

Advanced Databases / Database Tuning F2010

Introduction

Today

- Questionnaire
- Trends
- What you will learn and How
- Some logistics

Questionnaire

<https://spreadsheets.google.com/viewform?hl=en&formkey=dGhZaFI6Y2lzaVRwVU1pRTNhZVV1MIE6MA>

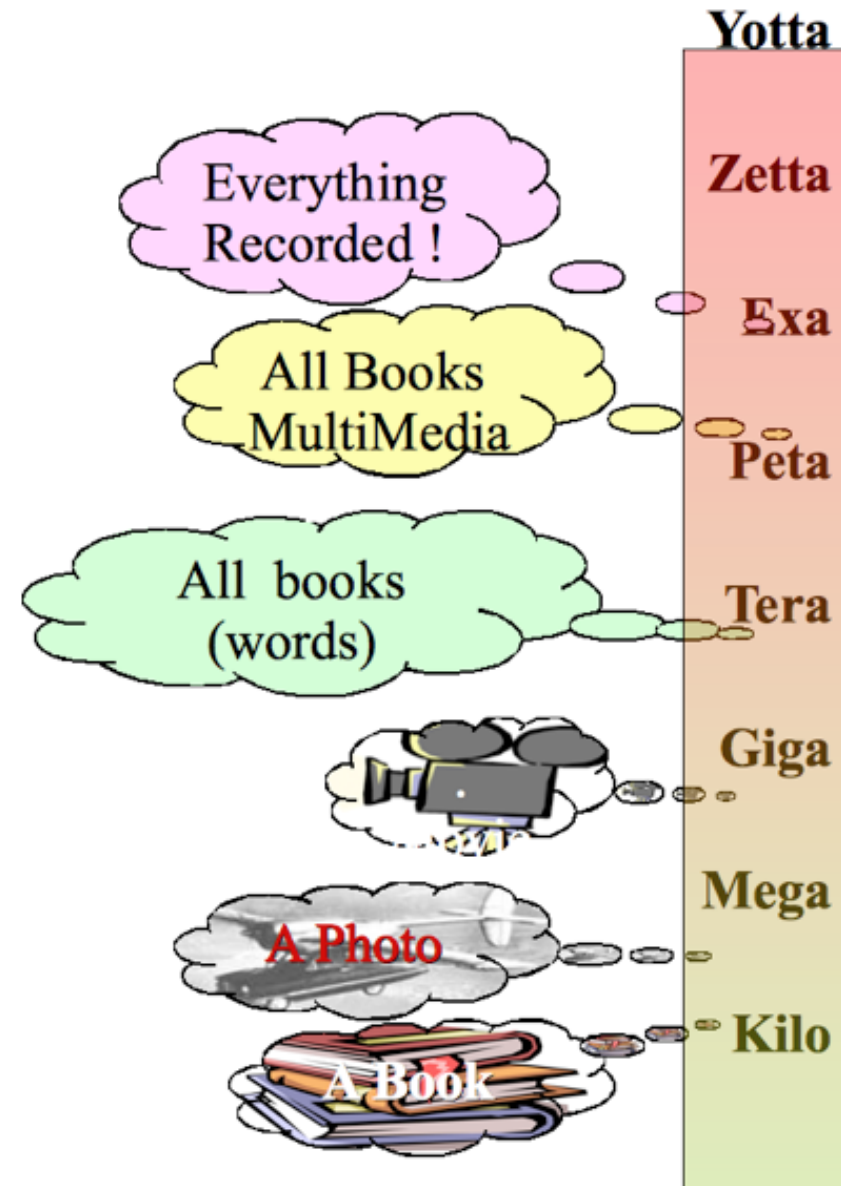
Today

- Questionnaire
- Trends
- What you will learn and How
- Some logistics

Big Data

How much data is there?

- 2007 ~ 281 EB (exabytes)
10E18 bytes
- 2023 ~ Avogadro's number
6x10E23 bytes
~ 0,1 yottabyte



An Exponential World

Over 10 years:

- 10x better access time
- 10x more bandwidth
- 100x more capacity
- 4000x lower media price
- Scan takes 10x longer
- Data on disk is accessed 25x less often (on average)

DBMS Timeline

File-Based Applications

- Ad-hoc
- Efficient
- Un-manageable

Hierarchical / Network Systems

- Imperative
- Navigational

Relational Systems

- Data Independence
- Transaction support
- New features
- System R, Oracle, DB2, SqlServer, Postgres, MySQL, ...

Analytics
(Map-reduce)

OLTP
(main-memory store)

OLAP
(Column stores)

Today

- Questionnaire
- Trends
- What you will learn and How
- Some logistics

Computer Systems [Saltzer&Kaaschak09]

A **System** is a set of interconnected components that has an expected behaviour observed at the interface with its environment.

Coping with complexity:

Modularity, Abstractions, Layering, Hierarchy, Iteration.

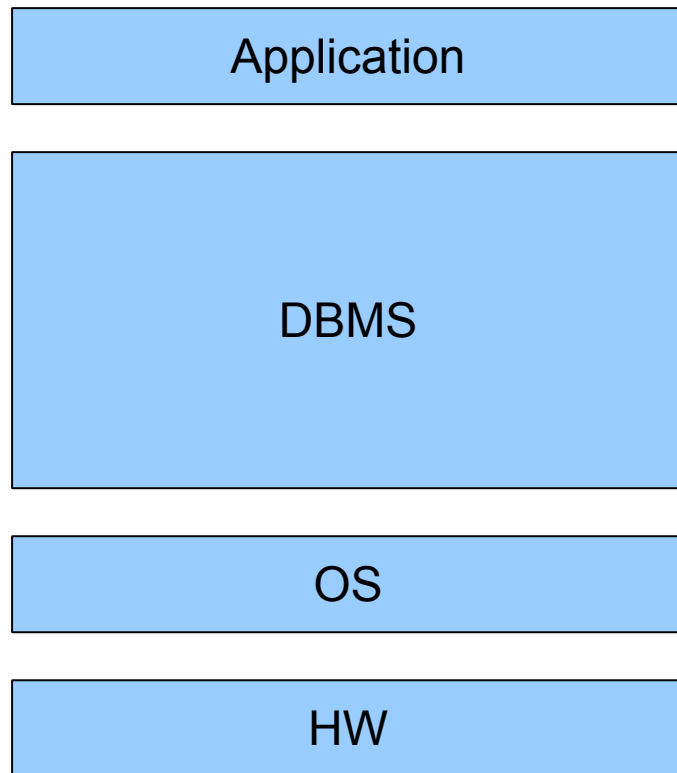
Three Fundamental Abstractions

Storage: read (name) -> value ; write (name, value)

Interpreter

Communication

Database Tuning



Activity of making a database application run faster:

- Faster means higher throughput (or response time)
- Avoiding transactions that create bottlenecks or avoiding queries that run for hours unnecessarily is a must

Why Database Tuning?

- Troubleshooting:
 - Make managers and users happy given an application and a DBMS
- Capacity Sizing:
 - Buy the right DBMS given application requirements
- Application Programming:
 - Coding your application for performance

What is great about DB2 v9.7

DB2 9.7 Feature Overview

- **Application Freedom**
 - SQL Procedure Language (SQL PL) enhancements
 - Oracle Application Enablement Made Easy
 - Application development enhancements and changes
- **Service Level Confidence**
 - Workload management enhancements
 - High availability and resiliency enhancements
 - Manageability enhancements
 - Performance enhancements
 - Database administration changes
 - Security enhancements and changes
- **XML Insight**
- **Storage TCO Reduction**
 - Enhanced Compression

How about DB2 9.7 Performance?

Performance Enhancements



- **Statement Concentrator**
 - Promotes sharing of access plans
- **LOB inlining**
 - Save storage space and improve I/O performance
- **Scan sharing**
 - Better workload concurrency and performance without requiring any expensive hardware upgrades
- **Distribution map supports larger data warehouses**
 - In DB2 Cobra, the distribution map has grown from 4096 (4 KB) entries to 32 768 (32 KB) entries in one column.
 - This increase greatly reduces the potential for data skew issues, thus enabling warehouses to grow far larger.

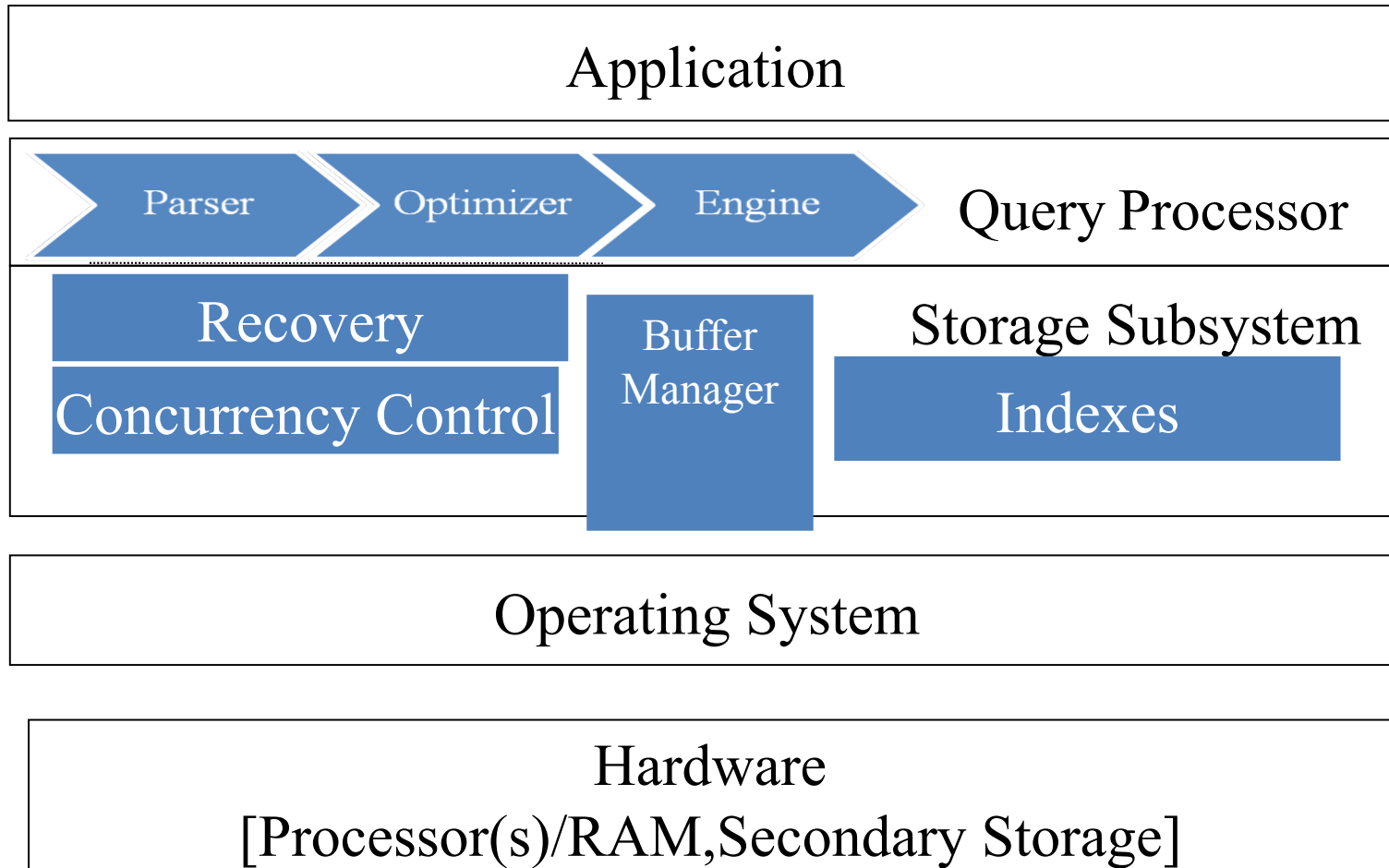
Simplistic Case

The following query runs too slowly

```
select *  
from R  
where R.a > 5;
```

What do you do?

Database Tuning



Principle # 1: Think Globally, Fix Locally

Intended Learning Outcomes

After the course the student should be able to:

- (1) Analyze the performance impact of design choices within a database system (indexing, query optimization, concurrency control and recovery)
- (2) Analyze and Compare the characteristics of different data models (relational, sequence, array)
- (3) Analyze and Compare the characteristics of distributed and parallel data management architectures (e.g., database management system, column store, cluster/grid data processing)
- (4) Design and execute experiments to evaluate a data management system performance

What you will learn

- Scientific Method (4)
 - Cycle of hypothesis, experimentation, analysis.
- Relational Database Internals
 - Architecture, data structures and algorithms.
 - Necessary for tuning purposes
- Tuning Principles for RDBMS (2)
- Beyond Relational DBMS
 - Architectures (3) and Data Models (2)
 - Much relevant for Scientific Data

How you will learn

- Lectures
 - Archi, Data model, Database internals, Tuning principles.
- Exercises
 - Paper. Understanding of lecture material.
- Experiments
 - Understanding of tuning principles.
 - Exercise of scientific method.
- Assignments (test as exercise / experiments)
- Project
 - sanitized tuning case study where you will use experiments to define/justify your tuning choices

Today

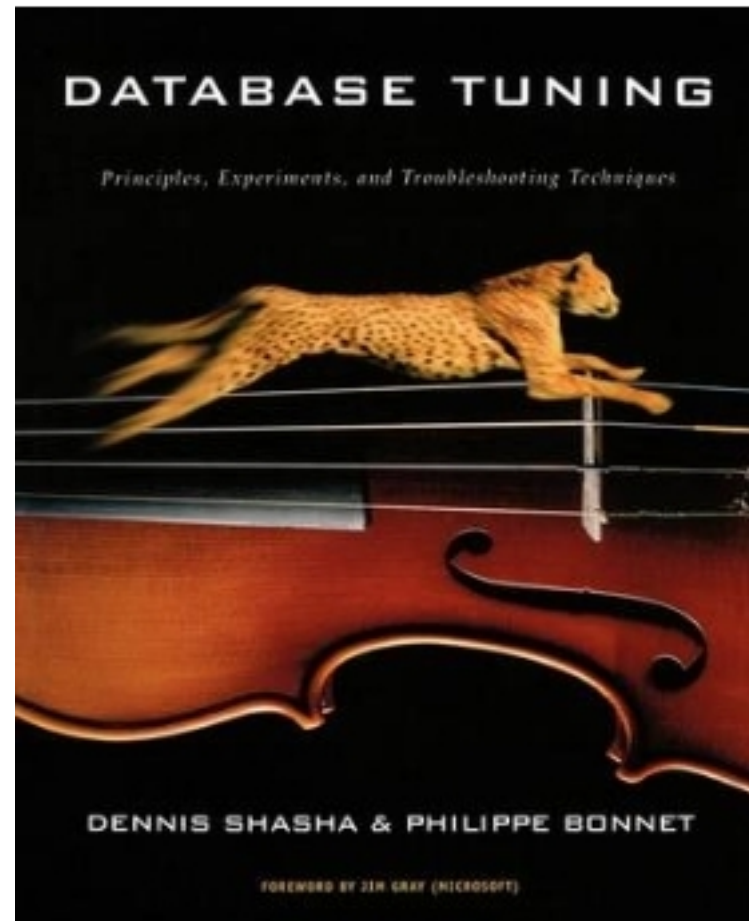
- Questionnaire
- Trends
- What you will learn and How
- Some logistics

Grades

- SDVN based on 3 assignments
- SDBT based on Project
 - 3 assignments are mandatory (need to have done to get a grade)

Text Book

- SL Books should have it (slbooks.dk)



Lectures and Exercises

- Lectures start at 9am
 - never again 8:30 from now on
- Class organized in units (mix of lectures and exercises/experiments):
 - Intro, architecture, data models, tuning the guts, index tuning, query tuning, tuning the application interface
 - Course plan on home page
- No units after easter
 - SDVM is done
 - SDBT full focus on project.

Amazon-IBM EC2

- Grant from IBM/Amazon to use IBM software on the cloud infrastructure from Amazon for experiments
 - More on this next time.
 - Docs on the course web page under Experiments.
- You need to create an account on Amazon WS <http://aws.amazon.com/>
 - You secure with a credit card
 - You get a \$100 credit voucher from Philippe so that you don't pay for the use that corresponds to this class.