Perfopticon: Visual Query Analysis for Distributed Databases

Dominik Moritz, Daniel Halperin, Bill Howe, and Jeffrey Heer
Computer Science & Engineering, University of Washington

CPSC 547
Thursday, November 12
By: Dmitry Tebaykin

Overview

1. Introduction into SQL and databases
2. Why is this paper important?
3. The 4 views of Perfopticon (with analysis and pictures)
4. Could you use Perfopticon?
5. Conclusions

1. Introduction into SQL and databases

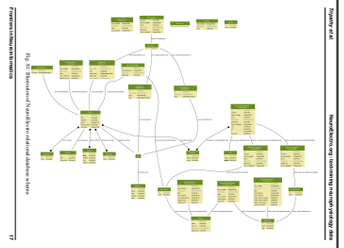
In our case:

Database - tables of data joined

SQL - language for talking to databases

Examples of questions:

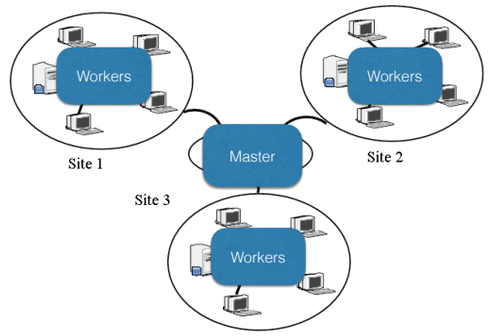
- "What is the age of every student in UBC?"
- "How many people are taking CS547 this term?"



http://in.physiology.org/content/113/10/3474

1. Introduction into SQL and databases

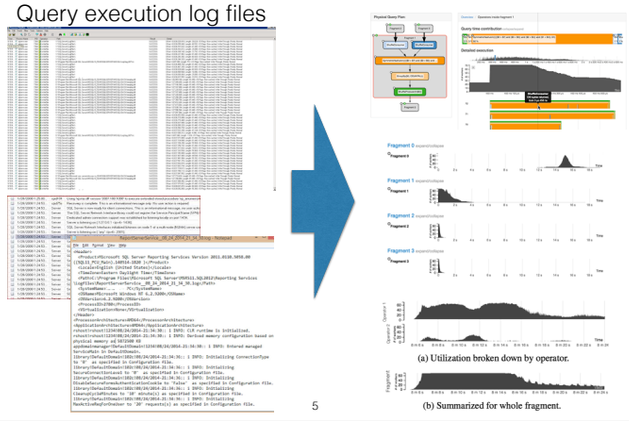
Distributed database system:



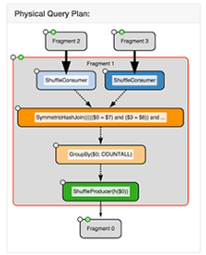
https://cnx.org/resources/0d203a416b87d2bed544825664c14614602f9385/graphics8.png

2. Why is this paper important?

Query execution log files

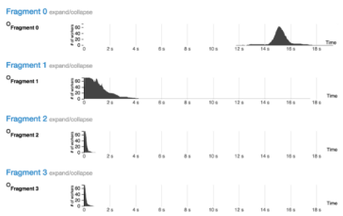


3. The 4 views of Perfopticon (with analysis and pictures)



View 1	Query plan view
What: data	Directed graph that represents: query plan for data access generated by DBMS
Why: tasks	Locate, identify, compare
How: encode	Shape marks for nodes (execution steps), connection marks for links
How: facet	Coordinate: linked highlighting and navigation with other views

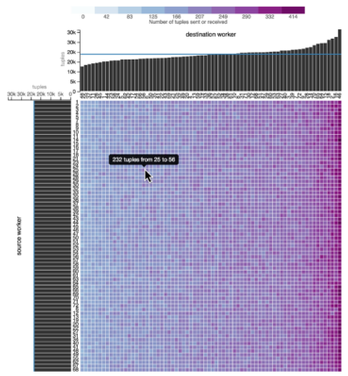
3. The 4 views of Perfopticon (with analysis and pictures)



View 2	Work distribution view
What: data	Tables from query log files
Why: tasks	Compare, identify outliers
How: encode	Histograms showing execution time of workers
How: facet	Partition: multiple views for each query fragment. Coordinate: linked highlighting and navigation with other views
How: reduce	Navigate

3. The 4 views of Perfopticon (with analysis and pictures)

View 3	Communication view
What: data	Table: two continuous variables (amount of data sent and received by workers)
Why: tasks	Compare, identify outliers, summarize
How: encode	2D matrix alignment of area marks, diverging colormap
How: facet	Coordinate: linked navigation with other views



3. The 4 views of Perfopticon (with analysis and pictures)

View 4	Local execution view
What: data	Tables from query log files
Why: tasks	Compare, identify outliers
How: encode	Histograms, bar charts (colour indicates active/ inactive/wait states)
How: facet	Partition: multiform views. Coordinate: linked highlighting
How: reduce	Navigation

4. Could you use Perfopticon?

- Built into Myria (Giant online database), requires log files for the query executions with slight modifications.
- Their example: Myria, added 3 lines to log file per query execution step.
- The tool has a front-end component, upload your query log files and view the results.

5. Conclusions

- Perfopticon can be used effectively for query and database optimization (Emma, the oceanographer, managed to speed up her query and Chu S. et. al created a better table joining algorithm).
- Provides the ability to spot underperforming or overtasked nodes and drill down into the problem.
- Might work for non-relational databases as well.
- Needs more validation.