Adaptive Data Skipping in Main-Memory Systems
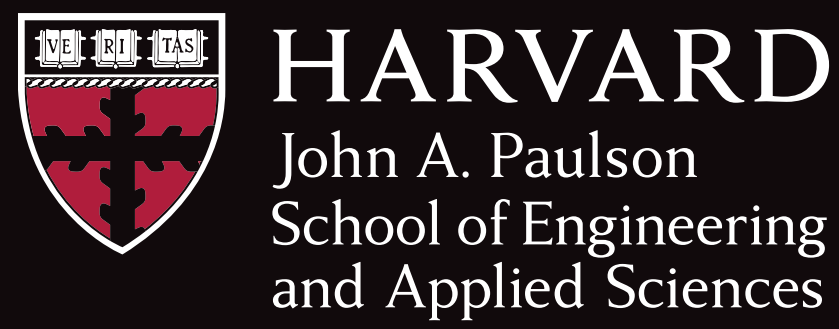
Wilson Qin, Stratos Idreos
Harvard University
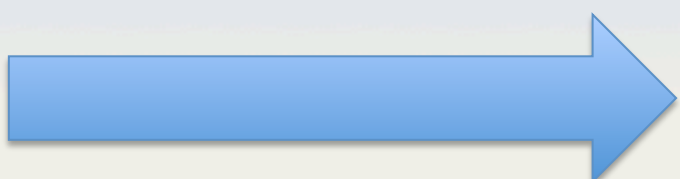wilsonqin@college.harvard.edu, stratos@seas.harvard.edu

DASlab

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

## Fast Data Access Patterns *enable* Fast Knowledge Extraction

Lightweight indexes such as **Zonemaps** logically segment data into contiguous zones, keeping metadata for each zone. Metadata informs database operations, indicating which data zones are non-pertinent to a query - this is called **data skipping**.

At scale, data skipping enables faster queries without the overhead of maintaining full indexes on the data. Faster data filtering operations during scans, ultimately means a faster knowledge discovery cycle for users.
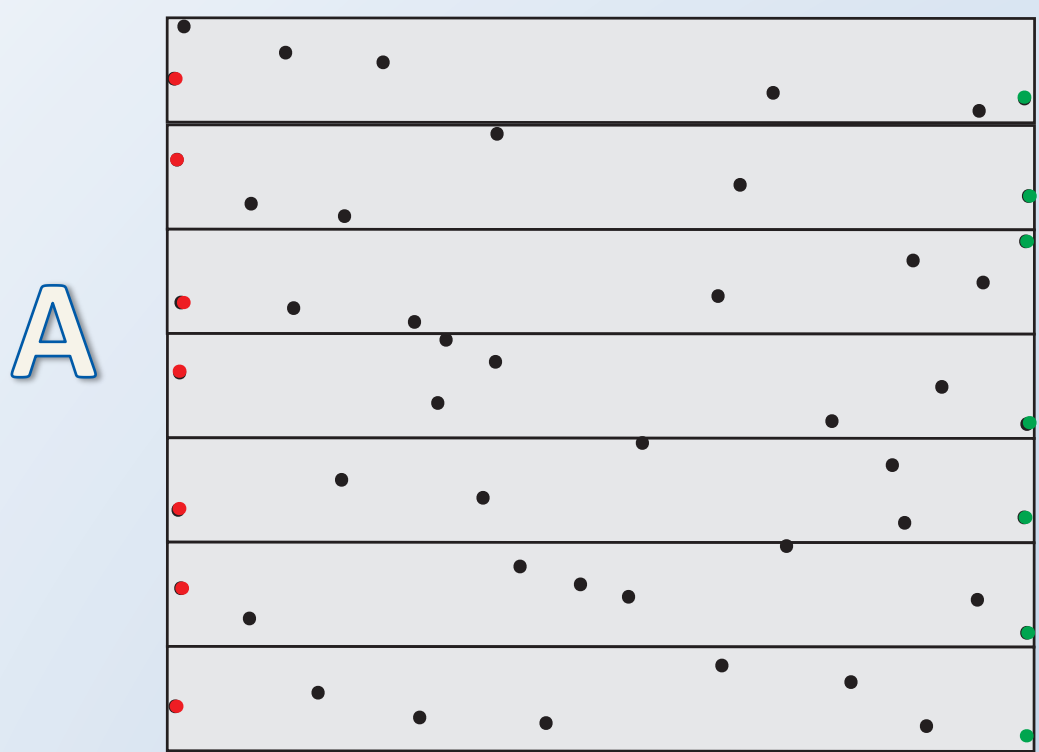

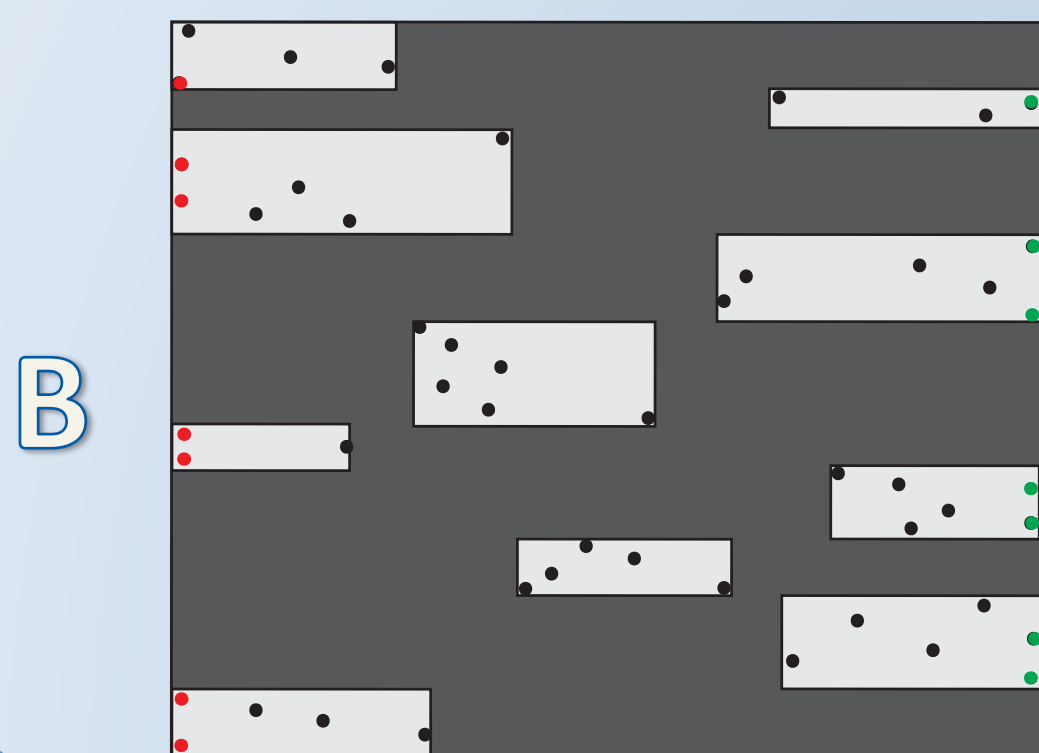Zonemaps keep minima, maxima statistics on fixed sized zones.

Metadata (Min, Max)   Data Values

50
100
90
55
75
65
70
75

50
100
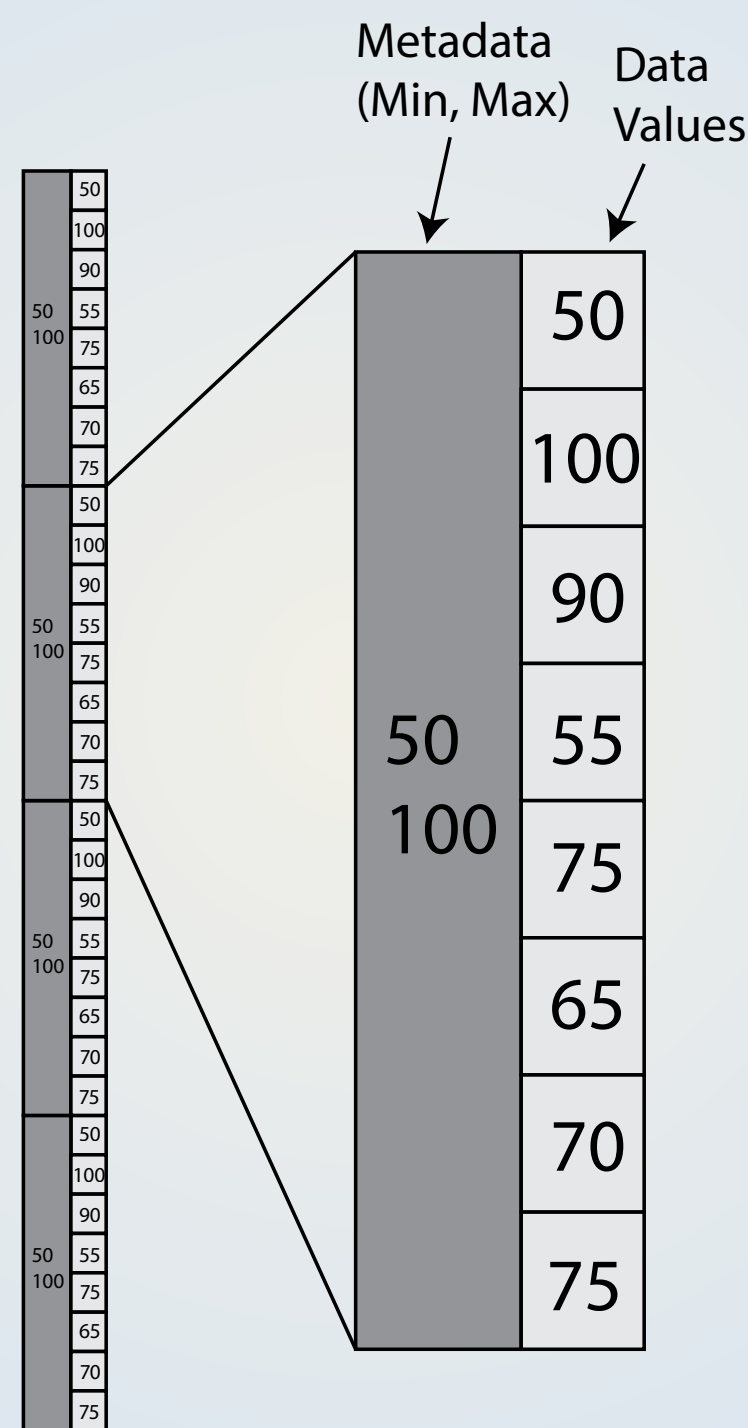
Target Column with Zonemap

Individual Zone

However, for certain datasets and query workloads, a zonemap layout can even be *less effective* than a scan.

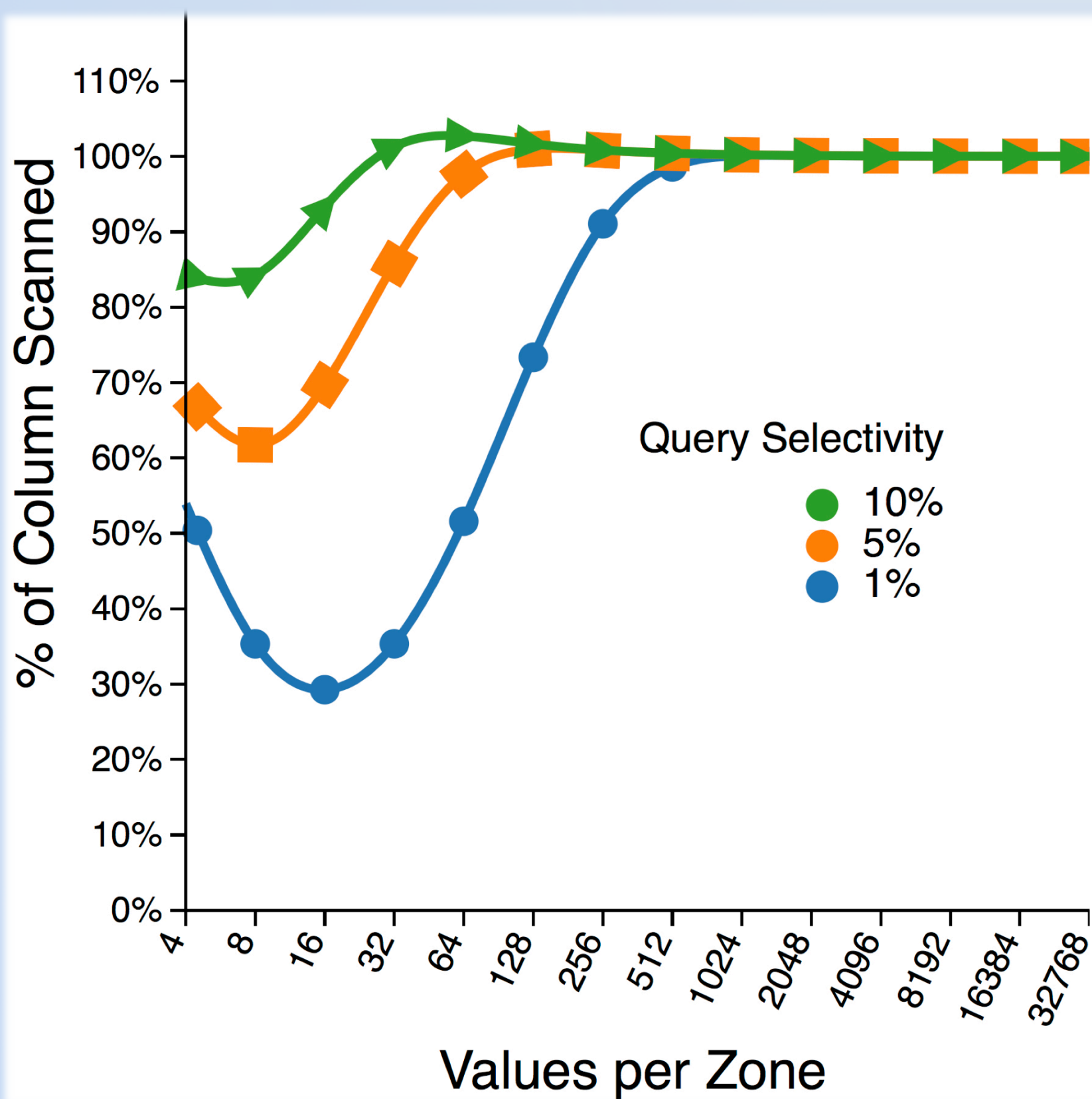**Are there lightweight methods to avoid zones with useless metadata (A), and create useful zones (B)?**

A — Zones with wide domains, and useless metadata.

B — Zones with tighter domains, and useful metadata.

No Static Layout Fits All

Query Selectivity
- 🟢 10%
- 🟠 5%
- 🔵 1%

y-axis: % of Column Scanned (0% – 110%)
*x-axis:* Values per Zone (4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768)

Variation in zonemap effectiveness by layout indicates the potential for *dynamic* structures beyond static layouts that can adapt to a changing query workload over time.

We introduce **Adaptive Data Skipping** as a robust data skipping framework applicable to a much wider array of data distributions and ad-hoc query workloads than standard data skipping methods.
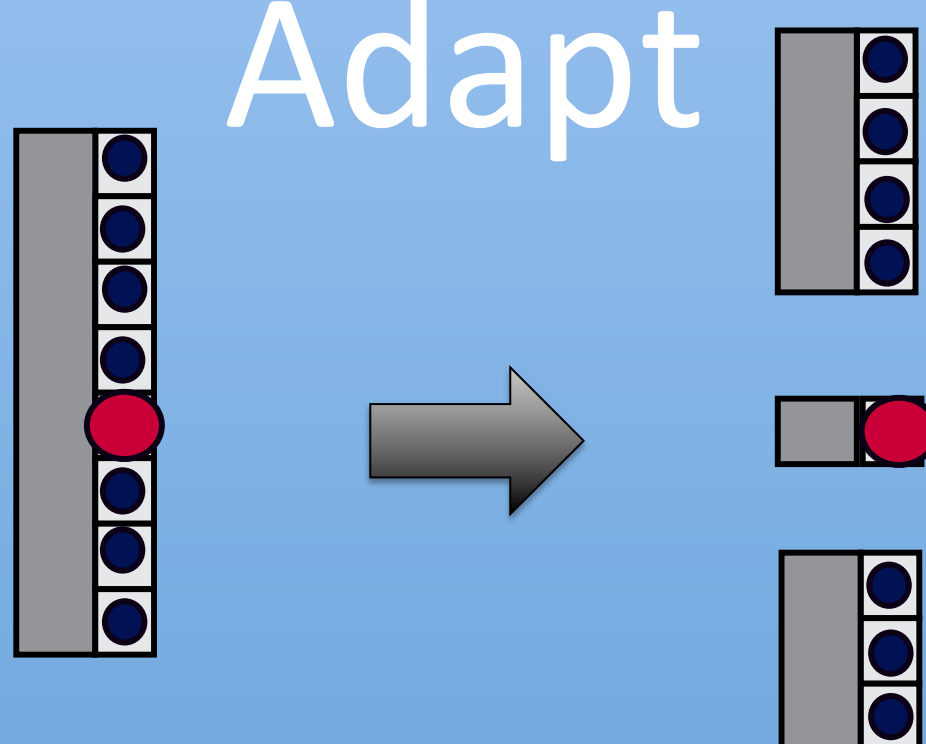
Incremental   On-Demand

Opportunistically improve relevant zones by gradually restructuring them during query time – as underlying data is being scanned.
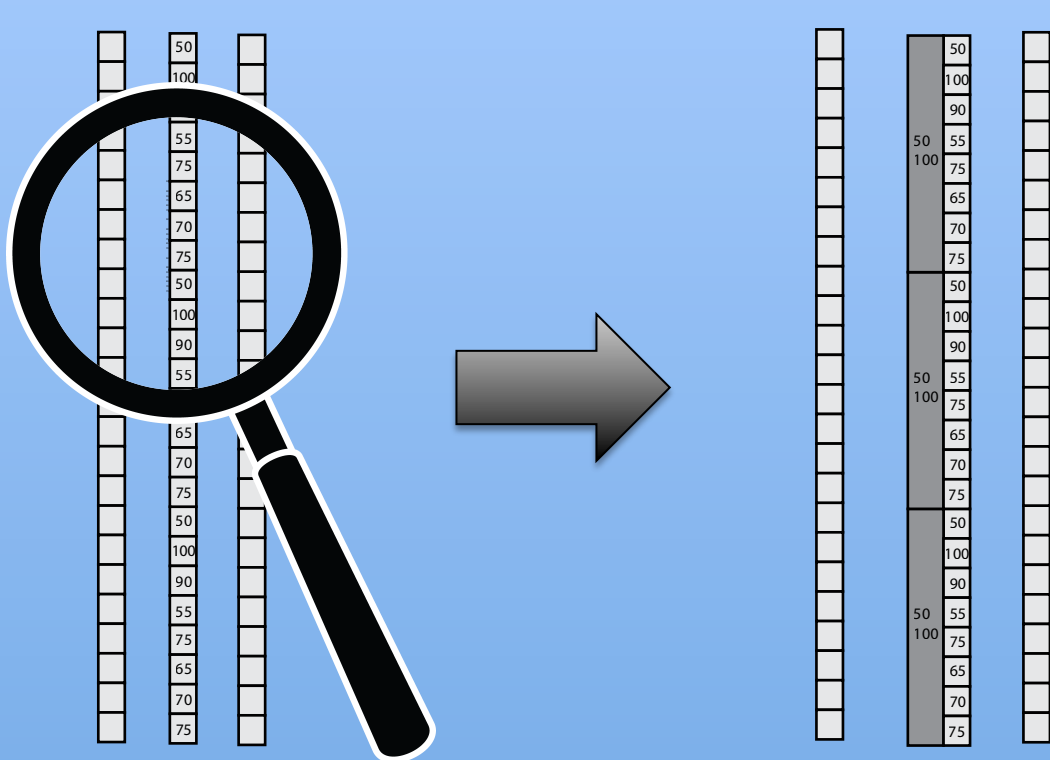
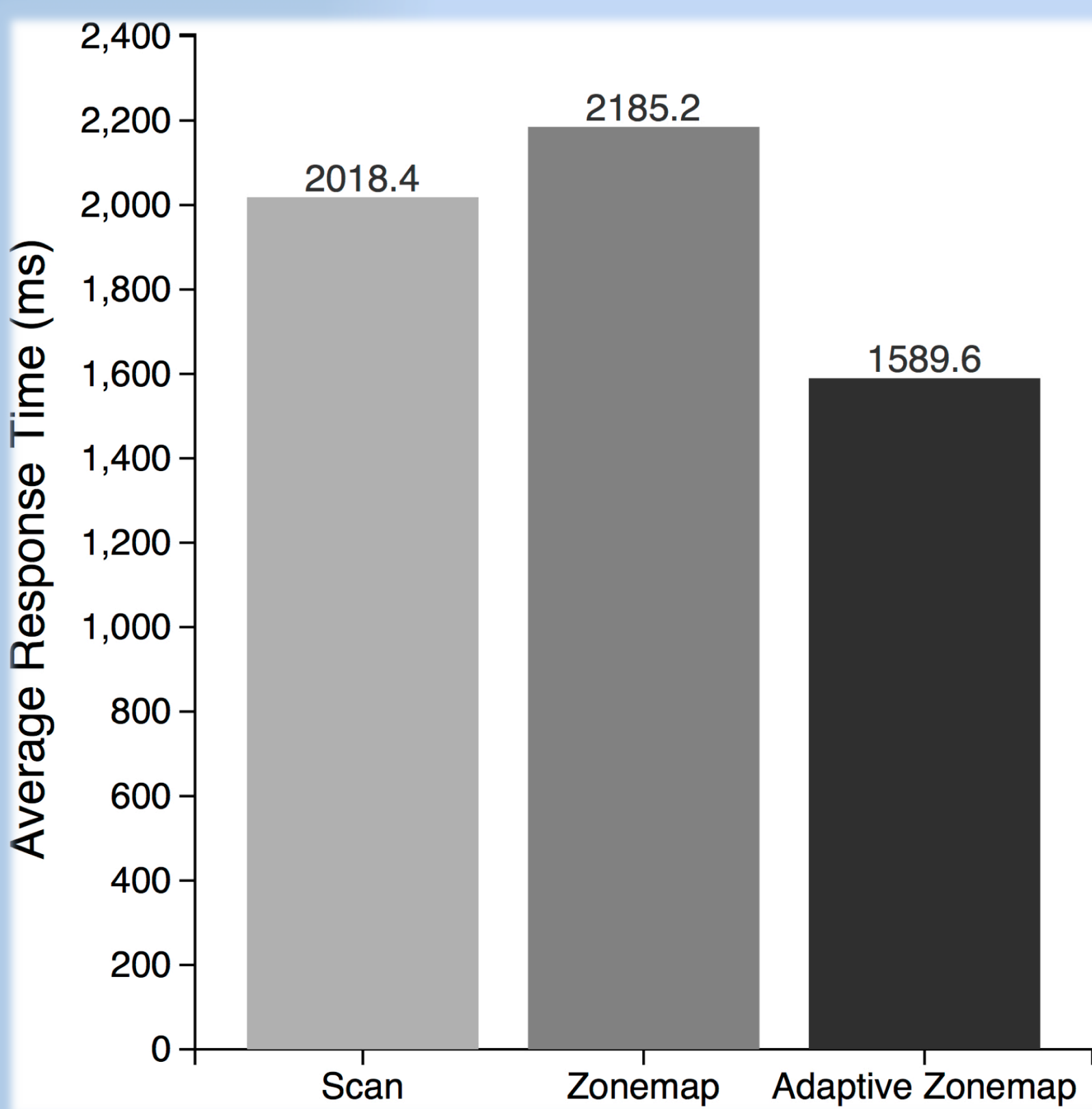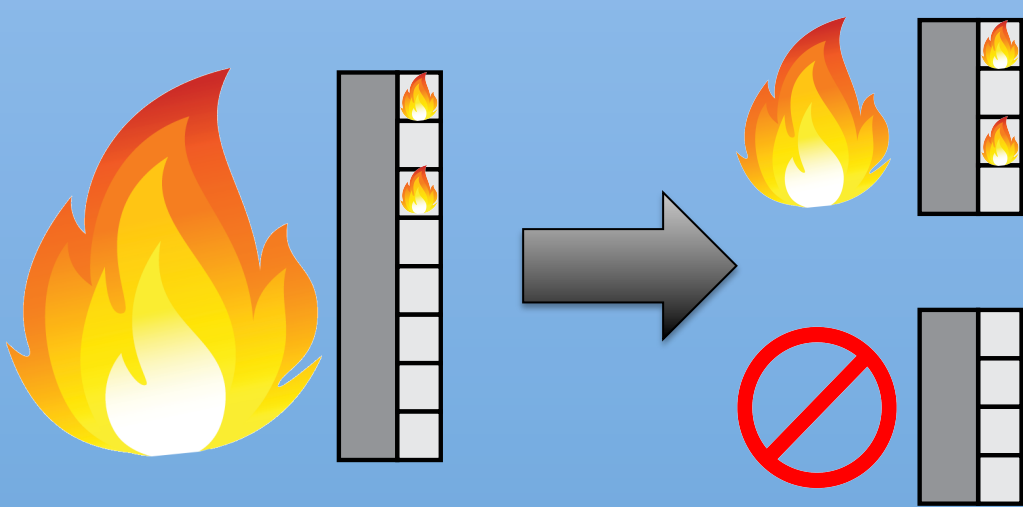## Adapt   Increase Skipping

---

Average Response Time (ms)

Scan	Zonemap	Adaptive Zonemap
2018.4	2185.2	1589.6

1.4X speed up Adaptive Zonemap prototype: early results for select operations.

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

daslab.seas.harvard.edu

DASlab