INFO20003 Database Systems

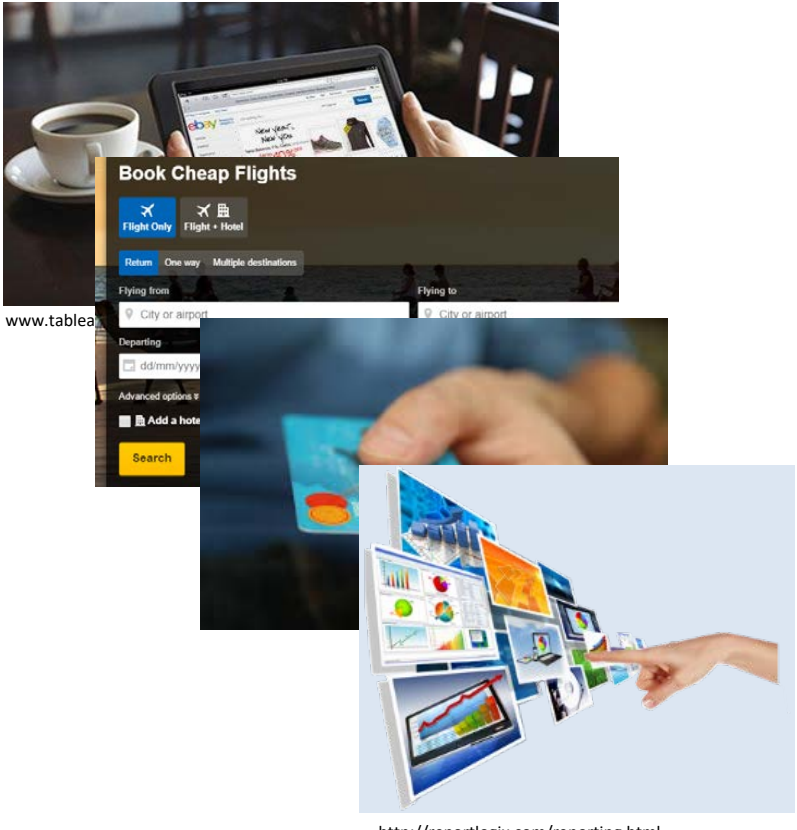
Dr Renata Borovica-Gajic

Lecture 22
Adaptive databases for the future
Introducing research avenues (non-examinable)

Week 11

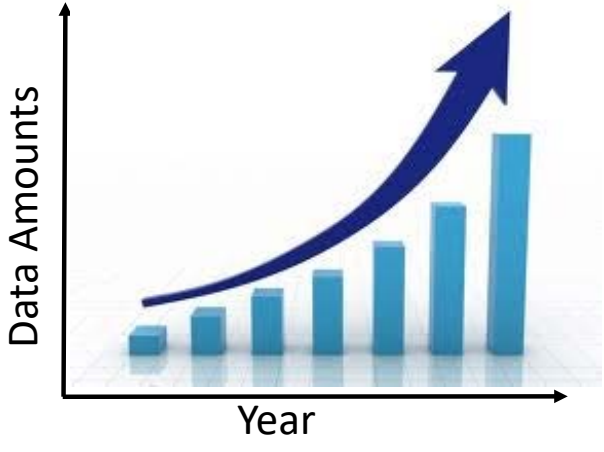
Data, data everywhere…

[The Economist]



www.tablea

http://reportlogix.com/reporting.html



50-fold from 2010-2020*

* "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", 2012, IDC

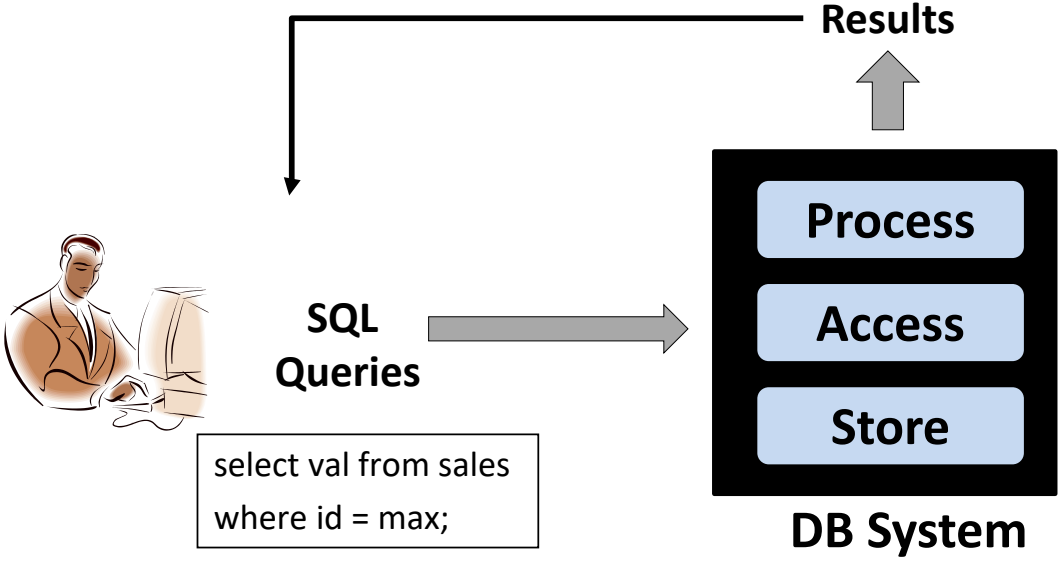
And grows exponentially…

Finding useful information

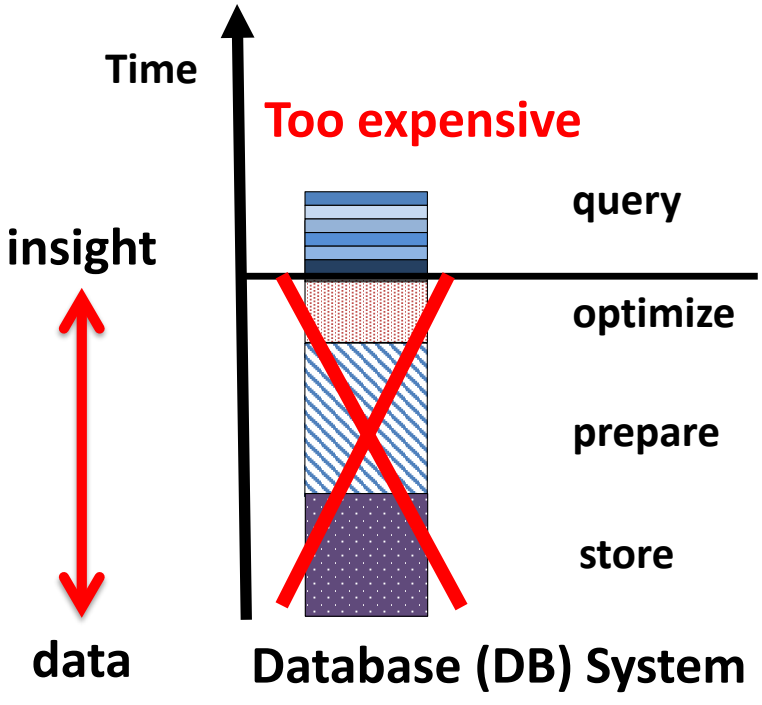
Equals to finding the needle in a haystack

Data analysis with databases

Database systems (DB):
4 decades of research,
predominant data analysis tool



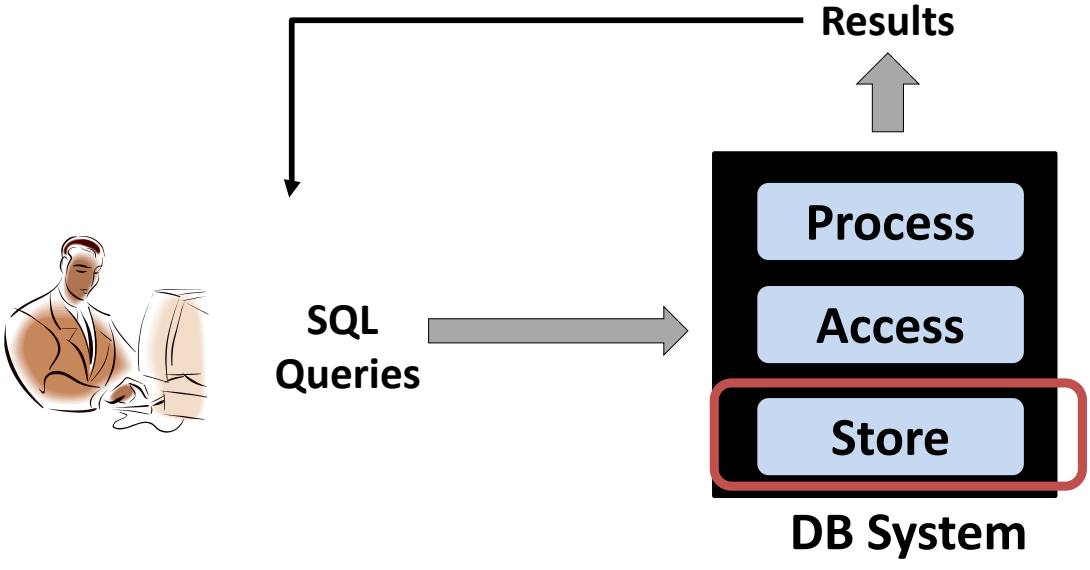
From data to knowledge



The luxury is long gone

Unless……

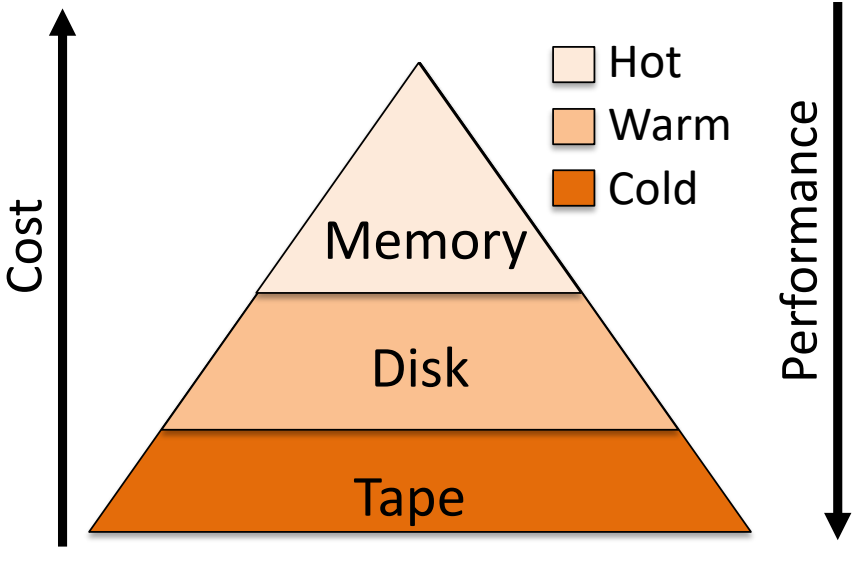
Data analysis with databases



Store data carefully

[VLDB'16, ADMS'17]

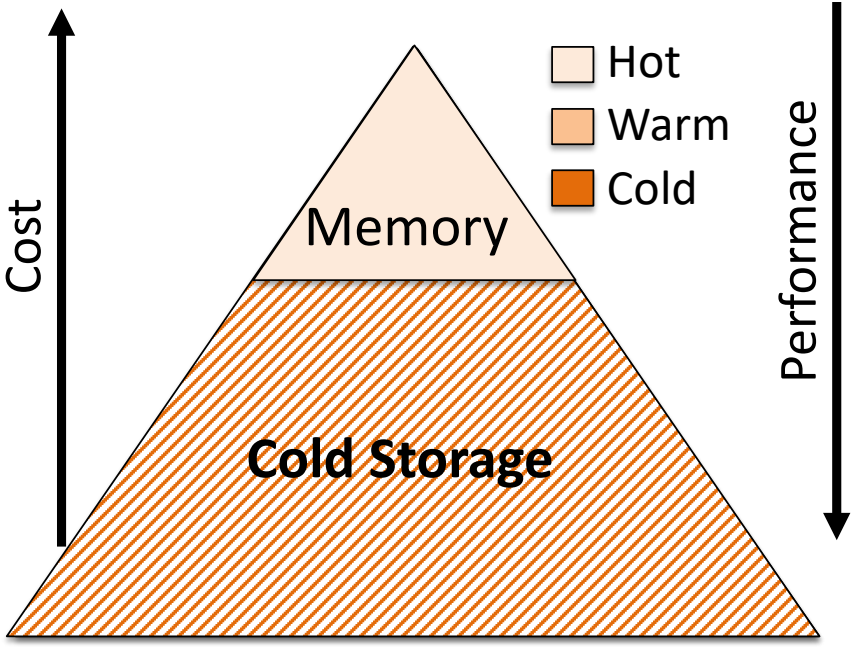
Storage tiering in private and public clouds



Store data carefully

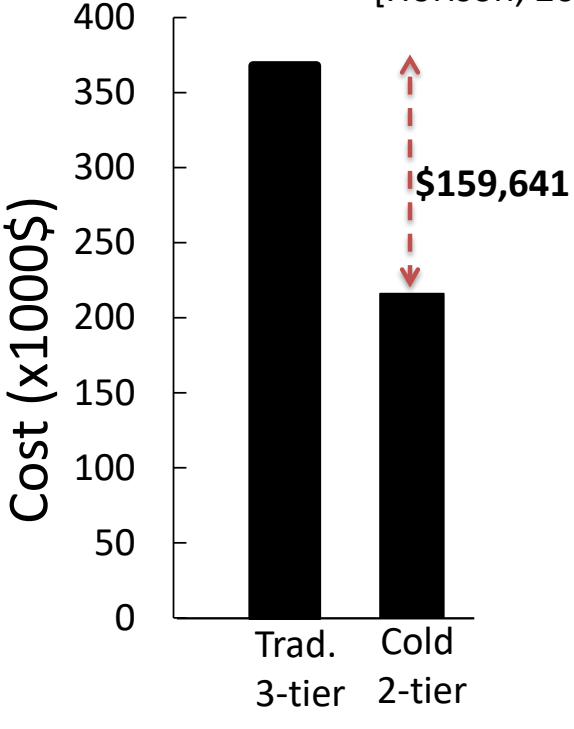
[VLDB'16, ADMS'17]

Storage tiering in private and public clouds



Storing 100TB of data

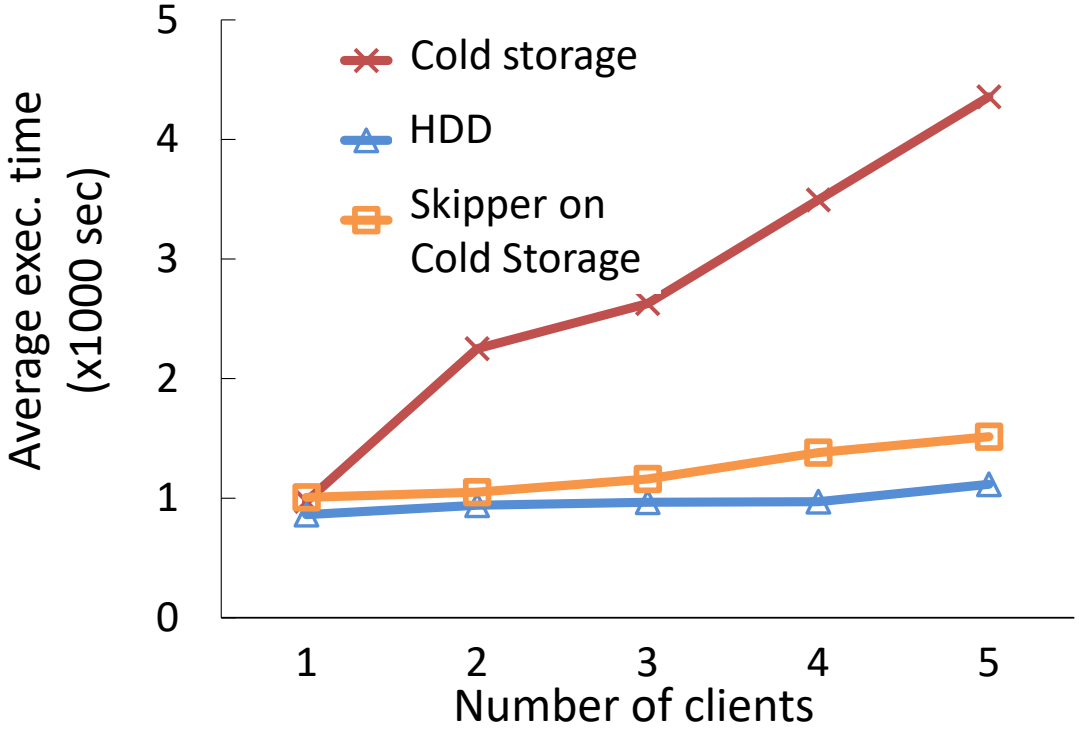
[Horison, 2015]



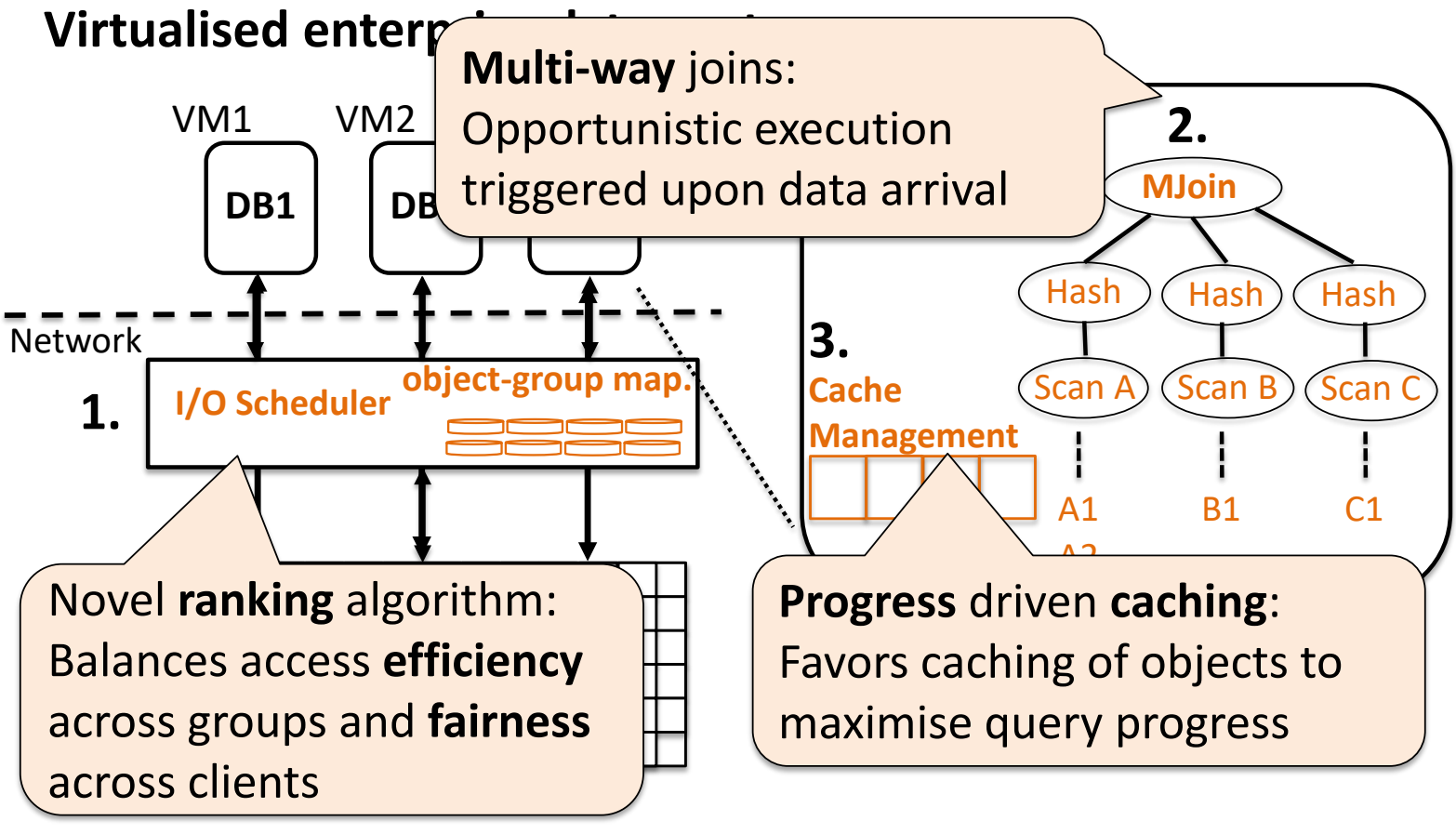
Embrace new technology

Cost benefit without performance penalty

Setting: multitenant enterprise datacenter, clients: TPCH 50, Q12, CSD: shared, layout: one client per group



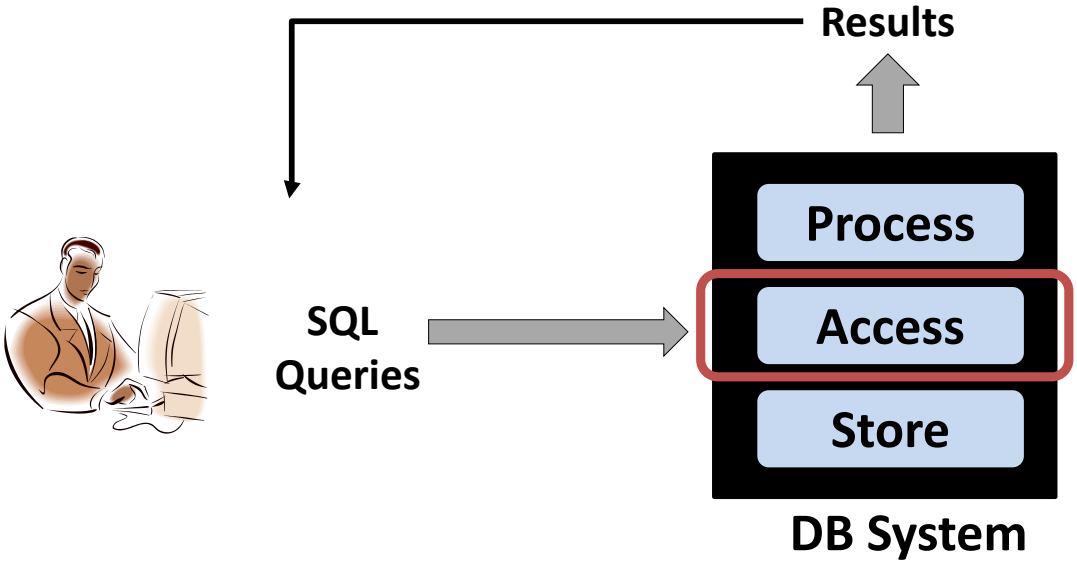
Skipper to the rescue



Lesson #1

Embrace new HW technology

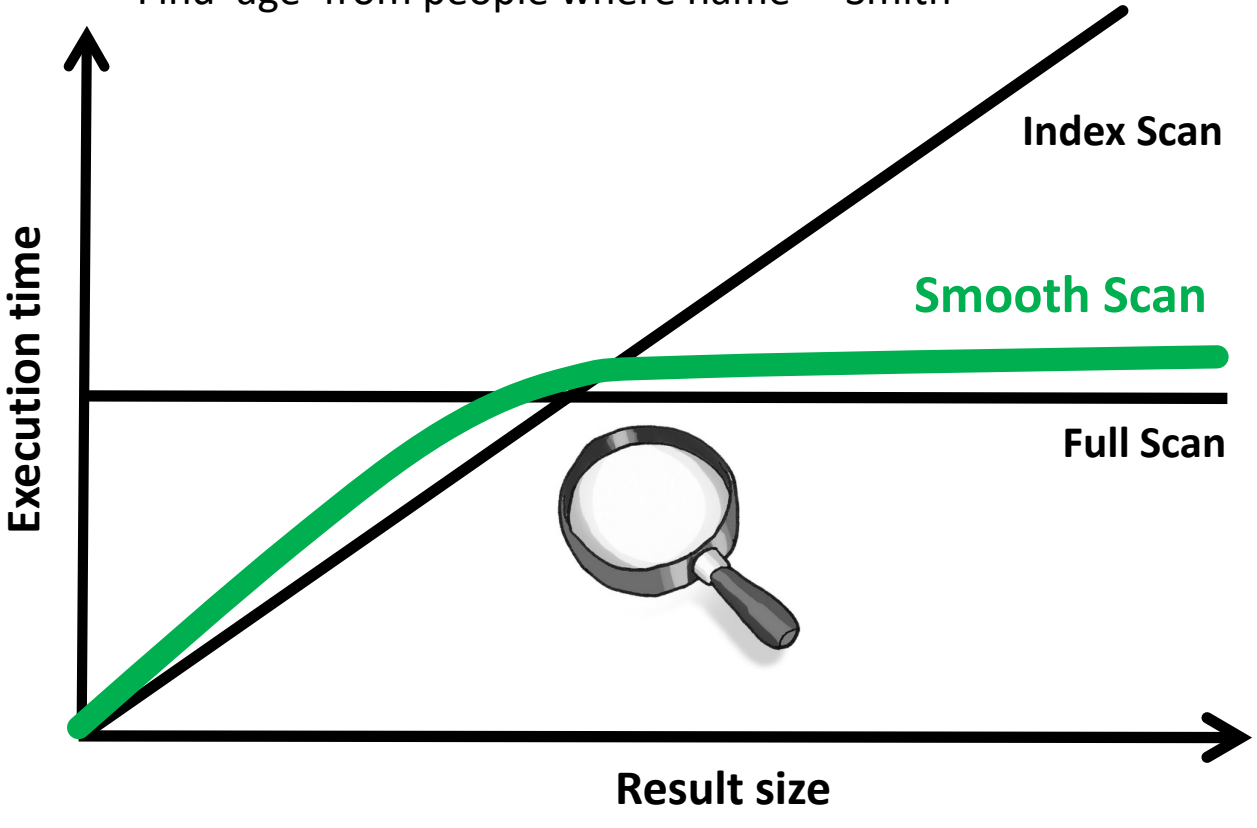
Data analysis



Choose access strategy on-the-fly

[DBTest'12, ICDE'15, VLDBJ'18]

Find 'age' from people where name = 'Smith'

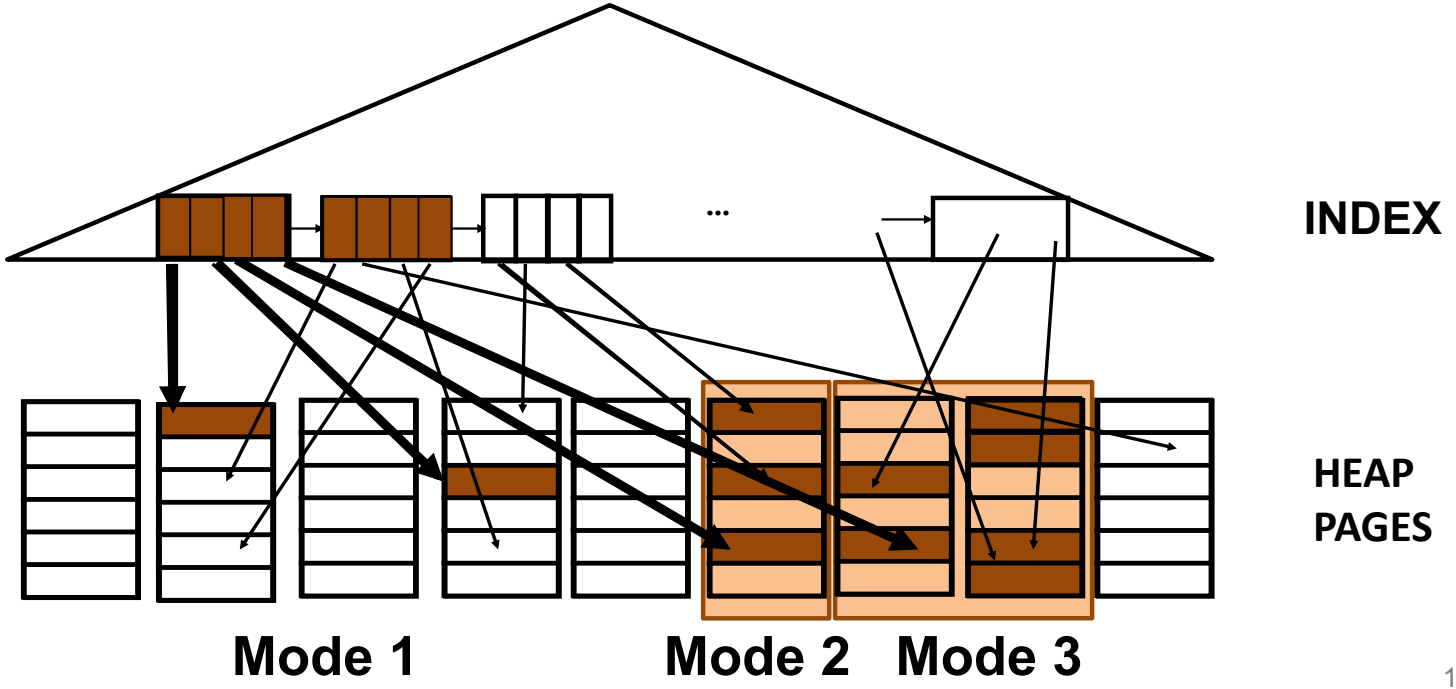


Adapt to data

Morphing mechanism

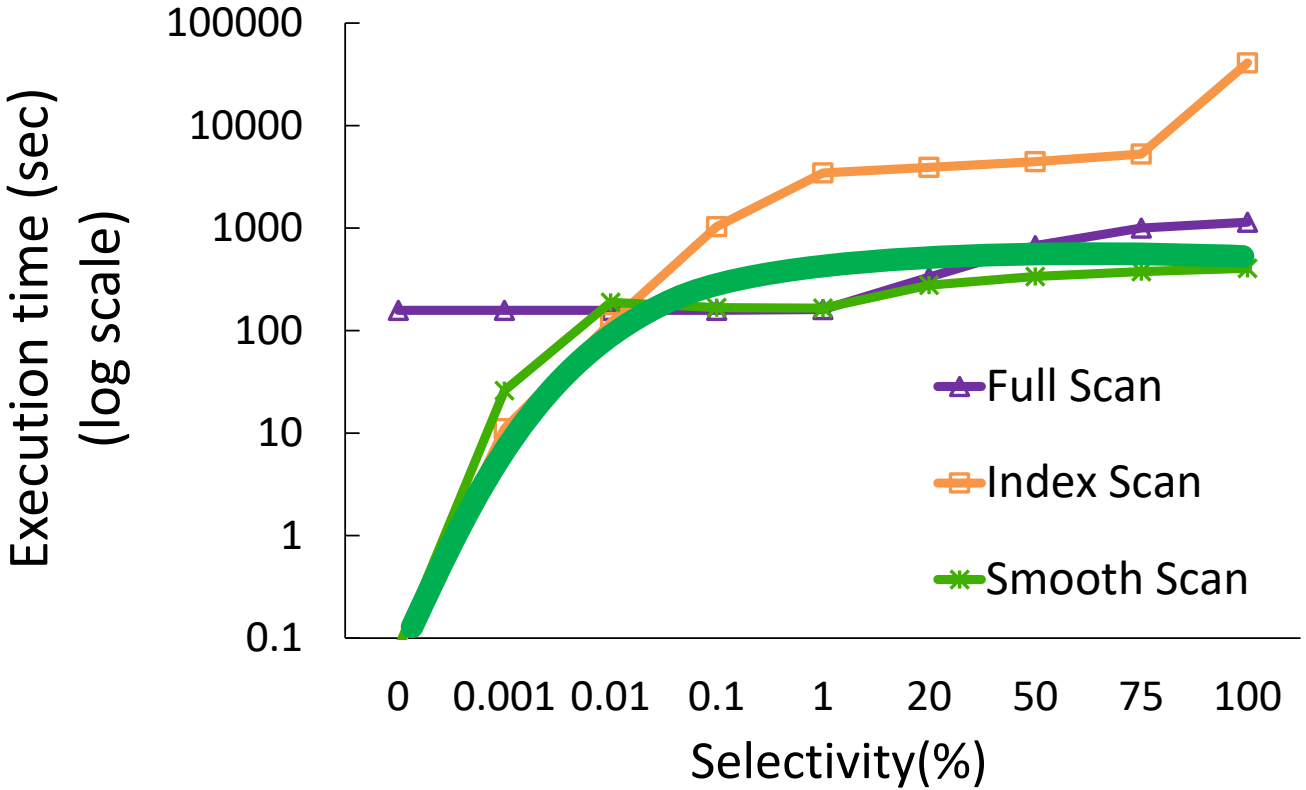
Modes:

1. **Index Access:** Traditional index access
2. **Entire Page Probe:** Index access probes entire page
3. **Gradual Flattening Access:** Probe adjacent region(s)



Smooth Scan in action

Setting: Micro-benchmark, 25GB table, Order by, Selectivity 0-100%

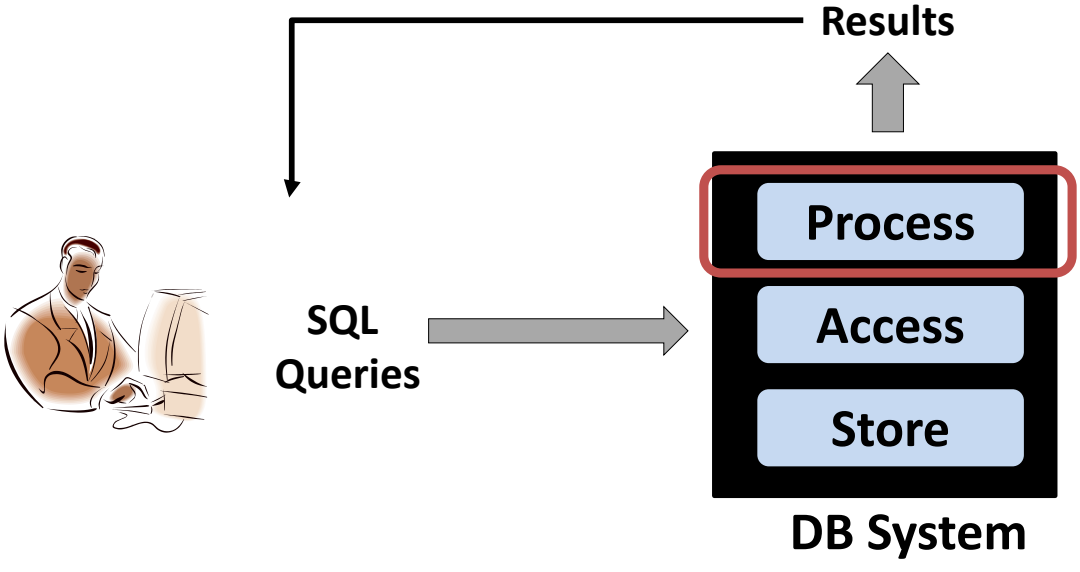


Near-optimal over entire selectivity range

Lesson #2

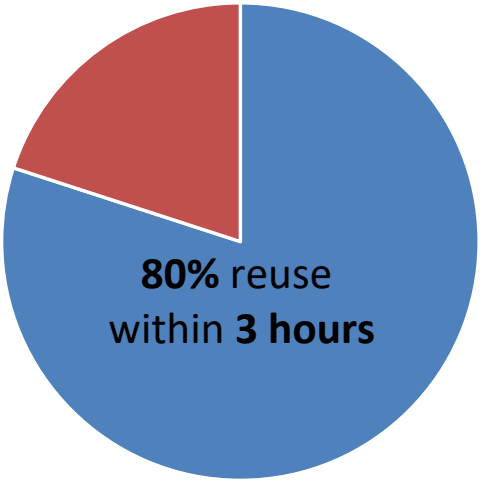
Learn from data

Data analysis with databases

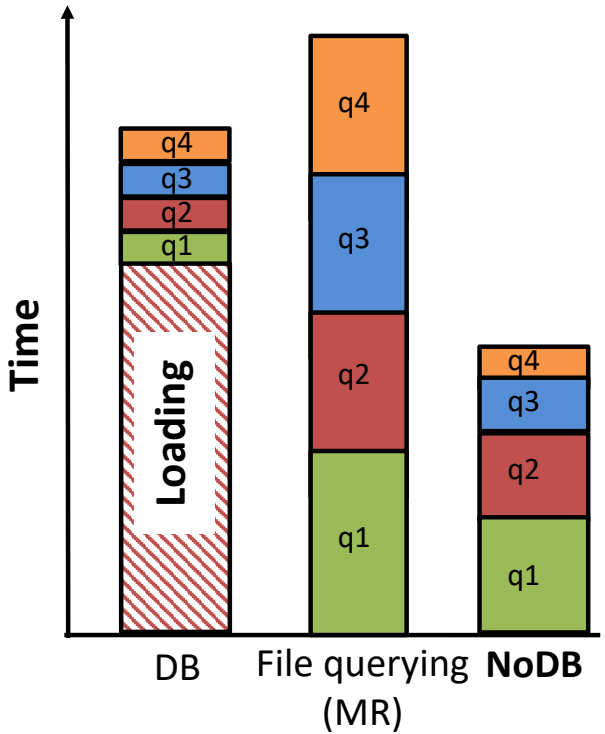


Process instantly

[SIGMOD'12, VLDB'12, CACM'15]

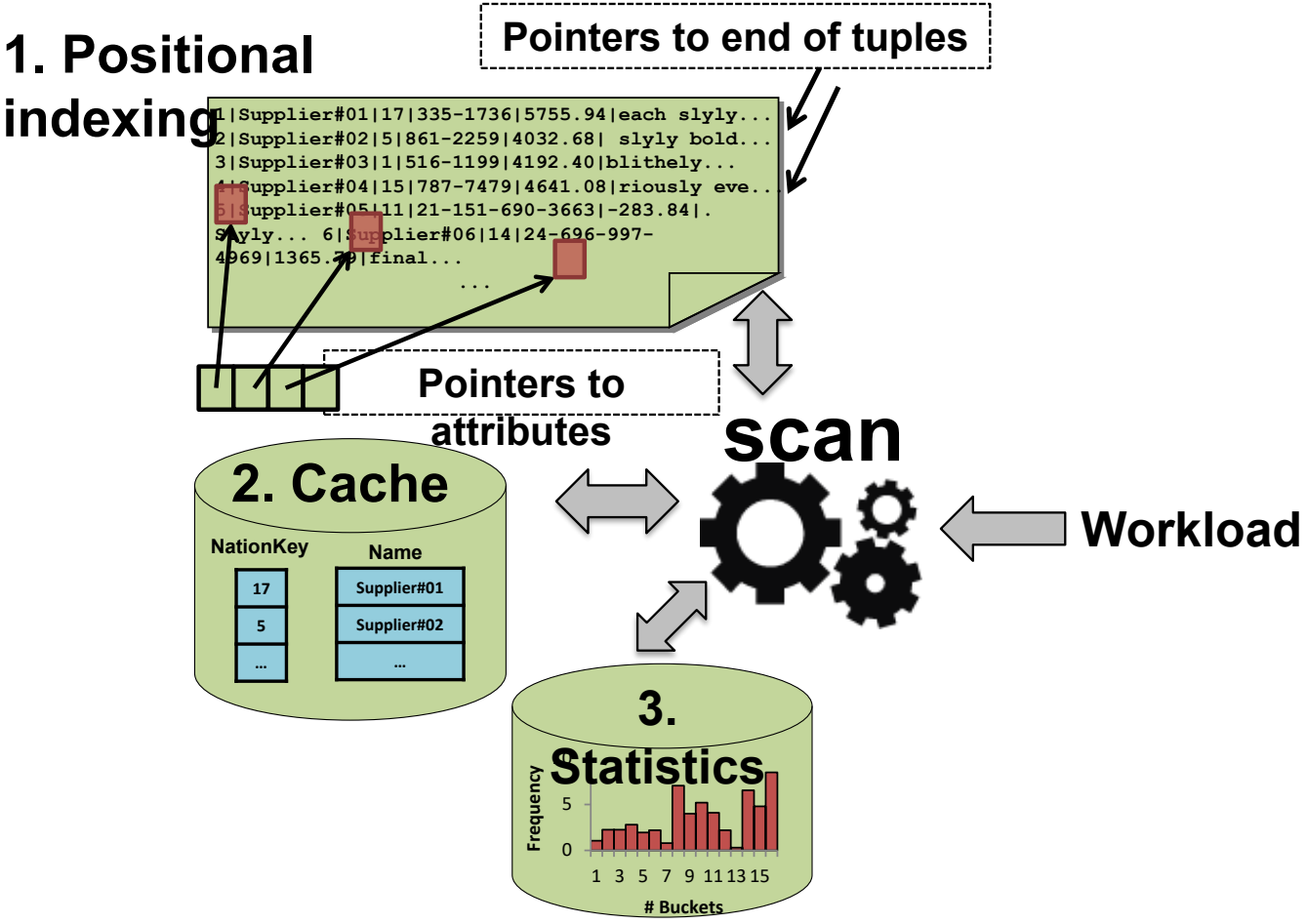


[Facebook]



Adapt to queries

PostgresRaw: NoDB from idea to practice

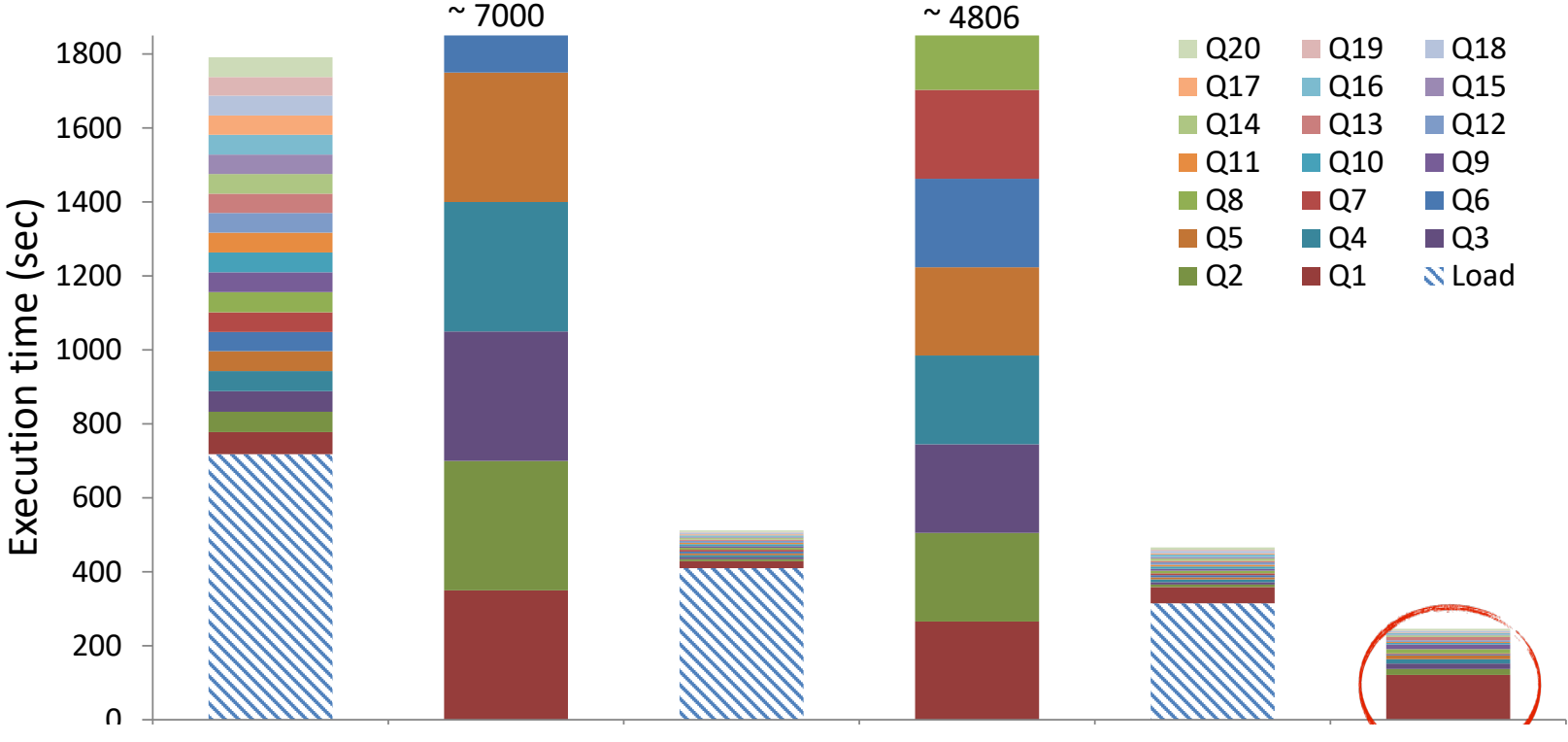


Adjust to queries = progressively cheaper

PostgresRaw in action

Setting: 7.5M tuples, 150 attributes, 11GB file

Queries: 10 arbitrary attributes per query, vary selectivity



Data-to-insight time halved with PostgresRaw

Per query performance comparable to traditional DBMS

Lesson #3

Learn from queries

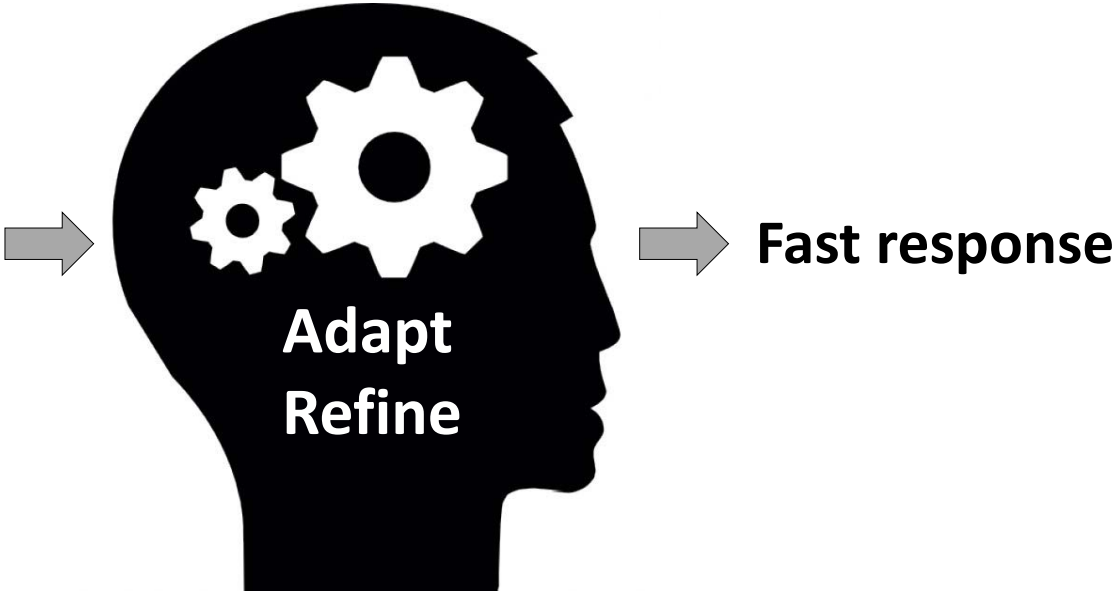
Self-designing systems for data analysis

"It is not the strongest species that survive, nor the most intelligent, but the ones most responsive to change." Charles Darwin

Queries
[SIGMOD'12]
[VLDB'12]
[CACM'15]

Data
[DBTest'12]
[ICDE'15]
[VLDBJ'18]

Hardware
[VLDB'16]
[ADMS'17]



DB System

Fast response

Thank you!

Anyone can be a data scientist with self-driving DB