Smooth Scan: Statistics-oblivious Access Paths

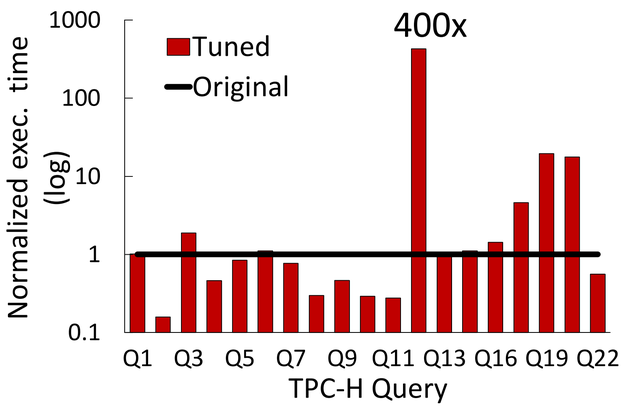
Renata Borovica-Gajic, Stratos Idreos, Anastasia Ailamaki, Marcin Zukowski and Campbell Fraser

renata.borovica@epfl.ch, stratos@seas.harvard.edu, anastasia.ailamaki@epfl.ch, marcin@snowflake.net, campbellf@google.com

The Need for Intra-Query Adaptivity

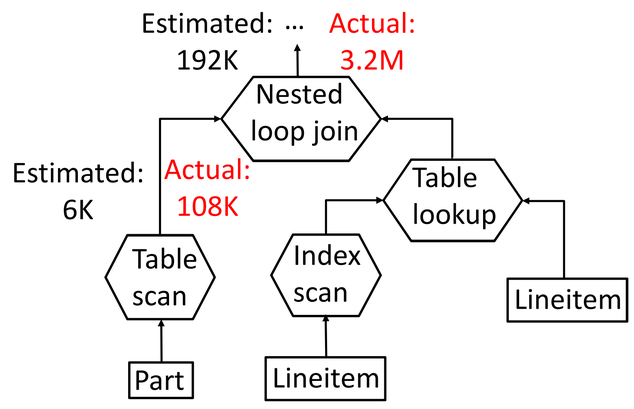
State of affairs in database systems

Setting: TPC-H, SF10, DBMS-X, Tuning tool 5GB space



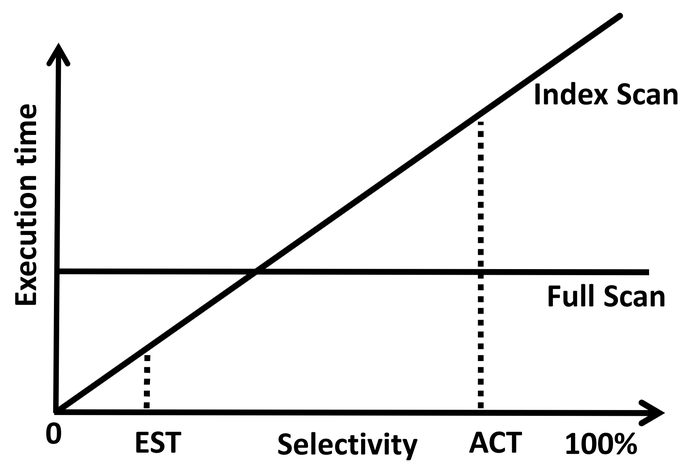
➢ **Degradation due to sub-optimal access path choices**

Cause for sub-optimal plans



➢ **Cardinality misestimates**

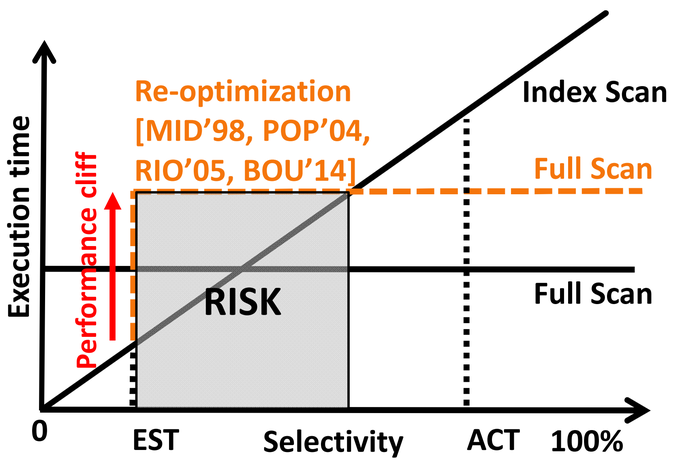
Access path selection problem



➢ **Statistics: unreliable advisor**

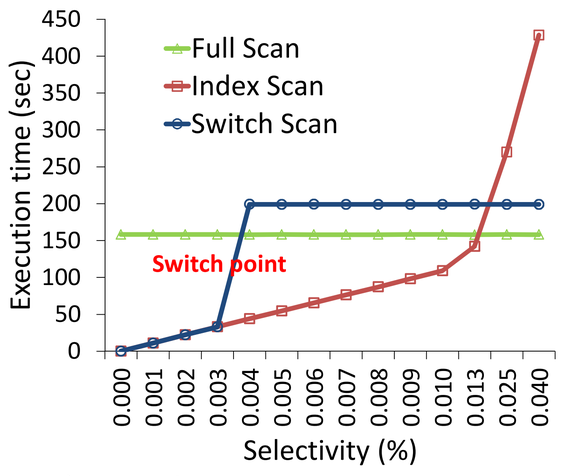
Adaptivity in Access Path Operators

Mid-query re-optimization



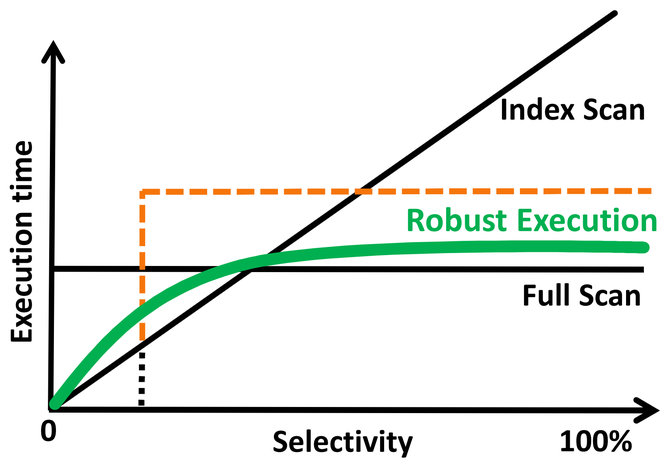
➢ **Re-optimization risky**

Re-optimization in action



➢ **Violate user expectation**

Quest for robust execution



➢ **Near-optimal for all inputs**

Smooth Scan in a Nutshell

Statistics-oblivious access paths

SMOOTH SCAN

No need for accurate statistics

Learn result distribution at run-time

Adapt as you go

DESIGN GOALS

Avoid **performance cliffs & risk**

Continuous, gradual and **smooth adaptation**

➢ **Adaptive, but smooth**

Adaptivity with Smooth Scan

INSIGHT: **Morph** between Index and Sequential Scan

HOW?

1. Index Access
2. Entire Page Probe
3. Gradual Flattening Access

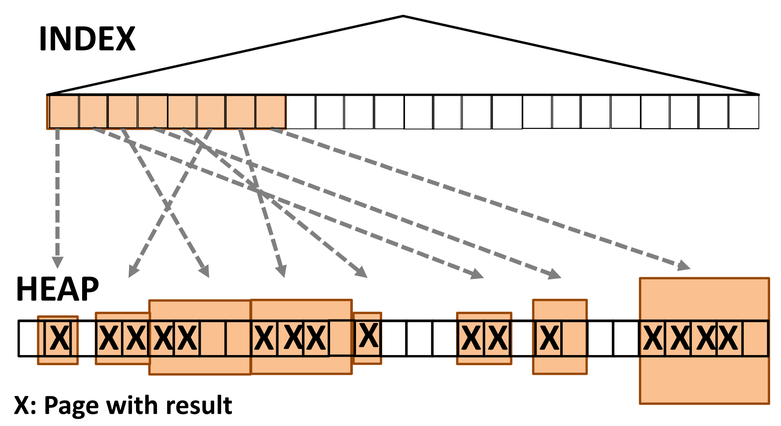
WHEN?

Selectivity increase -> Mode Increase

Selectivity decrease -> Mode Decrease

➢ **Data driven adaptation**

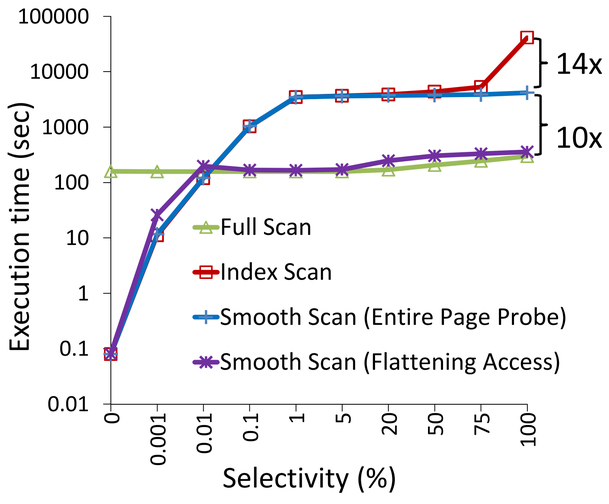
Region Snooping = Data driven adaptation



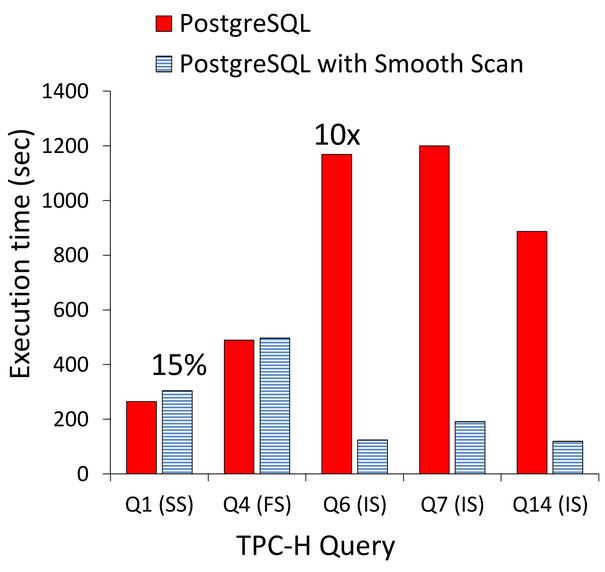
➢ **Avoid repeated access** ➢ **Less random I/O**

Smooth Scan in Action

Setting: 400M tuples, 25GB, Index(c2)
Query: select * from R where c2 < X%;



Setting: TPC-H, SF 10, PostgreSQL with Smooth Scan



Summary

Operator morphing from one form to another

+

Data driven **adaptation**

=

Robust query execution

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

snowflake

HARVARD School of Engineering and Applied Sciences

Google

AiAS DATA-INTENSIVE APPLICATIONS AND SYSTEMS