Topics for PhD research in data management and systems

Stratis D. Viglas

School of Informatics
University of Edinburgh
sviglas@inf.ed.ac.uk
Databases? Isn’t that the boring part of accounting?

Why study databases?

• Not really; database technology makes the world go round
• It’s not just about money from/to the bank
• Digital asset management (multimedia, data streams)
• Scientific applications: biological sciences, geosciences, DNA mapping
• Decision support, data analysis, data mining
• Big Data

Primary focus

Systems research; database systems are a well-understood area of immediate impact

Stratis D. Viglas
Why study databases?

Databases? Isn’t that the boring part of accounting?

- Not really; database technology makes the world go round
- It’s not just about money from/to the bank
  - Digital asset management (multimedia, data streams)
  - Scientific applications: biological sciences, geosciences, DNA mapping
  - Decision support, data analysis, data mining
  - Big Data

Primary focus

*Systems research; database systems are a well-understood area of immediate impact*
Storage technology

- Support for flash memory and solid-state drives
  - How does one optimise database systems for flash memory?
  - What sort of algorithms and techniques should there be in place?
- Disparate storage layer
  - Data structure optimisation
  - Placement in storage across media with widely different performance characteristics
- Persistent memory
  - It is not volatile memory; it is not a disk
  - Revisit all problems with a clean sheet
    - Algorithms, placement, concurrency, logging, etc.
Parallel processing

- **Small-scale**
  - Multicore and manycore setups
  - How does one orchestrate query processing in such environments?
    - Parallel algorithms, scheduling techniques, software support for parallel processing

- **Large-scale**
  - Cluster-based processing and distributed computing
  - Expressive file structures for storage
  - Algorithms for massive parallelisation of data analytics
  - Data-stream management on clusters
Code generation

- Front-end
  - Query
  - Results
  - Schemas
  - Statistics
  - Types
  - Preparator
  - Executor

- Back-end
  - Compiler
  - Linker
  - Evaluator
  - Buffer pool
  - Storage manager
  - Catalog
  - DB tables

- Code generation process:
  - Query
  - Parser
  - Syntax tree
  - Optimizer
  - Plan
  - Generator
  - Code
  - Compiler
  - Library file
  - Binary

- Schemata and Statistics:
  - Code
  - Schemas
  - Statistics
  - Types

- Pages and Binary:
  - Pages
  - Binary

- Database Components:
  - Catalog
  - DB tables
  - Buffer pool
  - Storage manager

- Stratis D. Viglas