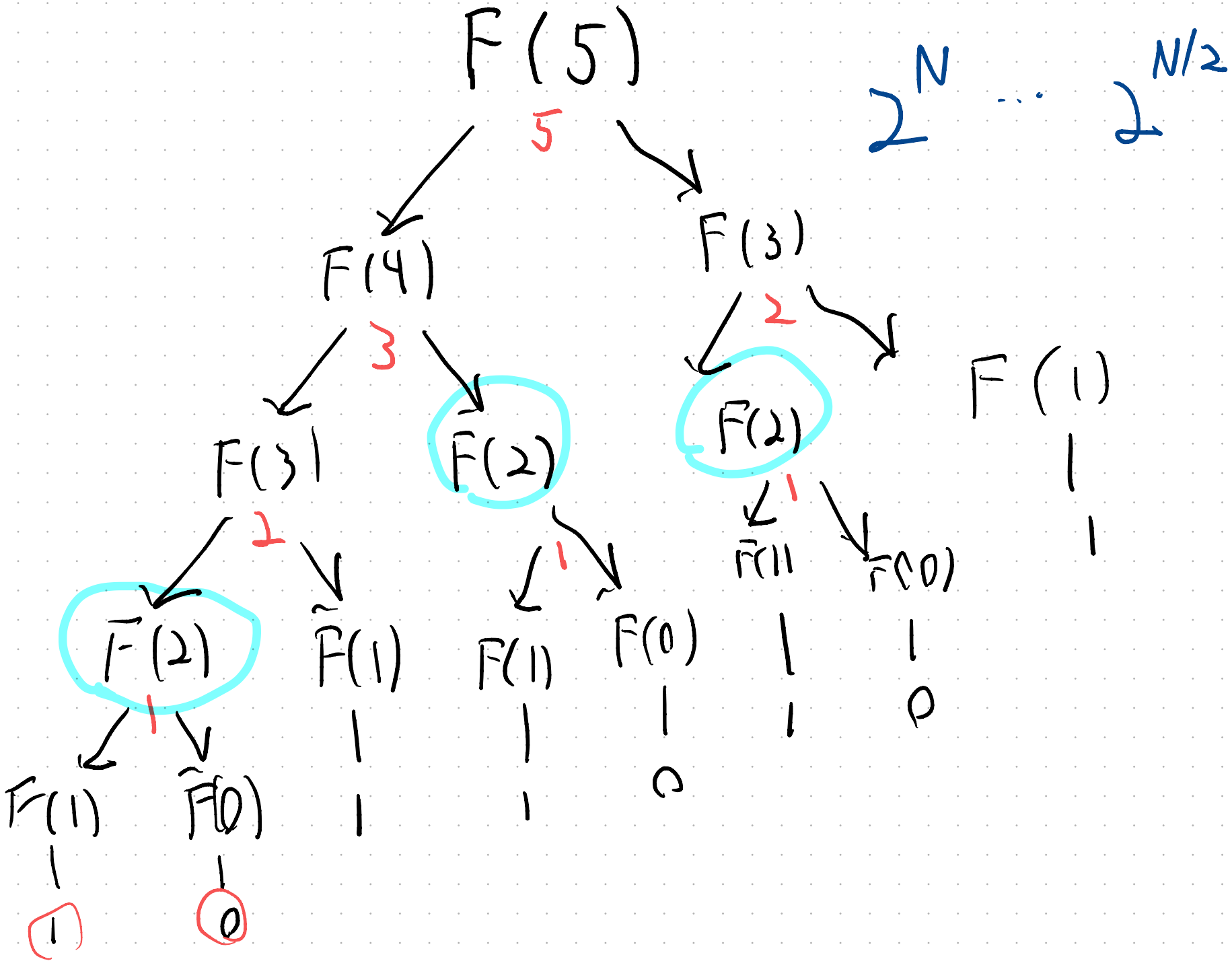
Algorithms For Big Data

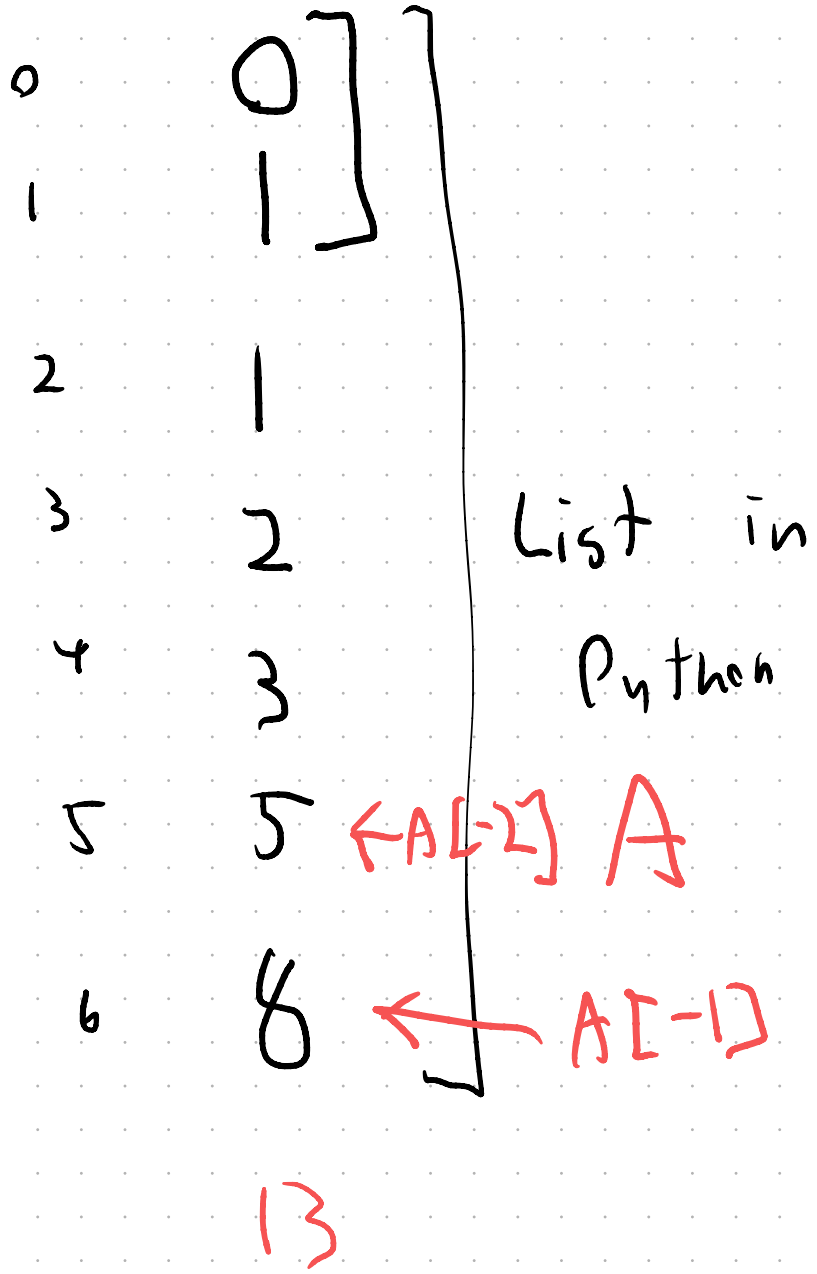
Prof John Iacono

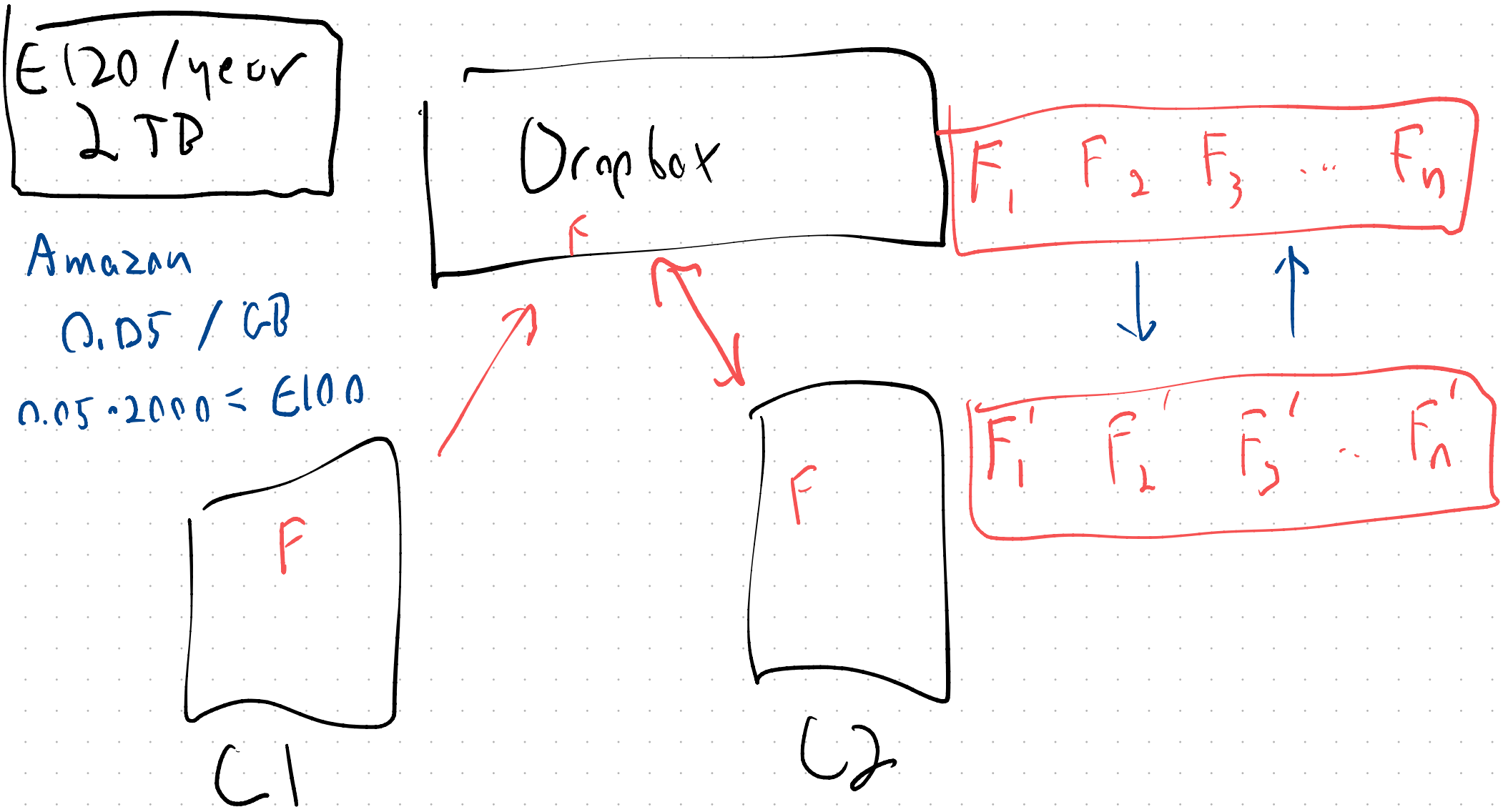
Starts at 8:05

0
1
1
2
3
5
6 ...

$$F(N) = \begin{cases} 0 & \text{if } N \text{ is } 0 \\ 1 & \text{if } N \text{ i } 1 \\ F(N-1)+F(N-2) & \text{otherwise} \end{cases}$$


F(5)
5
F(4)
3
F(3)
2
F(3)
2
F(2)
F(2)
F(2)
1
F(1)
F(1)
F(1)
F(0)
F(1)
F(0)
1
1
1
1
1
0
1
0
2^N ... $2^{N/2}$


0
1
2
3
4
5
6
0
1
1
2
3
5
8
13
List in Python
A[-2]
A
A[-1]


€120 / year
2 TB
Dropbox
F
F_1 F_2 F_3 ... F_n
Amazon
0.05 / GB
0.05 · 2000 = €100
F
C1
F
C2
F_1' F_2' F_3' .. F_n'

Data1 = [, , , ,]

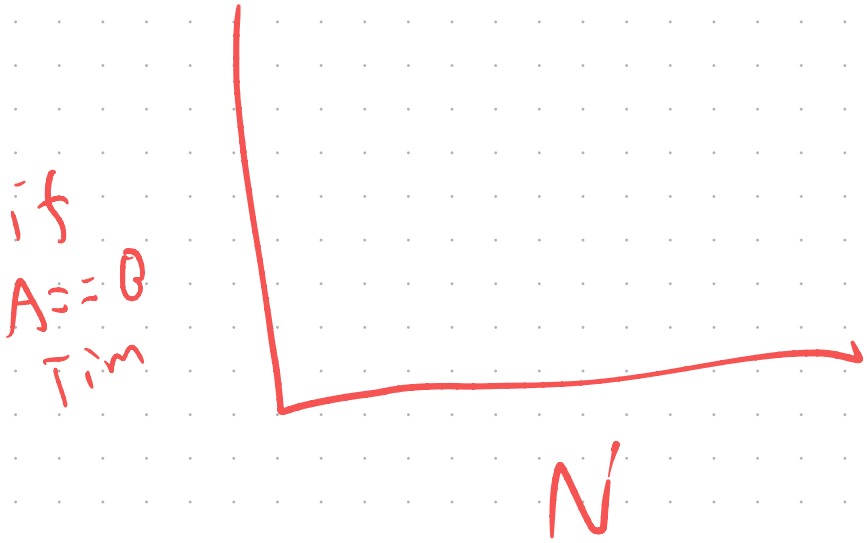
Data2 [, , , ,]

N

A = [, , , , ,]

B = [, , , , ,]

if A == B;



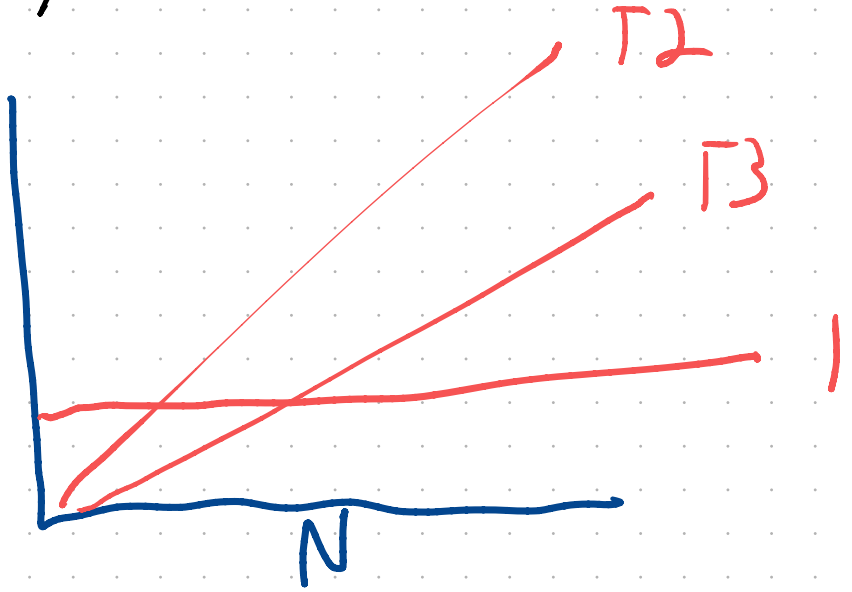
Test 1: Two completely different lists of size N

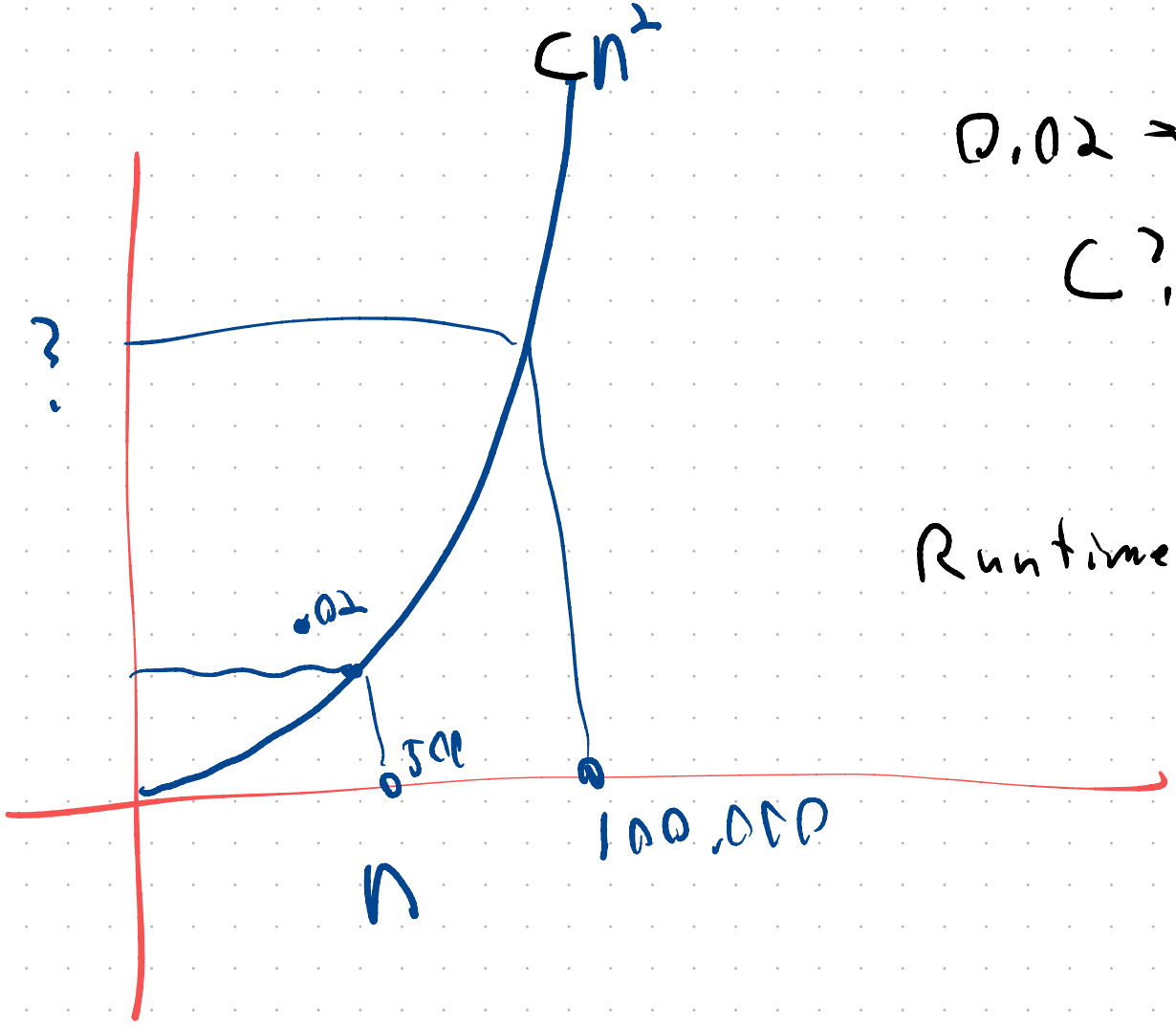
O(1)

Constant

Test 2: Two identical lists

Test 3: Two lists, first Half the same.





$0.02 = c\,500^2$

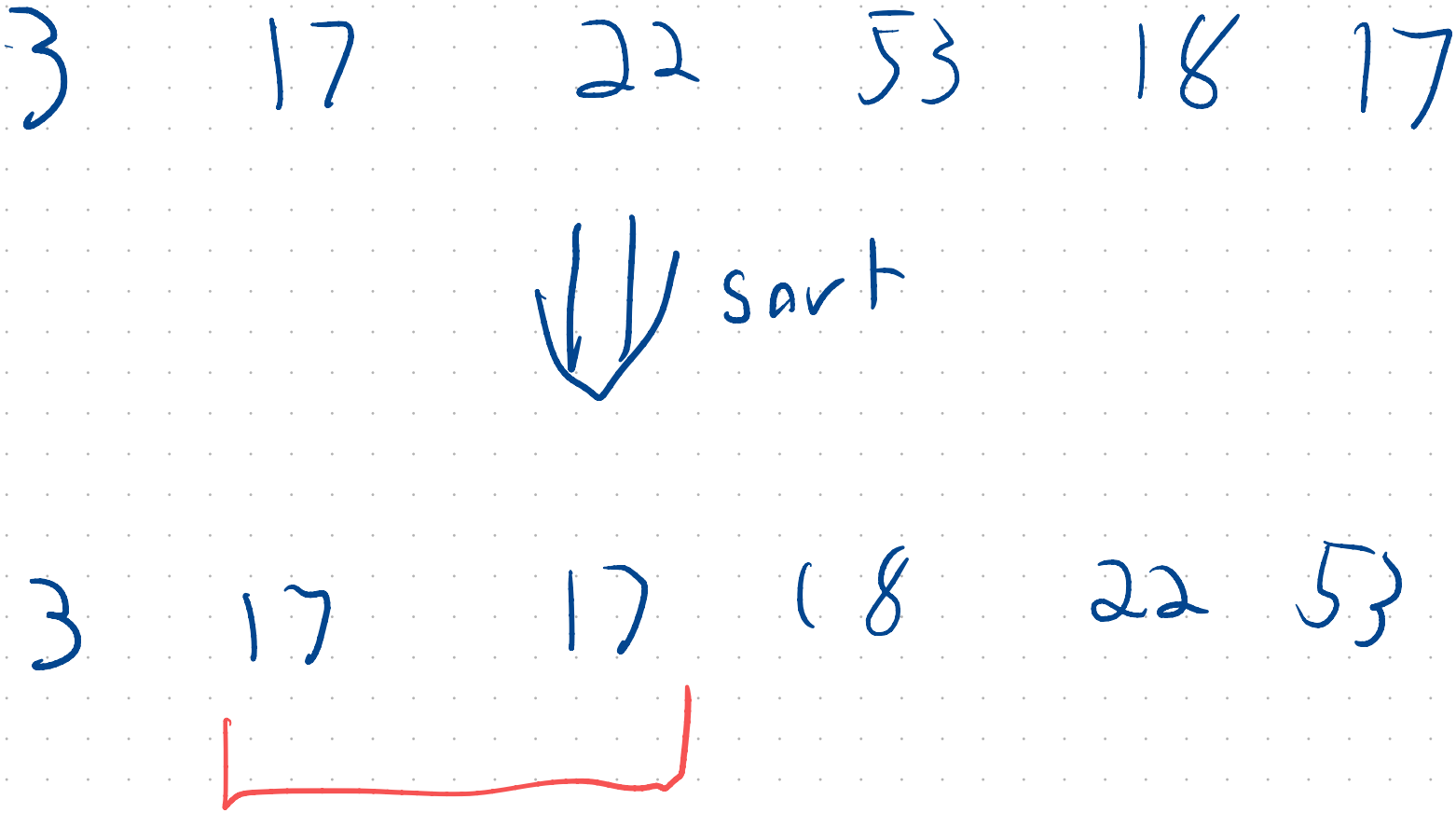
c? $\quad c = \frac{0.02}{500^2}$

Runtime on 100,000

$\frac{0.02}{500^2} \cdot 100{,}000^2$

$= 800$ seconds

12 minutes


3 17 22 53 18 17
sort
3 17 17 18 22 53

Hash Function

$h(x) \rightarrow$ Number in a certain range

range r

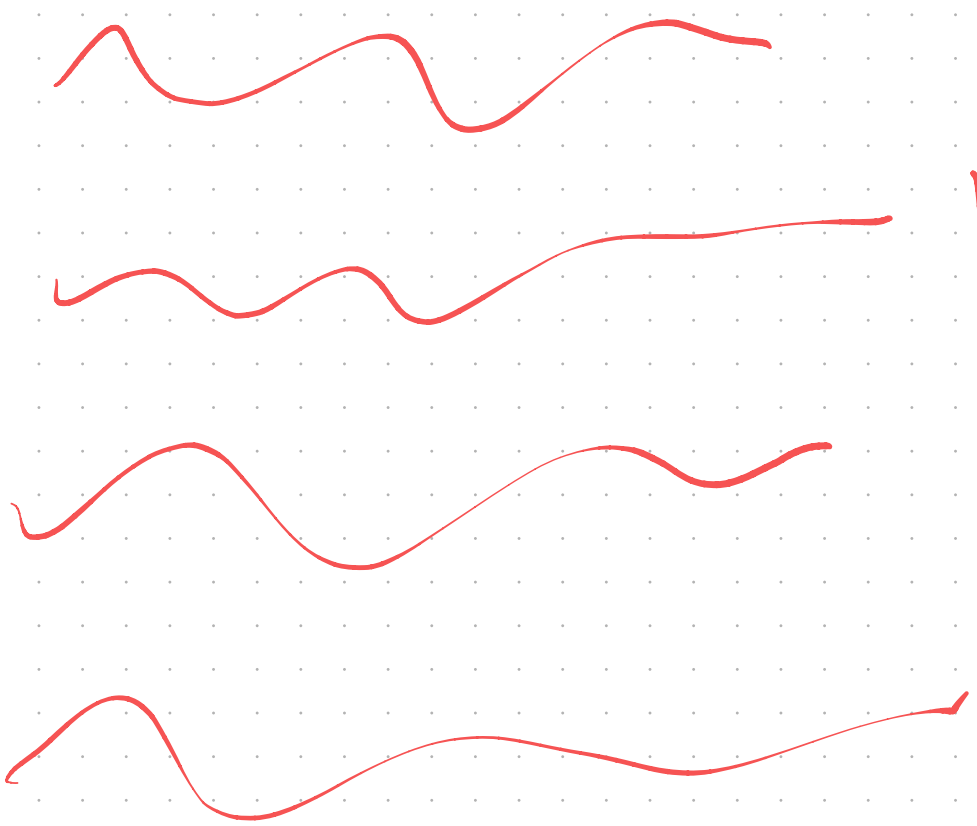
$0 \ldots r-1$

$h(x) = h(x)$

if $x \neq y$ $h(x) \neq h(y)$ (usually)

probability $< \frac{1}{r}$

h



"this" → 3

"that" → 3

"those" → 2

2 3 7