Testing the Accuracy of Query Optimizers

Zhongxian Gu	UC Davis
Mohamed A. Soliman	Greenplum/EMC
Florian M. Waas	Greenplum/EMC

Evaluating Query Optimizer

Evaluating Query Optimizer

How can we measure the accuracy of a commercial query optimizer?

Quality Assurance Challenge

- Functional testing
  - Correctness tests
  - Testing for crashes/asserts, internal errors
- Performance/Optimality
  - Typically eye-balled by expert for ‘suspicious’ plans
  - Tested with trivial examples where optimal plan is known

Optimizers are hard to test;
generally under-tested;

Example

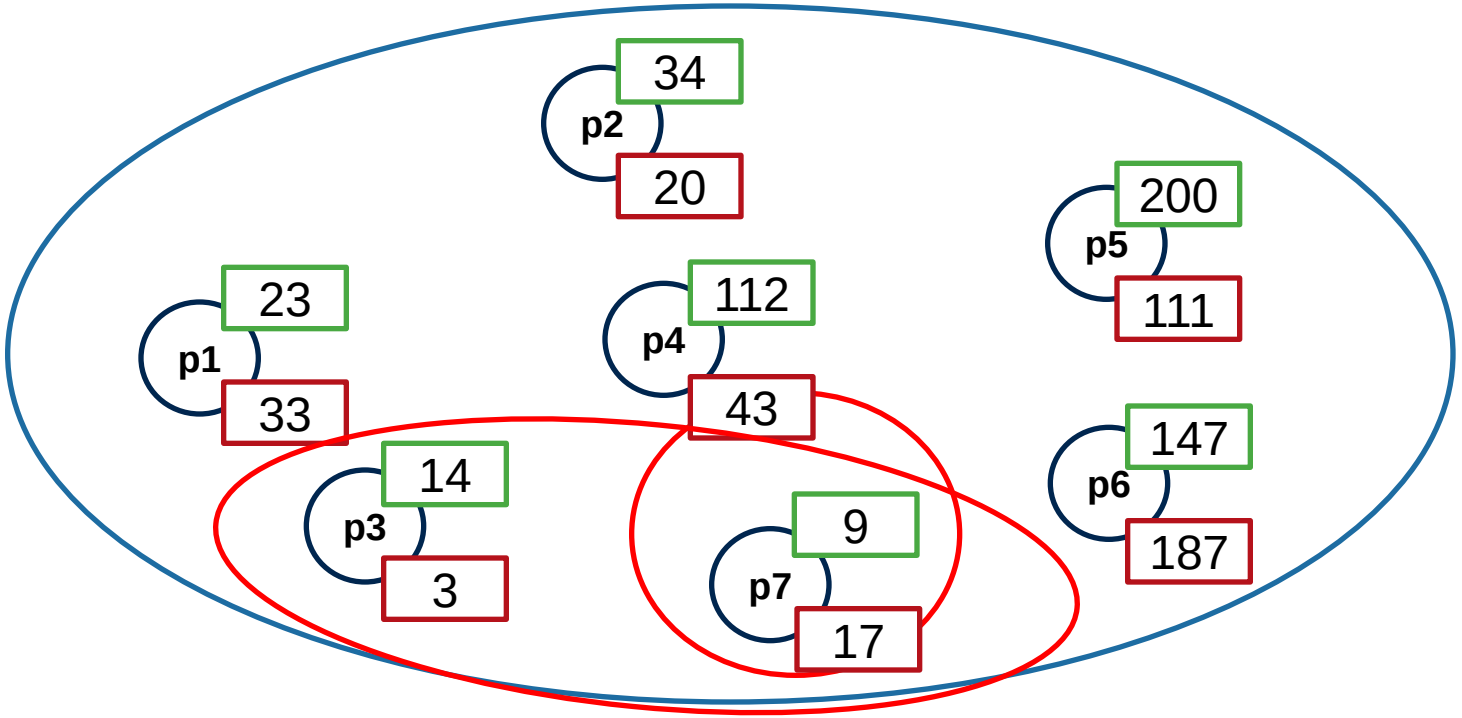
- Commercial DB Vendor
- Regular regressions
  - 400,000 test cases from real-world app stack
  - Used for both functional and performance testing
- Major optimizer extension/rework done
- Same regression test used
  - Ensures same functionality
  - Used to demonstrate significant performance improvements

Tests did not reveal sub-optimality for years!

Common Approaches

Approach	Pros	Cons
Eye-balling	• Intuitive • Suitable for troubleshooting	• Manual • Cannot be automated • Not exhaustive
“Freezing” of plans	• Finds all possible regressions	• Requires manual verification of changes • Not scalable • Hinders innovation in the long-term
Large test suites	• May represent customer workloads • Large coverage	• Merely cement current behavior • Do not test for optimality
Randomize tests	• Test corner cases	• Test “plumbing” only • Do not test for optimality • Do not test for correctness

Measuring Optimizer Accuracy

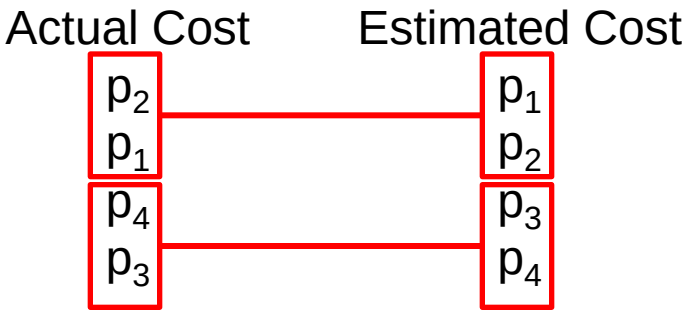


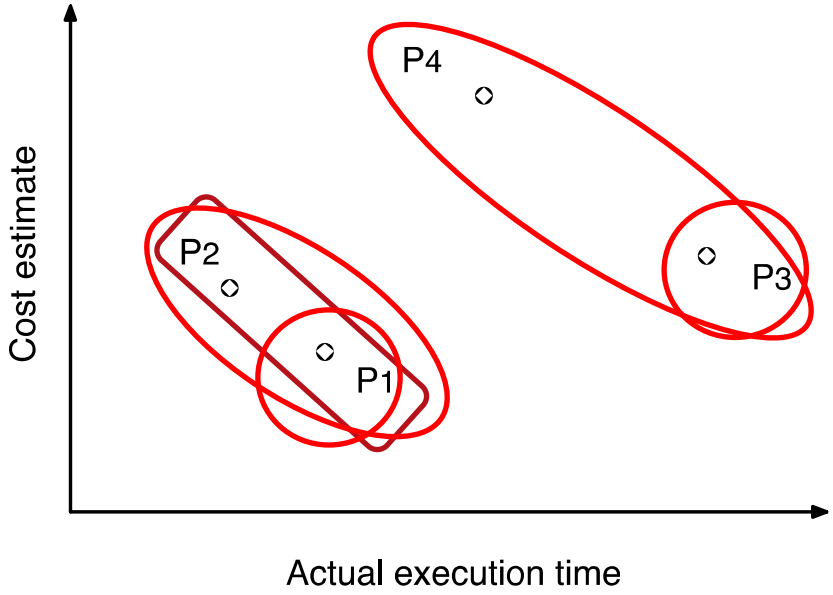
Plan search space

Given two plan alternatives, can the optimizer distinguish them based on cost?

Measuring Optimizer Accuracy

- Compute a correlation score between the plan ranking based on estimated cost and that of actual cost
  - Plan discordance
  - Plan relevance
  - Pairwise distance



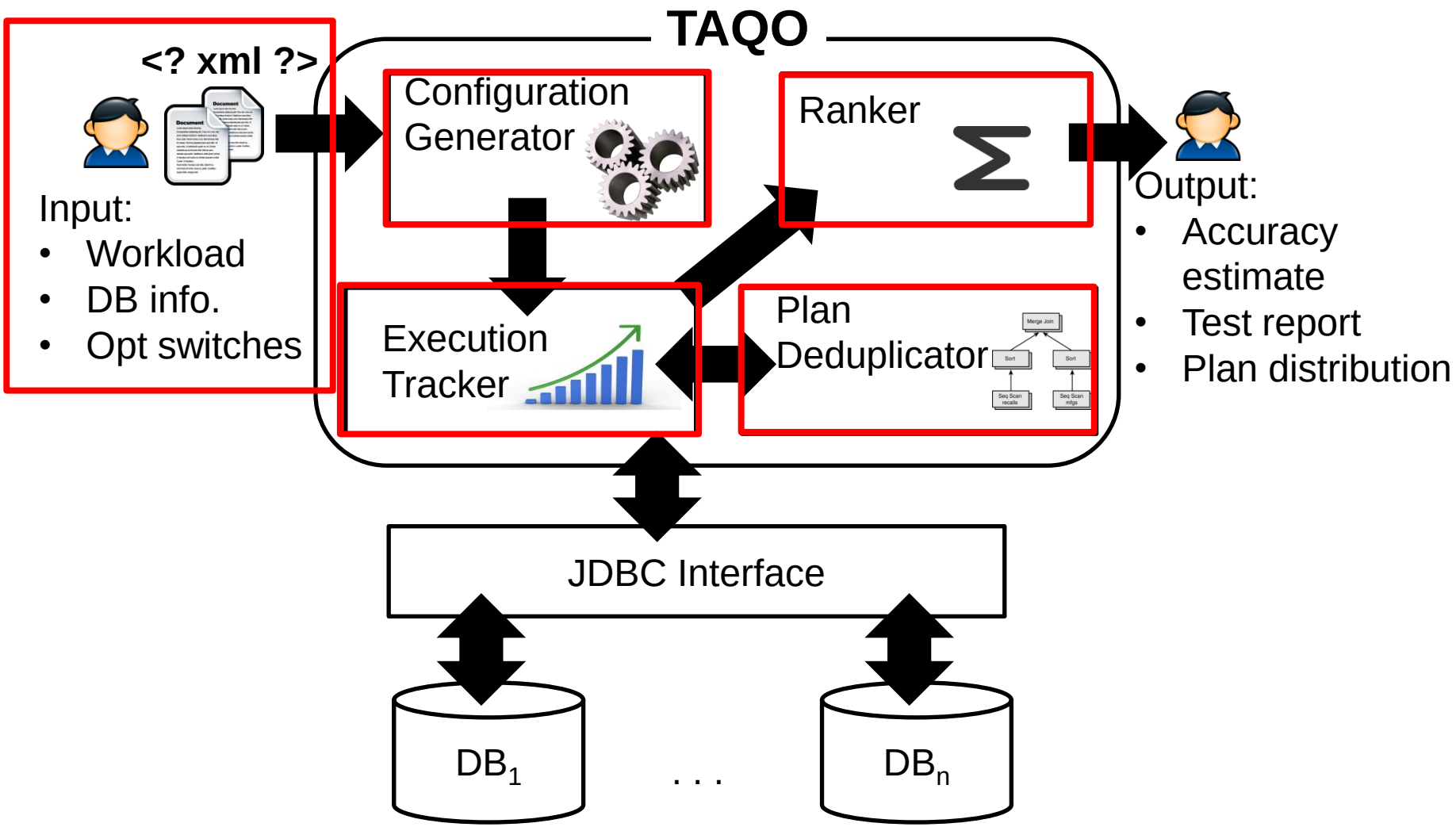


Generating Plan Sample

- Commercial optimizers provide switches to influence plan choice for a given query
  - Example: enable/disable Hash Join alternative
  - Usually achieved by issuing simple commands, or by incorporating special constructs into query text

- The existence of optimizer switches allows forcing a number of plan alternatives

TAQO



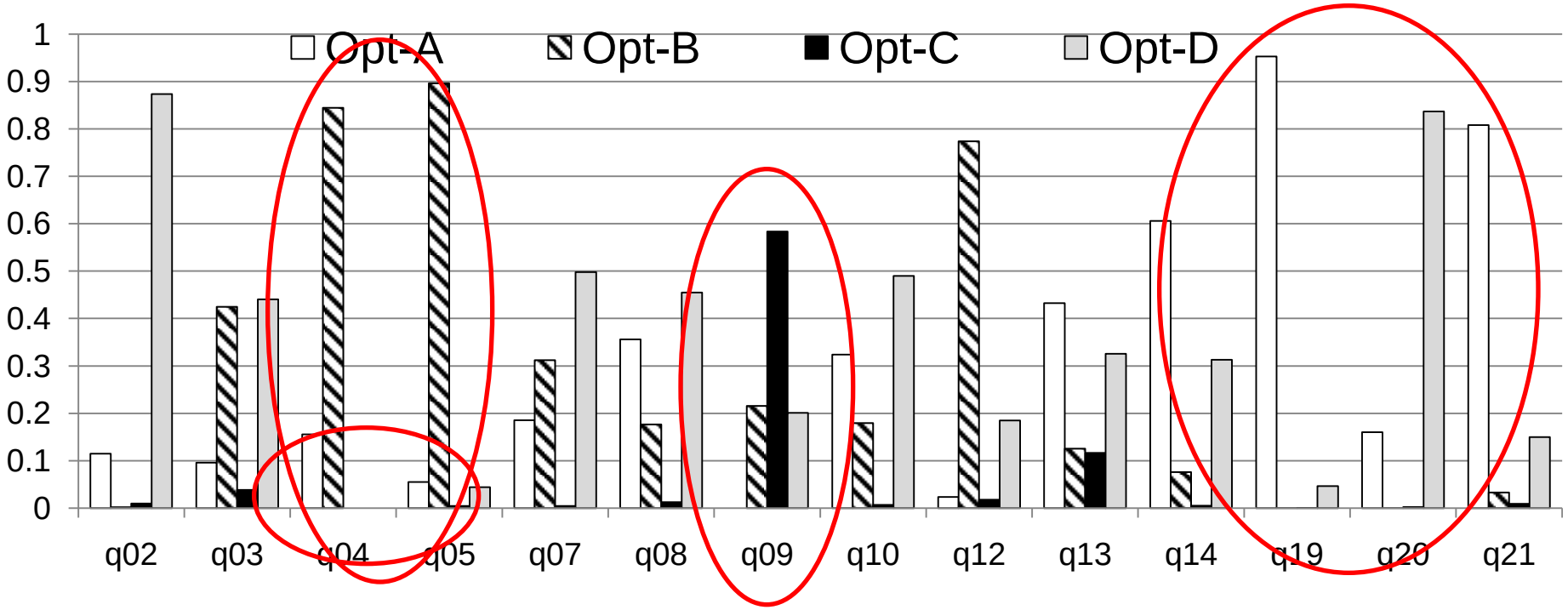
Evaluation

- 4 commercial optimizers
- TPC-H
- Compare results
  - Accuracy: Ranking of estimated vs. actual
  - Optimality: default plan is optimal
- The tests were automatically conducted using TAQO

Plan Samples

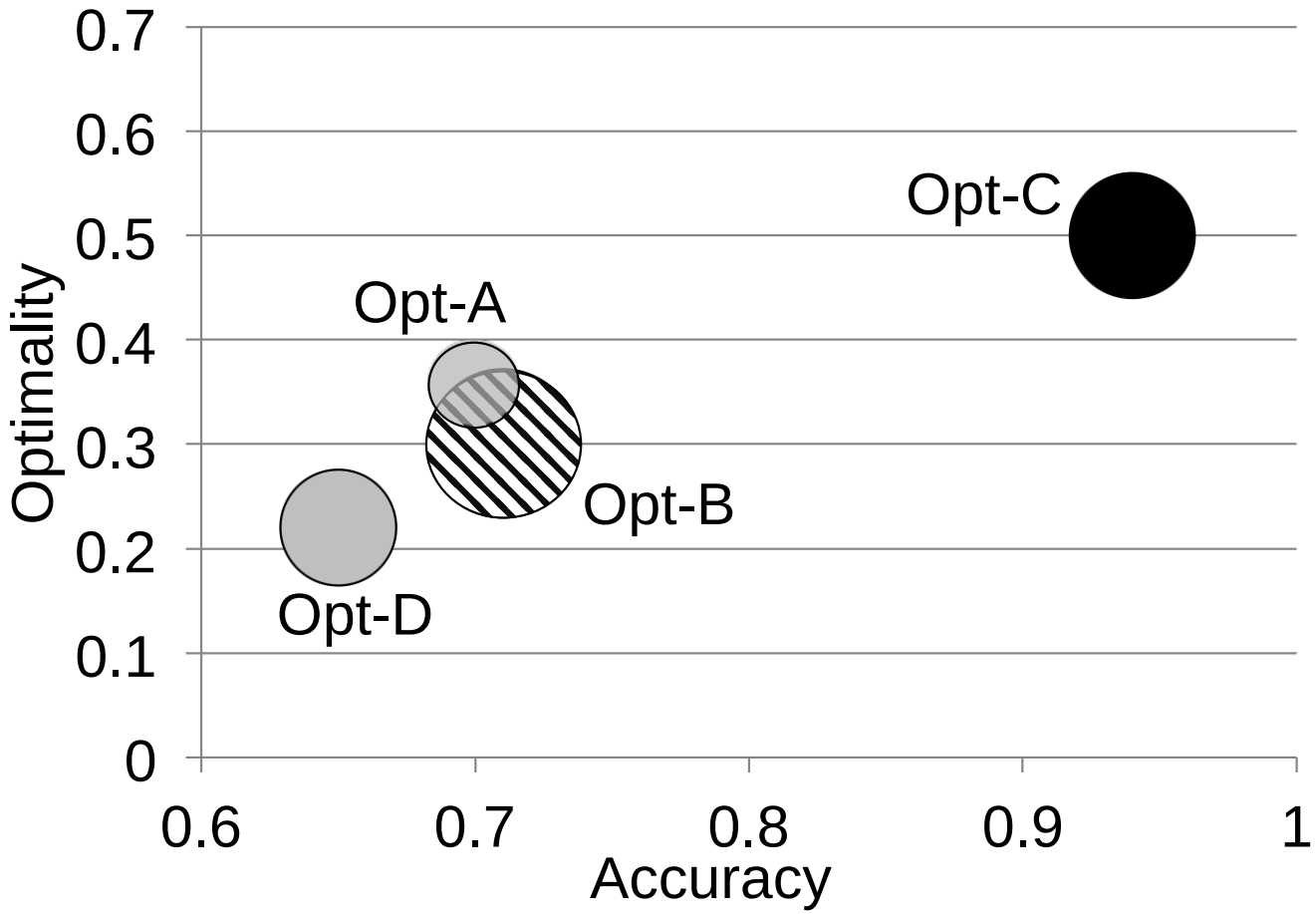
Query	Opt-A		Opt-B		Opt-C		Opt-D	
	OK	T/O	OK	T/O	OK	T/O	OK	T/O
Q1	1	5	1	1	31	2	6	10
Q2	73	6	161	40	236	120	87	14
Q3	24	8	19	37	27	67	49	15
Q4	26	2	2	2	24	64	49	15
Q5	34	8	22	133	32	57	140	36
Q6	3	0	1	1	16	0	4	0
Q7	34	6	45	105	36	57	126	26
Q8	20	18	25	163	18	44	65	32
Q9	4	46	14	144	3	50	96	23
Q10	34	6	65	98	29	64	56	9
Q11	36	2	173	23	1	0	33	12
Q12	15	3	8	13	60	31	80	24
Q13	19	13	14	2	28	6	60	24
Q14	13	1	11	4	29	14	27	4
Q15	6	0	2	0	1	0	69	8
Q16	54	6	62	9	1	0	24	23
Q17	6	18	10	14	1	0	24	23
Q18	31	19	28	84	1	0	53	13
Q19	12	2	11	13	86	1	26	6
Q20	83	5	61	87	178	58	94	49
Q21	18	38	75	59	98	70	98	72
Q22	12	8	28	4	1	0	37	12

Optimizers Comparison



Correlation scores for different TPC-H queries. Low score indicates high accuracy

Accuracy vs. Optimality



Summary

- TAQO: a new framework for testing the accuracy of query optimizers
- Leveraging rank correlation metrics for evaluating optimizer's accuracy
- Evaluation of major commercial query optimizers suggest that the abilities of different optimizers differ widely
- TAQO is part of Greenplum's development and test cycles

GREENPLUM

A DIVISION OF EMC

Rank Correlation Score

- Given a sample of plans, compute a rank correlation score based on the rank of estimate costs and actual costs

- Encompass quantitative value
  - Discordance of plan pairs

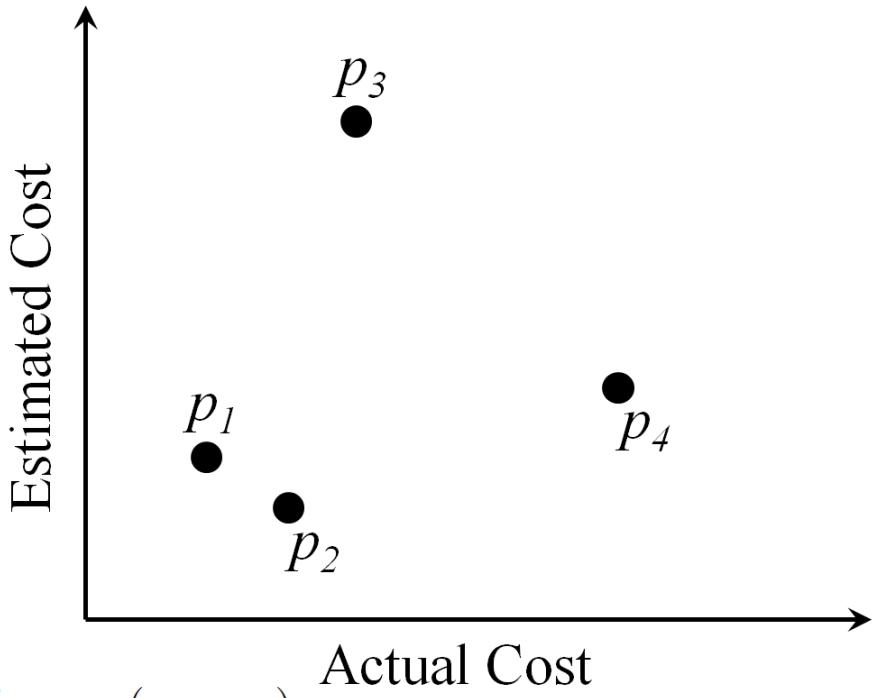
$$\tau = \sum_{i<j} \text{sgn}(e_j - e_i)$$

  - Relevance of plan

$$w_m = \frac{a_1}{a_m}$$

  - Pairwise of distance

$$d_{ij} = \sqrt{\left(\frac{a_j - a_i}{a_n - a_1}\right)^2 + \left(\frac{e_j - e_i}{\max_k(e_k) - \min_k(e_k)}\right)^2}$$



- Rank correlation score: $s = \sum_{i<j} w_i w_j d_{ij} \cdot \text{sgn}(e_j - e_i)$

Normalization Across Systems

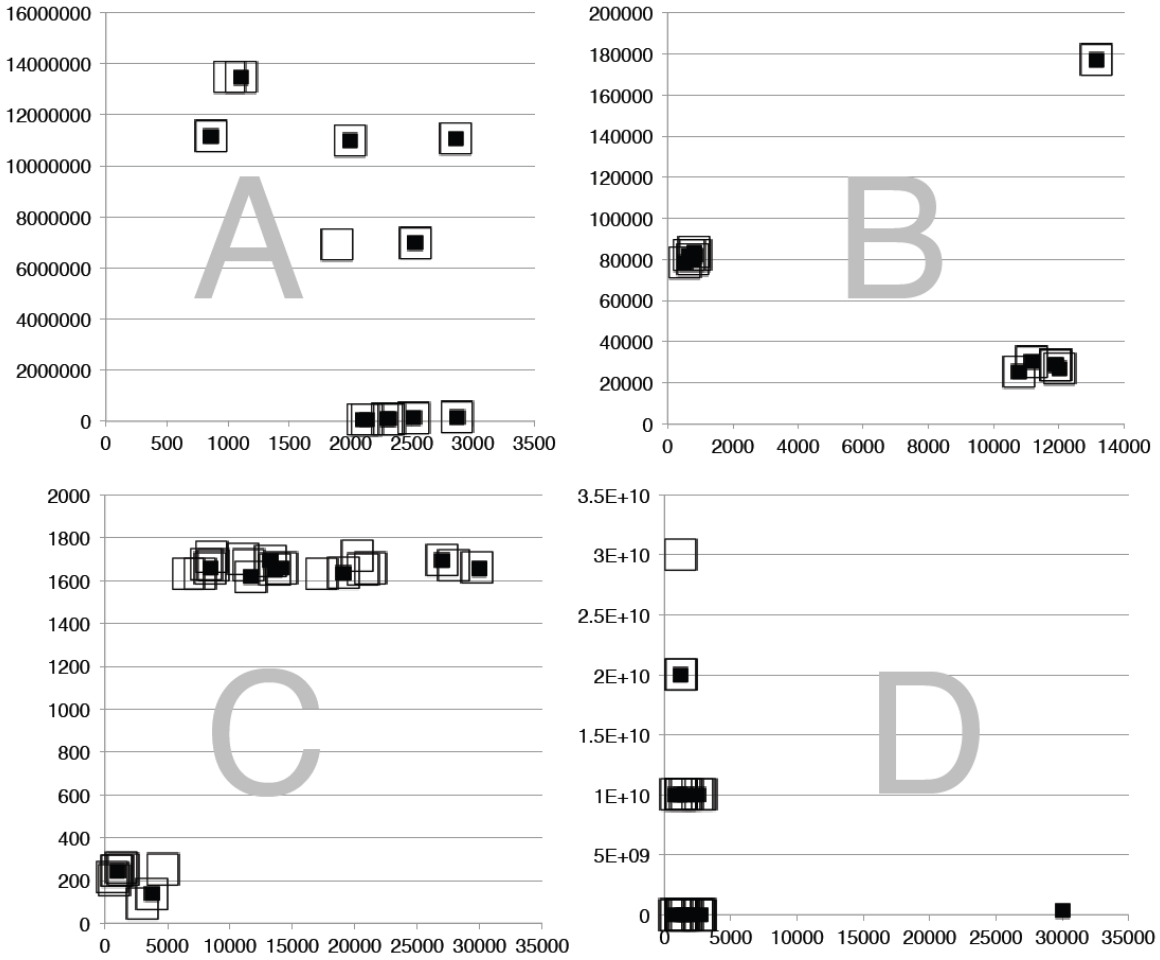
- Consider the actual number of plans

- Apply k-medoids algorithm to find the k representatives

$$k = \min_i(|S_i|)$$

  - Robust to noise and outliers

In-depth Analysis



Plan distribution plots for Query-14

TAQO- Testing Accuracy of Query Optimizers

- How can we measure the accuracy of an optimizer?
- How can we compare among different optimizers?
- TAQO provides an automatic and general approach to accomplish the job
- Evaluation on commercial database optimizers