Grading

- 2 written individual assignments (homeworks) - 20%
- 4 practical assignments - 30%
- 2h written exam, covering all taught content in the course - 50%

SQL – Structured Query Language

HW1: SQL

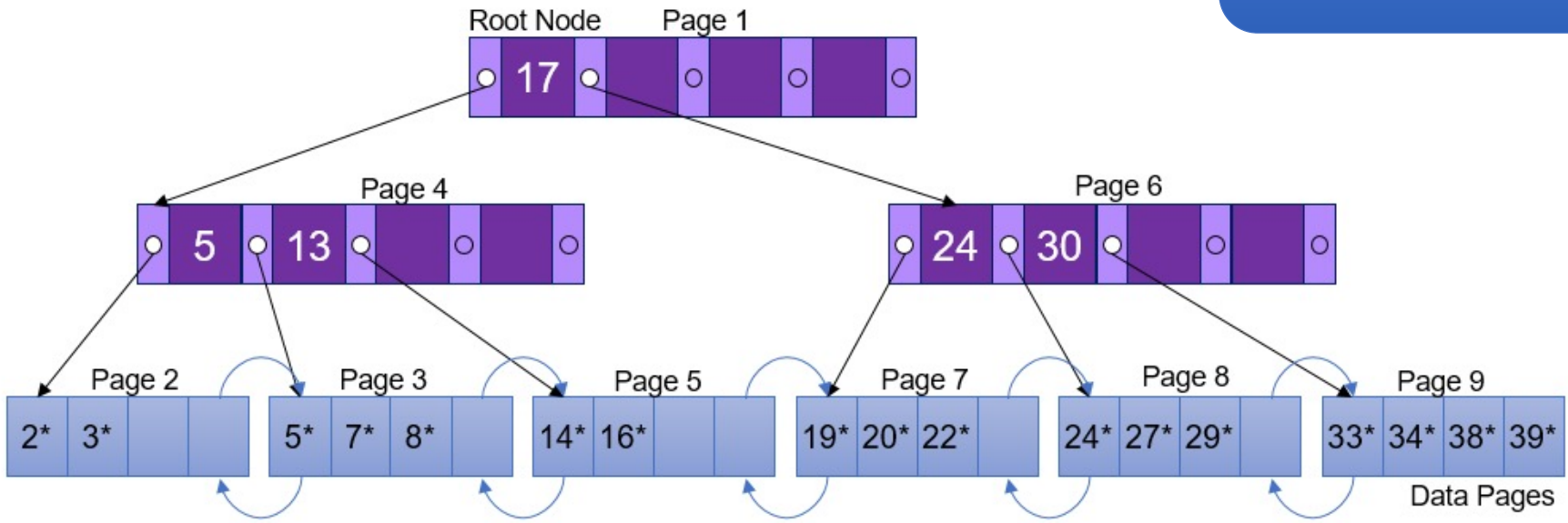
Language for computing on relations

- that separates the WHAT from the HOW
- enables the system to choose the best how given the *data* and its *layout*

HW1 posted today, deadline September 18, 23:59.

Indexing stored data

HW2:
Indexing & RA



Docker for practical assignments

- We’ll use a « virtualization » layer (Docker)
- Docker containers of DBMSs
  - Oracle
  - Cassandra
  - Cloudera Quickstart - Hive, SparkSQL, DataFrames, Impala)

  More details on Docker and setup will be posted soon.

Bridging the WHAT with the HOW

- Query optimization!
- Three stages
  - Plan space
  - Cost estimation
  - Search algorithm

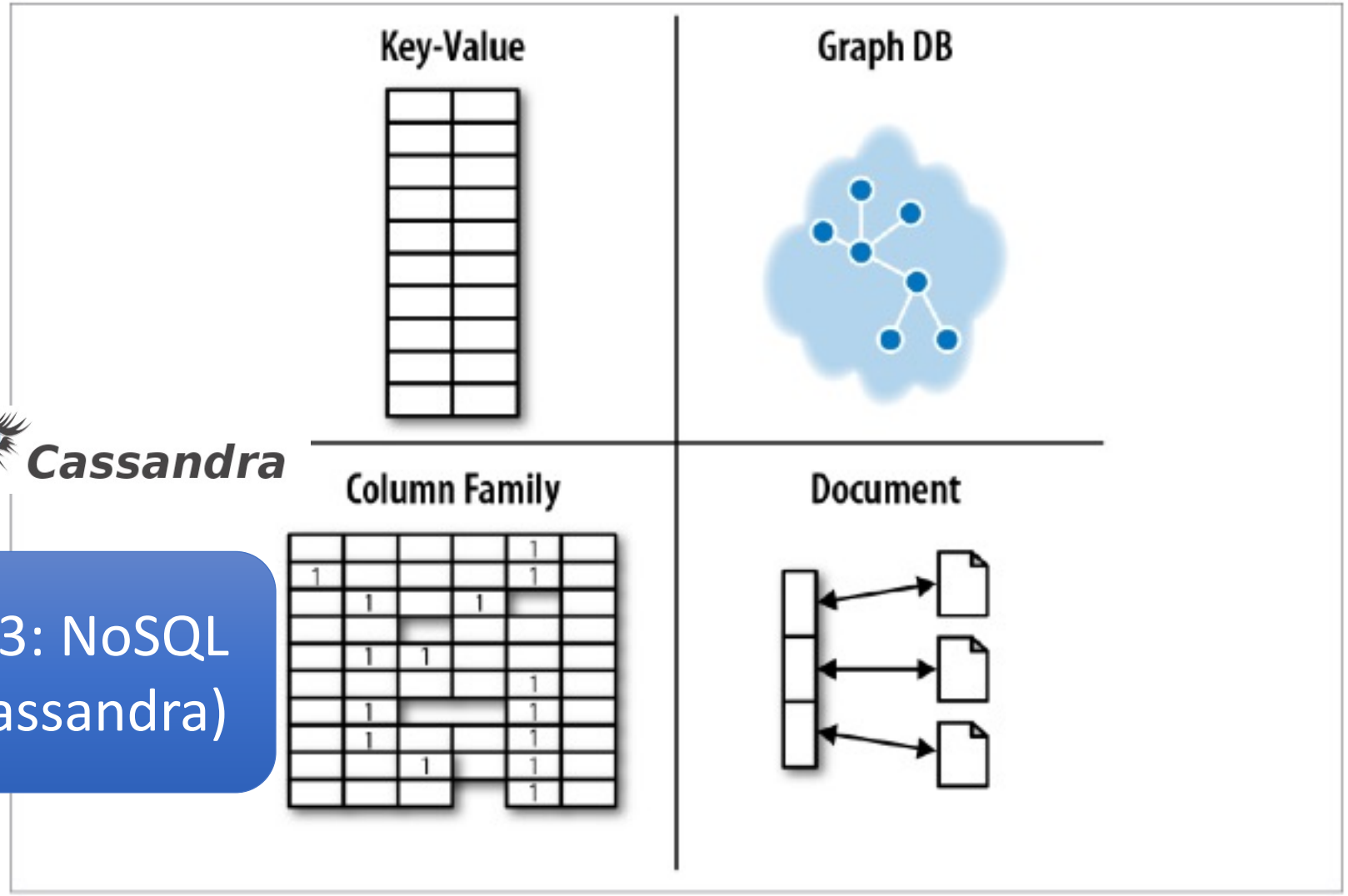
PA1: Query optimisation (Oracle)

Transactions and concurrency control

- Correct execution: serially-ordered
- Desire: interleave to maximize performance
- Risk: Disorder leads to data anomalies
- Allowable orders: (conflict) serializability
- Implementation: (strict)2PL

PA2: Transactions (Oracle)

Review of major types of NoSQL databases…



PA3: NoSQL (Cassandra)

Structured data management / SQL on HADOOP-SPARK

BDMS: deal with same class of analytical workloads as parallel systems, but different architectures, design decisions, implementations, impacting

- Query optimization
- Query execution
  - E.g., MapReduce algorithms for processing relational data
- Fault tolerance

PA4: SparkSQL, DataFrames, Hive, Impala (Cloudera)

Exam

- Open-book
- 25% questions with multiple choices (quizz)
- 75% short questions / answers
- Previous exams (with solutions) posted on Moodle

Feedback from teacher

- Homeworks and practical assigment work will be graded within 3 weeks from the submission deadline, and corrections will be posted online shortly after the submission deadline.
- After grading, overall class performance indicators will also be provided to students.
- Within 2 weeks of the publication of final course results, group feedback on overall student performance will be provided to students, including grade distribution.