

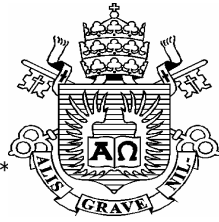


# Autonomic Index Management

Prof. Dr. Sérgio Lifschitz  
PUC-Rio

Marcos Antonio Vaz Salles\*  
ETH Zürich

[sergio@inf.puc-rio.br](mailto:sergio@inf.puc-rio.br)  
[marcos.vazsalles@inf.ethz.ch](mailto:marcos.vazsalles@inf.ethz.ch)



\*Work done while author was affiliated with PUC-Rio

1



# Motivation

```
(1) insert into sales (prodNum, s_date, qty, price)
values (4, current_timestamp, 20, 348);

(2) select prodNum, s_date, sum(price) as total
from sales
where price > 1500000 and
s_date between '20040101' and '20040131'
group by prodNum, s_date;
```

- What are the best indices for an application that uses these commands?

2



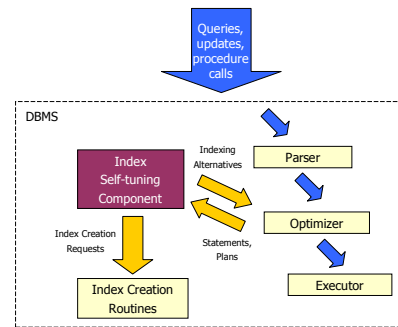
# Problem Statement

- Our work
  - Decide automatically what candidate indices to create
  - Tune indices during system operation
- Our assumptions
  - Candidate index selection work is reused
  - Objective is to maximize average throughput
  - Workload does not have unstable fluctuations

3



# Built-in Architecture for the Self-tuning Component

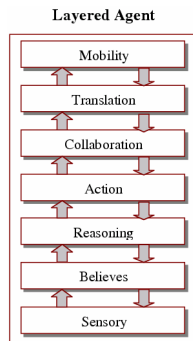


4



# Implementation Architecture

- Agent Architecture (Kendall et al 1999)
- Layered: ease of construction
- Well delimited points of interaction with DBMS
- C++ implementation and C interface to PostgreSQL DBMS



5



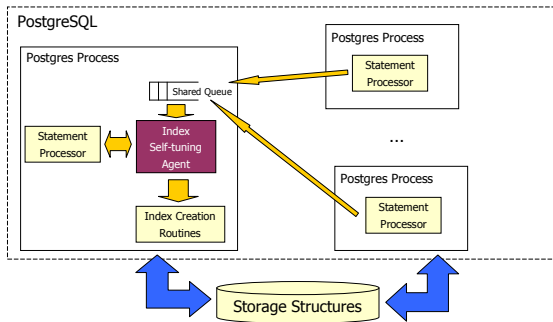
# Integration of Agent and DBMS

- PostgreSQL: one server process for each client
- Problems:
  - Capture statements processed
  - Interact with DBMS routines for index creation
- Our approach:
  - Shared statement queue
  - Indexing process registers as a regular server process

6



## Integration of Agent and DBMS



7



## Experimental Results

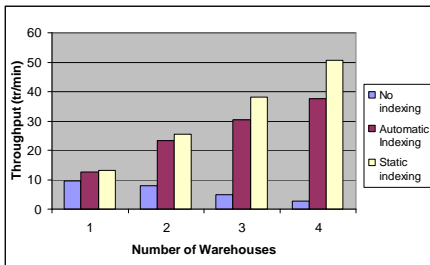
- Tests with OSDL's DBT-2 toolkit
- Workload that uses TPC-C transactions and schema
- Test scenarios:
  - No Indexing: database without indices and agent turned off
  - Automatic Indexing: database without indices and agent turned on
  - Static Indexing: database with toolkit designers' indices and agent turned off

8



## Throughput Evaluation

- # of Warehouses is a scaling factor for the database
- 90 min experiment: past the learning period



9



## Results Analysis

- The component could obtain good average throughput for the application evaluated
- The component goes through learning and steady-state phases
- The longer the component runs with a steady workload, the better the observed average throughput

10



## Publications and Prototype

- Publications so far related to this research project:
  - PhD and MSc Workshop, Brazilian Symposium on Databases (SBBD 2003)
  - CLEI Electronic Journal 2003
  - Tech report on database self-tuning (survey), PUC-Rio, 2004
  - Brazilian Symposium on Databases (SBBD 2004): demo, 2<sup>nd</sup> place prize out of 10 demos
  - International Workshop on Self-Managing Database Systems (SMDB, held in conjunction with ICDE 2005)
  - International Conference on Autonomic Computing (ICAC 2005): poster
- System Prototype
  - <http://www.inf.puc-rio.br/~postgresql>

11



## Remaining Challenges

- Deal with queries that have QoS requirements
- Deal with load surges and workload characterization
- Possible next steps
  - Evaluation against more complex workloads
  - Investigation of new heuristics for index materialization

12