

Space is big. Really big.

[We have a lot of data... Really a lot.
I'm not joking, a crazy amount.] $\times 1k$

2 computational frameworks: - map reduce (parallel)
- streaming (lost data)

Techniques:

- Partitioning
- Coresets (\sim sampling)
- Sketches (statistics)
- Locality sensitive hashing (similarity search)
↓
LSH

SLIDES!

The exam probably has one question about the homeworks

silvestri@dei.unipd.it

tutors (and prof) : bde-course@dei.unipd.it



Homework 2024