TDT4225 Very Large, Distributed Data Volumes

Svein Erik Bratsberg
Department of Computer Science (IDI), NTNU

Info

- Lecturer:
  - Svein Erik Bratsberg, office 209, email: sveinbra@ntnu.no
- Lectures:
  - Tuesdays 12.15-14.00 KJL5 (Kjelhuset)
  - Automical video recording using Panopto
- Exercises: Alexander Fredheim, Erling Moen, Simen Tengs
  - Time for guidance. Thursdays 11-12 KJL5 (Kjelhuset)
- Exam: 15. Des 2021. Hopefully normal university exam 50 %
  Permitted examination support material: D – No written and handwritten examination support materials are permitted. A specified, simple calculator is permitted.
- Exercises:
  - 2 with evaluation (25 % each)
  - 2 compulsory.
- Partial evaluations (delvurderinger)

Curriculumn (1)

- Martin Kleppmann: Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems
  - Chap 1-9. (383 pp).
- Håvard Dybvik: Evaluating the potential of LSM-trees to supersede B-trees in databases
  - Chapter 2.1 - 2.9 Survey of Storage, Indexing, and Database (31 pp)
- George Coulouris et al: Distributed Systems - Concepts and Design
  - Chapter 14 Time and global states (44 pp).

Curriculumn (2)

- Ongaro/Ousterhout: In Search of an Understandable Consensus Algorithm (RAFT), USENIX 2014 (16 pp)
- Dynamo: Amazon's Highly Available Key-value Store, SOSP '07, (16 pp).
- Spanner: Google's Globally-Distributed Database, OSDI 2012, (14 pp) Video.
- Dostoevsky: Better Space-Time Trade-Offs for LSM-Tree Based Key-Value Stores via Adaptive Removal of Superfluous Merging, SIGMOD 2018, (16 pp) Video

Lecture plan, temporary

Date	Tue 12-14	Theme
24.Aug	KJL5	Intro + LSM/SSD
31.Aug	KJL5	Kleppmann. Chap 1 and 2.
7.Sep	KJL5	Kleppmann. Chap 3 and 4.
14.Sep	KJL5	Kleppmann. Chap 5 and 6.
21.Sep	KJL5	Kleppmann. Chap 7 and 8.
28.Sep	KJL5	Kleppmann. Chap 9.
5.Oct	KJL5	Coulouris. Chap 14.
12.Oct	KJL5	RAFT
19.Oct	KJL5	Dynamo
16.Oct	Video	Spanner
2.nov	Video	Dostoevsky
12.nov		???

Exercises

- Compulsory. All 4 must be approved. Done in groups of up to 3.
- Exercise 1: Theory - SSDs, LSM-trees and Kleppmann stuff (17 Sep)
- Exercise 2: Programming - MySQL (8 Oct) (25 % eval)
- Exercise 3: Programming - MongoDB (22 Oct) (25 % eval)
- Exercise 4: Theory - Kleppmann stuff, Coulouris stuff and systems (RAFT/Dynamo/Spanner/Dostoevsky) (5 Nov)