CS 8803-MDS

Human-in-the-loop Data Analytics

Lecture 7

09/14/22

Logistics

Extra credit for discussing proposal outline

Thursday/Friday 11-12

Project proposal presentation (09/21)

https://bit.ly/3Bg0Kst

5 min per group (timed)

* Keep track of the division of work

Today's class

Experiences with Approximating Queries in Microsoft's Production Big-Data Clusters

Researcher: Abhinav

Database Benchmarking for Supporting Real-Time Interactive Querying of Large Data

Author: Hamsika

Archaeologist: Jingfan

PORTLAND

ACM SIGMOD/PODS International Conference on Management of Data
June 14-19th, 2020

ACM SIGMOD +PODS 2020

Portland, Oregon USA

SIGMOD PODS

ACM SIGMOD/PODS International Conference on Management of Data June 14 - June 19, 2020 Portland, OR, USA

Welcome

- Homepage
- News

Organization

- Experience Report
- Organization
- SIGMOD PC
- PODS PC

2020 ACM SIGMOD/PODS @ Portland, OR, USA

SIGMOD/PODS 2020 Experience Report now available

Coronavirus Updates

Updates on SIGMOD/PODS 2020 Registration (12 May 2020)

Calls For Submissions

- Important Dates
- Calls for Submissions

PODS Program

- Program Overview
- Detailed Program
- Research Papers
- Keynote Talk

Contribution/Strengths

First benchmark of its kind

- Benchmark is derived from real user traces
- Central to the paper's benchmark is this idea of workflows. The paper explains a workflow as a sort of summary that includes a study participant, a dataset, and a task.
- Unlike traditional benchmarks which focuses on throughput and resource usage, this benchmark has three measures: throughput, latency and accuracy.
- Through user interactions two general categories were defined for characterizing traces: gesture and filter speed, interaction rates and pacing.

Limitation/Weaknesses

Limitation of the study

- The participants are users who use databases regularly
- The number of users in the paper seem very low
- The systems being used by the participants was not ideal (since it did not run with a sub 100ms latency for all queries, especially for 10M rows.) So, the users would be using the system in a way that they would assume the system would perform good, which creates bias.
- User traces are also affected by screen and mouse refresh rate.
- Using crossfilter-styled visualization limits the number of attributes

Limitation/Weaknesses

Limitation of the benchmark

- The current way of collecting user traces are kind of expensive and not scalable
- Lack of description for the setup of the workflows and the translation of the workflows into the queries
- More comparisons with TPC-H, SSB benchmarks
- A comparison of amount of storage used would be good, as that is an important cost. There could be tradeoffs on storage vs performance.
- They considered only 5 databases and didn't include popular databases like Oracle, MongoDB and DynamoDB.
- Additional factors for performance differences: MonetDB and DuckDB use a columnar data layout; MonetDB is an in-memory database.
- Their assumptions that network transfer time and rendering are negligible can be questioned.

Future work

- How to streamline benchmark creation
- How to predict user action
- How to strategically throw-away non-essential queries and merge appropriate queries to help with peak query-load
- We should have a system of benchmarks that can run different benchmarks depending on the requirement - this will provide more desirable results rather than developers selecting benchmarks that are convenient or give good results only (biased).
- Is it possible to use answers (or internal states) from pervious queries to speed up the answers of later queries? For example, if a user drags a brush on the screen, then the difference between consecutive queries should be pretty small, because the move should be continuous.

Other thoughts

- I do have concerns about providing 'too much' cognitive ease to users, especially those conducting statistical analysis. I think its worthwhile that researchers investigate potential biases these exploratory tools may introduce into the thinking of the data scientist employing them. Could be the case that slowing down the user in some situations is more valuable than going fast and having a grand old time.
- I think a widely-accepted evaluation benchmark will direct research effort and resources. Benchmarks should be handled and proposed with extra care.
- The paper has a good reading flow.

Behavioral Measures

less obtrusive to participants, often more quantitative in nature

Direct Observation

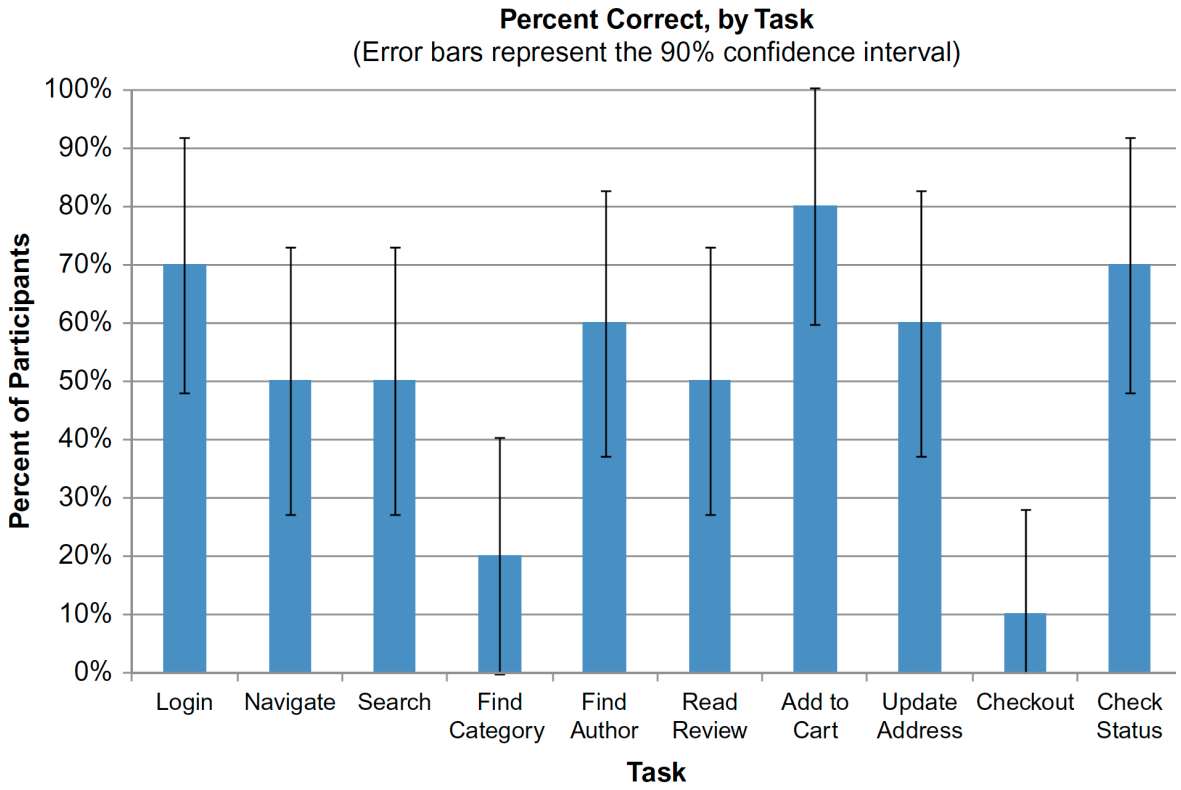
researcher observes and records participant activity

Indirect Observation

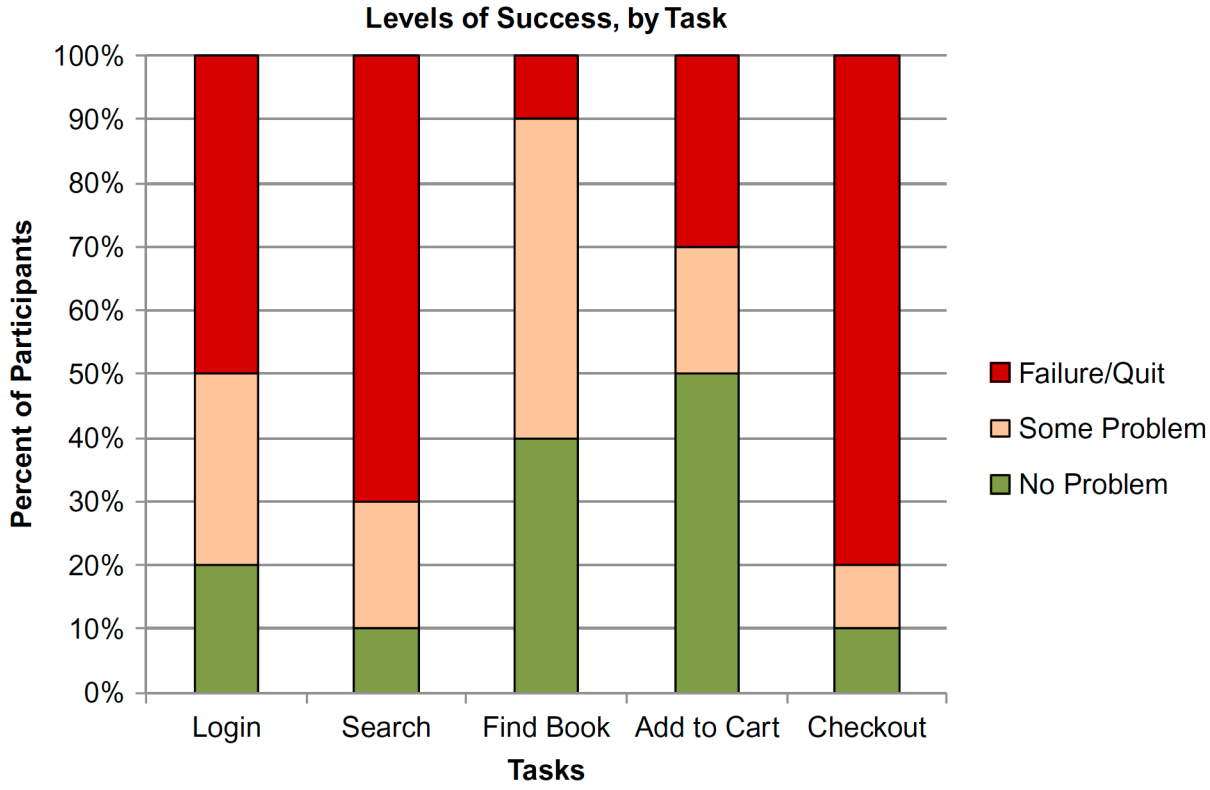
video recorded and analyzed later

participant records activity (e.g., diary entries)

How to measure: behavioral



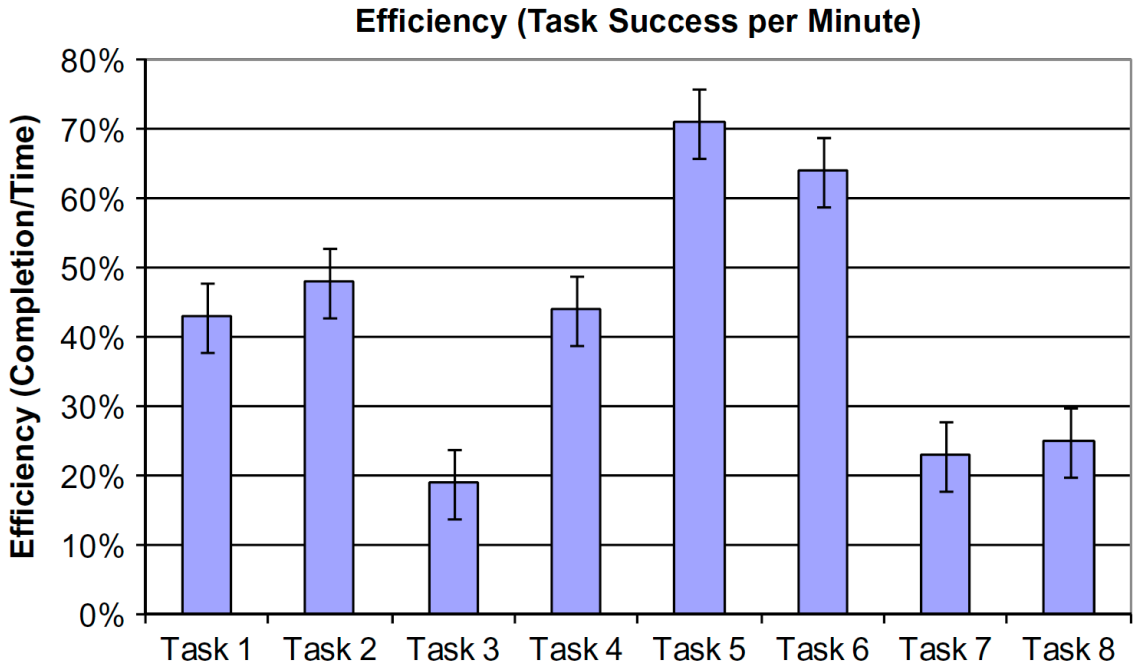
Binary success



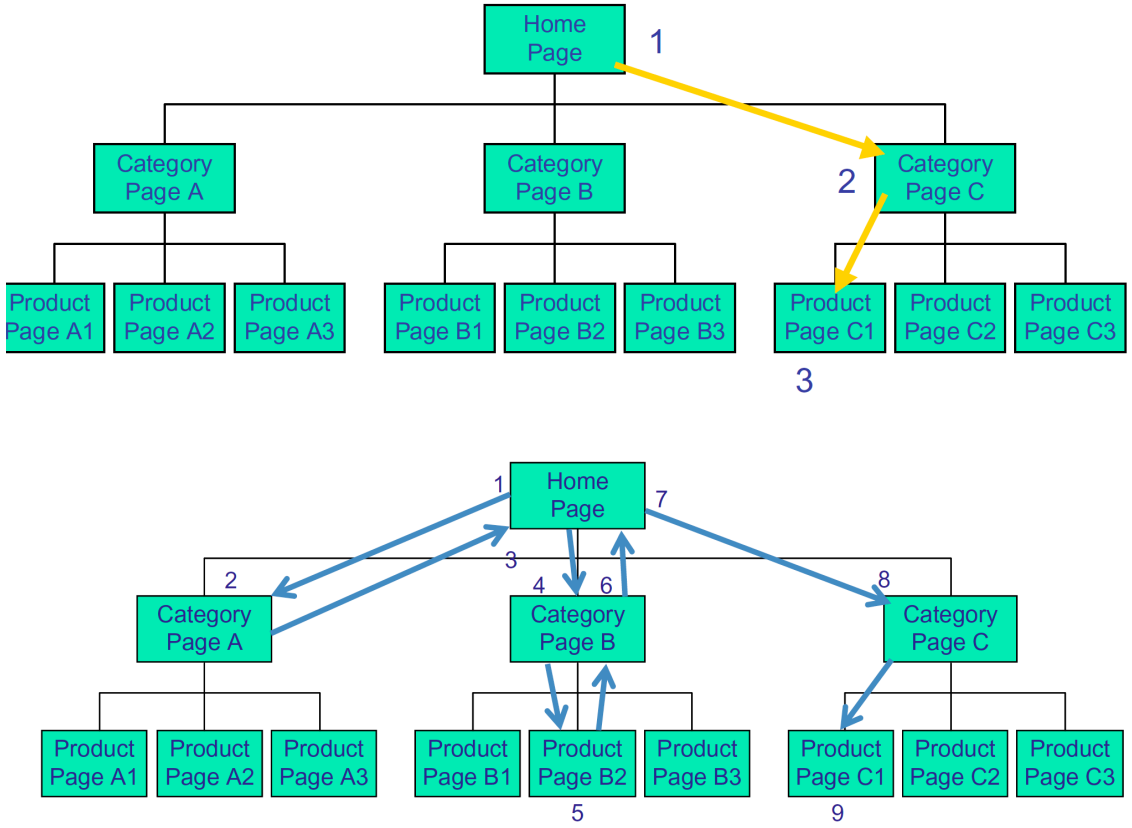
Levels of success

Source: Albert, Bill, and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

How to measure: behavioral



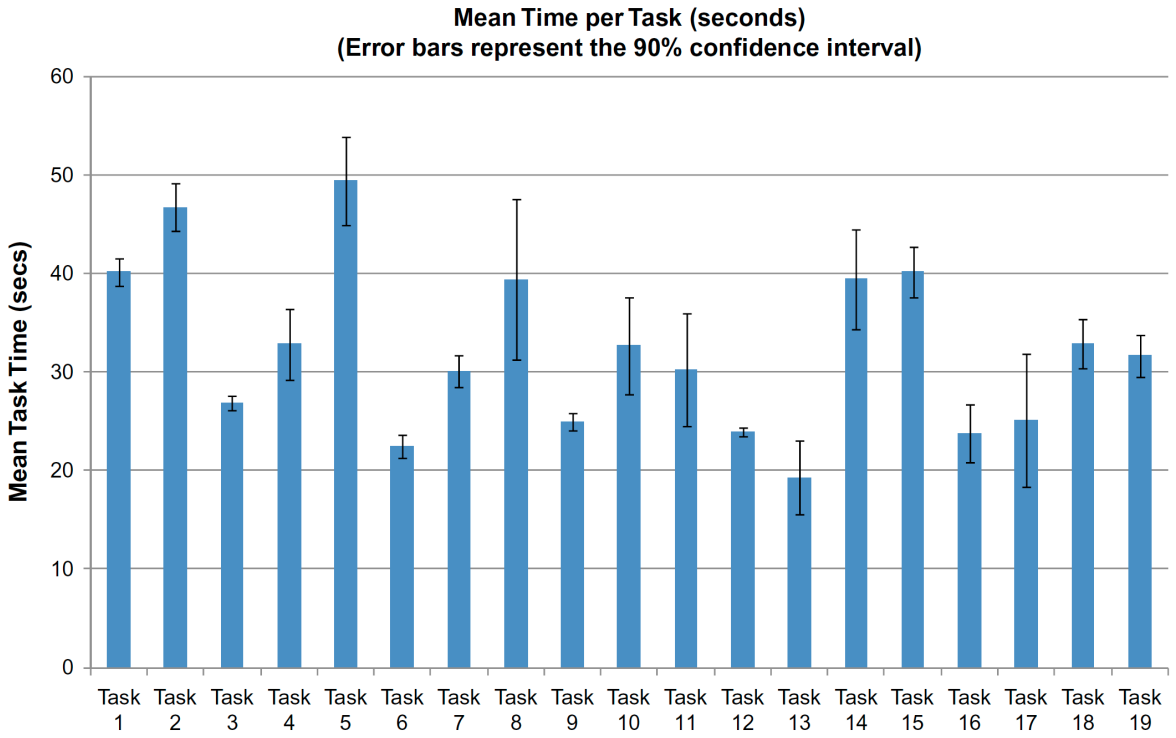
Efficiency



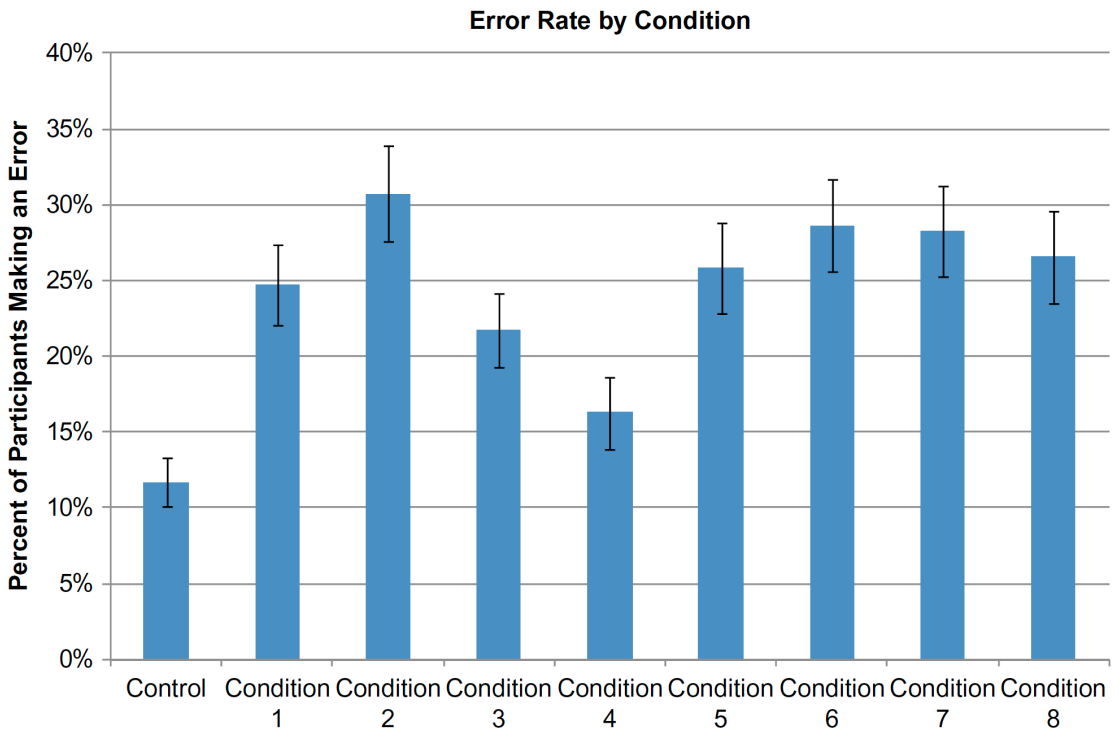
actions

Source: Albert, Bill, and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

How to measure: behavioral



Time on Task



Errors

Source: Albert, Bill, and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

How to measure: behavioral

Engagement patterns

heatmap.js

Dynamic Heatmaps for the Web

heatmap.js is a lightweight, easy to use JavaScript library to help you visualize your three dimensional data!

https://www.patrick-wied.at/static/heatmapjs/

IRB: Institutional Review Board

https://oria.gatech.edu/irb

Experiments conducted at universities require ethical oversight

Consider: risk to participants, data privacy etc.

Protocols must be reviewed and approved by IRB

Study might quality for Exempt Review (but still need to apply)

Next class

Hillview: A trillion-cell spreadsheet for big data

Author: Akshay

Archaeologist: Eric

Practioner: Ashmita

Researcher: Vishnu