Data on the Shore

How efficient use of hardware and good communication lead to the fourth paradigm

Anastasia Ailamaki
Data-Intensive Applications and Systems (DIAS)
Trends we would like to follow

Contexts/chip

- Pentium
- Itanium
- Intel Core2
- UltraSparc
- IBM Power
- AMD

Top ten warehouse size (TB)

Year


Scaling with number of cores

Year


Contexts/chip

- Pentium
- Itanium
- Intel Core2
- UltraSparc
- IBM Power
- AMD

Half the deal is eliminating stall time
Implicit and explicit parallelism
Data center cost breakdown

- Servers
- Networking Equipment
- Power Distribution
- Power
- Other infrastructure

From: http://perspectives.mvdirona.com/2010/03/18/

Must use all iron
Scalability: a LOT of effort

Concurrency != parallelism
Bitten by Amdahl’s law

The maximum benefit from a parallel system is given by

$$\frac{\text{new\_time}}{\text{old\_time}} \geq (1 - p) + \frac{p}{N}$$

where

- $p =$ parallel fraction of work
- $N =$ hardware parallelism

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!

Even a little serial code hurts a lot!
Guilty: Critical Sections

Unscalable

Unscalable

Fixed

Core
Core

Point-to-point communication

Core
Core
Core
Core

Logging

Group of cores

Core
Core
Core
Core

Unscalable → Fixed / Composable
Locking logical entities (e.g., records)

Transaction processing engine

Data

Locking = serial code
Typical Lock Manager

Higher HW parallelism $\rightarrow$ Longer Request Queues $\rightarrow$ Longer CSs $\rightarrow$ Higher Contention
Unpredictable access patterns

Partition the data?
Physical data partitioning

Distributed transactions, load imbalance
DORA: *Logical* partitioning

John  Anne  Chris  Niki

Move contention away from critical path
Predictable access patterns

![Graph showing database records vs. throughput and real CPU load](image-url)
How far can we go?

Sun Niagara T1
Insert-only workload

Scalability implies performance

*Shore-MT available at dias.epfl.ch
The LSST project expects to detect more than 100 billion objects, which is at least 10 times more than we’ve observed in the last 400 years of astronomy.
paradigm shift: yesterday...
theory, experiments, simulation
today...
Design/tuning automation
Support for complex structures
data-driven science
Automating scientific workflows
Sloan Digital Sky Survey, 2002

Observed objects

Q1    Q2

Index

TYPE CX CY E1 E2 E3

Unused part

Fast-moving galaxies?

Automating Database Design
Earthquake Simulation, 2004

Quake Group [www.cs.cmu.edu/~quake]

Scientist

Simulation Model Mesh

Simulation Output

fopen( ‘time050.dat’ )

100GB per time-step, 20000 time-steps
Performance? Access to complex structures?
Tetrahedral Mesh Models

- Queries
  - Point
  - Range
  - Feature

- Example: Visualization
  - Show ground velocity at Q
  - Draw the temperature of R

Efficiently process *geometric* queries
R-Tree indexing

Minimum Bounding Rectangle (MBR)

Disk page

Tight mesh connectivity hurts performance
Directed Local Search (DLS)

- Point queries
  - Step 1: Find a “nearby” element efficiently
  - Step 2: Follow path of adjacent elements
- No Minimum Bounding Rectangles
- Mesh topology (connectivity)
DLS will not work here
Indexing using R-trees

#reads increases with density
FLAT Indexing & Querying

• Indexing
  – group spatially nearby objects on disk pages
  – add links between neighboring groups
  – use traditional R-Tree to index disk pages

• Querying
  – *seed phase*: find random element inside range in R-Tree (not affected by overlap)
  – *crawl phase*: use seed element and recursively traverse all neighbors

**Meso-scale: from 10K to 1M neurons**
Touch Detection

Model Synapses
- electrical connections between axons and dendrites

Data Challenge
- 100K neurons => 5B synapses
- 30GB addl space to store synapse data
- Human Brain => 100B neurons ~PB space
- Current distance join inefficient

Grid Hash Join: 8 hours in 10 seconds
Broader Impact

• Aim to help researchers to organize, execute, analyze ever-bigger experiments
  – results put to use in real brain experiments
  – invent technology useful in mechanical engineering, medical analyses, earthquake simulations, etc.

Courtesy Cornell Fracture Group

Courtesy CMU Quake Group

Courtesy Cornell Fracture Group
RawDB: *in-situ* query processing

Give me your data as is

Give me your queries

(supports SQL + your tools)

Challenge: *full querying power*
Net-net

• Data-driven science needs robust data mgmt
  (we are not there yet)

• Exploit available hardware resources
  (scaling up is as important as scaling out)

• Communicate and build *many* bridges to sciences
  (as opposed to trying to sell one solution to all)
Thank you!
“Mr. Osborne, may I be excused? My brain is full.”