

# CS222/CS122C: Principles of Data Management

UCI, Fall 2019  
Notes #14

Volcano Optimizer, Other Indexes,  
Open Topics, and Wrap up

Instructor: Chen Li

# *Volcano Optimizer*

## ❖ Volcano optimizer

- Unified search engine based on rules.
- Cost-based optimization.
- Flexible and Extensible. Do not have multiple stages because everything is a transformation rule.

## ❖ Impact:

- Another very popular optimizer besides System-R style.
- Also often called “Cascades” optimizer, an improvement to Volcano made by the same author Goetz Graefe.
- Widely used: Microsoft SQL Server, Google F1 Query, Greenplum Orca, Apache Calcite (used by Hive, Flink, ...)

# Rule Example

## ❖ Logical Equivalence Rules:

- Transform Logical Operators into another logically equivalent form.

- Join commutativity

- Join associativity

- Filter pushdown

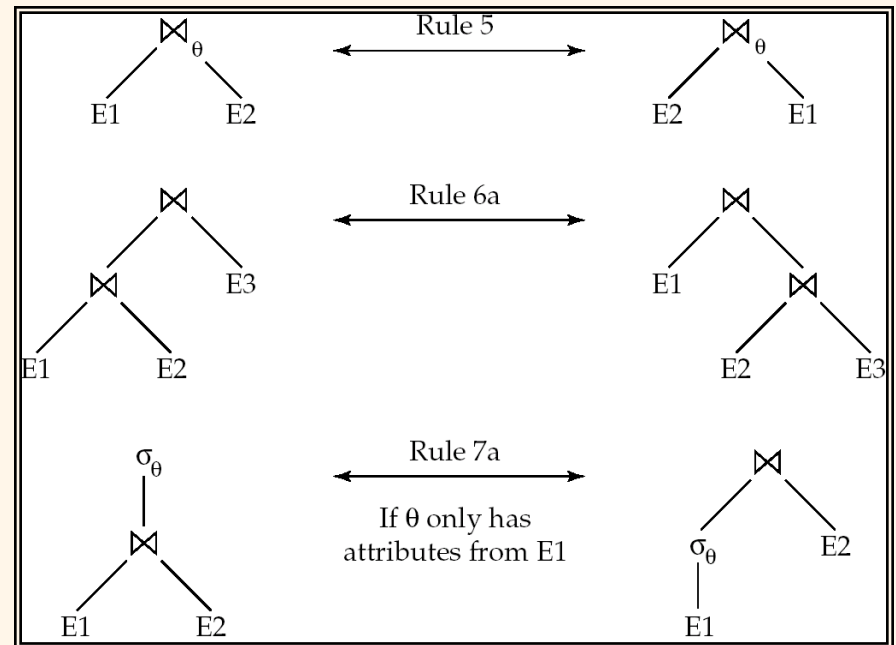
- Project pushdown

- Merge Projects

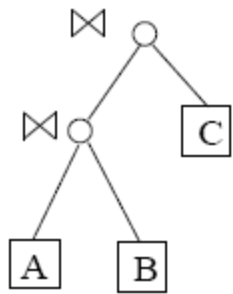
- Merge Union

- ...

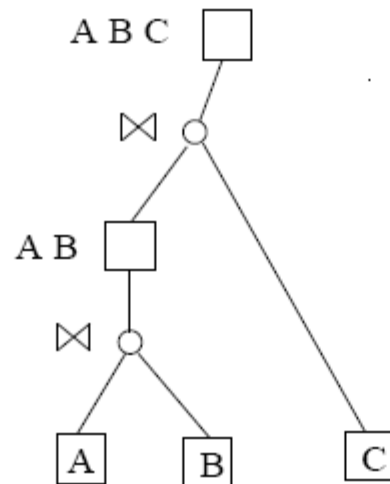
- Apache Calcite (open-source implementation of Volcano) currently has 100+ Logical Equivalence Rules!



# Memo (AND/OR tree)



(a) Initial Query

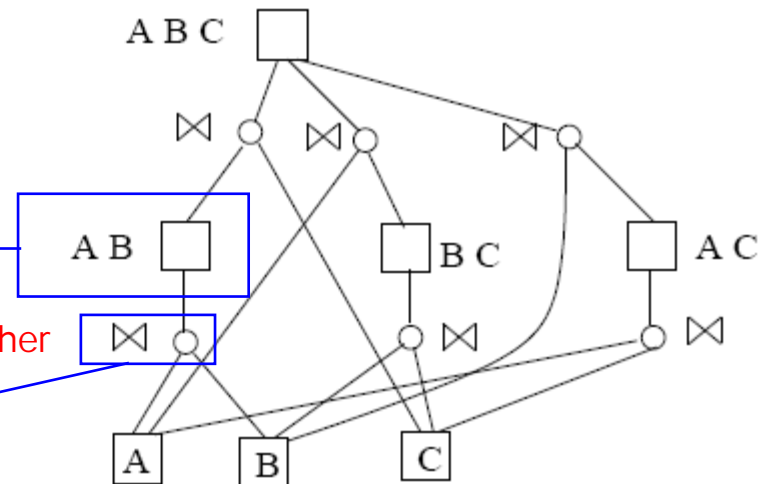


(b) DAG representation of query

logical equal:

A join B  
B join A

choose whether  
A join B  
or  
B join A



(c) Expanded DAG after transformations

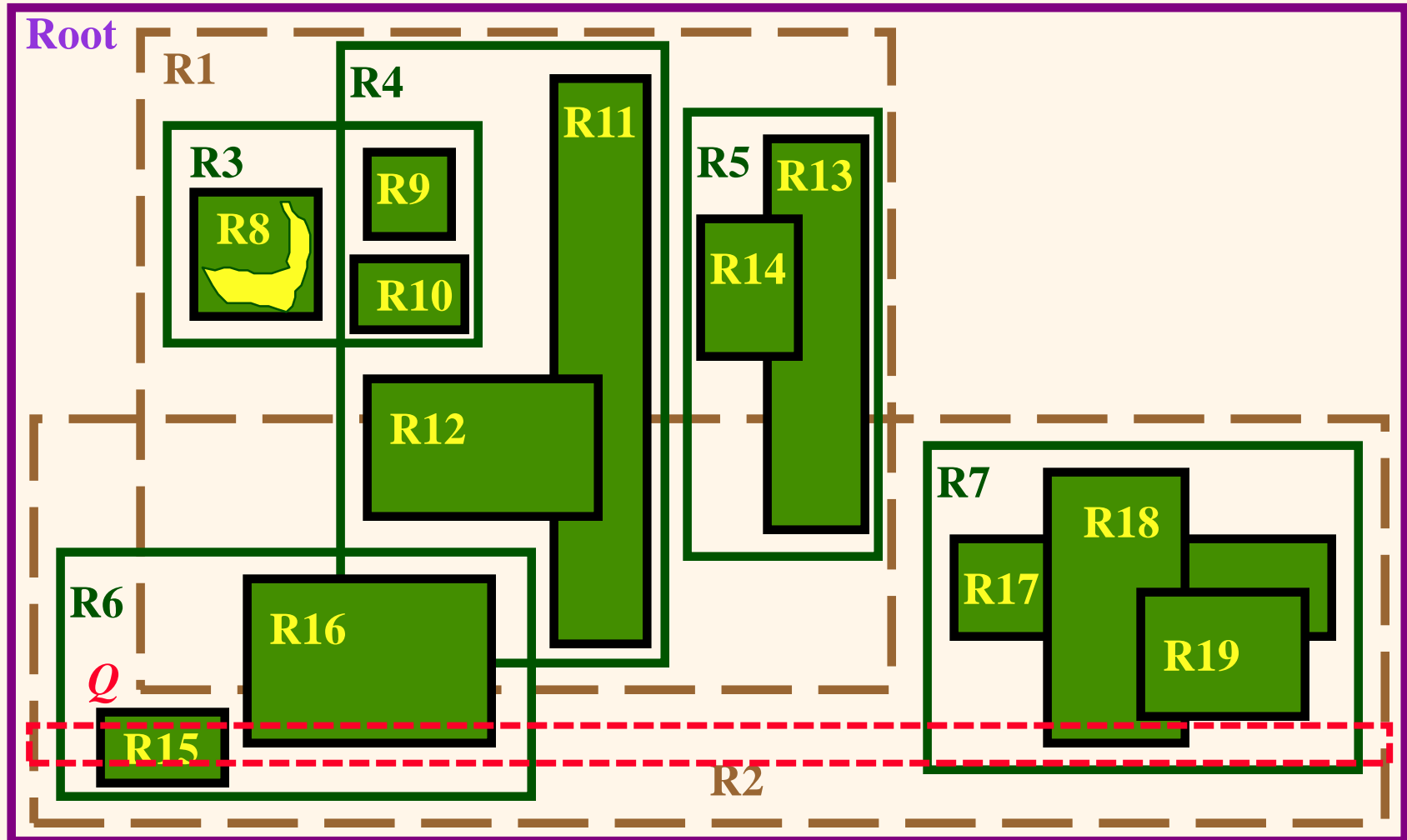
# *Volcano Optimizer summary*

- ❖ Use transformation rules to generate plan space and build logical and physical operators.
- ❖ Use AND/OR tree to compactly represent all alternative plans in search space.
- ❖ Use cost to evaluate a physical plan.
- ❖ Use Dynamic Programming to find the best plan.

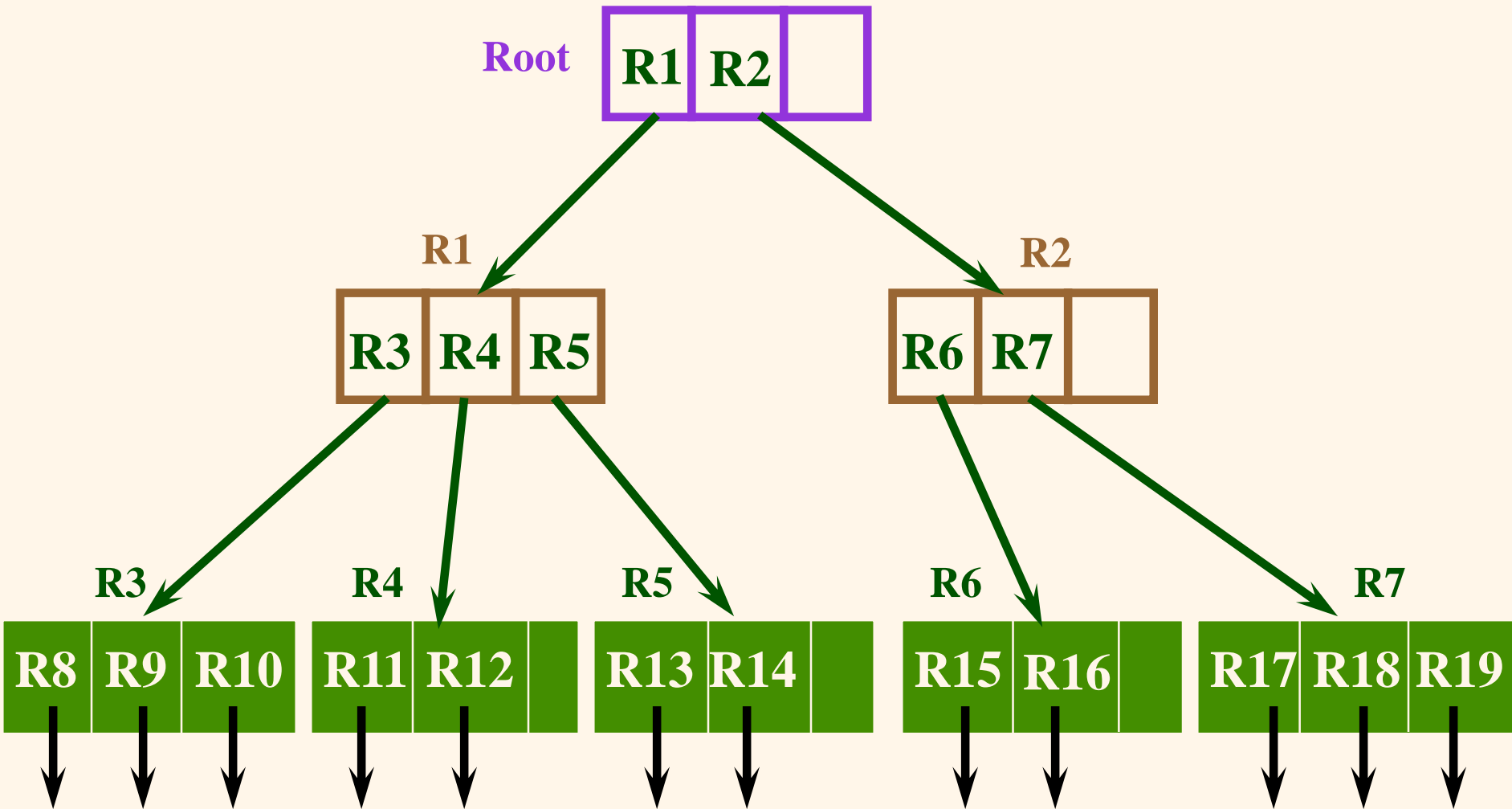
# Other indexing techniques

- ❖ Spatial: R-tree      R-Range
- ❖ Text: inverted index

# *R-Tree*

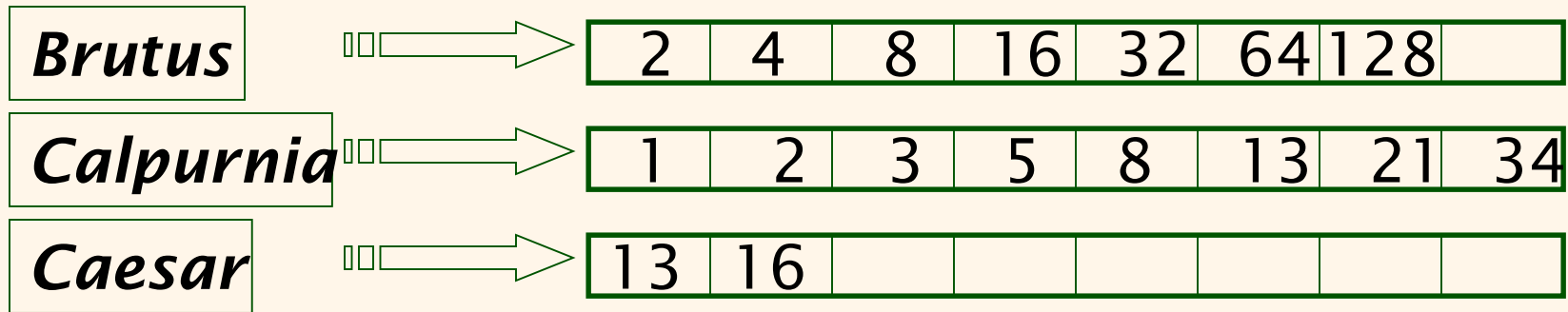


# *R-Tree*

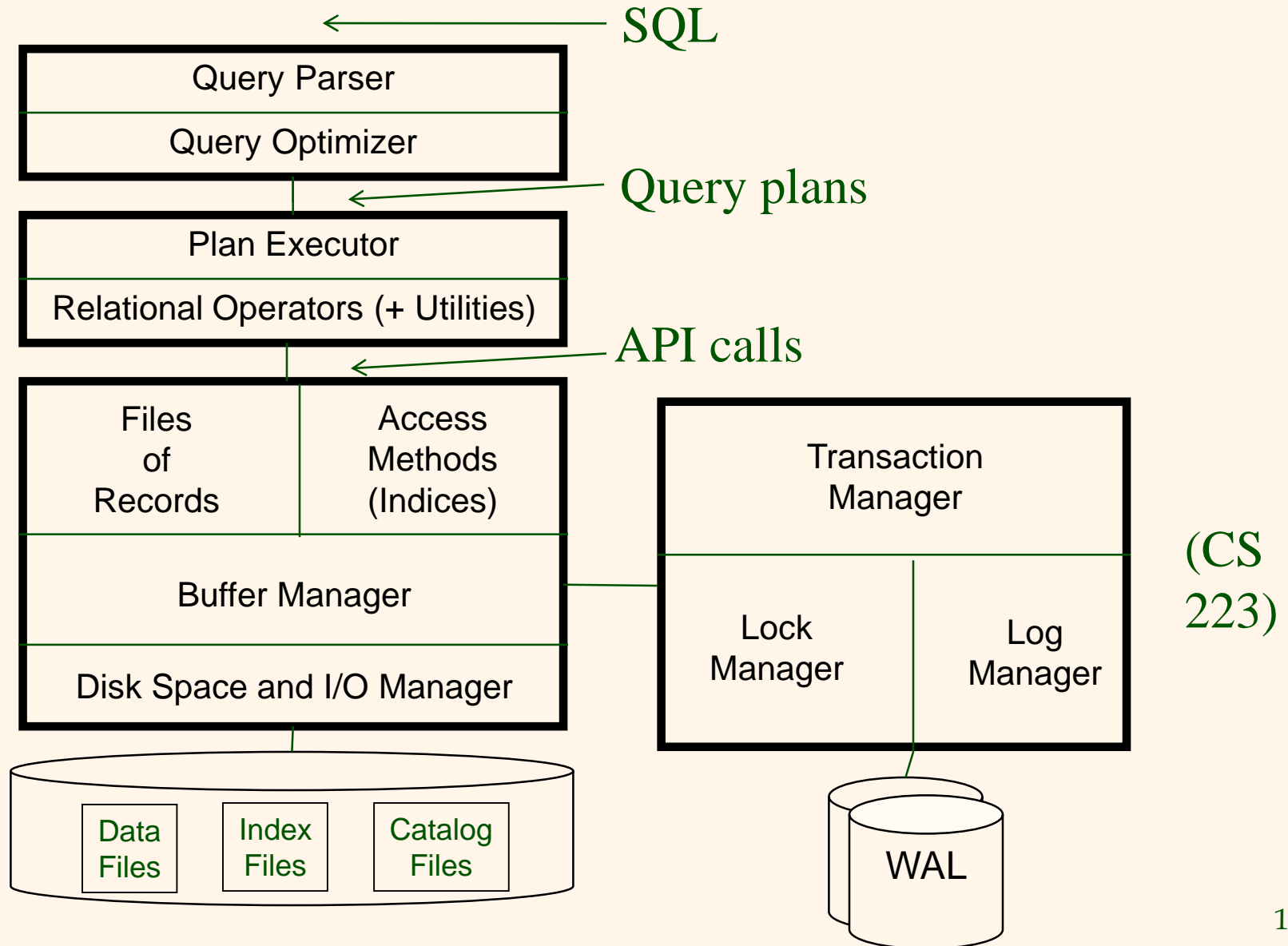




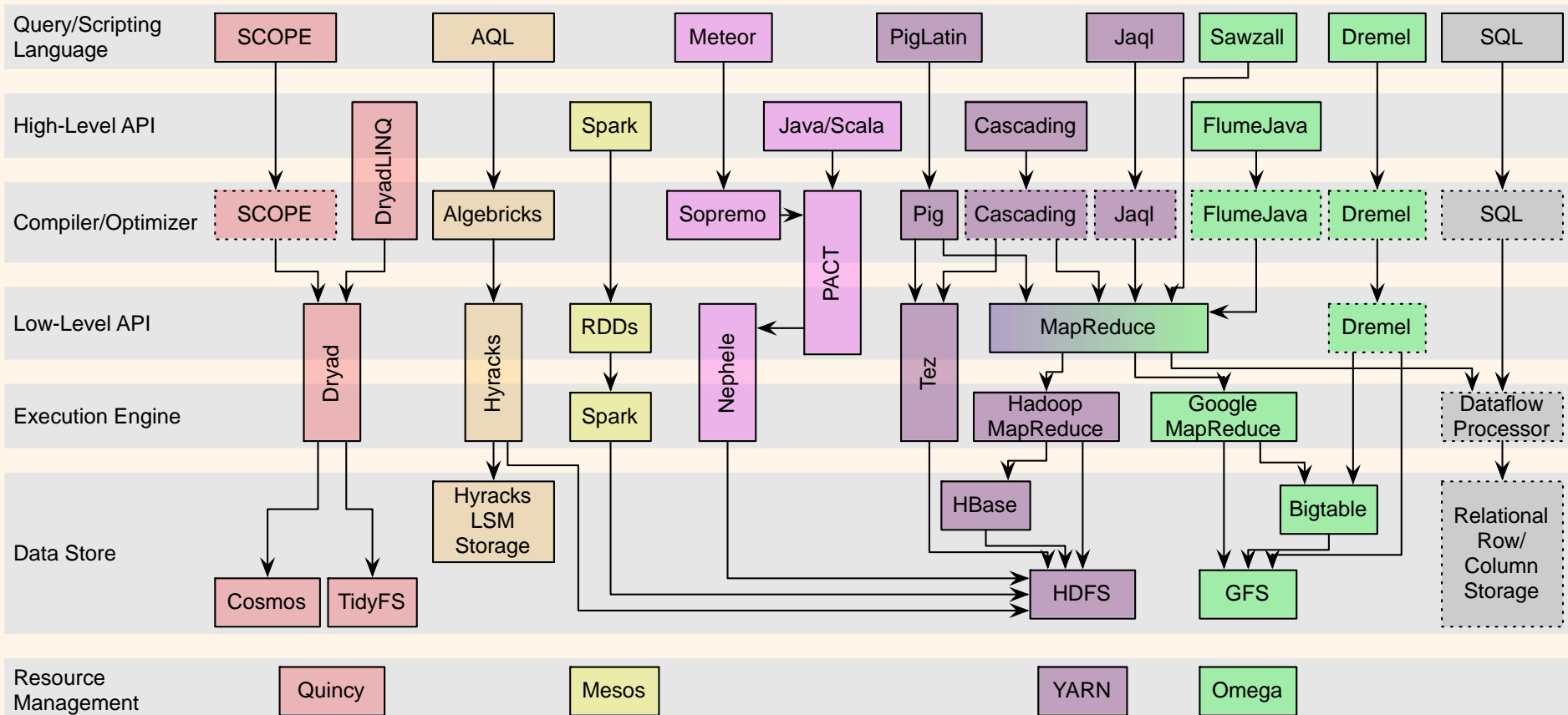
# *Text indexing: Inverted index*



# DBMS architecture revisited



# *“NoSQL” / Big Data*



**Map each system to components in a DBMS!**

# *So Where To From Here?*

- ❖ CS122D: Beyond SQL Database Management,  
<https://grape.ics.uci.edu/wiki/asterix/wiki/cs190-2020-winter>
- ❖ CS 223: Transaction Processing and Distributed Data Management
- ❖ Research projects
  - Apache AsterixDB: parallel DBMS for semi-structured data
  - Texera: Data analytics using GUI-based workflows
  - Cloudberry: Big data visualizations