

NoDB: Efficient Query Processing on Raw Data Files

Ioannis Alagiannis[‡]

Renata Borovica[‡]

Miguel Branco[‡]

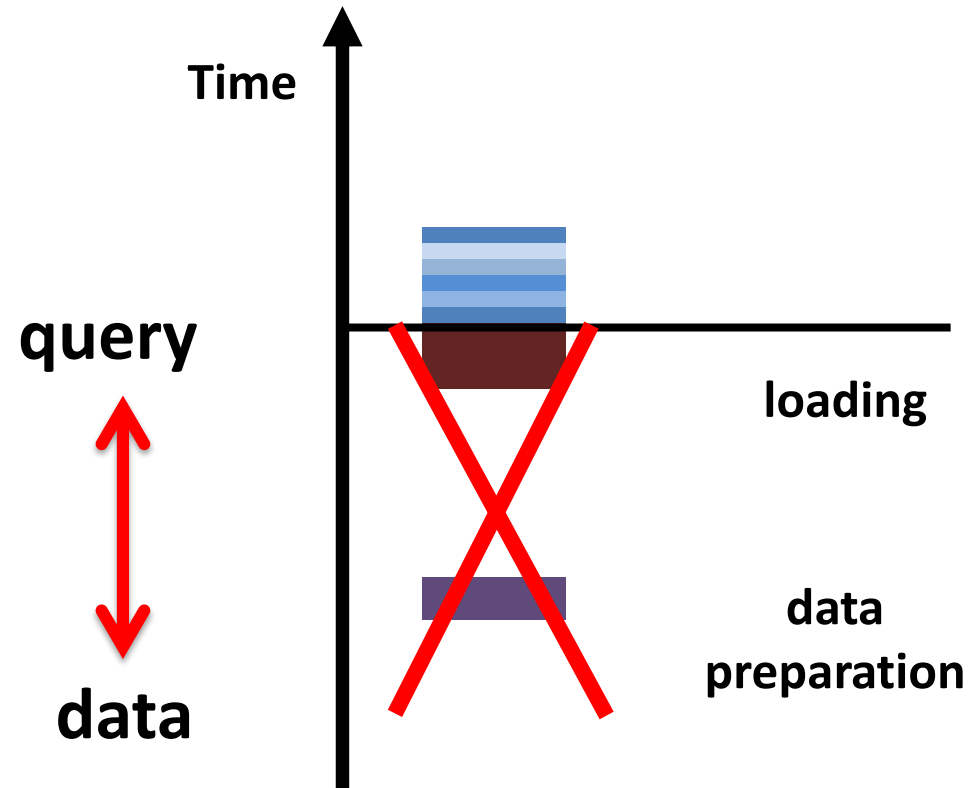
Stratos Idreos^{}*

Anastasia Ailamaki[‡]

[‡]École Polytechnique
Fédérale de Lausanne

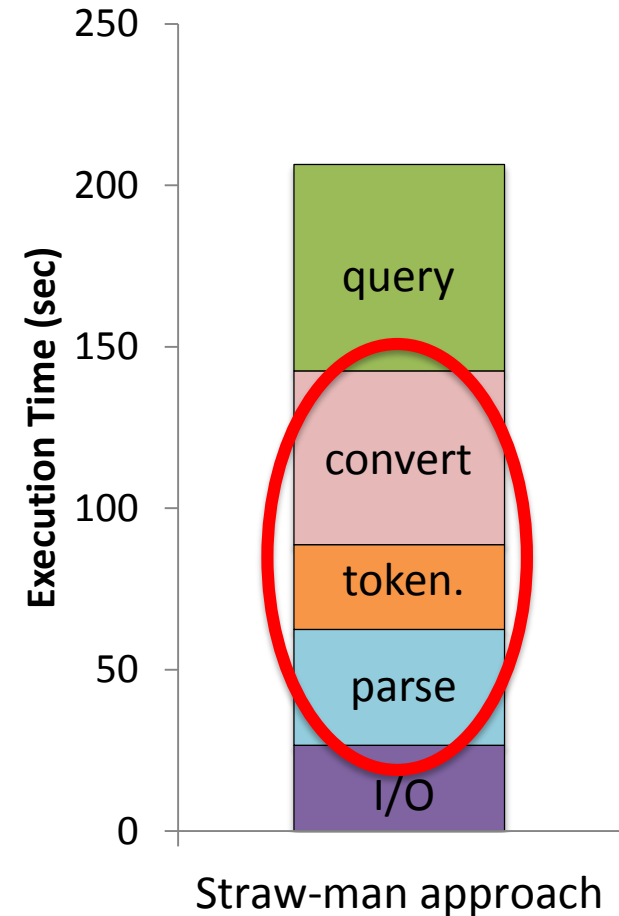
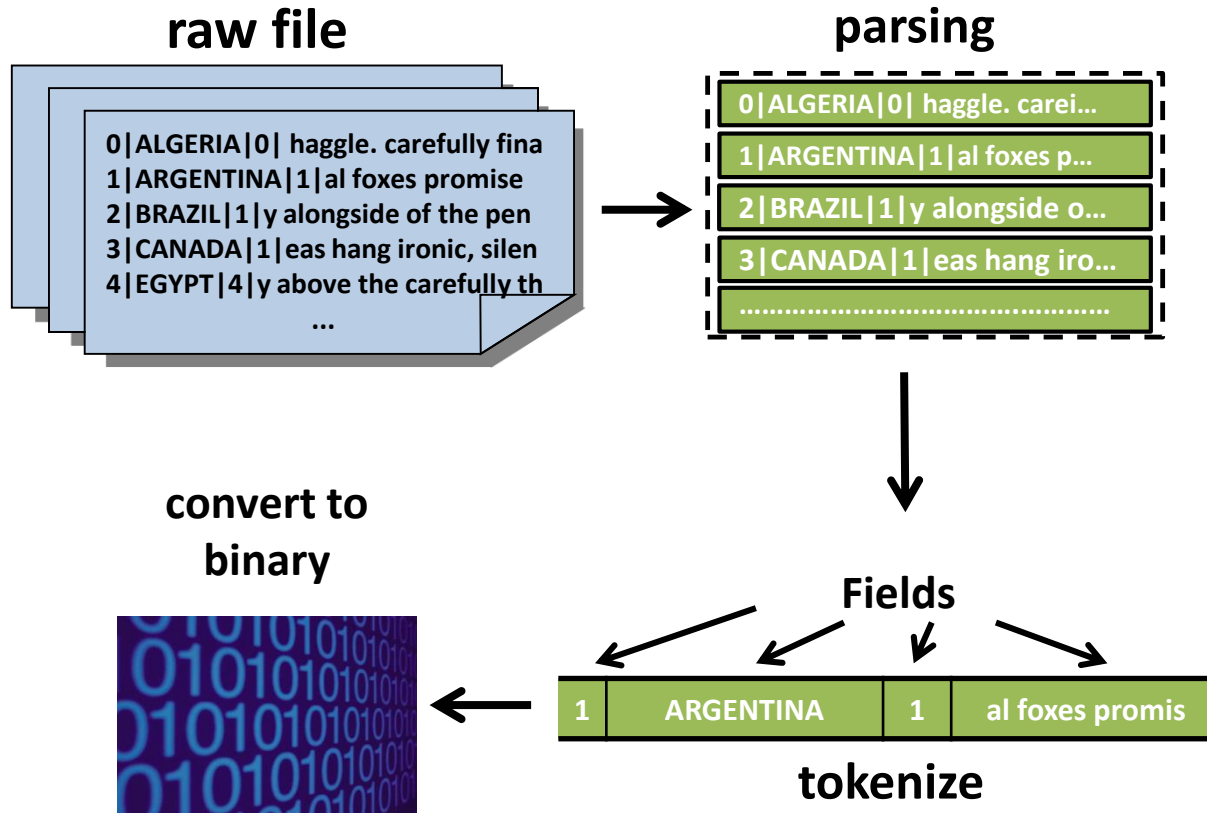
^{*}CWI, Amsterdam

From data to results



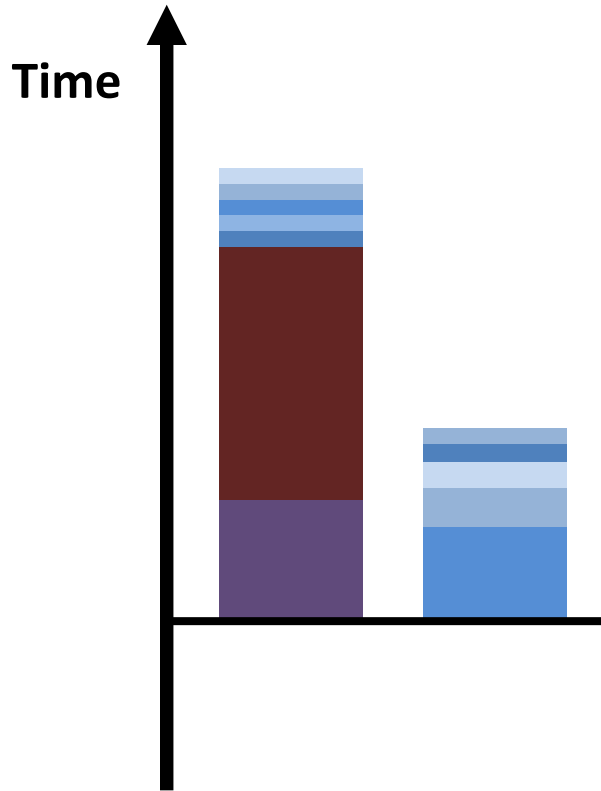
Reduce data-to-query time

Querying data *in situ*



Straw-man approach is slow

NoDB philosophy



No data loading

Instant gateway to data

Raw files first-class citizen

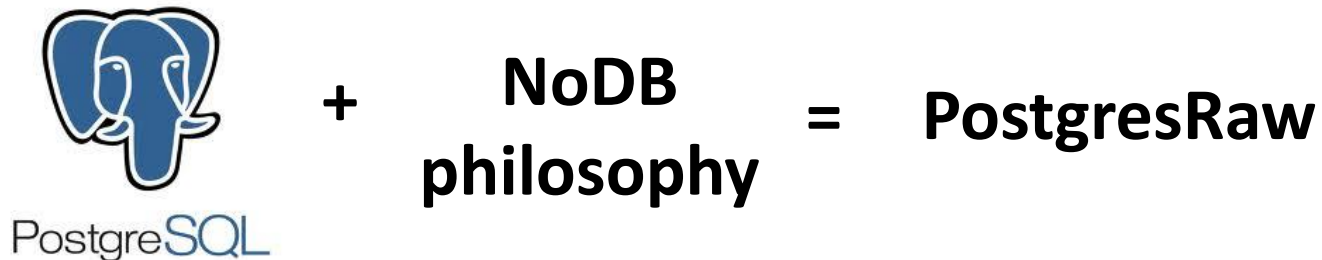
Driven by the workload

Adaptive in situ DBMS

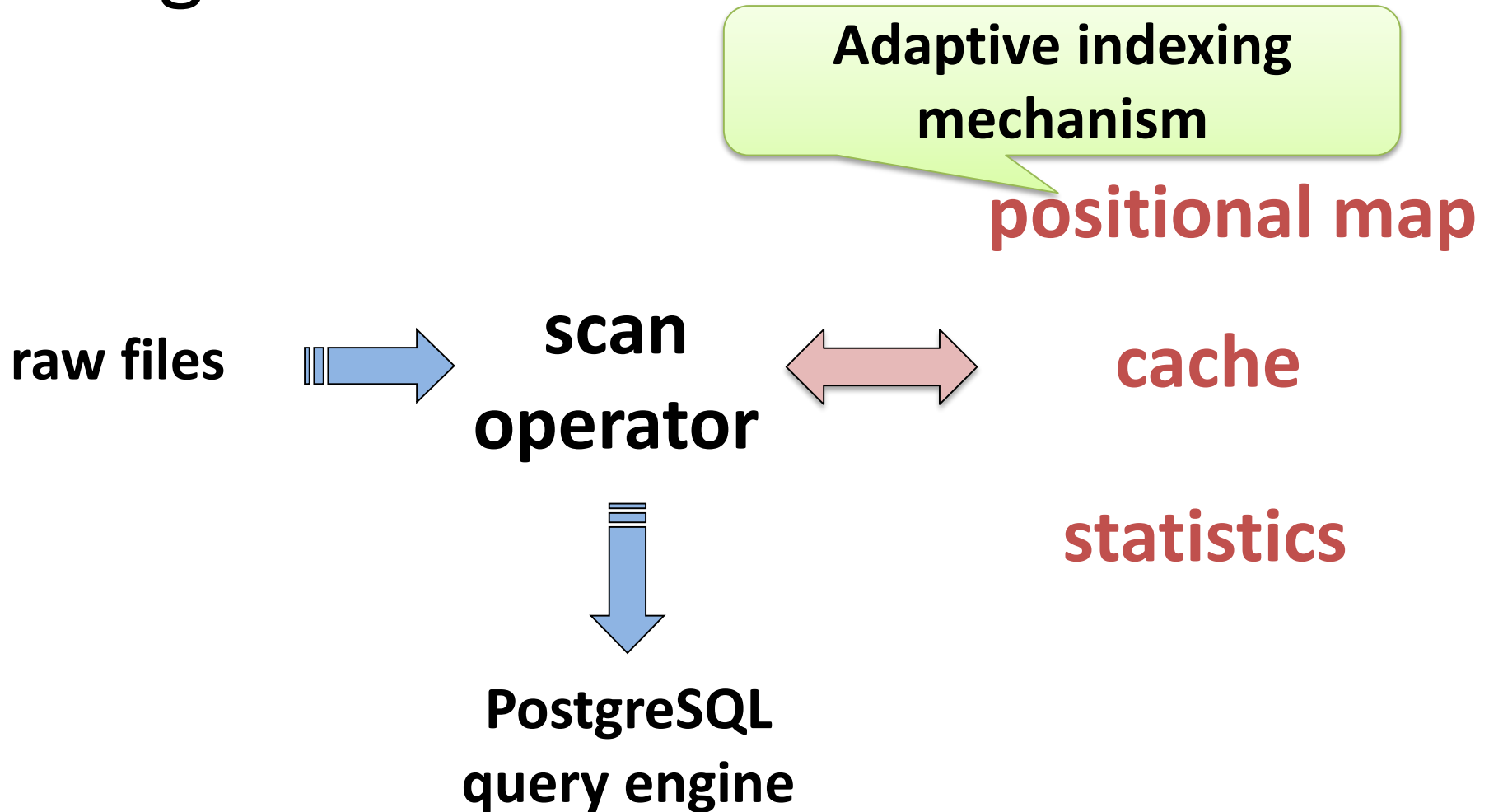
NoDB in practice

Efficient *in situ* querying

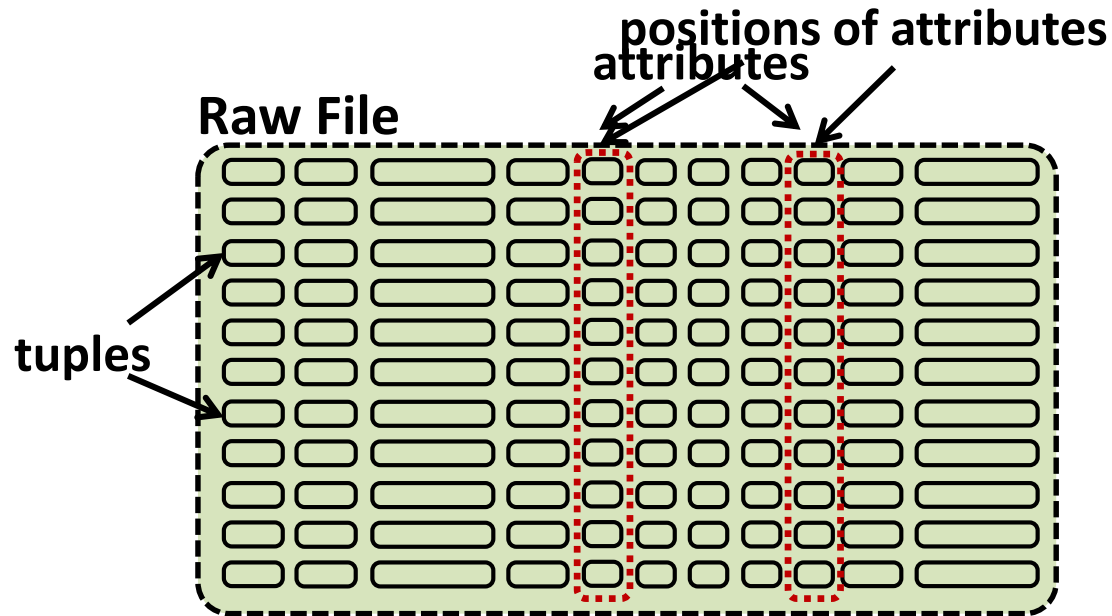
Minimal changes to the query engine



PostgresRaw



Positional map

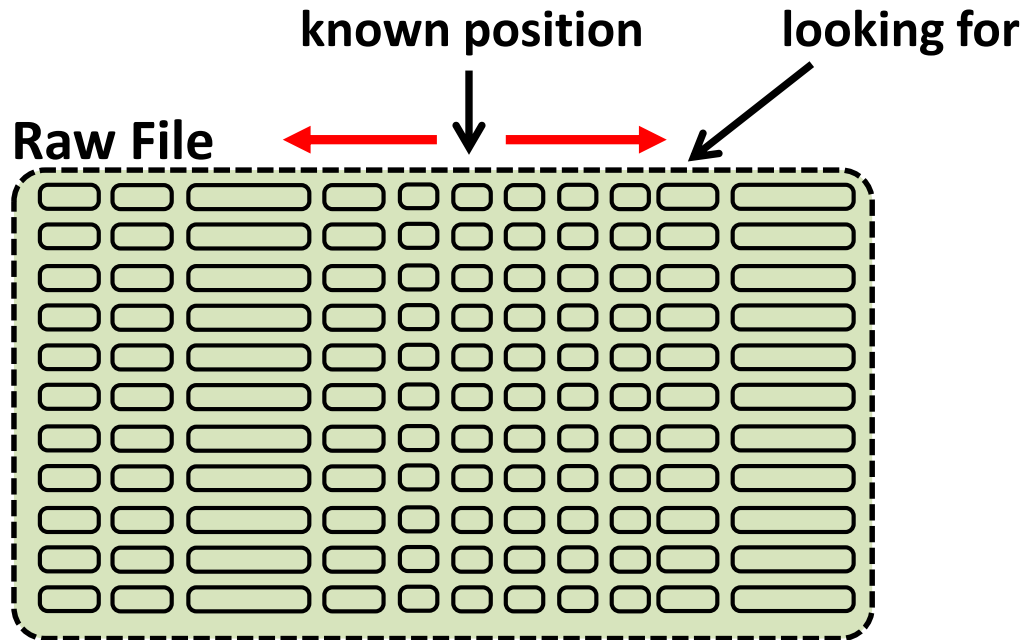


Reduce parsing

Reduce tokenizing

Created on-the-fly

Positional map

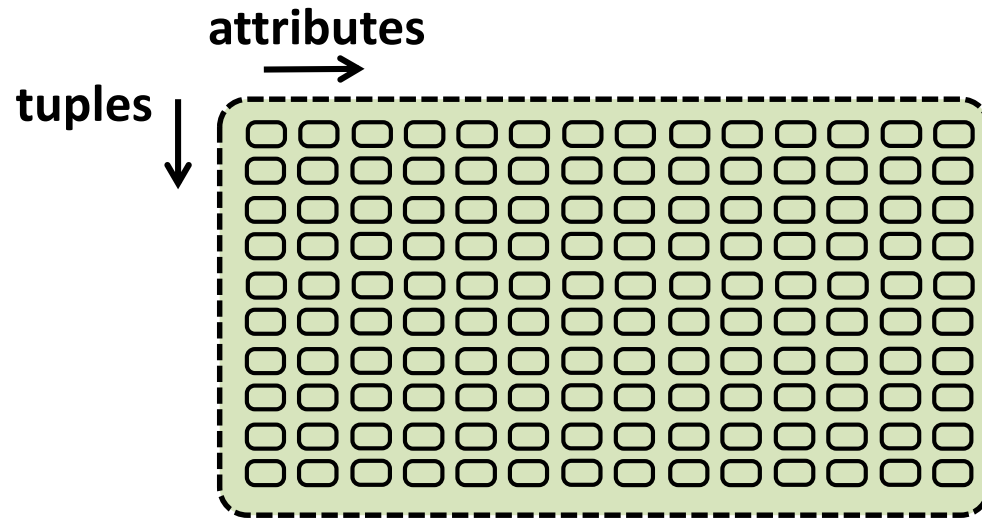


Reduce parsing

Reduce tokenizing

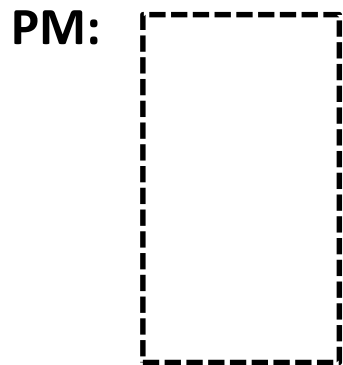
Created on-the-fly

Positional map in action

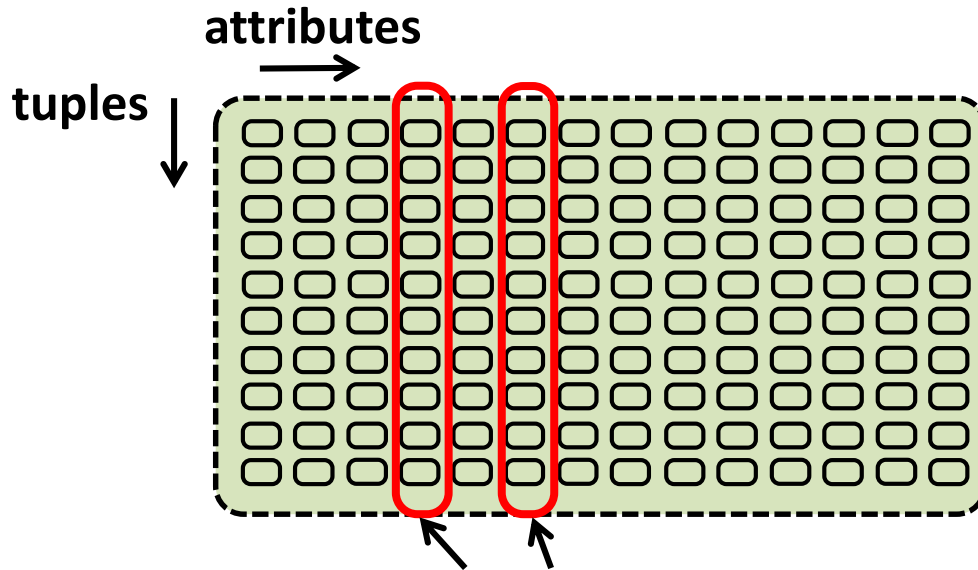


1. Positional map is empty

Indexed attributes:



Positional map in action



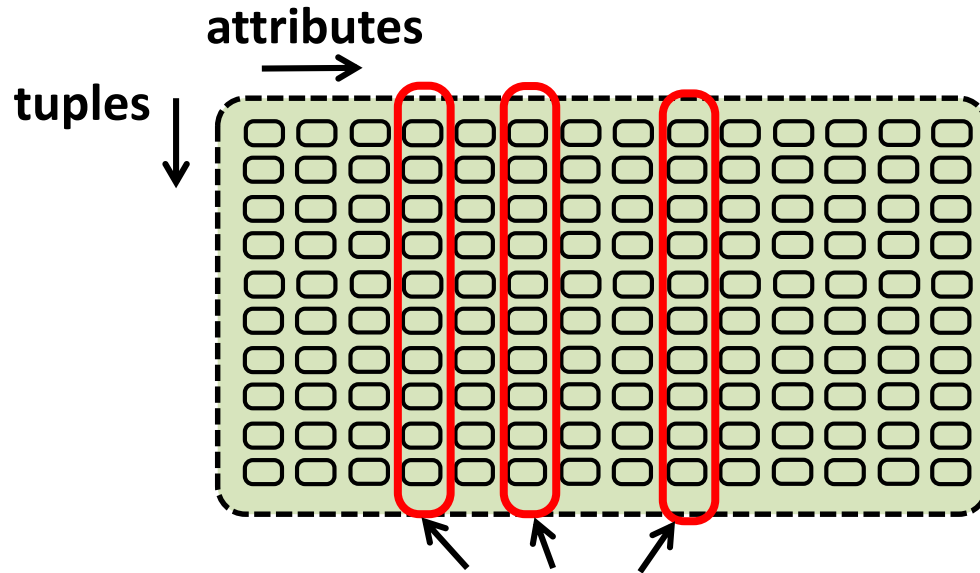
Indexed attributes: a4, a6

PM:

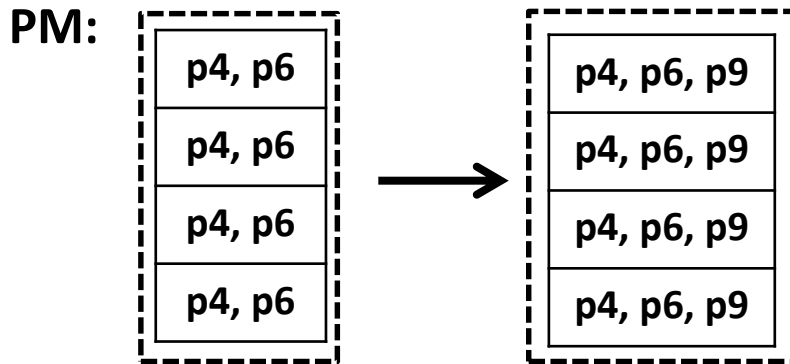
p4, p6
p4, p6
p4, p6
p4, p6

1. Positional map is empty
2. Q1 accesses *a4* and *a6*

Positional map in action



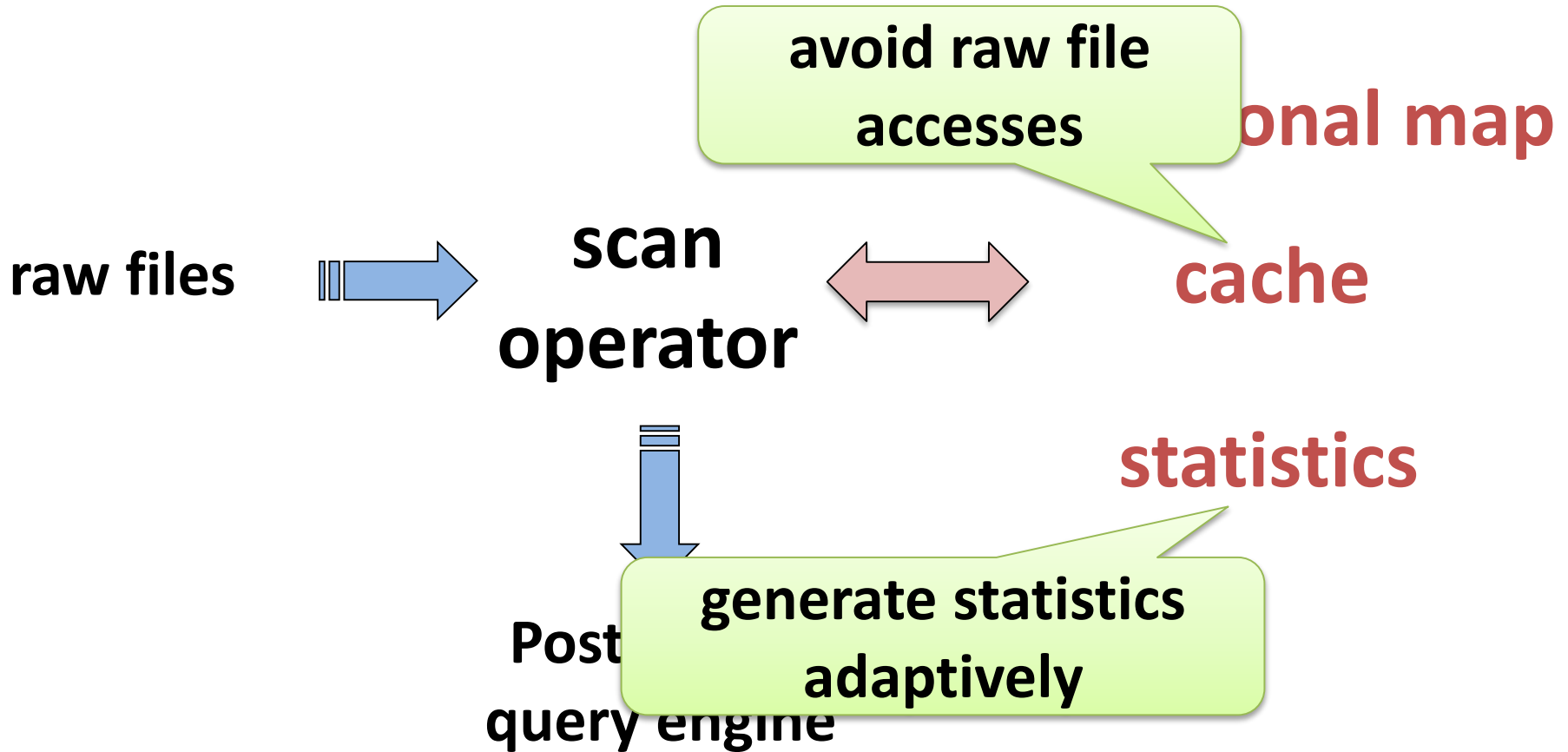
Indexed attributes: a_4 , a_6 , a_9



1. Positional map is empty
2. Q1 accesses a_4 and a_6
3. Q2 accesses a_4 and a_9

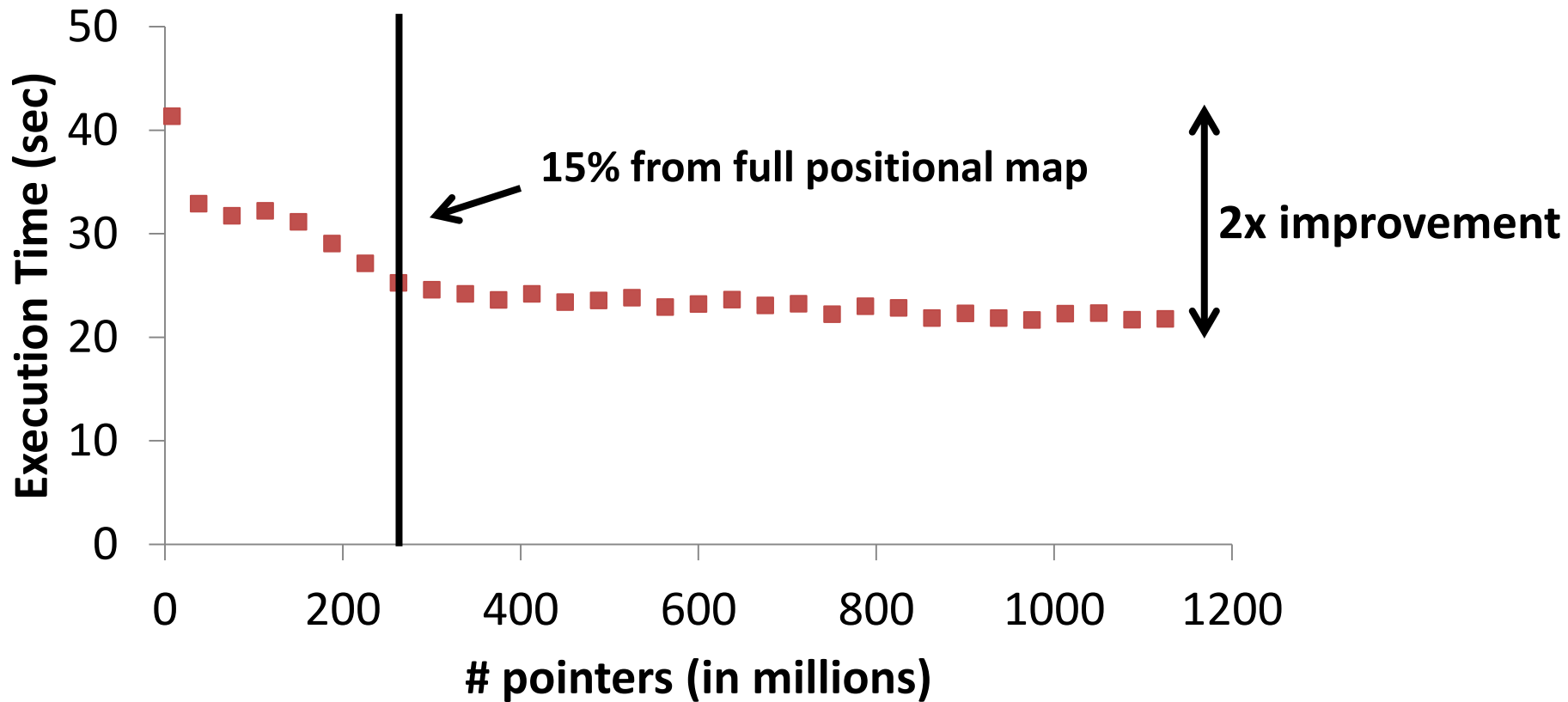
Make raw data access progressively cheaper

PostgresRaw



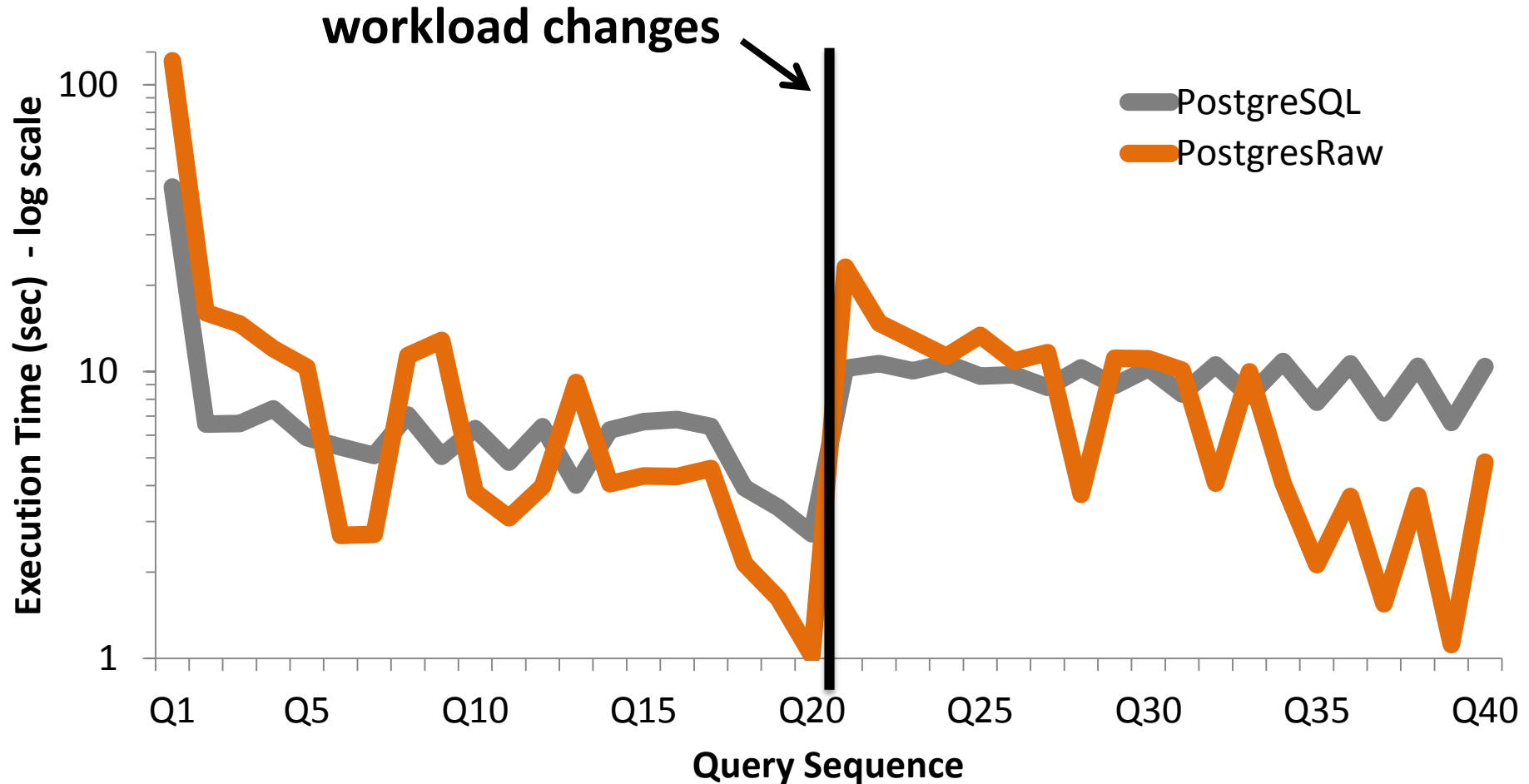
Impact of positional map

Random queries on 10 attributes
Vary storage capacity (15MB-2GB)



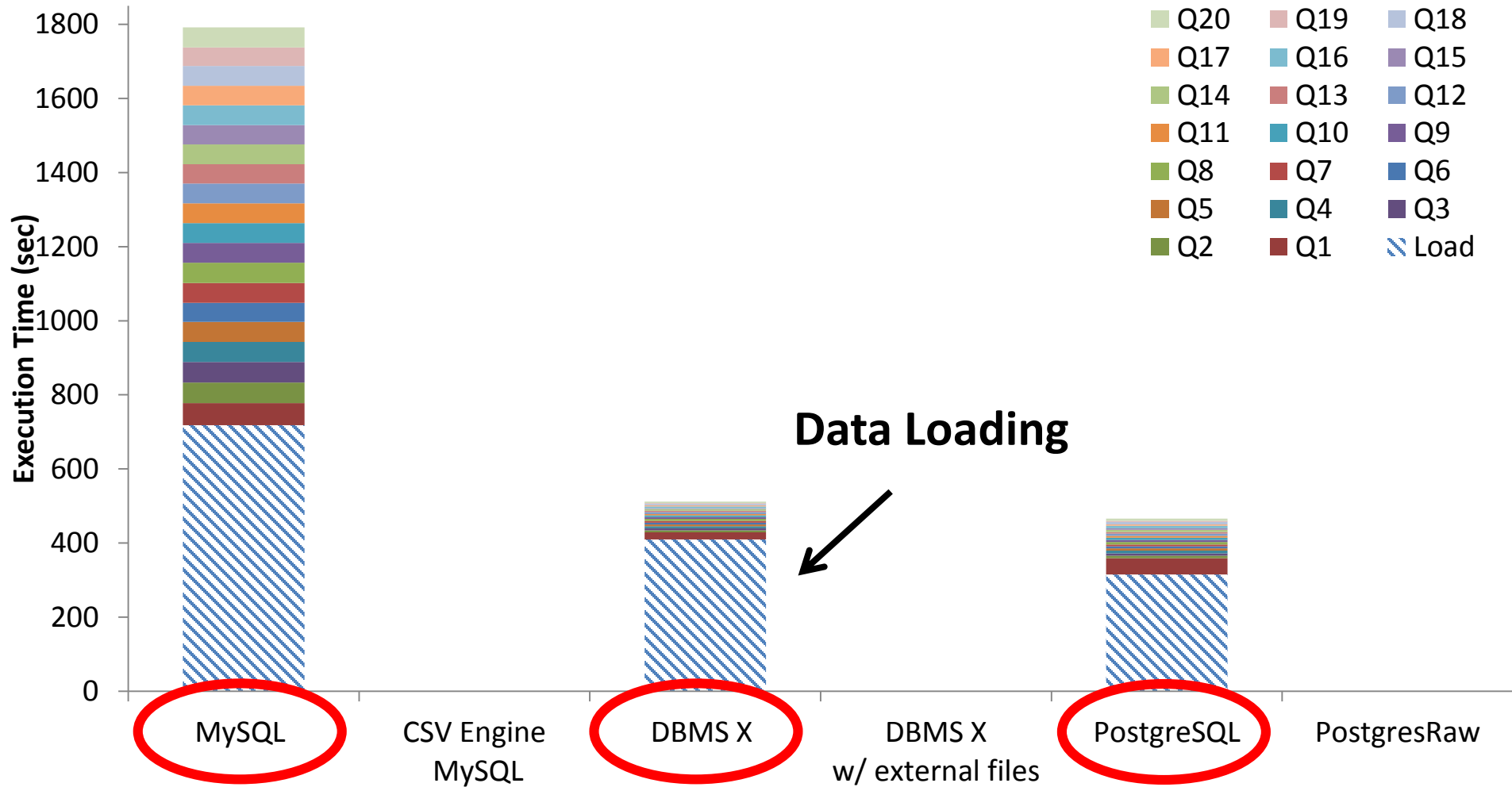
No need for the whole positional map

Adapting to changes

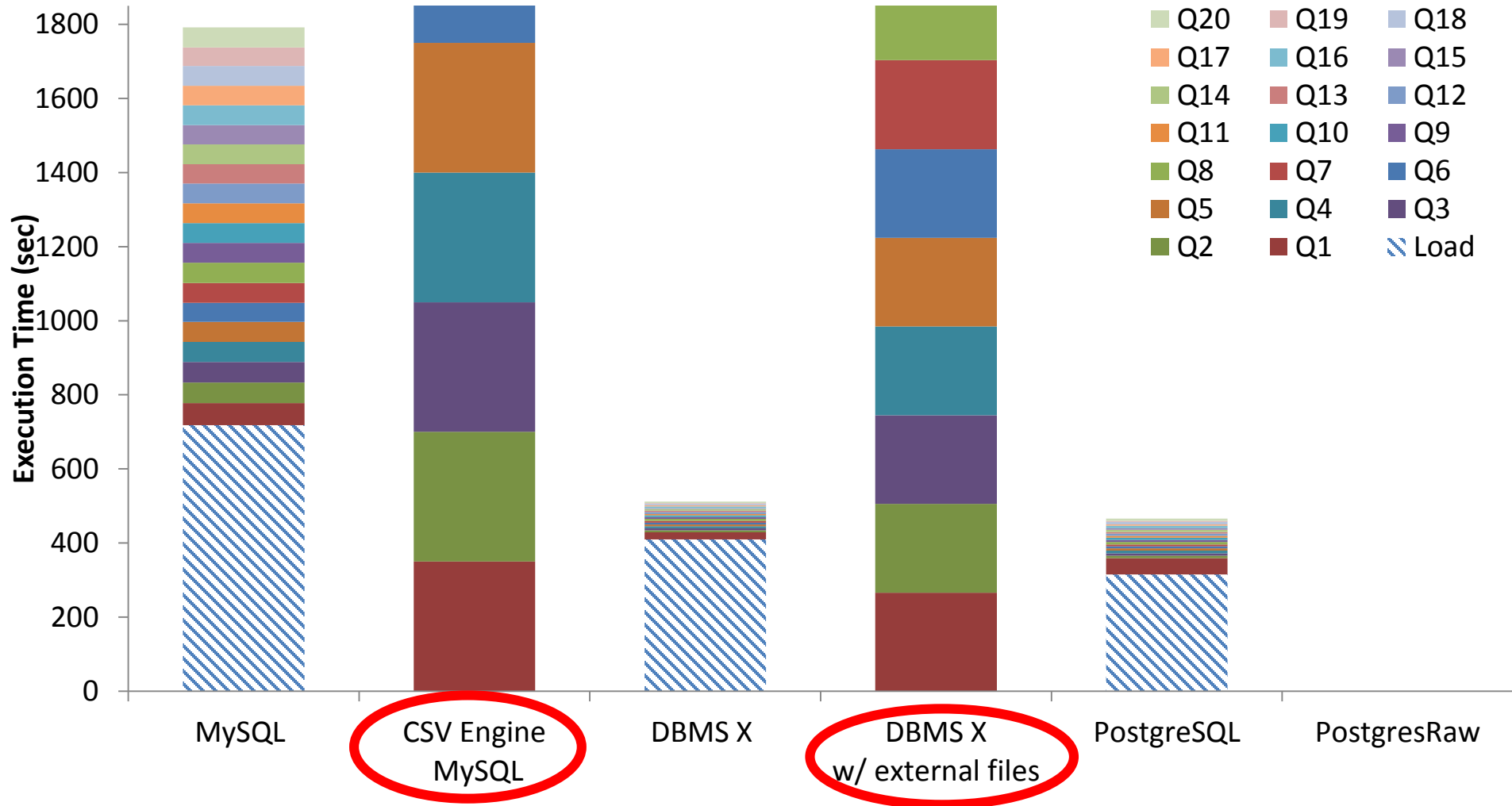


Graceful adaptation to workload changes

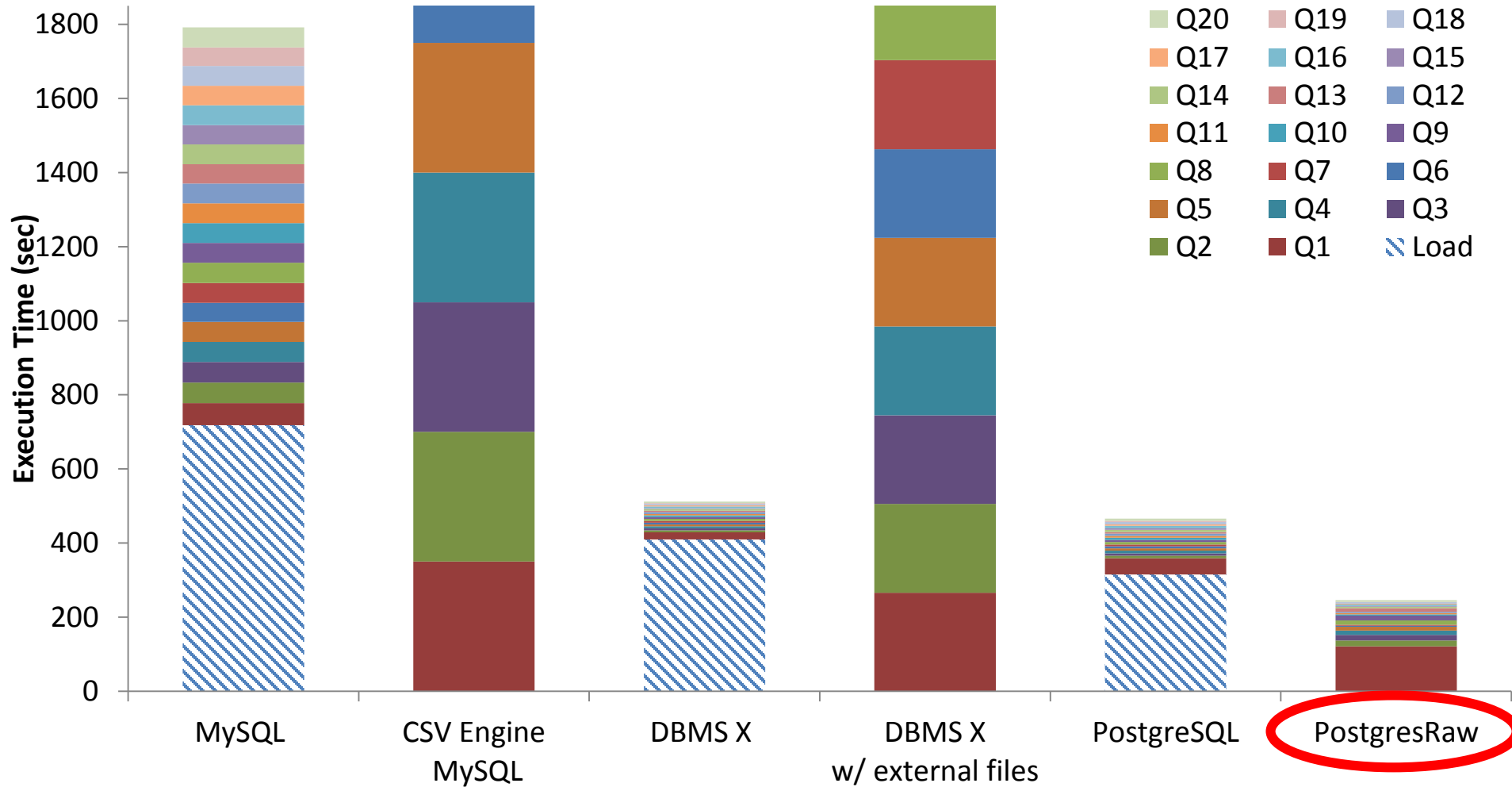
PostgresRaw vs. other DBMS



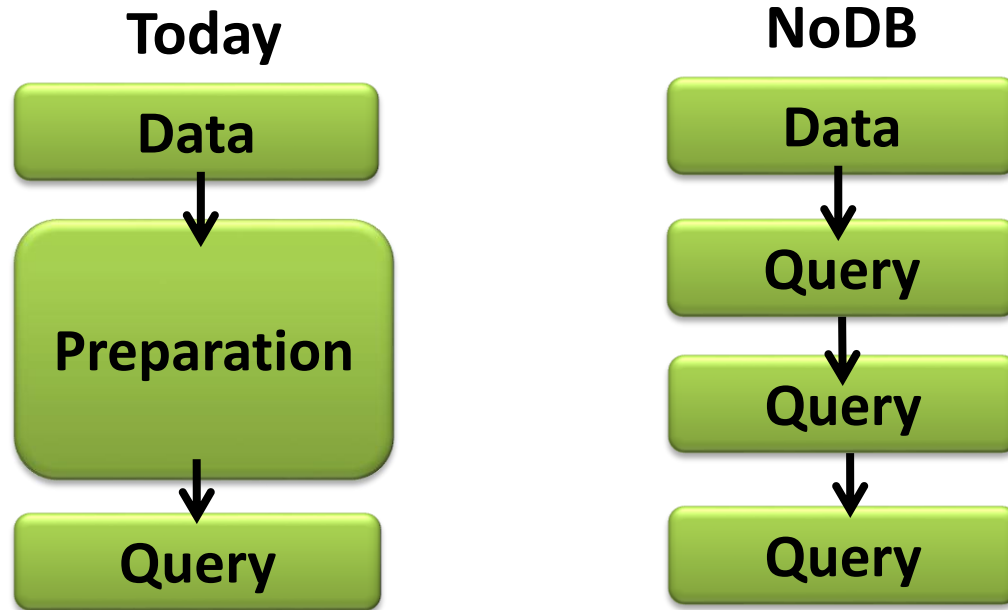
PostgresRaw vs. other DBMS



PostgresRaw vs. other DBMS



Competitive with conventional DBMS



Adaptive load - store - execute

Thank you!!