Agenda for Dagstuhl Seminar 12282: Database Workload Management

July 9 - 13, 2012

Goal: produce the following artifacts:

(1) Descriptions of the most significant database workload management challenges facing industry, each defined in terms of a rough specification of a target workload and its objectives.

(2) For some number of these challenges, a sample workload that would demonstrate the challenge, and that would allow solutions to the challenge to be validated and compared.

(3) Descriptions of the "best" workload management techniques and best practices (both proven and unproven) that might apply to these challenges (both in practice and in research), as well as a partially-annotated bibliography that lists the papers that discuss those techniques and that summarizes their potential benefits and limitations.

(4) Identify new synergies and opportunities for new techniques and new applications of existing techniques.

Sunday: arrival

Sunday afternoon: Everyone travels to Dagstuhl. We will create a twiki page where you can post your arrival/departure information so that people flying into Frankfurt at similar times to meet up at the airport train station and travel down together.

Monday: Challenges from Commercial Systems

Monday early morning

1. **Welcome and context setting**.
   Introduce goals, and present the agenda for the rest of the week.
   Audience challenge: map the challenges.
2. **Challenges from Commercial Systems**
   Introducing systems and describing key database workload management challenges.
   - 45 min Douglas Brown, on Teradata.
   - 20 Rao Kakarlamudi, on HP Seaquest
   - 20 Sivaramakrishnan Narayanan, on EMC Greenplum

Coffee break

Monday late morning

3. **Challenges from Commercial Systems, continued**
   10-20 minutes each, introducing systems and describing key database workload management challenges.
   - Russell Sears, on Yahoo
   - Robert Chansler, on LinkedIn
   - Archana Ganapathi, on Splunk
   - Jingren Zhou, on Microsoft

Lunch

Monday early afternoon

Workloads

- Michael Seibold on mixed workload benchmark
- Kai-Uwe Sattler on tractor pulling workload
- Yanpei Chen on Hadoop/MapReduce workloads

Cake break

Monday late afternoon

4. **Categorize and prioritize challenges from commercial systems**
    - **Goal:** identify which challenges are most significant and characterize the most significant challenges.
    - **Process:** Break into 3 groups assigning the morning's speakers to groups. Place morning's speakers in break-out groups, reflecting these groupings.
    - Assign people semi-randomly to break-out groups. Identify a moderator/facilitator and a scribe for each group. Each group will produce:
        1. An ordered list of the challenges from the morning, and a rationale for the order.
        2. Capture the known knowns, known unknowns, unknown unknowns, etc.
        3. A high level sketch of a workload that would demonstrate the top challenge(s), including metrics for success.

Monday dinner

Monday early evening (in the beer room)

5. **Drinks and snacks and conversation**
    - Organizers will present small prizes for insightful/clarifying questions.
    - Select favorite maps of the day.
    - Informal feedback about what's working and what's not working.

Tuesday: Research Context

Tuesday early morning

6. **Announcements and context setting for the day**.

   Audience challenge: map the current state of the art (what can be done today, regardless of whether it is done), as well as what cannot be done (but is needed for the commercial challenges).

7. **Presentations from yesterday's break-out groups.**

Coffee break

Tuesday late morning

8. **Performance modeling/prediction, algorithms**

   20 minutes each.

   - Archana Ganapathi
   - Y.C. Tay
   - Jens Dittrich on Hadoop++/HAIL
   - Norbert Ritter

Lunch

Tuesday early afternoon

9. **Mixed Workloads**

   20 minutes each, introducing techniques and describing benefits and limitations

   - Wendy Powley
   - Ashraf Abounalga
   - Barzan Mozafari
   - Shivnath Babu

Cake break

Tuesday late afternoon

10. **Robust Query Processing and Database Workload Management**

    20 minutes each, introducing techniques and describing benefits and limitations

    - Goetz Graefe
    - Awny Al-Omari
    - Kostas Tzoumas
    - Bradley Kuszmaul/Michael Bender on Work stealing/Scheduling

Tuesday dinner

Tuesday early evening (in the beer room)

11. **Drinks, snacks, conversation. Prizes.**

Wednesday early morning

12. **Announcements and context setting for the day**.
    Goal for the day is to select focus challenges and produce sample workloads that demonstrate those challenges .
13. Select focus challenges. Self-divide into breakout groups that will select a challenge to focus on and produce sample workloads for that challenge. Select a facilitator and a scribe for each group.
14. Break out into groups and select challenges.
15. Report back and say which challenges each group selected, and why.

Coffee break

Wednesday late morning

16. Continue breakout groups and produce sample workloads
    1. Develop a concrete workload specification that any new researcher will be able use to create an instance of a workload for that challenge. As part of this task, aspects like datasets, queries/transactions, temporal/spatial patterns, etc., must be identified. Will it be possible to generate a workload instance synthetically? Do workload specifications (e.g., TPC-H) exist already for that category, and are they adequate? Is workload characterization for this category still an unsolved problem? Are real traces available or can they be made available? ...
    2. Specify the workload management metrics and objectives for each challenge. Objectives should reflect the Service Level Objectives that will be defined over the metrics. If it is not clear, then it should be specified how the metrics as well as the adherence to objectives can be measured.
    3. What risks makes the sample workload potentially not useful/not practical/not publishable? How can we get around them?

Lunch

Wednesday early afternoon

17. **Hike! Or contingency plan for rain?**
    - To be led by Goetz. Wear appropriate shoes and bring water. Leave your laptop at the institute.

Cake (and water) break

Wednesday late afternoon

18. **Report back from break-out groups, presenting their results**
    - 10 minutes each to present their results.
    - Audience challenge: suggest what would make someone else's work more publishable.

Wednesday dinner

Wednesday early evening (in the beer room)

Snacks, drinks, conversations, prizes

Thursday: laying groundwork for test suites and papers

Thursday early morning

19. Announcements
    collect departure information.
20. Set context, then break into groups. **Goal is to produce 3-5 page abstracts with fake data that could be produced if the proposed workloads actually existed.**
21. Break into groups and write abstract for collaborative paper(s), and sketch graphs that would demonstrate the problem and evaluation of solutions. What data would you need that you anticipate might be difficult to get? What would make the results of this paper more (or less) significant? How does this work propose to push the state of the art? What would make this work difficult (or easy) for a commercial system to actually use, in real-life.

Coffee break

Thursday late morning

22. Regroup and present plans and insights. Discuss and collect feedback. Should any efforts merge? Are any of them not practical?
23. **Continue breakout groups.**

Lunch

Thursday early afternoon

24. **Continue breakout groups** to produce the 3-5 page abstracts. 2:30 pm deadline. Post the abstracts on the wiki.

Thursday late afternoon

25. **Abstract/paper reviews**
    - Each group should review the work of all the other groups.
    - Everyone should produce reviews and feedback.
    - Each group should meet individually like a PC and decide on "most promising practical idea," "most promising wild idea," "ripest low-hanging fruit," and "most difficult to achieve."

Thursday dinner

Thursday early evening (in the beer room)

26. **Gong Show!**

Friday early morning

1. Presentations by the top ranked papers.

Coffee break

Friday late morning

1. Conclusions and awards!! ☺

Final lunch

Friday early afternoon: almost everyone departs. (Pre-order taxi pools on Thursday to the train station.)