Seminar

Column-Oriented Database Management Systems

Summer Term 2012

Lehrgebiet Informationssysteme

Weiping Qu
qu@cs.uni-kl.de

AG Datenbanken und Informationssysteme

AG Heterogene Informationssysteme

Goals

- Familiarize yourself with a scientific topic independently
- Find scientific literature on web or in the library, cite correclty (ACM digital library, DBLP, Citeseer, Google)
- Prepare a written composition, presentation, and discussion
- Time Management
- Don’t copy and paste!

Written composition

Searching for Literature

- Some basic literature is provided by your supervisor
- You have to search for further literature
- Collect a list of related literature and send it to your supervisor

Annotated Table of Contents (TOC)

- Prepare an annotated TOC for your supervisor (approx. 1.5 pages)

Written Composition

- LNCS Layout
- PDF Format required
- Length: 6000–8000 Words (net.) ≅ 15–20 Pages
- Correct and complete bibliography

Links

Homepage

- http://wwwlgis.informatik.uni-kl.de/cms/courses/seminar/

LNCS

- http://www.springer.com/computer/lncs?SGWID=0-164-6-793341-0

Typography

- http://zvisionwelt.wordpress.com/downloads/#typokurz

ACM Digital Library

- http://dl.acm.org/

DBLP

- http://www.informatik.uni-trier.de/~ley/db/

Presentation

- Length: 60 Minutes (45 minutes for the presentation, 15 minutes for the discussion)
- Presentation:
  - Projector (private or one of our notebooks)
  - Overhead slides
- You must submit your presentation electronically, one week after you gave your talk
- Presentations (Room 36/336) are scheduled for
  - Monday, July 16th 2012, 10:00 a.m – 1:00 p.m and 2:00 p.m – 4 p.m.
  - Tuesday, July 17th 2012, 10:00 a.m – 1:00 p.m.

Column-Oriented Data Layout

- Another storage scheme compared with traditional row(tuple,record)-oriented layout
- Advantages: good compression, efficient read (prefetching), good cache performance, efficient processing(vectorized, SIMD), good indexing …
- Suitable for read-only, data-intensive application on large data repositories (queries that access only subset of whole data set)
- Often used in Data warehouse, Business Intelligence
- Products: C-Store (MIT, US), MonetDB (CWI, Netherland)

Topics

Topic 1: General comparison between different table layouts (row-/column-store, hybrid)

Topic 2: Column-oriented compression techniques

Topic 3: Query processing (vectorized, LLVM, MIL)

Topic 4: Tuple reconstruction with materialization strategies in column-store

Topic 5: Indexing column-store

Topic 6: Update mechanisms in column-store

Topic 7(Backup): One-size-fits-all databases (Hyper, OctopusDB)

Topic 8(Backup): Operating system for massively-parallel data processing (Barrelfish)

How you successfully pass the seminar

- **“Unbenoteter Schein”:**
- Meet the deadlines!
- Decent presentation
- Be present when others give their talks
- **“Benoteter Schein”**
- See above

Criteria for grading:

- Quality of your written composition
- Quality of your presentation (including your slides)
- Discussion
- Meeting the deadlines
- Overall impression of your supervisor
- ...

Schedule

- Thursday, April 19th: Kick-off meeting
- Thursday, May 3rd: Deadline for literature list
- Thursday, May 17th: Deadline annotated TOC
- Thursday, June 14th: Deadline written composition
- Thursday, June 28th: Deadline corrected written composition
- Monday, July 16th (10:00 – 13:00 & 14:00 – 16:00): Presentation I
- Tuesday, July 17th (10:00 – 13:00) : Presentation II

All deadlines are strict!

Thank your for your attention!

Questions

Select your topic...now!

Topic 1: General comparison between different table layouts (row-/column-store, hybrid)

Topic 2: Column-oriented compression techniques

Topic 3: Query processing (vectorized, LLVM, MIL)

Topic 4: Tuple reconstruction with materialization strategies in column-store

Topic 5: Indexing column-store

Topic 6: Update mechanisms in column-store

Topic 7(Backup): One-size-fits-all databases (Hyper, OctopusDB)

Topic 8(Backup): Operating system for massively-parallel data processing (Barrelfish)