Wintersemester 2013/2014

Seminare

Master Informatik
„Advanced Topics of Database Systems“

CS 5840 - Fachübergreifende Kompetenzen = englischsprachiges Seminar
CS 5480 - für den Bereich Software Systems Engineering = Seminar Software Systems Engineering

Privatdozent Dr. rer. nat. habil. Sven Groppe
groppe@ifis.uni-luebeck.de

Students' Duties

- Preparation of Slides
- Preparation of Handout
  - 2 to 3 pages, to be delivered to all participants and to the supervisors directly before the presentation
- Presentation
  - Approx. 1 hour (inclusive discussion)
- Attending presentations of all other students
  - Contributions to a lively discussion after each presentation

Timeline

participating in all other presentations and contributing to lively discussions

2 weeks

Assign-ment of topics

discussion with supervisor about presentation

electronic submission of pdfs of presentation and handout via Email (final from student's view)

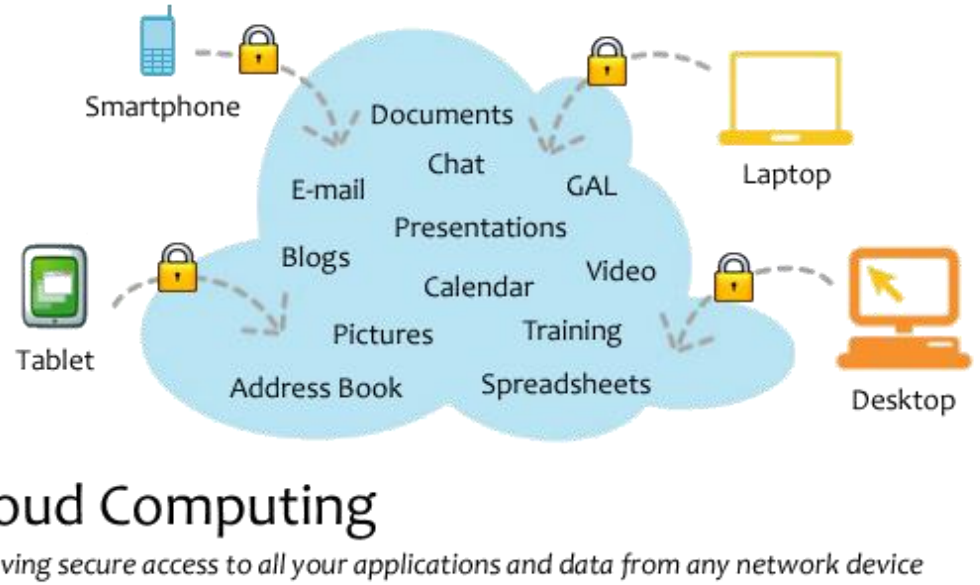
improving presentation and handout according to supervisor's comments and remarks

presentation, electronic submission of source files and pdfs of slides and handout to the supervisor via Email

Overview Topics

- Cloud Computing
- Semantic Web
- Recursive Queries
- Twitter
- Smart SSD

Cloud Computing



source: Wikipedia

- typically, cloud provider offers online-services
- multiple server-based computational resources via a digital network like internet
- applications are provided and managed by the cloud server
- data is stored remotely in the cloud configuration

Cloud computing - Topics

- **Yufei Tao, Wenqing Lin, Xiaokui Xiao, Minimal MapReduce Algorithms, SIGMOD 2013**
  - *minimal algorithm*: an algorithm that guarantees the best parallelization in multiple aspects at the same time, up to a small constant factor.
  - elegant minimal algorithms for a set of fundamental database problems
  - demonstration of their excellent performance with extensive experiments

Semantic Web

- Idea
  - "web of data" that enables machines to understand the semantics, or meaning, of information on the World Wide Web.
  - extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other
- Semantic Web databases can be seen as graph databases for labelled and directed graphs
- Additional implicit data is considered based on ontology inference
  - E.g. *Peter* has the *driverLicenseNo 456* (which is a fact) and only a person can have a *driverLicenseNo* (expressed in an ontology), then *Peter* is a *person* (implicitly present)

Semantic Web - Topics

- **Kai Zengy, Jiacheng Yang, Haixun Wangz, Bin Shaoz Zhongyuan Wangz, A Distributed Graph Engine for Web Scale RDF Data, VLDB 2013**
  - distributed, memory-based graph engine for web scale RDF data
  - Instead of managing the RDF data in triple stores or as bitmap matrices, RDF data is stored in its native graph form
  - much better query performance
  - the system can support other operations (e.g., random walks, reachability) on RDF graphs as well
- **Wangchao Le, Anastasios Kementsietsidis, Songyun Duan, Feifei Li, Scalable Multi-Query Optimization for SPARQL, ICDE 2012**
  - revisits the classical problem of multiquery optimization in the context of RDF/SPARQL
  - heuristic algorithms that partition the input batch of queries into groups such that each group of queries can be optimized together
  - portable across different RDF stores

Semantic Web - Topics

- **Souripriya Das, Seema Sundara,Matthew Perry, Jagannathan Srinivasan, Jayanta Banerjee, Aravind Yalamanchi, Making Unstructured Data SPARQL - Using Semantic Indexing in Oracle Database, ICDE 2012**
  - Semantic Indexing feature introduced in Oracle Database 11g Release 2 for indexing unstructured text (document) columns
  - enables searching for concepts (such as people, places, organizations, and events) with further options for sense disambiguation and term expansion by consulting knowledge captured in OWL/RDF ontologies
- **Leonid Libkin, Juan Reutter, Domagoj Vrgoc, TriAL for RDF: Adapting Graph Query Languages for RDF Data, PODS 2013**
  - navigational patterns over triples are beyond reach of both RDF languages and graph query languages that work on encodings of RDF
  - language that work directly on RDF triples, and have both relational and navigational querying facilities, just like graph query languages

Recursive Queries

- **Datalog as subset of prolog is a rule language for deductive databases**
- **Example:**
  Facts: parent(bill,mary).
  parent(mary,john).

  Rules: ancestor(X,Y) :- parent(X,Y).
  ancestor(X,Y) :- parent(X,Z), ancestor(Z,Y).

  Query: ?- ancestor(bill,X).

- **Topic:**

  K. Tuncay Tekle, Yanhong A. Liu, More Efficient Datalog Queries: Subsumptive Tabling Beats Magic Sets, SIGMOD 2011
  - top-down evaluation method with more reuse of answers than the dominant tabling strategy

Twitter

- **Michael Busch, Krishna Gade, Brian Larson, Patrick Lok, Samuel Luckenbill, Jimmy Lin, Earlybird: Real-Time Search at Twitter, ICDE 2012**
  - core retrieval engine that powers Twitter's realtime search service
  - A key requirement of real-time search is the ability to ingest content rapidly and make it searchable immediately, while concurrently supporting low-latency, highthroughput query evaluation
  - differences to indexes used in traditional web search

Smard SSD

- **Jaeyoung Do, Yang-Suk Kee, Jignesh M. Patel, Chanik Park, Kwanghyun Park, David J. DeWitt, Query Processing on Smart SSDs: Opportunities and Challenges, SIGMOD 2013**
  - Smart SSD will package CPU processing and DRAM storage to run user programs inside a Smart SSD
  - opportunities and challenges associated with exploiting this functionality of Smart SSDs for relational analytic query processing
  - initial prototype of Microsoft SQL Server running on a Samsung Smart SSD
  - significant performance and energy gains can be achieved by pushing selected query processing components inside the Smart SSDs

- **Cloud Computing**

1. **Yufei Tao, Wenqing Lin, Xiaokui Xiao, Minimal MapReduce Algorithms, SIGMOD 2013**
   http://www.cse.cuhk.edu.hk/~taoyf/paper/sigmod13-mr.pdf

- **Semantic Web**

2. **Kai Zengy, Jiacheng Yang, Haixun Wangz, Bin Shaoz Zhongyuan Wangz, A Distributed Graph Engine for Web Scale RDF Data, VLDB 2013**
   http://www.vldb.org/pvldb/vol6/p265-zeng.pdf
3. **Wangchao Le, Anastasios Kementsietsidis, Songyun Duan, Feifei Li, Scalable Multi-Query Optimization for SPARQL, ICDE 2012**
   http://www.computer.org/csdl/proceedings/icde/2012/4747/00/4747a666.pdf
4. **Souripriya Das, Seema Sundara,Matthew Perry, Jagannathan Srinivasan, Jayanta Banerjee, Aravind Yalamanchi, Making Unstructured Data SPARQL - Using Semantic Indexing in Oracle Database, ICDE 2012**
   http://www.computer.org/csdl/proceedings/icde/2012/4747/00/4747b405.pdf
5. **Leonid Libkin, Juan Reutter, Domagoj Vrgoc, TriAL for RDF: Adapting Graph Query Languages for RDF Data, PODS 2013**
   http://homepages.inf.ed.ac.uk/s1058408/data/trial.pdf

- **Recursive Queries**

6. **K. Tuncay Tekle, Yanhong A. Liu, More Efficient Datalog Queries: Subsumptive Tabling Beats Magic Sets, SIGMOD 2011**
   http://www.cs.stonybrook.edu/~liu/papers/RuleQueryBeat-SIGMOD11.pdf

- **Twitter**

7. **Michael Busch, Krishna Gade, Brian Larson, Patrick Lok, Samuel Luckenbill, Jimmy Lin, Earlybird: Real-Time Search at Twitter, ICDE 2012**
   http://www.computer.org/csdl/proceedings/icde/2012/4747/00/4747b360.pdf

- **Smart SSD**

8. **Jaeyoung Do, Yang-Suk Kee, Jignesh M. Patel, Chanik Park, Kwanghyun Park, David J. DeWitt, Query Processing on Smart SSDs: Opportunities and Challenges, SIGMOD 2013**
   http://pages.cs.wisc.edu/~jignesh/publ/SmartSSD.pdf

More topics upon request…