

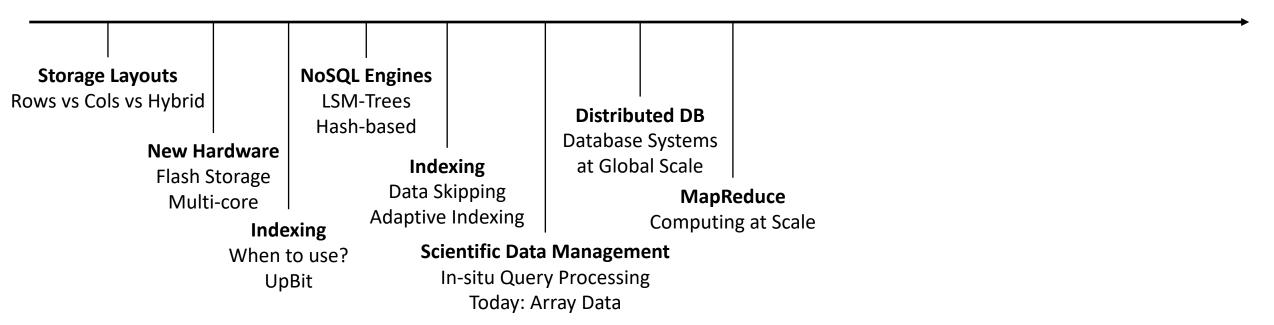
## CS 591: Data Systems & ML

Prof. Manos Athanassoulis

mathan@bu.edu

http://manos.athanassoulis.net/classes/CS591

## CS591 progress bar



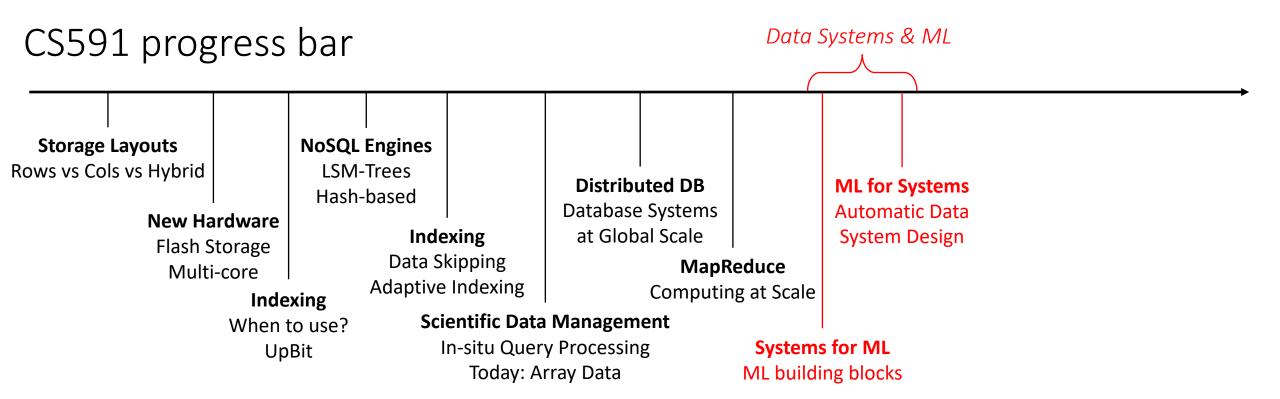


Data Systems & ML CS591 progress bar **Storage Layouts NoSQL Engines** Rows vs Cols vs Hybrid LSM-Trees **Distributed DB ML for Systems** Hash-based Database Systems **Automatic Data New Hardware** Indexing at Global Scale System Design Flash Storage Data Skipping MapReduce Multi-core **Adaptive Indexing** Computing at Scale Indexing **Scientific Data Management** When to use? In-situ Query Processing **Systems for ML** UpBit

ML building blocks

Today: Array Data



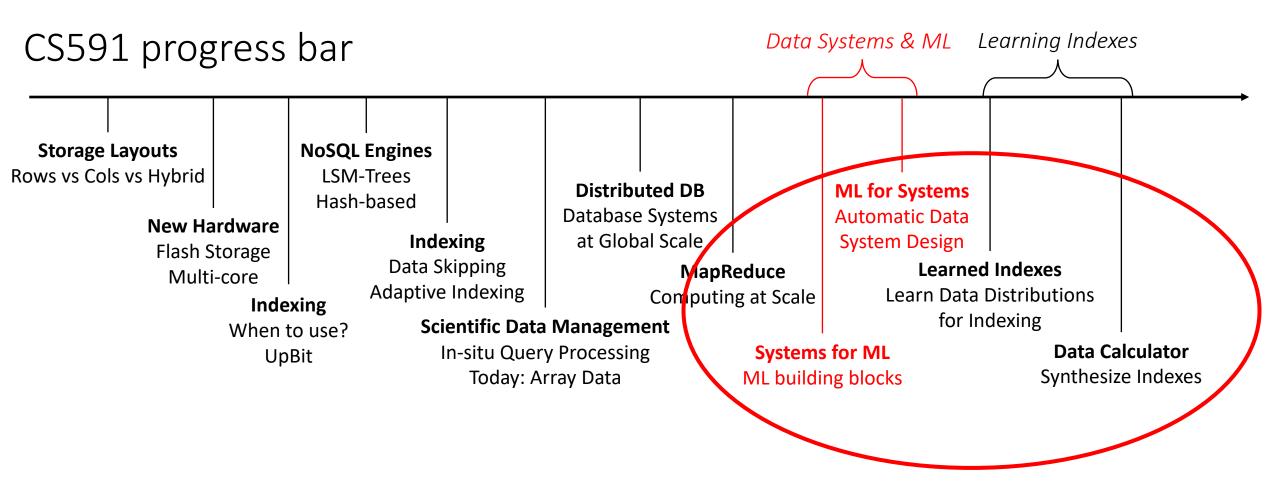


how can we efficiently support statistical queries on large datasets?

Data Systems & ML:

how can we use statistical analysis of data & queries to better tune data systems?





Added video presentations for the last four papers!



## Project Presentations



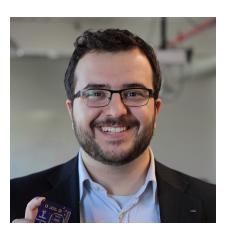
April 29<sup>th</sup>, 11:59pm: *submit project report and code* 

April 30<sup>th</sup> and May 2<sup>nd</sup>: 6 + 6 10-minute presentations (doodle link will be sent after class)

May 7<sup>th</sup>, 11:59pm (hard deadline): *send updated report (if needed)* (maximum project grade change 10%)



## Visitor: Charles Fracchia



CEO and Co-Founder, BioBright Inc.

Data Management at Scale for Scientific (focus on Biomedical) Discovery

Visits MiDAS group on May 3<sup>rd</sup>

Talk on Friday, May 3<sup>rd</sup>, at 11am (room TBA)

