ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

HY590.45
Modern Topics in Scalable Storage Systems

Kostas Magoutis
magoutis@csd.uoc.gr
http://www.csd.uoc.gr/~hy590-45

Απαιτήσεις του Μαθήματος

- Project (50%)
  - Μελέτη και πειραματική αποτίμηση συστήματος/ων
  - Ατομική ανάθεση
  - Περιοδικές αναφορές προόδου
- Δύο παρουσιάσεις σχετικής δουλειάς (15%)
  - Επιλογή από πρόσφατα συνέδρια (όπως FAST, SOSP, κλπ.)
- Ένα γραπτό κουίζ, Απρίλιος/Μάιος 2022 (5%)
  - Θέματα από ένα paper, θα ανακοινωθεί πριν
- Τελική εξέταση (30%)
  - Επιλογή τριών papers, θα ανακοινωθούν πριν

Syllabus (tentative)

Select project topic (by 2/3, prepare brief proposal, see instructions on site)

Select related work papers (by 14/3), recent publications @FAST, SOSP, OSDI, etc.

Syllabus

Date	Topic	Readings, notes
Mon 14/2	Course overview	-
Wed 16/2	Background	See recommended readings
Mon 21/2	Extending file systems over the network	Sandberg: Design and Implementation of the Sun Network Filesystem
Wed 23/2	NFS (contd.)	Macklem: Not Quite NFS, Soft Cache Consistency for NFS
Mon 28/2	Distributed coordination	Ongaro: In Search of an Understandable Consensus Algorithm
Wed 2/3	Raft (contd.)	*Proposals for project topics due*
Mon 7/3	Clean Monday	-
Wed 9/3	Distributed virtual disks	Lee: Petal: Distributed Virtual Disks
Mon 14/3	Petal (contd.)	*Proposals for papers (presentations I & II) due*
Wed 17/3	Distributed file systems I	Thekkath: Frangipani: A Scalable Distributed File System
Mon 21/3	Distributed file systems II	Ghemawat: The Google File System
Wed 23/3	Google file system (contd.)	-
Mon 28/3	Related work presentations I	-
Wed 30/3	Related work presentations I	-
Mon 4/4	Related work presentations I	-
Wed 6/4	Application-specific storage systems	Saito: Manageability, Availability and Performance in Porcupine: A Highly Scalable, Cluster-based Mail Service
Mon 11/4	Porcupine (contd.)	-
Wed 13/4	Structured data	Chang: A Distributed Storage System for Structured Data
Mon 18/4 - Fri 29/4	Easter recess	-
Mon 2/5	BigTable (contd.)	-
Wed 4/5	In-class quiz	Project progress reports
Mon 9/5	Related work presentations II	-
Wed 11/5	Related work presentations II	-
Mon 16/5	Related work presentations II	-
Wed 18/5	Distributed transactions	Corbett: Spanner: Google's Globally-Distributed Database
TBA	Final exam	-
TBA	Project presentations	-

Final report & presentation dates in June, exact date TBA

Course themes

- Fundamentals: Organization, metadata, journaling, Raft
- Distributed file systems: NFS
- Scalable virtual disks: Petal
- Shared-disk distributed file systems: Frangipani
- Separate data from metadata: Google file system
- Application-specific scalable storage: Porcupine
- Scalable storage of structured data: BigTable
- Scalable transactions: Spanner

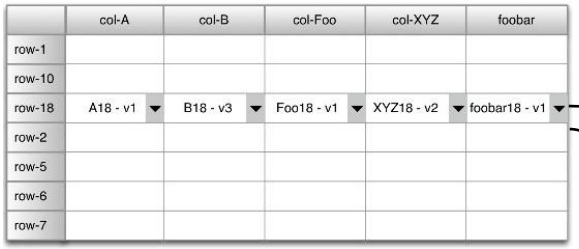
Scalable Storage Systems: Goals, Requirements

- Expandability
  - Increase system size (capacity) as needed
- Performance
  - Increase linearly with system size
- Availability
  - Survive failures gracefully
- Manageability
  - React to changes automatically

Concepts

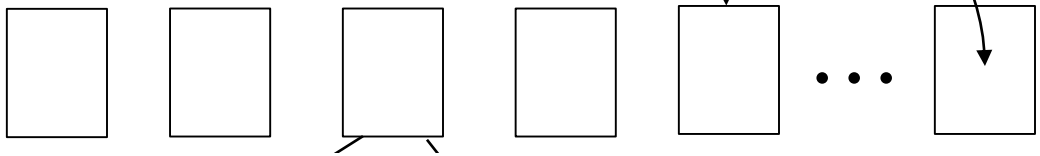
API, semantics

Data model



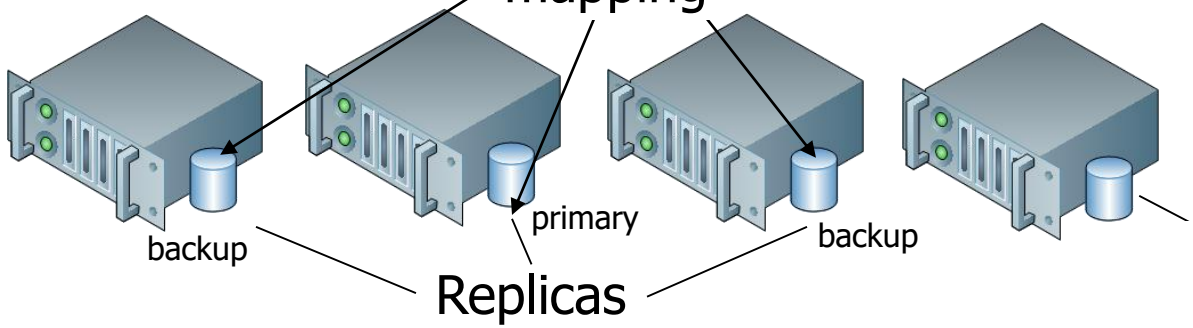
mapping

Horizontal partitions (shards)



mapping

Servers



Storage device & networking technologies

Leverage large-scale infrastructures, address challenges in doing so