  - What are the implications of what you did?
  - What did you learn?

THE END OF THE TALK…

- Conclusions as Takeaway Message
  - What are 2-3 things you want the audience to remember?
    - If you give them 6, they remember none.
  - Give them at least one number (“2X improvement”, “30% lower hardware complexity”, …)

CONCLUSION

- Interested in integrating APOLLO with more database systems
  - Improve the toolchain based on developer feedback
- Automation will help reduce labor cost of developing DBMSs
  - Developers get to focus on more important problems

THE END OF THE TALK… PART II

- The Post-Talk Questions
  - A bungled question is unfortunately very memorable…
- Prepare for them! They are part of the talk!
  - Hold practice sessions with a broad audience to get questions from researchers in slightly different areas
  - Have a friend record all questions asked (or video-record) so you can prepare backup slides.

THE POST-TALK QUESTIONS… PART II

- During the Question Session:
  - Repeat/rephrase each question asked
    1) Helps back of room hear what was asked
    2) Ensures that you actually understand the question and are answering what was asked
    3) Gives you time to formulate a good answer
- If they ask "Did you try XYZ…"
  - Not-so-good answer: "No."
  - Better answer "No, but we did try ABC and saw that it only helped by 5% which led us to surmise that XYZ would also perform similarly"

THE POST-TALK QUESTIONS… PART II

- Try to give things a short but complete answer and then move on. Don't ask "Did that answer your question?"
- When in doubt, "That's an interesting question, but perhaps it would be easier to take the answer offline"

PRACTICE, PRACTICE, PRACTICE!

- Build your confidence; get feedback; form a support group; return the favor

MORE HINTS

- Tape yourself and watch the tape
- Enroll in a public speaking class
  - Toast masters, community courses
- Memorize first 5 minutes of your talk
  - Helps start out if you are nervous
- Script the main ideas of the talk so you practice where to say key points.
  - Then throw the script away so your talk will not sound too robotic or pre-planned…

BODY LANGUAGE

- Eye contact, Fillers, Gestures
  - You should not avert eyes to show respect
  - Blocking screen will not add mystery
- Enunciation
- Voice modulation and emphasis

BODY LANGUAGE

- Speed of delivery
  - There's no prize for learning how to fit 20 words in 10 seconds
- Most of all, project your enthusiasm for what you are presenting!

LOGISTICAL DETAILS

- Redundancy/fault tolerance: make copies of your slides on a flash drive
  - Your computer may fail you
- Create versions in multiple formats for just in case
  - E.g., ppt and pdf
- Make sure you check the projection systems prior to your talk or session if at a conference!

LOGISTICAL DETAILS

- Turn off automatic time-based transitions in powerpoint.
- Plug in your laptop to avoid power-save modes or battery problems.
- Use your own laptop if at all possible!

ELEVATOR PITCH

THE ELEVATOR PITCH / HALLWAY CONVERSATION

- Scene 1: You step into an elevator and realize that {Bill Gates, Sergei Brin, …} just walked in. The door closes. You have ~30 seconds to explain to them what you do.
- Scene 2: You are at a conference and you have a chance to discuss your work with one of the research leaders of your field. You have ~30 seconds to start a conversation with them about what you do.
- What do you say?

EXERCISE

- Practice an elevator pitch or 30-second conversation with your table.
- Time it!
- Offer suggestions for improvements.

EXERCISE

- Remember these:
  - What is the problem?
  - Why is it important?
  - What have others done about it?
  - What am I doing about it?
    - That is useful, novel, interesting, different…

OTHER TIPS

GRAPHS & VISUALS

- A subset of your listeners may be color blind
  - Don't make bar charts with equally bright bars of red and green
  - Use stripes or something to distinguish the bars.
- If you put a chart on the screen you have to explain it
- Always label all axes of your graphs

GRAPHS & VISUALS

- Don't ask the audience to compare by memory to a graph from a previous slide.
- If you want them to compare 2 sets of results and see how much you improved things
  - Put them on the same slide so they can see the data side by side

DELIVERY

- The two things most amateurish about Powerpoint presentations:
  - Too much text
  - Inability to skip slides when pressed for time
- Start preparing your talk more than 48 hours in advance

DELIVERY

- Be energetic! A really jazzed presenter can help the audience get excited about the topic, and on the flip-side, if the presenter looks bored, you can guess how 95% of the audience must feel...

DELIVERY

- Record yourself when practicing, or have someone watch you
  - Look for odd body movements (rocking back and forth, waving hands)
  - Using "um" or some other noise to fill gaps
- Give a practice talk to get feedback

ANIMATION

WRITE-AHEAD LOGGING | Linear-Time Recovery

1 ANALYSIS → 2 REDO → 3 UNDO

WRITE-BEHIND LOGGING | Constant-Time Recovery

1 ANALYSIS

AVAILABILITY

VISUAL ELEMENTS

- Fonts
  - Myriad Pro, Bebas Neue etc.
- Color palettes
  - https://colorhunt.co/palettes/popular
- Icons
  - https://www.flaticon.com/packs/database-and-servers
- PPT or Keynote templates

SUMMARY

- Keep your audience and goals in mind.
  - Don't ramble or meander: The destination and route should always be clear.
- Just like playing tennis or piano, giving good presentations is a skill that can be practiced and improved!
  - Practice for your talks. Look for opportunities to give talks. Practice elevator pitches with your friends often.

SUMMARY

- Remember the Big Four Questions:
  - What is the problem?
  - Why is it important?
  - What have others done about it?
  - What have I done?

USEFUL RESOURCES

Oral:

David Patterson: How to Give a Bad Talk
http://pages.cs.wisc.edu/~markhill/conference-talk.html#badtalk

Mark Hill's "Oral Presentation Advice",
http://pages.cs.wisc.edu/~markhill/conference-talk.html

CRA-W, http://www.cra-w.org/gradcohort

http://www.randsinrepose.com/archives/2008/02/03/out_loud.html

http://www.slideshare.net/selias22/taking-your-slide-deck-to-the-next-level

http://www.presentationzen.com/

Written:

Strunk & White "The Elements of Style"

Gopen & Swan "The Science of Scientific Writing"
http://www.americanscientist.org/issues/feature/the-science-of-scientific-writing/9

Many schools provide many writing resources: Use them!
→ Writing center or tutor.

Also, it may be worthwhile to *pay* a writing tutor to help teach you and edit your work, in order to make your overall idea-to-paper process easier!